

Calculation of Performance Objectives in Contact Centers with Impatient Customers

Nikolay Anisimov, Nikolay Korolev, Vidas Placiakis, Herbert Ristock

Genesys Telecommunication Laboratories (an Alcatel company),
2001 Junipero Serra Blvd, Daly City, CA 94012, USA
E-Mail: {anisimov, korolev, vidasp, herbertr}@genesyslab.com

Abstract. This paper addresses the problem of calculating performance objectives in telephony contact centers used for optimization of workforce management. We consider a model with impatient customers that cannot wait very long for the next available representative. We present methods for calculating service level, average speed of answer, abandonment rate, and agent occupancy.

Key Words: Queuing System, Abandonment, Erlang-C, Erlang-A.

1 Introduction

Workforce Management systems are important parts of contemporary contact centers [7]. They help to optimize the performance of contact centers by creating comprehensive schedules for agents. This problem is also important because the human agents within contact centers are the most expensive resources and therefore should be optimally used while maintaining acceptable service level and quality.

The staffing of a contemporary contact center is a very complex task. It should take into consideration different aspects of contact center operation and structure. These are multi-site contact centers, contact centers with multi-skilled agents, blended contact centers with simultaneous inbound and outbound calls, and multi-channel contact centers with multimedia interactions like e-mail and chat.

In this paper we consider the basic problem of staffing a single-site contact center with single-skilled agents. This problem will be stated and solved in terms of the Queuing System Theory [4]. The classical approach to similar problems is to use the Erlang C formula for $M/M/m$ model. However for real-world contact centers the Erlang C approach is not satisfactory because it does not take into consideration abandonment of calls that results from customer impatience. This factor can substantially change the contact centers performance and impact agent staffing.

The Erlang C model enhanced with an abandonment is denoted as $M/M/m+M$ and called Erlang-A. It was studied in several academic papers and books, e.g., see [2, 3, 5, 6, 8]. A comprehensive description of these results and analysis of corresponding literature were given in [5, 9].

In this paper we suggest application-oriented analytical methods of calculating contact center performance objectives based on $M/M/m+M$ model. The methods are based on exact formula of waiting time distribution. Based on this distribution we derive the formulae for calculation service level, abandonment rate, agent occupancy, average speed of answer and estimated waiting time.

2 System and Problem Statement

2.1 System Description

Schematically the system of inbound call processing is shown in Figure 1. Customers call in to a contact center. The contact center contains a set of human operators called agents whose mission is to process inbound calls. At any moment the agent may only be processing a single call.

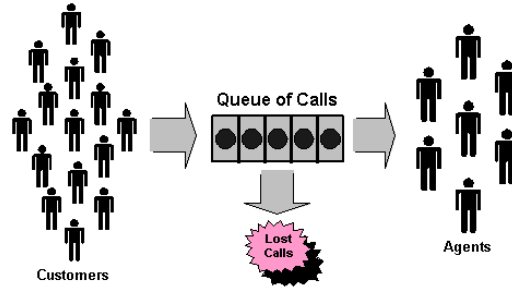


Fig. 1. Inbound Call Environment

Inbound calls are distributed to agents according to a routing strategy. Typically, the routing strategy is implemented as an ACD queue of a switch. If all agents are busy at the moment of call arrival, the call is placed into a queue where it waits for the next available agent. While the call is in the queue, an impatient customer can drop the connection by hanging up the receiver. This will result in removing a call from the queue. In a call model such calls are called *abandoned* because they are accompanied by the event **Abandoned**.

If an agent completes the processing of a call and a queue is not empty, he/she immediately receives the next call from the queue. Calls in a queue are processed according to a FIFO (First-In-First-Out) discipline. In this paper we will consider the case when the queue is unlimited.

2.2 Performance Objectives

One of main purposes of workforce management is to determine the optimal number of agents for processing inbound calls. Optimality is defined with respect to some performance objectives of the contact center. We use four such performance objectives simultaneously:

- *Service Level*. The service level is defined as the ratio of inbound calls that are answered within a specified time. Service level is defined as a tuple $\Lambda = (\gamma, \sigma)$ where γ is a ratio and σ is a specified time. Note that ratio γ takes into consideration all calls including abandoned calls. The Service Level objective is defined as the lowest boundary the ratio γ may reach while σ is fixed. Therefore we have the objective $\Lambda_{min} = (\gamma_{min}, \sigma)$.
- *Average Speed of Answer (ASA)*. ASA is the average time customers will wait for an answer from the contact center. For telephony calls this factor means the average waiting time calls wait in a queue for the next available agent. It should be stressed that only “successful” calls (or calls that reach agents) are taken into consideration. Abandoned calls are excluded from consideration. Average speed of answer will be denoted as W . The corresponding objective W_{max} means that the average speed of answer can not exceed some specified value.
- *Maximum Abandonment Rate*. The abandonment rate is defined as the rate of abandoned calls to all incoming calls. It will be denoted by letter A . The maximum abandonment rate will be denoted as A_{max} .
- *Maximum Occupancy Rate*. The occupancy rate is the ratio of time agents handle calls versus total active time. The active time includes only the time that agents wait for calls and handle calls. The handle time includes talk time and after-call work. Occupancy rate is usually referred to as agent utilization and will be denoted as U . The maximum occupancy rate that will be denoted as U_{max} defines an upper boundary for agent occupancy that cannot be exceed during contact center operation. This objective is needed so that agents are not overworked.

Typically a Workforce Management system uses these four objectives simultaneously. It is not hard to see that all of these objectives are not mutually contradictory. For instance, increasing the number of agents will result in increasing all four objectives.

2.3 Statement of Problem

Now the problem can be formulated as finding the number of agents $\bar{m} \in \{1, 2, 3, \dots\}$ so that all four objectives will be met. Moreover, this number should be minimal in the sense that the number $\bar{m} - 1$ will violate at least one objective.

Let us consider the statement of the problem more formally. Let $\gamma(\bar{m})$, $W(\bar{m})$, $A(\bar{m})$, and $U(\bar{m})$ be performance objectives at the number of agents $\bar{m}$. Then the problem is to find such $\bar{m}$ so that the following Boolean expression is true:

$$\begin{aligned} &(\gamma(\bar{m}) \geq \gamma_{min}) \wedge (W(\bar{m}) \leq W_{max}) \\ &\wedge (A(\bar{m}) \leq A_{max}) \wedge (U(\bar{m}) \leq U_{max}) = \text{true} \end{aligned} \tag{1}$$

The minimization of $\bar{m}$ is ensured by the fact that the expression (1) becomes not true for $\bar{m} - 1$.

2.4 Main Assumptions

It is not hard to see that the system described in the previous section can be represented with the aid of one of models of queuing system theory [4]. However, before choosing a model, let us make some important assumptions about the processes in the system. These assumptions will affect the model and corresponding calculations.

1. *Call arrival.* We will assume that call arrival is a stationary Poisson process with a parameter λ called average arrival rate.
2. *Service time distribution.* We will assume that distribution of service time is exponential with a parameter μ that is used in many queuing systems.
The parameter μ is called *an average service rate*. It is known that for telephony applications the exponential distribution fits quite well. Moreover, the exponential distribution is said to be “memoryless” which substantially simplifies calculations. The average service time is $T_S = 1/\mu$.
3. *Abandonment rate distribution.* The time that a customer will wait until abandoning the call is also assumed to be exponential with a parameter ν .

These assumptions lead us to multi-server model with exponential abandonment M/M/m+M. While call arrival can be successively modeled by a Poisson process, service time needs more careful consideration. Indeed, as analysis of real-world data shows [1, 2], the service time is better approximated by lognormal distribution. However, as pointed out in [1], the model M/M/m+M is quite robust and can give accurate calculations of service objectives.

3 Model of System without Nuisance Calls

Before considering a general model, let us consider a simplified case without abandoned calls that is usually denoted as M/M/m and is called the model with multi-server and with infinite queue. The service objectives for this model can be calculated using well-known formulae. See for example [4].

Processes of call arrival and call processing meet the assumptions made in the previous section: call arrival is a Poisson process with a parameter λ and service time is exponential with a service rate μ . Let m be the number of agents (servers). The length of the queue is assumed to be unlimited. For convenience let us define a new parameter

$$\rho = \frac{\lambda}{\mu} \quad (2)$$

called a traffic offer. For the system to be stable assume that $\rho < m$ or $\rho/m < 1$.

The key role in this model is played by the Erlang-C formula [4] that is defined as follows:

$$E_C(m, \rho) = \frac{\frac{\rho^m}{m!}}{\frac{\rho^m}{m!} + (1 - \frac{\rho}{m}) \sum_{k=0}^{m-1} \frac{\rho^k}{k!}} \quad (3)$$

Erlang-C defines a probability that all servers are busy, and therefore also the probability that an arriving call will have to wait.

The service objectives can be directly calculated using the following formulae:

Service Level The distribution of waiting time is a modified exponential, with the distribution:

$$P\{W \leq t\} = W(t) = 1 - E_C(m, \rho)e^{-(m\mu - \lambda)t} \quad (4)$$

Then the service level is defined as $A = (\gamma, \sigma)$ where

$$\gamma = 1 - E_C(m, \rho)e^{-(m\mu - \lambda)\sigma} \quad (5)$$

Average Speed of Answer ASA in this case is equivalent to the average waiting time in the queue and is calculated using the following formula:

$$W = \text{average waiting time} = \frac{E_C(m, \rho)}{\mu(m - \rho)} = \frac{E_C(m, \rho)}{\mu m - \lambda} \quad (6)$$

Abandonment Rate Clearly, in this case the abandonment rate is equal to zero.

Occupancy Ratio The occupancy ratio (or agent utilization) can be calculated by the next simple formula:

$$U = \frac{\lambda}{m\mu} = \frac{\rho}{m} \quad (7)$$

4 System with Nuisance Calls

In this section we consider a more complicated system where calls in a queue may leave the queue due to customer impatience. In the spirit of Kendal's notation usually this model is denoted as M/M/m+M and referred to as Erlang A see [2].

4.1 Model and Parameters

The model with nuisance calls is shown in the Figure 2. The queue length is not limited and therefore all arriving calls are placed into the queue.

Again, all processes in the system meet assumptions made above: arrival is a Poisson process with the parameter λ , service time is exponential with the service rate μ , and abandonment time is exponential with the rate ν . Let m be the number of agents (servers). The length of the queue is assumed to be unlimited.

For convenience we will use the following parameters:

$$\rho = \frac{\lambda}{\mu} \quad \text{and} \quad \beta = \frac{\nu}{\mu} \quad (8)$$

Note that this time we will not require $\rho < m$ because the system will be stable although $\rho \not< m$ due to the abandonment process.

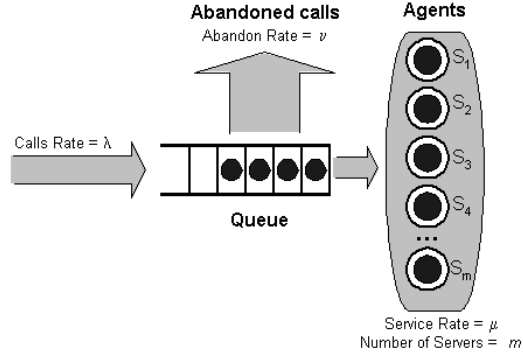


Fig. 2. Model with nuisance calls

4.2 Stationary Equations

The probability of k calls in the System is calculated with the aid of the formula:

$$P_k = \begin{cases} \frac{\rho^k}{k!} P_0 & \text{if } 1 \leq k \leq m \\ \frac{\rho^k}{m! \prod_{i=1}^{k-m} (m+i\beta)} P_0 & \text{if } k > m \end{cases} \quad (9)$$

where P_0 is the probability of zero calls in the System:

$$P_0 = \left[\sum_{k=0}^m \frac{\rho^k}{k!} + \sum_{k=m+1}^{\infty} \frac{\rho^k}{m! \prod_{i=1}^{k-m} (m+i\beta)} \right]^{-1} \quad (10)$$

These equations were derived using classical Birth-Death technique, see for example [3].

4.3 Some Auxiliary Formulae

In this section we present some auxiliary formulae that will be used for the calculation of performance objectives.

Average Queue Length The average length of a queue in the system can be calculated by the formula:

$$Q = \sum_{k=1}^{\infty} k P_{k+m} = P_0 \frac{1}{m!} \sum_{k=1}^{\infty} \frac{k \rho^k}{\prod_{i=1}^{k-m} (m+i\beta)} \quad (11)$$

Here we used the formula (9) for $k > m$.

Flow in the System The flow in the system is described by the obvious formula:

$$\lambda = \lambda_\mu + \lambda_\nu \quad (12)$$

The sub flow λ_μ is comprised of calls that reach the agent while the sub flow λ_ν relates to abandoned calls.

Distribution of Waiting Time for Persistent Calls To calculate service level $A = (\gamma, \sigma)$ we have to calculate the distribution function $P\{\tau \leq t\}$ of waiting time for persistent calls. We define persistent calls as calls that will wait for an agent regardless of what happens. In this case we suggest using the following formula¹:

$$P\{\tau \leq t\} = \sum_{k=0}^{m-1} P_k + \sum_{k=m}^{\infty} P_k P_k\{\tau \leq t\} \quad (13)$$

where

$$P_k\{\tau \leq t\} = \sum_{j=0}^{k-m} \frac{c_{j,k}}{\mu_j} (1 - e^{-\mu_j t}) \quad (14)$$

and

$$c_{j,k} = \mu_j \prod_{i=0, i \neq j}^{k-m} \frac{\mu_i}{\mu_i - \mu_j} \quad \text{and} \quad \mu_i = m\mu + i\nu \quad (15)$$

The first sum in the equation (13) is the probability that a call will not wait.

Distribution of Waiting Time for Successful Calls Notice however that the formulae above calculate PDF only for “persistent” calls, i.e. calls that cannot leave the queue. However, all calls can be abandoned and we need a distribution that takes this into consideration. We will call this a distribution of “successful” calls reflecting the fact that only call that reached agents will be taken into consideration. This distribution is calculated on the basis of a conditional distribution (provided the call is not abandoned):

$$P\{\tau \leq t \mid \tau < \xi\} = \frac{P(\{\tau \leq t\} \times \{\tau < \xi\})}{P(\tau < \xi)} \quad (16)$$

where

$$P(\{\tau \leq t\} \times \{\tau < \xi\}) = \sum_{k=0}^{m-1} P_k + \sum_{k=m}^{\infty} P_k P_k\{\{\tau \leq t\} \times \{\tau < \xi\}\} \quad (17)$$

$$P_k\{\{\tau \leq t\} \times \{\tau < \xi\}\} = \sum_{j=0}^{k-m} \frac{c_{j,k}}{\mu_j + \nu} (1 - e^{-(\mu_j + \nu)t}) \quad (18)$$

¹ The derivation of the formula is presented in Appendix A.2

and $c_{j,k}$ is calculated according to formulae (15).

The probability $P(\tau < \xi)$ can be calculated according the following formulae:

$$P(\tau < \xi) = \sum_{k=0}^{m-1} P_k + \sum_{k=m}^{\infty} P_k P_k \{\tau < \xi\} \quad (19)$$

$$P_k \{\tau < \xi\} = \sum_{j=0}^{k-m} \frac{c_{j,k}}{\mu_j + \nu} \quad (20)$$

Notice that the probability $P(\tau < \xi) = 1 - A$ where A is an abandonment rate and the formulae (19) and (20) can be used for the calculation of abandonment rate A .

The random value ξ is the time calls wait before abandoning (see Appendix A.2).

4.4 Abandonment Rate

Abandonment rate can be calculated using the formula:

$$A = \frac{\lambda_\nu}{\lambda} = \frac{\nu Q}{\lambda} \quad (21)$$

where the average queue length Q is calculated according to the formula (11). More strict derivation of abandonment rate is presented in Appendix A.3.

4.5 Service Level

Remember that we defined a service level parameter γ as the ratio of calls that reached agents within time σ to all calls including abandoned ones. More formally, we can define γ as a probability of two simultaneous events: the call reaches agents within time σ and the call is not abandoned. Therefore we can use the formula (17) to calculate the service level:

$$\gamma = \sum_{k=0}^{m-1} P_k + \sum_{k=m}^{\infty} P_k P_k \left\{ \{\tau \leq \sigma\} \times \{\tau < \xi\} \right\} \quad (22)$$

where

$$P_k \left\{ \{\tau \leq \sigma\} \times \{\tau < \xi\} \right\} = \sum_{j=0}^{k-m} \frac{c_{j,k}}{\mu_j + \nu} \left(1 - e^{-(\mu_j + \nu)\sigma} \right) \quad (23)$$

Here coefficients $c_{j,k}$ and μ_j are calculated according to formulae (15).

Note that these formulae are valid only if $\nu > 0$. In the case where $\nu = 0$ we suggest using formulae for the model M/M/m see Section 3.

4.6 Average Waiting Time of Persistent Calls

We define two average waiting times for incoming calls. First we consider “persistent” calls, i.e., calls that will wait for agents regardless of what happens. This value is very important and is known in the contact center industry as *estimated wait time*. It is used for announcing to a calling customer how long he/she would need to wait for the next available representative, and helps to increase the customers patience.

The average waiting time for these calls can be calculated by the formula that is derived from the distribution (13)-(14):

$$W = \sum_{k=m}^{\infty} P_k t_k \quad \text{where} \quad t_k = \sum_{j=0}^{k-m} \frac{c_{j,k}}{\mu_j^2} \quad (24)$$

4.7 Average Speed of Answer

The average waiting time discussed in the previous section relates to so-called “patient” customers. It can be used for estimating average waiting time where the system can say to the customer – “if you wait for an agent regardless of what happens, then you will wait x minutes”. However there is a probability (sometime substantial) that a waiting call can leave the queue due to abandonment. Such calls modify the distribution function of waiting time and therefore the actual average waiting time. To calculate this value called Average Speed of Answer we suggest using the formula derived from the distribution (16-20):

$$W = \frac{1}{1-A} \sum_{k=m}^{\infty} P_k t_k^* \quad \text{where} \quad t_k^* = \sum_{j=0}^{k-m} \frac{c_{j,k}}{(\mu_j + \nu)^2} \quad (25)$$

and A is the abandonment rate that can be calculated with the aid of the formula (21) or formulae (19) and (20).

4.8 Occupancy Ratio

The formula for occupancy ratio is the following:

$$U = \frac{\lambda - \nu Q}{m\mu} \quad (26)$$

where Q is calculated according to the formula (11). We can also express the occupancy ratio in terms of abandonment rate A :

$$U = \frac{\lambda}{m\mu} (1 - A) \quad (27)$$

5 Conclusion

In this paper we presented strict analytical methods for calculating performance objectives and optimal agent staffing for contact centers. The methods are easy to implement in software. We have considered only a simple case with a one-site, single-skilled inbound contact center. This solution will serve as the basis for more complicated cases like a multi-site and multi-skilled contact center environment.

References

1. Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Zeltyn, S., Zhao, L. and Haipeng, S. Statistical Analysis of a Telephone Call Center: A Queueing-Science Perspective. (2005) Journal of the American Statistical Association, Vol 100: 36-50.
2. Garnett O., Mandelbaum A. and Reiman M. Designing a Call Center with Impatient Customers. Manufacturing and Service Operations Management, 4(3), 208-227, 2002.
3. Gnedenko B.V., Kovalenko I.N. Introduction to Queueing System Theory. Moscow, Nauka, 1987 (in Russian).
4. Kleinrock L. Queueing Systems. Volume I. Theory. Wiley, New York, 1972.
5. Mandelbaum A. and Zeltyn S. Service Engineering in Action: The Palm/Erlang-A Queue, with Applications to Call Centers. Draft, March 2005.
Available at http://iew3.technion.ac.il/serveng/References/Erlang_A.pdf
6. Palm C. Research on telephone traffic carried by full availability groups. Tele, Vol. 1, 1957
7. Reynolds P. Call Center Staffing. The Complete, Practical Guide to Workforce Management. The Call Center School, 2003. ISBN 0-9744179-0-4.
8. Whitt W. Sensitivity of Performance in the Erlang A Model to Changes in the Model Parameters. Submitted to Operations Research. 2004.
9. Zeltyn S. Call Centers with Impatient Customers: Exact Analysis and Many-Server Asymptotics of the M/M/n+G queue. PhD Thesis, Technion, Haifa, Israel. Available at http://iew3.technion.ac.il/serveng/References/MMNG_thesis.pdf

A Appendix: Derivation of Formulae

A.1 Stationary Equations

To derive formulae for P_k for the system with an infinite queue and nuisance calls we will use Birth-Death technique [4]. General equations for Birth-Death system are:

$$P_k = \frac{\lambda_{k-1}}{\mu_k} P_{k-1} = \prod_{i=1}^k \frac{\lambda_{i-1}}{\mu_i} P_0 \quad (28)$$

$$P_0 = \frac{1}{1 + \sum_{k=1}^{\infty} \prod_{i=1}^k \frac{\lambda_{i-1}}{\mu_i}} \quad (29)$$

Here λ_k and μ_k are rates of birth and death respectively. In general, these rates are dependant on the number of calls in the system k .

Rates λ_k and μ_k for this system can be defined as follows:

$$\lambda_k = \lambda \quad (30)$$

$$\mu_k = \begin{cases} 0, & \text{if } k = 0; \\ k\mu, & \text{if } 1 \leq k \leq m \\ m\mu + \nu(k - m), & \text{if } k > m \end{cases} \quad (31)$$

Rewriting equations (28-29) for λ_k and μ_k defined in (30-31), we obtain the next equations of local equilibrium:

$$\begin{aligned} k\mu P_k &= \lambda P_{k-1}, \quad \text{if } 1 \leq k \leq m \\ (m\mu + (k - m)\nu) P_k &= \lambda P_{k-1}, \quad \text{if } k > m \end{aligned} \quad (32)$$

A.2 Waiting Time Distribution for Persistent Calls

Let us derive formulae (13-15) for waiting time distribution. We will use Laplace transformation techniques [4].

Calculate Laplace transformation of $P_k\{\tau \leq t\}$. Suppose a new call arrives to the system where k calls stay in the queue. The transportation of the call through the queue can be represented as a sequence of $k + 1$ servers, each i -th server having an exponential distribution with a parameter $\mu_i = m\mu + i\nu$, $0 \leq i \leq k$. The Laplace transformation will have the form:

$$F_k^*(s) = \prod_{i=0}^k \frac{\mu_i}{s + \mu_i}, \quad \text{where } \mu_i = m\mu + i\nu \quad (33)$$

Note that by definition $\mu_i \neq \mu_j$, $i \neq j$. To make the Laplace transformation possible to invert, let us transform the product (33) into a sum:

$$\prod_{i=0}^k \frac{\mu_i}{s + \mu_i} = \sum_{i=0}^k \frac{c_{i,k}}{s + \mu_i} \quad (34)$$

We have to determine coefficients $c_{i,k}$, $0 \leq i \leq k$. Multiply (34) on $(s + \mu_j)$ we obtain:

$$(s + \mu_j) \prod_{i=0}^k \frac{\mu_i}{s + \mu_i} = (s + \mu_j) \sum_{i=0}^k \frac{c_{i,k}}{s + \mu_i} \quad (35)$$

Making the multiplication we have:

$$\mu_j \prod_{i \neq j}^k \frac{\mu_i}{s + \mu_i} = c_{j,k} + \sum_{i=0, i \neq j}^k \left[\frac{c_{i,k}}{s + \mu_i} (s + \mu_j) \right] \quad (36)$$

Remembering that $\mu_i \neq \mu_j$, $i \neq j$ let us set $s = -\mu_j$. Then we obtain

$$c_{j,k} = \mu_j \prod_{i=0, i \neq j}^k \frac{\mu_i}{\mu_i - \mu_j} \quad (37)$$

Substituting μ_i and μ_j from (33) we have

$$c_{j,k} = (m\mu + j\nu) \prod_{i=0, i \neq j}^k \frac{m\mu + j\nu}{\nu(i - j)} \quad (38)$$

The probability density function (*pdf*) is obtained by Laplace transformation inversion:

$$p_k(x) = \sum_{i=0}^k c_i e^{-\mu_i x} \quad (39)$$

Finally, let us obtain the probability distribution function (PDF) by integrating *pdf*:

$$F_k(x) = \int_0^x \left[\sum_{i=0}^k c_i e^{-\mu_i t} \right] dt = \sum_{i=0}^k c_i \left[\int_0^x e^{-\mu_i t} dt \right] = \sum_{i=0}^k \frac{c_i}{\mu_i} (1 - e^{-\mu_i x}) \quad (40)$$

A.3 Waiting Time Distribution for Successful Calls

In this section we derive distribution formulae (16)-(18) for successful calls, i.e., calls that reached agents.

Let τ be a random value of the waiting time of calls that reached agents. The distribution function $F_\tau(t) = P\{\tau \leq t\}$ is defined according to formulae (13-15). We can easily calculate its PDF:

$$p_\tau(t) = \sum_{k=m}^{\infty} P_k P'_k \{\tau \leq t\} \quad \text{where} \quad P'_k \{\tau \leq t\} = \sum_{j=0}^{k-m} c_j e^{-\mu_j t} \quad (41)$$

Let ξ also be a random value related to the time calls spent in the queue before abandonment. According to our assumption it has an exponential distribution with the parameter ν : $F_\xi(t) = 1 - e^{-\nu t}$. Its PDF is as follows: $p_\xi(t) = \nu e^{-\nu t}$

We need to calculate a distribution function of a new random value $\zeta = (\tau \leq t \mid \tau < \xi)$. We are interested in those calls that have not been abandoned.

$$F_\zeta(t) = P\{\tau \leq t \mid \tau < \xi\} = \frac{P(\{\tau \leq t\} \times \{\tau < \xi\})}{P(\tau < \xi)} = \frac{\iint_{\mathcal{D}} p_\tau(x) p_\xi(y) dx dy}{\iint_{\mathcal{S}} p_\tau(x) p_\xi(y) dx dy} \quad (42)$$

Here $\mathcal{S}$ is a region defined as $\xi, \tau \geq 0$ and $\tau < \xi$. The region $\mathcal{D}$ is defined as $\xi, \tau \geq 0$, $\tau < \xi$ and $\tau \leq t$.

Let us calculate the denominator of the formula (42):

$$\iint_S p_\tau(x)p_\xi(y)dxdy = \int_0^\infty \int_x^\infty p_\tau(x)p_\xi(y)dydx \quad (43)$$

$$= \int_0^\infty \int_x^\infty \left[\sum_{k=m}^\infty P_k \left(\sum_{j=0}^{k-m} c_{j,k} e^{-\mu_i x} \right) \right] \nu e^{-\nu y} dydx \quad (44)$$

$$= \sum_{k=m}^\infty P_k \sum_{j=0}^{k-m} \frac{c_{j,k}}{\mu_i + \nu} \quad (45)$$

Let us calculate the numerator of the formula (42):

$$\iint_D p_\tau(x)p_\xi(y)dxdy = \sum_{k=m}^\infty P_k \left[\sum_{j=0}^{k-m} \frac{c_{j,k}}{\mu_j + \nu} \left(1 - e^{-(\mu_i + \nu)t} \right) \right] \quad (46)$$

Average Speed of Answering

$$W = \int_0^\infty t dF_\zeta(t) = \int_0^\infty t p_\zeta(t) dt = \sum_{k=m}^\infty P_k \sum_{j=0}^{k-m} \frac{c_{j,k}}{(\mu_i + \nu)^2} \quad (47)$$

Abandonment Rate Let us sum the equations (32). We obtain

$$\begin{aligned} \sum_{k=1}^\infty \lambda P_{k-1} &= \mu \sum_{i=1}^m i P_i + \sum_{i=m+1}^\infty (m\mu + (i-m)\nu) P_i \\ \lambda \sum_{k=1}^\infty P_{k-1} &= \mu \left(\sum_{i=1}^m i P_i + \sum_{i=m+1}^\infty m P_i \right) + \nu \sum_{i=m+1}^\infty (i-m) P_i \\ \lambda &= \mu \left(\sum_{i=1}^m i P_i + \sum_{i=m+1}^\infty m P_i \right) + \nu \sum_{j=1}^\infty j P_{j+m} \\ \lambda &= \mu \left(\sum_{i=1}^m i P_i + \sum_{i=m+1}^\infty m P_i \right) + \nu Q \end{aligned} \quad (48)$$