Journal de Mathématiques Pures et Appliquées de Ouagadougou Volume 4 Numéro 1 (2025)

ISSN: 2756-732X URL:https//:www.journal.uts.bf/index.php/jmpao

Modeling of Telecommunication Networks by $M/M/K$ $(K > 2)$ Queueing Systems with Impatient Customers

Finyori FAYAMA[†1]
Raogo Frank Emile 1er Jumeau KABORE[‡]
Abel ZONGO[†]
S. Pierre Clovis NITIEMA[†]

[†]Université Thomas SANKARA
[†] Laboratoire Sciiences et technologie (LaST),
Université Thomas SANKARA, 12 BP 417 Ouagadougou, Burkina Faso
‡ Département de mathématiques,
Université Joseph KI-ZERBO, 03 BP 7021 Ouagadougou, Burkina Faso

Abstract : The objective of this work is to apply the $M/M/K$ $(K > 2)$ queueing network model to better understand the problem of saturation and disruption in telecommunication networks. Assuming that the system has reached an equilibrium between arrivals and departures, we derive the system of equations associated with the model. By using the probability generating function method, we obtain the network availability probability as well as key performance metrics such as the average number of calls waiting for connection and the average call duration before connection (using Little's formula). Through numerical illustrations, we show, on the one hand, the influence of the parameters λ (arrival rate), μ (service rate), and ξ (impatience threshold) on the network availability probability $P_{0,0}$, and on the other hand, the impact of $P_{0,0}$ on the average call duration before connection.

Keywords : Queueing systems; $M/M/K$ network; Telecommunication networks; Availability probability
2020 Mathematics Subject Classification : 60K20, 60K25, 60K30
(Received: 28/10/2025) (Accepted 24/11/ 2025)

1 Introduction

[1]Corresponding author:finyorif850@gmail.com

Multiserver queueing networks are an extension of simple queueing models used to analyze complex systems in which multiple servers interact with each other. These networks are applied in many fields such as call centers, telecommunication networks, data centers and cloud services, airports, and healthcare or clinical services.
In this paper, we use the $M/M/K$ ($K > 2$) network model to help understand the problem of saturation and disruption in telecommunication networks.
First, assuming that the system has reached its steady-state statistical equilibrium, we derive the balance equations associated with the model. Second, by using the generating function method, we obtain the availability probability $P_{0,0}$ of the telecommunication network. Third, based on the probability generating function, we derive the network performance metrics such as the average number of calls waiting for connection and the average call duration before connection. Finally, with the help of Matlab software, we illustrate, on the one hand, how the model parameters affect the availability probability of the network, and on the other hand, how this availability probability influences the average call duration before connection.

2 Literature Review

From the very beginning, Erlang focused on the application of probability theory to challenges posed by telephone traffic. In 1909, he published his first work, entitled "The Theory of Probabilities and Telephone Conversations" [8], in which he demonstrated that randomly arriving telephone calls follow a Poisson distribution.

Furthermore, ERLANG A.K [7] addressed several significant probability theory problems related to automatic telephone exchanges.

Probability theory is particularly useful for solving various problems in telecommunications. These issues arise not only in call management but also in statistical studies that support the development of essential plans for research in physics as well as for the manufacturing of telephone devices [11].

The development of queueing theory has become an indispensable tool for analyzing the performance of complex computer and communication networks [3].

Mohammad Gharbieh et al. [10] developped a spatio-temporal, traffic-aware mathematical model for Internet of Things (IoT) devices using uplink cellular connectivity. Their model relies on stochastic geometry and queueing theory to account for the traffic demands of each IoT device, different transmission strategies, as well as mutual interference among devices.

Ahmed Nasrallah et al. [12] provided a comprehensive survey on queueing and scheduling mechanisms to support large-scale deterministic networks (LDNs). The main outcome of this survey is a clear organization of various research and standardization efforts related to queueing and scheduling mechanisms for LDNs, along with the identification of key development trends and interdependencies of these mechanisms.

Bernal Alvaro et al. [2] presented a methodology for accurate, scalable, and predictive estima-

tion of end-to-end KPIs with sub-second granularity, near real-time, in converged fixed-mobile networks. Specifically, they extended their CURSA-SQ methodology, originally developed for mobile network traffic analysis, to enable converged exploitation of fixed and mobile networks. CURSA-SQ combines simulation and machine learning, powered by monitoring data from real networks. Numerical results validated the accuracy, robustness, and usability of the proposed CURSA-SQ methodology in converged fixed-mobile network scenarios.
BANCHS Albert et al. [1] identified the main challenges related to the integration of AI into a next-generation network architecture and proposed a native-NI approach tailored to the requirements of 6G. As an illustration, they presented a case study focused on a representative feature: network capacity prediction. This feature, designed to fully leverage their NI architecture, demonstrates the benefits of intelligent and context-aware AI integration within the network.

Polese Michele et al. [13] highlighted the transformative potential offered by adopting new cellular architectures, transitioning from conventional systems to the progressive principles of Open RAN. This promises to make 6G networks more agile, cost-effective, energy-efficient, and resilient. It also paves the way for a multitude of new use cases, ranging from widespread support for autonomous devices to low-cost expansions in previously underserved regions.

Ruiz Marc et al. [14] presented a traffic analysis methodology aimed at helping operators calculate the expected traffic demand in their networks by combining well-known mass services with forecast scenarios for B5G services. Numerical results, based on forecasts and data provided by major European operators, indicate that multi-band networks will be necessary across all network segments, including metropolitan aggregation, metropolitan core, and transport (backbone) networks, by the end of this decade.

Satka Zenepe et al. [15] proposed a comprehensive and structured systematic review of existing research on TSN-5G integration. They outlined the planning, execution, and analysis phases of this review, while identifying technical trends, key features, and gaps in the current literature. Notably, 73 percent of the primary studies focus on time synchronization in TSN-5G integration, with approaches achieving accuracy ranging from hundreds of nanoseconds to one microsecond. Moreover, the majority of these studies aim to optimize communication latency, a critical performance metric for modern industrial applications such as automotive systems and real-time automation. Their results highlight the current state of the art while suggesting future research directions to advance TSN and 5G integration in next-generation industrial communication systems.

Singh Saroja Kumar et al. [16] studied the problem of change-point detection in inter-arrival times of an M/M/2 queue with heterogeneous servers. The model assumes the queue is in steady state, customers are served by the fastest available server, and there is no transfer of customers between servers. Maximum likelihood estimators were derived for the arrival rates before and after the change point, as well as for the service rates of the servers. Monte Carlo simulation results demonstrated the efficiency and performance of the proposed estimators.

Miquel Ferriol-Galmés et al. [9] presented RouteNet-Fermi, a customized Graph Neural Network (GNN) model that shares the objectives of queueing theory while achieving significantly higher accuracy under realistic traffic patterns. This model accurately predicts latency, jitter, and packet loss in a network. Experimental results showed that RouteNet-Fermi attains precision comparable to low-level packet simulators, while being more scalable for large networks.

Devigili Mariano et al. [5] investigated the feasibility of extracting information from IQ constellations and using it both for accurate transmission quality (QoT) estimation and effective fault management. They designed deep neural network (DNN) models for QoT estimation. Subsequently, specific DNN models and algorithms exploiting IQ constellation features were proposed for the detection, identification, and severity estimation of "soft" faults. Results from both simulations and experiments demonstrated notable accuracy in QoT estimation and in predicting faults affecting transmitters, optical filters, and amplifiers.

Dudin Alexander et al. [6] developed a framework for uniform algorithmic analysis of queueing systems with a Markovian arrival process and simultaneous service of a limited number of customers, described by a multidimensional Markov chain. This chain behaves as a quasi-death finite-state process between two successive service initiation instants, with jumps occurring at these instants. Such a description of the service process generalizes many known limited resource-sharing mechanisms and is well-suited for modeling various future mechanisms.

Franco Coltraro et al. [4] presented the formal derivation and mathematical properties of a continuous fluid-flow queueing model called the logistic queueing model. They provided extensions to model different characteristics of telecommunication networks, including finite buffer sizes and flow propagation.

Thapa Suman et al. [17] proposed a novel Markovian modeling approach for queueing systems with servers providing correlated services. They applied this approach to a queueing system with Poisson arrivals and two positively correlated exponential servers, which they denoted as the $M/M_D/2$ system. They first proved that the queueing process (i.e., the number of customers in the system) is a Markov chain and then provided an analytical solution for the stationary distribution of the process. This solution offers a better understanding of the impact of server correlation on system performance, compared to a system with independent services.

3 Model Description

Let us assume that customer calls arrive according to a Poisson process with rate λ, and that call durations are exponentially distributed with rate μ. The service discipline is FCFS (First-Come, First-Served), since when many calls arrive simultaneously, the first arrivals are served first.

Let θ denote the probability that a customer becomes impatient, and let T be the impatience threshold for each customer, which follows an exponential distribution with parameter ξ. Let γ be the probability that a customer abandons the system once the impatience threshold is exceeded, and let δ be the probability that a dissatisfied customer permanently leaves the system. Customer impatience is caused by network congestion or technical failures (such as cable or optical fiber issues). The network state depends on the number of connected users, the volume of data exchanged, and potential incidents or maintenance operations. Let $P_{0,0}$ denote the probability that the network is available.

Failure recovery times are exponentially distributed with rate α ($\alpha > 0$), and failure notification times follow a Poisson process with rate β ($\beta > 0$).

This system is modeled as a Markovian $M/M/K$ queueing network ($K > 2$).

4 Assumptions

We make the following assumptions regarding $P_{0,0}$ (the probability that the network is available):

- If $P_{0,0}$ increases, the network becomes increasingly available.
- If $P_{0,0}$ decreases, the network becomes less available, indicating a tendency toward network congestion.

Based on these assumptions, we will determine the performance parameters that allow us to analyze the network state and the impatience behavior of customers during calls.

5 Equilibrium Equations of the Model

Let $R(t)$ denotes the state of the network at time t, defined as:

$$R(t) = \begin{cases} 0 & \text{if the network is available,} \\ 1 & \text{if the network is busy.} \end{cases}$$

The following transition diagram is associated with the model:

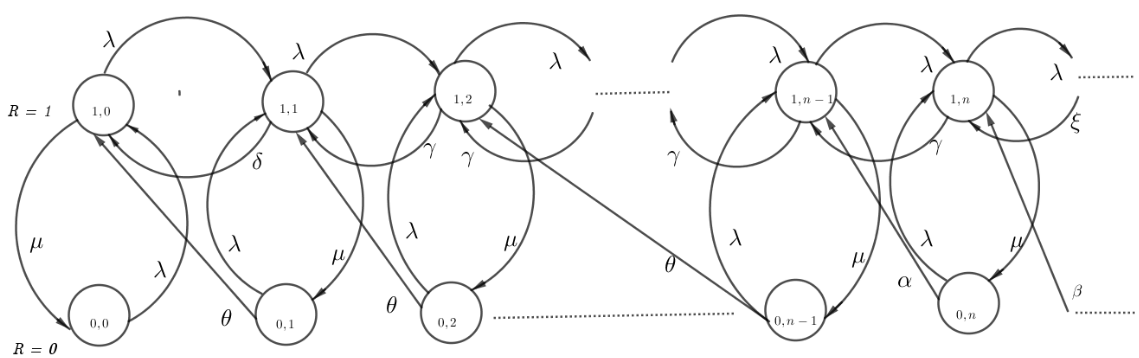

Figure 1: Transition diagram associated with the model

In Figure 1, the vertices represent the different states of the network, while the arcs describe the possible transitions. These transitions are characterized by strictly positive rates $(\lambda, \mu, \alpha, \beta, \xi)$ and probabilities (θ, γ, δ). The associated Markov chain is irreducible, since every state is accessible from any other with a nonzero probability in a finite number of steps. It is also aperiodic, because for each state there exists a positive probability of returning to that state at some point. Consequently, there exists a stationary distribution over the state space to which the chain converges regardless of the initial state. We therefore conclude that the chain is ergodic.

Theorem 5.1. *Under the established assumptions, the network availability probability, denoted by $P_{0,0}$, is given by the following expression:*

$$P_{0,0} = \frac{G\left[(\lambda+\alpha)(\mu+\gamma-\xi)-(\lambda+\beta)\mu\right]+(\lambda+\mu+\alpha)\left(\beta-\frac{\xi\theta}{\delta}\right)P_{0,1}}{(\lambda+\mu+\alpha)\left(\left[\lambda+\mu+\gamma-\xi-\frac{\xi}{\delta}(\lambda+\mu)\right]\frac{\lambda}{\mu}-\lambda-\beta+\frac{\xi\lambda}{\delta}\right)}.$$

where:

- G*: normalization constant of the network;*
- λ*: call arrival rate (Poisson process);*
- μ*: service rate associated with the exponential distribution of call durations;*
- θ*: probability that a customer becomes impatient;*
- δ*: probability that a dissatisfied customer permanently leaves the system;*
- α*: repair rate (exponentially distributed repair times);*
- ξ*: parameter of the exponential distribution of the impatience threshold;*
- γ*: probability of leaving the system beyond the impatience threshold;*
- β*: rate of fault notification (Poisson process);*
- $P0,1$ *: probability of transition from state 0 to state 1.*

Proof. By applying the principle of flow conservation (the outgoing flow of a service equals its incoming flow), the system of balance equations associated with the model is obtained:

- For node $(0,0)$:
$$\lambda P_{0,0} = \mu P_{1,0} \quad ①$$
- For node $(1,0)$:
$$(\lambda+\mu)P_{1,0} = \lambda P_{0,0} + \theta P_{0,1} + \delta P_{1,1} \quad ②$$
- For node $(0,n)$:
$$(\lambda+\alpha)P_{0,n} = \mu P_{1,n} \quad ③$$
- For node $(1,n)$:
$$(\lambda+\mu+\gamma)P_{1,n} = \lambda P_{1,n-1} + \lambda P_{0,n} + \xi P_{1,n+1} + \beta P_{0,n+1} \quad ④$$

Thus, the system of balance equations can be written as:

$$\begin{cases} \lambda P_{0,0} = \mu P_{1,0} & ① \\ (\lambda+\mu)P_{1,0} = \lambda P_{0,0} + \theta P_{0,1} + \delta P_{1,1} & ② \\ (\lambda+\alpha)P_{0,n} = \mu P_{1,n} & ③ \\ (\lambda+\mu+\gamma)P_{1,n} = \lambda P_{1,n-1} + \lambda P_{0,n} + \xi P_{1,n+1} + \beta P_{0,n+1} & ④ \end{cases} \tag{5.1}$$

Finally, since the sum of probabilities equals 1, we obtain the normalization condition:

$$\sum_{n=0}^{\infty} \left(P_{0,n} + P_{1,n}\right) = 1. \tag{5.2}$$

We solve the system using the method of generating functions.
Let $P_i(z)$ denote the generating function of the probabilities $P_{i,n}$, defined by:

$$P_0(z) = \sum_{n=0}^{\infty} z^n P_{0,n}, \qquad P_1(z) = \sum_{n=0}^{\infty} z^n P_{1,n}.$$

Substituting into the third and fourth equations of the system yields:

$$(\lambda + \alpha) \sum_{n=0}^{\infty} z^n P_{0,n} = \mu \sum_{n=0}^{\infty} z^n P_{1,n}, \tag{5.3}$$

and

$$\begin{aligned}(\lambda + \mu + \gamma) \sum_{n=1}^{\infty} z^n P_{1,n} = \lambda \sum_{n=1}^{\infty} z^n P_{1,n-1} + \lambda \sum_{n=1}^{\infty} z^n P_{0,n} \\ + \xi \sum_{n=1}^{\infty} z^n P_{1,n+1} + \beta \sum_{n=1}^{\infty} z^n P_{0,n+1}.\end{aligned} \tag{5.4}$$

As

$$\sum_{n=0}^{\infty} z^n P_{0,n} = p_0(z), \qquad \sum_{n=0}^{\infty} z^n P_{1,n} = p_1(z),$$

equation (3) becomes

$$(\lambda + \alpha) p_0(z) = \mu p_1(z), \tag{5.5}$$

and equation (4) becomes

$$\begin{aligned}(\lambda + \mu + \gamma) \left[p_1(z) - P_{1,0}\right] = \frac{\lambda}{z} p_1(z) + \lambda \left[p_0(z) - P_{0,0}\right] \\ + \frac{\xi}{z} \left[p_1(z) - P_{1,0} - z P_{1,1}\right] + \frac{\beta}{z} \left[p_0(z) - P_{0,0} - z P_{0,1}\right].\end{aligned} \tag{5.6}$$

From equation (5), we obtain:

$$p_0(z) = \frac{\mu}{\lambda + \alpha} p_1(z). \tag{5.7}$$

From equation (6), we solve for $p_1(z)$:

$$\begin{aligned}\left(\lambda + \mu + \gamma - \frac{\lambda}{z} - \frac{\xi}{z}\right) p_1(z) = \left(\lambda + \frac{\beta}{z}\right) p_0(z) + (\lambda + \mu + \gamma) P_{1,0} \\ - \frac{\xi}{z} P_{1,0} - \beta P_{0,1} - \lambda P_{0,0} - \frac{\beta}{z} P_{0,0} - \xi P_{1,1}.\end{aligned} \tag{5.8}$$

From the second equation of the system, we deduce:

$$P_{1,1} = \frac{(\lambda+\mu)P_{1,0} - \lambda P_{0,0} - \theta P_{0,1}}{\delta}. \tag{5.9}$$

Substituting $P_{1,1}$ into the previous equation and factoring partially with respect to $P_{1,0}$, $P_{0,0}$, and $P_{0,1}$, we obtain:

$$\begin{aligned}\left(\lambda+\mu+\gamma-\frac{\lambda}{z}-\frac{\xi}{z}\right)p_1(z) &= \left(\lambda+\frac{\beta}{z}\right)p_0(z)\\ &+\left[\lambda+\mu+\gamma-\frac{\xi}{z}-\frac{\xi}{\delta}(\lambda+\mu)\right]P_{1,0}-(\lambda+\frac{\beta}{z}-\frac{\xi\lambda}{\delta})P_{0,0}-(\beta-\frac{\xi\theta}{\delta})P_{0,1}.\end{aligned} \tag{5.10}$$

From the first equation, we have:

$$P_{1,0} = \frac{\lambda}{\mu}P_{0,0}. \tag{5.11}$$

Substituting $p_0(z)$ and $P_{1,0}$ gives:

$$\begin{aligned}&\left(\lambda+\mu+\gamma-\frac{\lambda}{z}-\frac{\xi}{z}-(\lambda+\frac{\beta}{z})\frac{\mu}{\lambda+\alpha}\right)p_1(z)\\ &=\left(\left[\lambda+\mu+\gamma-\frac{\xi}{z}-\frac{\xi}{\delta}(\lambda+\mu)\right]\frac{\lambda}{\mu}-\lambda-\frac{\beta}{z}+\frac{\xi\lambda}{\delta}\right)P_{0,0}-(\beta-\frac{\xi\theta}{\delta})P_{0,1}.\end{aligned} \tag{5.12}$$

Hence,

$$p_1(z) = \frac{\left(\left[\lambda+\mu+\gamma-\frac{\xi}{z}-\frac{\xi}{\delta}(\lambda+\mu)\right]\frac{\lambda}{\mu}-\lambda-\frac{\beta}{z}+\frac{\xi\lambda}{\delta}\right)P_{0,0}-(\beta-\frac{\xi\theta}{\delta})P_{0,1}}{\lambda+\mu+\gamma-\frac{\lambda}{z}-\frac{\xi}{z}-(\lambda+\frac{\beta}{z})\frac{\mu}{\lambda+\alpha}}. \tag{5.13}$$

Taking the limit $z \to 1$, we have:

$$p_1(1) = \lim_{z\to 1} p_1(z) = \frac{\left([\lambda+\mu+\gamma-\xi-\frac{\xi}{\delta}(\lambda+\mu)]\frac{\lambda}{\mu}-\lambda-\beta+\frac{\xi\lambda}{\delta}\right)P_{0,0}-(\beta-\frac{\xi\theta}{\delta})P_{0,1}}{\mu+\gamma-\xi-(\lambda+\beta)\frac{\mu}{\lambda+\alpha}}. \tag{5.14}$$

Similarly,

$$p_0(1) = \frac{\mu}{\lambda+\alpha}p_1(1). \tag{5.15}$$

The normalization condition gives:

$$\frac{1}{G}\left[p_0(1)+p_1(1)\right] = 1, \tag{5.16}$$

which leads to

$$\frac{1}{G}\left(\frac{\mu}{\lambda+\alpha}+1\right)p_1(1) = 1. \tag{5.17}$$

Substituting the expression of $p_1(1)$, we solve for $P_{0,0}$:

$$P_{0,0} = \frac{G[(\lambda+\alpha)(\mu+\gamma-\xi)-(\lambda+\beta)\mu]+(\lambda+\mu+\alpha)(\beta-\frac{\xi\theta}{\delta})P_{0,1}}{(\lambda+\mu+\alpha)\left([\lambda+\mu+\gamma-\xi-\frac{\xi}{\delta}(\lambda+\mu)]\frac{\lambda}{\mu}-\lambda-\beta+\frac{\xi\lambda}{\delta}\right)}. \tag{5.18}$$

□

Lemma 5.2. *Let C denote the maximum number of connections that a base station can support. If N represents the number of simultaneous calls, then:*

$$\lim_{N\to\infty}\mathbb{P}(\textit{call is connected}) = 0.$$

Proof. Let λ denote the arrival rate of calls to the network. We calculate the limit of $P_{0,0}$ (the probability that the network is available) as $\lambda \to \infty$:

$$\lim_{\lambda\to+\infty} P_{0,0} = \lim_{\lambda\to+\infty}\left[\frac{G\left[(\lambda+\alpha)(\mu+\gamma-\xi)-(\lambda+\beta)\mu\right]+(\lambda+\mu+\alpha)\left(\beta-\frac{\xi\theta}{\delta}\right)P_{0,1}}{(\lambda+\mu+\alpha)\left(\left[\lambda+\mu+\gamma-\xi-\frac{\xi}{\delta}(\lambda+\mu)\right]\frac{\lambda}{\mu}-\lambda-\beta+\frac{\xi\lambda}{\delta}\right)}\right].$$

Dividing numerator and denominator by λ and taking the limit, we get:

$$\lim_{\lambda\to+\infty} P_{0,0} = \frac{G(\gamma-\xi)+\left(\beta-\frac{\xi\theta}{\delta}\right)P_{0,1}}{\infty} = 0.$$

Indeed, the numerator remains finite:

$$\lim_{\lambda\to+\infty}\left[G\left((1+\frac{\alpha}{\lambda})(\mu+\gamma-\xi)-(1+\frac{\beta}{\lambda})\mu\right)+(1+\frac{\mu+\alpha}{\lambda})\left(\beta-\frac{\xi\theta}{\delta}\right)P_{0,1}\right] = G(\gamma-\xi)+\left(\beta-\frac{\xi\theta}{\delta}\right)P_{0,1}.$$

The denominator tends to infinity as $\lambda\to\infty$:

$$\lim_{\lambda\to+\infty}(\lambda+\mu+\alpha)\left(\left[\lambda+\mu+\gamma-\xi-\frac{\xi}{\delta}(\lambda+\mu)\right]\frac{\lambda}{\mu}-\lambda-\beta+\frac{\xi\lambda}{\delta}\right) = \infty \quad \text{for } \xi\neq\delta.$$

Hence, by the quotient rule:

$$\lim_{\lambda\to+\infty} P_{0,0} = 0.$$

We conclude that the network becomes **saturated** as λ becomes very large. □

6 Performance Metrics of the Telecommunication Network

Using the probability generating function, we calculate performance metrics of the telecommunication network such as the mean number of calls waiting for connection and the average call duration before connection (using Little's formula).

Let $G_X(z)$ denote the probability generating function. By definition:

$$\begin{aligned} G_X(z) &= \mathbb{E}(z^X) \\ &= \sum_{n=0}^{1} z^n P_n(z) \\ &= P_0(z) + zP_1(z). \end{aligned}$$

From equation (7), we have:

$$p_0(z) = \frac{\mu}{\lambda+\alpha} p_1(z).$$

Therefore:

$$\begin{aligned} G_X(z) &= \frac{\mu}{\lambda+\alpha} p_1(z) + z p_1(z) \\ &= \left(\frac{\mu}{\lambda+\alpha} + z \right) p_1(z). \end{aligned}$$

From equation (17), we have:

$$p_1(z) = \frac{\left(\left[\lambda+\mu+\gamma-\frac{\xi}{z}-\frac{\xi}{\delta}(\lambda+\mu) \right] \frac{\lambda}{\mu} - \lambda - \frac{\beta}{z} + \frac{\xi\lambda}{\delta} \right) P_{0,0} - \left(\beta - \frac{\xi\theta}{\delta} \right) P_{0,1}}{\lambda+\mu+\gamma-\frac{\lambda}{z}-\frac{\xi}{z}-\left(\lambda+\frac{\beta}{z}\right)\frac{\mu}{\lambda+\alpha}}.$$

By substituting $p_1(z)$ with its expression, we have:

$$G_X(z) = \left(\frac{\mu}{\lambda+\alpha} + z \right) \frac{\left(\left[\lambda+\mu+\gamma-\frac{\xi}{z}-\frac{\xi}{\delta}(\lambda+\mu) \right] \frac{\lambda}{\mu} - \lambda - \frac{\beta}{z} + \frac{\xi\lambda}{\delta} \right) P_{0,0} - \left(\beta - \frac{\xi\theta}{\delta} \right) P_{0,1}}{\lambda+\mu+\gamma-\frac{\lambda}{z}-\frac{\xi}{z}-\left(\lambda+\frac{\beta}{z}\right)\frac{\mu}{\lambda+\alpha}}. \tag{6.1}$$

- **Mean number of calls waiting for connection:**
  Let n_{mean} denote the mean number of calls waiting for connection. Differentiating $G_X(z)$ with respect to z and taking the limit as $z \to 1$, we obtain:

$$n_{\text{mean}} = A + \frac{\left(\frac{\mu}{\lambda+\alpha}+1\right)(\xi+\beta)P_{0,0}}{\mu+\gamma-\xi-(\lambda+\beta)\frac{\mu}{\lambda+\alpha}} - \frac{\left(\frac{\mu}{\lambda+\alpha}+1\right)\left[\lambda+\xi+\beta\left(\frac{\mu}{\lambda+\alpha}\right)\right] \times A}{\mu+\gamma-\xi-(\lambda+\beta)\frac{\mu}{\lambda+\alpha}} \tag{6.2}$$

  where

$$A = \frac{\left(\left[\lambda+\mu+\gamma-\xi-\frac{\xi}{\delta}(\lambda+\mu) \right] \frac{\lambda}{\mu} - \lambda - \beta + \frac{\xi\lambda}{\delta} \right) P_{0,0} - \left(\beta - \frac{\xi\theta}{\delta} \right) P_{0,1}}{\mu+\gamma-\xi-(\lambda+\beta)\frac{\mu}{\lambda+\alpha}}. \tag{6.3}$$

- **Average call duration before connection:**
  Let d_{mean} denote the average call duration before connection and let λ denote the arrival rate of calls per unit time. Using Little's formula, we have:

$$d_{\text{mean}} = \frac{n_{\text{mean}}}{\lambda}. \tag{6.4}$$

Substituting the expression for n_{mean}, we obtain:

$$d_{\text{mean}} = \frac{1}{\lambda}\left[A + \frac{\left(\frac{\mu}{\lambda+\alpha}+1\right)(\xi+\beta)P_{0,0}}{\mu+\gamma-\xi-(\lambda+\beta)\frac{\mu}{\lambda+\alpha}} - \frac{\left(\frac{\mu}{\lambda+\alpha}+1\right)\left[\lambda+\xi+\beta\left(\frac{\mu}{\lambda+\alpha}\right)\right]\times A}{\mu+\gamma-\xi-(\lambda+\beta)\frac{\mu}{\lambda+\alpha}}\right]. \quad (6.5)$$

7 Numerical Illustrations

Using numerical simulations with Matlab, we obtain the following curves:

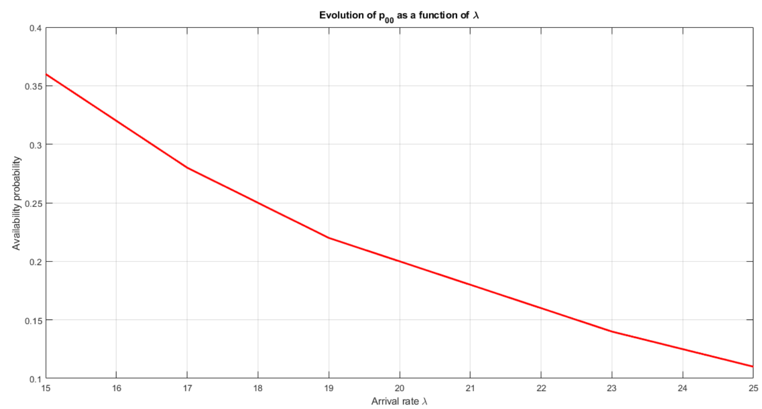

Figure 2: Influence of the arrival rate on the network availability probability.

From Figure 2, an increase in the call arrival rate λ leads to a decrease in $P_{0,0}$ (the network availability probability). This is explained by an increase in demand relative to capacity, meaning that when the number of incoming calls exceeds the available resources, the probability that the network is available decreases. The network becomes saturated more quickly, causing congestion, slower call processing, longer waiting times, and even call rejections. Some clients may therefore decide to leave the network without receiving service, leading to impatient behavior (abandonment, drop-off, and feedback).

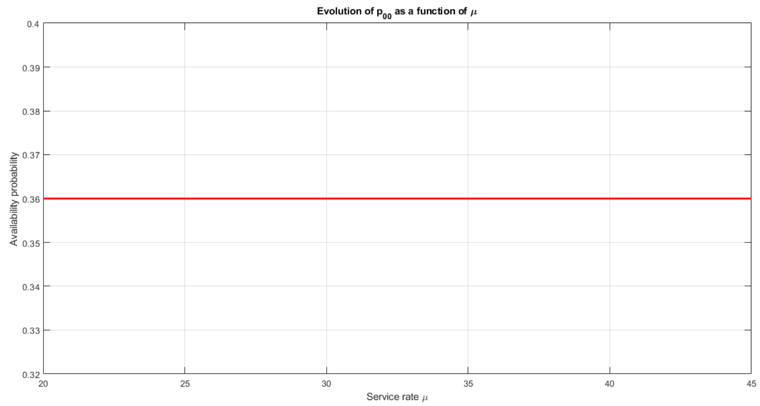

Figure 3: Influence of the service rate on the network availability probability.

Figure 3 shows that when call durations increase, the probability $P_{0,0}$ remains constant at 0.36. This can be explained by several factors: either the network has sufficient capacity to absorb the increase in call duration without degrading availability, meaning it can handle the increased load without causing congestion; or the network employs optimization mechanisms such as call prioritization and dynamic resource reallocation to maintain a good level of availability even with longer calls. Additionally, saturation is not reached, so the network can still accommodate additional or longer calls.

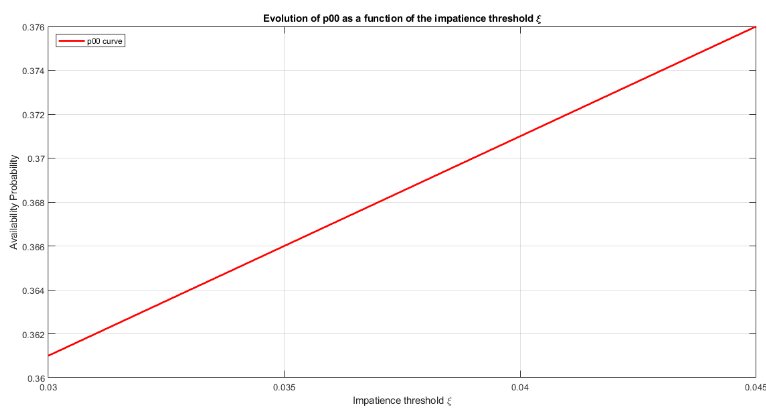

Figure 4: Influence of the impatience threshold on network availability probability.

Figure 4 shows that as the impatience threshold increases, the network becomes increasingly available. With a high impatience threshold, clients wait until a network resource becomes available, which can reduce repeated call requests (feedback). This reduces the instantaneous load on the network, allowing it to stabilize. As clients do not rush to retry service, congestion is reduced, resources are more smoothly allocated, and calls are served more efficiently, avoiding rapid saturation and decreasing the abandonment rate.

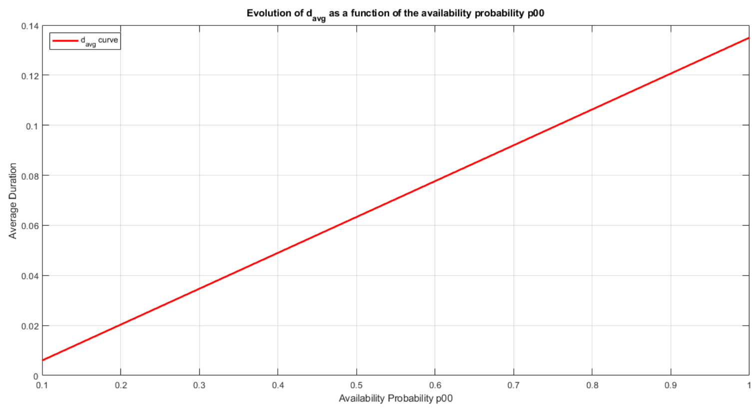

Figure 5: Influence of network availability probability on the average call duration before connection.

Figure 5 shows that as the availability probability $P_{0,0}$ increases, the average call duration before connection also increases. This means that with higher availability, calls take longer on average before being connected because the network can handle more requests, giving each call a better chance of completion. In other words, when the network is more available, the impatience effect diminishes, and clients wait longer before abandoning, indicating that more network resources are available to serve incoming calls.

8 Conclusion and Perspectives

In this work, we employed the M/M/K model (with K $>$ 2) to analyze the issues of saturation and disruption in telecommunications networks. The probability generating function method was used to solve the system of equilibrium equations associated with the model, which allowed us to determine the network availability probability $P_{0,0}$, the average number of calls waiting for connection, as well as the average call duration before connection, using Little's formula.

The study shows that increasing the number of service channels or servers (K $>$ 2) reduces waiting times, which in turn decreases customer impatience and consequently lowers the abandonment rate.

We recommend improving communication with waiting clients, for instance by providing an estimated waiting time, in order to minimize abandonments. Furthermore, it is important to continuously monitor parameters such as the arrival rate λ, the service rate μ, and the impatience threshold ξ, which will allow for rapid adjustments to the model and the adaptation of network configurations to changes in client behavior.

For future work, we plan to evaluate the network reliability index to help operators identify areas for improvement and enable regulators to monitor and compare the performance of different operators, particularly in highly competitive areas.

References

[1] BANCHS A., FIORE M., GARCIA-SAAVEDRA A. et al., *Network intelligence in 6G: Challenges and opportunities.* In: Proceedings of the 16th ACM Workshop on Mobility in the Evolving Internet Architecture. 2021. p. 7?12.

[2] BERNAL A. et al., *Near real-time estimation of end-to-end performance in converged fixed-mobile networks.* Computer Communications 150 (2020): 393?404.

[3] COHEN J. W., BOXMA O. J., *A survey of the evolution of queueing theory.* Statistica Neerlandica 39.2 (1985): 143?158.

[4] COLTRARO F., RUIZ M., VELASCO L., *The logistic queue model: theoretical properties and performance evaluation.* arXiv preprint arXiv:2405.17528 (2024).

[5] DEVIGILI M. et al., *Applications of the OCATA time domain digital twin: from QoT estimation to failure management.* Journal of Optical Communications and Networking 16.2 (2024): 221?232.

[6] DUDIN A. et al., *Framework for analysis of queueing systems with correlated arrival processes and simultaneous service of a restricted number of customers in scenarios with an infinite buffer and retrials.* Algorithms 17.11 (2024).

[7] ERLANG A. K., *Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges.* Post Office Electrical Engineer's Journal 10 (1917): 189?197.

[8] ERLANG A. K., *The theory of probabilities and telephone conversations.* Nyt. Tidsskr. Mat. Ser. B 20 (1909): 33?39.

[9] FERRIOL-GALMÉS M. et al., *RouteNet-Fermi: Network modeling with graph neural networks.* IEEE/ACM Transactions on Networking 31.6 (2023): 3080?3095.

[10] GHARBIEH M. et al., *Spatiotemporal stochastic modeling of IoT enabled cellular networks: Scalability and stability analysis.* IEEE Transactions on Communications 65.8 (2017): 3585?3600.

[11] MOLINA E. C., *The theory of probabilities applied to telephone trunking problems.* The Bell System Technical Journal 1.2 (1922): 69?81.

[12] NASRALLAH A. et al., *TSN algorithms for large scale networks: A survey and conceptual comparison.* arXiv preprint arXiv:1905.08478 (2019).

[13] POLESE M. et al., *Empowering the 6G cellular architecture with open RAN.* IEEE Journal on Selected Areas in Communications 42.2 (2023): 245?262.

[14] RUIZ M. et al., *Network traffic analysis under emerging beyond-5G scenarios for multi-band optical technology adoption.* Journal of Optical Communications and Networking 15.11 (2023): F36?F47.

[15] SATKA Z. et al., *A comprehensive systematic review of integration of time sensitive networking and 5G communication.* Journal of Systems Architecture 138 (2023): 102852.

[16] SINGH S. K. et al., *Change point estimation in an M/M/2 queue with heterogeneous servers.* Mathematics and Computers in Simulation 212 (2023): 182?194.

[17] THAPA S., ZHAO Y. Q., *Markov modelling approach for queues with correlated service times– the $M/M_D/2$ model.* arXiv preprint arXiv:2505.07648 (2025).