prof. dr hab. inż. Adam Grzech
Instytut Informatyki Technicznej
Politechnika Wrocławska
Wybrzeże Wyspiańskiego 27
50-370 Wrocław

Wrocław, 8 kwietnia 2013 roku

Recenzja rozprawy habilitacyjnej i ocena dorobku naukowego dr inż. Krzysztof Zatwarnickiego.

Recenzja rozprawy habilitacyjnej i ocena dorobku naukowego dr inż. Krzysztofa Zatwarnickiego została przygotowana w odpowiedzi na pismo prof. dr hab. inż. Antoniego Wilińskiego, Dziekana Wydziału Informatyki Zachodniopomorskiego Uniwersytetu Technologicznego w Szczecinie.

Recenzja i ocena zostały opracowane na podstawie przesłanych: rozprawy habilitacyjnej pt. „Systemy webowe z jakością usług uwzględniające kryterium czasowe. Problemy projektowania" oraz opracowania pt. *Zestawienie dorobku naukowego, dydaktycznego i organizacyjnego dr inż. Krzysztofa Zatwarnickiego.*

I. Recenzja rozprawy habilitacyjnej „Systemy webowe z jakością usług uwzględniające kryterium czasowe. Problemy projektowania"

I.1. Ogólna charakterystyka rozprawy

Rozprawa habilitacyjna dr inż. Krzysztofa Zatwarnickiego pt. *Systemy webowe z jakością usług uwzględniające kryterium czasowe. Problemy projektowania* została wydana przez Oficynę Wydawniczą Politechniki Opolskiej w 2012 roku.

Rozprawa, licząca 172 strony, składa się ze wstępu, trzech rozdziałów dotyczących modelu systemu obsługi żądań HTTP (*Hypertext Transfer Protocol*), systemów minimalizujących czasy odpowiedzi na żądania i systemów obsługujących żądania z zadanym czasem odpowiedzi, podsumowania (wnioski końcowe), spisu literatury przedmiotu przywoływanej w rozprawie oraz streszczeń rozprawy w języku polskim i angielskim.

W spisie literatury znajduje się 228 pozycji. Wśród cytowanych źródeł jest 49 prac Autora rozprawy (autorskich i współautorskich). Cytowane prace są reprezentatywne dla bardzo bogatego piśmiennictwa dotyczącego modelowania, projektowania i analizy systemów webowych, w tym metod optymalizacji obsługi żądań w takich systemach.

Recenzowana rozprawa jest podsumowaniem wieloletnich prac naukowo-badawczych i technicznych Autora z zakresu projektowania i analizy systemów zarządzania rozdziałem żądań HTTP w systemach webowych, których pierwszym etapem była – obroniona w 3002 roku - rozprawa doktorska pt. *Adaptacyjny algorytm dystrybucji żądań w klastrze serwerów www.*

I.2. Charakterystyka badań naukowych relacjonowanych w rozprawie habilitacyjnej

Opiniowana monografia pt. *Systemy webowe z jakością usług uwzględniające kryterium czasowe. Problemy projektowania* dotyczy modelowania, projektowania, analizy i badań symulacyjnych systemów zarządzania żądaniami w systemach webowych.

Przedmiotem rozprawy są systemy zarządzania rozdziałem żądań w jednoserwerowych i wieloserwerowych (scentralizowanych i rozproszonych) systemach webowych dla potrzeb realizacji dwóch zadań: minimalizacji średniego czasu obsługi żądań oraz minimalizacji średniego czasu obsługi przy jednoczesnym gwarantowaniu czasu obsługi nie dłuższego niż zadany, maksymalny czas obsługi żądania.

Deklarowany cel monografii to opracowanie odpowiednich algorytmów szeregowania i rozdziału żądań oraz różnych struktur decyzyjnych, w których algorytmy rozdziału i szeregowania żądań mogą być użyte do rozdziału żądań w jednoserwerowych oraz wieloserwerowych systemach webowych.

Podstawą proponowanych i opisanych w monografii algorytmów rozdziału i szeregowania żądań jest autorski, adaptacyjny model opóźnienia (czasu odpowiedzi) w systemie webowym oraz

znana z literatury przedmiotu metoda szeregowania EDF (*Earliest Deadline First*). Adaptacyjny model opóźnienia jest wykorzystywany do szacowania czasu odpowiedzi na żądania i podejmowania decyzji o lokalizacji serwera (lub klastra serwerów, a następnie serwera w klastrze) do którego skierowane zostanie żądanie w celu wykonania żądania. Algorytm szeregowania żądań jest wykorzystywany do wyznaczania kolejności obsługi żądań innej niż kolejność napływu żądań do systemu tak, aby żądania wymagające dłuższego czasu obsługi mogły być obsługiwane kosztem opóźniania obsługi żądań nie wymagających krótkiego czasu obsługi.

Opracowany algorytm szacowania czasu odpowiedzi (odpowiednio modyfikowany) został wykorzystany do opracowania propozycji jednopoziomowego systemu rozdziału żądań pomiędzy serwerami jednego klastra oraz dwupoziomowego systemu rozdziału żądań pomiędzy klastrami serwerów, a następnie serwerami wcześniej wybranego klastra. Zaproponowane rozwiązania dotyczą systemów webowych, w których celem systemu zarządzania jest minimalizacja średniego czasu obsługi żądań.

Wspomniane algorytmy szeregowania żądań i szacowania czasów obsługi żądań (odpowiednio modyfikowane) zostały wykorzystane do opracowania propozycji zarządzania kolejnością realizacji żądań w systemie jednoserwerowym, zarządzania kolejnością obsługi żądań przez klaster serwerów i rozdziału żądań pomiędzy serwerami klastra serwerów oraz do wyznaczania klastrów serwerów, kolejności obsługi przez klaster i serwera w klastrze. Zaproponowane rozwiązania dotyczą systemów webowych, w których czasy odpowiedzi są gwarantowane, a zarządzanie żądaniami może zwiększyć wartość odpowiednio zdefiniowanego wskaźnika jakości.

Proponowane w monografii rozwiązania zostały wyprowadzone z krytycznej analizy sposobów zarządzania żądaniami we współczesnych systemach webowych. Uzasadnienie potrzeby poszukiwania i implementacji nowych metod zarządzania żądaniami w systemach webowych jest przekonywujące i zbudowane na podstawie wyczerpującej analizy literatury przedmiotu, bardzo dobrej znajomości zagadnień związanych z systemami webowymi i siecią Internet oraz dużej wiedzy i doświadczeniu Autora z zakresu technicznych uwarunkowań realizacji rzeczywistych systemów webowych.

Opracowane i relacjonowane w rozprawie rozwiązania (nazywane przez Autora systemami) są przedstawiane według jednego schematu, zgodnie z którym opisywane są ogólne cele rozwiązania, formułowane są zadania projektowe oraz prezentowane są funkcje modułów funkcjonalnych i wyniki symulacyjnych badań ilościowych. Warte jest podkreślenia to, że w modelach symulacyjnych - zaprojektowanych i wykorzystywanych do badań ilościowych opracowanych metod zarządzania żądaniami – bardzo starannie i z dużym nakładem pracy próbowano odtworzyć rzeczywiste warunki realizacji usług w rozproszonych systemach webowych.

I.3. Zawartość pracy

Monografia składa się ze wstępu, trzech rozdziałów merytorycznych, podsumowania i spisu literatury przedmiotu.

Wstęp składa się z ogólnego wprowadzenia w tematykę systemów webowych oraz czterech części, w których opisano usługę www, zaproponowano klasyfikację systemów webowych z jakością usług, przedstawiono zagadnienia będące przedmiotem monografii oraz podano cel i zakres monografii.

Obszerne wprowadzenie we wstępie zawiera ogólne uwagi dotyczące systemów webowych, jakości usług takich systemów, możliwości dostarczania usług o różnej jakości we współczesnych systemach teleinformatycznych oraz metodyki projektowania systemów, w tym systemów webowych. Treści tej części mają różny poziom ogólności, a formułowane uwagi – najprawdopodobniej ze względu na przyjęty, informacyjny charakter tej części pracy – są często niejednoznaczne i mało precyzyjne.

W części pierwszej wstępu (podrozdział 1.1) opisano usługę www, serwis webowy (rozumiany jako *serwery www, urządzenia sieciowe niezbędne do funkcjonowania serwisu i oferowanych treści*) oraz system webowy (rozumiany jako *system składający się z serwisu webowego oraz odpowiednich urządzeń wraz z oprogramowaniem wspierającym obsługę oraz dostarczanie żądań od klienta i odpowiedzi HTTP z serwisu*). Z podanych opisów nie wynika jednoznacznie jakie są – według Autora - różnice pomiędzy systemem webowym i serwisem webowym, usługami dostarczanymi przez te wyróżniane systemy i serwisy oraz jaki jest cel wprowadzania tak niejednoznacznego i mało czytelnego podziału. Jeżeli serwis webowy składa się z serwerów, urządzeń sieciowych oraz treści, które udostępnia użytkownikom, to czym jest system webowy?

Część druga wstępu (podrozdział 1.2) dedykowania jest klasyfikacji systemów webowych z jakością usług. W omawianej części przedstawiono wyczerpującą analizę literatury z zakresu zarządzania rozdziałem zadań i alokacji zasobów stosowane w systemach webowych oraz zaproponowano podział systemów webowych (zamiennie używając terminu serwis webowy) z jakością usług na systemy webowe z kryterium biznesowym, z kryterium czasowym i z kryterium wydajności. Proponowaną, hierarchiczną klasyfikację systemów webowych Autor uzasadnia różnicami w podejściu do projektowania systemów webowych i ocenie jakości działania systemów webowych z trzech różnych punktów widzenia: operatora systemu, dostawcy treści i użytkownika. Proponując klasyfikację systemów webowych Autor popełnia błąd metodologiczny – system webowy jest systemem usługowym, którego projektowanie i analiza działania nie może nie uwzględniać jakościowych i ilościowych wymagań usługobiorcy oraz ograniczeń usługodawcy. Brak jest w pracy uzasadnienia dla potrzeby zaproponowanej klasyfikacji systemów webowych. Przedstawiane uzasadnienie takiego podziału nie jest ani przekonywujące, ani jednoznaczne (o czym Autor pisze w pracy), ani nie odpowiada obecnie stosowanym modelom biznesowym. Trudno sobie wyobrazić i znaleźć przykład systemu webowego, który mógłby być sklasyfikowany jednoznacznie zgodnie z zaproponowanym podziałem. W omawianej części pracy znajduje się wiele stwierdzeń, których prawdziwość jest co najmniej dyskusyjna. Być może wynika to z nadmiaru treści, które Autor chciał przekazać w możliwie syntetycznej formie. Zaproponowana klasyfikacja systemów webowych nie ma praktycznie żadnego wpływu na treści prezentowane w dalszych częściach pracy.

W części trzeciej wstępu (podrozdział 1.3) opisano zaproponowane przez Autora rozwiązania dotyczące dwóch sposobów zarządzania obsługą żądań użytkowników w systemach webowych, których celem jest minimalizacja średniego czasu odpowiedzi na żądania oraz dostarczanie odpowiedzi w zadanym czasie. Dwa wymienione sposoby zarządzania obsługą żądań dotyczą trzech różnych wersji systemów webowych: z jednym serwerem, z klastrem serwerów umieszczonych w jednej lokalizacji (system wieloserwerowy lokalnie rozmieszczony) i z wieloma klastrami serwerów rozmieszczonych w różnych lokalizacjach (system wieloserwerowy globalnie rozmieszczony).

Część ostatnia wstępu (podrozdział 1.4) to deklaracja celu i opis zawartości monografii. Zdeklarowany cel pracy, to *opracowanie grupy systemów webowych z kryterium czasowym, które umożliwiają podnoszenie jakości oferowanych usług poprzez minimalizację czasów odpowiedzi na żądania HTTP ... oraz utrzymywanie czasów odpowiedzi nie dłuższych niż założone dla całych stron www* ... dla różnych serwisów webowych (różniących się liczbą i rozmieszczeniem serwerów). Dla jednoznaczności celu pracy zasadne byłoby dopowiedzenie, co to znaczy opracować system webowy i jaka jest relacja pomiędzy systemem i serwisem webowym. Z lektury dalszych części pracy wynika, że *opracowanie systemu webowego* rozumiane jest jako opracowanie metod i algorytmów scentralizowanego i rozproszonego zarządzania rozdziałem żądań oraz weryfikacja ich jakości w środowisku symulacyjnym.

Rozdział drugi pracy dotyczy modelu systemu obsługi żądań HTTP, który jest wykorzystywany w opracowanych metodach zarządzania obsługą żądań HTTP (opisanych w dalszych rozdziałach rozprawy) do szacowania czasów odpowiedzi na żądania. W zaproponowanym przez Autora modelu systemu obsługi żądań wykorzystano klasyczne podejście do szacowania obciążenia rozproszonego systemu masowej obsługi poprzez analizę tzw. zwłoki czasowej – całkowitego opóźnienia obsługi żądania wnoszonego przez system obsługi. Przedstawiony adaptacyjny model systemu obsługi jest *de facto* modelem opóźnienia w systemie webowym.

We wstępie do tego rozdziału Autor opisuje wymagania stawiane modelom systemu webowego, których spełnienie przesądza o przydatności modelu w zadaniach projektowania i analizy systemów zarządzania obsługą żądań oraz uzasadnia przydatność opracowanego modelu rozmyto-neuronowego. Opisywany pokrótce w pracy (opisy szczegółowe można znaleźć w publikacjach Autora) adaptacyjny model systemu obsługi żądań HTTP – opracowany i rozwijany przez Autora od wielu lat - umożliwia szacowanie czasów odpowiedzi serwerów webowych na żądania HTTP z wykorzystaniem informacji o bieżących i przeszłych zdarzeniach w systemie.

W podrozdziale 2.1 wprowadzano oznaczenia wykorzystywane do zdefiniowania zadania modelowania obsługi żądań HTTP w systemie webowym oraz podano wymagania stawiane modelowi systemu obsługi żądań (w postaci zadania projektowego). W podrozdziale 2.2 przedstawiono kolejno: schemat modelu obsługi żądań, mechanizm estymacji czasu odpowiedzi na żądanie i mechanizm adaptacji wartości parametrów modelu. W procesie estymacji czasu odpowiedzi i adaptacji wartości parametrów modelu wykorzystywane są wartości miar obciążenia systemu, których przykładów ani sposobów wyznaczania w przedstawionym opisie nie podano.

W podsumowaniu rozdziału drugiego powtórzono wcześniej podane, ogólne uwagi dotyczące opracowanego modelu rozmyto-neuronowego do adaptacyjnego szacowania czasu odpowiedzi na

żądania w systemie webowym. Brak jest w podsumowaniu jakościowej i ilościowej oceny stopnia realizacji zadania projektowego, które zostało sformułowane w opisie zadania modelowania. W szczególności bark jest dyskusji jakości szacowania czasu odpowiedzi, tzn. jaka jest złożoność obliczeniowa algorytmów estymacji i adaptacji, w jakim stopniu gromadzenie i przetwarzanie danych o przeszłych zdarzeniach i obciążeniu serwerów poprawia dokładność szacowania czasu odpowiedzi, czy i jak opracowany model jest wrażliwy na rozmiar okna czasowego dla danych o przeszłych zdarzeniach, czy i jak model jest wrażliwy na niekompletność danych, jaka jest skalowalność modelu, itd.

W rozdziale trzecim przedstawiono trzy zadania zarządzaniem rozdziałem żądań, których celem jest minimalizacja średniego czasu odpowiedzi w systemach webowych z lokalnie rozmieszczonymi serwerami webowymi, globalnie rozmieszczonymi serwerami webowymi z serwerami pośredniczącymi oraz globalnie rozmieszczonymi serwerami webowymi bez serwerów pośredniczących. Wspomniane trzy sposoby zarządzania rozdziałem żądań różnią sposoby użycia opisanej w rozdziale drugim metody adaptacyjnego szacowania czasu obsługi żądania. Sposoby użycia modelu opóźnienia zależą od ich liczby i lokalizacji w systemie webowym (w odniesieniu do użytkowników i lokalnych podsystemów).

Część pierwsza rozdziału trzeciego (podrozdział 3.1) dotyczy systemu LFNRD (*Local Fuzzy Neural Request Distribution*) – systemu dystrybucji żądań do lokalnie rozmieszczonych serwerów webowych. We wprowadzeniu do tej części pracy przedstawiono komponenty wieloserwerowych systemów webowych, w których implementowane są algorytmy dystrybucji żądań wpływające na jakość obsługi żądań oraz ogólną charakterystykę algorytmów dystrybucji żądań. Dalej opisano system LFNRD w którym zastosowano metodę dystrybucji żądań (nazwaną metodą LFNRD), sformułowano zadanie projektowe dotyczące przełącznika webowego, przedstawiono przełącznik LFNRD i omówiono wyniki badań systemu LFNRD.

Sformułowane zadanie projektowe dotyczy opracowania metody wyznaczania - dla każdego nadchodzącego żądania - takiego serwera, dla którego szacowany czas odpowiedzi na żądanie będzie najmniejszy. Do oszacowania czasu odpowiedzi serwerów wykorzystywane są wartości aktualnego obciążenia serwerów (liczba równocześnie obsługiwanych żądań i liczba równocześnie obsługiwanych żądań dotyczących obiektów dynamicznych). Zadanie dotyczy rozdziału żądań przy założeniu, że wszystkie żądania należą do jednej klasy ruchu i każde z żądań może być obsłużone przez każdy serwer w klastrze. Zaproponowane rozwiązanie jest przybliżonym rozwiązaniem adaptacyjnego zadania równoważenia obciążenia w systemie z wieloma stanowiskami obsługi.

Do zbadania jakości pracy zaproponowanej metody rozdziału żądań wykorzystano autorski system symulacyjny sieci kolejek. Dla potrzeb skonfigurowania środowiska symulacyjnego przyjęto szereg założeń dotyczących m.in. średniej wielkości strony, rozkładów wartości parametrów opisujących żądania i zachowania użytkowników, struktury strumienia żądań (żądania statyczne i dynamiczne), klas żądań dynamicznych oraz konfiguracji serwerów w klastrze (cztery konfiguracje trzech i pięciu serwerów z jednakowymi i różnymi czasami obsługi żądań). Badania polegały na porównaniu średnich czasów obsługi żądań, generowanych w postaci samopodobnego strumienia żądań, w czterech konfiguracjach symulowanego systemu webowego, w których implementowane były różne metody rozdziału żądań – zaproponowana metoda LFNRD oraz metody i algorytmy referencyjne (FNRD (*Fuzzy Neural Request Distribution*?), LARD (*Locality Aware Request Distribution*), CAP (*Client Aware Policy*), WRR (*Weighted Round Robin*) i RR (*Round Robin*)).

Wyniki wybranych badań ilościowych metody LFNRD, prezentowanych w pracy, wskazują na to, że wykorzystane w metodzie LFNRD mechanizmy szacowania czasu odpowiedzi serwerów z wykorzystaniem informacji o obciążeniu serwerów i czasach obsługi przeszłych żądań, pozwalają na skrócenie – w porównaniu z metodami referencyjnymi – średniego czasu obsługi żądań. Skrócenie średniego czasu obsługi przy korzystaniu z metody LFNRD może być większe wtedy, gdy system webowy jest heterogeniczny, tzn. gdy czasy obsługi tych samych żądań w różnych serwerach klastra są różne. W szczególności, korzyści ze stosowania proponowanego algorytmu są duże, jeżeli porównywać działanie opracowanej metody LFNRD z *bez pamięciowymi* metodami rozdziału żądań. Analizując opis zaplanowanych i wykonanych badań symulacyjnych należy podkreślić staranność i troskę w odtworzeniu w środowisku symulacyjnym rzeczywistych warunków obsługi żądań w systemie webowym, w tym przyjętego modelu zachowań użytkowników systemu webowego (samopodobny strumień żądań). Należy jednak także zauważyć, że przyjęte w modelu symulacyjnym założenia dotyczące strumieni żądań, struktury ruchu, konfiguracji klastrów, itd. powodują, że prezentowane wyniki badań ilościowych metody LFNRD mają charakter ilustracyjny.

W podsumowaniu podrozdziału 3.1 (dotyczącego metody LFNRD) przedyskutowano możliwe praktyczne zastosowania i ograniczenia zaproponowanej metody rozdziału żądań, w tym złożoność obliczeniową metody i jej wpływ na skalowalność rozwiązania.

Część druga rozdziału trzeciego (podrozdział 3.2) dotyczy systemu GARDiB (*Global Adaptive Request Distribution with Broker*) dystrybucji żądań pomiędzy globalnie rozmieszczonymi serwerami webowymi. Różnica w zadaniach dystrybucji żądań w systemach LFNRD (jeden klaster) i GARDiB (zbiór klastrów) polega na tym, że system globalnie rozmieszczonych serwerów webowych jest złożony ze zbioru lokalnie rozmieszczonych serwerów webowych (nazywanych w pracy serwisami lokalnymi). Decyzje w miejscu obsługi żądań w systemie serwerów rozmieszczonych globalnie mogą być podejmowane dwuetapowo – najpierw wybierany jest jeden z podsystemów (podsystem lokalny będący zbiorem lokalnie rozmieszczonych serwerów webowych), a następnie żądania są dystrybuowane wewnątrz wcześniej wybranego lokalnego podsystemu webowego. Istotą systemu GARDiB jest to, że dystrybucja żądań dostępu pomiędzy lokalnymi podsystemami (lokalnymi serwisami) realizowana jest przez serwery pośredniczące, na wejściu których agregowany jest strumień żądań generowanych przez podzbiory zbioru użytkowników.

Różnica pomiędzy zaproponowanymi zadaniami projektowymi i koncepcjami rozwiązania zadań rozdziału żądań w systemach LFNRD i GARDiB polega na tym, że w systemie LFNRD wybierany jest serwer webowy, a w systemie GARDiB wybierany jest serwis lokalny. Serwery (w systemie LFNRD) i serwisy lokalne (w systemie GARDiB) wybierane są na podstawie oszacowania czasu odpowiedzi z wykorzystaniem tego samego modelu obsługi żądania. W systemie GARDiB – podobnie jak w systemie LFNRD – przyjmowane jest założenia, że każde generowane żądanie może być obsłużone przez każdy z serwisów lokalnych. Konstrukcja systemu GARDiB, w której występują serwery pośredniczące, niczego nie zmienia – w porównaniu z systemem LFNRD – w sposobach estymacji czasu odpowiedzi, adaptacji wartości parametrów modelu obsługi (modelu opóźnienia) żądań i podejmowania decyzji. W systemie GARDiB ta sama metoda dystrybucji żądań jest stosowana w serwerach pośredniczących i przełącznikach tworzących swego rodzaju drzewa decyzyjne. Przyjęcie dwóch nazw (przełącznik i serwer przełączający) dla rozwiązań realizujących te same funkcje nie ma większego uzasadnienia.

W omawianym podrozdziale opisano system GARDiB, sformułowano zadanie rozdziału dystrybucji żądań pomiędzy serwisami lokalnymi, przedstawiono moduły dystrybutora żądań (serwer pośredniczący) oraz zaprezentowano opis i wyniki badań ilościowych w modelu symulacyjnym systemu GARDiB, w którym wcześniej zaproponowana metody rozdziału żądań (LFNRD) używana jest dwukrotnie dla przekierowania każdego z żądań.

Badania ilościowe systemu GARDiB zostały przeprowadzone dla czterech konfiguracji systemu, w których udział strumienia żądań kierowanego poprzez serwer pośredniczący - do trzech lub pięciu różnie obciążanych podsystemów lokalnych - w całkowitym strumieniu ruchu wynosił od 16% do 40%. W badaniach symulacyjnych wykorzystano modele strumieni generowanych żądań i opóźnienia wnoszonego przez Internet dobrze opisujące rzeczywiste zachowania użytkowników systemów webowych i obsługę ruchu w Internecie. Do oceny jakości równoważenia obciążeń poprzez dwukrotne użycie opracowanej metody dystrybucji żądań (wcześniej opisanej metody LFNRD) przeprowadzono ilościowe badania porównawcze, w których jako metody referencyjne wybrano prosty algorytm karuzelowy (RR) i dwie odmiany ważonego algorytmu karuzelowego (WRR).

Przedstawione wyniki badań ilościowych wskazują na korzyści (skrócenie średniego czasu odpowiedzi) z użycia w modelu symulacyjnym proponowanego rozwiązania - w porównaniu z metodami rozdziału żądań wybranych jako referencyjne - wynikające z równoważenia obciążeń w rozproszonym systemie webowym z wykorzystaniem zaproponowanej adaptacyjnej metody szacowania czasów odpowiedzi w systemie webowym. Poprawa jakości obsługi żądań (skrócenie średniego czasu obsługi) jest tym większa, im większy jest udział strumienia żądań obsługiwanego przez serwer pośredniczący w całkowitym strumieniu żądań oraz im większe występują różnice w obciążeniu poszczególnych lokalnych podsystemów webowych. Otrzymane wyniki są jakościowo podobne do wyników uzyskanych dla symulacji systemu LFNRD – uwzględnienie danych o historii obsługi żądań w systemie webowym i aktualnym obciążeniu systemu webowego pozwala na podejmowanie lepszych – w sensie średnim – decyzji o rozdziale obciążenia (dystrybucji żądań).

Relacjonowane w tej części pracy wyniki, dobrze zaplanowanych i przeprowadzonych pracochłonnych eksperymentów, ilustrują potencjalne korzyści ze stosowania zaproponowanej metody dystrybucji żądań w hierarchicznym, dwupoziomowym systemie zarządzania. Wnioski formułowane na podstawie przeprowadzonych badań symulacyjnych są prawdziwe w odniesieniu do zakresu przeprowadzonych eksperymentów symulacyjnych oraz przyjętych założeń dotyczących liczności lokalnych podsystemów webowych i struktury obciążenia. Uogólnienie tych wniosków wymagałoby

dalszych badań, w tym m.in. dotyczących wrażliwości rozwiązania na różną liczbę serwerów pośredniczących w systemie webowym.

W podsumowaniu podrozdziału 3.2 (dotyczącego systemu GARDiB) przedyskutowano możliwe zastosowania i ograniczenia zaproponowanego sposobu dystrybucji żądań, jego złożoność obliczeniową oraz skalowalność rozwiązania. Zdaniem Autora proponowane rozwiązanie, tzn. serwer pośredniczący nie powinien być *wąskim gardłem* systemu webowego nawet przy większych obciążeniach systemu webowego. W takich przypadkach Autor sugeruje zwielokrotnienie liczby serwerów pośredniczących. W przypadku potrzeby oceny rzeczywistej przydatności proponowanego rozwiązanie niezbędne jest poszerzenie analizy o analizę korzyści *versus* koszt implementacji.

Część trzecia rozdziału trzeciego (podrozdział 3.3) dotyczy systemu GARD (*Global Adaptive Request Distribution*) dystrybucji żądań dla globalnie rozmieszczonych serwerów webowych. Jednym z uzasadnień dla proponowanej w tym podrozdziale metody dystrybucji i obsługi żądań jest to, że – zgodnie z opinią Autora - proponowana wcześniej metoda GARDiB może być kosztowna i może powodować wzrost opóźnień w przypadku występowania w systemie webowym dużych odległości pomiędzy użytkownikami i serwerami pośredniczącymi.

Celem proponowanego systemu GARD jest – podobnie jak w systemie GARDiB - minimalizacja czasu odpowiedzi na każde, nowo nadchodzące żądanie. W systemie GARD – w odróżnieniu do systemu GARDiB, w którym wykorzystywane są serwery pośredniczące lokalizowane po stronie użytkowników systemów webowych – proponowane jest użycie serwerów dystrybucji żądań lokalizowanych w lokalnych podsystemach webowych (lokalnych serwisach webowych). Istotnym elementem systemu GARD jest to, że jeżeli w odpowiedzi podsystemu lokalnego przekazywany jest obiekt HTML, to serwer dystrybucji żądań modyfikuje ten obiekt tak, że adresy obiektów zagnieżdżonych wskazują podsystemy lokalne, których czasy odpowiedzi będą najmniejsze. Do modyfikacji obiektu wykorzystywane są dane o aktualnym obciążeniu podsystemów lokalnych oraz obciążenia sieci.

W omawianym podrozdziale przedstawiono ogólny opis proponowanego systemu, wprowadzono oznaczenie z użyciem których sformułowano zadanie projektowe dotyczące serwera dystrybucji żądań, opisano proponowany serwer dystrybucji żądań oraz zaprezentowano wyniki badań symulacyjnych opracowanej metody dystrybucji żądań.

Badania symulacyjne systemu GARD zostały przeprowadzone w sposób podobny dla sposobu badania systemu GARDiB. W szczególności dotyczy to liczby lokalnych podsystemów webowych, struktury i charakterystyk strumieni żądań, charakterystyk sieci oraz charakterystyk użytkowników. Prezentowane wyniki ilościowych badań zaproponowanej metody dystrybucji żądań wskazują na to, że zastosowanie tej metody pozwala – w porównaniu z innymi, referencyjnymi metodami – na skrócenie czasów odpowiedzi na żądania. Dyskusja korzyści wynikających z ewentualnego zastosowania metody GARD powinna uwzględniać także koszty implementacji serwerów dystrybucji w lokalnych podsystemach webowych oraz koszt obsługi dodatkowego ruchu sieciowego, niezbędnego do działania systemu.

Otrzymane wyniki mają charakter ilustracyjny m.in. ze względu na to, że złożoność proponowanego rozwiązania wymusiła konieczność przyjęcia w modelu symulacyjnym wielu założeń i ograniczeń upraszczających, które znacząco wpływają na wyniki ilościowych porównań.

Złożoność obliczeniowa systemu GARD jest – w porównaniu ze złożonością systemu GARDiB – większa, co wynika m.in. z tego, że liczba serwerów dystrybucji żądań jest równa liczbie wyróżnionych podsystemów lokalnych i że każdy z serwerów dystrybucji żądań utrzymuje i aktualizuje dane dotyczące obciążeń sieci i wszystkich podsystemów lokalnych.

W podsumowaniu wyników badań systemu GARD Autor stwierdza, że – podobnie jak w podsumowaniu badań systemu GARDiB - serwer dystrybucji żądań nie powinien być *wąskim gardłem* systemu webowego nawet przy większych obciążeniach systemu webowego (samo pojęcie większego obciążenia nie jest definiowane). W takich przypadkach Autor sugeruje zwielokrotnienie liczby serwerów dystrybucji w lokalnych podsystemach webowych, ale zagadnienie zwielokrotnienia serwerów dystrybucji nie jest w pracy analizowane.

Brak jest w pracy bezpośredniego porównania systemów GARDiB i GARD, które – z założenia – dedykowane są do rozdziału żądań w takich samych systemach webowych (globalnie rozmieszczone serwery webowe).

Rozdział trzeci kończy podsumowanie, w którym wskazano na potencjalne korzyści z użycia zaproponowanego sposobu estymacji czasu obsługi żądań z wykorzystaniem adaptacyjnego modelu opóźnienia w systemach webowych w trzech różnych konfiguracjach systemu dystrybucji żądań. Celem opracowanych i analizowanych, lokalnych i rozproszonych, systemów dystrybucji żądań jest minimalizacja czasu obsługi żądań, które są obsługiwane w kolejności ich napływu.

W rozdziale czwartym przedstawiono trzy propozycje zarządzania rozdziałem ograniczonej liczby żądań w systemach webowych (maksymalne obciążenie systemu webowego jest zadane), w których skrócenie czasu obsługi żądań dotyczących stron złożonych (tak, aby nie został przekroczony dopuszczalny czas odpowiedzi) może odbywać się kosztem wydłużenia czasu obsługi żądań dotyczących prostych stron (tak, aby czasy obsługi były akceptowalne, ale jednocześnie aby nie zostały przekroczone).

We wprowadzeniu do rozdziału czwartego podano ogólne wymagania dotyczące systemów obsługi z gwarancją czasu obsługi oraz założenia dotyczące zadań rozdziału żądań - prezentowanych w dalszych częściach tego rozdziału – których dotyczą algorytmy heurystyczne szeregowania żądań w jednoserwerowym systemie webowym oraz dystrybucji i szeregowania żądań w wieloserwerowych (rozmieszczonych lokalnie i globalnie) systemach webowych.

W podrozdziale pierwszym rozdziału czwartego (podrozdział 4.1) opisano system WEDF (*Web Earliest Deadline First*), wprowadzono oznaczenia niezbędne do zdefiniowania zadania projektowego, zdefiniowano zadanie projektowe, przedstawiono moduły serwera WEDF oraz zaprezentowano wyniki badań symulacyjnych proponowanego sposobu zarządzania obsługą żądań. System WEDF jest przeznaczony do zarządzania obsługą żądań w jednoserwerowym systemie webowym, a jego celem jest zapewnienie czasu obsługi stron www w czasie nie większym od - zadanego dla każdego z żądań - czasu maksymalnego. Z analizy proponowanego opisu proponowanego sposobu obsługi żądań wynika, że ograniczona z góry liczba żądań przyjmowanych do obsługi gwarantuje, że żądania te będą obsłużone w wymaganym czasie (żądania nie będą odrzucane), a szeregowanie pozwoli na poprawę jakości usług (w sensie satysfakcji) w porównaniu z obsługą bez szeregowania.

Przedstawiony opis systemu WDF (*de facto* implementacja strategii szeregowania EDF) i jego modułów jest mało precyzyjny i mało czytelny, w szczególności we fragmentach dotyczących wymagań na maksymalną liczbę równocześnie obsługiwanych żądań i maksymalną liczbę żądań przyjmowanych do obsługi w systemie.

Przedstawione w końcu omawianego podrozdziału wyniki symulacyjnych badań ilościowych opracowanego heurystycznego algorytmu szeregowania żądań zostały przeprowadzone w sposób zbliżony dla opisanych wcześniej badań symulacyjnych. Jakość obsługi żądań w systemie WEDF została porównana z jakością obsługi żądań w systemach, w których do szeregowania żądań stosowane są referencyjne algorytmy szeregowania. Z analizy otrzymanych i zaprezentowanych wyników badań ilościowych wynika, że zaproponowany sposób szeregowania żądań, polegający na skracaniu czasu odpowiedzi na tzw. *ciężkie* żądania (żądania stron dynamicznych) kosztem wydłużania czasu odpowiedzi na tzw. *lekkie* żądania (żądania stron statycznych), pozwala na wzrost średniej wartości satysfakcji użytkowników, a tzn. na skrócenie średniego czasu obsługi żądań.

W drugiej części rozdziału czwartego (podrozdział 4.2) omawiany jest system MLF (*Most Loaded First*), którego zadaniem jest zapewnienie czasu odpowiedzi na żądania, mniejszego od danego maksymalnego czasu odpowiedzi w klastrze serwerów rozmieszczonych lokalnie.

System MLF jest podobny do opisanego wcześniej (podrozdział 3.1) systemu LFNRD. Różnica pomiędzy wspomnianymi systemami jest taka, że w systemie MLF – w przeciwieństwie do systemu LFNRF - rozdział żądań pomiędzy serwery klastra jest poprzedzany szeregowaniem żądań w kolejce. Celem tego szeregowania jest – podobnie jak we wcześniej omawianym systemie WDF – przyśpieszanie obsługi żądań mających dłuższe czasy obsługi kosztem opóźniania obsługi żądań wymagających krótszych czasów obsługi, ale tak aby czasy obsługi nie były większe od zadanych, maksymalnych czasów obsługi.. Podobnie jak w systemie WDF zakłada się, że liczba przyjmowanych do obsługi żądań jest ograniczona i taka, że możliwe jest obsłużenie każdego z nich w czasie nie dłuższym niż dany, maksymalny czas obsługi żądania. Celem szeregowania żądań (zgodnie ze strategią EDF) jest ustalanie takiej kolejności obsługi żądań, aby maksymalizować średnią wartość satysfakcji użytkowników.

W omawianym podrozdziale opisano założenia systemu MLF, sformułowano zadanie projektowe (powtórzenie zadań projektowych dotyczących systemu WEFD i LFNRD), przedstawiono moduły przełącznika MLF (złożenie serwera szeregującego WEDF i przełącznika LFNRD) i zaprezentowano wyniki badań symulacyjnych proponowanego rozwiązania.

Badania symulacyjne systemu MLF – do porównania jakości zaproponowanego sposobu zarządzania rozdziałem żądań i innych, referencyjnych metod rozdziału żądań - zostały wykonane w środowisku symulacyjnym wykorzystywanym wcześniej do symulacyjnych badań ilościowych systemu LFNRD.

Przedstawione wyniki badań ilościowych wskazują na korzyści (skrócenie średniego czasu odpowiedzi) - w porównaniu z metodami rozdziału żądań wybranych jako referencyjne - wynikające z równoważenia obciążeń w rozproszonym systemie webowym z wykorzystaniem zaproponowanej metody szacowania czasów odpowiedzi. Poprawa jakości obsługi żądań (wartość średnia satysfakcji)

jest tym większa, im większe są różnice w obciążeniu poszczególnych lokalnych podsystemów webowych. Otrzymane wyniki są jakościowo podobne do wyników uzyskanych dla symulacji systemu LFNRD – uwzględnienie danych o historii obsługi żądań w systemie webowym i aktualnym obciążeniu systemu webowego pozwala na podejmowanie lepszych – w sensie średnim – decyzji o rozdziale obciążenia (dystrybucji żądań) i tym samym skrócenie średniego czasu odpowiedzi. Wzrost wartości satysfakcji powodowany jest możliwością zmiany kolejności obsługi żądań.

Prezentacja wyników przykładowych, ilustracyjnych badań ilościowych systemu MLF opatrzona jest ogólnym podsumowaniem, podobnym do podsumowań wyników badań wcześniej opisywanych systemów zarządzania rozdziałem żądań. W podsumowaniu tym Autor stwierdza, że *skonstruowany przełącznik* - łączący wcześniej opisane metody szeregowania i dystrybucji żądań - *może obsługiwać wydajnie serwis webowy i nie powinien być wąskim gardłem systemu*. Bez uszczegółowienia terminów *może* i *nie powinien* trudno rozstrzygać jednoznacznie o skalowalności proponowanego rozwiązania.

W podrozdziale trzecim rozdziału czwartego (podrozdział 4.3) przedstawiono system GGARDiB (*Guaranteed Global Adaptive Request Distribution with Broker*), w którym do rozdziału żądań pomiędzy lokalne podsystemy systemy webowego wykorzystywane są serwery pośredniczące. Działanie systemu GGARDiB jest podobne do działania wcześniej opisanego systemu GARDiB (podrozdział 3.2). Różnica pomiędzy systemami GARDiB i GGARDiB polega na tym, że w drugim z wymienionych systemów, serwer pośredniczący dodatkowo wyznacza czas zakończenia obsługi żądania, a przełącznik webowy dodatkowo szereguje żądania przekazane do obsługi w podsystemie lokalnym. W proponowanym systemie GGARDiB serwer pośredniczący (najbliższym dla użytkownika generującego żądanie) wskazuje lokalny podsystem webowy i ustala czas zakończenia obsługi żądania, a następnie przełącznik webowy (rozdzielający żądania pomiędzy serwery podsystemu lokalnego) szereguje żądania tak, aby czas obsługi żądania nie był większy od czasu obsługi (ustalonego w serwerze pośredniczącym) i wyznacza serwer do obsługi żądania.

W omawianym podrozdziale opisano zasady obsługi żądań, wprowadzono oznaczenia niezbędne do sformułowania zadań projektowych serwera pośredniczącego i przełącznika webowego oraz ich modułów, a następnie przedstawiono wyniki badań symulacyjnych zaproponowanego rozwiązania zadania rozdziału żądań.

Badania ilościowe systemu GGARDiB zostały przeprowadzone w modelu symulacyjnym podobnym do tego, który był wykorzystany do badań ilościowych systemu GARDiB. Z prezentowanych wyników badań ilościowych wynika, że rozdział żądań w systemie GGARDiB pozwala na uzyskanie wyższej jakości obsługi żądań (w sensie wybranego wskaźnika jakości) w przypadku, gdy lokalne obciążenie podsystemów lokalnych różni się. W przypadku równomiernie obciążonych podsystemów lokalnych, jakość obsługi żądań – uzyskana w systemie webowym GGARDiB i systemach webowych z referencyjnymi metodami rozdziału żądań – jest zbliżona.

Prezentowane w monografii wyniki badań ilościowych mają charakter ilustracyjny. Wyciąganie daleko idących wniosków na podstawie wyników, uzyskanych w czasie bardzo pracochłonnych i starannie przeprowadzonych eksperymentów symulacyjnych, nie jest możliwe ze względu na ograniczone zakresy zmienności liczby podsystemów lokalnych, oczekiwanych czasów odpowiedzi na żądania, różnic w obciążeniach podsystemów lokalnych, opóźnień wnoszonych przez sieć, itd.

Rozdział czwarty kończy krótkie podsumowanie, w którym wskazano na potencjalne korzyści wynikające z użycia znanej strategii szeregowania żądań (EDF) i opracowanej metody szacowania czasu obsługi żądań w systemach webowych (autorski adaptacyjny model opóźnienia) do obsługi żądań w systemach webowych z trzema różnymi konfiguracjami systemu dystrybucji żądań (dla pojedynczego serwera, dla klastra serwerów rozmieszczonych lokalnie i dla klastra serwerów rozmieszczonych globalnie). Celem opracowanych i analizowanych, lokalnych i rozproszonych, systemów szeregowania i dystrybucji żądań jest minimalizacja czasu obsługi żądań w systemie (maksymalizacja satysfakcji), w którym – poprzez limitowanie obciążenia systemu (maksymalne obciążenie systemu webowego jest zadane) – możliwe jest gwarantowanie obsługi żądań w czasie nie dłuższym, niż dany, maksymalny czas obsługi żądań.

W kolejnym rozdziale (rozdział piąty) powtórzono ogólne opisy opracowanego, adaptacyjnego modelu opóźnienia w systemie webowym oraz sześciu systemów zarządzania rozdziałem żądań w systemach webowych – trzech dotyczących zadania minimalizacji średniego czasu odpowiedzi na żądanie i trzech dotyczących zadania minimalizacji czasu obsługi żądań przy zadanym, maksymalnym czasie obsługi żądań. W podsumowaniu Autor zwraca uwagę na ścisły związek opracowanych rozwiązań z teorią i praktyką współczesnych systemów webowych oraz ich potencjalne zastosowania.

W podsumowaniu wskazano – na dużym poziomie ogólności – dalsze, możliwe rozwijanie rozwiązań proponowanych w monografii w systemach webowych zorientowanych na usługi

(wykorzystujących paradygmat SOA (*Service Oriented Architecture*)) oraz w systemach typu CAN (*Content Aware Network*).

I.4. Uwagi

Podane dalej uwagi mają różny poziom szczegółowości. Uwagi te nasunęły się w czasie lektury pracy na potrzeby oceny oryginalności prezentowanych wyników badań oraz oceny sposobów prezentacji treści pracy.

1. Dla celów pracy przyjęto, że *system webowy z jakością usług jest to system webowy, w którym wykorzystywane są odpowiednie metody umożliwiające osiąganie zamierzonych celów w zakresie jakości oferowanych usług* (str. 13). Z tego stwierdzenia wynika, że każdy system webowy jest systemem webowym z jakością usług.
2. Rozważania Autora (str. 15) dotyczące usług typu *best effort*, usług przewidywalnych i gwarantowanych są niejednoznaczne i nieprecyzyjne. Dyskutowane typy usług, występujące w różnego rodzaju systemach masowej obsługi, mogą i są jednoznacznie definiowane w obszernej literaturze przedmiotu, w tym literaturze z zakresu systemów teleinformatycznych.
3. W ramach dyskusji typów usług (str. 15) napisano, że *W ramach usług typu best effort celem podejmowania decyzji jest najczęściej maksymalizacja jakości usług. W usługach przewidywalnych i gwarantowanych celem sterowania jest utrzymanie jakości na wymaganym poziomie. Systemy oferujące jakość usług przewidywalną lub gwarantowaną są systemami z zadanym wyjściem.* Każde z tych zdań wymaga komentarza lub poprawy. Po pierwsze - nie jest prawdą, że w systemach świadczących usługi typu *best effort* celem jest maksymalizacja jakości usług. Po drugie – zadanie utrzymania jakości usług dotyczy nie tylko systemów dostarczających usługi przewidywalne i gwarantowane – dotyczy to także systemów dostarczających usługi typu *best effort*. Po trzecie – systemami z zadanym wyjściem są wszystkie systemy dostarczające omawiane typy usług. Podobne uwagi można sformułować odnośnie całej dyskusji dotyczącej jakości usług sieciowych prowadzonej przez Autora we wstępie, a w szczególności uwag odnoszących się do możliwości wykorzystania protokołów obecnie stosowanych w sieciach.
4. W kolejnych akapitach wstępu Autor opisuje proces projektowania wydajnych systemów webowych (str. 16). Jest w tym opisie wiele niejednoznaczności spowodowanych używaniem nieprecyzyjnych i wcześniej nie zdefiniowanych pojęć (obiekt, system, proces, zakres podejmowanych decyzji, itd.). Opis ten wymaga wyjaśnień m.in. następujących wątpliwości.
   - Jak definiowany jest *wydajny system webowy*?
   - Stwierdzenie, o tym że *wykonanie projektu wymaga opracowania ... algorytmów podejmowania decyzji, w tym algorytmów dystrybucji lub też szeregowania żądań ... oraz metod umożliwiających osiąganie założonego w projekcie celu* dotyczy projektu systemu webowego, systemu zarządzania systemem webowym, czy algorytmów dystrybucji i szeregowania żądań w systemie webowym?
   - Jak rozumieć stwierdzenia, że *wymagane jest opracowanie odpowiednich algorytmów podejmowania decyzji, w tym algorytmów dystrybucji lub też szeregowania żądań* oraz że *usługi gwarantowane wymagają, aby jakość usług była gwarantowana od początkowego elementu w sieci poprzez wszystkie pośredniczące aż do końcowego, ...* ?
5. Jak należy rozumieć zdanie: *Opracowując metody funkcjonowania systemów webowych należy wskazać odpowiednie rozwiązania techniczne na poziomie protokołów sieciowych, a zwłaszcza protokołów HTTP, TCP oraz IP.* Co to są odpowiednie rozwiązania techniczne na poziomie protokołów sieciowych?
6. Z podanych w podrozdziale 1.1 opisów serwisu i systemu nie wynika jednoznacznie czym różni się serwis webowy od systemu webowego. Jeżeli serwis webowy składa się m.in. z *urządzeń sieciowych niezbędnych do funkcjonowania serwisu* (str. 21), to czy ma sens serwis webowy bez *odpowiednich urządzeń wraz z oprogramowaniem wspierającym obsługę oraz dostarczanie żądań i ... odpowiedzi* (str. 22)? Kontynuując - dlaczego *serwery pośredniczące przekierowujące żądania do klastrów serwerów webowych globalnie rozmieszczonych* (str. 22), będące urządzeniami niezbędnymi do funkcjonowania konkretnego serwisu webowego, nie są elementami serwisu? Poza tym zasadne staje się następujące pytanie - jeżeli serwis webowy jest częścią systemu webowego, to komponentami jakich usług systemu webowego są usługi serwisu webowego?

7. W tytule rozdziału 1.2 oraz w treściach tego rozdziału Autor wielokrotnie używa pojęcia *systemy webowe z jakością usług*. Czy używanie tergo terminu oznacza, że są systemy webowe bez jakości usług?
8. Opisując wybrane zagadnienia dotyczące jakości serwisu webowego Autor napisał (str. 22), że: *zgodnie ze starszym podejściem, stosowanym jeszcze kilka lat temu, na jakość serwisu webowego mają wpływ ... dostępność, ... wydajność, ... czas dostępu do usługi, ... bezpieczeństwo ... i osiągalność*. Czy to oznacza, że są nowsze podejścia i rozwiązania, w których dostępność, wydajność, czas dostępu, bezpieczeństwo i osiągalność nie mają wpływu na jakość serwisu webowego? Zagadnienia klasyfikacji, projektowania i monitorowania jakości usług systemów informatycznych (do których serwisy webowe niewątpliwie należą) jest bardzo dobrze opisana w literaturze przedmiotu i międzynarodowych normach (np. ISO/IEC 9126:2001, ISO/IEC 25000:2005, PN-EN ISO 9001:2009, PN-ISO/IEC 20000:2007) oraz standardach (ISO/IEC 9126, PN-EN ISO 9241, ISO 13407, ISO 13406, ISO 11064, itd.)
9. Proponowany w podrozdziale podział systemów (serwisów?) webowych na systemy z kryterium biznesowym, czasowym i wydajności nie ma przekonywującego uzasadnienia. Trzy różne, brane pod uwagę - przy proponowanej klasyfikacji systemów - perspektywy postrzegania systemów webowych (operator, dostawca treści, użytkownik) występują w przypadku każdego systemu informatycznego i wszystkie muszą być uwzględniane jednocześnie w procesie projektowania i udostępniania systemu. Przystępując do próby klasyfikacji systemów webowych nie można pominąć tego, że aktualnie w Internecie stosowanych jest kilkadziesiąt modeli biznesowych i w każdym z nich komplementarne interesy operatorów, dostawców treści i użytkowników mogą mieć różne znaczenie, ale nie są i nie mogą być przeciwstawiane.
10. Przedstawiona na rysunku 1.2 klasyfikacja systemów webowych z jakością usług nie jest czytelna, a używane nazwy nie są poprawne. Co np. oznacza, że w warstwie trzeciej dla systemów gwarantujących realizację zadań wyróżniane są systemy sterujące dostępem, systemy szeregujące, systemy wieloserwerowe lokalnie rozmieszczone i systemy wieloserwerowe globalnie rozmieszczone? Jeżeli projektowany jest system z gwarancją realizacji zadań, to będzie to zawsze system w którym niezbędne jest sterowanie dostępem i szeregowanie zadań niezależnie od tego, czy będzie to system wieloserwerowy rozmieszczony lokalnie, czy rozmieszczony globalnie.
11. Jaka jest podstawa stwierdzenia (str. 25), że *w systemach webowych z kryterium wydajności i kryterium czasowym wszyscy klienci traktowani są w jednakowy sposób niezależnie od korzyści, jakie mogą przynieść właścicielom serwisu webowego*? Nie jest to zdanie prawdziwe. Ze stwierdzeniem tym wiąże się m.in. pytanie, które dotyczy wielu innych stwierdzeń w pracy – czy wymagania klientów przesądzają o usługach systemów webowych i sposobach ich dostarczania, czy też sposoby dostarczania usług decydują o grupach użytkowników systemów webowych?
12. Zrozumienie zdania (str. 25) o tym, że *systemy z kryterium wydajności można podzielić na systemy maksymalizujące wydajność oraz systemy nie dopuszczające do przeciążeń* wymaga - w pierwszej kolejności - zdefiniowania pojęcia wydajności. Czy możliwy jest wydajny system, w którym dochodzi do przeciążeń?
13. Jakie jest uzasadnienie dla stwierdzenia (str. 30), że w *trakcie bardzo szybkiego rozwoju Internetu dopiero po upływie pewnego czasu dostrzeżono potrzeby klientów końcowych*?
14. W opisie zawartości monografii (str. 31) napisano, że *dla grupy systemów z zadanym czasem odpowiedzi ... opracowane zostały trzy systemy ... w których wykorzystane zostały techniki szeregowania żądań bez sterowania dostępem, co stanowi nowość*. Na czym polega ta wspomniana nowość? Proponowane rozwiązanie jest dobrze znane pod nazwą *Differentiated Services* – jednej z dwóch podstawowych strategii dostarczania jakości usług. Czy możliwa jest – w jakimkolwiek systemie masowej obsługi o skończonej ilości zasobów - obsługa żądań w zadanym przez użytkownika czasie bez sterowania dostępem, tzn. bez względu na wielkość strumienia zgłoszeń?
15. W akapicie dotyczącym modeli systemów webowych (str. 35) napisano m.in.: *w szczególności zastosowanie odpowiedniego modelu w systemach z kryterium czasowym może otworzyć perspektywy w opracowaniu całej klasy nowym metod i rozwiązań*. Co kryje się pod pojęciami: odpowiedni model, otworzyć perspektywy, cała klasa nowych metod (jakich?) i rozwiązań (czego?).
16. Podrozdział 2.1 (str. 36) rozpoczyna następujące zdanie: *na wstępie przedstawione zostanie zadanie projektowe, którego rozwiązanie stanowić będzie koncepcję modelu systemu obsługi żądań*. Dalej – w sformułowaniu zadania projektowego (str. 37) – napisano: *należy opracować*

model systemu obsługi żądań .... Czy zadanie dotyczy opracowania koncepcji modelu, czy modelu? Czy jest różnica pomiędzy koncepcją modelu i modelem?

17. Sformułowane w podrozdziale 2.1 (str. 37) zadanie projektowe jest opisem oczekiwanych własności modelu obsługi żądań. Sformułowanie tego zadania nie jest jednoznaczne, bowiem nie podano wcześniej co to są parametry U_i systemu obsługi żądań, nie ma wymagań ilościowych dotyczących różnicy pomiędzy szacowanym i zmierzonym czasem odpowiedzi, co to znaczy że model powinien się adaptacyjnie dopasowywać oraz co to jest wiedza o zmierzonych czasach odpowiedzi. Jeżeli tak sformułowane zadanie projektowe zostało zrealizowane, to jaka jest i jak jest oceniona jakość proponowanego modelu?
18. W opisie modelu wyszczególniane jest obciążenie systemu, które opisane zostało następująco (str. 37): *obciążenie systemu będzie reprezentowane przez dwie składowe ... wskazujące różne miary obciążenia systemu w zależności od omawianego przypadku.* Dlaczego dwie składowe, a nie mniej lub więcej? Jakie różne miary obciążenia mogą być użyte? Jak wyznaczane są wartości tych miar obciążenia, jeżeli – jak podano w opisie modelu – jedyną mierzoną w systemie wartością jest wartość czasu odpowiedzi na żądanie?
19. Charakteryzując model (str. 45) napisano, że *model rozmyty jest dość ogólny, dzięki czemu po przekształceniu w model rozmyto-neuronowy możliwe będzie zastosowanie go do czasów odpowiedzi dla systemów obsługi żądań HTTP o bardzo zróżnicowanych architekturach.* Jakie jest uzasadnienie dla takiego stwierdzenia? Co to są bardzo zróżnicowane architektury systemów obsługi żądań HTTP? W zaproponowanym modelu opóźnienia architektura systemu webowego nie jest w ogóle uwzględniana, a ewentualne różnice w systemach webowych „widziane są" pośrednio poprzez różnice w mierzonych czasach odpowiedzi na żądania.
20. W podsumowaniu rozdziału drugiego napisano (str. 49), że *przedstawiony model systemu obsługi żądań HTTP stanowi podstawę konstrukcji pięciu spośród sześciu rozwiązań prezentowanych w monografii, dlatego też można stwierdzić, że model ten jest elastyczny.* Nie przesądzając o elastyczności opracowanego modelu (tym bardziej, że pojęcia elastyczności nie zdefiniowano) to to, że model jest podstawą konstrukcji kilku rozwiązań może równie dobrze znaczyć o tym, że pięć spośród sześciu konstruowanych rozwiązań to rozwiązania bardzo podobne.
21. We wprowadzeniu do rozdziału trzeciego (str. 51) najpierw napisano że w rozdziale tym *przedstawione zostały projekty systemów*, a w zdaniu następnym że przedstawione zostaną *systemy.* Lektura treści pracy byłaby łatwiejsza, gdyby Autor – przedstawiając wyniki swojej pracy - doprecyzował co rozumie przez projekt systemu i system.
22. We wprowadzeniu do podrozdziału 3.1 podano kilka stwierdzeń, które wymagają wyjaśnień lub komentarzy:
    - Jak rozumiana jest wydajność serwisów webowych, których dotyczy zdanie (str. 51) o tym, że *stosowanie grupy lokalnie rozmieszczonych serwerów jest ... metodą zwiększania wydajności serwisów webowych*? Czy chodzi o to, że zwiększanie liczby serwerów w klastrze może powodować wzrost jakości usług?
    - Głównymi komponentami systemów webowych są (str. 51): *mechanizm ..., algorytm ... i egzekutor* ... ? Czy mechanizm i algorytm można nazwać komponentami systemu?
    - Co to są mechanizm przekierowywania żądań i architektura połączeń? Jak architektura połączeń stosowana w klastrze (str. 51) zależy od przyjętego mechanizmu przekierowania żądań?
    - Co to znaczy *przesłać żądanie użytkownika do wybranego przez algorytm dystrybucji żądań serwera zgodnie z mechanizmem przekierowania żądań*?
    - Jak zdefiniowana jest wydajność całego serwisu webowego (systemu webowego?) (str. 51), która zależy od *zastosowanego algorytmu dystrybucji żądań i mechanizmu przekierowania*?
    - Czy złożenie mechanizmu przekierowania żądań i algorytmu dystrybucji daje w efekcie metodę dystrybucji żądań (str. 52)? Czym różni się algorytm dystrybucji żądań od metody dystrybucji żądań? Czym różni się – w rozumieniu Autora – metoda od algorytmu?
    - Trudno się zgodzić z poprawnością stwierdzenia o tym, że *algorytmy mogą być podzielone na algorytmy czwartej warstwy sieciowej oraz algorytmy siódmej warstwy sieciowej modelu ISO/OSI* (str. 52). Po pierwsze – w modelu odniesienia ISO/OSI czwarta warstwa, to warstwa transportowa, a warstwa siódma to warstwa aplikacji. Po drugie – wspomniany podział dotyczy nie podziału algorytmów, ale ich lokalizacji

w protokołach warstw. Po trzecie – protokoły TCP i IP są protokołami w modelu TCP/IP, a nie w modelu łączenia systemów otwartych ISO/OSI.

- Podobnie jak wyżej, trudno zgodzić się ze stwierdzeniem (str. 52), że *algorytmy statyczne nie wykorzystują w swym działaniu żadnych informacji o stanie serwerów ani o żądaniu użytkownika*. Parametryzacja algorytmu statycznego wymaga na ogół dobrej znajomości zasobów i charakterystyk obsługiwanego ruchu, a statyczność nie jest synonimem niezmienności w czasie.

23. *Przygotowanie projektu systemu wymagało opracowania metody dystrybucji żądań HTTP oraz algorytmu dystrybucji żądań* (str. 52). Czy w tym zdaniu czegoś jest w nadmiarze, czy też czegoś brakuje?
24. Na liście oznaczeń (str. 54) podano, że *i* jest indeksem momentu (czasu?). Ten sam indeks *i* jest używany do indeksowania żądań. Przyjęcie tego samego indeksu do indeksowania czasu i żądań nie jest poprawne. Co oznacza *y* indeksowane i (momentem) i s (numerem serwera) i jaka jest tego interpretacja? Jaka jest interpretacja założenia podanego w zadaniu projektowym (str. 54) i sformułowanego z wykorzystaniem tego oznaczenia?
25. Na rysunku 3.2 przedstawiono schemat podejmowania decyzji, w którym wyróżniono *algorytm podejmowania decyzji LFNRD* oraz *system webowy LFNRD* (str. 55). Czy to oznacza, że algorytm LFNRD jest poza systemem LFNRD? Po raz kolejny – czym w rozumieniu Autora – różnią się algorytmy, metody, projekty systemów i systemy?
26. W uzasadnieniu dla systemu GARDiB napisano (str. 66), że *serwer DNS z odpowiednio zaimplementowanym algorytmem podejmowania decyzji może wskazać adres serwisu lokalnego* oraz, że *jedną z głównych wad takiego rozwiązania (serwer DNS) jest brak kontroli nad dystrybucją żądań klienta*. Czy serwer pośredniczący w systemie GARDiB jest uzupełnieniem dla serwera DNS? Czy zaplanowana do implementacji w serwerze pośredniczącym metoda dystrybucji żądań nie może być implementowana w serwerze DNS? Czy są jakieś sugestie dotyczące zależności liczby serwerów pośredniczących i liczby serwerów DNS w systemie webowym? Co to znaczy brak kontroli nad dystrybucją żądań klienta, i jak ten brak kompensowany jest w rozwiązaniach proponowanych przez Autora?
27. Jedyną różnicą w sformułowaniach zadań projektowych (sformułowania na str. 54 i 70), rozwiązaniach (rysunki 3.2 i 3.10) i schematach (rysunki 3.3 i 3.11) dla systemu (metody, algorytmu, przełącznika) LFNRD i systemu (metody, algorytmu, serwera pośredniczącego) GARDiB jest to, że w pierwszych z wymienionych są serwery, a w drugich serwisy lokalne indeksowane identycznie. Używanie identycznych oznaczeń w tych dwóch zadaniach powoduje, że przy pomocy tych oznaczeń nie jest możliwe sformułowania zadania dwupoziomowego rozdziału żądań (wspominanego na str. 66).
28. Na rysunku 3.10 przedstawiono schemat podejmowania decyzji w systemie GARDiB, w którym wyróżniono *algorytm podejmowania decyzji GARDiB* oraz *system webowy GARDiB* (str. 70). Czy to oznacza, że algorytm GARDiB jest poza systemem GARDiB?
29. W opisie serwera pośredniczącego GARDiB (str.71) jest następujący fragment – *Model serwisu lokalnego jest modelem systemu obsługi żądań ...* oraz *czas odpowiedzi szacowany jest na podstawie adresu ... żądanego obiektu oraz ...* . Przytoczony fragment jest ilustracyjny dla wielu innych stwierdzeń w pracy, które niczego nie wnoszą i nie są poprawne. Jak można szacować czas odpowiedzi na podstawie adresu?
30. Z lektury podrozdziałów 3.1 i 3.2 wynika, że w systemach LFNRD i GARDiB używana jest ta sama metoda dystrybucji żądań (LFNRD). W systemie LFNRD decyzja o miejscu (serwerze) wykonania żądania wymaga jednokrotnego użycia metody LFNRD dla każdego z żądań, natomiast w systemie GARDiB decyzja o miejscu wykonania żądania (podsystemie lokalnym i serwerze w podsystemie lokalnym wymaga dwukrotnego użycia metody LFNRD dla każdego z żądań. Wydaje się, że wprowadzenie różnych nazw dla metody dystrybucji żądań oraz dla różnych sposobów jej użycia byłoby z korzyścią dla czytelności pracy i oceny oryginalności dorobku Autora. Podobną uwagę można sformułować pod adresem schematów prezentowanych na rysunkach 3.3 i 3.10, z których pierwszy dotyczy *przełącznika LFNRD*, a drugi *serwera pośredniczącego GARDiB*. Te dwa schematy dotyczą różnych implementacji tej samej metody dystrybucji żądań i dlatego stosowanie dwóch różnych nazw nie ma większego uzasadnienia.
31. Analiza wyników prezentowanych na wykresach 3.13d i 3.14d wskazuje na to, że być może wzrost liczby lokalnych podsystemów webowych (lokalnych serwisów webowych) i ich równomierne obciążenie powoduje zmniejszenie korzyści ze stosowania zaproponowanej metody rozdziału żądań, a rozdział żądań z wykorzystaniem mniej złożonych (pamięciowo i obliczeniowo) metod daje zbliżone wyniki. Uwaga ta ma związek z konkluzją podaną na str.

80, zgodnie z którą *algorytm i metoda GARDiB mogłyby być z powodzeniem stosowane w dużych ... serwisach webowych*. Przy okazji pytanie – czym różnią się wspomniane w tym zdaniu algorytm i metoda GARDiB?

32. Na rysunku 3.16 przedstawiono schemat podejmowania decyzji, w którym wyróżniono *algorytm podejmowania decyzji GARD* oraz *system webowy GARD* (str. 84). Czy to oznacza, że algorytm GARD jest poza systemem GARD?
33. Jeżeli dobrze zrozumiałem opis działania systemu GARD, to każdy lokalny podsystem webowy jest uzupełniany o serwer dystrybucji żądań, którego zadaniem jest szacowania czasu odpowiedzi i modyfikacja wartości parametrów obiektów na podstawie aktualizowanych wartości parametrów opisujących wszystkie lokalne podsystemy webowe i sieć. Czy nie byłoby rozwiązaniem korzystniejszym – ze względu na ilość ruchu sieciowego, wymaganego do działania takiego systemu – użycie jednego, centralnego serwera dystrybucji żądań lub rozwiązania, w którym lokalne serwery dystrybucji byłyby koordynowane przez jeden nadrzędny serwer dystrybucji?
34. W opisie badań systemu GARD podano (str. 91), że *wykonane zostały dodatkowe badania przepustowości i czasów RTT dla 50 ośrodków(?) webowych ..., które były wykorzystywane do obliczania czasu przesyłania żądania ....* Jak obliczane były czasy przesyłania żądania – czy były one obliczane, czy predefiniowane? Dalej napisano, że *przed każdym eksperymentem symulacyjnym określany był procent użytkowników, dla których czas RTT ... jest najmniejszy*. Jak był określany ten procent i w jakim celu?
35. W podsumowaniu podrozdziału 3.3 napisano (str. 97), że *na podstawie wyników badań ... nad metodą GARD wnioskować można, że zastosowanie tej metody mogłoby podnieść wydajność serwisu, równocześnie podnosząc znacząco jakość obsługi użytkowników*. Ze zdania tego wynika, że podnosząc wydajność systemu można zwiększyć jakość obsługi żądań użytkowników systemu webowego. Czy przywoływane stwierdzenie nie stoi w sprzeczności z uzasadnieniem klasyfikacji systemów webowych zaproponowanych w podrozdziale 1.2 (str.24)?
36. We wprowadzeniu do rozdziału czwartego (str. 100) podano kilka stwierdzeń, które wymagają uściśleń i wyjaśnień.
    - Zdania *w rozwiązaniach proponowanych w tym rozdziale nie są wykorzystywane techniki sterowania dostępem i w jednakowy sposób obsługiwani są wszyscy klienci zgłaszający się do systemu webowego* oraz że proponowane rozwiązania będą działały *o ile napływ żądań nie będzie większy od możliwości obsługi żądań przez system webowy* wzajemnie się wykluczają – albo nie ma sterowania dostępem i napływ żądań do systemu może być większy od możliwości obsługi, albo jest sterowanie dostępem i napływ żądań jest nie większy od możliwości obsługi żądań. Jeżeli w systemie nie będzie sterowania dostępem, to jak inaczej można wymusić, aby *napływ żądań nie był większy od możliwości obsługi*? Dotyczy to wszystkich systemów obsługi, a nie tylko systemów webowych.
    - Sposób obsługi żądań, zgodnie z którym nie należy dopuszczać do przekroczenia zadanego, maksymalnego czasu odpowiedzi, a w przypadku jego przekroczenia należy minimalizować przekroczenie zadanego, maksymalnego czasu odpowiedzi nie jest rozwiązaniem gwarantującym czas odpowiedzi. Autor myli dwa podstawowe pojęcia dotyczące jakości usług w systemach sieciowych: *Guaranteed Service* [RFC2212] oraz *Controlled-load Service* [RFC2211]. Proponowany przez Autora system obsługi żądań (systemy WEDF, MLF i GGARDiB) można określić jako system dostarczania usług o jakości wyższej niż *best effort* (*better than best effort*).
    - Jak należy rozumieć zdanie, o tym że ... *wykorzystanie proponowanych metod nie podniesie wydajności serwisu ani nie przesunie granicy maksymalnej liczby obsługiwanych żądań w jednostce czasu przy dużym obciążeniu.*Dlaczego? Jak definiowana jest wydajność serwisu? Co to znaczy przesunięcie granicy maksymalnej liczby obsługiwanych żądań? Jak definiowane jest duże obciążenie?
37. Jeżeli zadaniem serwera szeregującego WEDF jest (str. 101) szeregowanie żądań *tak, aby czasy odpowiedzi całych stron www były równe lub krótsze od zadanego czasu maksymalnego*, to jedynym uzasadnionym kryterium jakości jest prawdopodobieństwo odrzucania (lub akceptacji) żądań. Użycie miary jakości obsługi jaką jest satysfakcja jest uzasadnione tylko wtedy, gdy zadaniem serwera WEDF jest jak najlepsze (w sensie satysfakcji) uszeregowanie żądań, których obsługa w dowolnej kolejności nie spowoduje przekroczenia dopuszczalnych czasów maksymalnych.

38. Przedstawiona w trzecim akapicie na str. 102 dyskusja dotycząca przepustowości systemu *X(ar)* i liczby jednocześnie obsługiwanych żądań *ar* oraz ich wzajemnych zależności nie jest ani precyzyjna, ani czytelna. Przyjęcie założenia, że dobranie wartości *ar* takiej, że średnia przepustowość systemu *X(ar)* będzie maksymalna nie gwarantuje tego, że chwilowa wymagana przepustowość systemu będzie mniejsza od maksymalnej przepustowości, a to oznacza że zdeklarowany cel szeregowania nie zostanie osiągnięty. Czego dotyczy rys. 4.2 i jaka jest interpretacja podanej tam wartości ar_{max}? Dlaczego w przypadku, którego dotyczy rys. 4.3, nie przyjąć wartości liczby równocześnie obsługiwanych żądań $ar_{max} = 4$ lub $ar_{max} = 6$? Czy konieczność ograniczenia od góry liczby żądań dopuszczanych do obsługi nie jest tożsama sterowaniu dostępem?
39. Co to znaczy (str. 102), *że zastosowanie metody WEDF nie powinno spowodować zmniejszenia wydajności całego serwisu*? Jeżeli wspominane jest zmniejszenie wydajności serwisu, to należy podać w odniesieniu do jakiej referencyjnej metody szeregowania możliwe jest zmniejszenie wydajności.
40. Czy zadanie projektowe dla serwera szeregującego WEDF dotyczy implementacji strategii obsługi EDF (*Earliest Deadline First*)? Czy jest – a jeżeli tak, to jaka – różnica pomiędzy procesem podejmowania decyzji w systemie WEDF, a obsługą zgodnie z metodą EDF?
41. Na liście oznaczeń (str. 104) i na rysunku 4.3 występuje $ar(\tau)$ – liczba żądań obsługiwanych równocześnie przez realizator w momencie τ. Czy τ nie powinno być indeksowane indeksem żądania?
42. W opisie wyrażenia 4.2 (str. 109) wprowadzono *współczynnik równoczesności*. Jak wartość tego współczynnika zależy od średniej liczby równocześnie obsługiwanych żądań? Czy wartość tego współczynnika zależy od czasu i liczby aktualnie obsługiwanych żądań?
43. W opisie przełącznika MLF (str. 118) napisano, najpierw, że *należy opracować projekt przełącznika webowego MLF* (str. 117), a następnie że ... *należy zaproponować konstrukcję przełącznika webowego, który będzie odpowiednio szeregował żądania i je dystrybuował w taki sposób, aby czasy odpowiedzi dla stron były krótsze od żądanego* ... i dalej – *należy wyznaczyć ... terminy ... tak, aby osiągnąć zamierzony cel*. Jaki jest ten wspomniany, zamierzony cel? Czy projekt i konstrukcja przełącznika to synonimy?
44. Czym uzasadnione jest zadanie projektowe dotyczące przełącznika MLF? To zadanie projektowe – będące złożeniem wcześniej sformułowanych zadań dotyczących projektów serwera szeregującego WEFD i przełącznika webowego LFNRD – zostało wcześniej sformułowane i rozwiązane.
45. W podsumowaniu wyników badań symulacyjnych systemów WEFD (str. 115) i MLF (str. 127) znajduje się to samo zdanie o tym, że *wyniki przeprowadzonych badań wydajności ... oraz analizy pesymistycznej czasowej złożoności obliczeniowej ... można stwierdzić, że ...* proponowane rozwiązanie (odpowiednio WEFD i MLF) ... *może wydajnie obsługiwać serwisy webowe i nie powinny być wąskim gardłem systemu*. Jak została zdefiniowana wydajność i gdzie znajdują się wyniki przeprowadzonych badań wydajności? Jak te dwa stwierdzenia mają się do wcześniejszych deklaracji (rys. 1.2) o tym, że wspomniane systemy nie są systemami z kryterium wydajności i do tego, że prezentowane w monografii wyniki badań dotyczą zależności satysfakcji w funkcji obciążenia oraz zależności czasu odpowiedzi w funkcji obciążenia?
46. Na rys. 4.17a przedstawiono *schemat procesu podejmowania decyzji w systemie GGARDiB*, w którym wyróżniono – oprócz systemu pomiarowego - *algorytm podejmowania decyzji GGARDiB* i *serwisy lokalne systemu GGARDiB*. Co to jest serwis lokalny systemu GGARDiB?
47. W opisie eksperymentów dotyczących systemu GGARDiB napisano (str. 141), że *dla systemu GGARDiB uzyskiwane są największe wartości satysfakcji, jeżeli czas transferu odpowiedzi na żądanie nie jest większy niż o jeden rząd od czasu odpowiedzi żądania w serwisie lokalnym* oraz dalej, że *ze względu na wskazania dotyczące zastosowań systemu GGARDiB przyjęto, że ... wszystkie czasy przesyłania danych zostały podzielone przez określony współczynnik (= 30) tak, aby czasy te nie były znacząco dłuższe od czasów obsługi żądań w serwisach lokalnych*. Jakie są wspomniane wskazania dotyczące zastosowań systemu GGARDiB? Czy konieczność podzielenia czasów transferu przez wspomniany wyżej współczynnik po to, aby otrzymać satysfakcjonujące wartości satysfakcji nie oznacza, że potencjalne korzyści nie kompensują złożoności i zasobochłonności systemu GGARDiB?
48. W podsumowaniu podrozdziału 4.3 napisano (str. 147), że ... *opracowanie metody GGARDiB wskazuje, że jest możliwe opracowanie wieloserwerowego systemu webowego realizującego obsługę żądań w taki sposób, aby zapewnić, by czasy odpowiedzi dla stron webowych były nie dłuższe od zadanej wartości*. Jeżeli opracowanie systemu webowego o wspomnianych

cechach jest możliwe i jeżeli system GGARDiB jest znaczącym krokiem na drodze do takiego rozwiązania, to jakie własności powinny cechować taką metodę? W komentarzu do tego stwierdzenia raz jeszcze pozwolę sobie powtórzyć, że gwarancja czasu odpowiedzi nie jest możliwa w żadnym systemie masowej obsługi, w tym w systemie webowym, bez stosowania metod kontroli dostępu i rezerwacji zasobów. W każdym innym przypadku można mówić tylko o podwyższaniu jakości usług w porównaniu do jakości usług dostarczanych w czasie realizacji strategii *best-effort*.

49. W podsumowaniu rozdziału czwartego (str. 147) napisano, że ... *zaprezentowane propozycje stanowią przegląd rozwiązań umożliwiających gwarantowanie usług* ... . Czy proponowane rozwiązania umożliwiają gwarantowanie czasu wykonania usług, czy gwarantują czas wykonania usługi? Jakie są warunki możliwości gwarantowania czasu wykonania usługi w proponowanych rozwiązaniach?

I.5. Oryginalność wyników badań naukowych przedstawionych w rozprawie.

W omawianej rozprawie habilitacyjnej przedstawiono oryginalne i spójne (podporządkowane deklarowanemu celowi rozprawy) wyniki badań naukowych Autora dotyczących modelowania, analizy i projektowania systemów zarządzania rozdziałem żądań w systemach webowych.

Zaproponowane i analizowane metody dotyczą zarządzania żądaniami w systemach webowych, których celem jest podnoszenie jakości usług webowych rozumianej jako minimalizacja średniego czas obsługi żądania w systemach bez ograniczania liczby żądań napływających do systemu oraz w systemach z ograniczoną liczbą obsługiwanych żądań.

W pracy przedstawiono uzasadnienia dla opracowanych rozwiązań, sformułowano odpowiednie zadania projektowe oraz opisano i przedyskutowano wyniki dobrze zaplanowanych, pracochłonnych badań ilościowych zależności przyjętych wskaźników jakości od wartości wybranych parametrów charakterystyk strumieni żądań oraz systemów webowych.

Metodyka przeprowadzonych i relacjonowanych w rozprawie badań naukowych jest poprawna.

Oryginalność rozprawy habilitacyjnej jest konsekwencją oryginalności formułowanych zadań badawczych, jak i zaproponowanych sposobów rozwiązywania tych zadań, których znaczenie i przydatność zilustrowano przykładowymi i ilustracyjnymi wynikami badań symulacyjnych.

Do najważniejszych wyników badań naukowych i technicznych Autora, prezentowanych w rozprawie i przesądzających o jej oryginalności, zaliczam:

- opracowanie adaptacyjnego modelu (rozmyto-neuronowego) opóźnienia w systemach webowych dla potrzeb szacowania czasów obsługi żądania,
- opracowanie systemu zarządzania rozdziałem żądań pomiędzy serwerami klastra serwerów w zadaniu minimalizacji średniego czasu odpowiedzi na żądania (system LFNRD),
- opracowanie hierarchicznego, dwupoziomowego systemu zarządzania rozdziałem żądań pomiędzy klastrami serwerów i serwerami klastra serwerów w zadaniu minimalizacji średniego czasu odpowiedzi na żądania (system GARDiB),
- opracowanie rozproszonego systemu zarządzania rozdziałem żądań pomiędzy klastrami serwerów i serwerami klastra serwerów w zadaniu minimalizacji średniego czasu odpowiedzi na żądania (system GARD),
- opracowania systemu szeregowania żądań do obsługi przez jeden serwer webowy w zadaniu minimalizacji średniego opóźnienia czasu odpowiedzi z jednoczesną gwarancją czasu obsługi przy założeniu, że obciążenie systemu jest ograniczone od góry (system WEFD),
- opracowania systemu zarządzania integrującego algorytmy szeregowania i rozdziału żądań obsługiwanych przez klaster serwerów webowych w zadaniu minimalizacji średniego opóźnienia czasu odpowiedzi z jednoczesną gwarancją czasu obsługi przy założeniu, że obciążenie systemu jest ograniczone od góry (system MLF),
- opracowania dwupoziomowego, hierarchicznego systemu zarządzania integrującego algorytmy rozdziału żądań (pomiędzy lokalne podsystemy systemu webowego) oraz szeregowania i rozdziału żądań w lokalnym podsystemie webowym w zadaniu minimalizacji średniego opóźnienia czasu odpowiedzi z jednoczesną gwarancją czasu obsługi przy założeniu, że obciążenie systemu jest ograniczone od góry (system GGARDiB),
- zaprojektowanie modeli symulacyjnych (reprezentatywnych dla systemów webowych działających w warunkach współczesnej sieci Internet) z implementowanymi - opracowanymi i referencyjnymi - metodami szeregowania i rozdziału żądań, dla potrzeb wyczerpujących i ilustracyjnych badań ilościowych jakości działania opracowanych systemów zarządzania rozdziałem żądań,

- przeprowadzenie badań symulacyjnych potwierdzających potencjalne możliwości użycia opracowanych metod zarządzania rozdziałem żądań w rzeczywistych systemach webowych,
- poprawną analizę otrzymanych wyników badań symulacyjnych oraz ograniczeń opracowanych metod zarządzania rozdziałem żądań.

Mocne strony opiniowanej monografii to demonstrowana przez Autora bardzo dobra znajomość współczesnych systemów webowych i literatury przedmiotu oraz dobre uzasadnienie dla potrzeby opracowania intuicyjnych rozwiązań zadań zarządzania rozdziałem żądań w systemach webowych o różnej strukturze, poprawnie zbudowane modele symulacyjne i ich użycie do porównawczej weryfikacji jakości zaproponowanych rozwiązań oraz poprawnie formułowane wnioski dotyczące potencjalnych korzyści z zastosowania opracowanych systemów w praktyce, ich ograniczeń i ewentualnych oraz potrzeby dalszych badań.

Słabe strony opiniowanej monografii to nieprzekonywująca klasyfikacja systemów webowych (zbędna z punktu widzenia celów rozprawy), brak ścisłych i jednoznacznych sformułowań zadań, w szczególności dotyczących systemów zarządzania w systemach webowych z gwarancją czasu obsługi żądań, brak odwołań do obszernej literatury z zakresu badań operacyjnych dotyczących systemów czasowo-optymalnych oraz redakcja rozprawy.

Lektury pracy nie ułatwia przyjęty sposób formułowania zadań i opisu proponowanych rozwiązań, w których ogólne i skądinąd proste zadania rozdziału żądań są nazywane i opisywane z użyciem nie zawsze precyzyjnych i konsekwentnie stosowanych terminów technicznych.

I.3. Konkluzje

W omawianej rozprawie Autor analizuje wybrane, dobrze uzasadnione zadania związane z zarządzaniem rozdziałem żądań w systemach webowych.

Zawartość merytoryczna opiniowanej rozprawy potwierdza dobrą znajomość przez Autora analizowanych zagadnień, w szczególności rzeczywistych warunków i ograniczeń sposobów obsługi żądań w rzeczywistych systemach webowych.

Większość wyników badań naukowych, przedstawionych w monografii, została wcześniej opublikowana przez Autora w czasopismach i referatach konferencyjnych, co oznacza że wyniki tych badań zostały zweryfikowane także w przez specjalistów.

Podsumowując wszystkie przedstawione wyżej uwagi stwierdzam, że dr inż. Krzysztof Zatwarnicki wniósł swoją pracą wkład w rozwój uprawianej dyscypliny wiedzy. Oryginalność tego wkładu w rozwój informatyki polega na umiejętnym wykorzystaniu i rozwinięciu dorobku naukowego w zakresie metod modelowania, projektowania i analizy systemów zarządzania rozdziałem żądań HTTP w systemach webowych.

Biorąc pod uwagę treści merytoryczne rozprawy, uzasadnienie dla zadań naukowo-badawczych omawianych w rozprawie, sposób formułowania zadań badawczych oraz sposoby ich rozwiązania i analizę otrzymanych wyników uważam, że rozprawa dr inż. Krzysztofa Zatwarnickiego pt. *Systemy webowe z jakością usług uwzględniające kryterium czasowe. Problemy projektowania* spełnia wymagania stawiane rozprawom habilitacyjnym i wnioskuję o jej przyjęcie.

II. Ocena dorobku naukowego dr inż. Krzysztofa Zatwarnickiego

Dorobek naukowy dr inż. Krzysztofa Zatwarnickiego dotyczy modelowania, analizy, projektowania i badań symulacyjnych rozproszonych systemów webowych, systemów zarządzania obsługą żądań w systemach webowych, jakości usług systemów webowych oraz projektowania dziedzinowych systemów informatycznych.

Na dorobek naukowy i techniczny Kandydata składają się publikacje naukowe, zgłoszenia patentowe, udział w realizacji projektów naukowo-badawczych oraz udział w organizacji konferencji naukowych.

Kandydat jest autorem i współautorem 65 publikacji, w tym 57 opublikowanych po uzyskaniu w 2003 roku stopnia naukowego doktora nauk technicznych w dyscyplinie informatyka.

Na wspomniane 57 publikacji (w tym 19 w czasopismach i wydawnictwach znajdujących się na ministerialnej liście czasopism punktowanych) składają się:

- 1 monografia (monografia habilitacyjna) w języku polskim,
- 3 artykuły (w tym 1 samodzielna) w czasopismach wyróżnionych w *Journal Citation Reports*, tym 1 praca w czasopiśmie *Cybernetics and Systems*, 1 praca w czasopiśmie *International Journal of Applied mathematics and Computer Science* i 1 praca w serii *Lecture Notes in Computer Science*,
- 16 prac (w tym 8 samodzielnych) prac w recenzowanych czasopismach krajowych lub zagranicznych wymienianych w wykazie ministra, w tym 4 w czasopismach krajowych (*Systems Science, Pomiary Automatyka Kontrola, Theoretical and Applied Informatics, Studia Informatica*) i 12 w seriach Springera: *Lecture Notes in Computer Science* i *Lecture Notes in Artificial Intelligence*,
- 2 referaty naukowe (w tym 1 samodzielny) na krajowych międzynarodowych konferencjach naukowych,
- 1 referatu (współautorskiego) na krajowej konferencji naukowej,
- 2 innych publikacji recenzowanych (w tym 1 samodzielna) (Zeszyty Naukowe Politechniki Opolskiej),
- 13 rozdziałów w monografiach w języku angielskim (w tym 5 samodzielnych) (materiały konferencyjne),
- 16 rozdziałów w monografiach w języku polskim (w tym 5 samodzielnych) (materiały konferencyjne).

Kandydat jest autorem jednego samodzielnego (w 2009 roku) i dwóch współautorskich zgłoszeń patentowych (w 2007 roku). Zgłoszenia patentowe dotyczą, odpowiednio: szeregowania żądań w systemach webowych, szacowania czasu odpowiedzi na żądania w systemie webowym oraz dystrybucji treści w systemach webowych. Przedmiotem zgłoszeń patentowych są rozwiązania przedstawiane w monografii habilitacyjnej.

Dorobek publikacyjny Kandydata jest spójny merytorycznie. Opublikowane prace dotyczą adaptacyjnego modelu opóźnienia w systemach webowych oraz jego wykorzystania do rozwiązywania zadań rozdziału żądań, różnych koncepcji dostarczania jakości usług webowych w lokalnych i rozproszonych systemach webowych, projektowania i symulacji systemów webowych oraz analizy sieci komputerowych.

Większość opublikowanych prac Kandydata - ściśle związanych tematycznie z przedmiotem rozprawy habilitacyjnej - została przywołana i syntetycznie zrelacjonowana w monografii habilitacyjnej.

Prace Kandydata były cytowane 80 razy (według Google Scholar) i 44 razy (według Web of Science). Indeks Hirscha wynosi 4 (według Web of Science) i 5 (według Google Scholar). Impact Factor publikacji dr inż. Krzysztofa Zatwarnickiego wynosi 1,444.

Kandydat uczestniczył – jako główny wykonawca i wykonawca – w realizacji dwóch projektach badawczych, odpowiednio w realizacji projektu pt. *Metody i algorytmy podejmowania decyzji w systemach webowych z jakością usług* (projekt badawczy własny w latach 2007 - 2009) i projektu pt. *Nowe technologie informacyjne dla elektronicznej gospodarki i społeczeństwa informacyjnego oparte na paradygmacie SOA* (Program Operacyjny Innowacyjna Gospodarka w latach 2009 - 2012). Wyniki prac badawczych i technicznych, realizowanych przez Kandydata we wspomnianych projektach, zostały wykorzystane w monografii habilitacyjnej.

Dr inż. Krzysztof Zatwarnicki, zatrudniony od 1999 w Politechnice Opolskiej, od lat współpracuje z Instytutem Informatyki Politechniki Wrocławskiej. Przejawami tej współpracy są m.in. udział w realizacji wspomnianych wcześniej projektów badawczych koordynowanych przez Politechnikę Wrocławską i organizacji kilku kolejnych edycji międzynarodowej konferencji *Information Systems Architecture and Technology*.

Kandydat nie ma żadnych doświadczeń w prowadzeniu międzynarodowej współpracy naukowo-badawczej.

Kandydat posiada doświadczenie w administrowaniu i we wdrażaniu komercyjnych systemów informatycznych pozyskane w czasie pracy i współpracy z przedsiębiorstwami.

Kandydat był recenzentem kilku prac zgłaszanych do czasopisma *International Journal of Applied Mathematics and Computer Science* oraz referatów zgłaszanych na konferencje: *International Conference on Computational Collective Intelligence Technologies and Applications* i *IEEE International Conference on Emerging Technologies and Factory Automation*.

Uzupełnieniem dorobku naukowego i naukowo technicznego Kandydata jest Jego dorobek dydaktyczny – zgromadzony w szczególności w czasie pracy w Politechnice Opolskiej nieprzerwanie od 1999 roku - obejmujący prowadzenie autorskich kursów w zakresie inżynierii oprogramowania, systemów operacyjnych, sieci komputerowych oraz analizy i modelowania systemów we wszystkich ich formach, prowadzenie prac dyplomowych (magisterskich i inżynierskich) na kierunku informatyka, udział w programowaniu dydaktyki i przygotowywaniu programów nauczania na kierunku informatyka oraz organizację i kierowanie laboratorium PlanetLab.

Uważam, że dorobek naukowy i naukowo-techniczny dr inż. Krzysztofa Zatwarnickiego spełnia wymagania stawiane kandydatom do stopnia naukowego doktora habilitowanego i że może być uznany jako upoważniający do wystąpienia o nadanie stopnia naukowego doktora habilitowanego nauk technicznych w dyscyplinie informatyka.

III. Podsumowanie

Podsumowując ocenę rozprawy habilitacyjnej i ocenę dorobku naukowego uważam, że dr inż. Krzysztof Zatwarnicki spełnia wymagania formalne przewidziane w Ustawie o tytule naukowym i o stopniach naukowych.

Wnioskuję o dopuszczenie dr inż. Krzysztofa Zatwarnickiego do dalszych etapów przewodu habilitacyjnego.