ORIGINAL ARTICLE

INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE

ISSN: 2692-5206, Impact Factor: 12,23
American Academic publishers, volume 05, issue 04,2025

Journal: https://www.academicpublishers.org/journals/index.php/ijai

PUBLIC SERVICE SYSTEMS: PROBABILITIES AND OPTIMIZATION

Mamatova Zilolakhon Khabibullokhonovna
Associate Professor at Fergana State University
PhD in Pedagogical Sciences
***ORCID**: 0009-0009-9247-3510*
***E-mail**: mamatova.zilolakhon@gmail.com*

Eshmamatova Nozimakhon Saydulla qizi
3rd-year student of Applied Mathematics,
Group 22-08, Fergana State University
***E-mail**: nozimaeshmamatova5@gmail.com*

Annotation: Queueing theory is a field of probabilistic systems that studies random demand flows and service processes. This theory was developed based on A. Erlang's research on telephone exchanges and is applied in various fields such as telecommunication, commerce, transportation, maintenance, and others. The main elements of the system include: incoming flow, queue, service node, and outgoing flow. The incoming flow represents the process of random demand arrivals, while the queue reflects the order in which services are awaited. The service node may consist of one or multiple channels, and the outgoing flow represents the served demands. The Poisson process is the most common type of incoming flow, characterized by stationarity, ordinariness, and lack of aftereffects. Service time often follows an exponential distribution. The efficiency of the system is assessed using indicators such as the utilization coefficient, waiting time in queue, time spent in the system by requests, and both relative and absolute throughput. Queueing systems can be open or closed, with or without waiting, and single-channel or multi-channel. The theory plays an important role in managing queues, optimizing resources, and improving service quality in practice.

Keywords: Queueing system, demand, arrival flow, queue, service node, departure flow, probabilistic system, stationary flow, ordinary flow, Poisson process, service time, transition probability, utilization coefficient, average number of requests, average waiting time in queue, average time a request spends in the system, relative throughput, absolute throughput.

Introduction. Operations research and optimal control are fields of study focused on decision-making and system optimization.

1. Operations research uses mathematical models, linear programming, game theory, and network optimization methods to efficiently allocate resources.
2. Optimal control deals with determining the best control strategies for systems. Its main methods include Pontryagin's Maximum Principle and Bellman's dynamic programming.

Literature Review: Queuing Theory

ISSN: 2692-5206, Impact Factor: 12,23
American Academic publishers, volume 05, issue 04,2025

Journal: https://www.academicpublishers.org/journals/index.php/ijai

A literature review on Queuing Theory is essential for understanding the fundamental principles, historical development, and practical applications of this field. Below, key literary sources and their contents are briefly analyzed, along with an overview of the general directions in current research.

A.K. Erlang's Works

Danish mathematician Agner Krarup Erlang (1878–1929) is considered the founder of queuing theory. His work on modeling traffic in telephone exchanges, particularly the 1917 paper "Solution of Some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges", significantly contributed to the field. Erlang's studies developed methods for determining system congestion and queue length based on models using Poisson arrivals and exponential service distributions. These works laid the foundation for classical queuing system models such as M/M/1 and M/M/c, and they remain relevant today in telecommunications and network analysis. Analysis: Erlang's contributions are fundamental in shaping the theoretical basis of queuing systems. However, his research primarily focused on telephone networks, and thus requires extension for application to modern complex systems.

L. Kleinrock's "Queueing Systems" (1975–1976)

Leonard Kleinrock's two-volume work "Queueing Systems" plays a crucial role in the modern development of queuing theory. The first volume (Theory) covers the mathematical foundations of queuing systems, including Poisson processes, Markov chains, and service time distributions. The second volume (Computer Applications) focuses on the application of the theory in computer networks and information systems. Kleinrock's work popularized important analytical tools such as Little's Law and Kendall's Notation. Analysis: Kleinrock's book provides a balanced approach between theory and practical application, greatly influencing the field of computer networks. However, its mathematical rigor may be challenging for beginners.

D. Gross and C.M. Harris – "Fundamentals of Queueing Theory" (1974, later editions up to 2013)

This book is considered one of the most important introductory texts on queuing theory. The authors explain models such as M/M/1, M/M/c, M/G/1, and also cover the statistical characteristics of queuing systems, simulation methods, and practical examples. Later editions of the book expand on modern issues, including the analysis of network systems and call centers. Analysis: The book is reader-friendly and enriched with examples and exercises. However, it mainly focuses on classical models and does not deeply explore complex dynamic systems.

Research Methodology

The research methodology in queuing systems focuses on analyzing random processes, queues, and service efficiency using mathematical modeling**,** statistical methods**,** and

ORIGINAL ARTICLE

INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE

ISSN: 2692-5206, Impact Factor: 12,23
American Academic publishers, volume 05, issue 04,2025

Journal: https://www.academicpublishers.org/journals/index.php/ijai

simulation**.** Below is a concise and meaningful analysis of research methodology suitable for the topic “Queuing Systems: Probability and Optimization.”

Findings and Analysis

The emergence and development of queuing theory stem from the need to study probabilistic systems arising in various fields of human activity. The formation of this theory was significantly influenced by the work of Danish scientist A. Erlang, whose research on improving telephone exchanges is of particular importance. Queuing theory uses requests arriving at random time intervals to provide quantitative characteristics of a system and develops methods for analyzing it. A queuing system involves both incoming requests and the service-providing entity. A key feature of such systems is the random nature of demand**,** meaning that the timing and volume of incoming requests are unpredictable. Additionally, the time required to service each request is also a random variable. As a result, a mismatch often arises between the rate at which requests arrive and the speed at which they are serviced. This leads to either the formation of queues or idle service facilities.

Examples of real-world queuing systems include:

- Long-distance telephone exchange services;
- Checkout counters in supermarkets;
- Vehicle traffic at signal-controlled intersections;
- Waiting lines in reception areas;
- Aircraft awaiting take-off clearance at major airports;
- Technicians servicing faulty equipment or machinery;
- Manuscripts awaiting publication in a publishing house.

Basic Elements of Queuing Systems

The key structural components of queuing systems include:

1. Incoming request stream**;**
2. Queue for service**;**
3. Service node**;**
4. Outgoing request stream**.**

Based on these, a queuing system model can be constructed, illustrated in **Figure 1**.

Incoming Stream

The incoming stream represents the process of service-requesting entities entering the system. Requests may arrive individually or in groups. Examples of individual arrivals include ships entering a port, customers entering a store, or clients visiting a repair workshop. Examples of group arrivals include freight train wagons arriving at a cargo station or diners entering a restaurant for meals.

INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE

ORIGINAL ARTICLE

ISSN: 2692-5206, Impact Factor: 12,23
American Academic publishers, volume 05, issue 04,2025

Journal: https://www.academicpublishers.org/journals/index.php/ijai

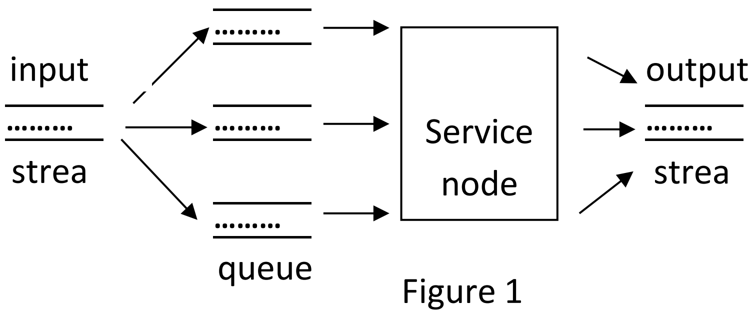

Figure 1

In some cases, requests may leave the system without being serviced. This may occur due to long waiting times, lack of space in the waiting area, or other reasons. If the system has no space for waiting, it is referred to as a loss system (or queue-less system). If all requests in the system are eventually serviced, it is called a queuing system**.**

The source of requests may be finite or infinite. If the system receives requests from an infinite source**,** it is known as an open queuing system**.**

Queue Discipline

The order in which requests are serviced plays a major role in queuing systems. In practice, the “First Come – First Served” (FCFS) discipline is most commonly used. Sometimes, “Last Come – First Served” (LCFS) or “Random Order of Service” (ROS) may also be applied. Queues may also be organized according to a priority rule, where some requests are serviced before others based on predefined criteria.

Service Node

Once it is their turn, requests proceed to the service node**.** The service node may consist of one or several service channels (servers). A system with only one server is called a single-channel system**,** while systems with multiple parallel servers are referred to as multi-channel systems. A typical example of a multi-channel system would be a large supermarket with several checkout counters**.** If servicing occurs through several stages in sequence, using different servers, the system is known as a multi-phase system**.** A typical example would be a manufacturing process where a product undergoes processing in a sequence of workshops.

Output Stream

Requests that have been serviced form the output stream. This output stream may serve as the input stream for another system, as seen in multi-phase systems. If serviced requests return to the system, we have a closed queuing system. An example of this would be

ISSN: 2692-5206, Impact Factor: 12,23
American Academic publishers, volume 05, issue 04,2025

Journal: https://www.academicpublishers.org/journals/index.php/ijai

the maintenance of machine tools in a manufacturing plant, where repaired machines are returned to the production line and may require servicing again later.

Poisson Process and Service Time

The study of queuing systems often begins with the analysis of the structure of the input stream. The simplest type is a regular stream, where requests arrive at fixed time intervals. However, regular input streams are rare in practice. In most real-life cases, the interarrival times of requests are random. The probability distribution of these random variables is a key characteristic of the input stream. Much of the research in queuing theory has been conducted on input streams that follow a Poisson process, also known as a simple arrival process, due to its frequent appearance in practical systems and the analytical convenience it offers.

The Poisson process has three key properties:

1. **Stationarity**: The probability of a certain number of arrivals in a time interval depends only on the length of the interval, not on its position on the time axis.
2. **Ordinariness**: In a sufficiently small time interval Δt , the probability of more than one arrival is negligibly small, i.e., it becomes a higher-order infinitesimal.
3. **Lack of Aftereffects (Memorylessness)**: The probability of a certain number of arrivals in a time interval does not depend on the number or timing of arrivals in previous intervals. In other words, arrivals in disjoint time intervals are statistically independent.

From these definitions, it is clear that:

- **Stationarity** implies the arrival process has time-invariant probabilistic characteristics.
- **Ordinariness** ensures that multiple arrivals at exactly the same moment are practically impossible.
- **Lack of Aftereffects** means that the arrival process in one time segment does not influence the process in another.

Let us denote the probability that k requests arrive during a time interval of length τ in a Poisson process as $p_k(t)$. For a Poisson arrival process:

$$p_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \qquad (1)$$

The distribution is valid, where $\lambda > 0$ is the flow parameter. This formula is known as the Poisson formula.
To understand the meaning of the λ parameter, let us find the expected value, i.e., the average number of arrivals within the time interval $(0, t)$

ORIGINAL ARTICLE

INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE

ISSN: 2692-5206, Impact Factor: 12,23
American Academic publishers, volume 05, issue 04,2025

Journal: https://www.academicpublishers.org/journals/index.php/ijai

$$M(t) = \sum_{k=1} kp_k(t) = \sum_{k=1} k\frac{(\lambda t)^k}{k!}e^{-\lambda t} = e^{-\lambda t}\lambda t \sum_{s=0} \frac{(\lambda t)^s}{s!} = \lambda t$$

Naturally, if $t = 1$, then $M(1) = \lambda$. This means that λ represents the average number of arrivals per unit of time.

Using the Poisson formula, we get:

$$p_1(\Delta t) = \lambda\Delta t + o(\Delta t) \quad (2)$$

Here, $o(\Delta t)$ is a quantity that becomes infinitesimally small compared to Δt

in other words, $\frac{o(\Delta t)}{\Delta t} \to 0$, $\Delta t \to 0$

From (2), when Δt is sufficiently small, we obtain the relationship:

$$p_1(\Delta t) \approx \lambda\Delta t$$

That is, in a sufficiently short time interval, the probability of an arrival is approximately $\lambda\Delta t$.

Let T denote the interarrival time between successive requests in the system. Then, $p(T < t)$ represents the probability that the arrival time of a new request in the system exceeds Δt . We calculate this probability as:

$$p(T < t) = \sum_{k=1} P_k(t) = \sum_{k=1} \frac{(\lambda t)^k}{k!}e^{-\lambda t} = 1 - e^{-\lambda t} . \quad (3)$$

Thus, equation (3) shows that if the arrival times of requests follow a Poisson process, then the interarrival times follow the exponential distribution.

To analyze a queueing system using mathematical methods, the service time distribution function of each service device in the node must be known.

Let $t_{x.k.}$ be the service time for a request, and $p(t_{x.k.} < t)$ be the probability that the service time does not exceed Δt . Then, $F(t) = p(t_{x.k.} < t)$ is the distribution function of the service time.

The service time distribution can vary. In practice, the exponential distribution is often used:

$$F(t) = 1 - e^{-\mu t}, \mu > 0, \quad (4)$$

ISSN: 2692-5206, Impact Factor: 12,23

American Academic publishers, volume 05, issue 04,2025

Journal: https://www.academicpublishers.org/journals/index.php/ijai

Here, μ is called the service rate (intensity). To understand the meaning of this parameter, we calculate the expected value of the service time:

$$M(t_{x.k.}) = \overline{t_{x.k.}} = \int_{0} tdF(t) = \int_{0} t\mu e^{-\mu t} dt = \frac{1}{\mu}$$

Thus, the μ parameter represents the average number of requests the service device can handle per unit of time.

Using the exponential distribution in (4), we can calculate the probability that a request will be served within time Δt as:

$$F(\Delta t) = p(t_{x.\kappa.} < \Delta t) = 1 - e^{-\mu \Delta t}.$$

$F(t) = p(t_{x.k.} < t)$ using the Taylor series expansion of the exponential function, we get:

$$p(t_{x.\kappa.} < \Delta t) = 1 - 1 + \mu \Delta t + o(\Delta t) \approx \mu \Delta t \qquad (5)$$

The probability that the service of a request will not be completed in time Δt is:

$$p(t_{x.\kappa.} \quad \Delta t) = 1 - p(t_{x.\kappa.} < \Delta t) \approx 1 - \mu \Delta t$$

Example

Customers with a high likelihood of making a purchase arrive at a retail store. Suppose we want to test the hypothesis that the arrival process of customers follows a Poisson distribution. To do this, we perform a statistical analysis in order to determine and construct the parameters of the empirical distribution function.

In building the mathematical model, we consider only those customers who are certain to make a purchase. Assume that the arrival time of each customer within the time interval T is fixed. We form a sample based on the number of customers who actually make purchases.

According to the statistical method, we divide the observation time interval T into n equal subintervals: $\Delta t = T/n$

We compute the parameter using:

$$\alpha \cdot \Delta t = \left(\sum_{i=1}^{r} i \cdot m_i\right) / \left(\sum_{i=1}^{r} m_i\right)$$

ISSN: 2692-5206, Impact Factor: 12,23
American Academic publishers, volume 05, issue 04,2025

Journal: https://www.academicpublishers.org/journals/index.php/ijai

Here, r is the last interval for which customer data exists (i.e., there are no intervals beyond r with customers). We then find the relative frequency Wi of observing i customers in Δt :

$$W_i = \frac{m_i}{\sum_{i=1}^{r} m_j}, i = 0,1,2,\dots,r.$$

Next, we calculate the theoretical probability p_i of having i arrivals in Δt based on the Poisson process:

$$p_i = \frac{(\alpha\Delta t)^i}{i!} e^{-\alpha\Delta t}, i = 0,1,2,\dots,r$$

Let's suppose the observation time is 300 minutes. We divide this time interval into n=100 equal subintervals, so Δt =3 minutes.

The hypothesis that customer arrivals at the store follow a Poisson distribution is not contradicted by the experimental data, and therefore, it can be accepted.

The obtained results are expressed in the following table format:

	The number of customers in the t= 3 min interval.	**The number of intervals with i customers, denoted as mi.**	**The probability value of the Poisson distribution, denoted as pi.**	**Relative frequency Wi**
1	0	0	0.010	0
2	1	5	0.043	0.05
3	2	11	0.106	0.11
4	3	13	0.163	0.13
5	4	22	0.187	0.22
6	5	18	0.172	0.18
7	6	14	0.132	0.14
8	7	9	0.087	0.09
9	8	4	0.050	0.04
10	9	2	0.026	0.02

ORIGINAL ARTICLE

INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE

ISSN: 2692-5206, Impact Factor: 12,23
American Academic publishers, volume 05, issue 04,2025

Journal: https://www.academicpublishers.org/journals/index.php/ijai

11	10	1	0.012	0.01
12	11	1	0.005	0.01
13	12	0	0.002	0

Explanation: In the example considered, the arrival rate of the process is given as α·Δt = 4, 6. The unit time is taken as t= 3 minutes. Therefore, the arrival rate per minute is α = 4, 6/3 ≈ 1,53 1/min per minute, assuming that the time unit is typically taken as 1 minute.

Conclusion:

Mass service systems are an important theoretical field that analyzes random demand flows and service processes, and they have been developed based on A. Erlang's research. These systems consist of the incoming flow, queue, service node, and outgoing flow, and they are characterized by probabilistic models such as Poisson flow and exponential service times. Open, closed, waiting, or non-waiting systems are used in various fields, such as communication, trade, transportation, and maintenance. The system's performance is evaluated using metrics like the utilization factor, waiting time, and throughput capacity. The theory plays a crucial role in efficiently managing queues, optimizing resource usage, and improving service quality.

References:

1. Erlang, A. K. (1917). Solution of Some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges. Post Office Electrical Engineers' Journal.
2. Kleinrock, L. (1975–1976). Queueing Systems, Volume 1: Theory & Volume 2: Computer Applications. Wiley-Interscience.
3. Gross, D., & Harris, C. M. (2013). Fundamentals of Queueing Theory (5th ed.). Wiley.
4. Medhi, J. (2002). Stochastic Models in Queueing Theory (2nd ed.). Academic Press.
5. Cooper, R. B. (1981). Introduction to Queueing Theory (2nd ed.). North-Holland.
6. Bhat, U. N. (2015). An Introduction to Queueing Theory: Modeling and Analysis in Applications (2nd ed.). Birkhäuser.
7. Taha, H. A. (2017). Operations Research: An Introduction (10th ed.). Pearson.
8. European Journal of Operational Research (Articles on queueing systems).
9. Queueing Systems (Articles on network queueing systems).
10. Scientific publications of Tashkent University of Information Technologies and Samarkand State University (articles on network systems and logistics).