

Modelowanie ruchu webowego dla potrzeb badania wydajności biznesowego serwisu WWW

Grażyna Suchacka

Wydział Elektrotechniki, Automatyki i Informatyki Politechniki Opolskiej

Artykuł dotyczy problemu modelowania ruchu webowego charakterystycznego dla serwisów WWW obsługujących aplikacje handlu elektronicznego typu B2C (ang. *Business-to-Consumer*), które w artykule określane są jako serwisy biznesowe. Problem ten wiąże się z szerszym zagadnieniem, jakim jest ocena i prognozowanie wydajności serwisów WWW na drodze eksperymentów symulacyjnych. Przedyskutowano specyfikę obciążenia biznesowych serwisów webowych, a następnie przedstawiono proponowany model obciążenia, opracowany na podstawie dostępnych w literaturze wyników badań charakteryzujących obciążenie rzeczywistych serwisów WWW. Zaproponowano połączenie modelu sesji użytkownika na witrynie B2C z modelem obciążenia na poziomie żądań HTTP, co pozwala na generowanie ruchu webowego charakterystycznego dla witryn B2C o dużej zmienności natężenia, co potwierdzają przedstawione wyniki badań symulacyjnych.

Słowa kluczowe: serwis WWW, handel elektroniczny, B2C, ruch webowy, eksperyment symulacyjny

Ostatnie lata przyniosły dynamiczny rozwój technologii teleinformatycznych, a zwłaszcza Internetu. Szczególnie intensywnie rozwijana jest obecnie najpopularniejsza usługa internetowa, World Wide Web (w skrócie WWW albo Web). Wiele aspektów codziennego funkcjonowania doczekało się implementacji webowych, wciąż zyskując rosnące rzesze zwolenników. Web okazał się szczególnie obiecującą platformą dla handlu elektronicznego, pozwalając użytkownikom Internetu na wyszukiwanie informacji o produktach, porównywanie cen oraz zamawianie produktów i usług za pośrednictwem globalnej sieci. Sklepy internetowe, zorganizowane jako witryny B2C, stały się atrakcyjną alternatywą dla tradycyjnych sklepów, pozwalając w dużym stopniu na przezwyciężenie ograniczeń związanych z czasem i przestrzenią.

Kluczowym czynnikiem sukcesu każdego sklepu internetowego jest wydajny serwis webowy obsługujący jego witrynę internetową. Serwis zbudowany jest zwykle z jednego lub wielu serwerów webowych, wspomaganych przez serwery zaplecza. Zadaniem serwisu webowego jest odbieranie żądań HTTP napływających od klientów przez Internet, generowanie odpowiedzi HTTP i wysyłanie ich do klientów. W związku z ograniczonymi zasobami sprzętowymi każdego serwisu, a także bardzo zmiennym i nieprzewidywalnym ruchem webowym, powstaje problem jakości usług ośrodków webowych QoWS (ang. *Quality of Web Service*). Polega on na występowaniu okresów przeciążenia serwisu webowego, podczas których czasy odpowiedzi serwisu dla żądań HTTP znacznie rosną, co z perspektywy użytkowników przejawia się zbyt długimi czasami oczekiwania na wyświetlenie żądanych stron w przeglądarce internetowej, a nawet brakiem odpowiedzi.

W związku z problemem QoWS prowadzone są badania nad charakterystyką obciążenia serwisów webowych, a także proponowane są nowe mechanizmy jakości usług, oparte na algorytmach klasyfikacji, kontroli przyjęć i szeregowania żądań. Jednocześnie rośnie zapotrzebowanie na modele i narzędzia umożliwiające ocenę i prognozowanie wydajności serwisów webowych. Modele analityczne mają ograniczone zasto-

sowanie do oceny wydajności serwisów WWW, przede wszystkim ze względu na ogromną złożoność tych systemów oraz ich obciążenia. Dlatego zwykle wykorzystuje się modele symulacyjne, umożliwiające odtworzenie przebiegu interakcji klientów z serwisem. Przeprowadzenie symulacji w kontrolowanych warunkach daje możliwość oceny zachowania konkretnego systemu w odpowiedzi na obciążenie o określonym charakterze i intensywności. Do badań można wykorzystać prototypowy serwis webowy, będący pewnym konkretnym rozwiązaniem sprzętowym i aplikacyjnym, lub też program komputerowy realizujący zdarzeniową symulację obsługi żądań w serwisie (ang. *discrete event-driven simulator*), implementujący pewien model serwisu.

Specyfika obciążenia biznesowych serwisów webowych

Można wyróżnić dwa podejścia do generowania ruchu webowego. Pierwsze polega na odtworzeniu obciążenia konkretnego serwera WWW na podstawie danych zapisanych w dzienniku (ang. *log*) serwera. W praktyce jednak dane te są trudno dostępne dla serwisów biznesowych ze względu na aspekt finansowy prowadzonego biznesu elektronicznego. Inne podejście polega na utworzeniu syntetycznego strumienia żądań na podstawie modeli matematycznych dla określonych cech obciążenia. Główne właściwości ruchu są określone w postaci rozkładów prawdopodobieństwa, dlatego można modyfikować cechy generowanego ruchu, m.in. przez modyfikowanie parametrów rozkładów. To podejście stwarza możliwość testowania badanego systemu pod kątem różnych scenariuszy obciążenia.

Wiele badań pokazało pewne charakterystyczne cechy ruchu webowego. Przede wszystkim ruch ten ma charakter „wybuchowy” (ang. *bursty*), co oznacza, że serwis może nieoczekiwanie ulec przeciążeniu na skutek otrzymania zbyt wielkiej liczby żądań w krótkim okresie czasu. Takie „wybuchy” ruchu można zaobserwować w różnych skalach czasowych – określają one „samopodobieństwo” (ang. *self-similarity*) ruchu we-

bowego i mają negatywny wpływ na wydajność serwisu.

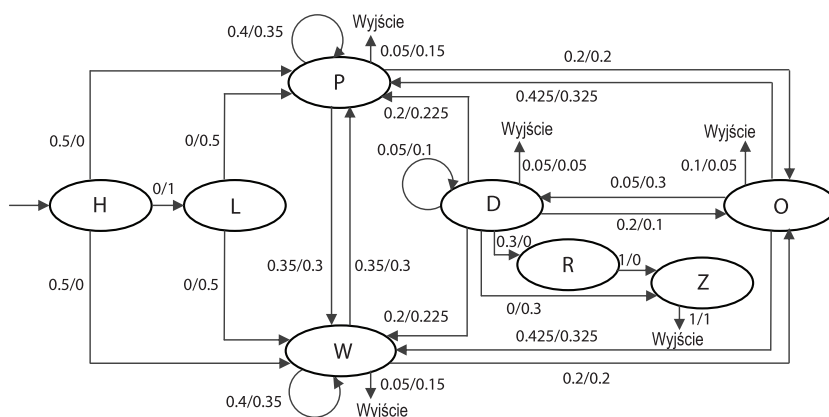
Jednocześnie zaobserwowano wzorce nawigacji użytkowników na witrynach sklepów internetowych w ramach sesji użytkowników. Sesja użytkownika definiowana jest jako sekwencja czasowo i logicznie skorelowanych żądań, nadesłanych przez użytkownika w trakcie pojedynczej wizyty na witrynie. Podczas sesji użytkownik wykonuje pewne typowe interakcje webowe – operacje biznesowe, takie jak wyszukiwanie produktów, dodawanie ich do wirtualnego koszyka zakupów itd. Z punktu widzenia właściciela e-biznesu, żądania HTTP nadchodzące do serwisu webowego mają różną wartość „biznesową” w zależności od wykonywanej przez użytkownika operacji biznesowej. Ponadto, rozpatrując problem modelowania ruchu webowego charakterystycznego dla witryn B2C, należy wyróżnić sesje użytkowników ograniczających się do przeglądania zawartości witryny oraz użytkowników zorientowanych na działania związane z zakupem produktów.

Uwzględnienie charakterystycznych cech obciążenia serwisów biznesowych oraz wzorców nawigacji użytkowników na witrynach B2C jest niezbędne przy konstruowaniu reprezentatywnego obciążenia syntetycznego, które ma być wykorzystane do oceny wydajności serwisu biznesowego na drodze eksperymentów, np. symulacyjnych. Dotychczas opracowano szereg generatorów ruchu webowego oraz tzw. benchmarków, z których najpopularniejsze to: httpperf, SPECweb99, SURGE, S-Clients, WebBench i WebStone. Programy te są jednak mało adekwatne w odniesieniu do biznesowych serwisów webowych, przede wszystkim ze względu na bardzo uproszczony model obciążenia, nieuwzględnienie sesji użytkowników ani przychodów uzyskiwanych przez właściciela e-biznesu. Z powyższych względów zdecydowano się na opracowanie modelu symulacyjnego serwisu webowego B2C na podstawie aktualnych danych literaturowych dotyczących charakterystyki obciążenia rzeczywistych serwisów tego typu. W artykule przedyskutowany został model obciążenia, zaimplementowany następnie w programie symulacyjnym umożliwiającym badanie wydajności biznesowego serwisu webowego.

Model obciążenia biznesowego serwisu webowego

Obciążenie serwisu WWW jest złożeniem wielu sesji użytkowników, którzy w danym czasie odwiedzają witrynę sklepu internetowego. W badaniach QoWS wykorzystywane są najczęściej dwa modele sesji: graf przejść stanów CBMG (ang. *Customer Behavior Model Graph*) [9] oraz diagram interakcji webowych zdefiniowany w specyfikacji TPC-W [8]. Oba modele zostały opracowane na podstawie analizy obciążenia reprezentatywnych witryn księgarń internetowych. Ze względu na rozróżnienie dwóch profili użytkownika oraz większą czytelność i łatwość rozbudowy modelu, zdecydowano się na wykorzystanie grafu CBMG.

Na podstawie danych dotyczących zachowań klientów sklepów tradycyjnych i internetowych [4, 11], definiujemy dwie klasy sesji: *OC* (ang. *ordinary customer*) i *KC* (ang. *key customer*).



Rys. 1. Model sesji użytkownika na witrynie B2C dla klasy OC/KC

Fig. 1. Model of a user session at a B2C Web site for the session class OC/KC

Proponujemy modelowanie sesji klasy *OC* i *KC* za pomocą rozbudowanego grafu CBMG odpowiednio dla profilu *occasional buyer* i *heavy buyer* (rys. 1). Wyróżniamy osiem stanów sesji: *H* – Wejście na stronę główną (ang. *home page*) witryny, *L* – Logowanie, *P* – Przeglądanie witryny, *W* – Wyszukiwanie produktów, *O* – Opis produktu, *D* – Dodanie produktu do koszyka, *R* – Rejestracja użytkownika oraz *Z* – Finalizacja transakcji zakupu.

Każdy węzeł grafu odpowiada jednemu stanowi sesji. Strzałka między dwoma stanami k i l oznacza prawdopodobieństwo przejścia $p_{k,l}$ ze stanu k do stanu l dla odpowiedniej klasy sesji. Strzałki *Wyjście* oznaczają spontaniczną decyzję użytkownika o opuszczeniu witryny. Każde przejście pomiędzy stanami k i l jest charakteryzowane przez średni czas namysłu użytkownika z perspektywy serwisu, $u_{k,l}$. Czas ten jest modelowany zgodnie z rozkładem wykładniczym, przy czym maksymalna wartość nie przekracza dziesięciokrotności wartości średniej [8]. Wartość $u_{k,l}$ wynosi 15 s dla wszystkich przejść z następującymi wyjątkami: $u_{W,O} = 30$ s, $u_{O,D} = 45$ s, $u_{D,R} = u_{D,Z} = 25$ s, $u_{R,Z} = 60$ s.

W oryginalnym grafie CBMG nie uwzględniono stanów związanych z rejestracją i logowaniem się użytkownika na witrynie, dlatego zaproponowano rozbudowanie grafu o dodatkowe stany *L* i *R*, aby w eksperymentach symulacyjnych uwzględniony był nadmiar obliczeniowy związany z wykonywaniem tych operacji. Założono, że każdy *zwykły użytkownik* (*OC*) prowadzi interakcję z witryną nie będąc zalogowanym; dopiero gdy zdecyduje się na dokonanie zakupu, musi się zarejestrować. Każdy *kluczowy użytkownik* (*KC*) loguje się od razu po wejściu na witrynę.

Kiedy nowa sesja jest inicjowana, zostaje jej przypisana klasa sesji *KC* lub *OC*, zgodnie z zadaniem parametrem Δ_{KC} określającym odsetek sesji kluczowych użytkowników w generowanym obciążeniu w oknie obserwacji.

Wszystkie sesje rozpoczynają się od stanu *H*. Każda kolejna interakcja webowa w sesji danej klasy zależy od prawdopodobieństw przejść z bieżącego stanu sesji do innych stanów, ale również od czasu odpowiedzi systemu uzyskanego dla ostatniej interakcji webowej w danej sesji. Jeżeli wydajność mocno obciążonego serwisu jest słaba i czas odpowiedzi dla strony okazał się dłuższy niż pewna wartość T_U (określająca maksymalny czas tolerancji opóźnienia strony przez użytkownika), zakładamy, że zniecierpliwiony użytkownik rezygnuje z kontynuowania wizyty na witrynie.

Przy badaniu wydajności biznesowego serwisu webowego konieczne jest uwzględnienie finansowego aspektu działań użytkownika na witrynie. Zakładamy, że każda wizyta do stanu D odpowiada dodaniu jednego produktu do koszyka zakupów. W rezultacie, liczba produktów zakupionych w ramach sesji zależy od tego, ile razy wystąpił w sesji stan D . Przyjmujemy za [10], że 20% produktów dodawanych do koszyka stanowią książki techniczne o cenie z zakresu [\$5, \$100], a pozostałe produkty to książki nietechniczne o cenie z zakresu [\$5, \$60]. Ceny generowane są zgodnie z rozkładem Gaussa ze średnią \$45 dla książek technicznych i \$18 dla książek nietechnicznych.

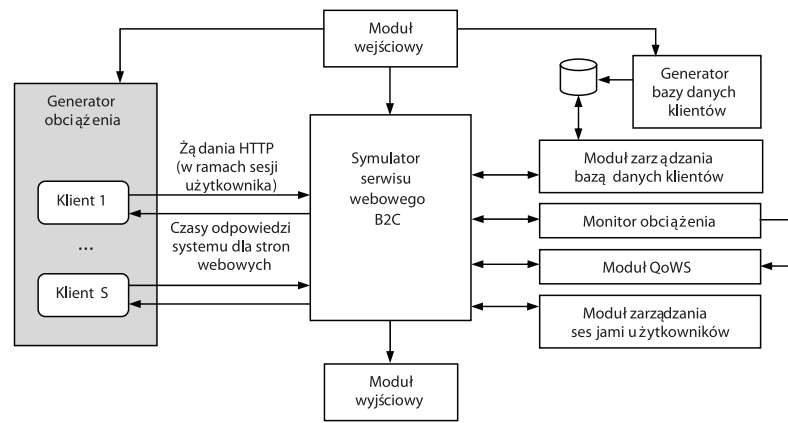
Wszystkie interakcje webowe modelujemy jako strony zawierające zagnieżdżone obiekty statyczne (jak np. obrazki) i dynamiczne (generowane na bieżąco). Przetworzenie każdej strony webowej w serwisie wymaga obsługi wielu żądań HTTP: pierwszego żądania o szkielet strony HTML oraz kolejnych żądań o obiekty zagnieżdżone na stronie. Według naszej najlepszej wiedzy, żaden z opracowanych dotychczas modeli obciążenia dla serwisów B2C nie charakteryzuje obciążenia na poziomie HTTP, dlatego zdecydowano się na połączenie wyników kilku badań obciążenia webowych serwisów biznesowych [12, 14] oraz niebiznesowych [2, 5, 6].

Tab. 1. Rozkłady prawdopodobieństwa zastosowane w modelu obciążenia

Tab. 1. Distributions and their parameters applied in a workload model

Kategoria	Rozkłady	Parametry
Liczba obiektów statycznych na stronę	Pareto	$\alpha = 1,33; k = 2$
Liczba obiektów dynamicznych na stronę	geometryczny (+1)	$p = 0,8$
Rozmiar obiektu HTML	logarytmiczno-normalny	$\mu = 7,63; \sigma = 1,001$
	Pareto	$\alpha = 1; k = 10240$
Rozmiar statycznego obiektu zagnieżdżonego	logarytmiczno-normalny	$\mu = 8,215; \sigma = 1,46$
Rozmiar dynamicznego obiektu zagnieżdżonego	Weibull	$\lambda = 0,0059; b = 0,9$
Czasy przybycia żądań zagnieżdżonych na stronie	Weibull	$\alpha = 7,64; k = 1,705$

Liczba obiektów statycznych na stronie jest modelowana zgodnie z rozkładem Pareto, a liczba obiektów dynamicznych zgodnie z rozkładem geometrycznym, zwiększana o 1 (aby zapewnić co najmniej jeden obiekt dynamiczny na stronie). Rozmiary plików HTML są uzyskiwane ze złożenia rozkładów logarytmiczno-normalnego oraz Pareto. Rozmiary zagnieżdżonych obiektów statycznych i dynamicznych są generowane odpowiednio zgodnie z rozkładem logarytmiczno-normalnym i Weibulla. Odstępów czasowych pomiędzy



Rys. 2. Architektura programu symulacyjnego

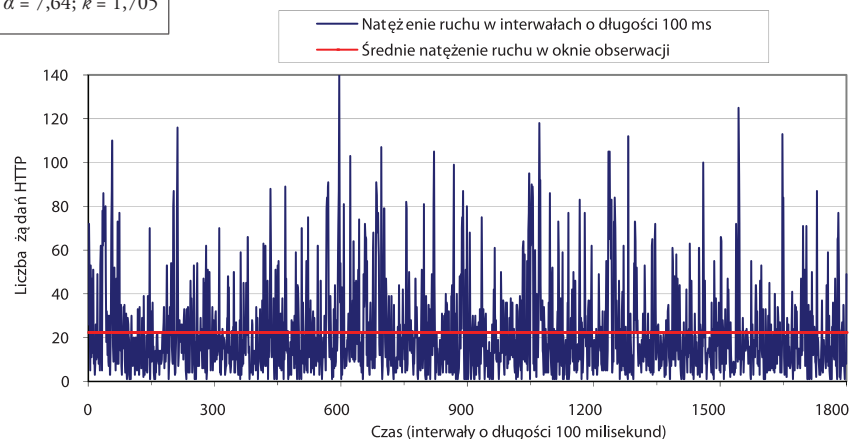
Fig. 2. Architecture of the simulation tool

chwilami przybyć żądań zagnieżdżonych na stronie są zgodne z rozkładem Weibulla (tab. 1).

Opisany wyżej model obciążenia został zaimplementowany w module generatora obciążenia programu symulującego działanie biznesowego serwisu webowego (rys. 2). Program został napisany w języku C++ przy wykorzystaniu profesjonalnego pakietu do modelowania złożonych systemów, CSIM19 [7]. Generator obciążenia wysyła do serwisu webowego sekwencję żądań HTTP, emulując zachowanie użytkowników witryny B2C. Na podstawie parametrów określonych w module wejściowym, w trakcie pojedynczego eksperymentu inicjowana jest zadana liczba λ_s nowych sesji na minutę. W eksperymentach symulacyjnych przyjęto, że $\Delta_{KC} = 10\%$ i $T_U = 8$ s. Każdy eksperyment obejmował 10-godzinny etap wstępny symulacji oraz 3-godzinny etap właściwy, dla którego rejestrowane były dane statystyczne.

Wybrane wyniki eksperymentów symulacyjnych

Przy użyciu symulatora przeprowadzono eksperymenty weryfikujące wybuchowość generowanego ruchu webowego. W trakcie pojedynczego eksperymentu inicjowana była stała liczba nowych sesji na minutę, dlatego natężenie obciążenia mierzone było dla interwałów rzędu sekund i milisekund. Wybuchowość w oknie obserwacji opisana jest parą parametrów (a ; b), gdzie a jest współczynnikiem między maksymalnym obserwowanym



Rys. 3. Natężenie generowanego ruchu w interwałach o długości 100 ms

Fig. 3. Web traffic bursts in slots of 100 ms

natężeniem przybyć żądań HTTP i średnim natężeniem w oknie obserwacji; b oznacza udział czasu, kiedy natężenie ruchu przekraczało średnią [1].

Liczby nadchodzących żądań HTTP były rejestrowane na wejściu serwisu webowego dla okresów o długości 100 ms oraz 1 s. Wyniki pokazały, że generowane obciążenie cechuje się dużą zmiennością. Dla mocno obciążonego serwisu, czyli dla $\lambda_s = 100$ sesji/min, wybuchowość ruchu dla okresów 100-milisekundowych charakteryzowana jest przez wartości (10,24; 34%). Oznacza to, że przez jedną trzecią czasu natężenie generowanego ruchu przekraczało wartość średnią, przy czym maksymalne natężenie ruchu przekraczało średnią ponad 10-krotnie. Wybuchowość ruchu dla interwałów 1-sekundowych charakteryzowana jest przez wartości (5,29; 36%). Stopień wybuchowości ruchu okazał się nieco niższy dla wyższych wartości λ_s , jednak dla wszystkich obserwowanych przypadków był znaczący. Na rys. 3 zilustrowane zostało natężenie żądań HTTP nadchodzących do serwisu w ciągu 3 min, rejestrowane w okresach 100-milisekundowych. Duże różnice w liczbie nadchodzących żądań widoczne na rysunku potwierdzają wybuchowość generowanego obciążenia.

Podsumowanie

Ocena wydajności biznesowych serwisów webowych jest złożonym problemem, przede wszystkim ze względu na trudno dostępne dane dotyczące obciążenia rzeczywistych serwisów tego typu oraz związany z tym brak odpowiednich modeli i narzędzi. Przedstawiony w artykule model obciążenia serwisu obsługującego witrynę sklepu internetowego stanowi próbę uzupełnienia luki w tym zakresie.

Wyniki przedstawione w artykule wpisują się w badania prowadzone na Wydziale Elektrotechniki, Automatyki i Informatyki Politechniki Opolskiej, dotyczące rozwoju symulacyjnych metod weryfikacji efektywności biznesowych serwisów WWW [3] oraz opracowania dla takich serwisów nowych mechanizmów jakości usług [13].

Bibliografia

1. Banga G., Druschel P.: *Measuring the capacity of a Web server*. USITS'97, Berkeley, CA, 1997, pp. 61–71.
2. Barford P. et al.: *Changes in Web client access patterns: characteristics and caching implications*. World Wide Web, Vol. 2, No. 1–2, June 1999, pp. 15–28.
3. Borzemski L., Suchacka G.: *Web traffic modeling for e-commerce Web server system*. CCIS, Vol. 39, Springer-Verlag, Berlin Heidelberg 2009, pp. 151–159.
4. Buystream: *E-Metric Research Group: Measure twice, cut once – metrics for online retailers*. www.techexchange.com/thelibrary/online_retail_metrics.html.
5. Cardellini V. et al.: *Web switch support for differentiated services*. ACM Performance Evaluation Review, Vol. 29, No. 2, 2001, pp. 14–19.
6. Casalicchio E., Colajanni M.: *A client-aware dispatching algorithm for Web clusters providing multiple services*. WWW'01, Hong Kong, 2001, pp. 535–544.
7. CSIM19, Development Toolkit for Simulation and Modeling, www.mesquite.com.

8. García D.F., García J.: *TPC-W e-commerce benchmark evaluation*. IEEE Computer, Vol. 36, No. 2, February 2003, pp. 42–48.
9. Menascé D.A. et al.: *A methodology for workload characterization of e-commerce sites*. ACM Conference on Electronic Commerce, Denver, CO, USA, November 1999, pp. 119–128.
10. Menascé D.A. et al.: *Business-oriented resource management policies for e-commerce servers*. Performance Evaluation, Vol. 42, No. 2–3, September 2000, pp. 223–239.
11. Pecaut D.K., Silverstein M., Stanger P.: *Winning the online consumer: insights into online consumer behavior*. Boston Consulting Group, March 2000, www.bcg.com.
12. Shi W., Collins E., Karamcheti V.: *Modeling object characteristics of dynamic Web content*. Journal of Parallel and Distributed Computing, Vol. 63, No. 10, October 2003, pp. 963–980.
13. Suchacka G., Borzemski L.: *Business-driven QoS management of B2C Web servers*. LNCS, Vol. 6236, ETM 2010, Springer, Heidelberg, 2010, pp. 93–100.
14. Xia C.H. et al.: *Web traffic modeling at finer time scales and performance implications*. Performance Evaluation, Vol. 61, No. 2–3, July 2005, pp. 181–201. ■

Web traffic modeling to evaluate efficiency of a business Web server system

The paper deals with the problem of modeling Web traffic which is characteristic of Web server systems hosting e-commerce B2C (Business-to-Consumer) sites. In the paper such systems are called business Web server systems. The problem is related to a broad issue of Web server system performance prediction and evaluation through simulation experiments. First, a specificity of business Web server system workload is discussed, followed by the proposal of a workload model, based on up-to-date results on real Web server workload characteristics. We propose combining a model of a user session at a B2C Web site with HTTP-level workload models both for business and non-business Web servers. Simulation results have shown that the generated Web traffic is highly variable and bursty.

Keywords: Web server, e-commerce, B2C, Web traffic, simulation experiment

mgr inż. Grażyna Suchacka

Ukończyła studia na kierunku Informatyka o specjalności Systemy Informacyjne na Wydziale Informatyki i Zarządzania Politechniki Wrocławskiej. Pracuje jako asystent w Katedrze Informatyki na Wydziale Elektrotechniki, Automatyki i Informatyki Politechniki Opolskiej. Zajmuje się zagadnieniami związanymi z jakością usług serwisów WWW, ze szczególnym uwzględnieniem aplikacji handlu elektronicznego.



e-mail: g.suchacka@po.opole.pl