SCIREA Journal of Information Science and Systems Science
http://www.scirea.org/journal/ISSS
May 15, 2023
Volume 7, Issue 3, June 2023
https://doi.org/10.54647/isss120287

Simulation Models of the Bi-Level Randomized Policy and *N*-Policy for Multi-Server Queueing Systems

Yuriy Zhernovyi
Faculty of Mechanics and Mathematics, Ivan Franko National University of Lviv, Ukraine
Email: yuriy.zhernovyy@lnu.edu.ua

Abstract

We consider a multi-server *G*/*G*/*n* queue that operates the bi-level randomized (p, N_1, N_2) -policy or the *N*-policy. This means that as soon as there are no more customers in the system, the server will be shut down immediately. If the number of arriving customers falls to a particular low threshold value N_1, the server will be activated for work with a probability of p or remain turned off with a probability of $1-p$. If the number of arriving customers reaches a specified high threshold value N_2 ($\geq N_1$), the server will start serving waiting customers until the system is empty again. When $p=1$ or $p=0$ or $N_1 = N_2 = N$, the (p, N_1, N_2) -policy becomes the classic *N*-policy. Using GPSS World simulation models, we studied the dependencies of system performance measures on the following parameters: threshold values N_1, N_2 or N, the load factor, coefficient of variation of inter-arrival times, and number of servers. We validated the simulation models by comparing the results with those obtained by an analytical method. We determined the simulation time required to obtain results corresponding to the stationary process. By utilizing the created simulation models, we can solve the problem

of minimizing a long-run expected cost rate by selecting appropriate values for the thresholds N_1 and N_2.

Keywords: *queueing system, N-policy, be-level randomized policy, simulation model, GPSS World*

1. Introduction

As it is widely understood, the majority of queueing models research focuses on optimizing the design and control of queues. The primary goal of studying controllable queueing systems is to reduce operational costs and enhance overall efficiency. Broadly speaking, there are several approaches to controlling the service, including the *N*-policy developed by Yadin and Naor [1], the *T*-policy introduced by Heyman [2], and the *D*-policy outlined by Balachandran [3]. The *N*-policy queue assumes that when the number of customers in the system reaches *N*, the idle server should immediately resume service. To address the issue of customer sensitivity to delays and to improve the flexibility of queueing systems, some researchers have proposed queueing systems that use joint control policies. For instance, Lee and Seo [4] studied the *M*/*G*/1 queueing system with the dyadic Min(*N*,*D*)-policy combined with the *N*-policy and the *D*-policy. Under this system, the server resumes service when either *N* customers are in the queue or the total service time of waiting customers exceeds *D*, whichever happens first.

In many practical production systems, determining the exact threshold for starting service is crucial for operating the systems in a cost-efficient manner. If the threshold is set too low, the system will experience frequent state switching and incur significant switching costs over an extended period. Conversely, if the threshold is set too high, customers will have to wait longer, resulting in lower satisfaction and potential loss of customers. In light of the situation described earlier, the paper [5] suggests a new *M*/*G*/1 queueing model that utilizes a bi-level randomized (p, N_1, N_2)-policy. This means that as soon as there are no more customers in the system, the server will be shut down immediately. If the number of arriving customers falls to a particular low threshold value N_1 (≥ 1), the server will be activated for work with a probability of p ($0 \leq p \leq 1$) or remain turned off with a probability of $1-p$. If the number of arriving customers reaches a specified high threshold value N_2 $(\geq N_1)$, the server will start serving waiting customers until the system is empty again.

One of the methods for studying queuing systems is the simulation method, when the model simulates the operation of a real system, that is, the model reproduces the process of functioning of a real system in time. In many cases, simulation becomes the most effective and often practically the only available method for studying systems. For example, an efficient analysis of a *G*/*G*/*n* multi-server queuing system by analytical methods is impossible, while such an analysis using simulation methods is not particularly difficult [6]. In this paper, we use the GPSS World simulation system [7, 8].

GPSS (General-Purpose Simulation System) is a general process-oriented simulation software environment. GPSS World is a Microsoft Windows application designed to run on various Windows operating systems.

The main contributions of this paper are as follows.

1) We construct the GPSS World simulation models of the *N*-policy and bi-level randomized (p, N_1, N_2)-policy for the *G*/*G*/*n* multi-server queueing system, which allows us to study the dependencies of the system performance measures on the following parameters: threshold values N_1, N_2 or N, the load factor ρ, coefficient of variation V of inter-arrival times, and number of servers *n*.

2) Using the constructed simulation models, we have the opportunity to obtain not only the average values of the system performance measures but also the distributions of all performance measures, as well as their graphical representations.

3) By utilizing the created simulation models, we can solve the problem of minimizing a long-run expected cost rate by selecting appropriate values for the thresholds N_1 and N_2.

2. Simulation Models

2.1. Basic Definitions and Assumptions

We consider a *G*/*G*/*n* multi-server queueing system in which both service times and the inter-arrival times have arbitrary distributions. A random variable *X*, the time to serve a customer, has a general distribution with a cumulative distribution function (CDF) $F_X(x)$, and $F_Y(x)$ is

CDF of inter-arrival time Y. We denote as $E(X)$ and $E(Y)$ the mean of the random variables X and Y, respectively. We assume that the service is organized according to the natural FIFO discipline.

The bi-level randomized (p, N_1, N_2)-policy consists of the following. Whenever the system is empty, all servers keep dormant in the system. If the number of customers reaches N_1 (≥ 1) in the system, all the deactivated servers are turned on, with probability p ($0 \leq p \leq 1$), or are still left off, with complementary probability $1-p$. If the number of customers reaches N_2 $(\geq N_1)$ in the system, the system starts to serve customers immediately. Furthermore, once the system is activated, it will keep providing service until the system becomes empty.

When $p=1$ or $p=0$ or $N_1 = N_2 = N$, the *G*/*G*/*n* queueing system with (p, N_1, N_2)-policy is equivalent to the *G*/*G*/*n* queueing system with the conventional *N*-policy [1].

Let us denote for the random variable Y the probability density function, variance and coefficient of variation as $f_Y(t)$, $D(Y)$, and V, respectively, then for the gamma distribution, we have

$$f_Y(t) = \frac{t^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)} e^{-\frac{t}{\beta}}, \quad \alpha > 0, \quad \beta > 0, \quad t \geq 0, \quad \Gamma(t) = \int_0^{\infty} x^{t-1} e^{-x} dx;$$

$$E(Y) = \alpha\beta, \qquad D(Y) = \alpha\beta^2, \qquad V = \frac{\sqrt{D(Y)}}{E(Y)} = \frac{1}{\sqrt{\alpha}}.$$

In this paper, we use a cost structure proposed in [5], that consists of a linear waiting cost with rate *h* and a fixed startup cost *R* for each busy period. Using the renewal reward theorem, for a cycle period that is defined as the finite interval between two consecutive system busy period ending instants, and the average number of customers in the system $E(L)$, we obtain the long-run expected cost rate given by

$$F = h \cdot E(L) + \frac{R}{E(C)}.$$

Here *C* denotes a busy cycle duration, which consists of the system idle period and the system busy period, denoted by *I* and *B*, respectively. The system idle period refers to the duration of time when there are no customers present in the system. On the other hand, the system busy period is the time interval that begins with the arrival of the first customer and ends when the system becomes empty again.

We denote the waiting time of customers as W and its average value as $E(W)$.

2.2. A System with the Bi-Level Randomized (p, N_1, N_2)-Policy

Below we provide a GPSS simulation model in the case when $n=5,\ N_1=9,\ N_2=10,$ $p=0.7,\ h=15,\ R=1500,$ the random variables X and Y have the uniform and gamma distributions, respectively. We assume that $E(X)=2.5,\ E(Y)=1,\quad \rho=E(X)/(nE(Y))=0.5,$ the random variable X is uniformly distributed on the interval $[0, 5]$, the parameters of the gamma distribution of the random variable Y are as follows: $\alpha=4/9,\ \beta=9/4,\ V=1.5.$ The simulation time $t_{\text{mod}}=4\cdot10^5.$

```
*****Definition of parameters and tables
Tmod EQU 400000  ;simulation time
Sys STORAGE 5
en1 EQU 9  ;value of N1
en2 EQU 10  ;value of N2
T1 VARIABLE 0  ;initial value of I
T2 VARIABLE 0  ;initial value of C
T3 VARIABLE 0  ;initial value of B
Dis TABLE (S$Sys+Q1),0,1,100  ;distribution of L
Wtime QTABLE 1,0,1,100  ;distribution of W
Ctime TABLE X$T2,0,10,100  ;distribution of C
Btime TABLE X$Tbusy,0,5,100  ;distribution of B
Itime TABLE X$Tidle,0,5,200  ;distribution of I
*****Tabulation of L
GENERATE 1
TABULATE Dis
TERMINATE
***** At the initial time t=0, the logic switch is set to the on state
GENERATE ,,,1
LOGIC S Key
TERMINATE
*****Flow of customers and service
GENERATE (Gamma(1,0,9/4,4/9))
GATE LR Key,LQ1
TEST E (S$Sys+Q1),(en1-1),LT1
TRANSFER .3,,LT1
SAVAIL Sys
SAVEVALUE T3,AC1  ;start of busy period
```

```
SAVEVALUE Tidle,(AC1-X$T1)  ;completion of idle period
TABULATE Itime
LOGIC S Key
TRANSFER ,LQ1
LT1 TEST E (S$Sys+Q1),(en2-1),LQ1
SAVAIL Sys
SAVEVALUE T3,AC1  ;start of busy period
SAVEVALUE Tidle,(AC1-X$T1)  ;completion of idle period
TABULATE Itime
LOGIC S Key
LQ1 QUEUE 1
ENTER Sys
DEPART 1
ADVANCE (Uniform(1,0,5))
LEAVE Sys
TEST E (S$Sys+Q1),0,TER
SUNAVAIL Sys
LOGIC R Key
SAVEVALUE T2,(AC1-X$T1)
SAVEVALUE Tbusy,(AC1-X$T3)  ;completion of busy period and busy cycle
TABULATE Ctime
TABULATE Btime
SAVEVALUE T1,AC1  ;start of idle period and busy cycle
TER TERMINATE
*****Completion of the simulation
GENERATE Tmod
SAVEVALUE Cost,(15#TB$Dis+1500/TB$Ctime)  ;calculation of F
SAVEVALUE BC,(TB$Btime/TB$Ctime)  ;calculation of E(B)/E(C)
TERMINATE 1
START 1
```

The simulation model is constructed as a sequence of blocks and called block segments. Block segments generally start with a GENERATE block that inserts transactions into the simulation model and ends with a TERMINATE block that removes transactions from the simulation model. Such a block segment specifies a process, i.e., a life cycle, for transactions.

We begin the model by setting the parameters and tables of distributions for the random variables that we aim to obtain through the simulation. The STORAGE command defines a Storage Entity named Sys with a total capacity of 5 units. By changing the capacity, we can set a desired value for the number of servers. The Tables are used to calculate the distributions for several random variables, including the number of customers in the system (L), the waiting time of customers in the queue (*W*), the duration of the system busy cycle (*C*), the duration of the system busy period (*B*), and the duration of the system idle period (*I*).

The main segment forms the basis of the model and is designed to simulate the process of arrival and service of customers. The last segment sets a simulation time, saves a value of the expected cost rate function *F*, the ratio of $E(B)$ and $E(C)$, and stops the simulation process.

The main segment of the model consists of the combination of blocks ENTER and LEAVE provides the operation of a Storage Entity. The Storage Entity models the operation of a multi-server system. We use the SUNAVAIL and SAVAIL blocks to model the unavailable of the system according to the $\left(p, N_1, N_2\right)$ -policy. The GATE and TEST blocks are also involved in the implementation of this policy, as well as the logic switch Key, controlled by the LOGIC blocks. The logic switch is in the on state, if the system is available; otherwise, it remains off. The GATE block checks the state of the logic switch, while the three TEST blocks are used to verify that the number of customers in the system reaches the values of $N_1, N_2,$ and 0, respectively.

The policy "If the number of customers reaches N_1 in the system, all the deactivated servers are turned on, with probability *p* or are still left off, with complementary probability $1-p$" is implemented using the TRANSFER block, which operates in Fractional Mode and is located after the first TEST block. The second TRANSFER block operates in Unconditional Mode and directs transactions to the QUEUE block.

The QUEUE and DEPART blocks are used to update the statistics associated with a queue. Transactions (customers) are delayed for a random service time using the ADVANCE block.

To obtain the distributions of the random variable *L*, we use the GENERATE block, which creates transactions through each unit of model time and directs them to the TABULATE block associated with the table of the distribution of this random variable (Dis TABLE).

To obtain the distributions of the random variables *I*, *C*, and *B*, we use the SAVEVALUE and TABULATE blocks, as well as the arithmetic variables T1, T2, and T3. The SAVEVALUE blocks are located in such a way that transactions enter them at the start and end of the busy cycle, the busy period, and the idle period, respectively. The start time of each period is fixed using the system numerical attribute AC1.

This model offers flexibility to easily change not only the number of servers but also the probability distributions of the random variables X and Y, which are defined in the GENERATE and ADVANCE blocks.

2.3. A System with the *N*-Policy

We can obtain a simulation model for a system with the N-policy from the previous model if we substitute $N_1 = N_2 = N$ in it. To reduce a calculation time, it is advisable to use the model shown below.

```
*****Definition of parameters and tables
Tmod EQU 400000  ;simulation time
Sys STORAGE 5
en EQU 10  ;value of N
T1 VARIABLE 0  ;initial value of I
T2 VARIABLE 0  ;initial value of C
T3 VARIABLE 0  ;initial value of B
Dis TABLE (S$Sys+Q1),0,1,100  ;distribution of L
Wtime QTABLE 1,0,1,100  ;distribution of W
Ctime TABLE X$T2,0,10,100  ;distribution of C
Btime TABLE X$Tbusy,0,5,100  ;distribution of B
Itime TABLE X$Tidle,0,5,200  ;distribution of I
*****Tabulation of L
GENERATE 1
TABULATE Dis
TERMINATE
***** At the initial time t=0, the logic switch is set to the on state
GENERATE ,,,1
LOGIC S Key
TERMINATE
*****Flow of customers and service
GENERATE (Gamma(1,0,9/4,4/9))
GATE LR Key,LQ1
TEST E (S$Sys+Q1),(en-1),LQ1
SAVAIL Sys
SAVEVALUE T3,AC1  ;start of busy period
SAVEVALUE Tidle,(AC1-X$T1)  ;completion of idle period
TABULATE Itime
LOGIC S Key
LQ1 QUEUE 1
ENTER Sys
```

```
DEPART 1
ADVANCE (Uniform(1,0,5))
LEAVE Sys
TEST E (S$Sys+Q1),0,TER
SUNAVAIL Sys
LOGIC R Key
SAVEVALUE T2,(AC1-X$T1)
SAVEVALUE Tbusy,(AC1-X$T3)  ;completion of busy period and busy cycle
TABULATE Ctime
TABULATE Btime
SAVEVALUE T1,AC1  ;start of idle period and busy cycle
TER TERMINATE
*****Completion of the simulation
GENERATE Tmod
SAVEVALUE Cost,(15#TB$Dis+1500/TB$Ctime)  ;calculation of F
SAVEVALUE BC,(TB$Btime/TB$Ctime)  ;calculation of E(B)/E(C)
TERMINATE 1
START 1
```

This model contains fewer blocks compared to the previous one, specifically only two TEST blocks, since according to the *N*-policy, it is necessary to verify that the number of customers in the system reaches the values of *N* and 0, respectively.

2.4. Checking the Simulation Models and Choice the Simulation Time

Let us use analytical results for the *M*/*M*/1 system in the case of bi-level randomized $\left(p, N_1, N_2\right)$ -policy [5] to test the constructed simulation models and to choose the optimal value of the simulation time.

We assume that $N_1 = 5,\ N_2 = 12,\ E(X) = 5/14,\ E(Y) = 5/3,$ and $p = 0.6.$ Let us denote by p_k and $p_{k(sim)}$ the stationary probabilities that there are *k* customers in the system, calculated by the analytical method [5] and using the simulation model, respectively. Given that $p_k \le 10^{-6}$ for $k \ge 18,$ we evaluate the accuracy of the results using the error calculated by the formula $\Delta = \sum_{k=0}^{17} | p_k - p_{k(sim)} |.$ We obtain the following values of Δ: $\Delta = 0.0420,\ \Delta = 0.0103,$ and $\Delta = 0.0035$ for the modeling time $t_{\text{mod}} = 10^4,\ t_{\text{mod}} = 10^5,$ and $t_{\text{mod}} = 10^6,$ respectively.

The modeling time on the interval $(10^4, 10^5)$ appears to be the optimal choice for further calculations due to its high accuracy and quick implementation time.

The GPSS World uses random number generators to sample random numbers for GENERATE and ADVANCE blocks. We can select which random number generator number is to be used as the source of the random number. The results obtained for different values of the random number generator may differ slightly from each other. In this work, we use the number of the random number generator, which is equal to 1.

3. Study of the *N*-Policy and Bi-Level Randomized Policy

3.1. Basic Designations

We consider the case when the random variables X and Y have the uniform and gamma distributions, respectively. We assume that the random variable X is uniformly distributed on the interval $[0, n],$ where n is a number of servers. The coefficient of variation of the gamma distribution is related to its parameter α by the relation $V = 1/\sqrt{\alpha},$ and we have $E(Y) = \alpha\beta,$ $\rho = E(X)/(nE(Y)).$ The above equalities completely define the relationship between the parameters of the input flow and service time. In the expression of the expected cost rate function *F* we take $h = 15$ and $R = 1500.$

3.2. Minimization of Function *F* under a Constraint of Customers Waiting Time

Because extended waiting times lead to customer dissatisfaction and reduced revenue for the system, it results in a situation where nobody wins. Thus, following the work [5], our aim here is to determine the optimal policy (N_1^*, N_2^*) such that the expected cost rate function *F* is minimized under the premise that the average waiting time of customers does not exceed the predetermined threshold $W_0.$

Let us consider the case when $n = 5,$ $V = 1.5,$ $\rho = 0.5,$ $p = 0.7,$ and $t_{\mathrm{mod}} = 4 \cdot 10^5.$ Figures 1−4 show dependencies of values of $F / F_{\min}$ and $E(W)$ on N_1 for different values of $N_2.$ Here, $F_{\min} = 129.421$ and this value is achieved when $N_1 = N_2 = 7.$ The graphs show that value of F decreases together with N_1 and $N_2,$ reach a minimum, and then increase together with N_1 and $N_2.$ The values of $E(W)$ increase together with N_1 and N_2 for values N_1 and N_2 not exceeding 6, but for values exceeding 6, these dependencies may have a minimum when N_1 is small.

By using the results of calculations for F and $E(W)$, we can determine the minimum value of function F while adhering to a given constraint on the average waiting time. For instance, if we set a limit of $W_0 = 2$ for the average waiting time, then the minimum value of function $F = 129.703$ is attained when $N_1 = 7$ and $N_2 = 8$. Similarly, if $W_0 = 1.5$, the minimum value of function $F = 130.438$ is achieved when $N_1 = N_2 = 6$. If we set $W_0 = 1$, then the minimum value of function $F = 135.293$ is obtained when $N_1 = 4$ and $N_2 = 5$.

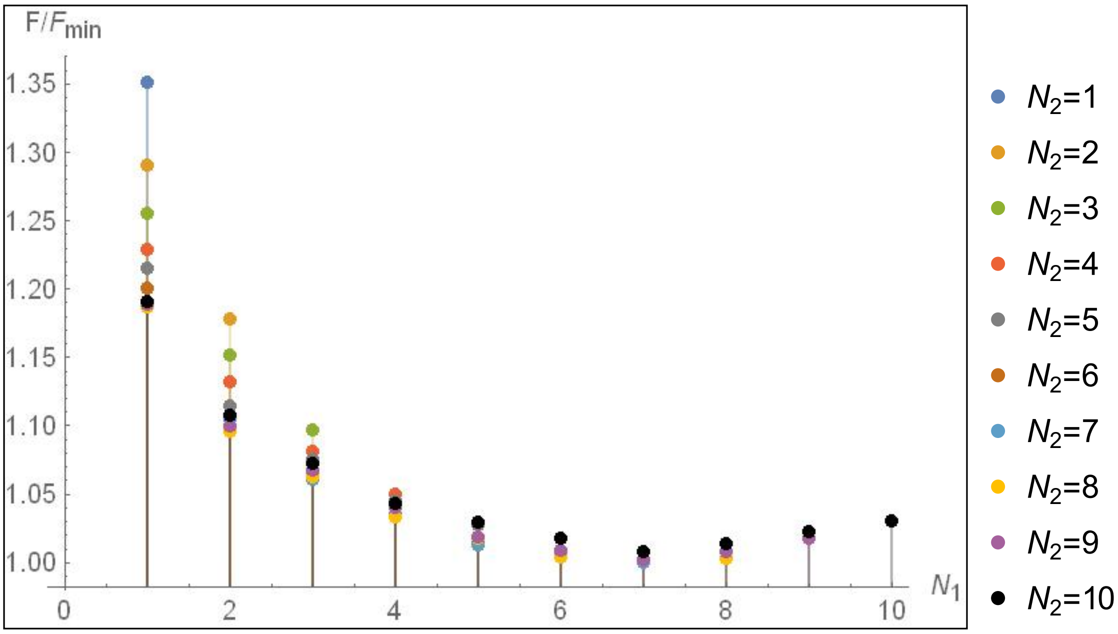

Figure 1. The dependence of $F / F_{\min}$ on N_1 for different values of N_2 in the case when $n = 5$, $V = 1.5$, $\rho = 0.5$ and $p = 0.7$

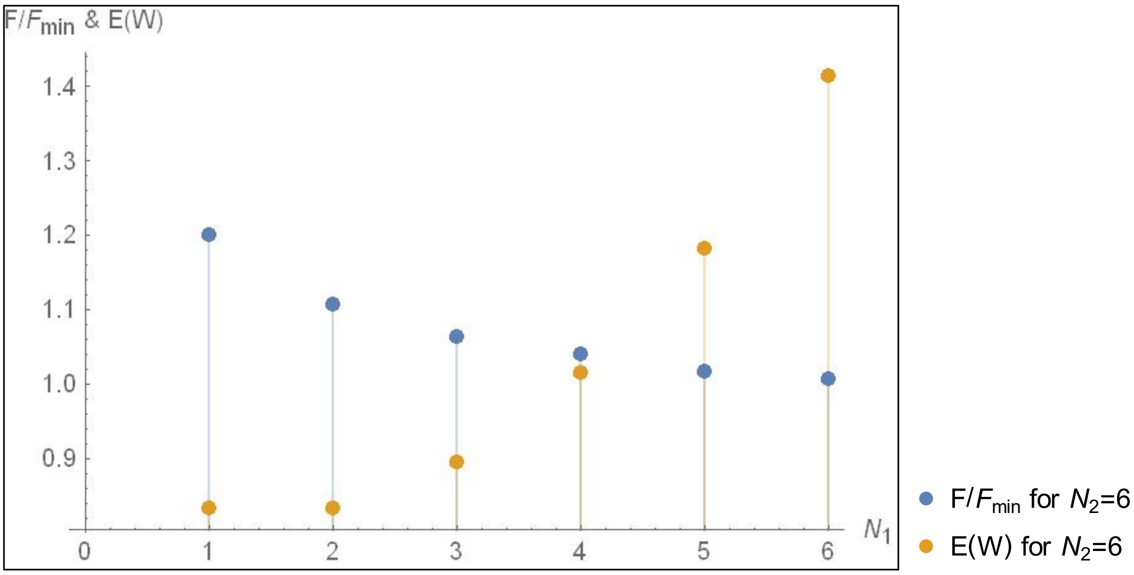

Figure 2. The dependencies of $F / F_{\min}$ and $E(W)$ on N_1 for $N_2 = 6$ in the case when $n = 5$, $V = 1.5$, $\rho = 0.5$ and $p = 0.7$

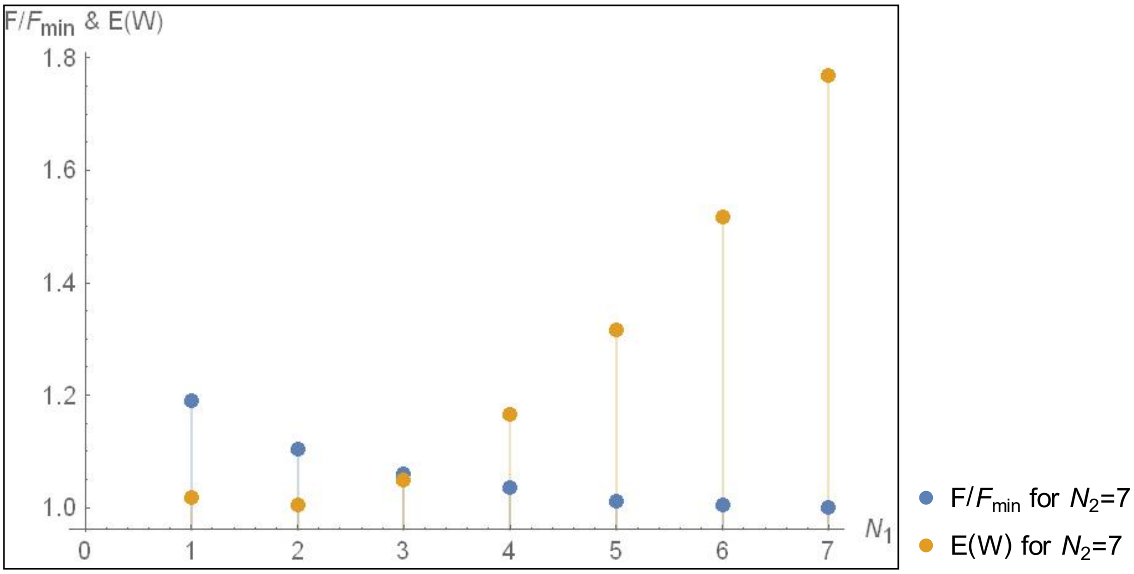

Figure 3. The dependencies of $F / F_{\min}$ and $E(W)$ on N_1 for $N_2 = 7$ in the case when $n = 5$, $V = 1.5$, $\rho = 0.5$ and $p = 0.7$

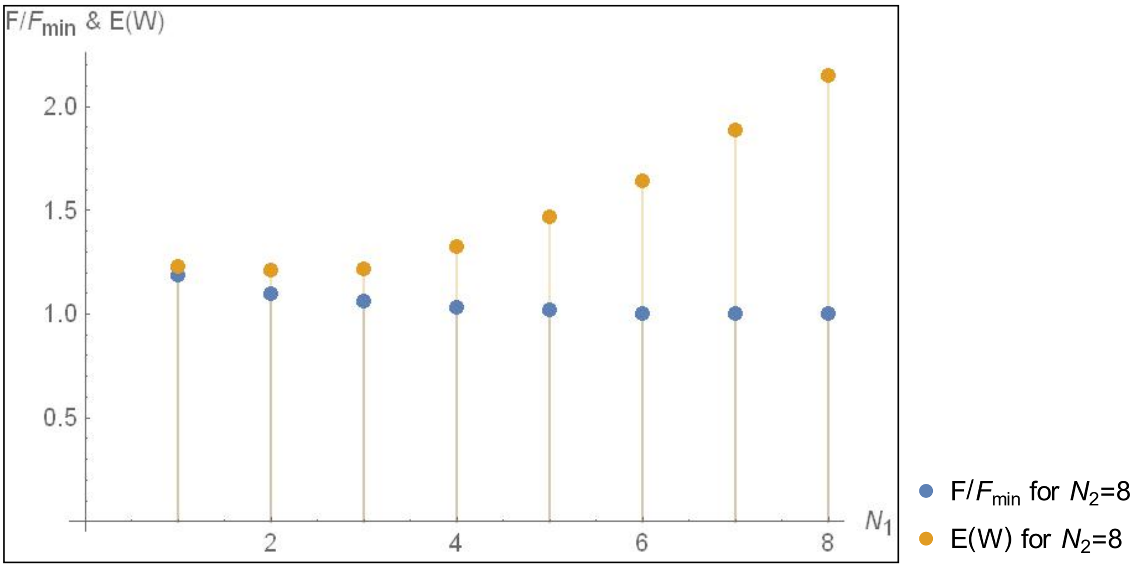

Figure 4. The dependencies of $F / F_{\min}$ and $E(W)$ on N_1 for $N_2 = 8$ in the case when $n = 5$, $V = 1.5$, $p = 0.7$ and $\rho = 0.5$

3.3. Dependencies of the System Performance Measures on *N* for Various Values of *ρ*

Let us consider the case of applying the *N*-policy when $n = 5$, $V = 1.5$, and $t_{\text{mod}} = 5 \cdot 10^5$. Figures 5−8 show dependencies of values of $E(C)$, $E(B)/E(C)$, $E(L)$ and F on N for different values of the load factor ρ.

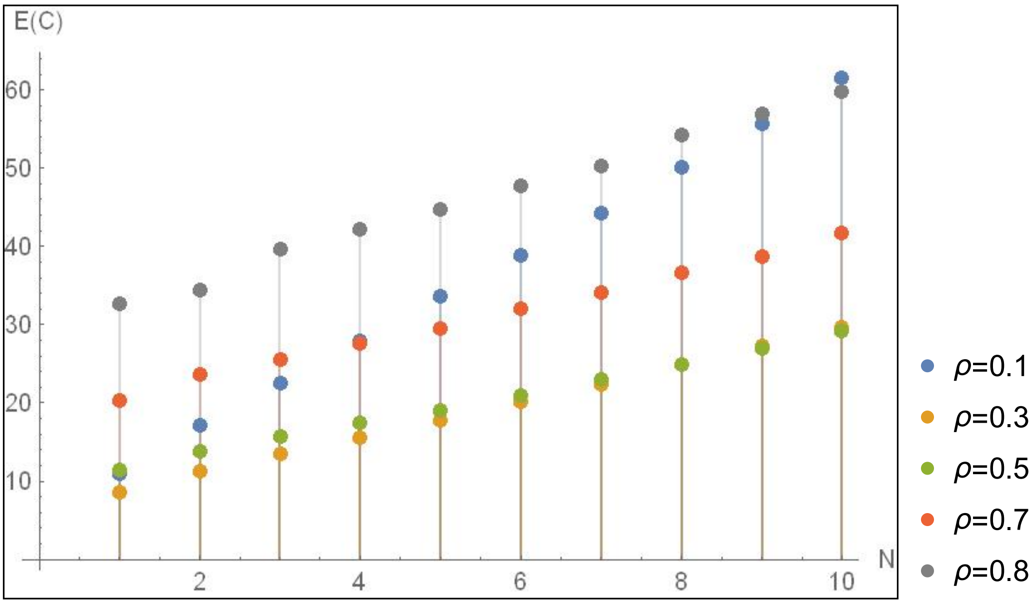

Figure 5. The dependencies of $E(C)$ on N in the case when $n = 5$ and $V = 1.5$

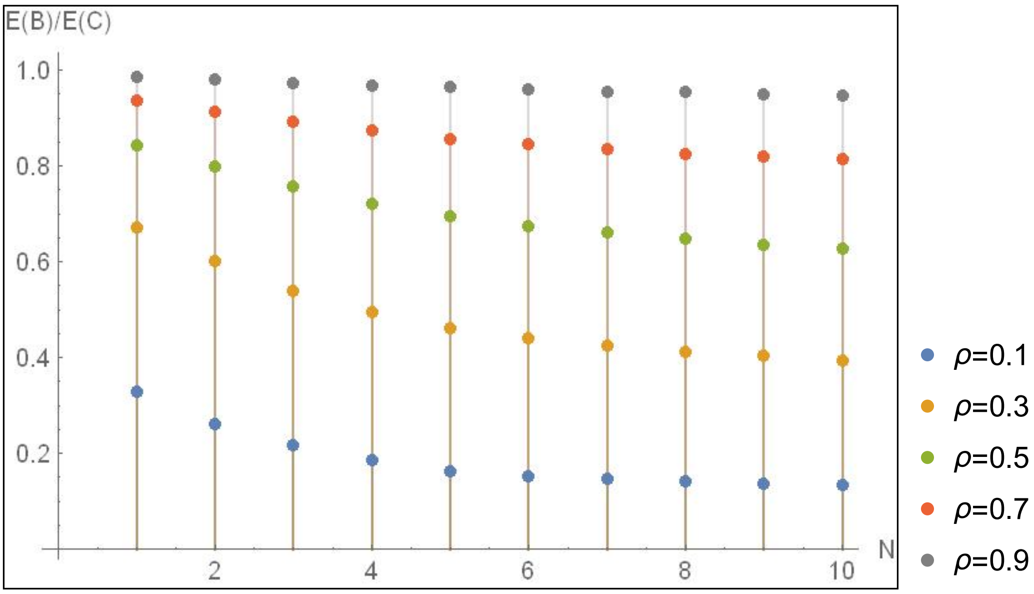

Figure 6. The dependencies of $E(B)/E(C)$ on N in the case when $n=5$ and $V=1.5$

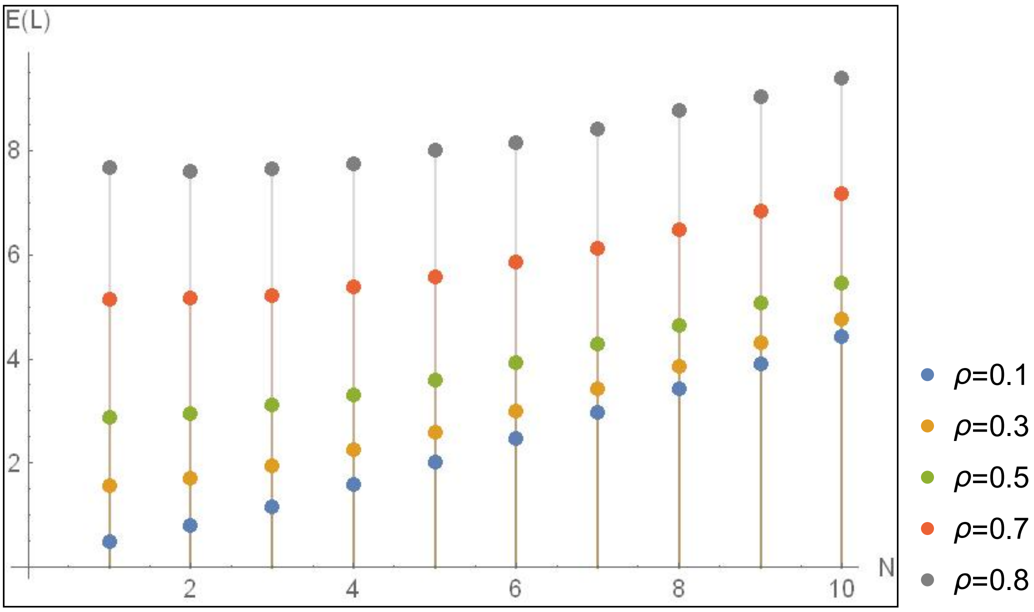

Figure 7. The dependencies of $E(L)$ on N in the case when $n=5$ and $V=1.5$

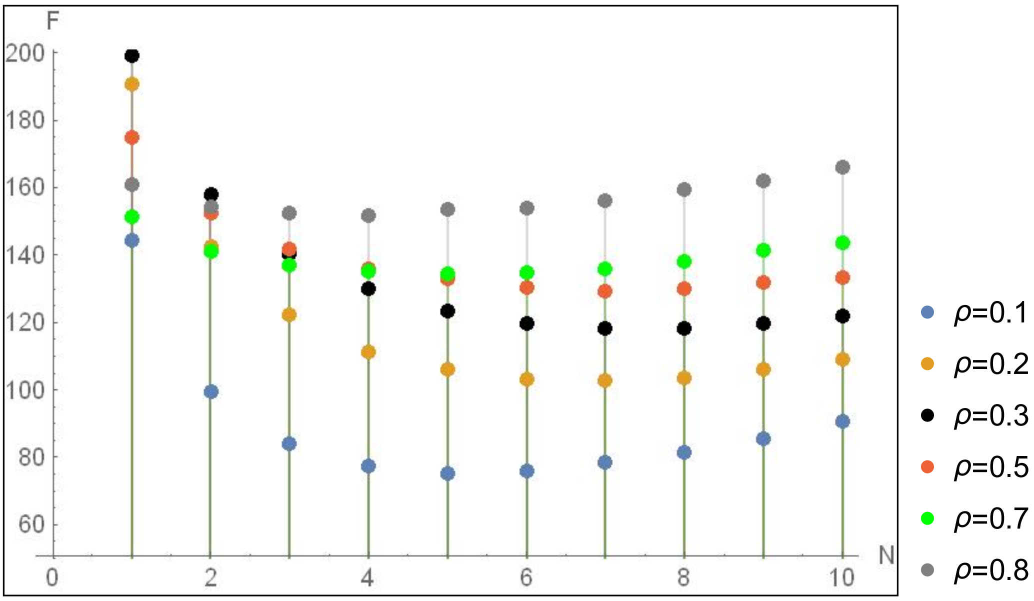

Figure 8. The dependencies of F on N in the case when $n = 5$ and $V = 1.5$

Analyzing the graphs, we can make the following conclusions: 1) as ρ increases, the values of all indicators increase; 2) the average cycle duration $E(C)$ significantly increases together with *N*, especially for $\rho = 0.1$; 3) as *N* increases, the ratio $E(B)/E(C)$ slightly decreases (where *B* is the duration of the system busy period); 4) the average number of customers in the system $E(L)$ slightly increases together with *N*; 5) the value of the function *F* decreases together with *N*, reaches a minimum, and then increases together with *N*.

3.4. Dependencies of the System Performance Measures on *N* for Various Values of *V*

Let us consider the case of applying the *N*-policy when $n = 5$, $\rho = 0.5$, and $t_{\text{mod}} = 5 \cdot 10^5$. Figures 9−12 show dependencies of values of $E(C)$, $E(B)/E(C)$, $E(L)$ and F on N for different values of the coefficient of variation of inter-arrival times, V.

Analyzing the graphs, we can make the following conclusions: 1) as V increases, the values of $E(L)$ and F increase, and the values of $E(C)$ and ratio $E(B)/E(C)$ decrease; 2) the values of $E(C)$ and $E(L)$ increases together with *N*; 3) as *N* increases, the ratio $E(B)/E(C)$ decreases; 4) the value of the function *F* decreases together with *N*, reaches a minimum, and then increases together with *N*.

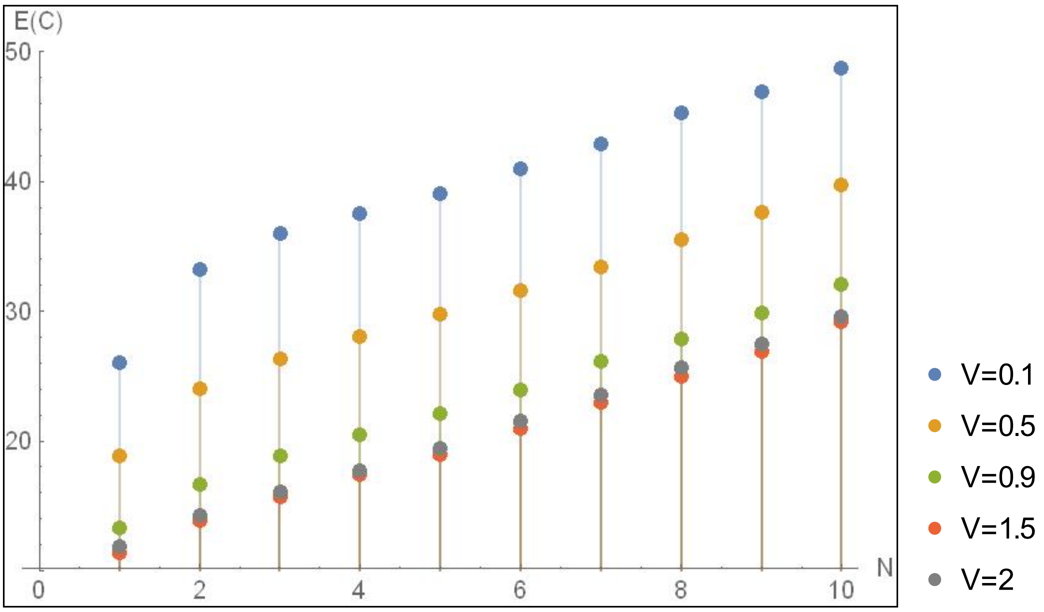

Figure 9. The dependencies of $E(C)$ on N in the case when $n=5$ and $\rho=0.5$

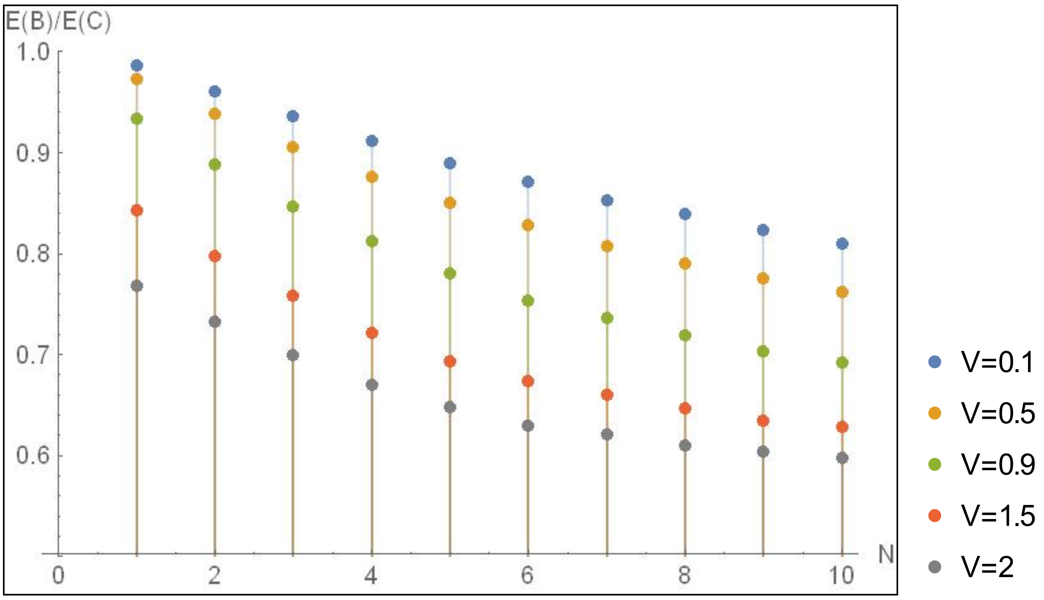

Figure 10. The dependencies of $E(B)/E(C)$ on N in the case when $n=5$ and $\rho=0.5$

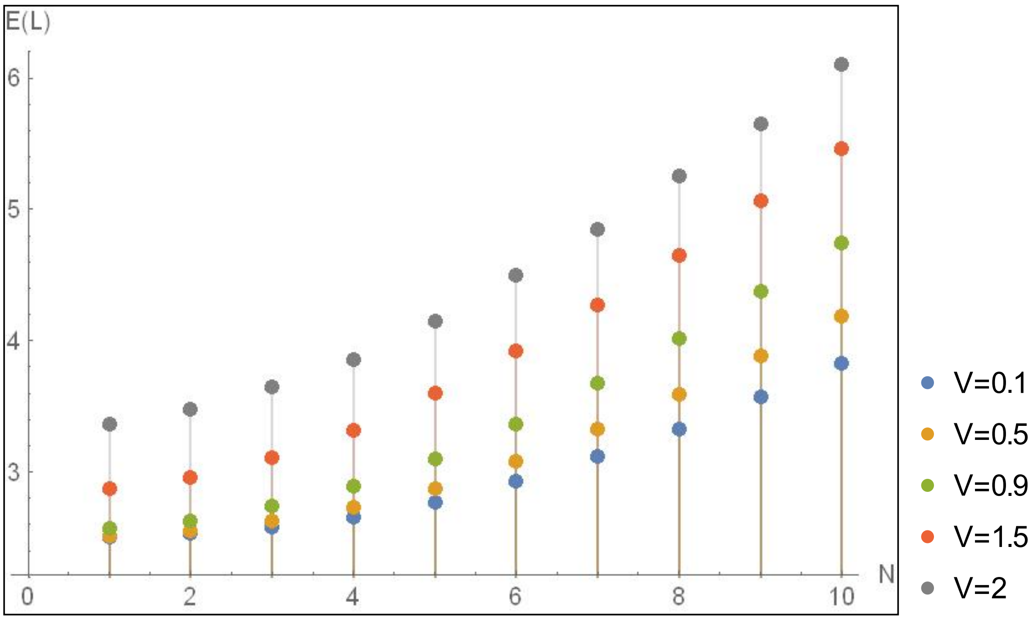

Figure 11. The dependencies of $E(L)$ on N in the case when $n = 5$ and $\rho = 0.5$

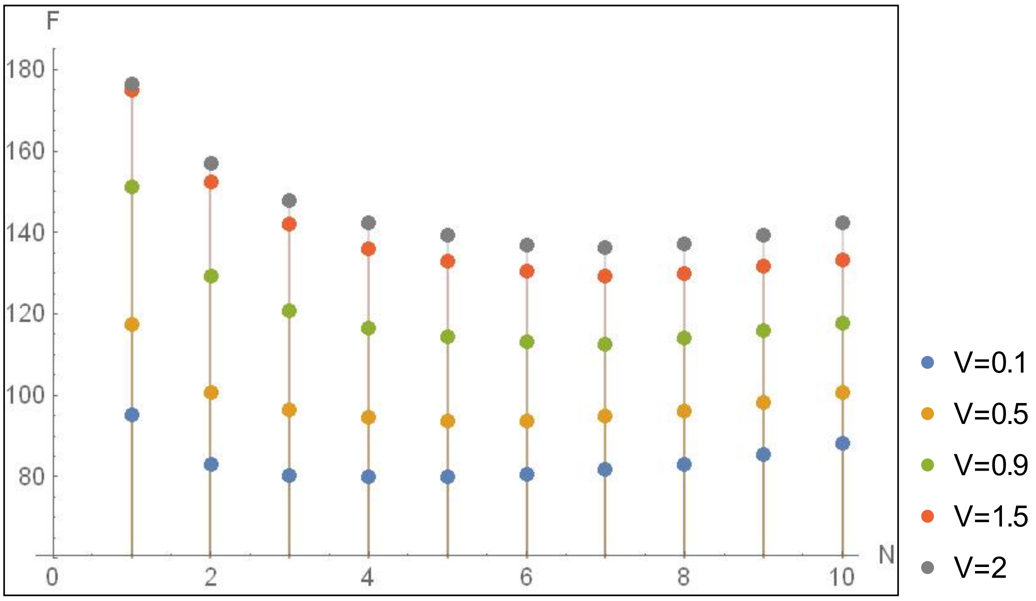

Figure 12. The dependencies of F on N in the case when $n = 5$ and $\rho = 0.5$

3.5. Dependencies of the System Performance Measures on *N* for Various Values of *n*

Let us consider the case of applying the *N*-policy when $\rho = 0.5$, $V = 1.5$, and $t_{\text{mod}} = 4 \cdot 10^5$. Figures 13−17 show dependencies of values of $E(C)$, $E(B)/E(C)$, $E(L)$, $E(W)$ and F on N for different values of the numbers of servers n.

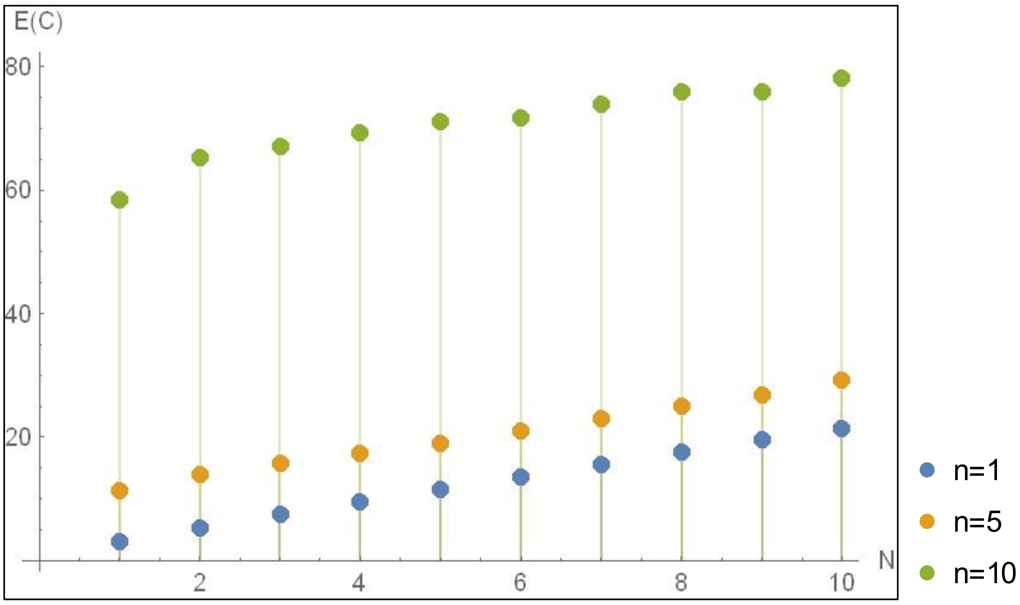

Figure 13. The dependencies of $E(C)$ on N in the case when $\rho = 0.5$ and $V = 1.5$

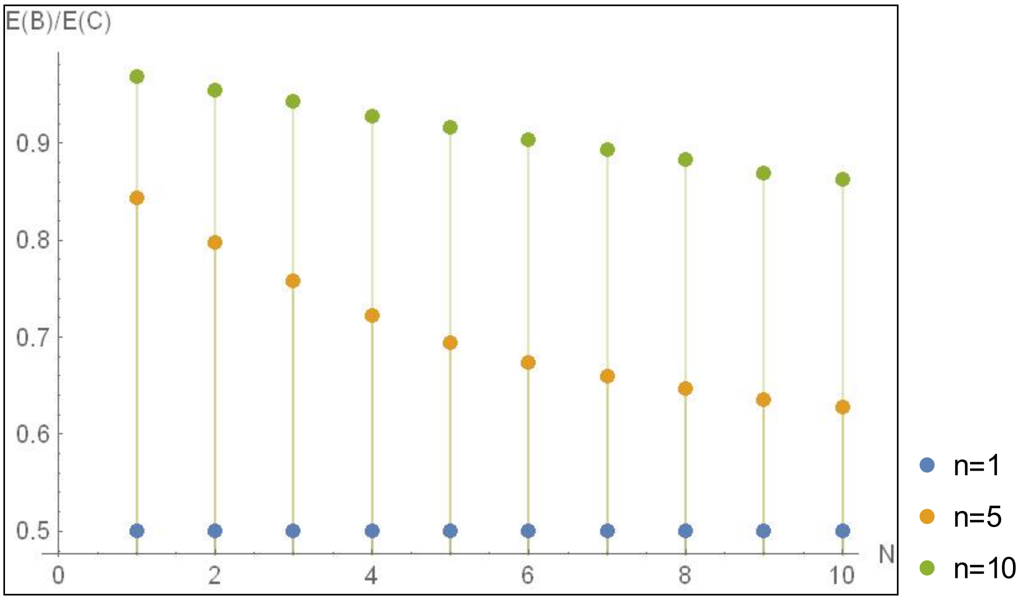

Figure 14. The dependencies of $E(B)/E(C)$ on N in the case when $\rho = 0.5$ and $V = 1.5$

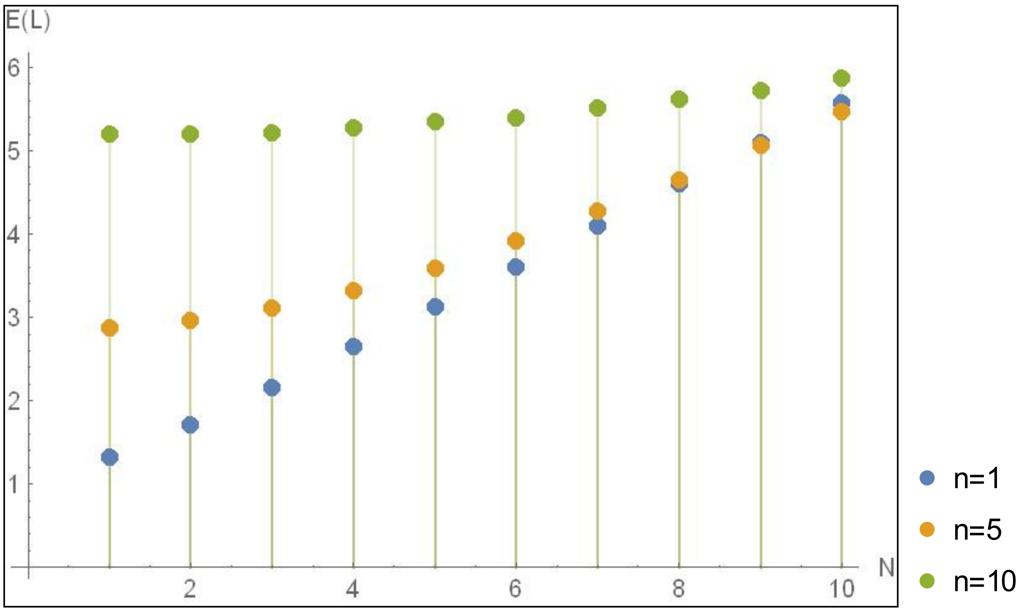

Figure 15. The dependencies of $E(L)$ on N in the case when $\rho = 0.5$ and $V = 1.5$

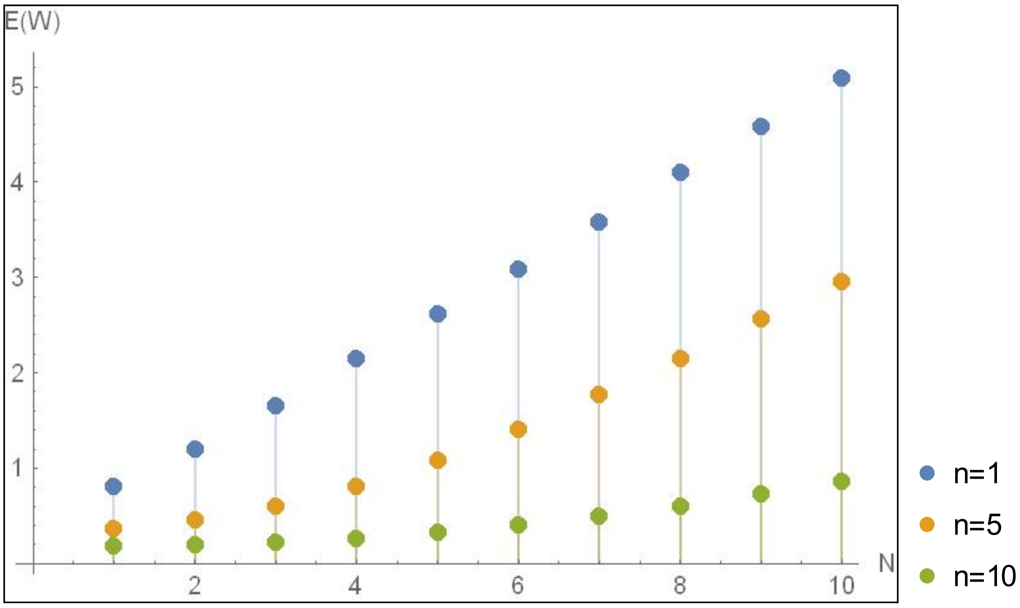

Figure 16. The dependencies of $E(W)$ on N in the case when $\rho = 0.5$ and $V = 1.5$

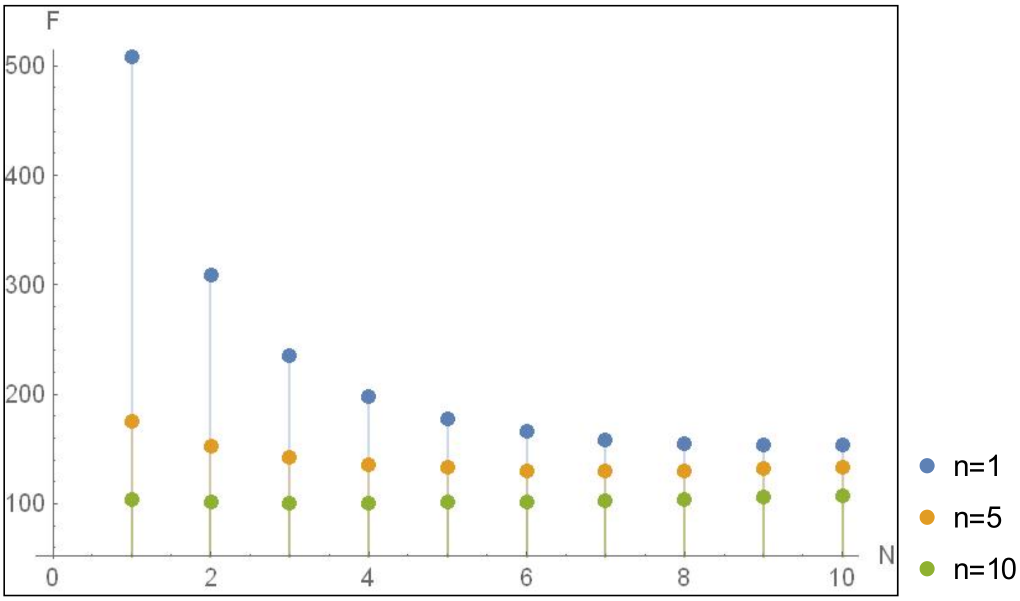

Figure 17. The dependencies of F on N in the case when $\rho = 0.5$ and $V = 1.5$

Analyzing the graphs, we can make the following conclusions: 1) as n increases, the values of $E(C)$, $E(B)/E(C)$ and $E(L)$ increase, and the values of $E(W)$ and F decrease; 2) the values of $E(C)$, $E(L)$ and $E(W)$ increases together with N, moreover, the ranges of values for $E(L)$ and $E(W)$ are the largest when $n = 1$; 3) as N increases, the ratio $E(B)/E(C)$ decreases, but for the case when $n = 1$, it practically does not change; 4) the value of the function F decreases together with N, reaches a minimum on the interval (1, 10), and then increases together with N. The range of values of the function F significantly narrows as the number of servers increases.

3.6. Examples of Obtaining Distributions of the System Performance Measures

We can get the distributions of those random variables for which tables are given in a simulation model. GPSS World tools make it possible to obtain graphic representations of distribution tables in the form of histograms and use graphs to track the dynamics of changes in random variables over time. The constructed simulation models contain tables to obtain distributions of the number of customers in the system (L), the waiting time of customers in the queue (W), the duration of the system busy cycle (C), the duration of the system busy period (B), and the duration of the system idle period (I).

Let us consider the case of applying the N-policy when $n = 5$, $\rho = 0.5$, $N = 7$, and $t_{\mathrm{mod}} = 5 \cdot 10^5$. We will analyze the cases of different coefficient of variation values for inter-arrival times, specifically $V = 0.5$ and $V = 1.5$

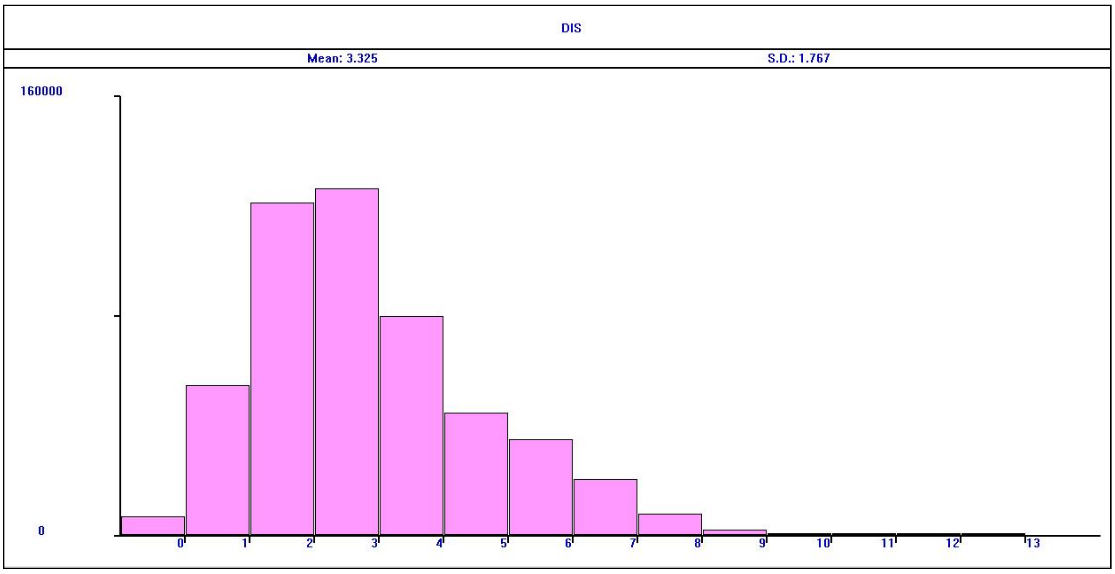

Figure 18. Distribution of L for the case when $n = 5, V = 0.5, \rho = 0.5,$ and $N = 7$

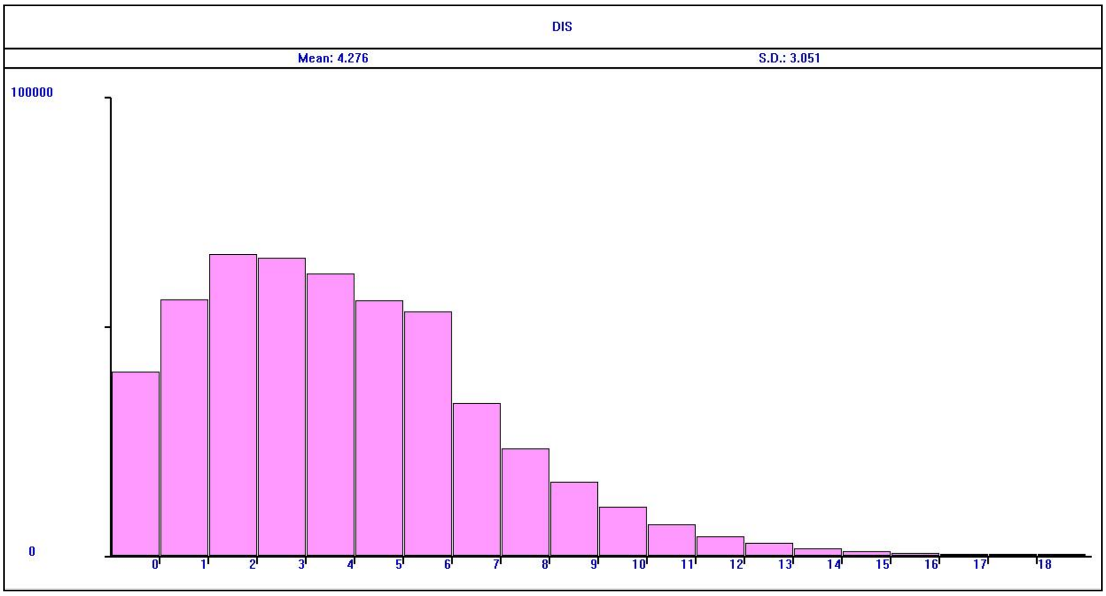

Figure 19. Distribution of L for the case when $n = 5, V = 1.5, \rho = 0.5,$ and $N = 7$

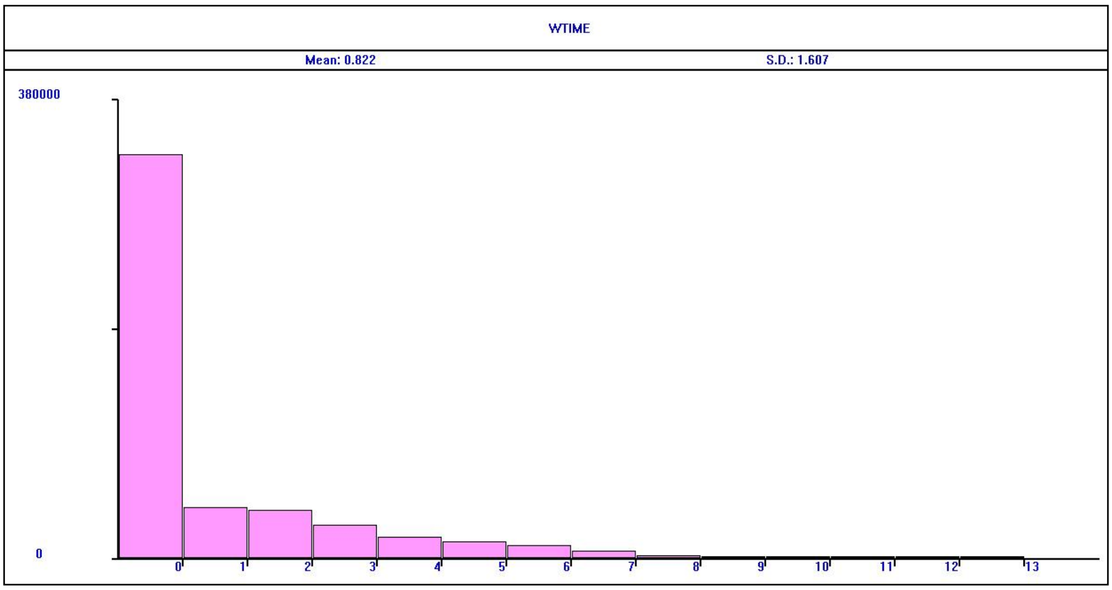

Figure 20. Distribution of W for the case when $n = 5, V = 0.5, \rho = 0.5,$ and $N = 7$

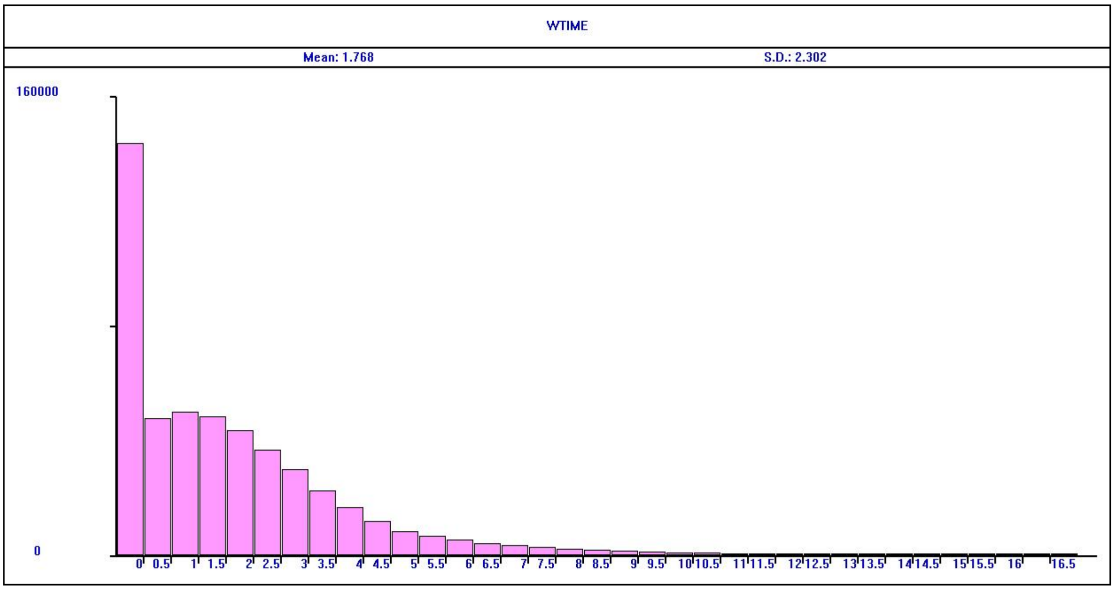

Figure 21. Distribution of W for the case when $n = 5, V = 1.5, \rho = 0.5,$ and $N = 7$

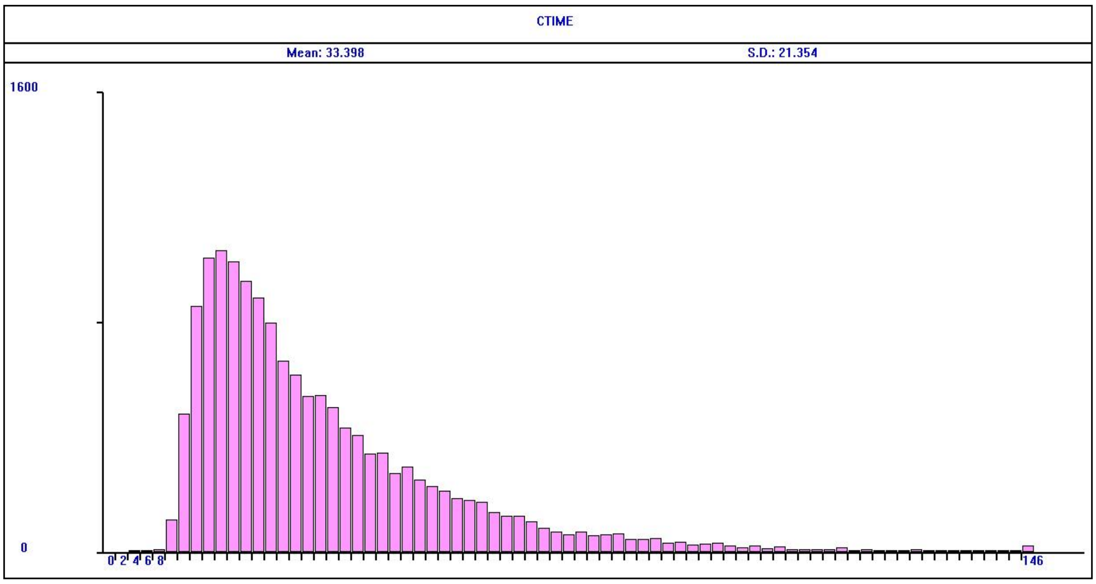

Figure 22. Distribution of C for the case when $n = 5, V = 0.5, \rho = 0.5,$ and $N = 7$

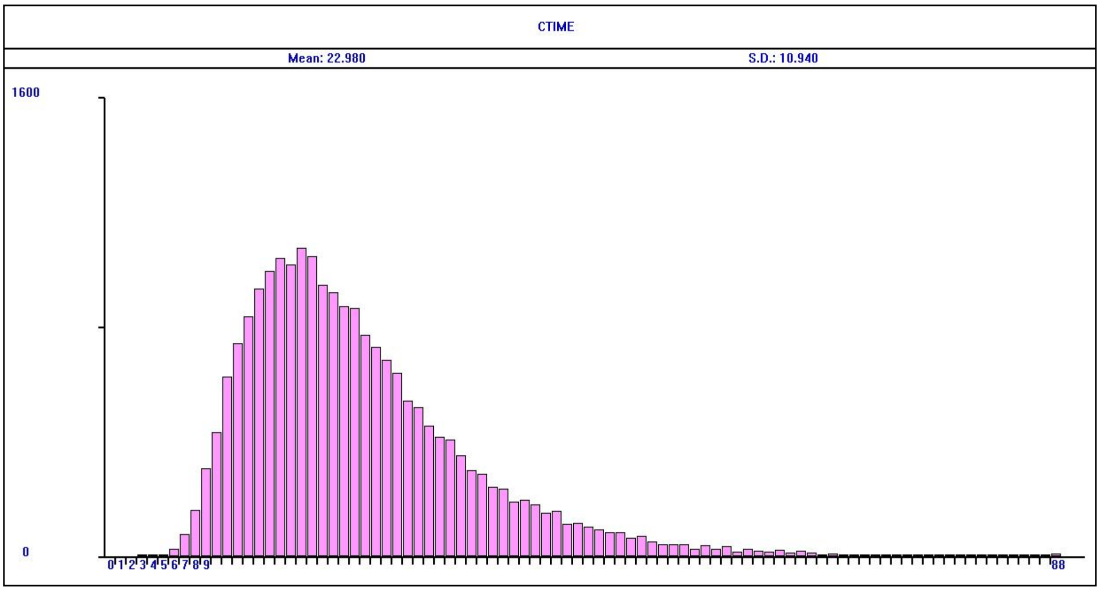

Figure 23. Distribution of C for the case when $n = 5, V = 1.5, \rho = 0.5,$ and $N = 7$

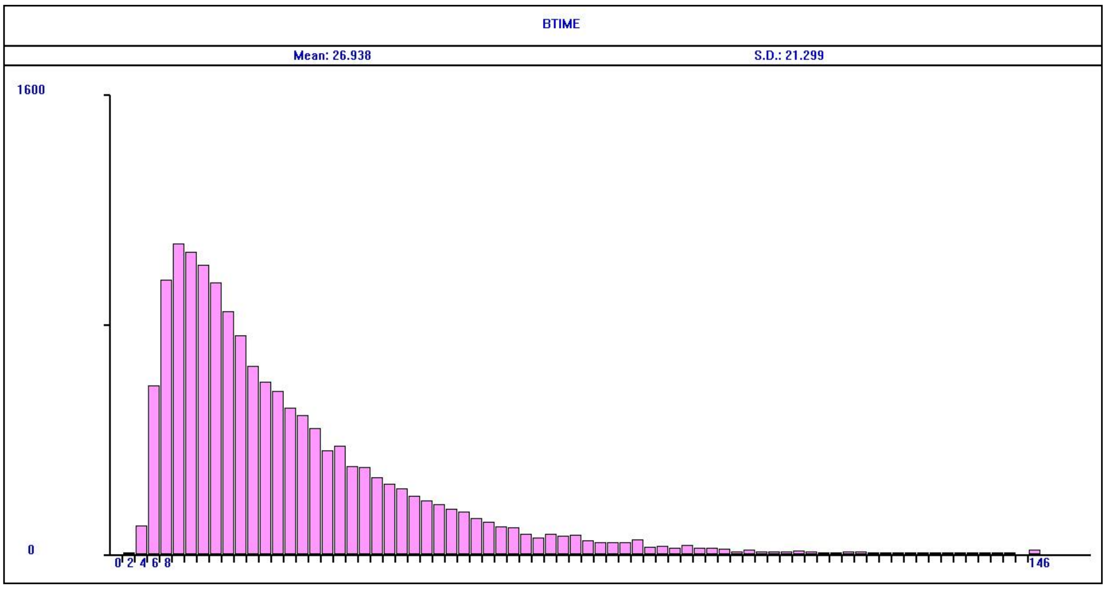

Figure 24. Distribution of B for the case when $n = 5, V = 0.5, \rho = 0.5,$ and $N = 7$

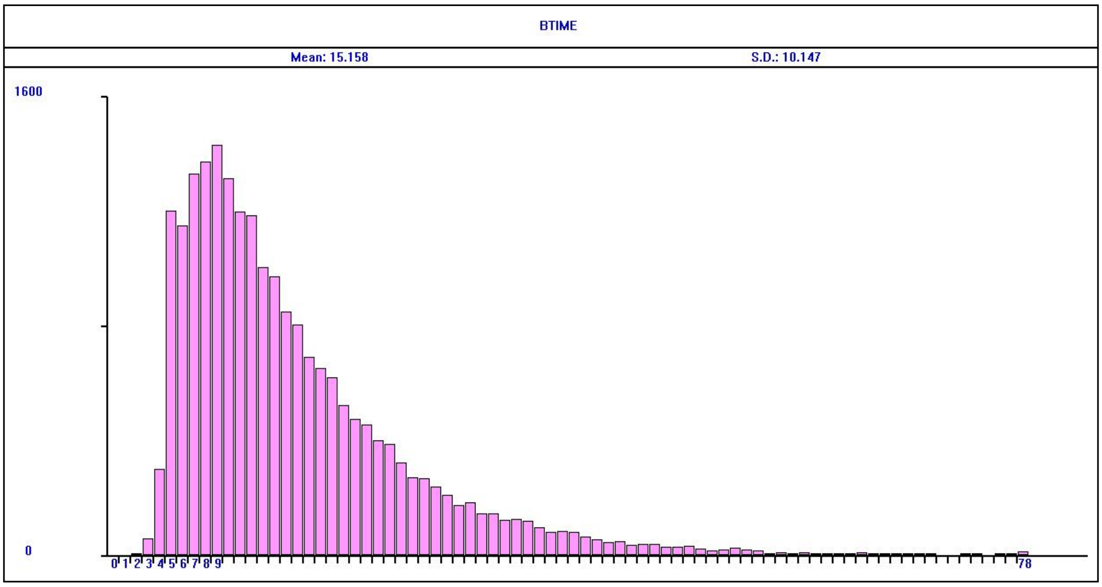

Figure 25. Distribution of B for the case when $n = 5, V = 1.5, \rho = 0.5,$ and $N = 7$

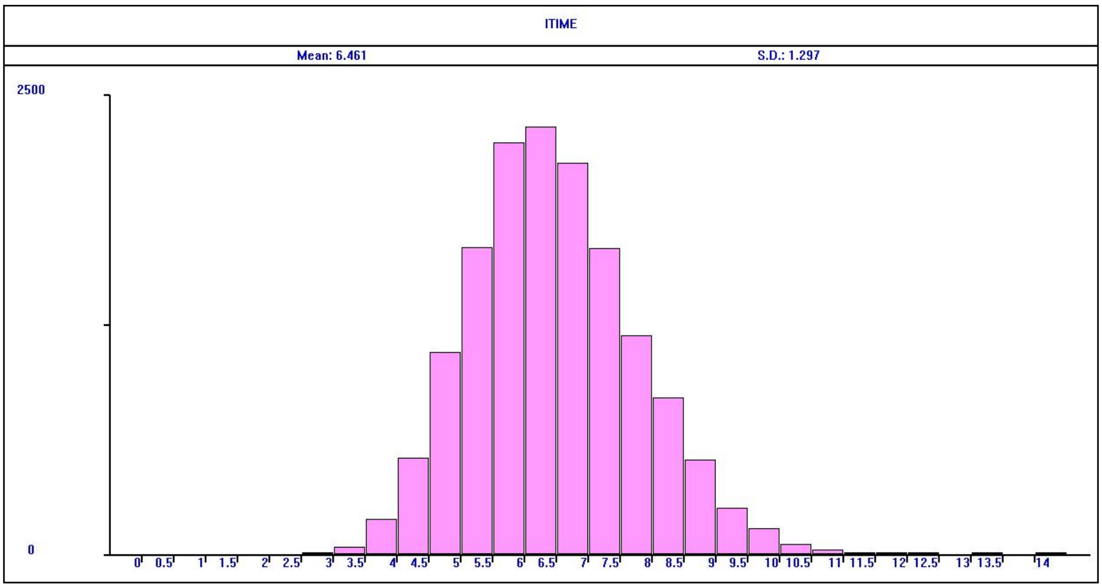

Figure 26. Distribution of I for the case when $n = 5, V = 0.5, \rho = 0.5,$ and $N = 7$

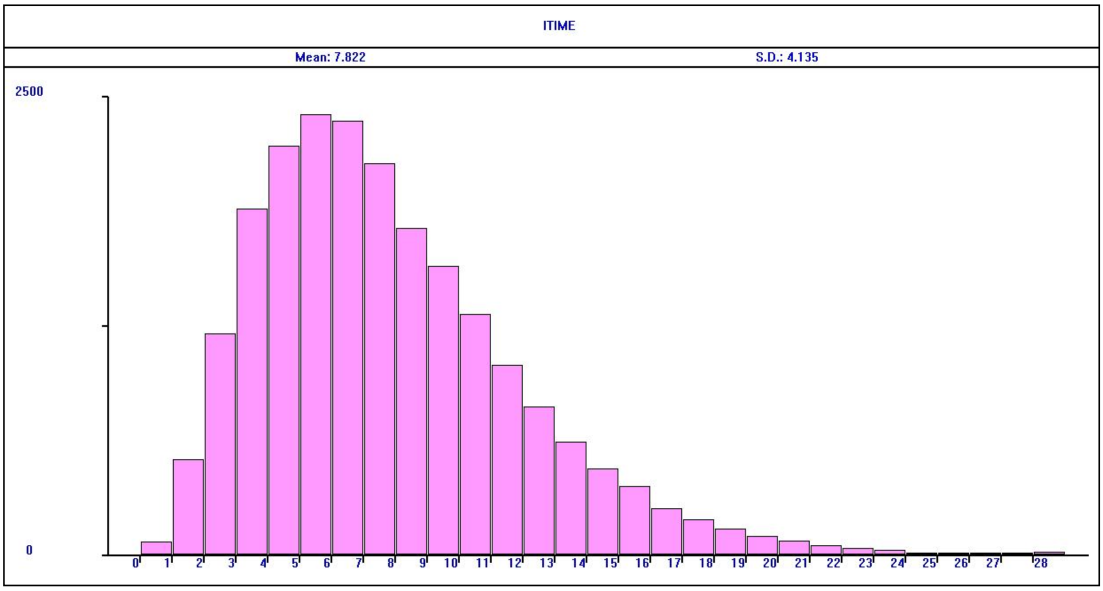

Figure 27. Distribution of I for the case when $n = 5, V = 1.5, \rho = 0.5,$ and $N = 7$

From the histograms presented in Figures 18−27, it follows that an increase in the coefficient of variation of inter-arrival times leads to an increase in the mean and variance of the random variables $L, W,$ and $I,$ while the mean and variance of the random variables C and B decrease. The histograms clearly demonstrate the impact of changing the coefficient of variation of inter-arrival times on the distribution shape of each of these random variables.

4. Conclusion

Simulation models are developed to describe the stochastic process of the *G*/*G*/*n* multi-server queueing system, which utilizes the *N*-policy or bi-level randomized (p, N_1, N_2)-policy. The constructed simulation model provides us with a fundamental opportunity to predict the impact of each input parameter on system performance measures for the *G*/*G*/*n* queue with arbitrary inter-arrival times and service time distributions. The calculations showed a good convergence of our simulation results with the results of the analytical model for the *M*/*G*/1 queue.

The results obtained for the *G*/*G*/*n* queue show a significant dependence of the system performance measures on the values of the coefficient of variation of inter-arrival times and the number of servers. This dependence increases even more if we consider the possibility of changing the coefficients of variation of service times. Therefore, it is difficult to take into account the practical application of the results of analytical models that were obtained only for *M*/*G*/1 queues.

References

[1] Yadin, M. and Naor, P., "Queueing system with a removable service station," *Journ. of the Operational Research Society*, 14 (4). 393-405. 1963.

[2] Heyman, D.P., "The *T*-policy for the *M/G*/1 queue," *Manag. Sci*., 23 (7). 775–778. Mar. 1977.

[3] Balachandran, K.R. "Control policies for a single server system," *Manag. Sci*, 19 (9). 1013–1018. May 1973.

[4] Lee, H.W. and Seo, W.J. "The performance of the *M*/*G*/1 queue under the dyadic Min(*N*,*D*)-policy and its cost optimization," *Perform. Eval*. 65 (10). 742–758. Oct. 2008.

[5] Xinyu Kuang, Yinghui Tang, Miaomiao Yu and Wenqing Wu, "Performance analysis of an *M*/*G*/1 queue with bi-level randomized $(p, N_1, N_2)-$ policy," *RAIRO-Oper. Res*., 56 (1). 395–414. Feb. 2022.

[6] Zhernovyi, Yu., "Simulation models of modified multiple vacation policy for multi-server queueing systems," *SCIREA Journal of Information Science and Systems Science*, 6 (1). 62–88. Feb. 2022.

[7] Birta, L.G. and Arbez, G., *Modelling and Simulation: Exploring Dynamic System Behaviour*, 3rd edition, Springer Nature, Switzerland, 2019, 491-520.

[8] Zhernovyi, Yu., *Creating models of queueing systems using GPSS World: Programs, detailed explanations and analysis of results*, LAP Lambert Academic Publishing, Saarbrücken, 2015, 220 p.