Master Thesis

Czech
Technical
University
in Prague

F3

Faculty of Electrical Engineering
Department of Telecommunication Engineering

Simulation Model of Customer Center

Bc. Ozgun Saydanoglu

Supervisor: Ing. Petr Hampl, Ph.D.
Field of study: Communications, Multimedia and Electronics
May 2019

MASTER‘S THESIS ASSIGNMENT

I. Personal and study details

Student's name: **Saydanoglu Ozgun** Personal ID number: **464325**

Faculty / Institute: **Faculty of Electrical Engineering**

Department / Institute: **Department of Circuit Theory**

Study program: **Communications, Multimedia, Electronics**

Branch of study: **Systems of Communication**

II. Master’s thesis details

Master’s thesis title in English:

Simulation model of customer center

Master’s thesis title in Czech:

Simulační model zákaznického centra

Guidelines:

Nowadays modern call centers, customers centers or contact centers typically process simultaneously inbound and outbound traffic. The inbound traffic is formed by customer requests and its first phase of service it is typically processed by Interactive Voice Response module. Then the inbound calls can be routed to dedicated groups of agents. The outbound traffic is formed by the connections initiated by agents of center to the costumers. Create a complex simulation model to determine key characteristics of those centers in terms of GoS. For the simulation model use the simulation environment OMNeT ++.

Bibliography / sources:

[1] STOLLETZ, R. Performance Analysis and Optimization of Inbound Call Centers. New York, Berlin: 2016. Springer-Verlag. 219 p. ISBN 3-540-00812-8.
[2] GROSS, Donald, John F. SHORTLE, James M. THOMPSON a Carl M. HARRIS. Fundamentals of Queueing Theory. Fourth edition. John Wiley & Sons, Inc., Hoboken, NJ, 2008. Wiley Series in Probability and Statistics. ISBN 978-0-471-79127-0.
[3] IVERSEN, Villy Bæk. Teletraffic engineering and network planning. DTU Fotonik, 2015. 382 p.

Name and workplace of master’s thesis supervisor:

Ing. Petr Hampl, Ph.D., Department of Telecommunications Engineering, FEE

Name and workplace of second master’s thesis supervisor or consultant:

Date of master’s thesis assignment: **10.10.2018** Deadline for master's thesis submission: **24.05.2019**

Assignment valid until: **28.02.2020**

Ing. Petr Hampl, Ph.D. Supervisor’s signature	doc. Ing. Radoslav Bortel, Ph.D. Head of department’s signature	prof. Ing. Pavel Ripka, CSc. Dean’s signature

III. Assignment receipt

The student acknowledges that the master’s thesis is an individual work. The student must produce his thesis without the assistance of others, with the exception of provided consultations. Within the master’s thesis, the author must state the names of consultants and include a list of references.

.

Date of assignment receipt

Student’s signature

Acknowledgements

I would like to thank to my supervisor, Petr Hampl who has guided and supported me throughout the work of this diploma thesis.

Declaration

I hereby declare that this master thesis presented is solely the result of my own work and I have only used the specified resources which are declared in the references.

In Prague, on 23.5.2019

....

Bc. Ozgun Saydanoglu

Abstract

The complexity of the recent customer center systems makes the usage of the simulation models for a performance measurement a potential solution. The service quality of a customer center is changed dynamically by the small variability. Implementation of the complex units such as interactive voice response, skill based routing and bidirectional - inbound and outbound - operation becomes a requirement. The goal of this thesis is to design a simulation model of a customer center by using a discrete event simulation tool $OMNeT++$ and to analyze important characteristics in terms of grade of service.

Keywords: customer center, skill based routing, inbound traffic, outbound traffic, interactive voice response

Supervisor: Ing. Petr Hampl, Ph.D.

Abstrakt

Klíčová slova: kontaktní centrum, smerování hovoru, príchozí provoz, odchozí provoz, interaktivní hlasová odezva

Překlad názvu: Simulacní model zákaznického centra

Contents

Figures

Tables

List of Notation

Symbol	**Meaning**
GoS	Grade of Service
ACD	Automatic Call Distributor
IVR	Interactive Voice Response
FIFO	First in First out
n_i	Total number of inbound flow types
n_o	Total number of outbound flow types
G	Total number of call agent groups
g_j	j^{th} call agent group
N	Total number of trunk lines
λ_i	Inbound call intensity
t_{ia}	Mean interarrival time between subsequent calls
C_o	Total number of inbound offered calls
C_b	Total number of inbound blocked calls
C_a	Total number of inbound admitted calls
B	Inbound call congestion
θ	Intensity of service in IVR
t_{ivr}	Mean service time in IVR
$C_{ivr\,j}$	Total number of inbound calls from j^{th} flow that served in IVR
p_j	Probability of an inbound call from j^{th} flow to be served in IVR
C_{ia}	Total number of calls that are routed from IVR
C_t	Total number of calls waited more than t seconds
S_j	The number of call agents from j^{th} agent group
ca_s	Status of an agent
$t_{qs_j,i}$	The time that an individual call from j^{th} flow enter to queue
$t_{s_j,i}$	The time that an individual call from j^{th} flow is allocated to an agent
$t_{w_j,i}$	Waiting time in the queue for an individual call from j^{th} flow
t_{wt}	Total waiting time in the queue for the inbound calls
C_{wj}	Total number of calls from j^{th} flow waited inside the queue
$E[W_o]$	Expected waiting time related to the offered calls
$E[W]$	Expected waiting time related to the admitted calls
$E[W_{ia}]$	Expected waiting time related to the calls routed from IVR

Symbol	**Meaning**
$E[W_w]$	Expected waiting time related to the calls encountered waiting in the queue
$P(W_o > 0)$	Probability of waiting related to the offered calls
$P(W > 0)$	Probability of waiting related to the admitted calls
$P(W_{ia} > 0)$	Probability of waiting related to the calls routed from IVR
$P(W_o > t)$	Probability of waiting more than t seconds related to the offered calls
$P(W > t)$	Probability of waiting more than t seconds related to the admitted calls
$P(W_{ia} > t)$	Probability of waiting more than t seconds related to the calls that are routed from IVR
C_{rj}	Total number of abandoned calls from j^{th} inbound flow
p_{rj}	Probability of a call from j^{th} inbound flow to abandon
t_{rj}	Time limit for j^{th} inbound flow to abandon the queue
R_j	Rate of abandonment for j^{th} inbound flow
μ	Intensity of service in the talk phase
t_a	Mean talk time
α	Intensity of service in the wrap up work phase
t_b	Mean wrap up time
β	Intensity of service in outbound serve phase
t_o	Mean service time for outbound flows
λ_{oc}	Outbound call intensity
t_{ioc}	Mean interarrival time between subsequent outbound calls
λ_{oe}	Outbound e-mail intensity
t_{ioe}	Mean interarrival time between subsequent outbound e-mails
γ_c	Outbound call congestion
γ_e	Outbound e-mail congestion
C_{osc}	Total number of served outbound calls
C_{ose}	Total number of served outbound e-mails
C_{oc}	Total number of generated outbound calls
C_{oe}	Total number of generated outbound e-mails
M	Exponential interarrival times
D	Deterministic interarrival times

Chapter 1

Introduction

The simple analytical model $M/M/N/R$ - Erlang-C - does not meet the expectations to use as a management tool for the recent customer center environments with the increasing complexity of these structures due to the fast developments in technology. Erlang-C model assumes a single agent group with a single flow type where all the incoming requests are resolved inside the interval that they arrived in which is no longer a valid assumption.

The recent customer center environments include all the mechanisms such as interactive voice response and automatic call distributor and comprise from skill based agent groups and categorized call flow types. It is a requirement to make a forecast about the performance measurements such as the service level by changing the traffic intensity per flow type or the staffing level based on the categorized agent groups. The forecast is required especially implementing any changes before going into the production. Finding the best time to operate the outbound flow is another important concern that needs to be estimated. The change in the staffing level or the traffic intensity might differ per interval and it is important to set these varieties to have an accurate estimation. Since there are many parameters that might impact the operation of the system, it is valuable to know the impacts of these changing parameters. Since there might be need for higher service level for a certain flow type than another, the outcome of the impacts when managing the structure is required to know before going into the production.

There are no general simulators to provide a solution for the modern contact center environments. The diploma thesis devotes to the design of a complex simulation model and analyzing the important characteristic of the customer center in the terms of grade of service (GoS). The design of the model features multi skill - classified agent groups and the categorized traffic flows - customer center operating on inbound, outbound and mixed - inbound and outbound simultaneously - directions. The discrete event simulation tool $OMNeT++$ is used to create the simulation model. The configuration of the simulation model should be done on the correct requirements and should produce accurate results.

$OMNeT++$ is an object-oriented discrete simulation tool and the core of the network is comprised from modules. The modules communicate with each other by passing messages and they are programmed in $C++$.

Chapter 2

Classification of a Customer Center

2.1 Basis of a customer center

A call center is defined as a system that provides services to customers by human agents over a telephone line. By the time with increasing demand to the call centers, it became inevitable to implement more enhanced structure including different kind of communication channels, extended size, categorized services and full time availability. Such type of service models are so called customer center or contact center [23]. Customer center is a complex and dynamic environment with lots of variability where the service is provided for all the requests via primarily telephone calls and various communication channels such as e-mail, chat, social media messaging and help desk software. Customer center is a service system and keeping a consistent service level is a priority while managing the staffing level efficiently [18].

Call agents and trunk lines are the fundamental resources of a customer center structure and a small change in the staffing level shows a significant impact on the cost of a customer center and the GoS [16].

A customer center can be homogeneous or heterogeneous based on the customer and the agent profile [9]. In the homogeneous type of customer center, the calls are collected in a common pool and the call agents are serving to all type of calls inside the pool. Figure 2.1 depicts a single call agent group that is handling all of the call arrivals from the general pool.

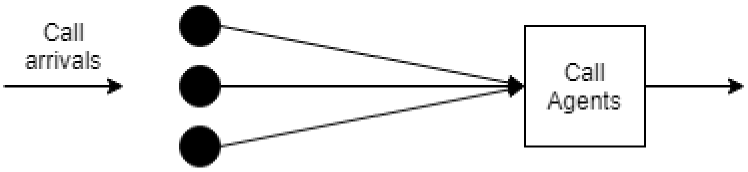

Figure 2.1: Homogeneous call arrivals and agents

In the heterogeneous customer center models, the call arrivals are classified into the classes and the call agents are trained to handle specific type of call requests in different categories. The call agents are classified into the groups according to their skill sets. For example as it can be seen in Figure 2.2 the call arrivals are divided into three types and call agent group 1 is serving only to arrival type 1 while call agent group 2 is serving only for arrival type 2 and agent group 3 is skilled for type 3 calls. In these types of customer center structures, the implementation of the skill based routing is essential. Routing the specific call types towards specific call agent groups is accomplished by the functionality of ACD (Automatic Call Distributor) system [19]. These types of customer center structures are so called multi skilled customer centers.

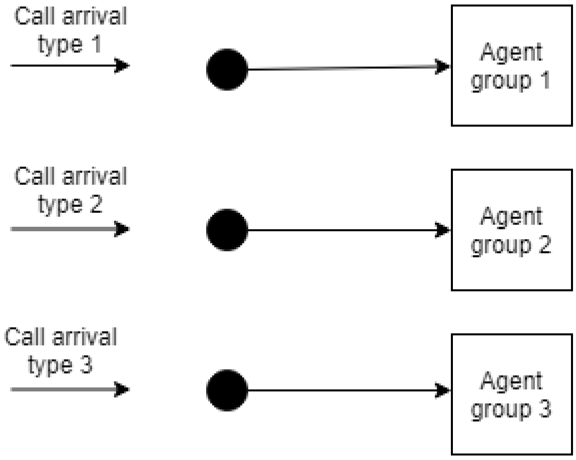

Figure 2.2: Heterogeneous call arrivals and agents

The communication is made between the customer and the customer center. Based on which side reaches out to another side customer centers are categorized to perform the service process to the traffic flow in three types

- Inbound customer center
- Outbound customer center
- Mixed customer center

An inbound customer center is provided by the companies and is operating for answering requirements, expectations and demands of the customers. These requests might include a vendor support, getting information about a product or services and help desk inquiries. In this type of customer center, the communication requests are incoming from the customers into the customer center facility. It is unknown to the service providers when and how many calls will arrive in a certain time. It is a random process and only a traffic forecast can be made about the arrival times. In an outbound call center, the approach to initiate the communication is made by the customer center individuals with the purpose of market researching and advertising products or services. In this type of model, agents who are reaching out to their target customers are planning the call based on their agenda. In mixed customer centers, inbound and outbound type of service is operated simultaneously [21]. An agent should be able to handle inbound requests and perform outbound services. In this type of environments it is also possible to divide some of agent groups as handling only the inbound traffic or outbound traffic and to allocate some types of agent groups to operate for both inbound and outbound.

2.2 Compounds of a customer center

The core modules that forms a customer center system in the current implementations are

- IVR (Interactive Voice Response)
- ACD
- Agent groups with specific skill sets
- Skill based routing

IVR is a mechanism that provides automated service via speech recognition or touch pad technology. The customers are able to navigate themselves through the IVR menu and seek the most suitable service category [2]. The agents are trained to handle different service categories and when the customer chooses the requested service category, the routing path is directed by the ACD system.

There are many factors that needs to be taken into account during the operation process. Figure 2.3 depicts these factors during the call transition.

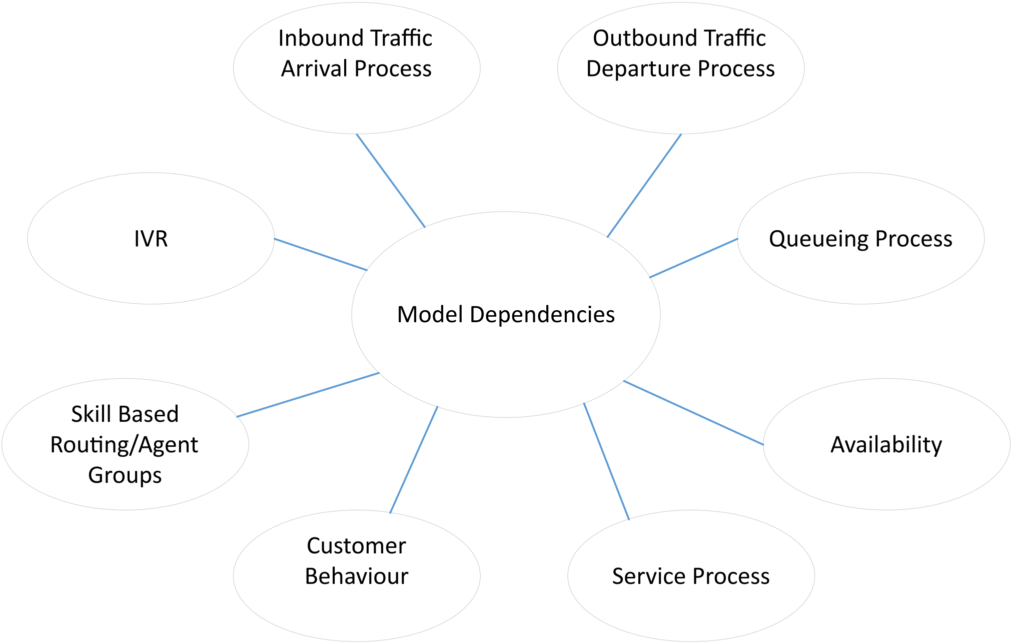

Figure 2.3: Fundamentals of a customer center model

The customer center structure can be considered as a queuing model when modeling the structure fundamentally. Arrival process and a flow character of an inbound and outbound flows are stochastic processes where a call could come at any time in a continuous time. The individual call arrivals are considered to be independent or partially dependent from each other. The system is not always available to accept the incoming requests or making the outgoing requests. The availability is determined by the utilization of the resources. The structure works as a delay and loss system [7] and when the resources are fully occupied the system is going to have a restricted availability until the service process is done for the events that are in progress.

The IVR functionality adds a self service progress into the service process [14] and gives a possibility for the customers to complete the requests without encountering agents. The queuing process can be with different queuing disciplines either preemptive or without preemptive queuing. The maximum queue capacity is based on the total number of trunk lines and the number of agents. The customers who are waiting in the queue should be considered as patient or impatient. The classification of the impatient customers might occur according to reneging or balking principles.

The call agents are divided into call agent groups. The classification is made by in which category the call agent is specialized with. The inbound calls should be categorized during the self service and should be routed to respective agent groups.

Chapter 3

Simulation Model of a Mixed Multi Skilled Customer Center

The complexity of the customer centers are making the usage of simulation models valuable. The simple analytical queuing model M/M/N/R [13] can be applied for the modeling of the customer center. However there are some restraints using the analytical model in comparison to the simulation model. The analytical model switches from one state to another at a time. By using a simulation model we can construct a more complex customer center structure. The complexity occurs since the system includes different call types, various agent groups, different handling times for different agent groups, different intensity for different call types and different handling times by the IVR for different call types.

In the analytical model the handling time is examined in one state where the wrap up work process is merged with the talking process. The state of the trunk line is considered as fully occupied during the entire service process. In this simulation model the handling time is divided into two states where the trunk line is only occupied during the first state of the service and these different phases are dynamically transitioned as it is depicted in Figure 3.3. Furthermore the analytical queuing model makes the modeling possible for only one type of agent group where the traffic flows are grouped into only one class. The analytical model is examined as a homogeneous system where the general call arrivals are handled by the general agent group. In contrast we can model skill based routing and skill based agent groups by using a simulation model [8].

This thesis develops a multi skilled and inbound and outbound blended customer center model with IVR by using discrete event simulation tool $OMNeT++$. In the analytical call center model as cited in [12], there is one type of incoming traffic flow and one type of call agent group. Since one of the challenges while designing a simulation model is to create the model with correct requirements so that it produces accurate results, the results are compared with a known model as cited in [12]. The comparison of the results are shown and explained in the Section 4.1.

In this thesis skill based agent groups are implemented where the arrival of the call flows are categorized into different types. The call flows are routed to the respective agent groups based on skills [20]. Furthermore in this thesis the call agent groups are able to be composed as serving only for incoming flow or only for outgoing flow or blended - incoming and outgoing flows simultaneously - and the mapping of the agent group patterns are depicted in Table 3.1 where the call agents from group A are handling only inbound and the call agents from group B handling only outbound services. However the call agents from group C are considered as skilled to handle both directional flows and while handling the incoming calls, also making outbound calls and e-mails. In this thesis the impatience of the customers are taken into consideration according to balking principle and the impatience customers abandoned the queue.

Agent Group Type	**Incoming Flow**	**Outgoing Flow**
A	X	
B		X
C	X	X

Table 3.1: Mapping of agent groups

3.1 Definition of the simulation model

In this chapter the implementation of the simulation model is explained by including the network topology, flowchart diagram, the parameter analysis and the details of the core modules that build the customer center network.

3.1.1 Network topology

The network topology is depicted in Figure 3.1.

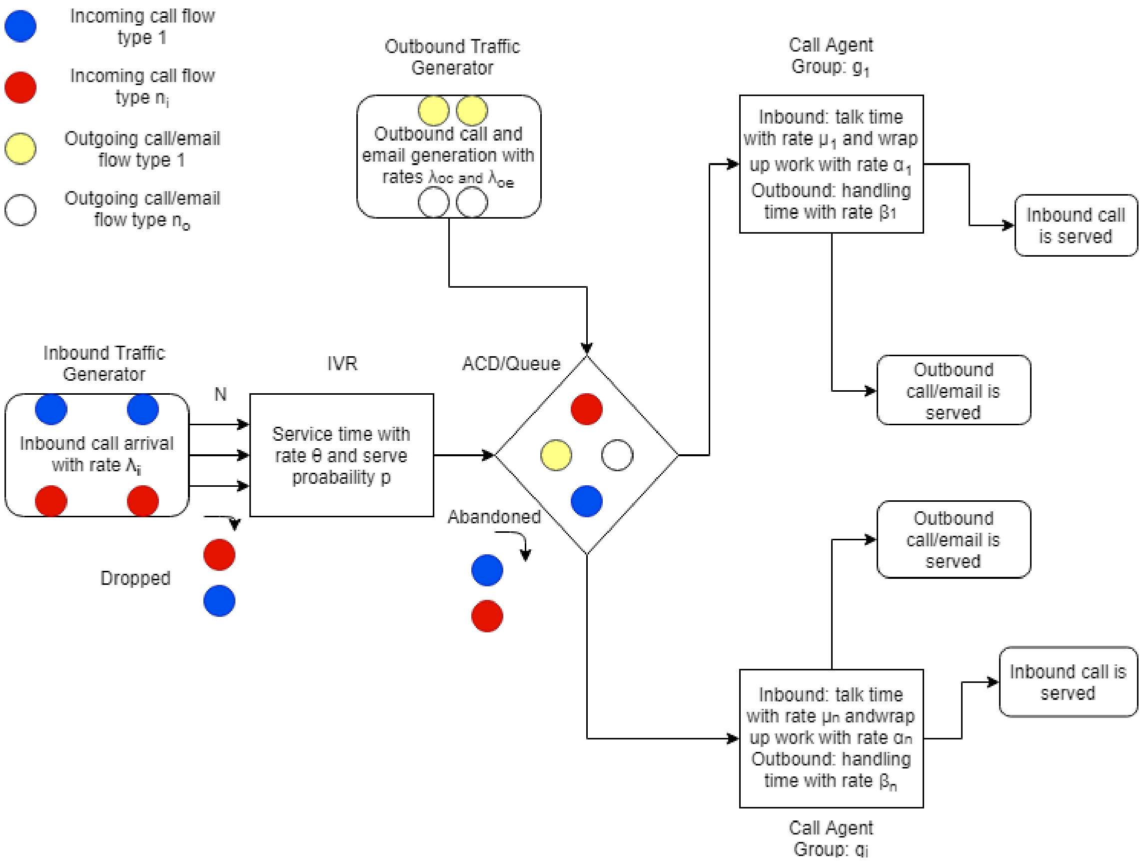

Figure 3.1: Network Diagram

The total number of incoming and outgoing flow types are represented with n_i and n_o, respectively. n_i and n_o gives the upper bound for the inbound and outbound traffic flows. The total number of agent groups are represented with G and a particular call agent group is identified by g_j where $j = 1, .., G$. Table 3.2 represents an example customer center system with $G = 7$, $n_i = 6$ and $n_o = 4$.

Agent Group	Agent Group Type
g_1	A
g_2	C
g_3	B
g_4	A
g_5	C
g_6	C
g_7	A

Table 3.2: Sample model with G=7

The core modules that compose the network are

- Inbound Traffic Generator

- IVR
- ACD (Routing and Queuing)
- Call Agent Groups
- Outbound Traffic Generator

.

Figure 3.2 illustrates the flow of the service process for the inbound and outbound requests.

The traffic is generated for all the categorized types of incoming call flows at a distinctive intensity [10]. The generated calls are transmitted into the IVR module and checked if the customer center is able to accept these calls. If there is any idle trunk line, the call is accepted into the system and starts with the first step of the service process. In case there are no idle trunk lines then the call is received a busy signal and blocked. The calls that are admitted start with the self service progress in the IVR module. There is a probability that is distinctive to each categorized flow types that a customer might resolve the incident by self service in IVR and exits the system. In case the incident is not resolved via self service then the call is routed into the queue module. Inside the queue module an idle call agent is searched upon arrival of the call. If there is any idle call agent then the conversation starts with the call agent without any waiting in the queue. If there is no call agent available upon call arrival then the call is waiting in the queue and as soon as one of the call agent becomes available, the service by an agent starts. If the customer is impatience then there is a probability that the customer is going to abandon the system after waiting more than mean waiting time and the call is going to be dropped from the system. The handling time of the call by an agent has two steps, first talk time where a conversation takes place between the agent and the customer. When the talk time stage is over the customer leaves the system. But the service for this incident is not yet completed. The call agent starts with the wrap up work for this request. Once the wrap up work is completed then the incident is completely resolved for this call and the agent becomes available from busy status.

Outbound traffic includes two sub-classes, call and e-mail flows, respectively. The generated calls and e-mails are transmitted to the queue module and stored in the outbound queue. If there is a call agent group B that is operating only for outbound direction services, then an idle call agent is searched and calls and e-mails are allocated immediately upon finding an idle call agent. However in case of the call agent group C that is operating for mixed type of services, then not only an idle agent is searched but also the buffer of the inbound queue is checked. If there is any incoming call is buffered and there is an idle agent then the outgoing call or e-mail is going to be stored in the outbound queue and the service is going to begin for the incoming call that is buffered inside the inbound queue [6]. The handling time for the outgoing calls and e-mails are equivalent to each other and distinctive for each flow type. The service process by the call agents for the outbound flows is treated in one phase in contrast to inbound calls.

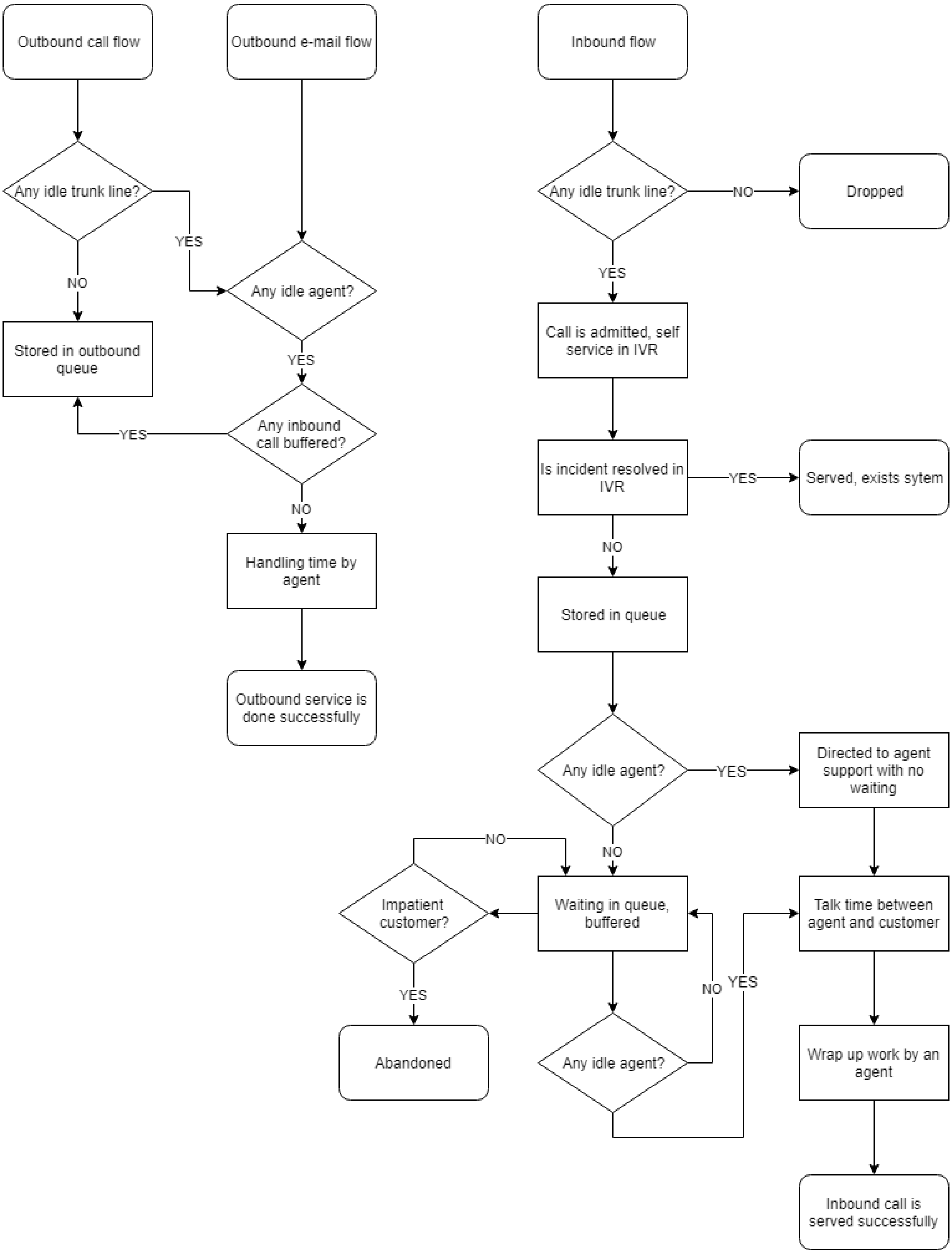

Figure 3.2: Flow Chart

3.1.2 Arrival process of the incoming flow

Considering the incoming call arrivals, the time between the subsequent calls which means the interarrival times are exponentially distributed [11]. There are number of calls could potentially arrive in a certain interval and the success is defined for each call whether or not the call is made. The success of a random variable could be at any time in continuous time based on the intensity and this doesn't give any information about the next call generation [3]. No matter how long is waited a new call arrival is considered as starting from fresh. Number of call arrivals in disjoint intervals are independent and identical from each other and defined by exponential random variables with the intensity λ_i [10] and the cumulative distribution function is as following

$$F(t;\lambda_i) = \begin{cases} 1 - e^{-\lambda_i t} = 1 - e^{-\frac{t}{t_{ia}}} & ,\ t \geq 0, \\ 0 & ,\ t < 0, \end{cases} \tag{3.1}$$

where t_{ia} represents the mean interarrival time between the consecutive events [18]. Ideally a day(s)/week(s) long interval is split into 15/30/60 minutes intervals when modeling a customer center simulation [1]. The disjoint intervals can be chosen with a constant time length or with different duration in each interval over the simulation run. In this model the parameters are updated in every interval dynamically. That makes it flexible to divide an interval into number of intervals each with particular λ_i.

The total number of calls that are counted or attempted into the system over an entire simulation run is represented by C_o

$$C_o = \sum_{j=1}^{n_i} C_{oj} \tag{3.2}$$

where C_{oj} represents the total number of offered calls for the j^{th} input flow type where $j = 1, .., n_i$. When the system receives an incoming call request, the outcome is determined whether the call is going to be accepted or rejected is based on the availability of the system. The availability of a customer center is dependent to the status of the trunk lines whether there is any free or all occupied [22]. The total number of trunk lines is represented by N. If there is at least one idle trunk line then the call is going to be accepted and is going to enter to the system. The call is going to be occupying a trunk line until it leaves the system. In Figure 3.3 the actions that occupy a trunk line and maintain it in the idle status during the service process for both inbound and outbound flows are illustrated.

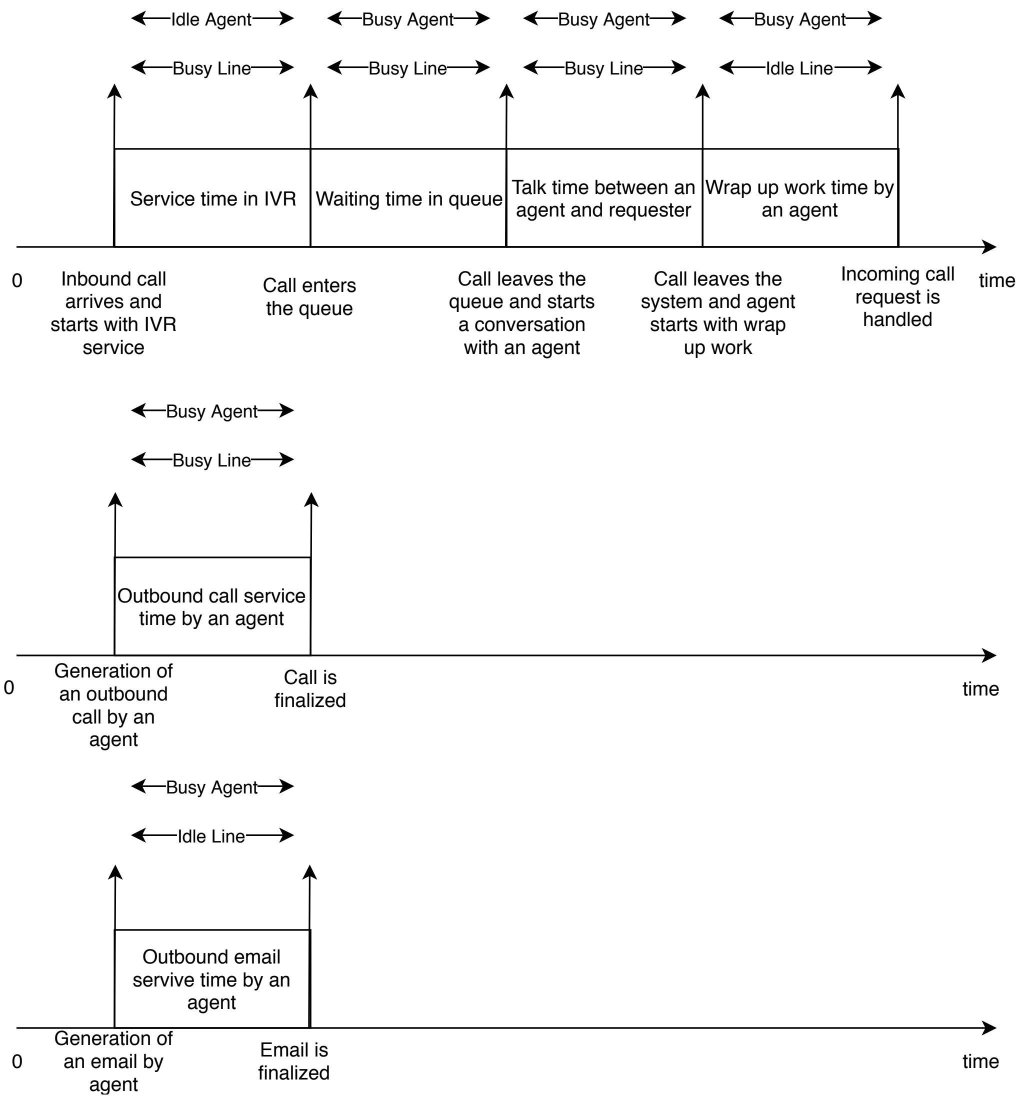

Figure 3.3: Trunk line and call agent utilization during the service process

Trunk line is in allocated status during the following events

- IVR service
- Queuing process
- Conversation time with an agent for the inbound call flow
- Service time of an agent for the outbound call flow

At the beginning of the following service processes the trunk line is not used and is going to make it available to carry a new incoming or outgoing service event.

- Wrap up time of an agent for the inbound call
- Service time of an agent for the outbound e-mail flow

3.1.3 IVR

A call is blocked if from N available trunk lines none of the trunk line is free at the time of the arrival and the caller is given a busy signal. Total number of calls that are blocked is represented with C_b where

$$C_b = \sum_{j=1}^{n_i} C_{bj} = C_o - C_a = BC_o \tag{3.3}$$

where C_a is the total number of calls that are admitted, accepted into the system. The number of calls that are admitted into the system is defined by

$$C_a = \sum_{j=1}^{n_i} C_{aj} = C_o - C_b \tag{3.4}$$

The inbound call congestion which means the ratio of the calls that are blocked from entering into the system is represented with B and defined by

$$B = \frac{C_o - C_b}{C_o} = 1 - \frac{C_b}{C_o} \tag{3.5}$$

When the call is accepted by the system it is switched into the IVR module and starts with the self service process. There are two likely outcomes that could happen at the end of the completion of self service which

- Inbound call is served by IVR and exits the system
- Inbound call is routed into the next stage for agent assistance

In the IVR module there is an automated service process for the inbound calls. Service time of an IVR is following exponential distribution with the mean intensity θ and the CDF

$$F(t;\theta) = \begin{cases} 1 - e^{-\theta t} = 1 - e^{-\frac{t}{t_{ivr}}} & , \ t \geq 0, \\ 0 & , \ t < 0, \end{cases} \tag{3.6}$$

where t_{ivr} represents the mean service time of the IVR module. In this model t_{ivr} is considered to be identical for all the different types of incoming call flows. Afterwards to the conclusion of the service, there is a possibility that the customer request will be resolved in this module without further assistance of a call agent. The call requests made by input group j are resolved inside the IVR module with a probability of p_j where $j = 1, .., n_i$.

Probability p_j is distinctive for each one of the inbound call flows and might be specified a different value from each other. The calls exits the system after they are served in the IVR module. The number of calls that are served by the IVR module from a i^{th} inbound flow is represented by C_{ivrj} where $j = 1, ..n_i$ and is defined by

$$C_{ivrj} = C_{aj} p_j \tag{3.7}$$

$1 - p_j$ is the probability that the call request from the inbound call group j is not resolved by the IVR module and is going to require to get an assistance from a call agent. Towards start of the contact with the call agent, the call is routed into the queue module. C_{iaj} represents the number of calls from i^{th} inbound flow that are routed into the next phase of service, queue module. C_{iaj} includes the number of calls that encountered a waiting in the queue before receiving a assistance from a call agent and number of calls that started the communication with a call agent without any waiting in the queue and is defined by

$$C_{iaj} = (1 - p_j)C_{aj} \tag{3.8}$$

3.2 Queue module

The inbound calls are passed into the queue module before the assistance of a call agent. The routing path of the calls are skill based according to the type of the call. The calls are distributed into separate queue channels. For instance, type 1 calls are stored in the channel 1 and the type 2 calls are stored in the channel 2. The call agent groups are categorized skill based. Individuals from call agent groups are handling only call types that are corresponding to their skill set. While an individual from agent group 1 is only serving to type 1 calls, an individual from agent group 2 is only serving to the type 2 calls. The waiting time of a call in the queuing module depends on the availability of the call agents. If there is at least one free call agent corresponding to the type the call belongs to, then the call is directed to get an assistance from a call agent immediately for the service. In this case the waiting time in the queue is calculated as zero. Otherwise, if there is no call agent available to start with the service, the call is stored in the queue and the waiting time is larger than zero.

The queuing discipline is in the means of FIFO (First in First out) as depicted in Figure 3.4. The call with the longest waiting time is popped from the queue when a call agent becomes available. When a call is arrived it is pushed to the back of the queue and positioned in the queue buffer. When a call is allocated to an agent, the buffer of the queue is cleared.

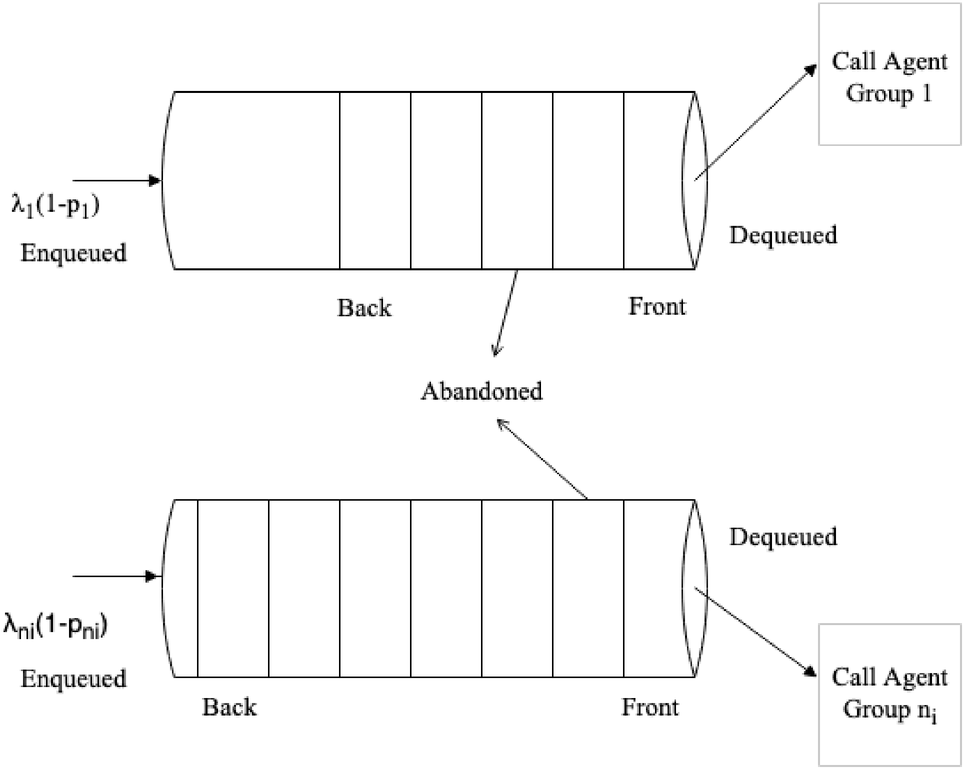

Figure 3.4: Queuing Discipline

One of the important part is allocating an incoming call to an available call agent. When a call is arrived into the queue module, it needs to be assigned to one of the available agents for the service. The allocation function works in a way that it selects a random agent from the range of total number of agents where the number of agents is represented with S_j where $j = 1, .., G$ and the number of agents is defined distinctively for each call agent group. After the random call agent is selected, the availability of this agent is checked. The status of an individual agent is represented with ca_s. If ca_s for a particular agent equals to 0 then the agent is in available status, otherwise the agent is in busy status. If the random agent that the allocation function finds is in free status then the call is assigned for this particular agent. Otherwise the allocation function will seek through other agents and is going to assign the call to the first random agent who is available. Until there is a success, this action will be recurring and as long as no agents are available, the call keeps waiting in the queue. When a call agent individual completes handling a job then he/she is going to check for pending jobs and starts with the handling of this job if any.

The current time that i^{th} call entered to the queue from a distinctive flow j is represented with $t_{qs_j,i}$ where $j = 1, .., n_i$ and the current time that a call is allocated to the call agent is defined by $t_{s_j,i}$ where $j = 1, .., n_i$.

The waiting time in the queue for i^{th} call from j^{th} flow is defined by $t_{w_j,i}$ where $i = 1, .., Cw_j$ and $j = 1, .., n_i$. Waiting time is greater than zero if only there is no available call agent upon call arrival into the queue. $t_{w_j,i}$ is defined by

$$t_{w_j,i} = t_{s_j,i} - t_{qs_j,i} \tag{3.9}$$

t_{wt} represents the sum of the total waited time inside the queue from all the calls from all n_i input flows and it is defined by

$$t_{wt} = \sum_{j=1}^{n_i} \sum_{i=1}^{C_{w_j}} t_{w_j,i} \tag{3.10}$$

where C_{wj} represents the total number of calls that are queued with waiting time greater than zero from particular flow types where $j = 1, .., n$.

The expected waiting time related to the offered calls represented with $E[W_o]$ where

$$E[W_o] = \frac{t_{wt}}{C_o} \tag{3.11}$$

The expected waiting time related to the calls that are entered to the system and served by IVR module or the call agent is represented with $E[W]$ and is defined by

$$E[W] = \frac{t_{wt}}{C_o - C_b} \tag{3.12}$$

The expected waiting time related to the calls that are served by call agents is represented with $E[W_{ia}]$ where

$$E[W_{ia}] = \frac{t_{wt}}{C_{ia}} \tag{3.13}$$

The expected waiting time related to all the calls that are served by call agents and faced a waiting in the queue is represented with $E[W_w]$ where

$$E[W_w] = \frac{t_{wt}}{C_w} \tag{3.14}$$

The probability that a call is going to wait in the queue related to the offered calls which means just before attempted for entering to the system is represented with $P(W_o > 0)$ where

$$P(W_o > 0) = \frac{C_w}{C_o} \tag{3.15}$$

The probability that a call is going to wait in the queue related to the admitted calls is represented with $P(W > 0)$ where

$$P(W > 0) = \frac{C_w}{C_o - C_b} \tag{3.16}$$

The probability that a call is going to wait in the queue after passing the IVR service stage is represented with $P(W_{ia} > 0)$ where

$$P(W_{ia} > 0) = \frac{C_w}{C_{ia}} \tag{3.17}$$

The probability of waiting after a certain time value is an important decision result parameter regarding service level. This time value is represented with t. The number of calls that are counted with a waiting greater than t is represented with C_t where

$$\begin{cases} C_t & , \ t > E[W_w], \\ 0 & , \ t \leq E[W_w], \end{cases} \tag{3.18}$$

The probability that a call will wait more than t seconds related to offered calls is represented with $P(W_o > t)$ where

$$P(W_o > t) = \frac{C_t}{C_o} \tag{3.19}$$

The probability that a call will wait more than t_w seconds related to admitted calls is represented with $P(W > t)$ where

$$P(W > t) = \frac{C_t}{C_o - C_b} \tag{3.20}$$

The probability that a call will wait more than t seconds after passing the IVR service stage is represented with $P(W_{ia} > t)$ where

$$P(W_{ia} > t) = \frac{C_t}{C_{ia}} \tag{3.21}$$

3.2.1 Abandonment from the queue

In this model the impatience of the customers who are making inbound calls are taken into account in the basis of reneging. For each type of incoming flow there is some probability that a customer will abandon the queue if a given time limit value is greater than the waiting time of an individual call [5]. The mechanism for checking the condition is run once after the $E[W_w]$ is computed. The number of calls that are lost from j^{th} inbound flow type due to abandonment from the queue is represented with C_{rj} where $j = 1, .., n_i$ and is defined by

$$C_{rj} = \begin{cases} p_{rj} C_{wj} & , \ t_{rj} > E[W_w], \\ 0 & , \ t_{rj} \leq E[W_w] \end{cases} \tag{3.22}$$

where t_{rj} represents the limitation time value and p_{rj} represents the probability of abandonment from j^{th} inbound flow type where $j = 1, .., n_i$, respectively. The rate of abandonment for j^{th} flow is represented with R_j where $j = 1, .., n_i$ and defined by

$$R_j = \frac{C_{rj}}{C_{iaj}} \tag{3.23}$$

The customers who abandoned the queue considered to be lost and not retry calling back immediately upon hang up.

3.3 Call agent groups

In this simulation model three types of call agent groups are modeled

- Agent group handling only inbound traffic
- Agent group making only outbound traffic
- Agent group handling inbound and making outbound traffic

In this section the descriptions are made considering the agent groups are skilled for both direction of the traffic. All of the agents in particular groups are able to handle inbound traffic and make outbound traffic.

In this model the number of agents are updated in every time interval dynamically. For instance in case of the changes in staffing level in disjoint intervals the number of agents are updated dynamically.

3.3.1 Process of handling inbound flow by the agent

The handling time of an incoming call by a call agent is divided into two steps

- Talk time
- Wrap-up time

The service process for an incoming call starts with the talk time. When an agent is assigned for the call the communication with the customer starts. During the talk time the agent is busy and the inbound call occupies a trunk line. Service time of an agent for this stage follows an exponential distribution with the intensity μ [17] and the cumulative distribution function is defined as

$$F(t;\mu) = \begin{cases} 1 - e^{-\mu t} = 1 - e^{-\frac{t}{t_a}} & , \; t \geq 0, \\ 0 & , \; t < 0, \end{cases} \tag{3.24}$$

where t_a represents the mean service time of a call agent during the talk time stage. After the talk time is completed the phone is hanged up and the customer leaves the system which makes the trunk line free. However the call is not yet served completely, the call agent is still busy for this call. The utilization of the call agents is depicted in Figure 3.3. The call agent starts with second stage of the service which is wrap-up work. Service time of an agent for wrap-up work time follows an exponential distribution with intensity α and the cumulative distribution function is defined as

$$F(t;\alpha) = \begin{cases} 1 - e^{-\alpha t} = 1 - e^{-\frac{t}{t_b}} & , \; t \geq 0, \\ 0 & , \; t < 0, \end{cases} \tag{3.25}$$

where t_b represents the mean service time of a call agent during wrap-up stage. When handling of the call is finalized then the call agent looks for a new inbound call waiting in the queue, receive this call and starts with service process. In case there are no inbound calls waiting in the queue, the call agent starts with making outbound traffic if there is any outbound service is scheduled.

3.3.2 Process of making outbound flow by the agent

The call agents are handling outgoing flows in two different categories, calls and emails. The call gent is busy while providing services in both categories. For both of the categories of services the service time is considered to be equal and follows an exponential distribution with intensity β and the cumulative distribution function is defined as

$$F(t;\beta) = \begin{cases} 1 - e^{-\beta t} = 1 - e^{-\frac{t}{t_o}} & , \ t \geq 0, \\ 0 & , \ t < 0, \end{cases} \tag{3.26}$$

where t_o represents the mean service time for outbound service for the calls and e-mails.

The number of calls and e-mails that are made by the call agents are represented with C_{osc} and C_{ose} respectively and γ_c and γ_e represents the outbound call congestion and outbound e-mail congestion, respectively

$$\gamma_c = 1 - \frac{C_{osc}}{C_{oc}} \tag{3.27}$$

$$\gamma_e = 1 - \frac{C_{ose}}{C_{oe}} \tag{3.28}$$

where C_{oc} represents the total number of outbound calls that are generated and C_{oe} represents the total number of outbound e-mails that are generated.

3.4 Generation process of the outbound flow

There are two sub types of categorizes for each outbound flows. The first category is to make outgoing calls and the second is to make other kind of services such as email. Trunk line is busy while agent is making an outgoing call whereas while agent is working on sending emails the trunk line is not occupied. Both of the traffic flows are generated following Poisson process. The consecutive interval lengths are considered to be identical as inbound traffic arrival. When a outgoing traffic is generated it is passed into queue module. When the queue module receives the outbound flow it stores this event into the outgoing queue buffer. The outgoing traffic event is processed if only the call agent becomes in available status and there would be no incoming call waiting in the inbound queue buffer.

The number of call generations in disjoint intervals are independent and identical from each other and defined by exponential random variables with arrival intensity λ_{oc} and the cumulative distribution function is as following

$$F(t;\lambda_{oc}) = \begin{cases} 1 - e^{-\lambda_{oc} t} = 1 - e^{-\frac{t}{t_{ioc}}} & , \ t \geq 0, \\ 0 & , \ t < 0, \end{cases} \tag{3.29}$$

where t_{ioc} represents the mean interarrival time between the consecutive outbound call events.

The number of e-mail generations in disjoint intervals are independent and identical from each other and defined by exponential random variables with arrival intensity λ_{oc} and the cumulative distribution function is as following

$$F(t;\lambda_{oe}) = \begin{cases} 1 - e^{-\lambda_{oe}t} = 1 - e^{-\frac{t}{t_{ioe}}} & , \ t \geq 0, \\ 0 & , \ t < 0, \end{cases} \tag{3.30}$$

where t_{ioe} represents the mean interarrival time between the consecutive email events.

The total number of calls, C_{oc}, and total the number of e-mails, C_{oe}, that are generated are defined by

$$C_{oc} = \sum_{j=1}^{n_o} C_{ocj} \tag{3.31}$$

$$C_{oe} = \sum_{j=1}^{n_o} C_{oej} \tag{3.32}$$

Chapter 4

Numerical Results

In this chapter the case studies and the results in terms of GoS of the customer center system is analyzed. One of the most important factor that impacts the performance measure is the randomness of the call arrivals [4]. In these simulation runs the entire simulation length for one run is taken as five days dividing into disjoint intervals and each simulation is repeated multiple times with different seeds. The confidence interval with the 95% is calculated and presented on figures with error bars and on tables. The results are analyzed by using *pandas* and *numpy* libraries and the figures are plotted by using the *matplotlib* library of the *Python*.

There are various measures that is important to the domain of the GoS from both customer and system perspective [15] [23]

- The waiting time,
- Call congestion for the incoming calls,
- Outbound traffic congestion,
- The utilization of the agents,
- Abandonment rate.

4.1 Verification of the simulation model with a known model

In this section the correctness of the network model that is designed by using $OMNeT++$ is compared and shown as it is mentioned in Chapter 3. The results are computed at 95% confidence level.

The parameters are configured identically in comparison to the analytical model cited in [12]. The number of inbound call flows, n_i, and the number of call agent group types, G, is equal to 1. The number of outbound call flows, n_o, is equal to 0. The traffic intensity for inbound call flow, λ_{i1}, is equal to 0.1818 s^{-1}. The total number of trunk lines, N, is equal to 100. The service time in the IVR module, t_{ivr} and the probability of call to be served in IVR module, p_1 are equal to 100 s and 0.3, respectively. The talk time stage, t_a

and wrap up time stage,t_b for the call agent are equal to 360 s and 180 s, respectively.

$G = n_i = 1, n_o = 0, \lambda_{i1} = 0.1818\ \mathrm{s}^{-1}, N = 100, t_{ivr} = 100\ \mathrm{s},$ $S_1 = 70, p_1 = 0.3, t_{a1} = 360\ \mathrm{s}, t_{b1} = 180\ \mathrm{s}$		
Parameters	**Known Analytical Model [12]**	**Simulation Model**
B	0.01074	0.01080 ± 0.003
$E[W_o]$	58.9930 s	$59.1036\ \mathrm{s} \pm 0.674$
$P(W_o > 0)$	0.49549	0.4950 ± 0.003
$P(W > 0)$	0.50087	0.5004 ± 0.003
$E[W_{ia}]$	85.1908 s	$85.3608\ \mathrm{s} \pm 0.983$
$P(W_{ia} > 0)$	0.71553	0.7149 ± 0.004
$E[W_w]$	119.060 s	$119.2421\ \mathrm{s} \pm 0.867$
$E[W]$	59.6336 s	$59.7574\ \mathrm{s} \pm 0.695$

Table 4.1: Result comparison with a known model

Table 4.1 shows that results of simulation model corresponds with the analytical model presented in [12].

4.2 Impact of increasing inbound intensity

In this section the impact of the incoming traffic intensity λ_i on performance measure is shown with different system configurations. The arrival processes for traffic flows are configured with two different distributions where one is following Poisson (M) process and the other is with deterministic (D) distribution in two different simulations.

4.2.1 System with single call agent group processing inbound traffic only

In this subsection the performance measures of the system are analyzed with a setup of single agent group that is operating only for single type inbound call flows without abandonment and no operation is provided on the outbound direction. In this example λ_{i1} is taken as a change parameter. The parameters are considered for this simulation model are $G = n_i = 1, n_o = 0, S_1 = 19, t_a = 300\ \mathrm{s}, t_b = 180\ \mathrm{s}, t_{ivr} = 120\ \mathrm{s}, p = 0.3, N = 33.$

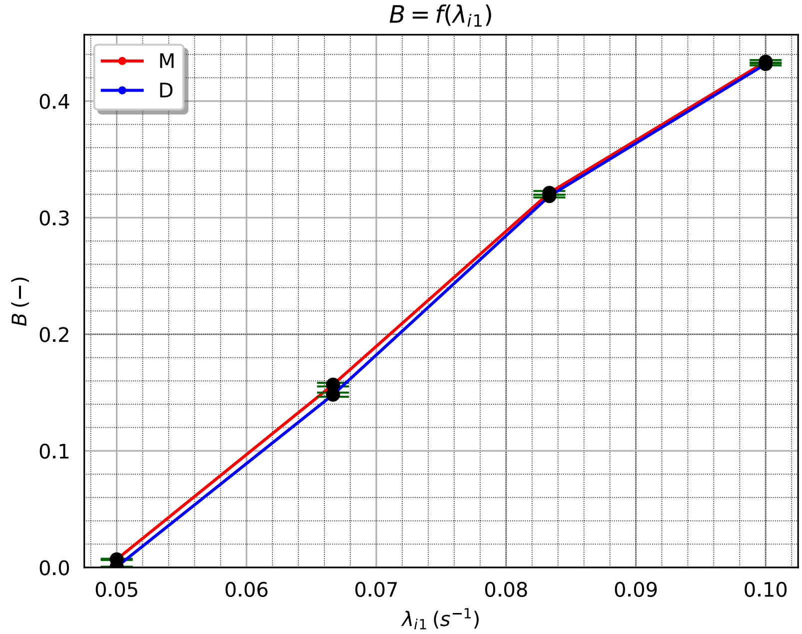

Figure 4.1: Influence $B = f(\lambda_{i1})$ with $G = n_i = 1, n_o = 0$

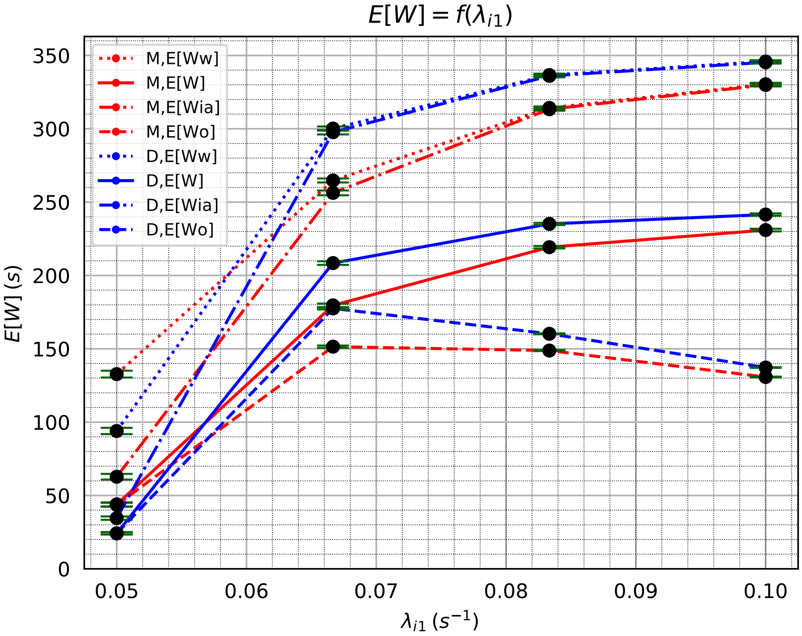

Figure 4.2: Influence $E[W] = f(\lambda_{i1})$ with $G = n_i = 1, n_o = 0$

Figure 4.1 shows the dependence of inbound call congestion to increasing traffic intensity. As the intensity of the incoming traffic is increasing the call congestion is increasing. While B is 0.0069 at the intensity of 0.05 s^{-1} for the arrival with the distribution M it becomes more than 40 percent at the intensity of 0.1 s^{-1} for both of the arrivals with the distribution M and D. This is because the resources of the system becomes fully utilized and the call congestion is increasing steadily. At a lower intensity the call congestion is observed to be lower with the distribution D than the M. As the traffic intensity is increasing the call congestion of the system with the distribution D is converging to the system with the distribution M.

Figure 4.2 shows the dependency of expected waiting times to increasing traffic intensity. While $E[W_w]$ is greater than $E[W_{ia}]$ at a lower traffic intensity, they start to become equivalent at the intensity of 0.083 s^{-1} for the system with the distribution M arrivals. This is because at low intensities the resources of the system is not fully utilized. It's possible that when some of the calls are arrived the queue will be empty and there will be an available agent. However at higher intensities in dependence to size of the system, the queue and system becomes always fully utilized and there is no possibility to get a service from agent without any waiting in the queue. It can be seen that the expected waiting times at a starting value is lower for the arrivals with the distribution D than M. However as the intensity is increasing the expected waiting times becomes higher for the arrivals with the distribution D than the arrivals with the distribution M. While $E[W_o]$ is increasing till the intensity of0.083 s^{-1} then it starts decrease. This is because the system is fully utilized but the number of calls that are offered is keep expanding which makes the expected waiting time related to all offered calls to decrease at a higher intensity.

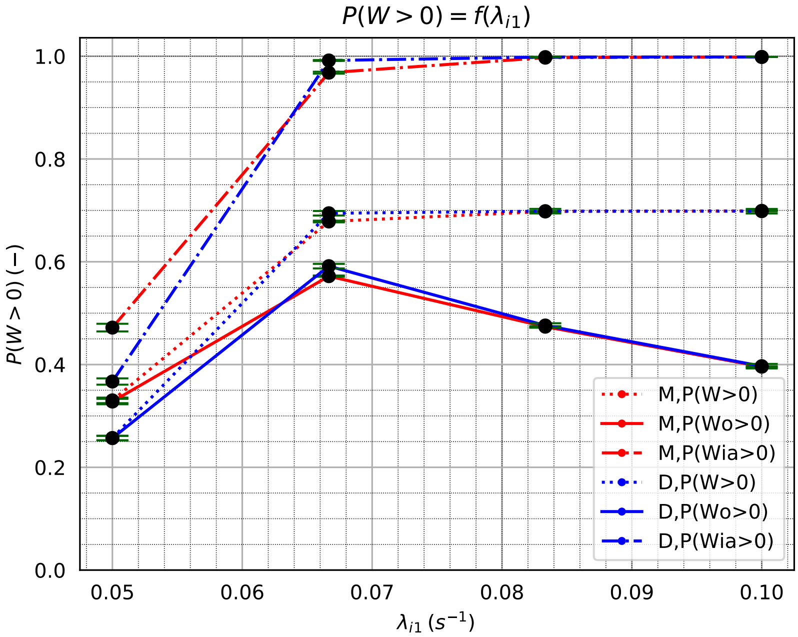

Figure 4.3: Influence $P(W > 0) = f(\lambda_{i1})$ with $G = n_i = 1, n_o = 0$

Figure 4.3 shows that the probability of waiting in the queue is converging to 1 at the intensity of 0.066 s^{-1} for the arrivals with the distribution D and at 0.083 s^{-1} for the arrivals with the distribution M and as the intensity is increasing. This is because the queue of the system becomes always full and the agents becomes fully utilized always. A similar trend can be seen for the waiting probability related to the calls passed the IVR phase. At the higher intensities $P(W_{ia})$ is shown at the value 0.68.

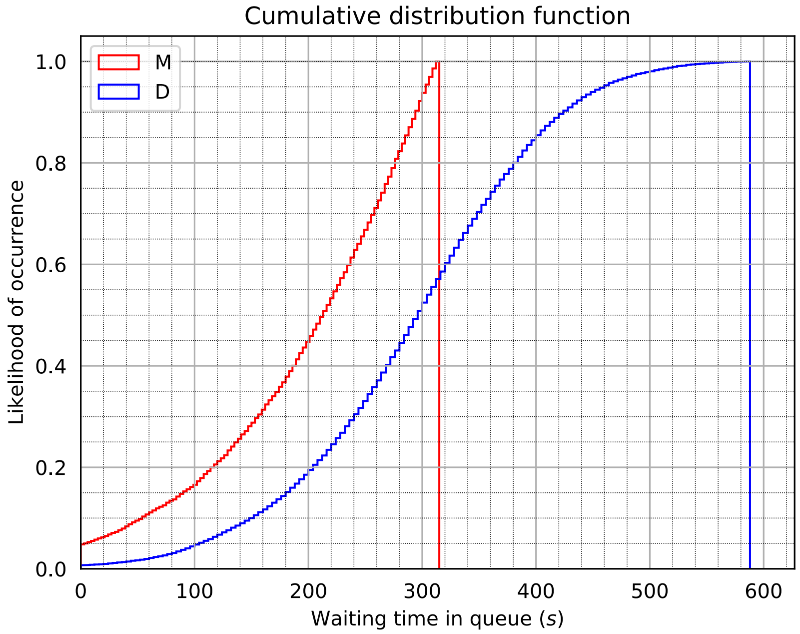

Figure 4.4: CDF of waiting time in the queue related to W_{ia} with $G = n_i = 1, n_o = 0, \lambda_{i1} = 0.06 \text{ s}^{-1}$

Figure 4.4 shows the CDF[1] related to of the waiting times in the queue for the arrivals with the distribution D and M. It shows that the scale of waiting time is wider for the arrivals with the distribution of D. It can be seen that there is a higher probability of there will be no waiting in the queue module for the arrivals with the distribution of M.

4.2.2 System with single call agent group processing inbound and outbound traffic

In this subsection the performance of the system is analyzed with the addition of an outbound flow to the system presented at the Section 4.2.1 to observe the efficiency of blending the outbound flow to a system operating only on inbound direction. There is a single call agent group operating on both inbound and outbound directions without abandonment and with the configuration of $G = n_i = n_o = 1, S_1 = 19, t_a = 300 \text{ s}, t_b = 180 \text{ s}, t_o = 150 \text{ s}, t_{ivr} = 120 \text{ s}, p =$

---

[1] The vertical line is drawn by default by the *matplotlib*.

$0.3, N = 33, \lambda_{oc1} = \lambda_{oe1} = 0.0292\ \mathrm{s}^{-1}$. The incoming call intensity λ_{i1} is taken as a change parameter and the results are observed with a system inbound and outbound arrivals with the distribution M and D distributions.

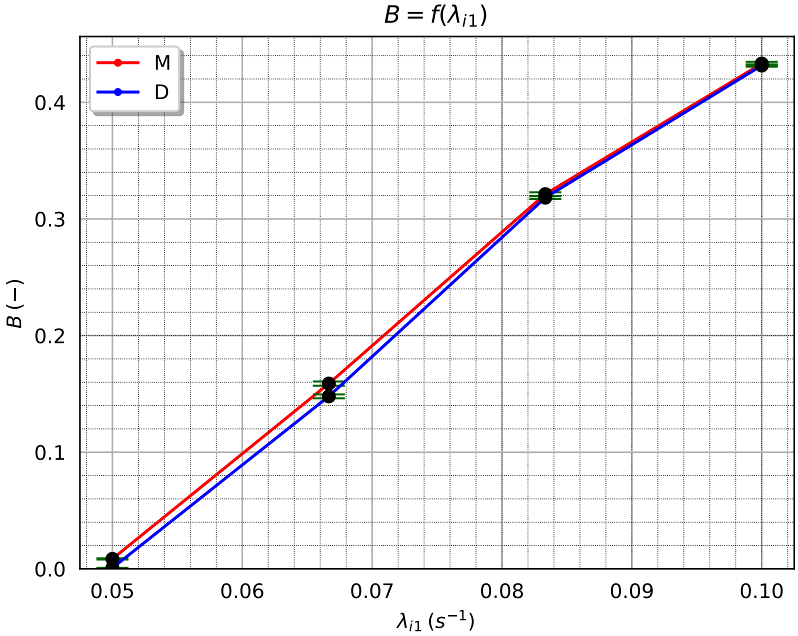

Figure 4.5: Influence $B = f(\lambda_{i1})$ with $G = n_i = n_o = 1$

Figure 4.5 shows the dependency of inbound call congestion to increasing inbound traffic intensity. The starting value of B with the distribution of M is at the value of 0.0085 which is greater than the observed values at the 4.2.1 since the addition of an outbound traffic flow to the system. When the call agents are in idle status with the absence of incoming calls the calls agents are going to start with serving outbound flows which will make the agents in busy status and if an incoming is going to arrive at the moment the agent is handling an outbound flow this is going to make the incoming call to be rejected and this is going to result in increase in the call congestion related to incoming calls.

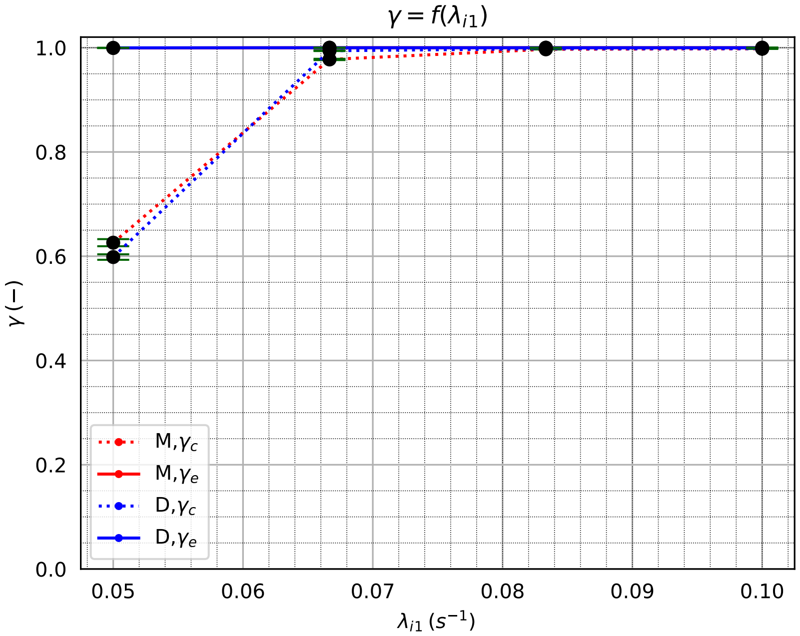

Figure 4.6: Influence $(\gamma_c; \gamma_e) = f(\lambda_{i1})$ with $G = n_i = n_o = 1$

Figure 4.6 shows the dependency of outbound call and e-mail congestion related with the increasing inbound traffic intensity. In the system with arrivals of M, the outbound e-mail congestion γ_e is at 0.99 from low intensity to high intensity values. The outbound call congestion γ_c is 0.625 at the low level of input intensity λ_{i1}) and starts to increase and converges to 1 for the arrivals with the distribution M. This is because since the intensity of incoming call flows is low at the start the call agents are not fully utilized with the incoming calls and almost every time they try to make an outbound service they are able to find a free trunk line to. However as the traffic intensity is increasing the buffer of the incoming queue is always occupied and the outbound call congestion is converging to 1 at a high incoming traffic intensity. There is a same trend is seen related to the outbound call congestion with the arrivals of D distribution in comparison to M distribution where the value is at 0.99. Outbound call congestion related to the D distribution starts at a lower value 0.59 but converging to 0.99 faster than the distribution with M as the intensity increases. The outbound e-mail congestion starts at 1 since the call agents are able to find a free trunk line due low inbound traffic intensity.

While call agents can perform outbound call flow at a lower incoming traffic intensity, handling outbound traffic becomes inefficient at a higher intensity since the outbound traffic congestion converges to 0.99 level based on size of the system.

Figure 4.7 shows the dependency of expected waiting times with increasing traffic intensity. Considering the system that arrivals following the distribution

M, $E[W_{ia}]$ is 30.6 percent higher than the system from Section 4.2.1 but the $E[W_w]$ is 6.2 percent lower. Since the incoming traffic intensity is low and the call agent are making outbound traffic in the absence of incoming call in the queue, the number of calls that are admitted into the system are decreased. However since the number of agents serving for incoming traffic flow and the number of incoming calls are decreased proportionally the expected waiting time in the queue is not changed dramatically. The expected waiting times related to arrival with the distribution D is increased at the low level of input intensity λ_{i1}) in comparison to the system from Section 4.2.1.

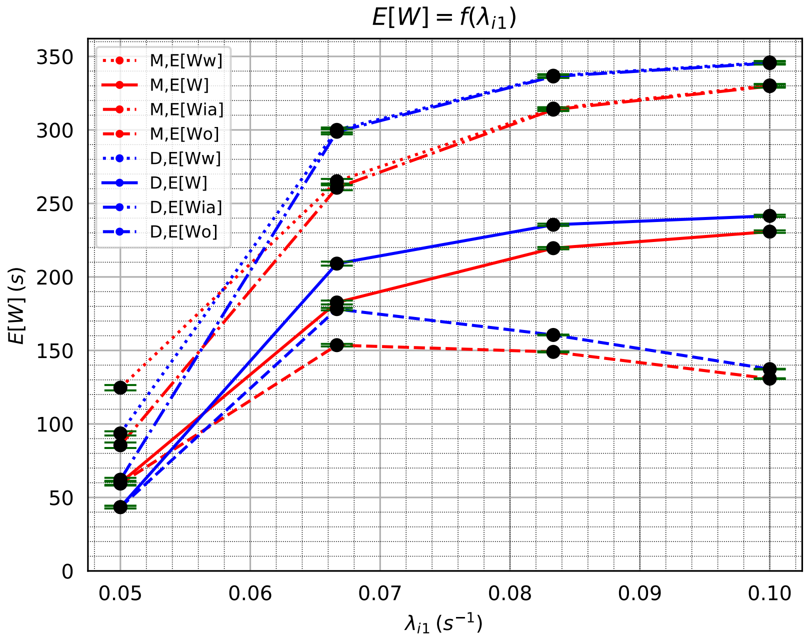

Figure 4.7: Influence $E[W] = f(\lambda_{i1})$ with $G = n_i = n_o = 1$

Figure 4.8 shows the CDF of the waiting times in the queue. In the system with only inbound operation as shown in Figure 4.4 the system with the distribution M has a lower scale. In contrast in this system the scale is larger. While the scale of waiting times larger in the system at Section 4.2.1, in this system an opposite distribution is seen.

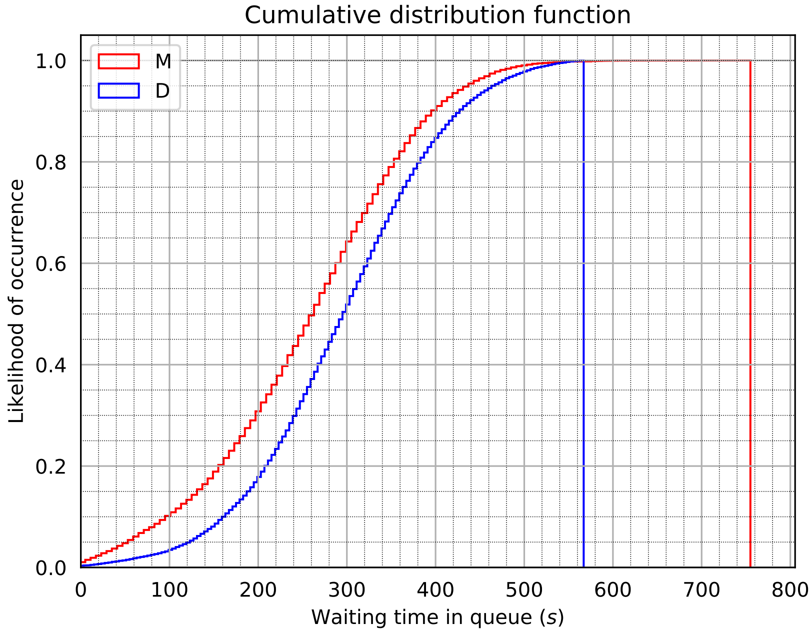

Figure 4.8: CDF of waiting time in the queue related to W_{ia} with $G = n_i = n_o = 1,\lambda_{i1} = 0.06\ \mathrm{s}^{-1}$

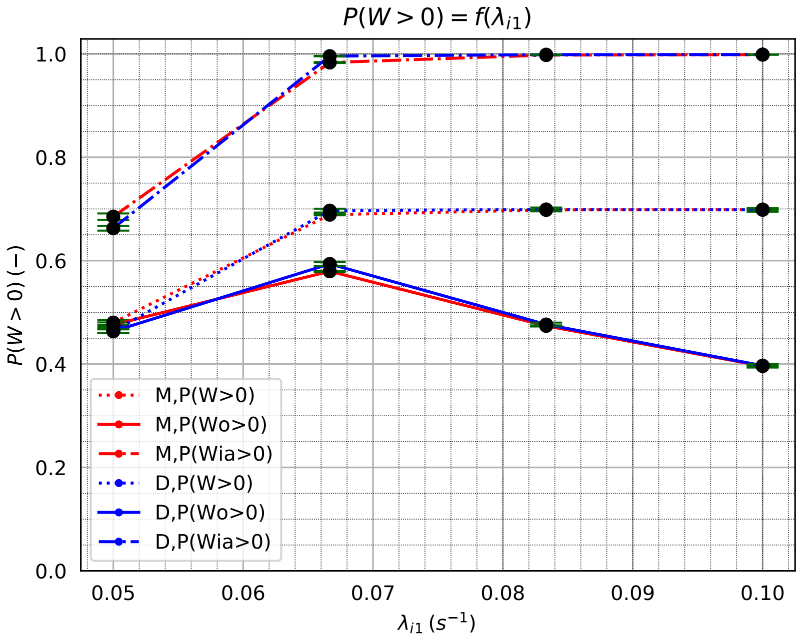

Figure 4.9: Influence $P(W > 0s) = f(\lambda_{i1})$ with $G = n_i = n_o = 1$

Figure 4.9 shows the dependency of probability of waiting in the queue. The admitted calls are converged to 1 faster than the system represented in

Section 4.2.1. This is because the probability of finding an available agent is decreased since the call agents started to work on outbound direction while waiting for a incoming call to arrive.

4.2.3 System with multiple call agent groups processing inbound and outbound traffic

In this section the performance measure is analyzed with a system of multi skilled call agent groups operating on inbound and outbound direction. There are three call agent groups operating on both inbound and outbound directions without abandonment and with the configuration of $G = n_i = n_o = 3, S_1 = S_2 = S_3 = 28, t_{a1} = t_{a2} = t_{a3} = 270$ s$, t_{b1} = t_{b2} = t_{b3} = 180$ s$, t_{o1} = t_{o2} = t_{o3} = 180$ s$, t_{ivr} = 90$ s$, p_1 = p_2 = p_3 = 0.3, N = 117, \lambda_{oc1} = \lambda_{oc2} = \lambda_{oc2} = \lambda_{oe1} = \lambda_{oe2} = \lambda_{oe3} = 0.038$ s$^{-1}, \lambda_{i2} = \lambda_{i3} = 0.06$ s^{-1}. The incoming call intensity λ_{i1} is taken as a change parameter. The simulation is executed with arrivals following Poisson and deterministic processes.

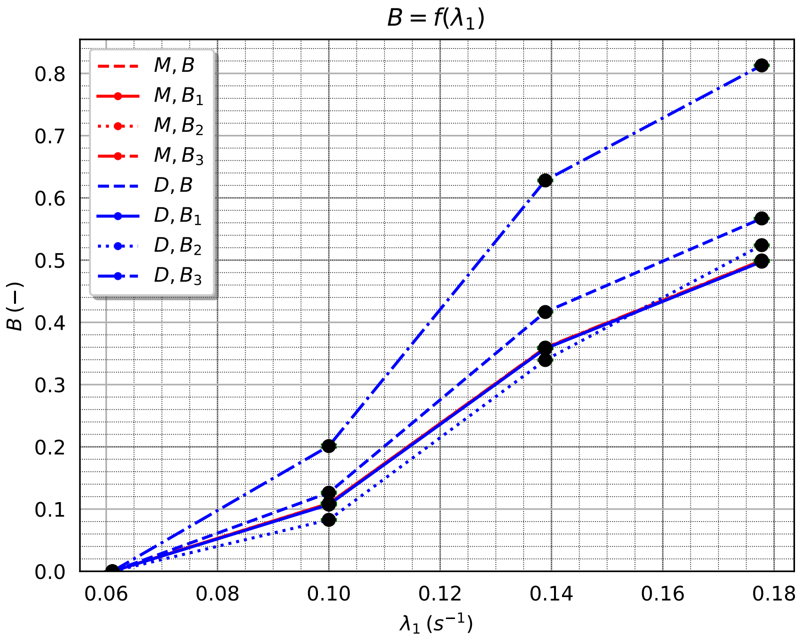

Figure 4.10: Influence $B = f(\lambda_{i1})$ with $G = n_i = n_o = 3$

Figure 4.10 shows the dependency of incoming call congestion with the increasing λ_{i1}. Considering the arrivals with the M process, the call congestion is 0 at the lowest intensity and it increases to 0.49 at the intensity of 0.177 s^{-1}. It can be seen that the call congestion is almost equivalent for all the types of call flows. The reason is that even though the traffic intensity is increasing only for the type 1 flows the system is prevailed by the type 1 call flow and the rate of calls that are served and dropped stays proportional to each other. It can be seen in Figure 4.12 that the system is dominated by the type 1 calls since while the outbound call congestion is heavily decreasing the e-mail congestion is decreasing for type 2 and type 3 flows. Considering the system with deterministic arrival process the call congestion shows the same characteristic with the system with stochastic arrival process. It can be seen that all the different call types have different call congestion while type 1 call types have the equivalent call congestion with the distribution M and D.

Figure 4.11 shows that at the starting traffic intensity for the system with stochastic arrival process, $E[W_w]$ is 44.2 s,43.7 s and 44.15 s for type 1, type 2 and type 3 flows, respectively. Since the traffic intensity is increasing, the system is heavily utilized by the type 1 call flows as it is mentioned. That's why the queue is stored by the calls from type 1 and the expected waiting time is increasing heavily for the type 1 traffic flows with the increasing traffic intensity. It can be seen that as the intensity is increasing the waiting times related to type 2 and type 3 flows is converging to $0s$. This is because since the intensity of the traffic is low for these types the incoming calls are starting to being served without waiting in the queue. The waiting time is expanding greater at system with deterministic arrival process.

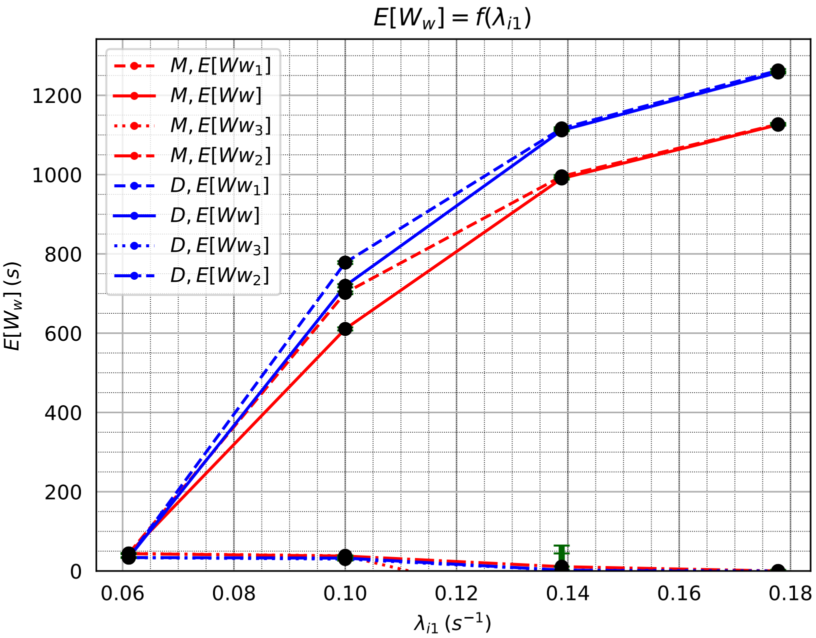

Figure 4.11: Influence $E[Ww_w] = f(\lambda_{i1})$ with $G = n_i = n_o = 3$

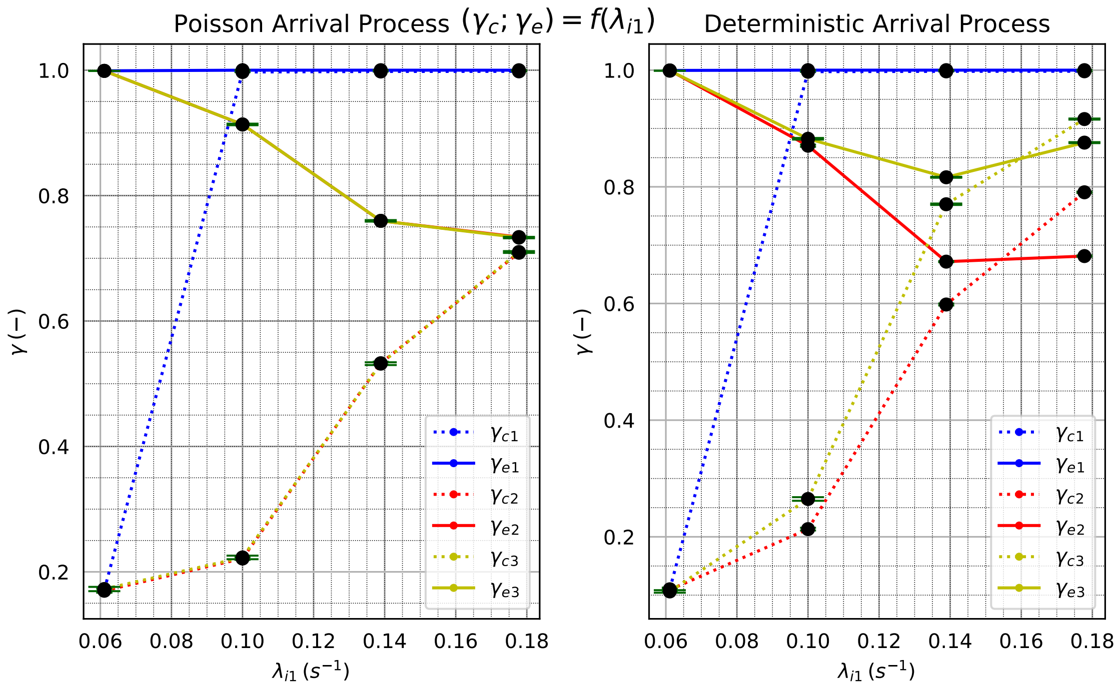

Figure 4.12: Influence $(\gamma_c; \gamma_e) = f(\lambda_{i1})$ with $G = n_i = n_o = 3$

Figure 4.12 shows the dependency of outbound traffic congestion with increasing incoming traffic intensity. With respect to the system following Poisson arrival process, the outbound call congestion for type 1 flow is 0.17 at the beginning intensity of $0.06s$ s^{-1} and as the incoming intensity increases the outbound call congestion rise up and converges to 1. This is because the system is loaded with the incoming type 1 call flow and the outbound call is not served when the system is loaded with incoming traffic. The outbound e-mail congestion stays at 1 since arrivals with lowest incoming intensity which means the call agents are able to find a free trunk line every time they become in idle status and there is no incoming call buffered in the queue since the traffic intensity is low but once the input intensity λ_{i1} starts to increase agents from g_1 are not able to serve any outbound service. However since the incoming intensity is constant for the type 2 and type 3 flows, the outbound congestion increases for the call flows and decreases for the e-mail flows. This is becasue the possibility of finding a free trunk line decreases but since the system is dominated by the input intensity λ_{i1} the number inbound calls from type 2 and type 3 decrease and since the agents from groups g_2 and g_3 are in available status they start with handling e-mail flow. g_2 and g_3 agent groups are still able to maintain making outbound service while g_1 is only handling inbound call flow.

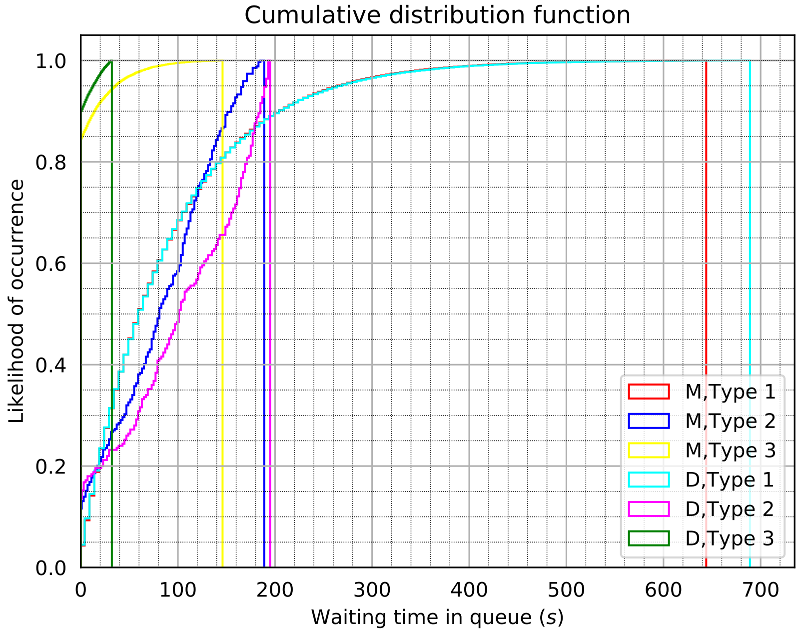

Figure 4.13: CDF of waiting time in the queue related to W_{ia} with $G = n_i = n_o = 3$,$\lambda_{i1} = 0.1$ s^{-1}

Figure 4.13 shows the CDF of waiting time in the queue at $\lambda_{i1} = 0.1$ s^{-1}. From the scale the distribution shows that the probability of call arrivals from type 3 flow are going to have $0s$ waiting in the queue with a high probability. It can be seen that the scale of waiting time is larger for the type 1 call flows since the input intensity λ_{i1} is increasing and more calls are waiting in the queue of type 1 calls.

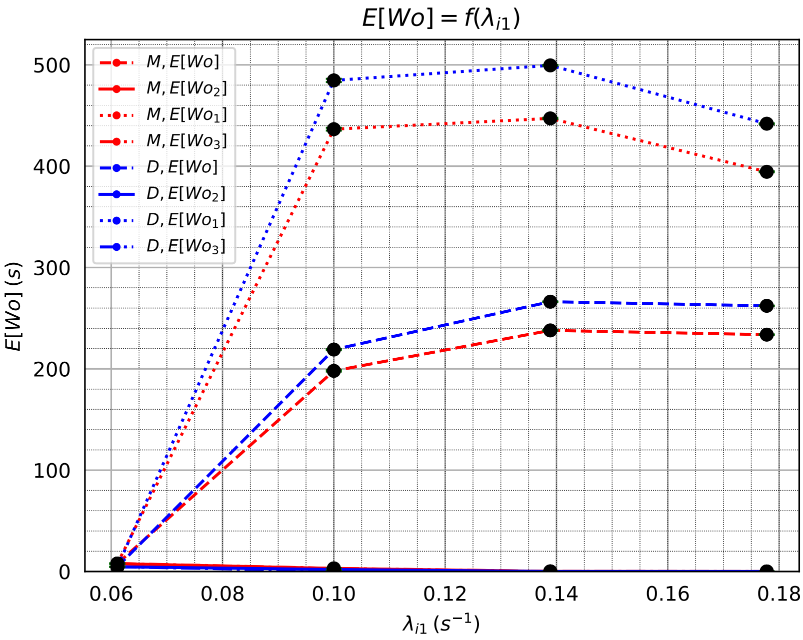

Figure 4.14: Influence $E[Wo] = f(\lambda_{i1})$ with $G = n_i = n_o = 3$

Figure 4.14 shows that the expected waiting time converges to $0s$ for the type 2 and type 3 flows since there is going to be no waiting in the queue. The expected waiting time related to offered calls is higher at the system with deterministic arrival processes than the system with Poisson process.

The probability of waiting in the queue related to admitted calls is 0.177, 0.177 and 0.178 for the flow types 1,2 and 3, respectively at the low input intensity λ_{i1} for the system with distribution M as it is shown in Figure 4.15. Since the waiting time in the queue related to type 1 flows is increasing, the probability of waiting related to admitted calls is increasing to 0.7 with increasing intensity λ_{i1}. Since the agents are fully utilized, it is converging to 0 for the other type of call flows as the intensity of type 1 flow is increasing.

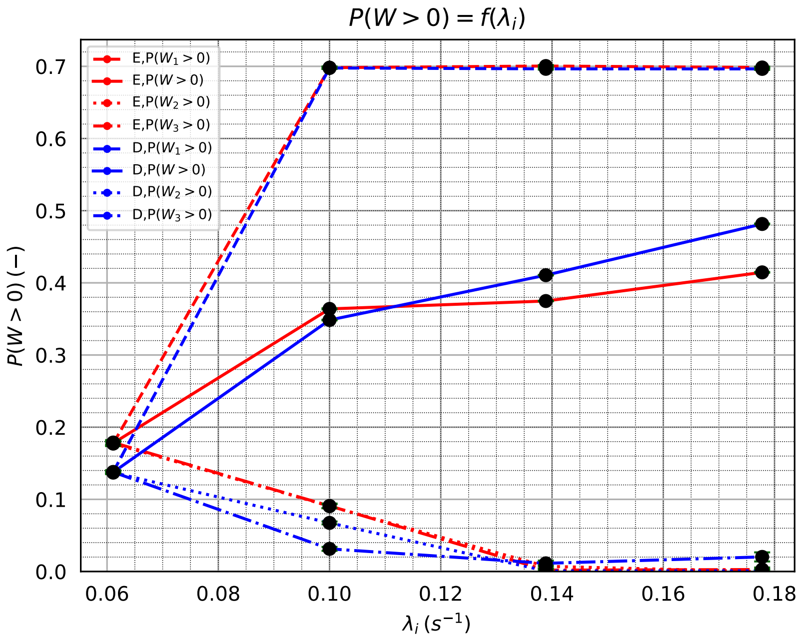

Figure 4.15: Influence $P(W > 0) = f(\lambda_{i1})$ with $G = n_i = n_o = 3$

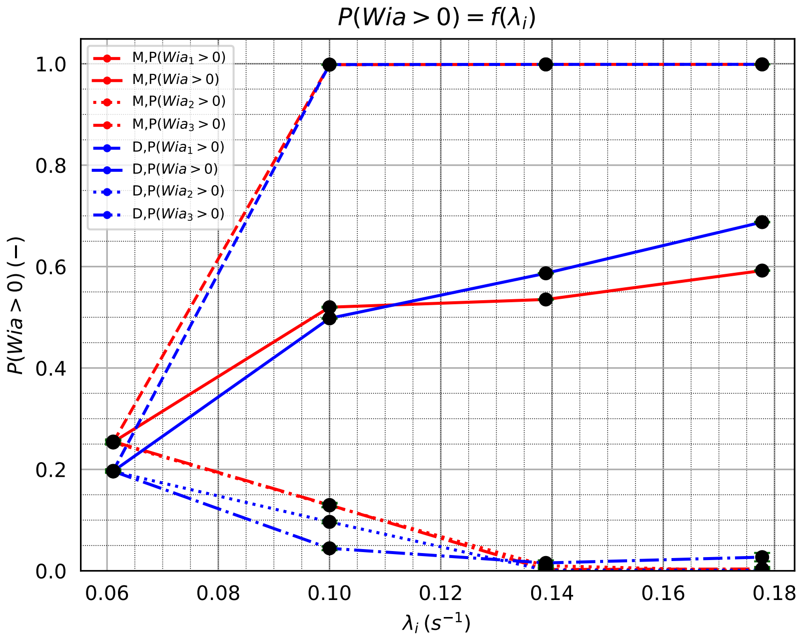

Figure 4.16: Influence $P(W_{ia} > 0) = f(\lambda_{i1})$ with $G = n_i = n_o = 3$

Figure 4.16 shows that while the probability of waiting after completing the service in IVR converges to 1 for the type 1 call flows, it is converging to 0 for the type 2 and type 3 call flows as the incoming traffic intensity increases.

Since the large proportion of the system is loaded with the calls from type 1 flow, there is going to be a always waiting in the queue.

4.3 Impact of increasing number of trunk lines N

In this section the impact of N on the performance measures is analyzed. The system is configured with three multi skilled agent groups operating on both inbound and outbound directions and without abandonment and N is chosen as a change variable. The configuration of the parameters are considered as $G = n_i = n_o = 3, S_1 = S_2 = S_3 = 15, t_{a1} = t_{a2} = t_{a3} = 240 \text{ s}, t_{b1} = t_{b2} = t_{b3} = 150 \text{ s}, t_{o1} = t_{o2} = t_{o3} = 180 \text{ s}, t_{ivr} = 105 \text{ s}, p_1 = p_2 = p_3 = 0.3, \lambda_{oc1} = \lambda_{oc2} = \lambda_{oc2} = \lambda_{oe1} = \lambda_{oe2} = \lambda_{oe3} = 0.037 \text{ s}^{-1}, \lambda_{i1} = \lambda_{i2} = \lambda_{i3} = 0.058 \text{ s}^{-1}$.

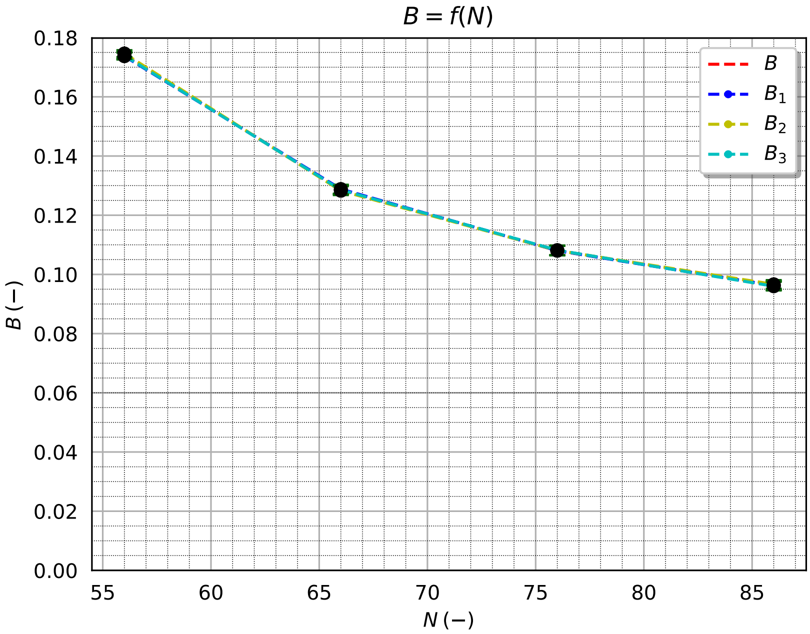

Figure 4.17: Influence $B = f(N)$ with $G = n_i = n_o = 3$

Figure 4.17 shows that the values of the call congestion are changing very closely to each other for the different type flows. While the starting value for B is 0.17 it drops to 0.09 where there is a 61.5 percent difference when the N is increased by 40.

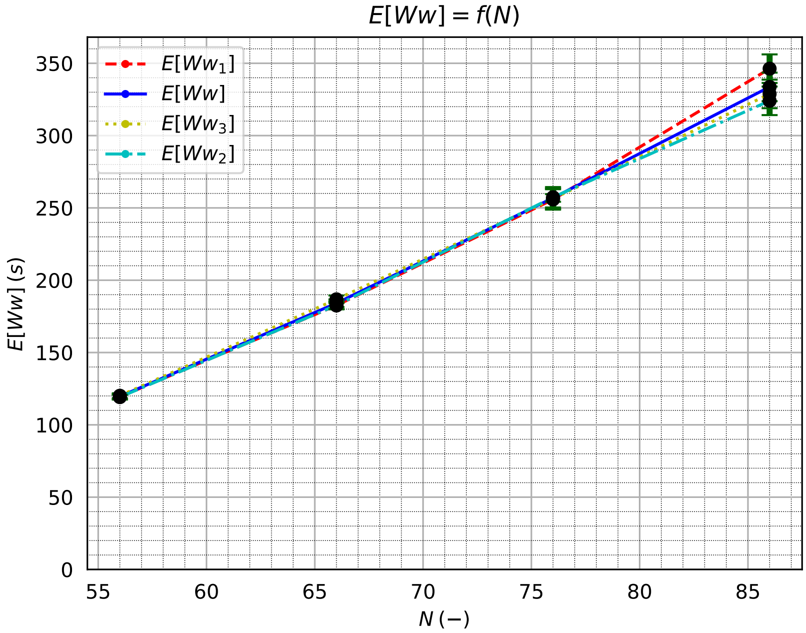

Figure 4.18: Influence $E[W] = f(N)$ with $G = n_i = n_o = 3$

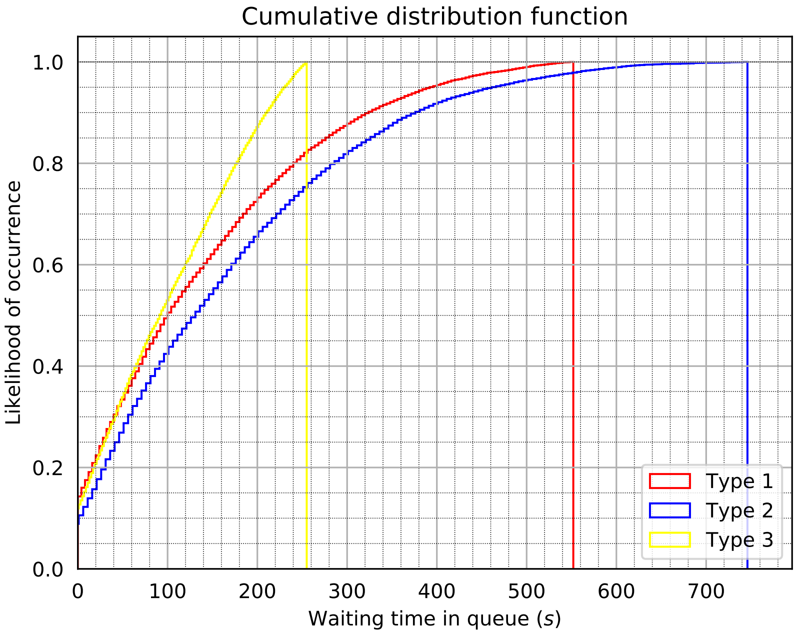

Figure 4.19: CDF of waiting time in the queue related to W_{ia} with $G = n_i = n_o = 3$,$N = 66$

The dependency of expected waiting times $E[W_w]$ and $E[W_{ia}]$ to increasing N is shown in Figure 4.18. Again, just like the change characteristic of B, the expected waiting times from different types are changing very closely to each other. This is because the traffic intensity and the staffing is equal for all the types. It can be seen that as the N is increasing, the expected waiting times are increasing steadily since the size of the system and the number of calls waiting in the queue increases in parallel but the resources related to agents stays constant. $E[W_w]$ is increasing 94.4 percent from 119.6 s to 333.6 s when the N is increased by 40. This shows that while GoS measurement is improving related to B, the $E[W]$ is deteriorating. While the waiting times are very close to each other at the low N, as the N is increasing the waiting times are becomes widening from each other. The CDF related to waiting times for different call types is depicted in Figure 4.19.

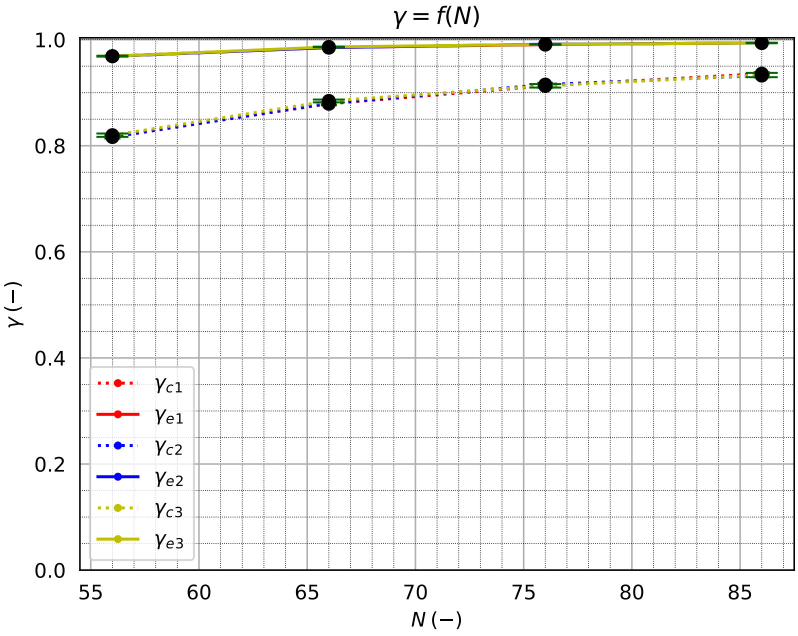

Figure 4.20: Influence $(\gamma_c; \gamma_e) = f(N)$ with $G = n_i = n_o = 3$

Figure 4.20 shows the dependency of outbound call congestion to increasing N. γ_c is going up to 0.94 from 0.81 and γ_e start at the value 0.94 and converges to 0.99 for all the types of call flows. The reason of call congestion is increasing when the N is increased is that the maximum capacity of the queue increases for the inbound traffic. There are more calls are stored in the incoming queue. The call agents are checking the incoming queue before making an outbound flow, possibility of finding the queue empty decreases since the number of calls admitted into system is increased and call agents are heavily utilized with inbound flows.

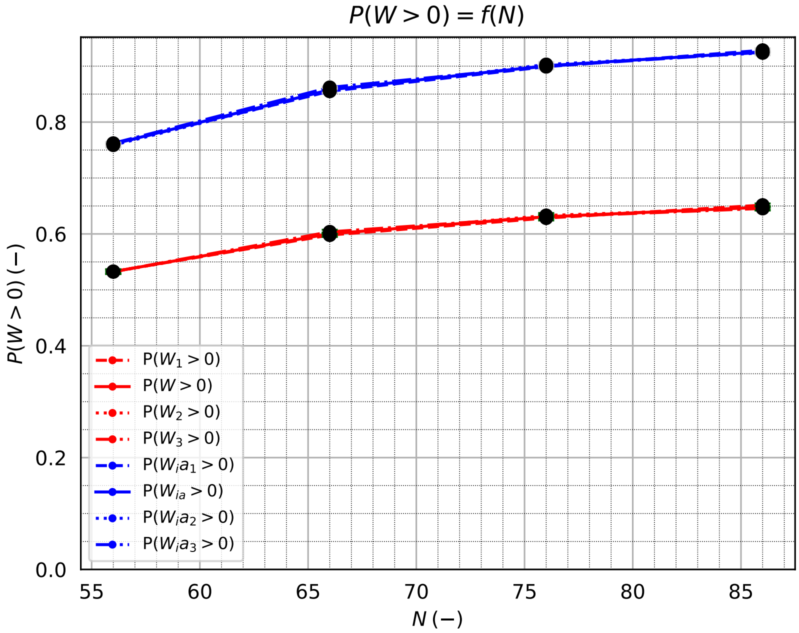

Figure 4.21: Influence $P(W > 0) = f(N)$ with $G = n_i = n_o = 3$

Figure 4.21 shows that the probability of waiting in the queue related to offered and admitted calls are increasing as the N is increasing since the queue size is increased with an expanding system size and number of calls that are served without waiting is decreased.

4.4 Impact of increasing number of agents in fist group S_1

In this section the impact of staffing level on performance measures is analyzed with a system of multi skilled call agent groups operating on inbound and outbound direction. The system is arranged with a configuration of $G = n_i = n_o = 3, S_2 = S_3 = 12, t_{a1} = t_{a2} = t_{a3} = 300$ s$, t_{b1} = t_{b2} = t_{b3} = 180$ s$, t_{o1} = t_{o2} = t_{o3} = 180$ s$, t_{ivr} = 105$ s$, p_1 = p_2 = p_3 = 0.3, N = 53, \lambda_{oc1} = \lambda_{oc2} = \lambda_{oc2} = \lambda_{oe1} = \lambda_{oe2} = \lambda_{oe3} = 0.0125$ s$^{-1}, \lambda_{i1} = \lambda_{i2} = \lambda_{i3} = 0.033$ s^{-1}. The number of agents from agent group 1, S_1, is taken as a change parameter.

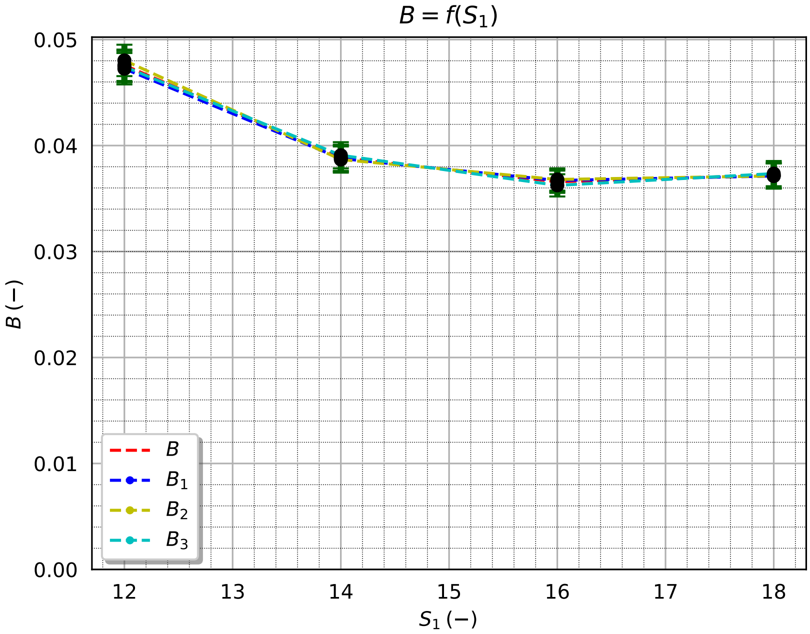

Figure 4.22: Influence $B = f(S_1)$ with $G = n_i = n_o = 3$

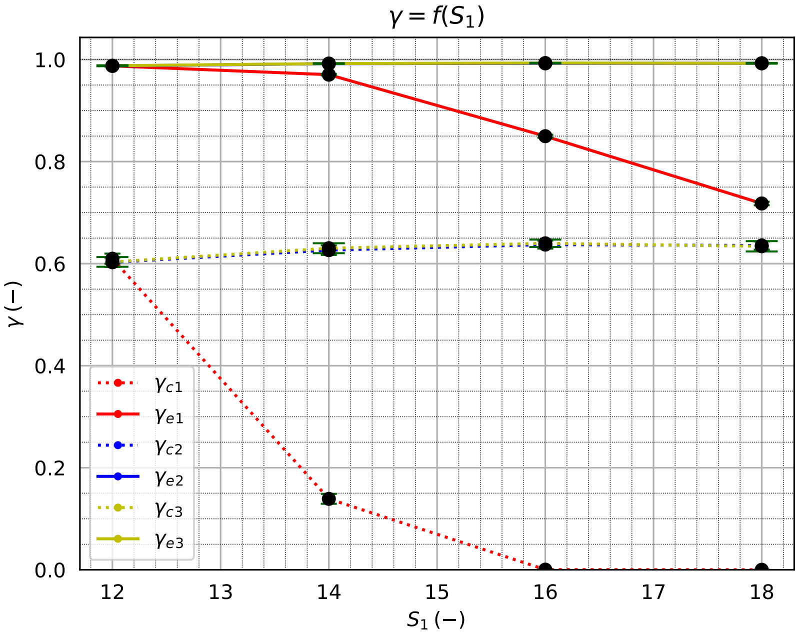

Figure 4.23: Influence $(\gamma_c; \gamma_e) = f(S_1)$ with $G = n_i = n_o = 3$

Figure 4.22 shows that the inbound call congestion is decreasing for flows from all the types and then increases. From the starting point where S_1 equals to 12 for the type 1 flow, B is decreasing by 24.6 percent when S_1 equals to

16. By the end point it increases by 0.8 percent when the S_1 changes from 16 to 18. It is important to consider that the intensity and the number of trunk lines are constant. Also, there is an important correlation between the inbound call congestion and outbound call congestion where it is depicted in Figure 4.23.

At the start since the incoming traffic intensity is not high for the system the call agents from all the types are also handling outbound flows and if there is an inbound call, it is going to be dropped if all the agents are utilized. When the S_1 increases, there is a higher probability that a call agent will be free and it can be seen that the B_1,γ_{c1} and γ_{e1} decreases for the type 1 flows which means that the agents are allocated by a larger volume of incoming traffic. As the number of agents from type 1 increases since there is no expand on N, the call agents from type 1 starts to handle e-mail flows which makes them busy and increase the B to some degree since there might be no free call agent upon incoming call arrival and all the trunk lines are allocated. ,γ_{c2} and γ_{c3} stays in the same level.

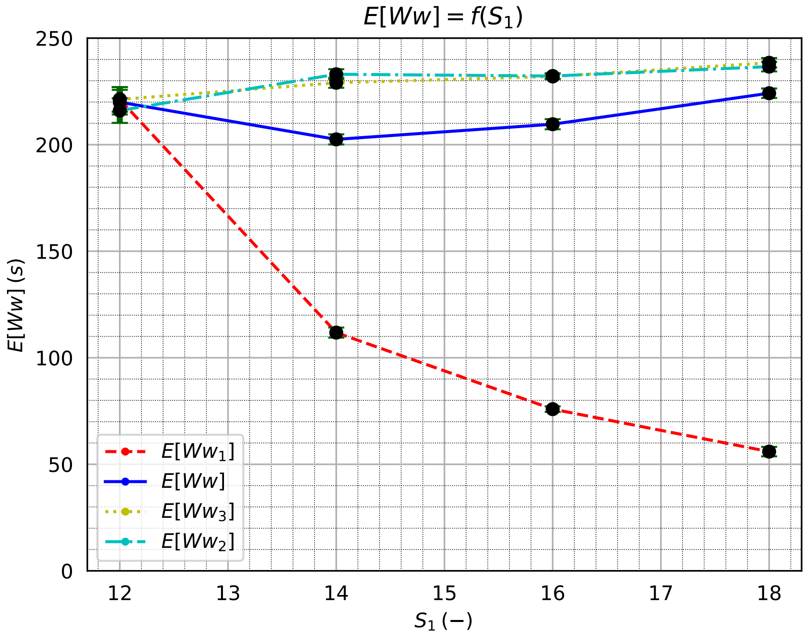

Figure 4.24: Influence $E[Ww] = f(S_1)$ with $G = n_i = n_o = 3$

Figure 4.24 shows that there is a 119.2 percent decrease in the $E[Ww_1]$ from start point to end point. Since the S_1 is increasing and the N is constant, it is expected that waiting time is going to decrease since the probability of finding a free agent increases. It can be seen that $E[Ww_3]$ and $E[Ww_2]$ is fluctuating. This is because the queue is moving for type 1 flows in a faster trend than the other flow types and this makes the system not having to split the number of calls to split proportionally from different flows. Figure 4.25 shows the CDF related to the waiting times in the queue and it can be seen that the probability of not waiting in the queue is largest for type 1 flows.

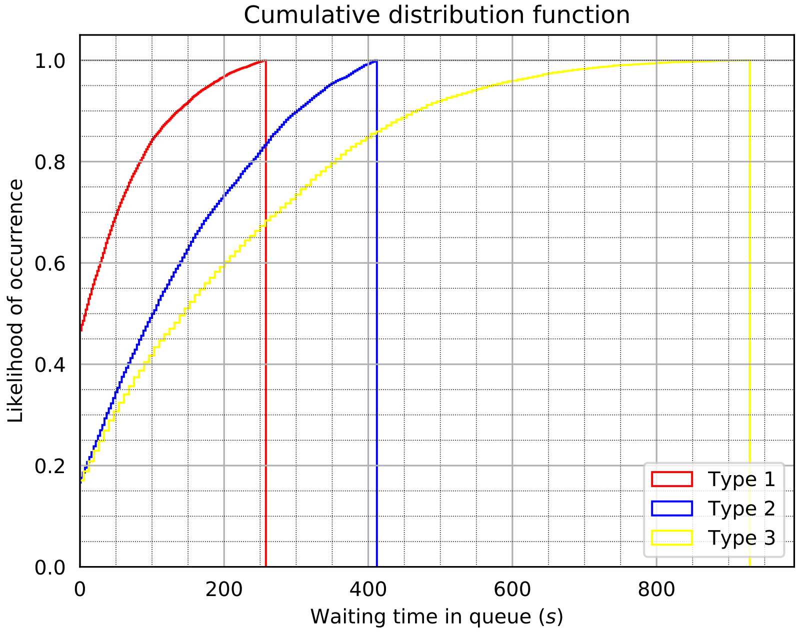

Figure 4.25: CDF of waiting time in the queue related to W_{ia} with $G = n_i = n_o = 3$,$S_1 = 14$

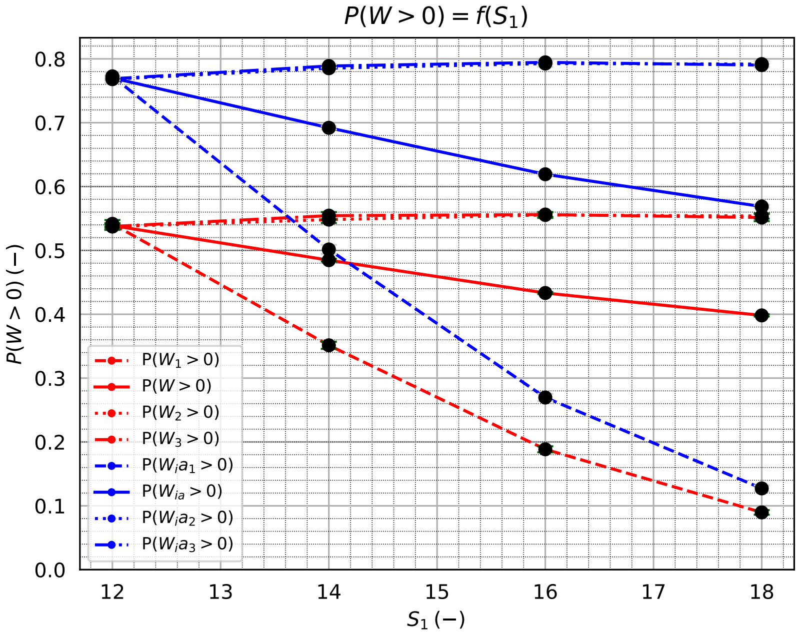

Figure 4.26: Influence $P(W > 0) = f(S_1)$ with $G = n_i = n_o = 3$

The probability of waiting in the queue is decreasing heavily for type 1 flows while it is slightly changed for the flow types 2 and 3 as it is depicted in Figure 4.26.

Chapter 5

Conclusion

The work in this diploma thesis is denoted to create a simulation model of a customer center and analyzing the important characteristics in terms of GoS.

In the first part, the general concept of a customer center is explained. The important terms that needs to be considered when designing the simulation model such as skill based routing, interactive voice response and automatic call distributor are explained.

In the Chapter 3, the core and the working mechanism of the simulation model which is created by using $OMNeT++$ are defined. The computation of the result parameters which is required for the case studies are defined.

In the Chapter 4, the case studies are analyzed with different simulation scenarios. All of the simulation case studies are configured with only one changing parameter and all the other parameters are constant in all of the intervals and over the simulation runs. First, to prove the correctness of the simulation model, a comparison is made with a known model which comprise from a single agent group and a single flow with inbound traffic only. It is observed that the results are corresponding to the analytical model.

Second, the impact of the increasing inbound traffic intensity is analyzed by using a single agent group and a single flow type with inbound traffic only. Two different simulation configurations are made by using the exponential and deterministic distribution of interarrival times of the input flow. It is observed that there is a great influence of the increasing traffic intensity on call congestion and expected waiting time. The system becomes fully loaded and it is not convenient to keep a reasonable service levels at high intensities. It is also observed that the system with the deterministic distribution of interarrival times of the input flow is resulting on lower call congestion and higher waiting times than the system with the exponential distribution of interarrival times of the input flow. By adding an outbound flow into the same system in the next scenario, it is observed that the operation of outbound flow is only possible at low input inbound intensity. The disadvantage is that probability of waiting in the queue increases but the mean waiting time is only changed slightly. At the high inbound traffic intensity, it is not convenient for the system to provide outbound services. In the next scenario the system is configured with multi skilled agent groups and multi flow types. Increasing the traffic intensity of only one flow type results that the load of the system is dominated by the inbound flows with the increasing intensity. It also results in a drastically increased waiting time for this type and drastically decreased waiting times for the other types of flows. It is noticeable that the inbound traffic congestion is increasing proportionally for all of the flow types. While there is no outbound operation for the flow type with increased traffic intensity, the outbound e-mail congestion is decreasing for the other types and the outbound call congestion is increasing. This is not very convenient for the system because the number of inbound call requests that receive the busy signal is increasing from the types with constant intensity. To avoid this kind of situation dedicated trunk lines related to each flow type could be configured so that the number of customers that are blocked from different types would stay in a consistent level.

In the next case study, it is observed that increasing the number of trunk lines in a multi skilled system results in a decreased inbound call congestion but in an increased expected waiting time. However, the outbound traffic congestion is increasing and comes to a degree that there is almost no operation on outbound direction. To increase the efficiency of the outbound traffic the solution requires increased number of call agents in parallel to increased number of trunk lines or to allocate a call agent group that is only working for the outbound flows.

In the last case study,the number of call agents for one of the agent group in a multi skilled system is increasing. It results in a decreased inbound call congestion but only until some point because of the limited system capacity and the change in call congestion is proportional for each flow types. While the mean waiting time in the queue is decreasing drastically for the flow type with increased number of agents, there is no heavy change for the other flow types. One of the important change is that the agent group with increased number of staff is starting to operate outbound call traffic at a maximum rate while it is consistent for the other agent groups. Increasing the staffing level shows a higher improvement in comparison to increment in trunk lines.

The results shows that the operation of outbound services can't be maintained when there is high input inbound intensity. However at lower inbound intensities, the operation of the outbound services can be processed to some degree. There can be dedicated call agent groups created for inbound and outbound directions but this depends on the importance of making an outbound flow for the operations. In this work to get the best effort from the call agent groups while considering the inbound flow as priority operation and making outbound services when it is possible is accomplished by the design of a model with multi skilled call agent groups and categorized inbound and outbound directional traffic flows. Moreover, one of the most important outcome observed by using the simulation model is the simplicity of switching between distribution models for the arrival and service times.

Bibliography

[1] J. Atlason, M. A. Epelman, and S. G. Henderson. Optimizing call center staffing using simulation and analytic center cutting-plane methods. *Management Science*, 54(2):295–309, 2008.

[2] A. N. Avramidis and P. L'Ecuyer. Modeling and simulation of call centers. In *Proceedings of the Winter Simulation Conference, 2005.*, pages 9–pp. IEEE, 2005. ISBN: 0-7803-9519-0.

[3] J. K. Blitzstein and J. Hwang. *Introduction to probability.* Chapman and Hall/CRC, 2014. ISBN: 978-0716771098.

[4] A. Brezavšček and A. Baggia. Optimization of a call centre performance using the stochastic queueing models. *Business Systems Research Journal*, 5(3):6–18, 2014.

[5] L. Brown, N. Gans, A. Mandelbaum, A. Sakov, H. Shen, S. Zeltyn, and L. Zhao. Statistical analysis of a telephone call center: A queueing-science perspective. *Journal of the American statistical association*, 100(469):36–50, 2005.

[6] W. Chan, G. Koole, and P. L'Ecuyer. Dynamic call center routing policies using call waiting and agent idle times. *Manufacturing & Service Operations Management*, 16(4):544–560, 2014.

[7] W. C. Chan. The erlang loss and delay systems. *Performance Analysis of Telecommunications and Local Area Networks*, pages 224–229, 2002. ISBN: 978-0-7923-7701-6.

[8] L. A. G. Franzese, M. M. Fioroni, R. C. Botter, and P. J. de Freitas Filho. Comparison of call center models. In *Winter Simulation Conference*, pages 2963–2970. Winter Simulation Conference, 2009. ISBN: 978-1-4244-5771-7.

[9] N. Gans, G. Koole, and A. Mandelbaum. Telephone call centers: Tutorial, review, and research prospects. *Manufacturing & Service Operations Management*, 5(2):41–49, 2003.

[10] D. Gross. *Fundamentals of queueing theory.* John Wiley & Sons, 2008. ISBN: 978-0471791270.

[11] K. Gustavsson. *Stochastic Modeling and Management of an Emergency Call Center: A Case Study at the Swedish Emergency CallCenter Provider, SOS Alarm Sverige AB.* PhD thesis, Mid Sweden University, 2018.

[12] P. Hampl. Analytical call center model with voice response unit and wrap-up time. *Advances in Electrical and Electronic Engineering*, 13(4):273–279, 2015.

[13] V. B. Iversen. Teletraffic engineering and network planning. pages 232–240, 2015.

[14] P. Khudyakov, P. D. Feigin, and A. Mandelbaum. Designing a call center with an ivr (interactive voice response). *Queueing Systems*, 66(3):215–237, 2010. ISSN: 1572-9443.

[15] R. Klungle and J. Maluchnik. The role of simulation in call center management. In *MSUG Conference*, pages 1–10, 1997.

[16] G. Koole and A. Mandelbaum. Queueing models of call centers: An introduction. *Annals of Operations Research*, 113(1-4):2–10, 2002. ISSN: 0254-5330.

[17] P. L'Ecuyer. Modeling and optimization problems in contact centers. In *Third International Conference on the Quantitative Evaluation of Systems-(QEST'06)*, pages 145–156. IEEE, 2006. ISBN: 0-7695-2665-9.

[18] B. G. Lewis, R. D. Herbert, and W. J. Chivers. Modelling service levels in a call centre with an agent-based model. *World Review of Science, Technology and Sustainable Development*, 7(1-2):2–12, 2010.

[19] S. Q. Liao. *Staffing and shift-scheduling of call centers under call arrival rate uncertainty.* PhD thesis, Ecole Centrale Paris, 2011.

[20] T. A. Mazzuchi and R. B. Wallace. Analyzing skill-based routing call centers using discrete-event simulation and design experiment. In *Proceedings of the 2004 Winter Simulation Conference, 2004.*, volume 2, pages 1812–1820. IEEE, 2004. ISBN: 0-7803-8786-4.

[21] J. Pichitlamken, A. Deslauriers, P. L'Ecuyer, and A. N. Avramidis. Modelling and simulation of a telephone call center. In *Proceedings of the 35th conference on Winter simulation: driving innovation*, pages 1805–1812. Winter Simulation Conference, 2003. ISBN: 0-7803-8131-9.

[22] R. Srinivasan, J. Talim, and J. Wang. Performance analysis of a call center with interactive voice response units. *Top*, 12(1):91–110, 2004. ISSN: 1863-8279.

[23] R. Stolletz. *Performance analysis and optimization of inbound call centers.* Springer Science & Business Media, 2003. ISBN: 978-3-540-00812-5.