



Rafał Marjasz

Polska Akademia Nauk
Instytut Informatyki Teoretycznej i Stosowanej
Zespół Internetu Rzeczy
rmarjasz@iitis.pl

OPTYMALIZACJA WYBRANYCH CHARAKTERYSTYK MODELU KOLEJKOWEGO PRZY UŻYCIU METOD DYNAMIKI SYSTEMOWEJ

Streszczenie: W artykule zaprezentowano wyniki badań symulacyjnych i optymalizacyjnych wykonanych na modelu kolejkowym typu $M^X/G/1/N$ z polityką wielokrotnych okresów przestoju (okresów bezczynności stanowiska obsługi). Model został wykonany w oprogramowaniu Vensim przy użyciu metody Dynamiki Systemowej, co pozwala na wspomaganie podejmowania decyzji w zakresie ustalenia optymalnej długości czasu trwania pojedynczego przestoju (co bezpośrednio wpływa na uzyskanie optymalnych charakterystyk całkowitego okresu bezczynności i zajętości stanowiska obsługi względem zadanej intensywności ruchu, np. strumienia pakietów). W tym celu wykorzystane zostały wbudowane w aplikację Vensim narzędzia do symulacji i kalibracji, a rezultaty badań zostały zaprezentowane w postaci wykresów prezentujących otrzymane charakterystyki. Na koniec przedstawiono możliwe zastosowania wraz z kwestiami otwartymi dla dalszej pracy badawczej.

Słowa kluczowe: teoria kolejek, symulacja, Dynamika Systemowa, Vensim.

JEL Classification: C61, C63.

Wprowadzenie

Modele kolejkowe o skończonej pojemności bufora mają szerokie zastosowanie w wielu dziedzinach nauki, spośród których wyszczególnić należy zarządzanie, ekonomię i sieci komputerowe, będące obszarem zainteresowań autora. W przytoczonych modelach można odwzorowywać szereg ograniczeń w dostępie do stacji roboczej lub zamiennie stanowiska obsługi (np. kasy w sklepie bądź router w sieci komputerowej) obsługującego kolejki zasobów (ludzi, pakie-

tów w sieciach komputerowych), co stanowi potencjał do przeniesienia rzeczywistych problemów związanych z oszczędzaniem energii (w tym elektrycznej oraz ludzkiej pracy) na grunt teorii kolejek. Traktując każdy okres obsługi przez stację roboczą (kasjerkę w sklepie, router w sieci) pojedynczego zgłoszenia (klienta, pakietu w sieci) jako czas pracy systemu oraz analogicznie każdy okres bezczynności stacji roboczej (brak klientów, pakietów w kolejce) jako czas oszczędzania energii systemu, dostrzegamy potrzebę zoptymalizowania działania modelu kolejkowego pod kątem zachowania jak najwyższej skuteczności w obsłudze zgłoszeń przy jednoczesnym minimalnym okresie zajętości stacji roboczej.

W literaturze z zakresu modeli kolejkowych istnieje wiele różnych mechanizmów dotyczących problemu oszczędzania energii, a jednym z nich jest interesująca autora polityka wielokrotnych okresów przestoju. Główną cechą tej polityki jest sukcesywne przechodzenie stacji roboczej w stan spoczynku za każdym razem, gdy wszystkie zgłoszenia oczekujące w kolejce zostaną obsłużone. Stan spoczynku składa się z wielu pojedynczych okresów bezczynności, a każdorazowo po zakończeniu takiego okresu sprawdza się, czy w międzyczasie uformowała się kolejka zgłoszeń (co najmniej jedno) do obsługi – tylko wówczas stacja robocza wznawia pracę. Mechanizm powtarzalnych okresów bezczynności w modelach kolejkowych typu M/G/1 jest szerzej opisany w pracach [Kempa, Marjasz, 2018; Kempa, 2016; Mancuso, Alouf, 2012; Alouf, Altman, Azad, 2008; Seo i in., 2004; Boxma, Schlegel, Yechiali, 2002].

Optymalizacji dokonano z wykorzystaniem dwóch metod. Pierwszą jest metoda Powella, ściślej: metoda kierunków sprzężonych Powella – to algorytm zaproponowany przez M.J.D. Powella [1964] do znajdowania lokalnego minimum funkcji. Funkcja nie musi być różniczkowalna, natomiast musi być funkcją o wartościach rzeczywistych oraz ustalonej liczbie danych wejściowych. Algorytm jako parametry początkowe przyjmuje punkt początkowy, a także zestaw początkowych wektorów wyszukiwania. Kolejną wykorzystaną metodą jest metoda Monte Carlo, użyta przez autora do zamodelowania modelu kolejkowego (będącego procesem stochastycznym Markova) w celu wyznaczenia przedziałów ufności. Inspiracją takiego podejścia do problemu jest praca [Berg, 2004]. Istotną rolę w tej metodzie odgrywa losowanie (wybór przypadkowy) wielkości charakteryzujących proces.

1. Dynamika Systemowa

Pojęcie Dynamiki Systemowej (DS) związane jest na poziomie semantycznym z głównym aspektem tej metody, którym jest ujęcie dynamiki świata w odwzorowującym ją modelu symulacyjnym. W celu przybliżenia zagadnienia autor

odwołuje się do wymienionych w literaturze książek opisujących tę tematykę. W książce [Łukaszewicz, 1975] w zostały wyróżnione trzy podstawowe elementy modelowanych obiektów:

- poziom – będący stanem wyróżnionego elementu systemu/zasobu (np. liczba klientów, pakietów oczekujących w kolejce na obsługę), wielkość chwilowa poziomu równa jest zakumulowanej różnicy strumieni dopływu i odpływu (ludzi/pakietów);
- strumień – określa szybkość, z jaką zasoby (ludzie, pakiety w sieci) uzupełniają bądź uszczuplają zawartość przyporządkowanych im poziomów, a wielkość chwilową strumienia reguluje stanowisko decyzyjne;
- stanowisko decyzyjne (formuła decyzyjna) – reguluje wielkość strumieni w zależności od informacji o chwilowych stanach systemu.

Podstawową strukturą najczęściej wykorzystywaną w modelach Dynamiki Systemowej są pętle sprzężenia zwrotnego zaprezentowane m.in. w [Kasperska, 2005]. Modele kolejkowe, podobnie jak większość modeli DS, składają się z wielu powiązanych ze sobą przyczyn i skutków tworzących strukturę systemu i wpływających na jego dynamikę, a każda taka składowa stanowi właśnie pętlę sprzężenia zwrotnego. Należy tutaj zwrócić uwagę na istotę dynamicznego zachowania, jaką jest samo otoczenie systemu. Przykładem może być zmienna charakterystyka napływu klientów/pakietów do stanowiska obsługi, determinująca okresy bezustannej wzmożonej obsługi, co prowadzi często do oscylacyjnego charakteru odpowiedzi systemu.

2. Model typu $M^X/G/1/N$

W teorii kolejek powszechnie stosowana jest notacja Kendalla opisująca podstawowe parametry modelu kolejkowego. Zgodnie z tą notacją poszczególne składowe prezentowanego w artykule modelu $M^X/G/1/N$ oznaczają:

- M^X – złożony proces Poissona opisujący napływ ludzi/pakietów, gdzie X oznacza zmienną liczbę osób/pakietów, które mogą wejść jednocześnie do systemu;
- G – dowolny rozkład czasu obsługi;
- 1 – jedno stanowisko obsługi;
- N – pojemność modelu kolejkowego wyrażona poprzez maksymalną liczbę ludzi/pakietów czekających w kolejce wraz z osobą/pakiem aktualnie obsługiwanym.

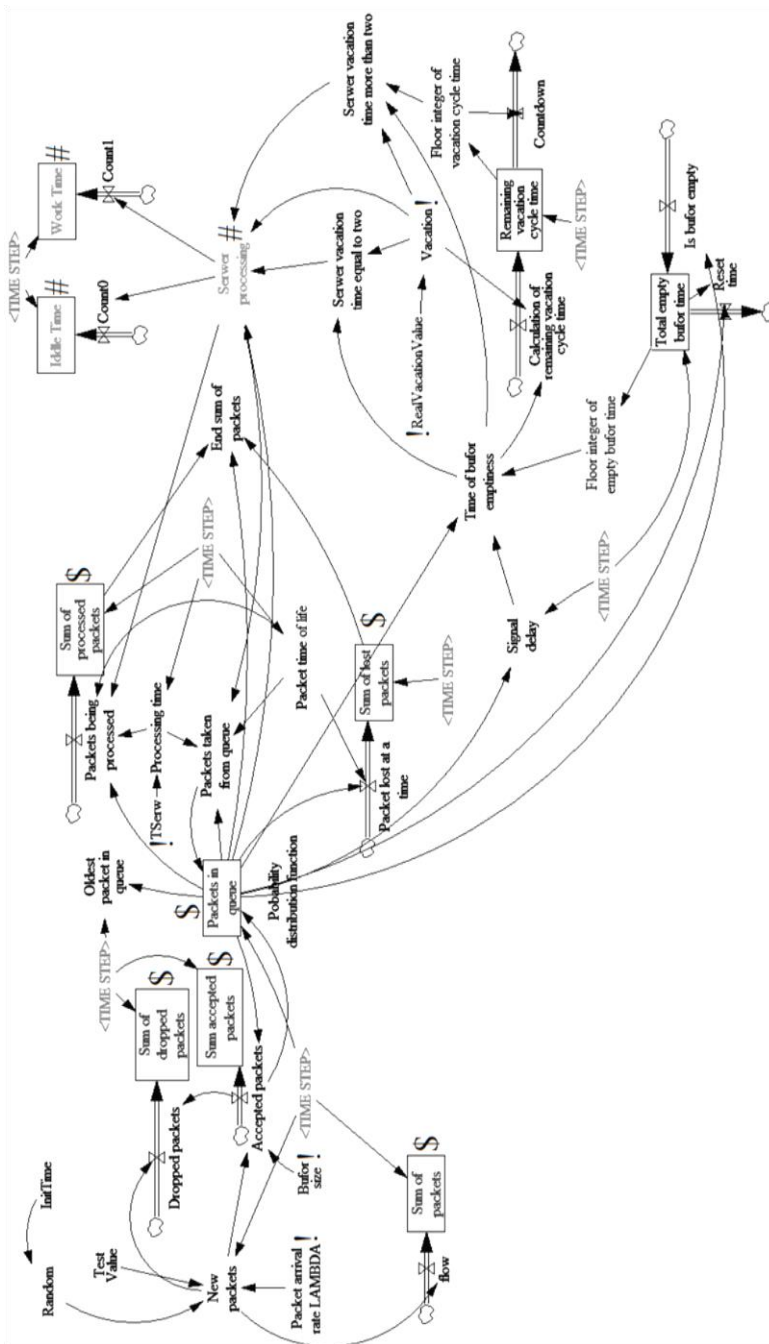
W rozpatrywanym modelu klienci/pakiety napływają do kolejki według złożonego procesu Poissona z intensywnością λ , która reprezentuje średnią liczbę

bę osób/pakietów pojawiających się w jednostce czasu. Przyjęto założenie, iż osoby/pakiety są obsługiwane pojedynczo jeden po drugim według dyscypliny FIFO (kto pierwszy wejdzie, ten pierwszy będzie obsłużony). Pojemność bufora akumulującego (kolejki) to $N-1$ osób/pakietów. W momencie obsłużenia wszystkich zgłoszeń (klientów/pakietów) stanowisko obsługi przechodzi w stan spoczynku o wielokrotnie powtarzanym pojedynczym okresie beczynności, którego stała długość jest jednym z parametrów systemu. Każdorazowo po upływie tego okresu system sprawdza, czy w kolejce oczekują zgłoszenia do obsłużenia, jeśli nie, to system rozpoczyna kolejny okres beczynności, jeśli tak – system wznowia obsługę.

3. Model kolejkowy w programie Vensim

Do symulacji użyte zostało oprogramowanie Vensim [*Vensim User's Guide*, 2019]. Strukturę modelu prezentuje rysunek 1. Dla ułatwienia w dalszej nomenklaturze ograniczono się do kolejkowania pakietów sieciowych obsługiwanych przez stanowisko obsługi / stację roboczą (którą w sieciach komputerowych może być router lub serwer), pamiętając, że analogicznie można by opisać zarządzanie kolejką w innej dziedzinie nauki (np. ludzi oczekujących na obsługę w kasie sklepowej lub informacji/dokumentów oczekujących na przetworzenie w systemach typu workflow – przykładem modelowania przepływu pracy jest kompleksowe podejście do kwestii akceptacji transakcji w firmie analizowane w pracy [Trąbka, 2013], w której kolejkowaniu podlega ścieżka wyrażania zgody na zakup). Poniżej opisane zostaną najważniejsze zmienne i parametry modelu w ujęciu teorii kolejek oraz Dynamiki Systemowej.

Obiekty wyróżnione na rysunku 1 znakiem \$ w kolejności od lewej strony stanowią poziomy (czyli chwilowe stany): sumy pakietów napływających do systemu; sumy pakietów, które nie dostały się do systemu z powodu przepełnienia bufora kolejki; sumy pakietów, które trafiły do bufora; pakietów aktualnie znajdujących się w kolejce; sumy pakietów utraconych z powodu przekroczenia maksymalnego czasu obsługi (TTL znany z sieci komputerowych); sumy pakietów obsłużonych. Wartości wymienionych poziomów stanowią będą pośrednio o jakości przeprowadzanych symulacji, gdyż oczywistym jest dążenie do sytuacji, gdy suma pakietów napływających do systemu będzie bliska bądź równa sumie pakietów obsłużonych przez stację roboczą.



Rys. 1. Struktura modelu

Źródło: Opracowanie własne wykonane w programie Vensim.

Obiekty wyróżnione znakiem ! reprezentują parametry wejściowe dla symulacji, którymi zarządzano celem otrzymania ich optymalnych wartości maksymalizujących wydajność całego systemu kolejkowego. Kolejno są nimi: intensywność napływu pakietów LAMBDA λ , rozmiar bufora (kolejki), czas obsługi przez stację roboczą, długość pojedynczego okresu przestoju (bezczynności stanowiska obsługi).

Obiekty wyróżnione znakiem # stanowią cel badań, którym jest oszczędność energii wskutek zminimalizowania czasu pracy stanowiska obsługi. Sumaryczny czas beczynności i czas pracy stacji roboczej jest akumulowany w wyszczególnionych prostokątami poziomach.

Należy nadmienić, że w badaniach uwzględniono odpowiednie zależności między parametrami i charakterystyką samego modelu. Oczywiście przypadki zyskiwania czasu beczynności stanowiska obsługi kosztem utraty pakietów wymagających obsłużenia zostały wyeliminowane.

4. Symulacje i optymalizacja

Optymalizację przeprowadzono pod kątem minimalizacji czasu pracy stacji roboczej, liczby odrzuconych, a także utraconych pakietów. W wykonywanych symulacjach przyjęto wspólne dla wszystkich założenia, iż symulowanym okresem czasu jest jedna sekunda, a krokiem czasowym jedna milisekunda. Mając na względzie zmienność losową złożonego procesu Poissona, w badaniach optymalizacyjnych przyjęto następującą strategię:

- wykorzystując wbudowaną w oprogramowanie funkcję modelującą zachowanie złożonego procesu Poissona dla zadanej intensywności λ , losowo uzyskano jeden konkretny rozkład wpływu pakietów do systemu kolejkowego względem rozpatrywanej w modelu rozpiętości czasu;
- wykonano pojedynczą symulację dla uzyskanego rozkładu wpływu pakietów, uwzględniając także pozostałe założone początkowe wartości parametrów wejściowych, stanowiącą punkt odniesienia dla procesu optymalizacji;
- w procesie optymalizacji otrzymano optymalne wartości pozostałych parametrów wejściowych względem uzyskanego na początku konkretnego rozkładu wpływu pakietów;
- przetestowano optymalność otrzymanych parametrów wejściowych poprzez wykonanie wielu symulacji z losowymi rozkładami wpływu pakietów dla zadanej intensywności λ i ocenę rezultatów tych symulacji.

Strategia w głównej mierze wykorzystuje probabilistyczny charakter złożonego procesu Poissona, implikujący wysokie prawdopodobieństwo znalezienia optymalnego rozwiązania na podstawie pojedynczego rozkładu wpływu pakie-

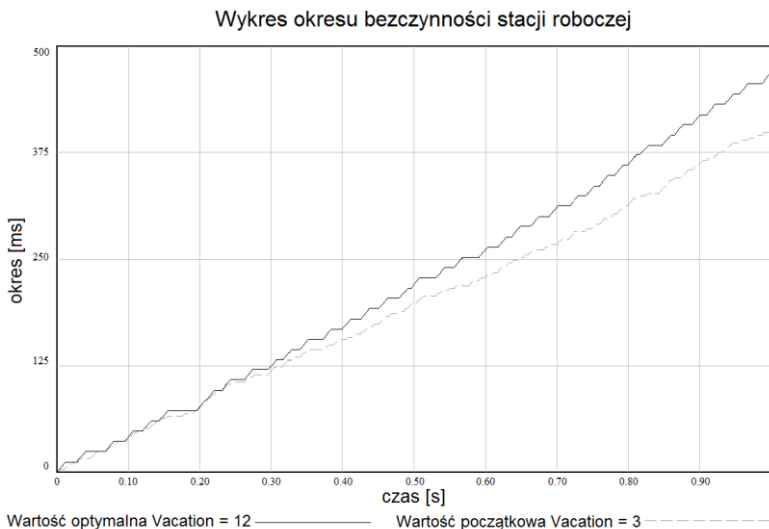
tów mającego cechę wspólną z pozostałymi w postaci intensywności λ opisującej ten proces stochastyczny. Testowanie optymalnego rozwiązania ma także na celu całkowite wyeliminowanie szansy otrzymania optimum dla możliwych skrajnych przypadków w rozkładzie wpływu pakietów (np. wszystkie pakiety sieciowe średnio napływające do systemu w pewnej jednostce czasu są zgrupowane na początku bądź pod koniec okresu czasu napływu pakietów).

5. Minimalizacja czasu pracy stacji roboczej

Parametry wejściowe modelu kolejkowego mają następujące wartości:

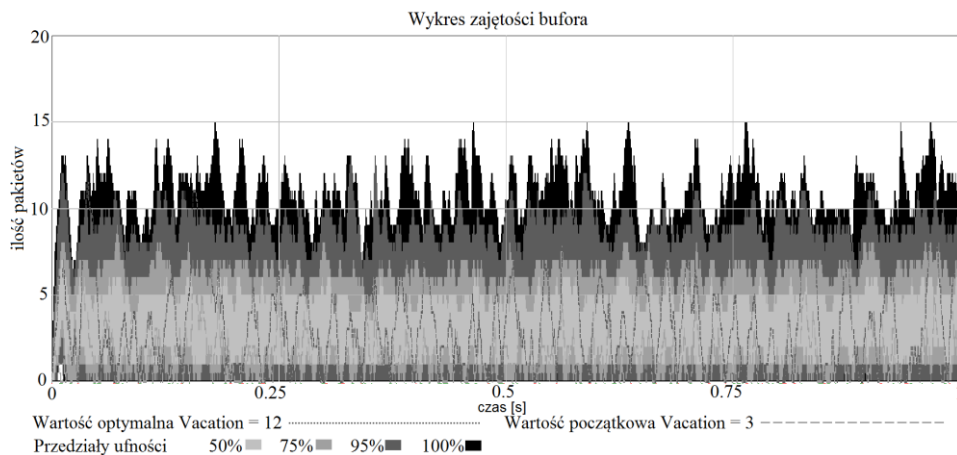
- intensywność napływu pakietów $\lambda = 500$ pakietów/sekundę,
- rozmiar bufora *Buffer size* = 15 pakietów,
- czas obsługi przez stację roboczą *TSerw* = 0,001 sekundy,
- długość pojedynczego okresu przestoju *Vacation* = 0,003 sekundy.

Wykorzystując wbudowany w oprogramowanie Vensim optymalizator z zaimplementowaną metodą Powella, wyznaczono optymalną wartość pojedynczego okresu przestoju przy jednoczesnej minimalizacji czasu pracy stacji roboczej oraz liczby odrzuconych i utraconych pakietów (rys. 2). Następnie przeprowadzono 200 symulacji testujących znalezione optimum względem losowych rozkładów wpływu pakietów o wspólnej zadanej intensywności λ i za pomocą wbudowanej funkcjonalności metodą Monte Carlo wyznaczono przedziały ufności dla zajętości bufora (rys. 3) i okresu bezczynności stacji roboczej (rys. 4).



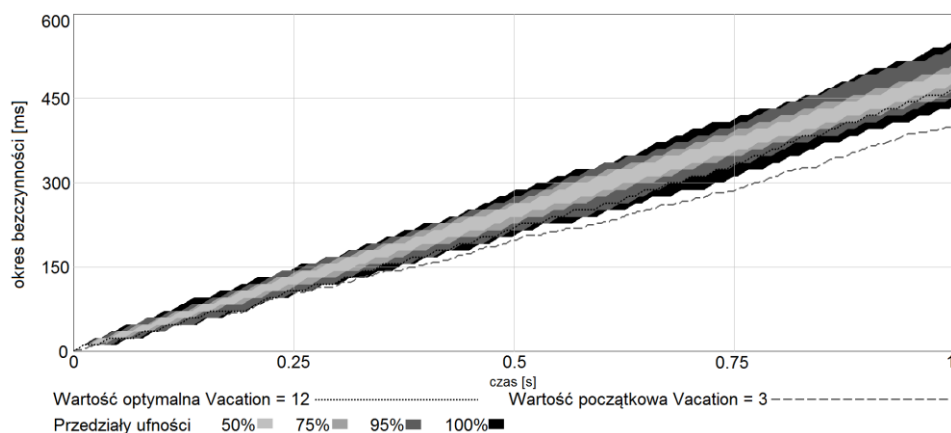
Rys. 2. Okres bezczynności stacji roboczej

Źródło: Opracowanie własne wykonane w programie Vensim.



Rys. 3. Przedziały ufności dla zajętości

Źródło: Opracowanie własne wykonane w programie Vensim.



Rys. 4. Przedziały ufności dla okresu bezczynności stacji roboczej

Źródło: Opracowanie własne wykonane w programie Vensim.

Podsumowanie

Przeprowadzone badania optymalizacyjne potwierdziły skuteczność zaproponowanej przez autora strategii. Czas pracy stanowiska obsługi w okresie jednej sekundy wzrósł średnio o 68 milisekund. Analiza wykresu przedziałów ufności na rysunku 3 wskazuje na zbliżone do idealnego wykorzystanie bufora akumulującego i kolejującego pakiety, który to tylko sporadycznie jest wypełniony

w całości bądź pusty, w większości pozostając zapełnionym w połowie swej pojemności.

Uzyskany wynik ma swoje przełożenie na pewną oszczędność energii stacji roboczej nawet w przyjętym stosunkowo niewielkim horyzoncie czasowym jednej sekundy. Kwestia ta jest bardzo istotna w szczególności w zakresie bezprzewodowych sieci komputerowych, wszelka optymalizacja zużycia energii dla urządzeń ma fundamentalne przełożenie na wydajność i sprawność m.in. projektowanych obecnie infrastruktur Internetu rzeczy inteligentnych miast. Stosowana metodyka bardzo mocno wpisuje się w dynamiczne środowisko komunikacji internetowej, a same środowisko i jego ograniczenia implikują kierunek dalszych prac badawczych. Przykładem może być technologia LPWAN (*Low Power Wide Area Networks*) wykorzystująca darmowe pasma radiowe ISM i standard LoRaWAN. Urządzenia końcowe (głównie czujniki wysyłające pomiar), komunikujące się w tym standardzie z antenami zbierającymi sygnały, są najczęściej zasilane baterią lub akumulatorem i aby wysłać dane, wybudzają się z trybu czuwania.

Kwestią otwartą pozostaje implementacja i badania optymalizacyjne nad nieuwzględnionymi w artykule cechami modeli kolejkowych (np. opóźnienie kolejkowe bądź czas wypełnienia bufora).

Literatura

- Alouf S., Altman E., Azad A. (2008), *M/G/1 Queue with Repeated Inhomogeneous Vacations Applied to IEEE 802.16e Power Saving*, "Proceedings of ACM SIGMETRICS", "Performance Evaluation Review", Vol. 36, s. 451-452.
- Berg B.A. (2004), *Markov Chain Monte Carlo Simulations and Their Statistical Analysis (With Web-Based Fortran Code)*, World Scientific, Hackensack.
- Boxma O.J., Schlegel S., Yechiali Y. (2002), *A Note on an M/G/1 Queue with a Waiting Server, Timer and Vacations*, "American Mathematical Society Translations", Series 2, Vol. 207, s. 25-35.
- Kasperska E. (2005), *Dynamika systemowa, symulacja i optymalizacja*, Politechnika Śląska, Gliwice.
- Kempa W.M. (2016), *Transient Workload Distribution in the M/G/1 Finite-Buffer Queue with Single and Multiple Vacations*, "Annals of Operations Research", Vol. 239(2), s. 381-400.
- Kempa W., Marjasz R. (2018), *Transient Queueing Delay in a Finite-Buffer Batch-Arrival Model with Constant Repeated Vacations* [w:] P. Gaj, M. Sawicki, G. Suchacka, A. Kwiecień (eds.), Proceedings of 25th International Conference, CN 2018, 19-22 June, Gliwice, Springer International Publishing, Cham, s. 311-320.

- Łukaszewicz R. (1975), *Dynamika systemów zarządzania*, PWN, Warszawa.
- Mancuso V., Alouf S. (2012), *Analysis of Power Saving with Continuous Connectivity*, "Computer Networks", Vol. 10, s. 2481-2493.
- Powell M.J.D. (1964), *An Efficient Method for Finding the Minimum of a Function of Several Variables without Calculating Derivatives*, "The Computer Journal", Vol. 7, Iss. 2, s. 155-162.
- Seo J., Lee S., Park N., Lee H., Cho C. (2004), *Performance Analysis of Sleep Mode Operation in IEEE 802.16e*, Proceedings of the 60th Vehicular Technology Conference, VTC2004-Fall, Los Angeles, Vol. 2, s. 1169-1173.
- Trąbka J. (2013), *Specific Analytical Perspectives in the Modelling of Workflow Systems* [w:] M. Pańkowska, H. Sroka, S. Stanek (eds.), *Cognition and Creativity Support Systems*, University of Economics, Katowice, s. 133-147.
- Vensim User's Guide* (2019), Ventana Systems, <http://www.vensim.com/documentation> (dostęp: 11.09.2019).

OPTIMIZATION OF SELECTED QUEUEING MODEL CHARACTERISTICS USING SYSTEM DYNAMICS METHOD

Summary: The article presents the results of simulation and optimization tests performed on a type $M^X/G/1/N$ -type queueing model with multiple vacation policy (periods of inactivity of the service station). The model was made in the Vensim software using the System Dynamics method, which allows to support decision making in determining optimal lengths of individual holiday duration in order to obtain selected optimal queueing characteristics in relation to a given traffic intensity (e.g. packet stream). For this purpose, the simulation and calibration tools built into the Vensim application were used, and the test results were presented in the form of graphs presenting the obtained characteristics. Finally, possible applications are presented along with open issues for further research work.

Keywords: queueing theory, simulation, System Dynamics, Vensim.