Louvain School of Management

and

Instituto Superior Técnico de Lisboa

Improvement of a Queuing System through Simulation and Lean Methodologies

Project Master's Thesis submitted by
Carlos Guilherme Almeida de Frias

with a view of getting the degrees
Master 120 credits in Industrial Engineering and Management
Master 120 credits in Business Engineering, professional focus

Supervisor at Instituto Superior Técnico
José Rui Figueira

Supervisor at LSM
Per Joakim Agrell

Academic Year 2016-2017

People would rather live with a problem they cannot solve

than accept a solution they cannot understand.

Robert Woolsey

Abstract

This Dissertation presents a real problem to which a consultancy firm, KICG, was confronted during a project in EDP Stores. As it will be detailed, customers, who currently visit these stores, must endure high waiting times which affects their satisfaction-levels. Not being aligned with EDP's corporate desire, the company aspires for a more efficient service of its clients. Thus, a tailored framework is hereby proposed to assist on tackling the identified causes and convenes a process that improves the overall performance of the system.

The present Dissertation gathered the knowledge from more than 100 sources that enabled a good characterization, understanding and modelling of the system and contributed to the conception of a tailored framework that resorts to a symbiotic result of three research areas: Queuing Theory, Simulation, and Lean.

After reading this work, it shall be clear that queues impact everyone on a daily basis, and that, consequently, have a crucial impact on both the welfare and the economy of a country. The proposed framework aims at improving any system's performance, but, in this specific case, it was implemented in EDP where the goal was to simultaneously increase customers' satisfaction levels and the system's performance (monitored by the market's regulator). Having proved to mitigate existing hazards and to improve the system's performance, the tailored framework's output is now being implemented by the firm. Its impact on Bragança's store performance was simulated after adopting best-practices and redesigning the system's capacity, leading to an estimated reduction of the waiting-time of 76%.

Key-Words: Queuing System; Waiting Time Reduction; Queuing Discipline; Discrete Event Simulation; Lean Improvement.

Acknowledgements

This work encompasses my Dissertation Thesis of both my Masters (Industrial Engineering and Management, and Business Engineering). Now that the all its 103 pages have been written, designed, and redesigned, it is with great pleasure and a greater feeling of accomplishment that I am publishing this work. Being able to, after 5 years of Engineering school, tackle a real issue of the biggest Portuguese company seems unbelievable. Yet, looking back, every given step towards this moment, makes it more be reasonable. After all, thank God, I have had a family supporting and inciting me to constantly do more and better. Being raised in this environment, where good moments are cherished and bad ones are taken as lessons, allowed me to focus and to enter two of the best European universities. I have to thank every single graduate Professor I had, for all the knowledge he, or she, may have passed to me, but also to my undergraduate (and while I am doing it, pre-university) Professors who have taught me to discipline my work and to strive for the best attendance possible, as «if to study is your work, do it well». Throughout my studies I have interacted with many international initiatives, cultures and people to whom I also want to thank, as not only they allowed me to train my language skills, but, far more important, to improve as a person. To conclude, I also want to acknowledge some concrete people who I can call friends for having helped me throughout this journey. Firstly, to both my supervisors who have been the Professors with whom I have identified the most during my Masters, by inciting me to always reach higher levels and helping me to discover new academic fields. For this and for all their feedback and support throughout the thesis, I want to thank to Professor José Rui Figueira and to Professor Joakim Per Agrell. Secondly, I want to thank Professor Marta Gomes, who have sought, from the beginning, to help me in this quest of improving the queueing system of a considerably large system. I still remember to have first met her without knowing a single thing of Simulation, and designing the first drafts of what would be the thesis with her. Thirdly, I want to acknowledge both involved firms for their availability, interest and shown appreciation for the work. Fourthly, I want to thank Joana Lemos for having invested/wasted one night of her life reading my paper and giving me a global appreciation of someone who is not, by all means, familiar with the nomenclature, hence providing me with a readability assessment. And, finally, I want to thank both my parents for having invested, supported and helped me throughout all my life with priceless teachings and love, acknowledging and thanking especially to my father who has effortlessly read and commented my work in both deliveries.

Table of Contents

Table of Figures

Table of Tables

Acronyms

AST – Average Service Time

AWT – Average Waiting Time

BP – Best-practice

CBS – Number of Customers Being Served

CEER – Council of European Energy Regulators

CIS – Customers In-System

CT – Called Time

CV – Coefficient of Variation

D – Deterministic Distribution

DEDS – Discrete Event Dynamic Simulation

DES – Discrete Event Simulation

DK – Daily Kaizen

EDP – *Energias de Portugal*

Ek – Erlang Distribution

EMEA – Europe, the Middle East and Africa

ERSE – *Entidade Reguladora dos Serviços Energéticos* (Portuguese Energy Market Regulator)

ET – Entrance Time

EU – European Union

FIFO – First in First Out

FT – Fast-track

FTE – Full-Time Equivalent

G – Generic Distribution

IEA – International Energy Agency

IMF – International Monetary Fund Home Page

KBS – Kaizen Business Model

KICG – Kaizen Institute Consulting Group

KPI – Key Performance Indicator

LIFS – Last in First served

LK – Leaders Kaizen

M – Exponential Distribution

MC – Markov Chain

MHE – Material Handling Equipment

MWh – Megawatt-hours (a derived unit of energy equal to 3.6 megajoules, times 10^3)

NVA – Non-Value-added (usually related with Lean process mapping)

PD / PL – Priority Discipline / Priority Level

PDCA – Plan, Do, Check, Act

PK – Project Kaizen

QCD – Quality, Cost, Delivery

QCD – Quality, Cost, Delivery

QI – Quality Indicator

ROI – Return on Investment

SBP – Service Best Practice

SDCA – Standardize, Do, Check, Act

SIRO – Service in Random Order

SK – Support Kaizen

SOP – Standard Operating Procedures

SPT – Shortest Processing Time

TNA – Total Number of Arrivals

TNC20 – Total Number of Customers who have waited more than 20 minutes

TT – Ticket-Type

TWh – Terawatt-Hours (a derived unit of energy equal to 3.6 megajoules, times 10^9)

UN – United Nations

VA – Value-added (usually related with Lean process mapping)

VSM – Value Stream Mapping

WIP – Work in Progress

WT – Waiting Time

1 Introduction

In this introductory Chapter, a summary of this Dissertation's content will be presented. Overall, after studying of three research areas, the analysed case study presents improvement opportunities to raise its efficiency levels, if this work's learnings are taken into consideration. Thus, to better understand the scope of this work, the next 4 sections will begin by contextualizing this Dissertation's problem and then presenting its goals, methodology and structure.

1.1 The context of the problem

This Dissertation will focus on a problem that is part of everyone's daily life – queuing. Being formed whenever the demand surpasses a system's service capacity, waiting lines are witnessed everywhere (*e.g.,* hospitals, industry, traffic, shops, *etc.*) (Asmussen, 2008). In economic terms, queuing impacts the national welfare by reducing the efficiency level of its economy. After all, the time that could be used productively is, instead, harming or even disrupting the economy, as both citizens and organizations will be wasting time waiting for almost every task they are planned to do (Hillier & Lieberman, 2010).

Research about Queuing Systems began with Agner Erlang's application on telecommunications in 1918, and it has since then been developed by many other academic fields, such as: Operational Research, Industrial Engineering, Social Sciences, or Medicine. Whenever they are not properly handled, queues will severely harm any system's efficiency, justifying the broad interest in mitigating their impact. Examples include systems where queues were responsible for up to 90% of the lead time in industrial environments, or even for severe health complications and complaints in hospitals (Ndukwe *et al.*, 2011).

Thus, both governments and organizations are prioritizing efforts to carefully analyse the queuing phenomenon, and to seek solutions that would mitigate them. As explained by Nosek & Wilson (2001), this institutional commitment has led to several corporate achievements such as an increased delivery-speed of pharmaceutical companies, or a better staff levelling, leading to significant financial savings and higher customers' satisfaction levels, as they will be the biggest beneficiaries, whenever their waiting period for a given service is reduced (Zehrer & Raich, 2016). For a given demand, this may be done by increasing either the system's capacity, or the overall queue management efficiency (Gutacker, 2016).

Aiming at improving the service level in its stores, the Portuguese firm *Energias de Portugal* (EDP), a vertically integrated utility company with a relevant presence in the global energy market, has hired the Kaizen Institute Consulting Group, a Lean consultancy firm, to better balance its stores capacity with the stores' demand. As it will be seen during the Dissertation, even though stores activity is heavily regulated, there will be room for improvement. In fact, while complying with the legislation, the proposed framework will reduce the average waiting time (AWT) of a specific store by 67.6%, hence proving its worthiness to be applied in the remaining stores of EDP. This will be accomplished by fostering efficient processes that comprise less wasteful activities based on best-practices (BP), and by proposing changes to the current management process of the queuing systems. All the uncovered solutions will have their impact simulated and analysed, after guaranteeing their compliance with both the regulator and EDP's corporate views.

1.2 Goals of the Dissertation

After meeting with EDP's performance team, the ambition of increasing the customers' satisfaction became clear. Being a crucial moment for the firm, as the market is witnessing a steep increase of competitors with the liberalization of the Energy market, EDP is seeks to simultaneously reduce the clients' AWT in-stores and increase its Service Quality Indicator, which is continuously monitored by ERSE (the market regulator).

Therefore, the goal of this Dissertation is to develop a framework that identifies improvement opportunities and sets reasonable targets based on both demand patterns and store layouts. The framework will present the following three functionalities using three software (MS Access, MS Excel, and Simul8):

1. Identification of the best servers in each store: Those servers (employees) who have lower service-times and practice a robust process (*i.e.,* with a low variability) should be identified and benchmarked. Not only will their processes be mapped and shared among their colleagues, but will also set a realistic performance target to the entire team. By sharing these best-practices across the different stores, EDP's support team will simultaneously raise the productivity level of their employees and decrease their variability.

2. Layout Improvement: As it will be explained in Chapter 3, customers are characterized by ticket-types upon their arrival to the store. Stores are currently managed by Inline®'s ticket system that offers the possibility of automatically calling the customers to each service-facility (either a table, or a counter). This allocation follows a set of predefined priority levels that would allegedly reduce the customers' AWT according to the forecasted length of their service and the number of customers who are still waiting in-store. However, most servers prefer to ignore this feature and to manually calling their clients. According to their experience, the automatic mode deteriorates the performance of the store. Thus, new proposals for allocating the tickets to the service-facilities should be identified and tested, estimating their impact on the AWT of the store.

3. Sensitivity Analysis: As its last functionality, the framework should study the AWT change upon the hiring, or the dismissal, of one employee. This way, EDP could start forecasting the required investment to reach a specific AWT (*i.e.,* the number of servers to be hired), whenever the two last functionalities would not prove to be sufficient.

1.3 Methodology

To ensure the achievement of every Dissertation goal, the workflow was planned with two stages.

In the first one, Chapters 2 to 5, a theoretical approach would be taken, describing both the system's stakeholders and its problem. This would enable a clear definition of the project's scope and of its ambition. A literature review would follow highlighting relevant academic knowledge when characterizing and understanding the system in-hands. Furthermore, existing frameworks that could potentially improve the system's performance should also be presented. Considering these findings, a tailored framework would be proposed through the consolidation of different papers, authors, and research fields to avoid a silo-minded solution.

The second stage, Chapter 6, would implement the previously designed framework, after consolidating and analysing the stores' data, ultimately leading to the achievement of its three goals. Once all the results are consolidated and verified, an objective comparison between the "as is" and the "to be" state will lead to robust proposals for improving the system's performance.

Together, these two stages will convene every step required to generate a substantiated solution to be discussed in Chapter 7.

1.4 Structure of the Dissertation

Following the previously described methodology, a clear structure segments the present Dissertation Thesis into 7 Chapters:

Chapter 1 summarises the entire Dissertation, explaining the motivation behind its subject, the goals defined for the Dissertation, and the methodology through which they are going to be reached.

Chapter 2 presents Kaizen Institute Consulting Group and its business model. Throughout this description, the motivation of implementing some Lean concepts will arise as they will present themselves as having symbiotic effects on this work.

Chapter 3 presents EDP and the Energy market to which it belongs to. With a promising future, EDP is focused on providing and assuring a better customer experience in its stores. After describing the stores' environment, depicting their business and stating the relevant legal restrictions, queues will become a problem to tackle by developing a future ambition where they are mitigated.

Chapter 4 explains and summarises the Literature Review that was conducted to understand and tackle the problem in-hands. To that end, the state-of-the-art of three academic areas will be presented in individual sections: Queuing Theory, Discrete Event Simulation, and Lean Theory. The first two sections will assist in understanding, characterizing, modelling and simulating the stores' system, whereas the last section will aim at learning how to minimize the variability of service-times, and to successfully implement unveiled solutions in the real environment.

Chapter 5 plans the methodology that is to be followed during the implementation of the framework. By following its work-steps, answers to the Dissertation goals will arise and a comparison between the current and the proposed future state will be presented.

Chapter 6 summarizes the findings of the implementation of the previously proposed Framework, aiming at concept-proving its methodology and its ability of unveiling interesting improvement opportunities.

Finally, Chapter 7 concludes this work, by referring its main takeaways, future-work development possibilities, and by advocating the Dissertation's impact, feasibility and implementability in a real system.

2 The Consulting Firm - Kaizen Institute Consulting Group

The scope of this project was developed during an internship at Kaizen Institute Consulting Group (henceforth referred to as KICG). Thus, before defining the client's problem, it is important to clearly understand the philosophy of this firm and how some of its tools might present useful for this Dissertation.

2.1 Company's History and Mission

For 23 years, KICG's founder, Masaaki Imai, travelled around the world with Taiichi Ohno, known as the father of the Toyota Production System. While spreading Lean's philosophy among Western CEOs, Imai realised the existence of a knowledge-gap between the Western and Eastern societies that was hampering the economic growth. Thus, he founded KICG in Switzerland in 1985 to mitigate this gap (KICG, 2017a).

Since its inception, KICG's mission has been to share the benefits of Lean and to facilitate its implementation across virtually every business sector. Nowadays, KICG is present in over 40 countries, fostering the development of consultants who not only seek tangible financial results, but also develop people in these organizations, hence supporting the sustainability of the new achievements through the adoption of a new business culture. After all, only by transferring its knowledge and expertise, will KICG enable its clients to continue their quest for permanent self-development and improvement (KICG, 2017a), following the intrinsic values of Lean's fundamentals presented in the literature review (Chapter 4).

2.2 Kaizen Business System – Managing change in an organization

It is KICG's belief that companies who base their management on excellence, people development and Lean will obtain a sustainable competitive advantage. To that end, KICG has developed the Kaizen Business System (KBS), presented in Figure 1, that has been progressively applied to different business sectors and has contributed to KICG's leading position as a Lean consultancy firm (KICG, 2017b).

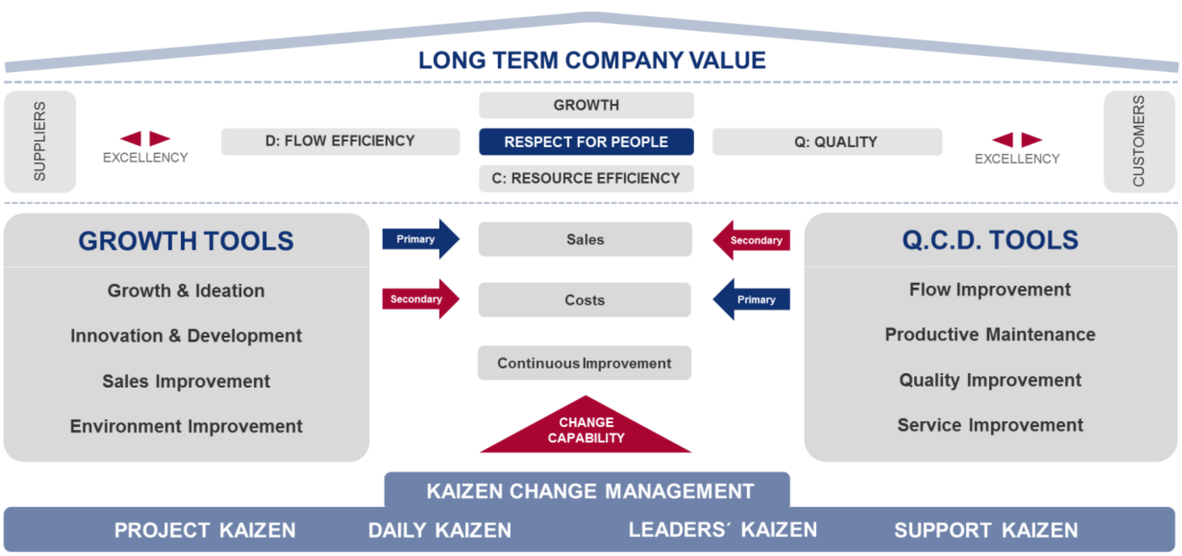

Figure 1: Kaizen Business System (KBS)

2.2.1 The Strategic Vision of the Kaizen Business System (KBS)

KBS should be seen as a house whose base is the set on Lean's principles (*cf.* subsection 4.3.1), and the roof is KICG's ultimate goal of obtaining a sustainable cycle of continuous improvement. In each company, the chosen strategy shall always be based on: the respect for people; growth; quality; resources efficiency; and flow efficiency. All these goals will ultimately foster a good relation among the firm, its clients and its suppliers (KICG, 2017b).

In order to reach this strategic vision, KBS proposes several tools that may be chosen according to the challenge in-hands. These tools are grouped in two operational models: the QCD Model with a primary impact on efficiency; and the Growth Model that aims at increasing sales. Both are supported by a robust change management system (the Change Management Model) that incites people to not only embrace, but also propose change in their business environment. Together, these three models will sustain KBS' strategic vision, and thus, they are known as the pillars of Kaizen (*cf.* subsection 2.2.2) (KICG, 2017b).

2.2.2 The Tools of the Kaizen Business System (KBS)

This subsection will develop the three previously mentioned KICG's models. It is important to emphasise that any combination of tools is valid, as long as every facet of the problem is tackled (KICG, 2017b).

2.2.2.1 Growth Model

Aiming at improving the company's strategy, the four tools of this model will increase the company's sales through both innovation and improvement of its processes (KICG, 2017d):

- Growth & Ideation (GIP) supports the company to manage its disruptive innovation initiatives by mapping its ideation process and defining its business plan. Both will consider the new product's life cycle and its impact on the company's portfolio;
- Research & Development (IDM) fosters agile software development aiming at optimizing any digital department of the company. This will reduce R&D's time-to-market (the time length between developing an idea and its launch), and it will lower production costs through waste (*Muda*) elimination;
- Marketing & Sales (MSI) provides solutions to accurately record and manage customer experiences, improving the effectiveness of the firm's marketing initiatives. This is accomplished by highlighting areas for improvement, optimization of pricing and margins, and dimensioning of the sales team;
- Environment Management (TEM) identifies and tackles current environmental hazards by mitigating, or even neutralizing them. The main goal of this transversal tool is to reduce the company's environmental impact by adopting energy-saving solutions, or an efficient use of resources.

2.2.2.2 Quality, Cost, Delivery (Q.C.D.) Model

Aiming at improving quality levels, this model will increase the efficiency of the processes by reducing costs, and improving service levels (*i.e.,* on-time-deliveries) thanks to four tools (KICG, 2017e):

- Flow Management (TFM) increases the efficiency of the flow on both logistic and production channels. This is achieved through a VSM (*cf.* subsection 4.3.2.3) that will map all the encompassed activities in a process. After identifying and removing wasteful activities from the process, its future vision can be designed with better efficiency levels and often shorter processes;
- Productive Maintenance (TPM) increases the equipment productivity thanks to a focused problem-solving methodology (*Kobetsu Kaizen*). In contrast with TFM that focus on the efficiency of the flow, TPM optimizes the efficiency of a certain equipment by: implementing staff development initiatives, raising awareness about preventive maintenance, and fostering the use of SDCA/PDCA cycles;
- Quality Management (TQM) aims at increasing quality levels, through standardization (SDCA cycle), trainings, and six sigma tools;
- Service Management (TSM) aims at increasing the efficiency of service flows, using TFM tools.

2.2.2.3 Change Management Model

Finally, in order to ensure a sustainable implementation of the previously mentioned tools, KICG offers four programmes that aim at establishing the Lean culture across its clients. These programmes are segmented by audience and by the service requirements (KICG, 2017f):

- Daily Kaizen (DK) develops and sustains people's growth, usually tailored to the Gemba teams;
- Leaders Kaizen (LK) ensures the top-management involvement and the adoption of Lean philosophy;
- Breakthrough Kaizen (BK) assists on the implementation of disruptive solutions during Projects;
- Support Kaizen (SK) offers specialized guidance on ongoing improvement initiatives.

2.3 Chapter's Conclusions

This Chapter has presented KICG, a Lean consulting firm, who, after 30 years implementing Lean, has developed a set of frameworks convened in its KBS model. The importance of Lean tools such as those convened in KICG's TQM is stressed in the literature, where its adoption is considered a strategic factor for the long-term success of a company. After all, for example, under 6-sigma, processes disturbances may only shift by as much as 1.5 standard deviations off-target, hence promoting a variability reduction that leads to less than 3.4 defects per million opportunities (*DPMO*) (Montgomery and Woodall, 2008).

Lean proves to be impactful in any company, or system, as it increases its performance through systematic processes which are not only performed by employees, but also adopted culturally, assuring its long-term sustainability. These principles will later prove to have motivated further research on Lean methodologies and to have shaped the approach of this project by fostering objective problem-solving analyses and by continuously seeking higher efficiency levels without sacrificing neither the customers' experience, nor employees' motivation.

By introducing KICG, this Chapter performed as preamble of the KICG's client presentation. Since this Dissertation will propose a solution for EDP's problem, the next Chapter must first address its several dimensions and then converge to concrete goals that shall be achieved during the Dissertation.

3 Case Study

This Chapter aims at presenting KICG's client, *Energias de Portugal* (EDP). As the Dissertation will tackle an existing problem in EDP stores, it is important to first understand its business through a brief market contextualization and, only then, deep-diving on the two facets of the problem: operational and regulatory.

3.1 The Client – EDP

This section will convene a general presentation of the Energy market and its players in Portugal. At the end, one should understand the business of EDP and be aware of the challenges presented by the regulator and the competitors.

3.1.1 Company's History & Facts

EDP is headquartered in Lisbon since its inception in 1976, when it was incorporated as a public enterprise, comprising 13 recently-nationalized companies. Forty years later, EDP succeeded in gaining a significant international presence, being present in 14 countries. It relies on its 12 thousand employees to handle its seven business units: Electricity Generation, Renewable Energies, Electricity Distribution, Electricity Supply, International Consulting, Gas in Iberian Peninsula, and Brazil Operations (EDP, 2016).

As a vertically integrated utility company, EDP is clearly devoted to be a world leader among the renewable-energy providers. In 2016 alone, 64% of the total generated electricity (70TWh) was obtained from renewable sources (*cf.* Figure 2). Being especially invested in Wind energy, EDP is currently the 5th largest wind power operator world-wide, with 197 wind farms spread among 12 countries as result of many business ventures (The Wind Power, 2017).

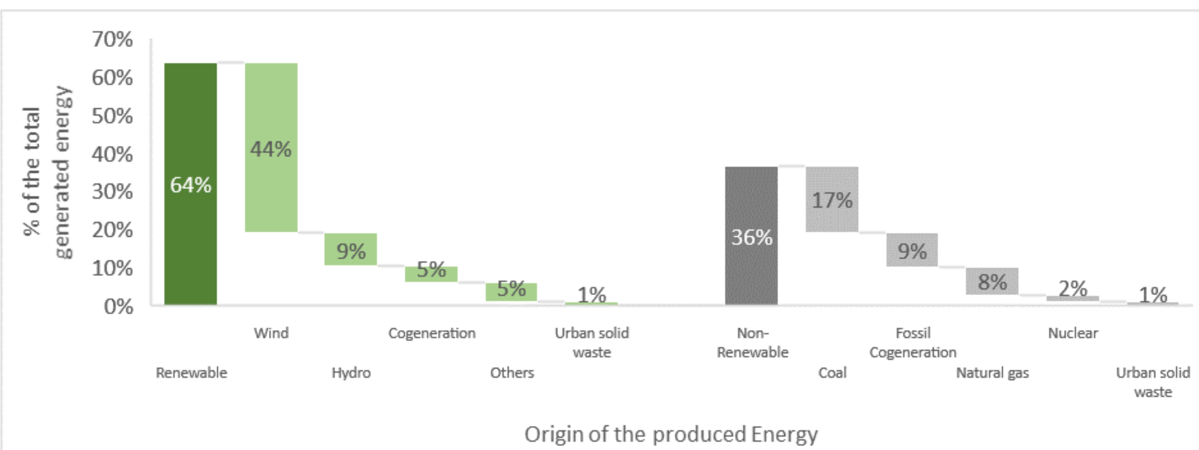

Figure 2: EDP Energy Sources in 2016 (adapted from EDP, 2017)

When considering all its business units, EDP generates, distributes and supplies electricity to 9.8 million customers and it distributes and supplies gas to 1.5 million customers, being the 3rd largest electricity generation company in Iberia for 3 years in a row (*cf.* Figure 3a), one of this region's largest gas distributers, and its 3rd largest electricity marketer for more than 10 years (*cf.* Figure 3b) (EDP, 2016).

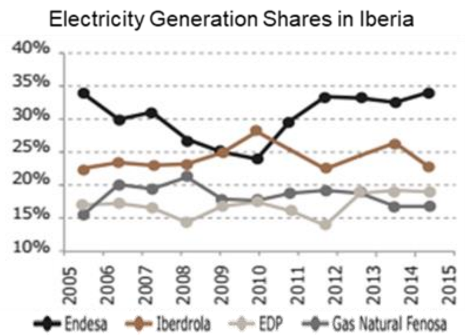

Figure 3a: Generation Shares (EDP, 2015)

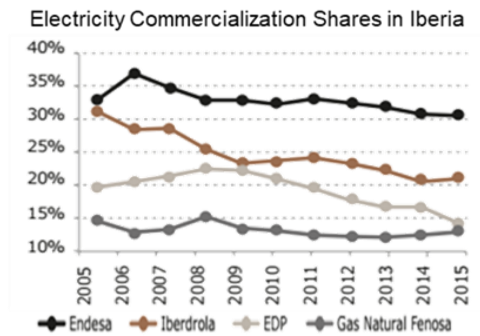

Figure 3b: Commercialization Shares (EDP, 2015)

Overall, EDP ranks among the best in the world regarding its financial performance, having the 56th top-performance worldwide, and the 5th when only considering the EMEA region and Electric Utilities companies (Platts, 2016). To support this growth, EDP efficiently manages all its business units by dividing itself in 7 different entities that may be consulted in Table 1. However, potentially influencing this performance, it should be emphasized that the Energy market was liberalized in 2000 and the process is still ongoing. On the one hand, EDP ceased to be the sole electricity provider in Portugal, currently having 17 competitor providers in the Portuguese electricity market. On the other hand, EDP started being able to commercialize gas nation-wide and is one of the 10 players in the gas market. This, together with the fact that EDP's last resort services (*i.e.,* its operations in the old regulated market) are to be extinct, will potentially transform the Energy market landscape, as explained in the next subsection (ERSE, 2017).

To serve its clients, EDP offers a network of stores that are present across the Portuguese territory. As the stores' focus is the residential clients and SMEs, not every listed entity is represented in the stores. In fact, analysing Table 1, stores' personnel will not be able to handle any process related with the two generation entities. Moreover, it is interesting to notice that each represented entity has a specific code in EDP's ticket-managing system Inline® (*cf.* subsection 3.2.2).

Table 1 : EDP Group and its presence in EDP Stores

EDP Companies	Short-Description	Inline® Code
EDP Production	Electricity Generation	*N/A*
EDP Bioelectric Production	Electricity Generation under a special biomass regime	*N/A*
EDP Distribution	Electricity Distribution across the country	ORD
EDP Gas Distribution	Gas Distribution across the country	Portgas
EDP Universal Service	Electricity Commercialization under the regulated market	CUR
EDP Gas Universal Service	Gas Commercialization under the regulated market	EDPGasSU
EDP Commercial	Electricity and Gas Commercialization under the liberalized market	EDPC

3.1.2 Overview of the global Energy Market

As a preamble of the Portuguese Energy market, this subsection presents the global energetic trends as they will ultimately set regional paces. In the latest study available, Energy Demand is forecasted to increase on a yearly average of 1.4% until 2040 (IEA, 2016), while global economy is expected to grow 3.4% per year (IMF, 2016). As seen in Figure 4, a significant reduction of the energy intensity is therefore expected, meaning that energy efficiency is estimated to increase. The Paris Agreement signals that trend, as 143 countries have pledged to challenging changes, assuring a minimum of 27% increase in energy efficiency by 2030 (UN, 2017).

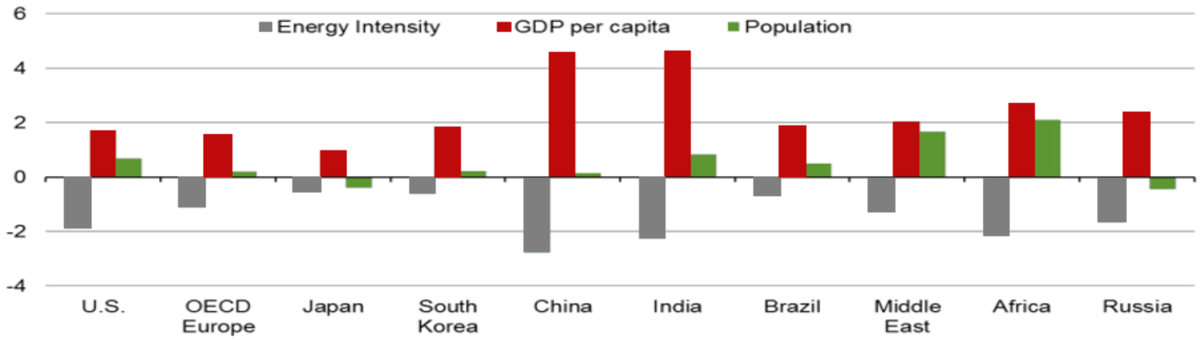

Figure 4: Forecasted average annual percent change between 2012 and 2040 (IEA, 2016)

According to that same study, a new energy source-mix will be witnessed as well. As seen in Figure 5, Natural Gas and Renewables will grow 13% and 42%, becoming the 2nd and 4th main global energy sources, respectively, whereas both Oil and Coal will have their share decreased (IEA, 2016).

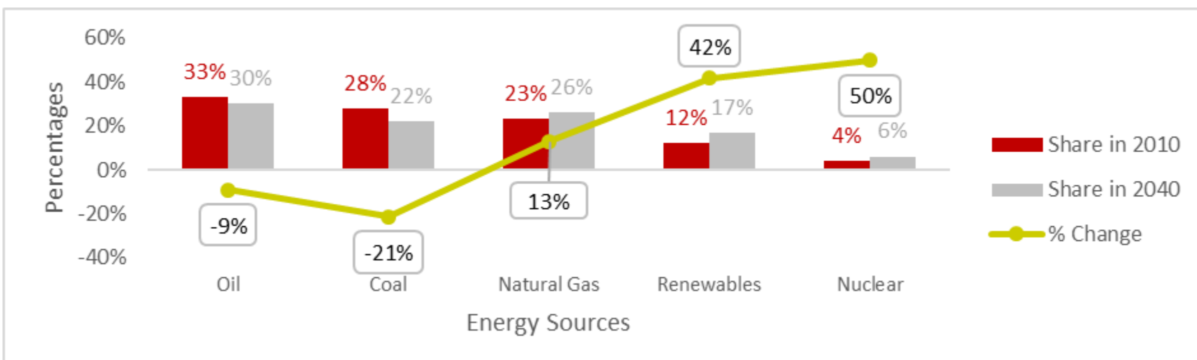

Figure 5: Global Energy Supply Sources – Forecasted growth between 2010 and 2040 (adapted from IEA, 2016)

Leading these global trends, the European Energetic Policy set by the EU Commission aims at: reducing Greenhouse emissions by 40%; having at least 27% of its energy generated from renewable sources; and finally, improving energy efficiency by 30% until 2030. At the same time, it is being witnessed a reinforced desire to interconnect all the EU energetic markets and implement a governance model that would coordinate their efforts on implementing the above-mentioned targets (European Commission, 2016).

3.1.3 Portuguese Energy Market and its Regulation

In line with both previously mentioned EU targets and IEA forecasts, in 2016, Portugal has increased its electricity consumption by 0.6%, reduced its coal-sourced generation by 14%, and increased its natural gas consumption by 7%. Thanks to a lower residual thermal demand in 2016, power prices have decreased to 40 €/MWh (a 20% decrease when compared with the 2015's price) (EDP, 2016).

Regarding market's regulators, the Portuguese Electricity market has been regulated by ERSE since 1997. It was not until 2002, though, that ERSE started regulating the gas market too, as part of a European Commission instruction (Law Decree nr. 97/2002). As a member of the Council of European Energy Regulators (CEER), ERSE's mission is to promote transparency and to monitor every company in the Portuguese Energy market, regulating it on behalf of the consumers' interests (ERSE, 2017).

Following its mission, one of ERSE's most impactful initiatives was the extinction of regulated tariffs. The liberalization of the market in 2000 aimed at a greater choice, better prices and better quality through competition among new players. Even though the transition was free and mandatory, there are still some consumers to be transferred to the liberalized market, being allowed to remain in the regulated one until the end of the transitional period (several times extended, but currently set to 2020). So far, 92% of the total electricity (in volume) is already commercialized by the liberalized companies (EDP, 2016).

Another impactful ERSE's initiative aimed at raising the quality of all the provided services. Thus, according to the Law Decree nr. 455/2013, every energy provider must guarantee a minimum service-level and it is liable for any wrongdoing, or any harm caused to the client. Moreover, this law clearly characterizes which services and information must be guaranteed and given to the clients by the providers, respectively.

3.2 Description of the Problem

After presenting the Energy market trends, its regulations and EDP's business of in the last section, a focused description of EDP Stores will be conducted next to clearly describe the problem in-hands.

3.2.1 Company Owned Stores Operation and its legal constraints

As previously explained, most of EDP entities are represented in stores across the country (*cf.* Table 1). Currently, EDP owns 41 stores, and rents another 25 to different agents. Yet, none of them are operated by EDP. Instead, they are operated by three distinct service providers, to which EDP has transferred the daily management of the store and its staff. These providers must guarantee that this fact is not perceivable to the customers by following the same operational norms. Every store is open during business days (Monday to Friday, excluding national holidays), from 8h30 to 16h00, and will serve any customer who enters the store while it is still open (the staff is in-store from 8h00 until 17h00). During the design of the stores network infrastructure, each store was built with a specific layout according to its expected demand. The layout comprised a specific number of service stations, split among counters and tables.

With more than 2.3 million visitors in 2016, the distribution of EDP stores layouts and their respective demand curve may be seen in Figure 6. After analysing this figure, interesting conclusions may be taken. For example, EDP has two stores with 12 service-stations that clearly do not follow the tendency line of clients per station. In this case, each one of the two stores has served, on average, 76.5 thousand customers in 2016 which is below the trendline by about 23 thousand clients than it would be expected to be. This may indicate either an overestimation of the demand, or visually highlight a shift of the expected demand patterns (*e.g.,* a decrease of clients in that region, or even a change of habits that result on going to a second store in the region).

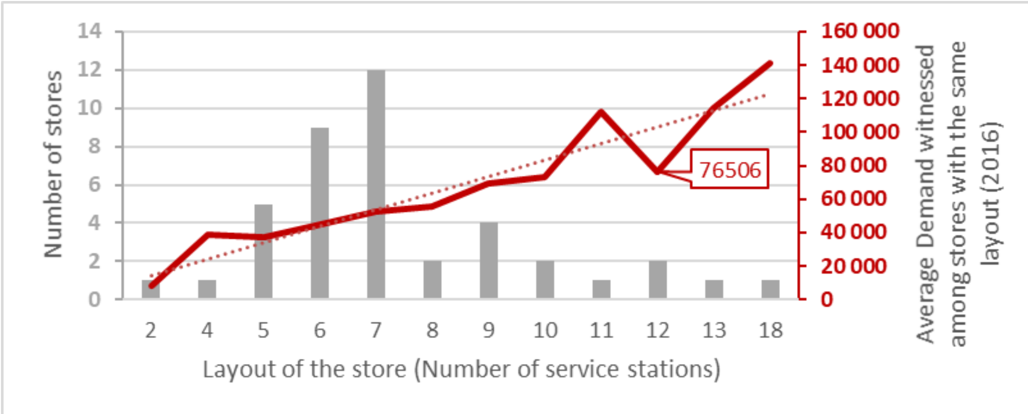

Figure 6: Distribution of stores with a certain layout and their average demand in 2016

As long as the current legislation is in force, these stores must remain open. The previously mentioned Law Decree nr. 455/2013 mandates EDP to provide a physical customer service, since it is a last resort provider. Moreover, as it serves more than 5000 clients per year, EDP must also assess its performance by recording the waiting time (WT) per customer. By law, the WT is defined as being the elapsed period of time from the customer's arrival to the moment when he is called by his final server, meaning that, whenever a customer is transferred to second server, these interactions will be still considered. The performance assessment follows Equation 1 (Law Decree nr. 455/2013, 2013):

$$QI = \frac{C_{<20}}{C_T} \tag{1}$$

Where QI, the quality indicator, equals the quotient of the number of served clients under 20 minutes of waiting time ($C_{<20}$), by number of the total served clients (C_T). The QI disregards clients who have abandoned the queue, since it could mistakenly include those clients who have just taken more than one ticket. The WT is influenced by the average service time (its reduction will allow to serve more clients per hour), the customers demand (its increase implies a higher congestion in-store), and the number of assistants (decreasing the number of servers, decreases the number of clients being served at a certain moment). Finally, the most recent impacting law decree (nr. 58/2016) grants priority to pregnant women, the elderly, disabled citizens and parents carrying young children and must be enforced in every store.

3.2.2 Inline® – Ticket management system

To accurately record the waiting time, EDP implemented a ticket management system in all its stores in 2006 (Tomás, 2017). EDP opted for Inline® solution, from Tensator, a UK consulting firm whose tools have emerged from a benchmark as exemplar in managing "customer journeys". Being present in over 150 countries, Tensator designs tailored solutions to record the customers' interactions with a high ROI rate (Tensator Group, 2017). Convening EDP's needs, a final solution was implemented across EDP stores, after a prototyping-phase. The solution encompassed 5 different components depicted in Figure 27 in Appendix (Tomás, 2017). Overall, this system is now recording every step of the customer journey in EDP stores, with an especial focus on the client's waiting time.

3.2.3 Current operation of the System

To understand the stores operation, this subsection describes the journey of each client in-store and its main events, both represented in the flowchart of Figure 7.

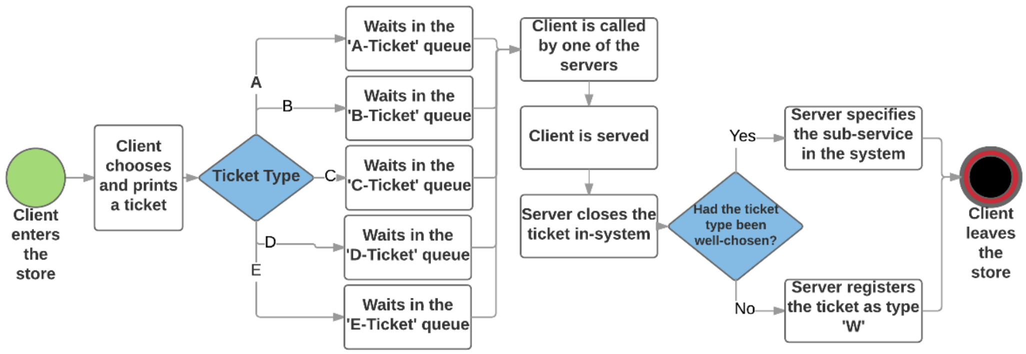

Figure 7: The client's journey in-store

Whenever clients enter an EDP store, they will classify themselves into one of the five service-types (or ticket-types) according to the purpose of their visit ('A' for Contracts; 'B' for Information; 'C' for Complaints; 'D' for Technical Matters; and 'E' for Payments). Then, clients must wait until their ticket is called by one of the store's employee. The server is then responsible to solely serve the client, since there is no forwarding of customers among colleagues and only occasionally will the server request assistance from the store manager on extraordinary situations. After serving the client, the server assigns a specific sub-service to the client through his ticket that will entail more information about the provided service and will enable the monitoring of the WT by entity (e.g., for ticket A, there are 13 possible sub-services, including several types of contracts, terminations and sales for each entity of EDP). Moreover, there is also a virtual 'W' ticket to handle clients who have not printed their tickets, either due to a system breakdown, or because they had priority in accordance with the law decree nr. 58/2016. By definition, 'W' tickets have a WT equal to zero, since they are "printed" by the server himself.

Thanks to the implementation of this system, EDP is fulfilling its obligation to record the waiting time per client and per type of inquiry, since from that moment on, any event of the client's journey is recorded in Inline®. Moreover, this solution also allows to automatize the calling process, so that, on their end, servers (*i.e.,* assistants) do not know which client they need to call, but solely focus on their function – to serve. Inline® prioritizes clients based on their service-type by grouping them into 3 levels (L1, L2 and L3). Even though the levels composition varies according to the service station, the rule is the same. Each service station serves the oldest ticket in its L1 (composed by one or more service-types) until there are no clients left. Only then, will the service-types in L2 will be called, followed by L3. The diversification of the ticket-types that each station has on its priority levels aims at guaranteeing service flexibility, since there will be no dedicated stations to a specific ticket-type.

3.2.4 Current performance of the system

To guarantee the above depicted process, stores' layout ensures that the customers' WT is recorded from the moment they enter, as required by law. This is achieved by placing ticket dispensers right next to the entrance door (see Figure 28 in Appendix). Inline® was also designed to allocate ticket-types to the previous three levels according to a scenario where seated service stations (tables) call for tickets with high service times (As and Ds), whereas quicker services would be called to the counters (where the client would be served while standing), hence striving for a quicker clients' rotation, without compromising their comfort. Therefore, usually tables have in their L1 service-types with high average durations, whereas counters will only call them if there are no other clients in-store (*i.e.,* L3), and vice-versa.

Two main problems arise in this process, though. The first one is due to clients' misperception of the nature of their own inquiries. As the law states clients must freely print their ticket, servers cannot assist on this task. Thus, they will only realise this situation after calling the client. At that moment, and even though, by law, clients would need to take a new ticket and go back to the queue (as each type of ticket has specific waiting-time targets), servers often ease the situation by manually opening a virtual ticket 'W' and resuming the service (*ca.* 5% of the services). In fact, the same solution is often used when a customer has more than one issue to solve. The first ticket is correctly classified, having its specific WT recorded, while the second has a virtual WT of zero seconds, as the client was already being served. This way, both WTs and service-times will be recorded. The second problem is that employees are not relying on the automatic mode, as according to their experience, Inline® mishandles the process, either by failing to maintain the AWT under-control, or by potentially harming their individual scores. Thus, they call the oldest ticket (*i.e.*, with the highest WT), often bypassing the planned ticket prioritization and allocation of service-types to the service stations.

Overall, the current operational performance in 2016, and following Equation 1, EDP's national QI is 93.1%, having almost 150 thousand clients with WTs greater than 20 minutes. Moreover, as shown in Figure 8, from the stores' demand (*ca.* 2.3 million printed tickets) 88% of the was served. Analysing EDP's demand-mix, 77% of the clients are seeking the 'A', 'C', and 'E' service-types.

In Figure 9, considering that the AWT has the official target of 7min30s, it is important to notice that, on average, only tickets 'A', 'C' and 'D' must wait more than what has been established. In fact, most clients will spend almost as much time waiting as being served in-store. The situation aggravates when considering that 'B' and 'E' clients spend more time waiting than in actually being served.

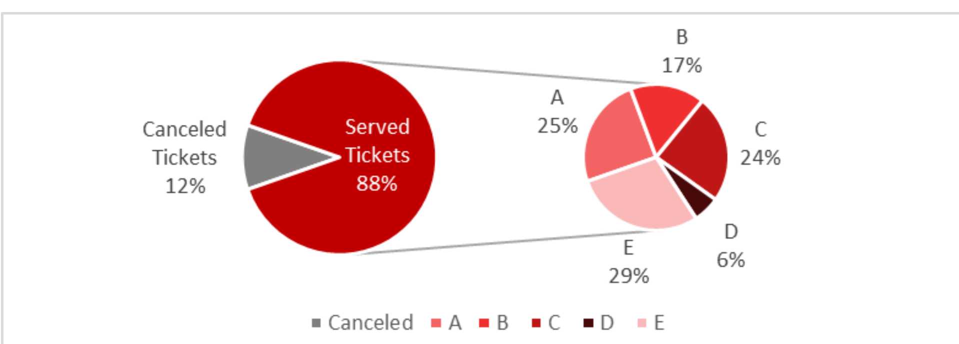

Figure 8: Characterization of the printed tickets in 2016

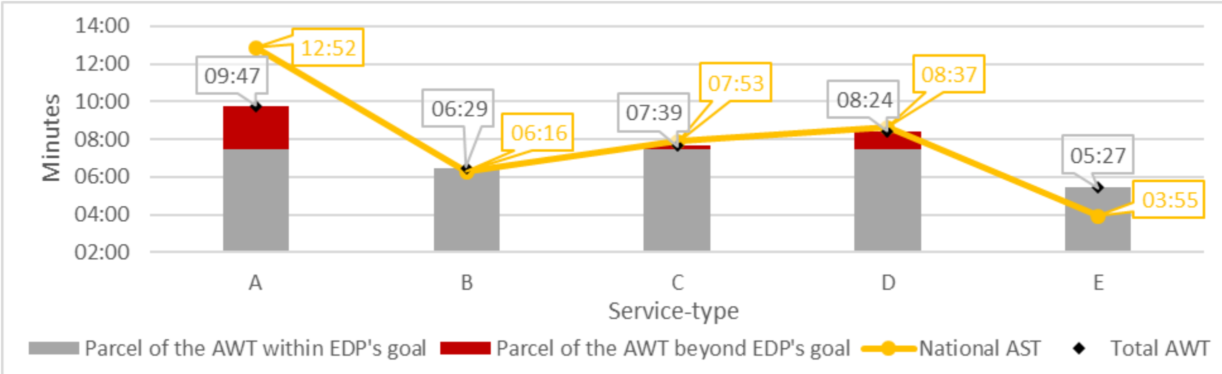

Figure 9: Recorded Nacional AWT and AST in 2016

Nah (2004) highlights the importance of this kind of data analysis, as customers have different tolerance levels. In fact, they will be willing to wait more time whenever they are expecting a complex, or long service, whereas when seeking a store to solve a quicker inquiry they will presume that their waiting time will be proportional, as they will be longing for a shorter stay in-system than otherwise.

3.2.5 Future State Ambition

Having been informed of the current state, EDP's Director of Operations understood that the system was no longer answering to EDP's needs and sought a future-state where EDP stores would have a higher QI and a lower AWT (Ribeiro et al., 2017). It was stressed, though, that this performance increase should not compromise the current AWT of clients with a service-type 'A' (*i.e.,* contractual issues, mainly sought by new clients) as these are crucial to EDP's mission of increasing its market share.

Analysing the literature, this desire to reduce the AWT is often presented and justified with the fact that clients become more satisfied when faced with shorter waiting periods. Indeed, sales will be potentially increased just by reducing client's waiting time, as explained by Zehrer & Raich (2016).

Considering the variables influencing the clients' WT (*cf.* subsection 3.2.1), EDP has proven keen to seek new solutions that would improve the stores performance. Following the methodology presented in Chapter 5, the goal is to adapt, or even change the current system to decrease service-times and to increase the number of servers available at each moment by better managing the store's system. It was also clear that EDP did not want to interfere with the distribution of the demand, the third WT variable, neither by developing appointment services, nor by advertising the time periods with less congestion.

Overall, the three following functionalities were ambitioned: identify service best-practices; improve the layout; and study the impact of hiring, or dismissing one employee. Throughout the implementation of the framework, best-practices should be identified to be subsequently shared across the stores and their staff. Using more than one software, these exemplar performances in terms of service (*i.e.,* servers with a consistently low service-times) and queue management (*i.e.,* implementation of a certain calling order) should be identified. After all, their standardization would reduce the AST, the AWT and both their variabilities.

Therefore, the framework should first seek, test and propose a better way of allocating the 5 ticket-types to the store's service stations, mitigating the problems described in subsection 3.2.4 and increasing the turnover of clients in the store. Then, it should estimate the impact of the adoption of the previously identified service best-practices. And finally, assess the staff-levelling policies impact on the KPIs of the stores (*i.e.,* assessing the impact of the hiring, or dismissal of one employee), especially interesting whenever the previous two steps induced improvement would not prove sufficient.

All these functionalities became goals of the Dissertation, as described in Section 1.2.

3.3 Chapter's Conclusions

This Chapter provided a brief glimpse of EDP's business and the legislation that is impacting its daily operations. By briefly presenting the company and the Energy market trends, EDP has proved capable of tackling the challenges that arise in a competitive environment, and to be succeeding securing leading positions by investing on renewable energy sources world-wide.

Nevertheless, it was also shown that, as in any organization, there is room for improvement. Namely, on EDP stores, the problem of high waiting-times was described through the depiction of the stores daily operations and the laws that are regulating them. After its careful analysis, the focus of this Dissertation became clear and a future vision was set.

Knowing the goal of reducing the average waiting-time, the different dimensions of the problem that needed to be tackled were identified. Thus, the need for potential solutions and frameworks that would assist doing so, arose. In the next Chapter, it will be presented a literature review on different research areas longing for a symbiotic solution that may be developed next in the Methodology.

4 State of the art

After presenting EDP's queuing system and its inherent challenges, this Chapter will connect the academic research work of three distinct areas (Queuing Theory, Simulation, and Lean), hence condensing several decades of knowledge with one goal: to the develop a symbiotic solution (*i.e.,* the proposed framework) that, supported by scientific publications characterises, models and unveils best-practices in the system in-hands. Towards that end, the three ensuing sections will all start with a broader introduction, and only then a focused review on EDP's system to avoid a silo-minded analysis on later stages of this Dissertation.

4.1 Queuing Theory – Characterizing the problem

Queuing Theory is credited to have started with Erlang's first try on modelling a queuing system in the beginning of the 20th century. Erlang (1918) was able to model the Copenhagen telephone network, and to develop the loss and waiting formulas, that are still being used to determine the probability of a full system.

Nowadays, literature on Queuing theory is extensive, but according to Newell (2013) it lacks pragmatism, since complex models are being developed everyday searching for exact and optimal solutions, while a relaxed solution (*i.e.,* a solution that ignores some system's specificities) would often be sufficient. To tackle this issue, assumptions are often presented when modelling a problem to ease its posterior optimization, further improving their overall applicability (Strang, 2012). In this section, this same reasoning is going to be applied, by presenting queues' main mathematical structural elements and their possible assumptions.

4.1.1 Introduction of Queuing Systems

Queues, or waiting lines, are part of everyone's daily life. From going to a shop, to a doctor's appointment, or even in traffic, people will have to wait for their time to either pay, be treated, or cross the city, respectively. The same scenario is witnessed on production systems, whenever a breakdown of a machine occurs. Overall, great inefficiencies may be generated whenever a system is not dimensioned or designed to minimize queues, since entities, or resources, will have to wait. Thus, the role of queuing models is to find solutions that raise the system's performance by modelling and analysing it (Hillier & Lieberman, 2010).

Queuing models are then crucial to organizations, since they provide valuable data and understanding of the system. For instance, demand and waiting-times can be estimated, avoiding the existence of unused service-capacity or a growing number of unserved customers, respectively (Hillier & Lieberman, 2010).

To introduce the terminology used throughout this Dissertation, a generic Queuing System is depicted in Figure 10. As observable, entities (*e.g.,* customers) will arrive to the system from the calling population (*input source)* at a specific rate, remaining in line (the *queue*) until they are called to a service-facility. The entrance in the *service mechanism* follows a specific order (the *queue discipline*) through which the server (human, or machine) calls a client and begins to serve him over a certain duration (*service time*). After completion, the customer leaves the system (Foster *et al.*., 2010).

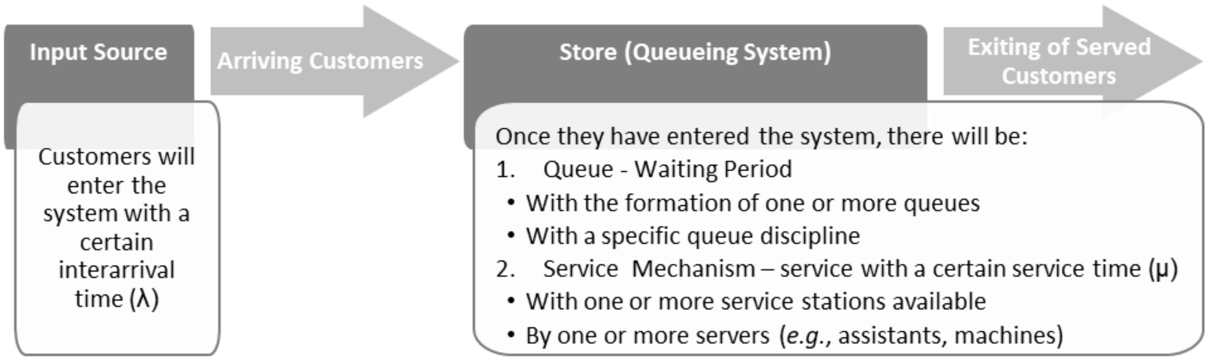

Figure 10: Conceptual Model of a Queuing System in a service store (adapted from Strang, 2012)

To describe queuing systems, Kendall has introduced the A/S/c notation with three fields. The first one describes the arrival of customers, the second the service mechanism, and the third the number of servers (Kendall, 1953). Since then, this notation has been extended to A/S/c/k/QD, further standardizing the description of the queue's capacity and its discipline, respectively (Gautam, 2008). In Table 2, classification examples are presented, introducing notations characterised in the next subsections.

Table 2: Kendall's Notation (adapted from Gautam, 2008)

Field	Meaning	Examples of Possible Classifications
A	Arrival Distribution	Exponential (M), Generic (G), Erlang (Ek), Deterministic (D), *etc.*
S	Service Distribution	Exponential (M), Generic (G), Erlang (Ek), Deterministic (D), *etc*.
c	Number of servers	1, 2, …, ∞
k	Queue Capacity	Default assumption: ∞; it may also be a constant (1, 2, 3, …)
QD	Queue discipline	Default assumption: FIFS; it may also be LIFS, SIRO, SPTD, PD, *etc.*

4.1.2 The Input Source

This subsection will characterize the input source of a queuing system. This component describes the arrival of the customers in terms of both *size* and *interarrival time*. On both components, assumptions may be made to ease its implementation and the solution-seeking, as previously explained (*cf.* Section 4.1).

4.1.2.1 The Size of the Calling Population

The calling population may be assumed as being either finite, or infinite. Since calculations are far easier, finite populations are often assumed to be infinite whenever their size is considerable (Jaiswal, 1968; Strang, 2012). Otherwise, if it is considered to be finite, the number of remaining customers in the source will influence the number of arrivals to the system, as the source will become gradually emptier. Fortunately, several authors have proposed models to handle this correlation. Pardo & De la Fuente (2008), for instance, highlight some applications for finite-population models and propose solutions to mitigate the uncertainty of the arrival, based on the work of Sztrik's (2001) and Almási *et al.* (2005).

4.1.2.2 The Interarrival Time

The time between two consecutive arrivals is referred as the interarrival time. Since they are usually considered random due to their Markovian (or no-memory) property proved by Gross & Harris (1998), interarrival times will often be described by an exponential distribution in infinite sources (Pardo & De la Fuente, 2008). However, in finite populations, more complex considerations must be made, since, as explained before, future arrivals will be influenced by previous entrances.

Considering that the system in-hands is infinite (*cf.* subsection 4.1.4), further research was conducted, showing that, since the system witnesses many independent entrances, they may be modelled by a Poisson's distribution, given that their sum will tend to a Poisson process (Khinchin *et al.*, 2013). Further confirming that hypothesis, Lariviere & Van Mieghem (2004) show that Poisson's distribution will efficiently model service systems where the population is large and analysed time horizon is long. In these conditions and knowing that the interarrival time of a Poisson process may be modelled by an exponential distribution, it becomes easier to model it having limited impact on the model's precision (Creemers & Lambrecht, 2010).

4.1.3 The Queuing System

In this subsection, the queuing system (henceforth referred as system) will be characterized hence completing the characterization of the model represented in Figure 10.

The queuing system models the entire customer journey in-system (*i.e.,* from his entrance to his exit). This journey may consist in several interactions with the service mechanism and may include queues, whenever the service-rate is lower than the arrival-rate (Baumann & Sandmann, 2017).

Observing Figure 11, a queuing system characterization may be summarized resorting to two main attributes: the number of *queues* (vertical axis), and the composition of the *service mechanism* (horizontal axis) (Banks *et al.*, 2005). Describing the four presented quadrants and analysing their impact on the total length of the queue, Joel *et al.* (2000), Ndukwe *et al.* (2011) and Maister (1984) have further developed these two to a total of five dimensions that will be described in the next subsections.

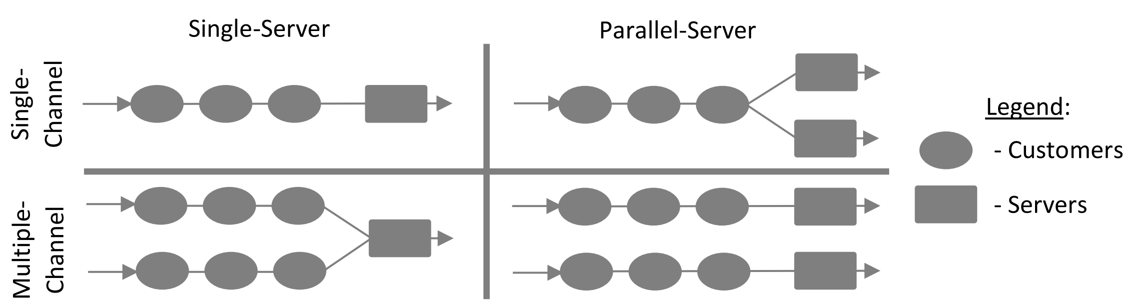

Figure 11: Representation of different Queuing Systems and their classification under the adopted terminology

4.1.3.1 The Layout of the Queue

Queues may be classified concerning its capacity and its number. As explained by Argon *et al.* (2016) the former will have a crucial impact in the system, since customers will be lost due to a certain length of the queue(s), if finite-queues are considered. Regarding the latter dimension, as depicted in Figure 11, one may define a queue as having one or multiple channels (Joel *et al.*, 2000). Even though this choice is often linked with the number of servers (*i.e.,* a single line, if there is a single server), one may adopt a single line even under a parallel line system to mitigate, for example, the *jockey effect* (*cf.* next subsection).

4.1.3.2 Customers' Behaviour in the Queue

While they are waiting, the customers' behaviour can also be described. The most common assumption is that the customer is *patient* and, thus, he will wait for his turn. Alas, most customers are impatient according to Maister (1984) and, whenever this is the case, the probability of losing him due to systems' inefficiencies is higher (Al Hamadi *et al.*, 2015). Some phenomena include *balking*, *reneging*, and *jockey effects*. The two first ones were analysed by Yue *et al.* (2006), and consider the possibility of a customer refusing to remain in the system, whenever either the queue or the waiting period is too long, respectively. Whereas *jockey effect*, presented by Zhao & Grassmann (1995) and modelled by Tarabia (2008) and Sakuma (2010), embraces the possibility of a customer being constantly shifting queues to join the shortest line.

4.1.3.3 The Queue Discipline

Concerning the order in which customers are called (*i.e.,* the queue discipline), several assumptions were found. Most literature assumes a first-in-first-out (FIFO) policy, since it is the easiest to model. Nevertheless, it will often fail to minimize the customers' welfare loss resultant from the queue's existence (Platz & Østerdal, 2017). Therefore, many other rules were embraced by the literature:

- The retrial discipline considers a model where the queue is never physically formed, as customers will randomly repeat their arrival until they find an available server (Morozov & Phung-Duc, 2017);
- Service in Random Order (SIRO) is suggested to enhance welfare in concave waiting-cost curves, by the random selection of a customer in line in Caulkins' paper (2010);
- Angel *et al.* (2008) propose to iteratively call the customer who has the shortest processing time (SPT) at each instant, reducing the overall customers' permanence in system;
- Last-in-first-served (LIFS) determines that the last customer entering the system is the first one to be called. It has been shown by Breinbjerg *et al.* (2016) that LIFS incentivizes customers to arrive more smoothly and, thus, mitigate the existence of peaks in demand;
- Finally, the Priority Discipline (PD) considers that each customer has a different priority-level (PL), according to its intents, or characteristics. Following this rule, those who are classified with the highest PL are called first, and only when the entire level is cleared will the lower PLs be called (Gómez-Corral, 2002). Furthermore, and particularly common in hospitals, the model may be sub-classified as being preemptive. If so, a server must interrupt its service to the current customer, whenever a higher PL enters the system, resuming, afterwards, the initial service (Brandwajn & Begin, 2017).

4.1.3.4 System Arrangement of the Service Mechanism

Regarding its arrangement, there are two dimensions to be considered: the *arrangement of service facilities* and the *number of service channels*. On the one hand, the former distinguishes the systems where the customer is gradually served across a sequence of service stations (*e.g.,* school canteens with facilities in series), from those where the customer is completely served at a single station. On the other hand, the latter (*cf.* Figure 11) distinguishes systems with a single server at each level, from those that have parallel servers (*e.g.,* high-way tolls). Any combination among these two dimensions may be made, influencing not only the system's, but often the queues' layout as well (Joel *et al.*, 2000).

4.1.3.5 Service-Times of the Service Mechanism

The Service Mechanism is also characterized according to the duration of the provided services (the *service-time*). This may be either similar, or distinct, among different tasks and servers (Nosek & Wilson, 2001). Moreover, as pointed out by Argon *et al.* (2016), whenever services are dependent on each other (*e.g.,* service facilities in series), service-times should include the travel-time between stations, since clients must travel along the service chain to be served, and they may be impacted by different cadencies (or service rates) along the chain, since a client that has already been served by the first server, may have to wait until the second station is ready to call him (*i.e.,* the latter has an higher service-time than the former).

4.1.4 The impact of Queuing Theory

Thanks to the previously presented research, where the several facets of a queuing system were characterized, it became clear that the described problem in the last Chapter can be modelled. Indeed, the queuing system of EDP Stores may be characterized as:

- having an infinite population, as it does not seem plausible to consider the impact of one customer's entrance on the size of the remaining population. Thus, the interarrival time may be described by an Exponential Distribution;
- having an infinite queue, since there is no official limit to the number of waiting customers. However, in future works, it would be interesting to further investigate the queue length impact on the customer' willingness to remain in the store, since the customer may refuse to stay in the system whenever the queue is too long;
- being visited by patient customers whose call is processed following a non-preemptive Priority Discipline, since calling is officially ruled by Inline®'s system and different priority-levels are assigned to each ticket-type;
- having a Single-channel and a Parallel-server system, since every store has at least 2 service stations, but only one queue in store that is managed by Inline® (*c.f.* Section 3.2).

Deep-diving on relatable literature, some papers were found studying similar systems. Xu *et al.* (2016) and Shi (2016) analyse the impact of different queue priorities on a firm's revenue, as a higher priority level may incite the customer to buy either at a higher price, or in bigger quantities, ultimately becoming loyal and returning to the store. There is also a considerable amount of research about queuing models in Health-care, presenting either optimization models, or better priority-allocation, that enhance the overall wellbeing of their patients that may be considered in the system in-hands (Cayirli & Veral, 2003; Lakshmi & Iyer, 2013).

Still, the most relatable research-area found was the telephone call-centres. With similar characteristics, the high-dimensionality that a M/M/c/k/PD model (the same as EDP Stores) imposes on the Markov-chain (MC) due to the several (*m*) priority levels, hampers the solvability of these models. Harchol-Balter *et al.* (2005) explain that, when handling the different states, the MC will become infinite in all the *m* dimensions, which is largely intractable. As Brandwajn & Begin (2017) state, in their paper that models preemptive PD, Brown *et al.* (2005) were able to derive exact solutions to a M/M/c queue, by assuming that servers had exponential service-times that were identical among different PLs. Three recent studies have tried to lift at least one of these three assumptions at a time. Firstly, Al Hanbali *et al.* (2015) modelled a non-exponential model (M/G/c) with K non-preemptive PLs, but still assuming identical service-times. Secondly, Wang *et al.* (2015) handled different service-times among servers, but with an exponential distributed model (M/M/c). Finally, Harchol-Balter *et al.* (2005) modelled a non-exponential service-time with non-identical service-rates, but had to interactively reduce the MC to one dimension, which incurred in small inaccuracy rates, otherwise avoidable, if a Simulation technique had been applied instead.

At this point, and with the queuing system fully characterised, the motivation to proceed by analysing Simulation arose, since it became clear that the queue in-hands (M/M/C/k/PD) quickly becomes unmanageable when different priority levels are considered (Brandwajn & Begin, 2017). Clearly, Harchol-Balter *et al.* (2005) suggestion of proceeding with simulation had to be considered.

4.2 Simulation – Simulating the problem

This section aims at presenting Simulation as a suitable choice to model and solve the problem in-hands, now that Queuing Theory has successfully characterised the system, but has failed to model it in a tractable manner. Once modelled, improvement opportunities may be inferred and implemented to increase the overall performance of EDP's stores system.

4.2.1 Motivation to Simulate

As explained in the last section, Simulation was referred as being a fitting choice to model the queue of EDP stores. In fact, simulation serves many purposes and has been used to help many companies improving their decisions after applying queuing theory (Nosek & Wilson, 2001). Examples include using Simulation to: get a better understanding of the system; compare different techniques, or system designs; assist in improving systems efficiency by solving their problems; and positively influencing the agents' behaviours by showing how an altered system may behave (Bhattacharjee & Ray, 2014).

According to Pidd (2004), modelling systems have become a trend, simply because they help decision-makers to easily explore a set of different feasible options whenever there are several solutions for the same problem. Even though, models may be also studied through direct experimentation, or through mathematical modelling, the benefits of simulation appear to outweigh both. Firstly, it has a lower implementation cost than the former, since it takes less time to study different options, and it avoids the need of repeating the same experience, often with different-skilled analysts, to be able to infer data. Secondly, it presents a less challenging solution to design dynamic systems than the latter option, by overcoming the problems identified in the last section. Overall, the use of simulation proves to be appropriate in this Dissertation, being its research further developed in the next sections. Still, the disadvantages summarized by Banks *et al.* (2005) will be kept in mind, mitigating them whenever it is possible by adopting a robust methodology. Indeed, simulation requires special training and careful interpretation of the results. Moreover, as there are many ways to model the same system, the most realistic one should be built.

4.2.2 Discrete Event Simulation

Several simulation approaches may be chosen. In this Dissertation, Discrete Event Simulation (DES) is proposed to be a good fit, as the queue system in-hands (Banks *et al.*, 2005; Tako & Robinson, 2010):

- is as a Discrete System, having state variables (*e.g.,* number of clients in-system) that change at discrete moments, opposed to a Continuous System where state variables change continuously (*e.g.*, simulating the evaporated percentage of water when it is being boiled);
- represents a dynamic (not static) system that changes through time;
- is stochastic, in opposition to deterministic, containing random variables, such as the service-times;
- comprises concrete operational problems that may be analysed quantitatively.

Thus, the following subsections will focus on researching articles that may help to simulate EDP stores with a DES model. All the findings will be then developed in Chapter 5, where the proposed Methodology will implement a Discrete Event Simulation considering every mentioned assumption so far.

4.2.2.1 Description of the Model

Indeed, DES can cope with different constraints, and still handle their implications in detail (Sachidananda *et al.*, 2016). As complex as it can be, fortunately, most of the structural components of a discrete system have already been described in Section 4.1. Therefore, its brief presentation will take place next.

Simulation mimics real environments thanks to two main components: the dynamic inputs, called *entities*, and the system's *resources*. While *entities* represent units of traffic (*e.g.,* clients, or information such as e-mails), *resources* simulate anything that is capacity-constrained (*e.g.,* machines, or workers) considering, for instance, their setup times, different skill-levels, and even shift patterns (Sachidananda *et al.*, 2016).

Entities have individual *attributes* that are crucial to measure the performance of the simulated system, since they record every interaction between *entities* and *activities* (*i.e.,* a set of actions they may undertake). These interactions will result in *events* that alter the system's *state* by changing the defined *global variable*. For instance, if the considered *global variable* (quantifying a specific dimension of the problem) is the total number of customers in-system, every event (in this case, new entrances, or exits) will change the system's state, generating a new simulation output. Moreover, since more than one *event* is usually planned, an *event list* schedules every *event* and its concrete time of occurrence for each simulation (White & Ingalls, 2015).

Throughout its journey, an entity may be in five different states. Generally, it is *active* whenever it is moving through the system, *conditioned-delayed* when it is on a queue, *ready* when it is available to be processed, and *time-delayed* when it is being served for a certain service-time. It can also be in a *dormant* state whenever it is not constrained neither by time, nor a condition, becoming *ready*, by definition, at a specific moment of the simulated time (Rossetti, 2015).

Convening both paragraphs, it shall be clear that an *entity* is assumed active until it reaches an *activity*, and that this interaction will ultimately change both the *entity's* and the *system's* states. Still to be characterised, *activities* have three main categories: queues, delays, and logic (Schriber *et al.*, 2012). The first one includes activities that place the entity in *ready*, or *conditioned-delayed* states. The second one represents all activities where entities are being served, and therefore on a *time-delayed* state, until their movement is resumed. And lastly, the third one models logic conditions that allow *dormant* entities to generate specific events (*e.g.,* skip the line, or resume its movement) at a given instant, or condition.

After understanding the general components of a discrete model, the next two subsections will detail two specific dimensions that must be considered whenever the environment to model in a discrete event simulation is stochastic (subsection 4.2.2.2) and dynamic (subsection 4.2.2.3).

4.2.2.2 Handling the randomness of the system

Whenever a DES model aims at replicating a real system, inputs will have some associated random variability, such as the exogenous variable *number of arrivals*, or even their *nature of inquiry*. In fact, being the most common situation in simulation, it is crucial to understand how to handle randomness. Nova (2008) states that, even though, ideally, statistical inference methods would be applied to fit the sample-data to a statistical model, simulation does not require that. In fact, thanks to Monte Carlo techniques simulation handles this uncertainty by generating artificial inputs that will respect the original characteristics of the statistical model (Rubinstein & Kroese, 2016). This enables the existence of a virtual environment that mirrors real stochastic events, as long as the software is able to generate *pseudo-random* numbers that verify the two properties of random numbers: *uniform distribution* and *independence* (Banks *et al.*, 2005).

4.2.2.3 Simulation Clock and Stopping Condition

Moreover, being dynamic, a DES (or, in this case, DEDS) model must have a time-keeping mechanism, the *simulation clock*, to keep track of the simulation timeline and to be able to pin planned future events on specific *clock times* (Rubinstein & Kroese, 2016).

Thanks to this mechanism, the execution of a DEDS may be conducted in a two-phases loop (the EMP and the CUP phase). Firstly, it processes every scheduled event for the current *clock time*, updating the *system state* and updating the *future event list* by loading(/deleting) entries to(/from) the event list. And secondly, the *clock*-time is skipped to the next scheduled event (Schriber *et al.*, 2012). This loop will continue until a *stopping condition* is met, being either the inexistence of active entities in-system (White & Ingalls, 2015); a specific time condition, such as the store closing-time (Daellenbach *et al.*, 2012); a randomly generated probability (Hernández, 2001); or a certain number of loops (Masud & Ravindran, 2008).

4.2.2.4 Verification of model assumptions

After understanding how a DES model works, it is also important to briefly present how to interpret its results. Strang (2012) explains that, as useful as simulation is, allowing to capture and manipulate variables, otherwise impossible to do without disrupting production, its outcomes may lead to misleading conclusions. Thus, it is crucial to verify the model's assumptions before and after the simulation, since, otherwise, the simulation outcome may be unreproducible, or unrealistic, whenever data has not been properly sampled to support the chosen theoretical assumptions (Strang, 2012).

Therefore, to foster the implementability of the simulated solution, the next section will present Lean since it has shown to reduce the lack of human commitment by fostering a better process of change and best-practices identification. Moreover, a careful data analysis will be planned in Methodology (Chapter 5) to guarantee that the input data will support every assumption made throughout the Dissertation.

4.3 Lean Philosophy – Implementing the solution

In this section, the fundamentals of Lean and some of its techniques will be presented to achieve two goals. Firstly, it will be explained how this Japanese-born philosophy differs from the Western improvement process together with some of the concepts that must be applied throughout a Lean methodology. Secondly, an assessment of Lean's impact on a queuing system will be conducted in two terms: implementability of solutions and best-practices identification. Analysing both, a special attention to supporting the employees and eliminating wasteful activities will be noticed (Beckers *et al.*, 2016).

4.3.1 Introduction of Lean

Ever since Toyota Motor became the top automotive manufacturer in the world, organizations have followed its example by fostering the adoption of Lean's philosophy and by advocating its concept of Kaizen (改善), a Japanese word that means "continuous improvement".

Lean contrasts with the Western culture of disruptive innovation, since it strives to promote improvement initiatives that entail relatively small, incremental, and low-risk investments. Innovation, on the other hand, often involves new disruptive technologies that require higher financial investments and higher attention as it may involve a complete redesign of the *status-quo*. Still, despite entailing small investments, Lean has proved to provide impactful results by increasing the overall efficiency of a process by two percent digits (Smith *et al.*, 2012); or by even improving an entire system by simultaneously reducing costs, lead-times and materials consumption (Rivera & Chen, 2007). Moreover, as it is applicable to its entire value stream, it has been sought by many companies (Womack & Jones, 2010). Nevertheless, some concepts must be promoted to successfully implement a Lean environment (Imai, 2012):

Lean and Management

Lean suggests that job functions should be reduced to two main activities: Improvement and Maintenance. The latter will ensure that everyone is performing according to the established standards (SOPs), whereas the former will aim at increasing the current job-standards. As described in Figure 12, everyone must be responsible for continuous improvement (Kaizen), hence guaranteeing that the company status is never taken as settled and that everyone will be constantly looking for improvement opportunities. Innovation, though, will only involve the upper-management and often encompassing long-term measures.

Overall, the goal is to incite everyone to be committed to sustain the current standards, while being simultaneously aware of their surrounding improvement opportunities. Clearly being a crucial concept of Lean, it became KICG's logo (*cf.* Figure 29 in Appendix).

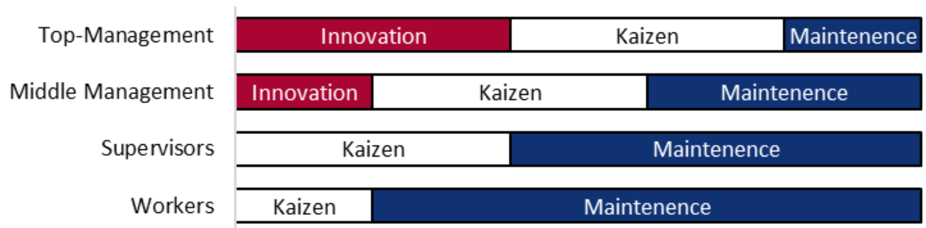

Figure 12: Lean Job perception (adapted from Imai, 2012)

Process *versus* Result

Management teams should become process-orientated, instead of focusing solely on results. Thanks to Lean, robust processes will lead to better results on the long run, since human efforts will be highlighted and management will commit itself to foster an appropriate environment.

Follow the PDCA+SDCA cycles

By implementing a Lean's philosophy, a company should never be satisfied with the *status quo*. Instead, better standards for its processes must be continuously sought (*cf.* Figure 30 in Appendix) and whenever a problem is identified, efforts should be focused on reducing the process variability first through Standardization (SDCA). After mitigating the variability, if the process outcome is still unsatisfactory, then the organization should implement a PDCA cycle, following these stages: (1) Plan, setting new targets and planning the set of actions that will promote the accomplishment of those goals; (2) Do, implementing that plan; (3) Check, inferring if the implementation of the plan is on track; and, finally, (4) Act, by spreading the new procedure and creating a new SOP, if the plan proved to be successful.

Putting Quality first

Whenever companies are analysing the existing trade-off among quality, cost and delivery (QCD), they often tempt themselves to compromise quality levels to, either deliver products on-time, or to meet cost reduction goals. According to Lean's philosophy, this will shorten the business life and, thus, managers should commit themselves to constantly prioritize quality.

Control to Act

Lean proposes the adoption of a problem-solving mind-set. Thus, it must be ensured that the data of a certain problem is objective, prior to tackling it. This will avoid the implementation of measures based on subjective reasoning. Instead, robust solutions will arise after consolidating, verifying and analysing appropriate parameters (KPIs).

Internal Customer

Following a Lean methodology, every activity included in a process is regarded as a customer on its own (an *internal customer*). Therefore, it must be assured that, neither defects, nor mistakes, are passed to the next work stage through a swiftly identification and rejection of any error. In fact, Lean advocates presenting only high-quality products, or services, to the final external customer.

Overall, Lean strives to foster robust processes among everyone in the organization and that, what may be perceived as short-term costs, will actually be long-term investments, since these initiatives will continuously raise the organization's standards and quality levels.

4.3.2 Application of Lean

To accomplish a sustained improvement of the organization, the previously mentioned concepts rely on the following six main transversal principles of Lean (KICG, 2017c).

4.3.2.1 Create Customer Value

Lean defends that the primary goal of a company should be satisfying its customers by maximizing the utility of their products and striving for their lowest possible price. To that end, a company should begin by understanding its customers' needs, and tailoring its offer towards their targeted market. An effective way to achieve it is to adopt a pull production system, since, as explained by Sundar *et al.* (2014), when implemented, the company's production will be solely based on the customer requirements. In this scenario, a downstream flow of information will be generated, since the client's needs will be transmitted iteratively to the closest internal supplier from the source (*i.e.*, whoever is in contact with the client). When this feedback reaches the last link of the chain, the inception of a product that is not only brand-new but also tailored begins, thus creating an upstream production flow. Finally, and after tailoring the product, it is suggested to also tailor both the quality levels and the delivering times to the needs of the client. This will further increase the product's value and continuously reduce the non-valuable activities in the operation.

4.3.2.2 Eliminate MUDA, MURA and MURI

To serve its clients, a company must conduct a set of operations. Lean proposes to begin by classifying each one of their activities as being either value-adding (VA), or non-value adding (NVA), whenever they are assessed to increase, or not the company's products value, respectively (Mostafa *et al.*, 2015).

Whenever they are assessed as NVA, they are *Muda* (wasteful in Japanese). Thus, efforts should be taken to reduce them, as this will decrease the overall cost and increase the overall efficiency of the process. Together with *Muda, Mura* (variability) must be also mitigated according to Wang & Disney (2016), since the unevenness of demand will become the root cause of both *Muda* and *Muri* (overburden of people, or equipment), by generating the need for inventory and an excessive strain on the process (bullwhip effect). To further acquaint the reader with these concepts, the 7 types of *Muda* are presented next (Melton, 2005):

- Excessive Production: considered the most impactful type of waste, producing more than what is demanded by the customers will aggravate the effect of all the following enumerated wastes;
- Materials Transportation: increasing employees' and MHE occupation, it may aggravate the lead-time (time elapsed from the start to the end of a process), the risk of an accident, and the labour costs;
- Inventory: increasing the need for warehousing solutions, it will raise capital costs. Furthermore, by temporarily concealing critical problems due to the existence of buffers (such as preventing the timely detection of a possible deterioration of stocks, or even upstream operational problems), it may even disrupt the supply chain of which the company belongs to;
- Motion: by requiring people to move to perform their task, productive time will be wasted;
- Waiting: mostly due to set-up times and queues, people and equipment become idle, adding no value;
- Over Processing: usually caused by a misperception of the client's needs, excessive work may be conducted on a certain product even though it will not add any value, as the customer was already satisfied by the original version;
- Defects and Errors: requiring re-work and leading to complaints, whenever errors are not fixed before they reach the final customer, they decrease the overall process efficiency and its service-level.

4.3.2.3 Value Stream Mapping (VSM)

An easy way to identify the three previously mentioned concepts is by assessing the current state of the system. Throughout a VSM, every process will be mapped and each critical (*i.e.,* with *Mura* or *Muri*) and NVA activity will be identified. Thus, the goal of this technique is to design a leaner system that, ideally, only includes balanced VA activities, by eliminating the wasteful activities and designing SOPs that are spread among employees, hence raising the overall standard of work (Piercy & Rich, 2009). Lean encourages companies to resort to a VSM whenever they are tempted to increase their efficiency solely through automation. Unfortunately, this solution will only further conceal the real inefficiencies of the system, since that course of action simply automates the process, without first assessing the worthiness of the convened activities (Bortolotti & Romano, 2012).

4.3.2.4 Gemba Effectiveness

To easily identify the *Muda*, according to Bicheno & Holweg (2000) Lean also aims at implementing a culture where decisions are taken based on objective and observable data. Thus, decision-makers are invited to go to the *Gemba* (the place where the work is conducted) and to assess the situation together with their team, instead of just reading the reports from their offices. By immersing themselves in the *Gemba*, not only new improvement opportunities will be unveiled, but also workers will be more motivated and involved. Overall, this will increase efficiency levels and avoid unrealistic solutions.

4.3.2.5 People Involvement

Since the two previously mentioned initiatives will require the involvement of everyone in the organization, it is crucial to foster a “no blaming, nor judging” environment, where by investing on training development programmes, knowledge and participation levels will become balance among team members (Imai, 2012).

4.3.2.6 Visual Management

Furthermore, Lean seeks raising the team awareness about its current performance. To that end, a transparency-policy must be adopted, where every process is mapped and every problem is visually highlighted. Using visual aids, such as graphics, the progression of the Key Performance Indicators (KPIs) is tracked and promptly presented to the team, promoting discussion and brainstorming among team members about how to solve problems. The goal is to enable the team to quickly find solutions, even if only a mitigating short-term measure is found. Moreover, thanks to these visual aids, not only complex issues are properly escalated to the upper-management, but its resolution process is tracked as well (Detty, 2000).

4.3.3 Impact of Lean on queuing models

After understanding the main principles of Lean, it was also important to link it to queuing systems. In fact, the queuing effect is often mentioned on Lean literature. Whenever material, or information, flows are designed, frameworks try to mitigate queues by tackling constraining points in the process (*bottlenecks*) (Sundar *et al.*, 2014) and to balance the demand with the utilization rate of each server, avoiding the need for extra-hours and fostering the standardization of work (Liker & Morgan, 2006).

Examples include: Chiang & Urban (2006) who have balanced productivity along a production line to produce less work-in-progress (WIP); the work of Pourvaziri & Pierreval (2017) that models a manufacturing system and simulates their Lean solutions for a dynamic facility layout, improving the system’s productivity and reducing the WIP; and, finally, the paper of Omogbai & Salonitis (2016) where it is suggested to rely on DES, whenever testing Lean hypothesis and their forecasted impacts.

Lean’s role on reducing the variability and standardizing work is also mentioned, together with its philosophy of fostering better communication channels and connecting managers and employees. Studies like Baril *et al.* (2016) point out its crucial impact on a project’s implementability. Being able to reduce the existing internal barriers towards change by ensuring that every team-member is part of the solution-seeking process in transversal team-meetings and that he understands both the main steps and the benefits of the project, employees will be more likely to embrace, support and implement the uncovered solutions.

4.4 Chapter's Conclusions

Throughout this Chapter, an extensive literature review was performed on three research areas, achieving its purpose of guaranteeing a holistic understanding of the entire Methodology presented next.

Briefly summarizing it, the Chapter began by presenting the Queuing Theory and helped to characterize the problem in-hands and its system. Supported by the research developed by the mentioned authors, the model components were identified and their attributes were characterised. Throughout this characterization, several papers were found stressing the limited applicability of mathematical modelling for a system like the one in-hands. Therefore, the need for Simulation, an alternative solution to model EDP stores, arose.

After classifying the system, a specific Simulation model – Discrete Event Modelling (DES) – was presented and assessed to be a good fit for this Dissertation. It was also understood that by performing small conceptual adaptations (*e.g.*, adding global variables), the previously proposed characterization of the model could be used in DES models to characterize their components. Moreover, some software requirements, such as the generation of *pseudo-random* numbers, and the need to make some assumptions, if the model were to be finished in a reasonable amount of time, were uncovered.

Finally, it was analysed Lean's impact on mitigating a system's variability and on possible methodologies that would ease the implementation of the uncovered solutions by the methodology. As explained, Lean will lead to the creation of robust processes that reduce both the variability of the processes and the overburden of their constituting elements (*e.g.,* employees) through the adoption of concepts that ultimately strive for a small-paced, but ever-changing environment. Considering these findings, the motivation to standardize the service processes and to implement the VSM technique on store operations arose. This way, not only would they map the current processes highlighting operational improvement opportunities, but they would also set realistic AST targets through service *best-practices* identification. The expected reduction of the service times variability would enhance the possibilities of a better levelling between service capacity and demand, which by itself would minimize the AWT, as proven by Hines *et al.* (2008).

Overall, with this Chapter, the Dissertation fulfilled its commitment of providing the necessary information to avoid a silo-minded analysis, since, throughout this work, more than 100 references (summarized in Table 3) have been mentioned. Furthermore, the three research areas proved to have a symbiotic effect by promoting complementary concepts and techniques. In fact, academia itself has suggested the development of more works that would embrace the benefits of connecting Simulation and Lean McDonald *et al.* (2002).

Table 3: References mentioned during the Literature Review

Subject	<2000	2000-2005	2006-2010	2011-2015	2016-2017
Queuing Theory Review (16 references)	Erlang (1918) Kendall (1953) Mann (1970)	Nosek & Wilson (2001) Nah (2004) Harchol-Balter et al. (2005)	Asmussen (2008) Gautam (2008) Hillier & Lieberman (2010) Foster et al. (2010)	Li & Hensher (2011) Strang (2012) Newell (2013)	Legros (2016) Gutacker (2016) Zehrer & Raich (2016)
Modelling the Input Source (15 references)	Jaiswal (1968) Maister (1984) Gross & Harris (1998)	Joel et al. (2000) Sztrik (2001) Lariviere & Van Mieghem (2004) Almási et al. (2005)	Pardo & De la Fuente (2008) Creemers & Lambrecht (2010)	Ndukwe et al. (2011) Strang (2012) Khinchin et al. (2013)	Baumann & Sandmann (2017)
Modelling the Queue (16 references)	Maister (1984) Zhao & Grassmann (1995)	Joel et al. (2000) Gómez-Corral (2002) Banks et al. (2005)	Yue et al. (2006) Angel et al. (2008) Tarabia (2008) Sakuma (2010) Caulkins (2010)	Al Hamadi et al. (2015)	Argon et al. (2016) Breinbjerg et al. (2016) Brandwajn & Begin (2017) Platz & Østerdal (2017) Morozov & Phung-Duc (2017)
Modelling the Service Mechanism (12 references)		Joel et al. (2000) Nosek & Wilson (2001) Cayirli & Veral (2003) Brown et al. (2005) Harchol-Balter et al. (2005)		Lakshmi & Iyer (2013) Al Hanbali et al. (2015) Wang et al. (2015)	Argon et al. (2016) Shi (2016) Xu et al. (2016) Brandwajn & Begin (2017)
Simulation Review (6 references)		Nosek & Wilson (2001) Pidd (2004) Banks et al. (2005)	Tako & Robinson (2010)	Sargent (2013) Bhattacharjee & Ray (2014)	
DES Model description (8 references)		McGregor & Cain (2004) Balci (2003)	Chiang & Urban (2006) Cardoen et al. (2010)	Schriber et al. (2012) White & Ingalls (2015) Rossetti (2015)	Sachidananda et al. (2016)
Handling the randomness of the system (3 references)		Banks et al. (2005)	Nova (2008)		Rubinstein & Kroese (2016)
DEDS Components and Verification of model assumptions (6 references)			Masud & Ravindran (2008)	Daellenbach et al. (2012) Schriber et al. (2012) Strang (2012) White & Ingalls (2015)	Rubinstein & Kroese (2016)
Lean Review (18 references)	Detty (2000) Bicheno & Holweg (2000)	Toriizuka (2001) Baines et al. (2005) Mason et al. (2005) Melton (2005)	Engström et al. (2006) Rivera & Chen (2007) Montgomery & Woodall (2008) Womack & Jones (2010) Mostafa et al. (2015)	Imai (2012) Smith et al. (2012) Sundar et al. (2014) Ward & Sobek II (2014)	Baril et al. (2016) Beckers et al. (2016) Wang & Disney (2016)
Value Stream Mapping (4 references)		McDonald et al. (2002)	Piercy & Rich (2009) Hines et al. (2008)	Bortolotti & Romano (2012)	
Lean applications on queuing models (5 references)			Chiang & Urban (2006) Liker & Morgan (2006)	Sundar et al. (2014)	Omogbai & Salonitis (2016) Pourvaziri & Pierreval (2017)

Following this Chapter on literature review, the next Chapter will now propose a tailored methodology to implement the so-far convened knowledge on the problem in-hands, followed by the assessment of its overall applicability.

5 Proposed Methodology

Having convened a literature review in the previous Chapter, this Methodology Chapter will focus on developing a framework that merges four complementary frameworks found in the literature. Throughout the process hereafter mapped, the Dissertation goals (*cf.* Section 1.2) will be tackled and accomplished.

Overall, this Methodology proposes developing a tailored framework whose outputs may highlight improvement opportunities in any ticket-based store system. The next two sections will present an explanation of existing frameworks to: reduce the AWT, AST, and their variability following Lean methodologies (section 5.1); and to accurately simulate discrete event systems (section 5.2). They will be followed by a concluding section, where the resultant framework will be ultimately presented and the symbiotic outcomes among the frameworks will be highlighted.

5.1 Designing a Lean System

Mason *et al.* (2005) have shown that servers have different performance levels, due to different skill-sets that are not perfectioned at the same rate, especially when they call different quantities of each service-type, and each type has still several sub-services associated. Nevertheless, following Lean Fundamentals (*cf.* Section 4.3.1), conducting a standardization of the best-practices can minimize both the variability and the average of the service-times.

Towards the goal of reducing the AST, the proposed methodology will begin by identifying the system's best servers. As every interaction between customers and servers is recorded by Inline®, quantitative data can be used to identify servers' best-practices (*cf*. Section 5.2). This implies, though, an objective criterion to determine which processes will become standards. To that end, Engström *et al.* (2006) propose to use the coefficient of variation (CV) to sort different performances and adopt the best (*i.e.,* the lowest) one. The coefficient of variation will measure the dispersion of the recorded performances of employee 'i' on sub-service 'x' by dividing its standard deviation by its mean, as follows in Equation 2:

$$CV_{(i,x)} = \sigma_{(i,x)} / \mu_{(i,x)} \quad (2)$$

For each sub-service, servers will be sorted and ranked by increasing CV. The lower the $CV_{(i,x)}$ is, the more robust will be server 'i' when he performs the sub-service 'x'. Two problems could arise though:

1. Servers with an enduring high AST will be set as example (as their CV will be low, due to a high mean and a low standard deviation);
2. Servers who have rarely served a ticket may also be mistakenly taken as examples (as their CV may be extremely low, if he served once or twice a ticket with a similarly short AST).

To avoid these two scenarios, two qualifying criteria must be set before identifying best-practices. To qualify as a potential best-practitioner, a server must have lower AST than the national average, and have performed more sub-services 'x' than the national average per server. MS Excel will be the chosen software to import the MS Access database, and to filter, sort, and rank the servers by their CV.

Some papers, such as Baines et al. (2005) and Toriizuka (2001), propose further performance shapers besides the server's actual process. In fact, they have mentioned that either the environment, the workspace, or even the machinery can have a crucial impact on servers' performance. However, for the purpose of this work, it shall be assumed that these, or any other conditions, are similar across the different stores during the identification of the best-practices, even though it could be inferred in future works whether there are any potential performance shapers that may stand-out among different stores.

These same Lean techniques may be applied when comparing different stores upon quantitative data. If a group of stores has similar parameters (*e.g.,* historical-demand, or layout), but one of them has a lower AWT, a carefully analysis may unveil best-practices to share across that same cluster.

Together with the adoption of exemplary performances, the system should also be redesigned to comprise fewer non-value-added activities. As proposed by Hines *et al.* (2008), Lean events will be conducted for each service-type and every process will be mapped following the presented VSM technique. Alas, this will not be part of the scope of the Dissertation, since those meetings will take time and, afterwards, upper-management would still need to approve them prior to the implementation. Nevertheless, it must be stated that further reduction of AWT is to be expected, when more efficient and standardized processes are shared among servers. Until now, an initial audit was already performed in May towards this goal, listing every process that is be executed, either in the front-office (32) when serving a client, or in the back-office (18) to guarantee smooth daily operation of the store. Its form may be found in Figure 32 in Appendix.

5.2 Simulating a Discrete Event System

Regarding the implementation of the DES, three symbiotic frameworks will be considered. An initial conceptual phase (steps 1 to 6), based on Sachidananda *et al.* (2016), and a simulation phase (steps 7 to 9), based on Cardoen *et al.* (2010) and Balci (2003).

5.2.1 Conceptual Phase

During the conceptual phase, the entire model development process should be planned in order to accomplish its purpose. Therefore, these next six steps will shape the simulation, guaranteeing that it be conducted in a focused way and avoiding possible deviations from the work objectives.

<u>Scope of the Model</u>

The first step is to define the scope of the model - in this case, the operation of an EDP store. After all, whenever its scope is not properly defined, the model will produce erroneous results. Thus, it is already clear that it will be necessary to characterise this system's components based on data collected by Inline®. Some assumptions shall also be made to tighten the scope of the simulation, as ignoring back-office activities (*i.e.,* only considering the set of activities undertaken to serve customers) and disregarding the special treatment based on the customer's age, or physical condition granted by law decree nr. 58/2016 (*cf.* subsection 3.2.1).

Simulation objectives

Having defined its scope, the objective of the simulation should also be stated. Concretely, the simulation shall address the goals of this Dissertation (*cf.* Section 1.2), by estimating the AWT change with: different ticket allocations; adoption of best-practices; the hiring, or the dismissal, of one employee.

Model components

The third step is to determine the model components and the required assumptions, in order to plan the level of abstraction and the amount of information considered during the simulation. In this case, the arrivals of the entities (*i.e.,* customers) will be based on the 2016's demand, and the layout of the stores will determine the number of servers (*i.e.,* assistants). Any other resource (*e.g.,* machines) will be assumed to be always available to the server. Moreover, any other possible entity, such as corporate emails waiting to be answered, will be ignored, since they are usually served on the back-office and the focus of the store should always be to serve first the customers who have entered that system. Also, as justified on subsection 4.1.4, the queue will be assumed infinite (*i.e.,* there shall not exist a maximum number of people waiting), and, once entering the system, entities will wait until they are called following a non-preemptive Priority Discipline (*i.e.,* customers shall not give up waiting).

As in White & Ingalls (2015), servers will be assumed free to call a new customer after completing the current service. Nevertheless, it will be computed an efficiency ratio to properly handle periods in which the server is free (*i.e.,* not serving a client), but is not available due to a warm-up period (time spent between closing a ticket and calling another, especially after a long service where the server requires a small pause), breaks (*i.e.,* a one hour lunch-break plus a ten minutes break per day), or even wrap-up work (time spent recording all the service details and filing all provided information after having served a client).

Constraints

After modelling the components of the system, their constraints must be considered as well. However, it is acceptable to make some assumptions to ease the implementation of the model, after a considered analysis of each unveiled restriction about whether it is, or not, crucial for the accuracy of the model (Newell, 2013).

Regarding the servers, on the one hand, it will be modelled that a server may only serve one customer at a time; he will only be available for service during his work-schedule; and he will have a daily 1-hour lunch-break, plus small breaks whenever required. On the other hand, yearly vacations will be ignored as the service providers (*cf.* subsection 3.2.1) guarantee that every service station has an assistant throughout the year, even if that means hiring a temporary assistant. Furthermore, as the skills level of the server will clearly impact its service-time, service-times must be characterized by a probability distribution based on historical data, or based on the standardization of identifiable best-practices.

Regarding the clients, it must be guaranteed that every customer is planned to be served. This translates in having every ticket-type assigned to at least one service-station, no matter the given level of priority.

Input data

The data that is introduced into the model will determine its accuracy on the long-term, as it will model how each one of the previously described components behaves under the modelled constraints. Therefore, a rigorous data collection and analysis must be performed, before importing it into the model. In this case, this must be performed by analysing the Inline® database and conducting interviews with store managers to describe: the arrivals of each entity type; number of available servers in each store; servers' availability (*i.e.*, considering breaks); AST per service type and per assistant; implemented queue discipline.

Definition of the KPIs

The last component of the model to plan is the KPIs, that will be used to compare the performance of different models. In this case, this simulation of EDP stores will prove valuable, if it is able to improve both the AWT and the Quality Indicator in each EDP store (*cf.* subsection 3.2.1). It is important to stress, though, that the servers' utilization rate will not be a performance indicator. As in Chiang & Urban (2006), this model will only focus on the improvement of the flow, thus focusing on the entity's point-of-view, and not on maximizing a certain server's number of calls.

5.2.2 Simulation Phase

Following the conceptual phase, the model will be implemented and tested. By following what has been planned and defined, this combination of frameworks may systematically analyse any system. Thanks to an earlier characterization of the system, this focused implementation will have as main target the fulfilment of all the 6 previously mentioned steps and the identification of improving scenarios.

Computational implementation

To build and implement the model Simul8® has been chosen. According to Simul8 Corporation (2017) its discrete event simulator is used by a wide range of business sectors who seek an affordable, but high-quality software to help their decision making. Since the software can be linked to MS Office, data may be easily imported between software. Moreover, this software presents an environment focused on building blocks. Thus, it presents a user-friendly interface, which is important as this tool might also be used by store managers in the future to proactively seek new improvement opportunities. To do so, the user is only required to characterize his current state (*e.g.,* arrival and AST distributions; queue discipline) and run different scenarios (McGregor & Cain, 2004).

Model Validation

After implementing the model, it is crucial to verify and validate that it successfully mimics the real environment (*cf.* subsection 4.2.2.4). This will avoid a scenario where a solution is mistakenly taken as appropriate for a system, estimating a new performance that will not be attainable given the system in-hands.

Therefore, to consider a system valid, the characteristics and assumptions that have been listed in the conceptual phase must be verified. This implies individually testing every component of the model, and checking if the desired behaviour is witnessed. Whenever an error is detected, the model must be corrected (Balci, 2003). Then, when all the components are verified, the simulation itself must possess a satisfactory range of accuracy within its domain of applicability to be valid. During the early stages of the implementation phase, this range may be large, but throughout the experimental period higher accuracy levels should be targeted in order to increase the confidence in the model (*i.e.,* that the simulated performances will be witnessed when implemented in the real system). Nevertheless, it is important to keep in mind that while the value of the model for the user will increase as this confidence level increases, its cost of implementation will increase as well (cf. Figure 13). Thus, it is important to define a satisfactory accuracy target (*e.g.,* 95%) from the beginning, as the end-user (EDP) will always desire the highest value possible (Sargent, 2013).

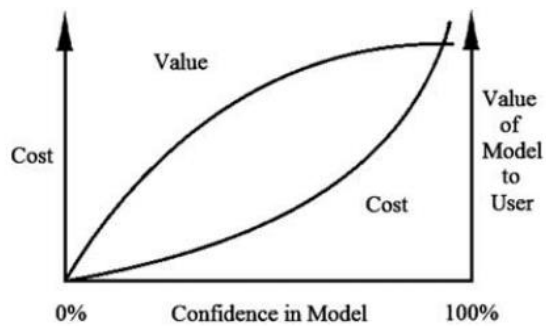

Figure 13: Impact of the Confidence Level on the Model's Cost and Value (Sargent, 2013)

<u>Assessment of the overall performance</u>

The last step of the proposed framework, consists in assessing the simulated performance and comparing it with the current state of the system. Since the model is already verified and validated, solutions may be considered as applicable and KPIs may be compared to those of the newly-proposed scenarios. Whenever the KPIs record better performances, the new scenario shall be assumed to improve the system and have its implementation recommended (Cardoen *et al.*, 2010).

5.3 Chapter's Conclusions

This Chapter has presented existing methodologies found in the literature. The first section was devoted to reviewing methodologies that can identify best-practices. Moreover, a framework to raise the system's efficiency through a VSM has been proposed, even though it will not be part of the scope of the Dissertation, as it will require more time to implement than the available time-span for this work. The second section described a nine-steps framework that was built based on three authors. Every step was explained, highlighting both the work-stages and the assumptions that it will comprise throughout the Dissertation.

This Methodology aimed at designing a tailored framework that could be applied in either this case study, or in similar systems. The resulting four-stages framework convenes methodologies whose both theoretical foundations and processes have been presented throughout this Chapter. As presented in Figure 14, these sequential stages have the main goal of assessing the overall impact that unveiled best-practices may have on the system's performance (*e.g.,* analysing its KPIs variation). This goal is accomplished through a robust process where every aspect of the real system is addressed, characterized and tested, hence leading to a realistic set of solutions from which one, or more, may be proposed for implementation.

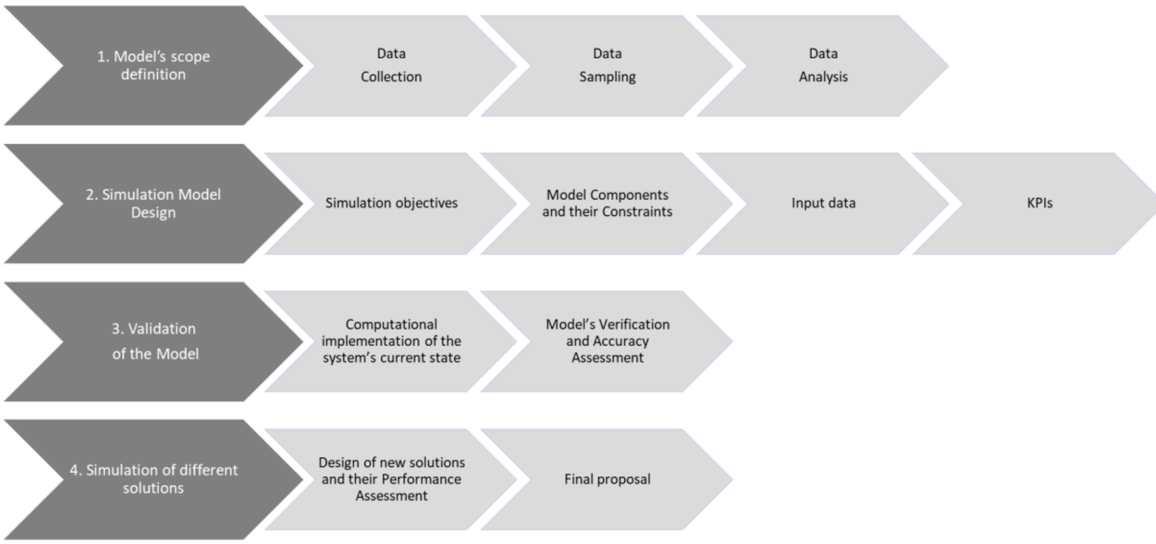

Figure 14: Proposed Methodology - A four-stages framework

Since the framework was primarily designed to be applied to systems that are composed by several business-units (*e.g.,* EDP stores), it begins by identifying the best-practices among each previously defined cluster. For that purpose, an initial collection of data is conducted in order to identify which are the components of the system and the key differentiators among units. By analysing a database such as Inline®, this can be performed enabling the clustering of data to gather similar and comparable units. These two steps will avoid the comparison of performances among distinct realities, and will find benchmarks (*i.e.,* standards of excellence) among each cluster after sorting it by a specific KPI.

From the moment when a benchmark is identified within a cluster, the exemplar performance may be set as a target to any unit in the cluster, after further analysing the data and proving its comparability through both quantitative and qualitative data. Once that is proven, best-practices impact on the system's KPIs may be simulated in a model mimicking that specific unit (*i.e.,* designed with its components and constraints). To validate the model and its forecasted performance, the simulation first run an as-is scenario to not only verify the model's components behaviour, but also to compute its accuracy range. If proven not to be satisfactory, the simulation should be then adjusted until the real behaviour is simulated and a desired accuracy is reached, since they are both crucial factors when advocating a solution. Once the model behaviour is verified and validated, new scenarios based on either best-practices, suggestions, or even new business concepts may be finally implemented and simulated in order to assess the estimated impact of these changes on that specific unit's performance.

The ensuing Chapter will implement this framework in a specific EDP store unveiling potential solutions that, as it will be explained next, will decrease the system's performance (AWT) by up to 76%.

6 Results and Discussion

In this Chapter, the previously proposed methodology will be tested in EDP stores to assess its impact. As depicted in Figure 14, the Framework will seek to improve this system based on data analysis and the adoption of BPs. To prove this concept, its implementation will focus on two comparable stores with diametric performances that will be identified by collecting and sampling Inline®'s data. Further analysing their data, their similarity will be confirmed and their components will be characterized. After assessing the validity of the model, the worse-performing store will be simulated in new scenarios to estimate the potential impact of the unveiled solutions (ultimately resulting, or not, in their implementation in the real system). For confidentiality purposes, all the ensuing data was multiplied by a distortion coefficient.

6.1 Model's scope definition

To better understand the system in-hands and to define a proper scope for this concept-proving implementation, a careful collection, sampling and analysis of EDP's stores data will be conducted in this section. This process will result on the identification of two comparable stores with opposite performances in which the designed framework will be implemented highlighting its potential benefits.

6.1.1 Data Collection

To understand EDP's system, the entire 2016's database in Inline® was imported to a MS Access file. This unprocessed file not only contained every client journey in-stores of 2016, but also the information records of: stores layout; possible combinations of provided services and subservices; and EDP's servers.

To collect workable data from the four previously mentioned tables, they were merged after identifying their key field (a structural component of a database table that guarantees the uniqueness of each entry). In this case, the key field was the *ID_Record* as a unique ticket code is attributed to each visitor. This finding was expected, considering that the system aims at recording a client's journey, hence revolving every recorded data around each specific ticket (*cf.* subsection 3.2.2).

Besides merging the data, a posterior data treatment was also performed, since there were 36 fields (out of 48) in the database that contained fallacious records (*i.e.,* random sample data) representing non-included functionalities in EDP's acquired version of Inline® (*cf.* Subsection 3.2.2). Overall, it resulted in a table with 12 renamed fields, where each one of its lines corresponded to the journey's record of a EDP store's visitor in 2016. This table convened all the required data to fully analyse and understand the system (*cf.* Figure 15) characterizing the store's: demand; layout; AST; and its current AWT.

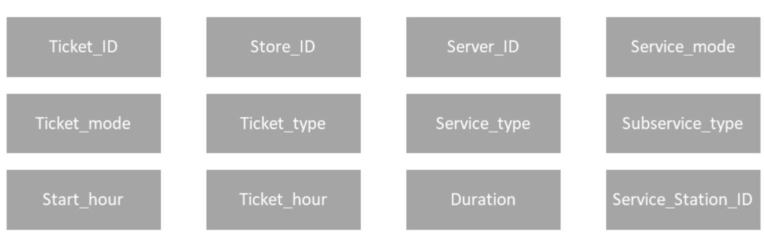

Figure 15: Collected Fields from Inline® system

Furthermore, the client's journey (*cf*. Figure 7) would also be fully characterized. Indeed, as a customer arrives to an EDP Store (*Store_ID*), he prints a ticket (*Ticket_ID*) choosing the nature of his inquiry (*Ticket_type* - numeric field: 0,1, *etc.*; and *Service_Type* - string field: 'A','B', *etc.*). A time-stamp (*Ticket_Hour*) registers that moment on the ticket with both the date and the time. Since he will wait until he is called (*Start_hour*), his waiting-time may be computed by subtracting these two date/time parameters. Furthermore, it is registered the timespan between being called to the counter/table (*Service_Station_ID*) and having been fully served. The field *Duration* includes both. Upon closing the client's ticket, the server (*Server_ID*) closes the ticket, validates the service type (*Service_type*) and classifies the subservice (*Subservice_type*). Automatically, the system will characterise that ticket as: having been served (*Service_mode*), differentiating it from no-shows and failed services; and as having been a printed ticket (*Ticket_mode*), differentiating it from the virtual 'W' tickets.

6.1.2 Data Sampling

At this point, due to the size of the system in-hands, after collecting data, records from over 2 million visitors and 41 stores were to be analysed. Therefore, to enable the implementation of the framework in time, the scope had to be restricted by focusing the analysis on two specific stores and assessing the potential impact of this framework in a specific environment.

With the goal of improving the system, Inline®'s database of served clients with a printed a ticket was exported to MS Excel and consolidated by store. This way both no-shows and 'W' tickets would be disregarded as they are not monitored by ERSE's KPI (QI). Every field was gathered according to the type of information that was sought (*e.g.,* a count of Ticket_IDs would convene the store's demand, and an average of each client's waiting-time would reveal the store's AWT).

Two main take-aways were taken from this step. Firstly, the average demand of the stores presented in Figure 6 was characterized. EDP stores average demand in 2016 was of 51 572 clients with a standard deviation of 21 897 (henceforth referred as σ_{EDP}). Secondly, and following Equation 1, EDP's stores were ranked by QI as shown in Table 4 where only an excerpt is presented for confidentiality purposes.

Table 4: Excerpt of EDP stores and their two main parameters sorted by QI

Rank	Store	Demand (2016)	Number Servers	Number of clients with a WT>20 min.	AWT (min.)	QI
1st	Guarda	37 472	5	538	03:41	98.56%
2nd	Loures	50 138	9	925	03:24	98.15%
3rd	Figueira da Foz	36 534	5	955	03:27	97.39%
			...			
39th	Bragança	34 180	5	3 734	08:26	89.08%
40th	Amadora	95 629	13	11 738	08:13	87.73%
41th	Penafiel	31 331	6	3 999	08:43	87.23%

Analysing Table 4, diametric performances will be found. On the one hand, the top-ranking stores (*i.e.,* with the highest QI) present a good queue management, having a higher ratio of clients called before 20 minutes of wait than the others. On the other hand, stores on the lower-end of this table have an opportunity to improve their queuing systems by using a proper benchmark as a target.

However, not all stores could be directly compared. To properly compare stores, clusters were created so that they would only be compared with each other if they had the same capacity (number of servers) and a similar demand. Therefore, a cluster consisted on stores that simultaneously had the same number of servers and a demand within an interval of $\pm 1\sigma_{EDP}$ from the top-ranking store's demand. In fact, even if store 'A' and 'B' had the same number of servers, but the demand of 'B' was outside the interval [Demand('A') $\pm \sigma_{EDP}$] they would not be compared. Instead, each one would generate its own cluster. After clustering stores, whenever a store would not achieve the top-ranking QI within its cluster, the respective benchmark's QI should be set as a target to be achieved.

Guarda, the best performing EDP store in Table 4, was taken as an example and its cluster was created following the previous described rules. Analysing Table 5, improvement opportunities could be directly inferred. For instance, even though Bragança's store had 91.2% of Guarda's demand, it had almost 7 times the number of clients with a WT greater than 20 minutes, proving that, after analysing clusters, two or more stores with contrasting performances were going to be easily identified.

This reality is critical for EDP, since ultimately the exemplary performances of stores like Guarda will not be verifiable on the national EDP's QI (*cf.* subsection 3.2.4) whenever there are stores with the same size, like Bragança, averaging up the number of clients that had WTs above ERSE's limit (20 minutes).

Table 5: Cluster of EDP stores with 5 service stations and a demand within $1\sigma_{EDP}$ sorted by QI

Rank	Store	Demand (2016)	Number Servers	Number of clients with a WT>20 min.	AWT (min.)	QI
1st	Guarda	37 472	5	538	03:41	98.56%
3rd	Figueira da Foz	36 534	5	955	03:27	97.39%
17th	Covilhã	34 816	5	2 087	06:04	94.00%
30th	Beja	25 002	5	2 160	07:13	91.36%
39th	Bragança	34 180	5	3 734	08:26	89.08%

As they seemed to be ideal candidates to concept-proof this Framework, these two stores became the scope of this project. The goal would be to reduce Bragança's AWT and to increase its QI, upon examining and implementing the identified best-practices of Guarda in terms of service and queue management.

6.1.3 Data Analysis

The final step of the *Model's scope definition* stage will analyse the sampled data fields to further prove their similarity and to unveil best-practices that may be set as a benchmark for the entire chosen cluster.

6.1.3.1 Quantitative Data Analysis

In this subsection, the two stores' performances (Guarda and Bragança) will be analysed and compared based on four dimensions (waiting-time, arrivals, demand-mix and service-time). These analyses will provide crucial insights to model Bragança's store and to unveil the key-differences between both systems that may explain the difference between their performances.

Quality Index

As the QI is, by law, solely based on the percentage of clients with WTs lower than 20 minutes, it was decided to begin this analysis by studying the two stores' WT histograms and by confirming the benchmark status of Guarda when compared with Bragança.

To that end, these histograms were built and overlapped (*cf.* Figure 16), with both stores' QI (*i.e.,* the cumulative probability on the [10;20[time-bin) reiterated. From this figure, it is witnessable that even though both distributions are right-skewed, Bragança presents a longer tail (*i.e.,* has a larger portion of occurrences far from the mode). In fact, as expected, while Guarda's store reaches a 99.8 cumulative percentage on the [20;30[time-bin, Bragança's distribution will only reach a similar value two time-bins ahead (99.85% on the [40;50[time-bin).

These findings suggest a better management of the queue in Guarda where, on average, customers will endure shorter waiting-periods and a lower variation of WT among clients.

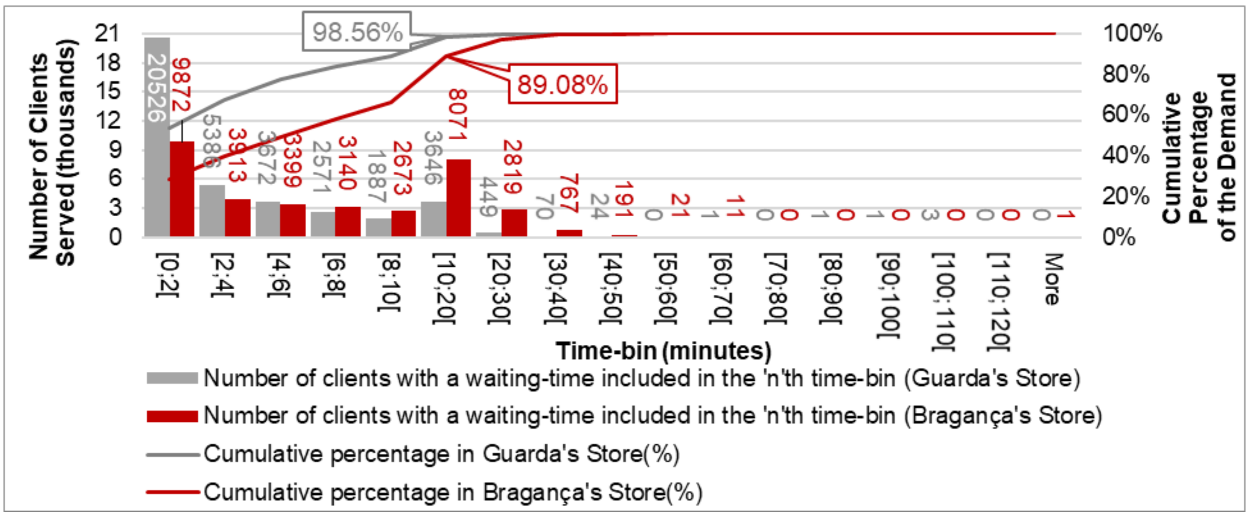

Figure 16: Overlapped Histograms of Guarda and Bragança's waiting-time

Customers' Arrivals

To further prove the reasonability of comparing both stores, the distribution of the demand was also compared. This was a key step, since it would not be reasonable to compare the system's performance in a store to which the customers arrive orderly and well-distributed along the day with a store where a high portion of the demand arrives in one period, hence congesting the system.

As the overall demand (*cf.* Table 5) does not transmit the different daily picks, nor the seasonality that may exist, an analysis of the number of clients entering per hour and per month was performed. A table in MS Excel was created for each ticket-type in both stores (5 tables per store), as exemplified in Figure 17.

Month	A: 08	09	10	11	12	13	14	15	Grand Total
jan	26	101	137	122	42	45	118	101	699
feb	26	93	105	100	50	42	84	88	588
mar	24	116	87	104	40	51	104	107	635
apr	34	83	102	93	47	47	88	78	575
mai	33	78	116	86	37	33	97	78	558
jun	41	125	130	92	61	46	115	106	716
jul	42	126	107	94	60	36	99	95	660
aug	45	149	155	161	69	35	112	132	861
set	39	82	137	94	27	33	86	79	577
oct	24	82	90	76	38	35	77	67	489
nov	22	56	99	68	26	34	77	51	437
dec	9	59	69	54	33	29	60	51	365
Grand Total	365	1150	1334	1144	530	466	1117	1033	7160

Figure 17: Number of arrivals of clients seeking 'A' tickets to Guarda's store per hour and per month

After a careful analysis of the previously mentioned tables, both stores were found to be comparable since their distribution of arrivals were similar. In Table 6, the demand for a specific ticket-type in both stores is summarized together with the yielded coefficient of variation (CV). Columns 3 to 5 present the arrivals of 2016 on a monthly basis (January to December), whereas columns 6 to 8 characterize these same arrivals on an hourly basis (from 8h30 to 16h). These arrivals were computed based on similar tables to the one depicted on Figure 17. In fact, the monthly average of 'A' tickets in Guarda (Table 6) may be computed by averaging *grand total* column, whereas the hourly average is based on the *grand total* row.

Since, so far, both stores prove to be comparable, it is also important to observe that that the demand presents different variability-levels considering different distributions (*cf.* CVs in Table 6). To simulate a real system, the most interesting distribution would be the one that is more propitious to have demand peaks. Therefore, while in Section 5.1 the goal was to minimize variability, in this step the worse (*i.e.,* the most variable) scenario is sought. Analysing the different CVs per ticket in Bragança (the store to be modelled), the hourly distribution was found to be the most variable one (its CV is on average 3.4 times bigger than when considering a monthly distribution). Thus, it should be the one considered when modelling the store's real environment.

Table 6: Arrivals comparison between Guarda and Bragança stores, following a monthly and an hourly distribution

Store	Ticket Type	Monthly Avg.	Monthly Std. Dev.	Monthly CV	Hourly Avg.	Hourly Std. Dev.	Hourly CV
Guarda	A	584.73	124.36	0.21	779.64	421.02	0.54
	B	670.16	93.77	0.14	893.54	466.25	0.52
	C	604.25	71.07	0.12	805.67	404.47	0.50
	D	280.85	68.41	0.24	374.47	200.40	0.54
	E	982.70	127.26	0.13	1 310.26	656.67	0.50
Bragança	A	415.77	62.23	0.15	554.35	288.63	0.52
	B	519.73	65.15	0.13	692.97	359.26	0.52
	C	690.25	90.26	0.13	920.33	451.57	0.49
	D	213.72	35.55	0.17	284.96	145.99	0.51
	E	1 008.91	174.63	0.17	1 345.21	665.13	0.49

<u>Stores performance on the remaining recorded KPIs</u>

The remaining component to be analysed and that would validate if both stores were comparable, was the demand-mix, hence in Table 7 a summary of the demand-mix of both stores and their KPIs is presented.

Table 7: Comparison of both stores in terms of their Demand-mix, AST and AWT per ticket

Ticket Type	Guarda Store			Bragança Store		
	Demand (%)	AST (min.)	AWT (min.)	Demand (%)	AST (min.)	AWT (min.)
A	19%	12:28	04:27	15%	09:49	08:38
B	21%	05:36	03:06	18%	06:49	08:06
C	19%	06:03	03:16	24%	06:47	08:29
D	9%	07:37	03:55	8%	06:59	08:24
E	31%	03:35	03:49	35%	05:34	08:31
Total	100%	06:31	03:41	100%	06:49	08:26

Both stores presented a similar demand-mix, where ‘E’ tickets were the most sought and ‘A’ and ‘D’ tickets were the less representative. Nonetheless, it must be pointed-out that ‘B’ is the most sought ticket in Guarda, whereas in Bragança the second ranked ticket is ‘C’. Fortunately, since ‘B’ and ‘C’ tickets have similar ASTs, this small difference in the demand-mix can be disregarded making both stores comparable.

Consequently, the two mentioned KPIs in Table 7 can be confronted. On the one hand, even though both stores had similar demand patterns and the same number of servers, Bragança’s AWT (8m26s) was 2.3 times higher than Guarda’s (3m41s). This was already expected, as Guarda’s distribution of WTs had a shorter tail where almost 70% of its clients were called under 4 minutes, ultimately leading to a higher QI (*cf.* Figure 16). On the other hand, Guarda’s Store does not present a considerable lower AST. In fact, even though it is substantially faster (-36%) serving ‘E’ tickets, it also takes more time serving ‘A’ (+27%) and ‘D’ (+9%) tickets. This point-out the existence of service best-practices in both stores and of queue management best-practices in Guarda.

<u>Service Best-practices (SBP)</u>

The previous findings motivated an identification of the service best-practices (SBPs). Therefore, both stores’ individual service performances were compared aiming at creating a common list with all SBPs. This list could then be shared among every employee leading to simultaneously shorter and less variable service-times whenever followed by a server.

To that end, an initial assessment of the current situation was performed. By creating a table that listing the offered services and their corresponding subservices (a total of 33 possible combinations), servers’ individual performances were detailed in terms of the average service-time (AST), the standard of variation of service-times (σ_{ST}) and the coefficient of variation, as presented in Table 8. It was considered that, as servers rotate among all service-stations on a weekly basis, their average service-times in the end of the year could be directly compared, disregarding the travel-duration component that it also comprises (*cf.* subsection 6.1.1), as it will be constant, especially in stores with the same dimension (*i.e.,* same cluster).

Table 8: Excerpt of the servers’ performance on each provided subservice

Ticket Type	Subservice Type	Server ID	Count of Ticket IDs	AST (min.)	Std. Dev. of AST (min.)	CV
A	0	1406	141	11:06	07:34	0.68
A	0	1407	291	08:47	04:56	0.56
A	0	1408	126	15:25	08:25	0.55
A	0	1409	781	09:25	06:17	0.67

After listing all the performances, a selection of best-practices could be conducted. Following the proposed methodology, servers were first ranked by increasing CV in each combination, as a smaller CV meant a more robust process. After sorting them, top performances would be considered exemplary, if, and only if, they simultaneously met two qualifying criteria. These criteria (explained in section 5.1) were formulated in Excel, through which a server's performance would be tested against criterium 1 (Equation 3) and criterium 2 (Equation 4).

Resorting to a pivot table, each server 'i' was listed with its respective AST_i, $\sigma_{ST(i)}$, and the number of clients he had served of each subservice type 'x'. Then, each server was confronted with the stores' average values. To be considered a best-practice, his AST on subservices 'x' had to be lower than his colleagues' (Equation 3) and had to have served more clients of subservice 'x' than his colleagues' average (Equation 4). Both Excel equations are represented next. It should be stressed that, in the latter equation, Excel does not compute an average of an average which would lead to an incorrect number. Being encoded into the pivot table, Equation 4 will average all the services duration individually.

$$\text{AVERAGEIF}([\textit{Subservice\_type}]\ ;\ x\ ;\ \text{Average}[\textit{Duration}]) \quad (3)$$

$$\text{AVERAGEIF}([\textit{Subservice\_type}]\ ;\ x\ ;\ \text{Count}[\textit{Ticket\_ID}]) \quad (4)$$

After implementing these equations, best-practices were listed in a table. Table 9 presents an excerpt (for confidentiality purposes) of that same table, where only the performances that met both qualifying criteria are listed. For each subservice (*e.g.,* subservice 0 of ticket A), one or more servers were found to serve consistently better than his colleagues.

Table 9: Excerpt of the identified best-practices on each subservice, together with their respective qualifying criteria and the store in which the server is currently working

Ticket Type	Subservice Type	Server ID	Count of Ticket IDs	AST (min.)	Std. Dev. of AST (min.)	CV	C1 (min.)	C2	Store
A	**0**	1407	285	08:47	04:56	0.56	10:04	253	Guarda
A	**0**	1409	765	09:25	06:17	0.67	10:04	253	Guarda
A	**1**	1407	14	08:23	04:48	0.57	08:45	10	Guarda
A	**2**	1524	16	11:53	05:47	0.49	16:30	13	Bragança
A	**3**	1410	444	12:34	06:17	0.50	14:06	420	Guarda
A	**3**	1280	428	09:15	08:40	0.94	14:06	420	Bragança
				...					
E	**0**	1539	1 360	03:59	03:09	0.79	04:22	1118	Guarda
E	**1**	1410	1 190	02:56	02:41	0.91	03:07	888	Guarda

Analysing this outcome, it should be stressed that while Bragança QI is lower, it still had 35% of the identified best-practices, showing that even though this dissertation will focus on assessing the impact on Bragança of adopting the remaining 65%, Guarda's store is also expected to further increase its performance by adopting Bragança's SBP. Moreover, summarizing the results by ticket type, a new AST can be estimated. If these service best-practices were to be standardized, taught and applied by every server in-store, it would be the expected to witness (in both stores) the service performances in Table 10.

Table 10: Expected AST after standardization and implementation of the best-practices

Ticket Type	AST (min.)	Std. Dev. of AST (min.)
A	08:03	03:31
B	05:33	01:59
C	05:55	03:31
D	07:38	03:34
E	03:56	02:17

6.1.3.2 Qualitative Data Analysis

In the previous subsection, a quantitative analysis of the records of Bragança and Guarda's stores proved that their system's main components are similar. Yet, they had divergent performances. Knowing the ambitioned functionalities of the model, service best-practices were quantitatively identified. In this subsection, a similar process will be followed, but, this time, exposing the current practices of the remaining component – queue discipline (*cf.* subsection 4.1.3.3). Thus, individual meetings were scheduled to fully describe and comprehend both stores management and operations prior to modelling them.

<u>Guarda's Store</u>

Guarda's store (*cf.* Figure 33 in Appendix) does not rely on the automatic mode of Inline®. Instead, the store manager has implemented her own solution. Designed by Gonçalves (2017) and her staff, this solution was discussed in 2015 with the goal of reducing the store's AWT in 2016. Taking into consideration both the perceived demand-mix and the store's AST per ticket-type, the solution consisted on implementing a fast-track. Thanks to the prioritization of fast, but frequent, services, the solution would reduce the store's AWT.

On the one hand, the current system has four service stations following a pure FIFO queue discipline. Calling the oldest ticket in-store, servers are instructed not to follow any other specific ticket type prioritization. On the other hand, the server in the 5th service station is instructed to call first the oldest client with 'B', 'C' and 'E' tickets (*i.e.,* FIFO in a specific subsystem). Only in the unlikely event of not having any clients waiting with these three ticket-types, may the server call the remaining types.

The decision of overlooking Inline®'s prioritization levels was taken unanimously in 2015, and has since then allowed a decrease of the store's AWT by 15%. As the main causes for Inline®'s inapplicability, Gonçalves (2017) argues that setting 3 static priority levels (Table 11) is too strict when the demand presents several daily peaks.

Table 11: Inline®'s Prioritization Levels in Guarda

Service Station:	**1**			**2**			**3**			**4**			**5**			**Sum of Stations**
Priority Level:	**N1**	**N2**	**N3**	**N1**	**N2**	**N3**	**N1**	**N2**	**N3**	**N1**	**N2**	**N3**	**N1**	**N2**	**N3**	
Ticket A	x			x			x					x				4
Ticket B		x			x			x		x				x		5
Ticket C		x				x				x				x		4
Ticket D	x			x			x					x				4
Ticket E			x								x		x			3
Sum of tickets	2	2	1	2	1	1	2	1	0	2	1	2	1	2	0	

Following Inline®, ‘E’ tickets (35% of the demand), for instance, would be served by three service-stations that could also call ‘A’, ‘C’, or ‘D’ tickets (tickets with high ASTs). This meant that, on the worst-case scenario, clients with quickly servable inquiries, would have to endure high WTs until the station became vacant. Moreover, Inline® lacks the flexibility of calling a specific client who, despite having a non-prioritised ticket-type, is waiting for an unusual amount of time. In fact, this is the reason why this store manager strives to be always in the front-office (which by itself is a best-practice that could not be realised without this interview). Acting like a “maître” and welcoming customers, she understands their visit purposes. Thanks to this policy, not only will she be able to increase the number of ticket-types that are well-chosen, but she will also manage the queue in real time assisting servers whenever they are faced with difficult problems, or assigning a specific client who has waited too much to an available server.

Finally, regarding servers’ daily pauses, Gonçalves (2017) explains that the store follows its legal obligation of 1-hour lunch breaks (2 servers at 12h, and the remaining 3 at 13h), plus any service-break that a server finds necessary (assuming a total of 10 minutes per day). Servers arrive to the store at 8h, while the store manager arrives 30 minutes earlier to open the store and guarantee that everything is ready to welcome the customers.

Bragança’s Store

Likewise to Guarda’s, Bragança’s store (*cf.* Figure 34 in Appendix) does not entrust Inline®’s priority levels (Table 12) to manage the prioritization of its queue. Even though the main cause for the system’s disregard is the same (lack of flexibility), the solution differs Guarda’s. Instead, Bragança’s store follows a FIFO discipline on all its service stations. Clearly aiming at a non-differentiating strategy, Braga (2017), the store manager, explained that this approach has been in place since 2015 to minimize clients’ discontent when they see a customer with a lower waiting time (*i.e.,* entered the store after them) being called first.

Therefore, every customer is called by their entrance order to the first available service station, regardless his ticket-type. Even though some improvement opportunities were acknowledged, Braga (2017) reinforced the current focus on reducing the customers’ complaints, since he believes that this is the ultimate indicator of the client’s satisfaction.

Finally, regarding servers’ daily pauses, Bragança’s store follows the same exact lunch and break policies as Guarda’s, which is not surprising as all stores must obey the same regulation (*cf.* section 3.2.1).

Table 12: Inline®’s Prioritization Levels in Bragança

Service Station:	**1**			**2**			**3**			**4**			**5**			**Sum of Stations**
Priority Level:	N1	N2	N3	N1	N2	N3	N1	N2	N3	N1	N2	N3	N1	N2	N3	
Ticket A	x			x				x			x					4
Ticket B		x			x		x			x			x			5
Ticket C		x			x		x						x			4
Ticket D	x			x				x			x			x		5
Ticket E		x			x		x			x			x			5
Sum of tickets	2	3	0	2	3	0	3	2	0	2	2	0	3	1	0	

EDP's Corporate view

At this point, there were two clear takeaways: none of the stores relied on the Inline® prioritization system; and two different queue disciplines were being enforced (one focused on reducing the AWT, and the other decreasing the customers' complaints). Thus, a prior clarification of the corporate policies was required to select a queue management best-practice.

Confronted with the stores' indifference towards Inline®, Martins (2017) acknowledged the need for reviewing this component, as it was clear that it was not answering EDP's needs. Nevertheless, regarding the two different managing strategies, Martins (2017) stressed that EDP's focus should be on reducing the customers' AWT and its distribution tail (*cf.* Figure 16). Therefore, his corporate view when choosing the preferable solution was leaning towards Guarda's solution.

This preference was based on two main principles. Firstly, the idea of having clients waiting more than others is not only acceptable, but it should also be witnessed on a daily basis. After all, clients who visit the store seeking a fast-service are likely less willing to wait, than those who seek a prolonged service (consistent with Nah's (2004) findings in subsection 3.2.4). Secondly, Martins (2017) counterposed that Bragança's strategy may not actually have the expected impact on the customer satisfaction, since there are many other indicators that should be considered when assessing the customers' satisfaction besides the number of complaints.

In fact, supporting his view, some indicators were analysed during the meeting. Considering the information summarised in Table 13, it is understandable why EDP supports Guarda solution.

Table 13: Set of indicators comparing the impact of both stores strategies

Store	Indicator	2015	2016	Variation (%)
Guarda	**AWT (min.)**	05:06	03:40	-28%
	NPS (between -100 and 100)	27.83	59.19	+113%
	Recommendation Score (out of 10)	7.64	8.43	+10%
	Satisfaction Score (out of 10)	8.04	8.82	+10%
Bragança	**AWT (min.)**	04:06	08:27	+106%
	NPS (between -100 and 100)	30.09	32.14	+7%
	Recommendation Score (out of 10)	7.74	8.04	+4%
	Satisfaction Score (out of 10)	8.13	8.43	+4%

As result of their individual strategies, and assuming that there is no other unknown change in both systems during that transitional period (*ceteris paribus*), Guarda was able to decrease by 28% its AWT, while Bragança increased it by 106%. Moreover, Bragança's strategy does not seem to be impacting customers' satisfaction, nor their recommendation scores, as much as Guarda's AWT reduction (4% *vs.* 10 %). In fact, while Bragança proclaimed a reduction of complaints, its Net Promotor Score (NPS) – a metric that indicates the percentage of promoters among the served (Keiningham et al., 2008) – has also increased significantly less than Guarda's (7% *vs.* 113%). The NPS is based on the likelihood of a customer recommending EDP's service, where only customer who have a recommendation score superior to 9 are considered promoters. In view of these results, and especially EDP's corporate view, it was decided to consider Guarda's queuing management policies as BPs to be adopted by Bragança's store.

6.2 Designing the simulation model

Throughout the last section, the detailed process of collecting, sampling and analysing the data highlighted the improvement opportunities in Bragança's store. After proving that Guarda was comparable and had best-practices that could be adopted, or even developed, the need for modelling and verifying the solutions applicability in Bragança arose. This section will then model Bragança's store by following the several steps proposed in the framework.

6.2.1 Simulation objectives

After defining its scope, this simulation shall address the three goals of this Dissertation (*cf.* Section 1.2) on Bragança's store. Throughout the process, it should estimate the AWT variation upon the adoption of: a different queuing discipline; service best-practices; and, finally, of different layouts (*i.e.,* hiring, or dismissing one employee).

6.2.2 Model components and their constraints

Simul8 is based on building blocks that represent the several components of the system. Each block has several attributes through which different levels of complexity may be reached in representational terms.

The design of the model begun by simply identifying the key activities, entities and queues in the system. Considering the customer journey (*cf.* subsection 3.2.3) and Bragança's store specificities (*cf.* subsection 6.1.3.2), the model of the store (*cf.* Figure 18) may be split into four areas: the Input Source, the Queue, the Service Mechanism, and the Simulation Clock. All of them will be described and characterised individually in the ensuing four subsections.

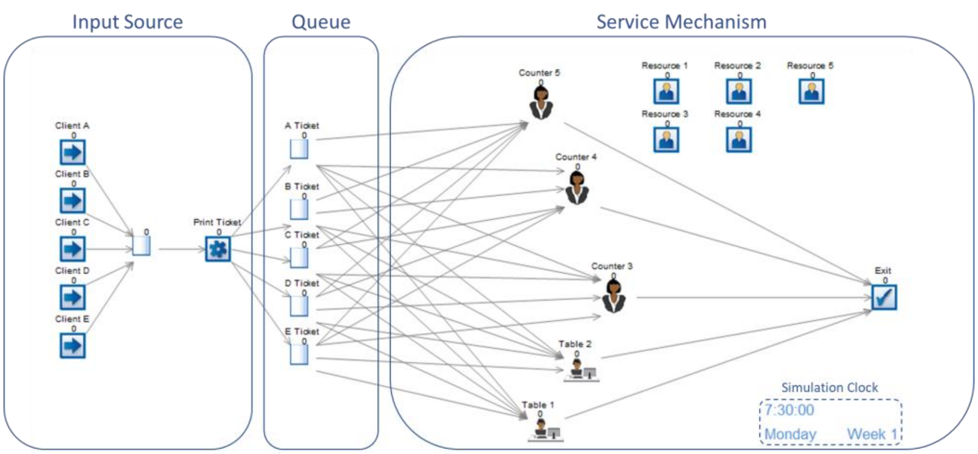

Figure 18: Simul8 Model of Bragança's store (current state)

6.2.2.1 Input Source

In the left-most area of Figure 18, 5 *start-points*, 1 *queue* and 1 *activity* may be witnessed. Together they model the entrance of entities in-system (*i.e.,* the clients entering the store) and all its specificities.

By creating 5 entrance points, an early differentiation among entities is achieved allowing entities to enter the system with specific arrival rates and with tailored attributes according to their ticket-type. In fact, in subsection 6.2.3.1, each block will have a specific exponential distribution describing the arrival of each ticket-type customer. Simultaneously, two attributes (modelled by Simul8 through labels) are set to every entity who enters the system. Once an entity enters, the first label will define the route through which each entity is going after printing the ticket ("L_TR"), whereas the second will characterize the sought service ("L_TS") (*cf.* Figure 36 in Appendix).

Since the objective of this model is to simulate ERSE's monitored time spent by a client in-store (after printing a ticket), the duration of the "Print Ticket" *activity* was set to zero and its capacity to infinite. Consequently, the transition between the entity's entrance and its arrival to the respective queue is instantaneous, and neither the activity, nor the queue will impact the simulation, being solely represented virtually. Nevertheless, it was decided to model them as well to enable future work on the entire customer journey and not only on the monitored journey. Alas, customers will often form a queue in front of the ticket dispenser and spend some time printing their ticket which may affect the real (counterposed to the monitored) customer satisfaction.

6.2.2.2 Queue

In the queue area of Figure 18, the five different created *queue blocks* represent each one of the ticket-type queues. Entities will enter one of these queues according to the embedded value on the "L_TR" label, after setting the routing-out properties of the Print ticket activity accordingly (*cf.* Figure 37 in Appendix).

Every queue has the same characterization: unlimited capacity and a FIFO sorting of entities (*i.e.,* in each queue, the first entity to arrive is the first to be called). Moreover, for now, all queues are connected to the five service stations, since currently all Bragança's service-stations call the entire range of ticket types.

6.2.2.3 Service Mechanism

The right-most area in Figure 18 consists of five service stations (three counters and two tables), the stores' five servers (the *resource*s) and the system's *exit block*.

All five service stations (*activity blocks*) have, for now, the same characteristics. Firstly, they all follow a FIFO strategy calling the oldest entity in-system (*i.e.,* the entity with the highest WT), and have a service duration that is estimated by a distribution based on the embedded value of the "L_TS" label. This distribution will allocate to each service the appropriate service-time distribution (to be defined in subsection 6.2.3.3) according to the ticket-type (*cf.* Figure 38 in Appendix). Secondly, every station requires the presence of one resource prior to calling an entity, and sets a standard efficiency level for each server (to be defined in subsection 6.2.3.4). Finally, a set of rules was determined to mimic the real system, where: a maximum of one entity can be served at each station at a time; there are no set-up times, and servers cannot leave the station until the client is fully served (*cf.* Figure 39 in Appendix).

The system's five resources (servers) are allocated to a specific service station. This allocation was performed randomly and did not interfere with the results, since none of the stations' parameters is dependent on the server (instead, they are based on the entities' labels). It was also crucial to model the lunch-breaks as they are time-bounded and have pre-defined slots. A shift availability system was defined with four shifts (8h-12h;13h-17h; 8h-13h;14h-17h) to meet that goal. Knowing that resources 1 and 2 have a lunch-break at 12h, they are available on the first and second shift; while the rest will wait until 13h, being only available to serve during the third and fourth shift (*cf.* Figure 40 in Appendix). Regarding the small breaks, they will be modelled through the previously mentioned efficiency ratio as they are not scheduled.

6.2.2.4 Simulation Clock

Considering that the system in-hands is dynamic, most of the previously mentioned components will vary along the day. Namely, the number of available resources and the arrival rates will have very distinct values during lunch-break hours, or according to the time of the day, respectively. To mimic this reality, Simul8 provides the option of setting a simulation clock according to the specific characteristics of the model.

Thanks to Braga's (2017) testimony, it is known that the store manager is the first to arrive to the store (at 7h30), followed by his colleagues (at 8h). The store opens to public at 8h30, from which moment it will welcome customers until 16h, moment when the store closes its doors to public. From that moment on, servers will either serve any remaining client, or perform other back-office duties until 17h (*cf.* subsections 3.2.1 and 6.1.3.2). The decision of modelling the entire journey of the servers including the before-opening and after-closing periods, not only allowed the existence of services after closure (*i.e.,* clients who were not called before the store closed and must still be served according to the Portuguese law), but it also made this model applicable to future work where the goal would be to raise the servers' efficiency.

Therefore, the simulation clock was set to run from 7h30 to 17h with a time-unit of minutes for five days a week (*cf.* Figure 35 in Appendix). Furthermore, as seen in Figure 18, this simulation clock was set to be always shown in the simulation environment with a full level of detail (*i.e.,* hour, weekday, week-number).

6.2.3 Input data

In this subsection, data that may still be lacking will be computed and presented. After this step, every component of the current model shall be completely characterised to run the simulation.

6.2.3.1 Entities' Arrivals

Knowing that the store is visited by clients with five different types of inquiries, it is crucial to mimic the real demand peaks per ticket-type 'h' when computing the expected interarrival time (λ_h). Based on the literature review (*cf.* subsection 4.1.2.2), it was known that this parameter should follow an exponential distribution. However, the significant variation of the number of arrivals witnessed on the hourly distribution (*cf.* subsection 6.1.3.1) motivated a dynamic parameter according to the simulated time. To ease its implementation, four time-slots (t) were created (8:30-9h; 9-12h; 12-14h; 14-16h) characterizing the opening, morning, lunch and afternoon arrival blocks, respectively.

The interarrival time for each ticket-type and for each time-slot ($\lambda_{h,t}$) was calculated following Equation 5. Parameter $\lambda_{h,t}$ is equal to the multiplicative inverse of the quotient of the average hourly number of arrivals of clients with ticket-type 'h' in time-slot 't' ($\mu_{h,t}$), by 60 (considering the time-unit of minutes).

$$\lambda_{h,t} = (\mu_{h,t}/60)^{-1} \tag{5}$$

After computing and modelling the 20 exponential distributions (*i.e.,* the possible combinations of $\lambda_{h,t}$) presented in Table 14, two extra periods of one hour each had to be modelled considering that the simulation clock included time periods during which no client was allowed in-store. To these two time-slots named "before opening" and "after closure" two fixed distributions were created with a value of 61, hence guaranteeing that no customer would enter the system during the 60 minutes in which these distributions were modelling the customers' arrivals (*cf.* Figure 41 in Appendix).

Table 14: Arrivals characterization in Bragança's Store

Ticket Type	$\mu_{h,t}$				$\lambda_{h,t}$			
	8:30-9h	9-12h	12-14h	14-16h	8:30-9h	9-12h	12-14h	14-16h
A	1.68	3.49	2.09	2.73	35.67	17.19	28.70	21.94
B	2.28	4.48	2.36	3.00	26.30	13.41	25.46	19.98
C	2.32	5.51	3.17	4.00	25.84	10.88	18.93	15.01
D	1.53	2.11	1.48	1.74	39.20	28.39	40.64	34.43
E	4.12	8.12	3.93	5.60	14.56	7.39	15.27	10.72

6.2.3.2 Travel Times

In Simul8, it is possible to set distinct travel times, or distances, between two components (blocks). Considering that once a customer is called, his server is already in the service-station, all the servers' travel times were set to zero. The same value was set to every link between the blocks in the input source (*cf.* Figure 18), as it was not in the scope of the current work. Moreover, regarding the distance clients must travel between their queues and the service-stations, travel times had to be considered to be zero, as the recorded AST by Inline® already includes this length of time (*cf.* subsection 6.1.1 and Figure 42 in Appendix).

6.2.3.3 Service Times

Regarding the duration of the services, the only recommended distribution found in the literature was the exponential distribution (*cf.* subsection 4.1.4). Nevertheless, after a first analysis of the histograms, it was concluded that no specific probability distribution seemed to accurately describe these five distributions. The histogram of the service-times of 'A' tickets is shown in Figure 19, as an example to prove that same statement (the rest of the histograms are in Appendix - Figure 43 to Figure 46).

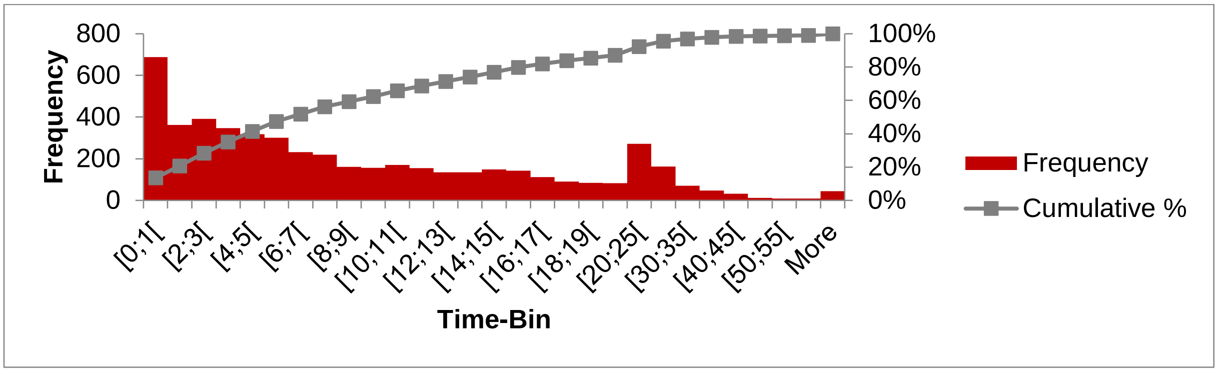

Figure 19: Histogram of the service times of tickets 'A'

Therefore, it was decided to resort to the real histograms and to create a tailored distribution in Simul8 for each ticket-type (*cf.* Figure 38 in Appendix) that would mimic the witnessed service-performances. Due to labour-law restrictions, these distributions were not tailored to each server, but this would have been possible to do it considering the collected data (*cf.* subsection 6.1.1).

Instead, all five tailored distributions were modelled into the service-stations. Upon the arrival of one entity, the simulation model would check the entities' label (*i.e.,* its ticket-type) and choose the respective distribution (label-based distribution).

6.2.3.4 Servers Efficiency

After modelling their AST, servers' efficiency levels were the remaining input to be set in the properties *of the activity blocks*. This was crucial to mimic the real behaviour of the system, since servers will not be always available. In fact, it would be unrealistic to consider that the store had 5 servers available at any moment throughout the day considering their lunch-breaks and pauses (*cf.* subsection 5.2.1).

To model this component, Simul8 (*cf.* subsection 5.2.2) offers the possibility of setting an efficiency index to each server. Nevertheless, due to legal labour restrictions in Portugal, only an average index (*i.e.,* the store's efficiency) could be computed. Considering that the server's mission is to serve clients, he would only be considered as active whenever he was serving a customer. Otherwise, he would be considered idle, hence decreasing the store's efficiency level.

An algorithm (*cf.* Figure 47 in Appendix) was created in MS Excel to compute the average number of active servers in 5-minutes slots throughout the store's opening hours ($A_{s,t}$). Considering the collected data (*cf.* subsection 6.1.1), each server 's' was classified, per time-slot, as being active ($A_{s,t}$=1), if he had called or served a client in that slot, or idle ($A_{s,t}$=0), otherwise. In the unlikely event of not having any clients to serve, the server was also classified as active. After all, an empty store should not have an impact on the servers' efficiency, as this occupation ratio (outside the project's scope, *cf.* subsection 5.2.1) could be directly inferred by Simul8.

Replicating this computation for the entire set of days (the 250 work-days of 2016), an overall efficiency index of 80% was obtained, highlighting the crucialness of this step, since if servers were always working besides legal breaks, the theoretical efficiency would be of 89.5%.

But considering that lunch-breaks had already been modelled in Simul8 (*cf.* subsection 6.2.2.3), the efficiency index of the store disregarding the one-hour lunch break had to be computed. The resulting efficiency (90%) was embedded in the model (*cf.* Figure 48 in Appendix) so that a server would only be 90% of his service-time available.

6.2.4 Definition of the KPIs

The remaining component to be set in this step of the framework was the KPIs. While EDP only has three KPIs (AWT, AST and QI), other metrics were created to verify and validate the systems' behaviour. Considering the simulation objectives, three types of indicators had to be set (*cf. Table 15*). Firstly, two high-level metrics were set to quantify both the total number of customers entering the system and their average time in-system. Then, focusing on the service-stations, the total amount of clients served per service-station was set. Finally, the last group of created metrics concerned the customer journey per ticket-type: number of clients with that inquiry, AWT, Standard Deviation of WT (σ_{WT}), Maximum WT, and percentage of customers served under 20 minutes (*i.e.,* the QI per ticket-type). Unfortunately, although it was not possible to create any metric recording the entities' AST per ticket, as Simul8 does not provides that option. However, the current metrics proved to be sufficient when verifying the model (*cf.* subsection *6.3.2.1*)

Table 15: Created metrics to record the simulated performance of the model

Metric's Scope	Performance Measure	EDP's respective KPI
System	Number of visitors	*N/A*
	Visitors' Average Time in System	*N/A*
Service-station	Number Completed Jobs	*N/A*
Customers journey per ticket-type	Average Waiting Time	AWT
	St. Dev. of Waiting Time (σ_{WT})	*N/A*
	Maximum Waiting Time	*N/A*
	Number of visitors with that inquiry	*N/A*
	% Visitors with a smaller AWT than the time limit (20 min.)	QI

By recording and collecting all these metrics, the performance of the new scenarios could be easily assessed and compared with Bragança's current performance. To do so, firstly the variation of the relevant KPIs of EDP (AWT and QI) would be analysed. The AST analysis was disregarded, as, in this model, it was an input (*i.e.*, a parameter, not a variable). A good scenario would decrease the AWT and increase the QI.

If both were forecasted to improve, the concerns of Bragança's store manager had to be taken into consideration too, knowing that solutions would be autonomously implemented by the store. Namely, his focus on reducing the σ_{WT} among different ticket-types. As AWT was forecasted to decrease, once again the need to resort to the CV arose. After all, only the solutions where the CV would not significantly increase could be expected to be welcomed by a store manager who had sacrificed his own KPIs to reduce the customers complaints.

Finally, and considering EDP's ambitioned future state, a scenario could not increase the overall system performance by deteriorating the AWT of the 'A' tickets. As a lower store's AWT could be achieved by, on average, reducing all the other tickets' AWT.

6.3 Validation of the Model

After tailoring the model's design to its scope, the present model should be run, verified and validated to assess its accuracy when compared with Bragança's real performance in 2016.

6.3.1 Computational implementation of the system's current state

To run a model in Simul8, after implementing all its components, constraints, distributions, and parameters, three simulation variables must be defined: the *Warm-up Period*, the s*imulation Data Collection Period*, and the desired *Number of Runs*.

The warm-up period allows the model to run during a certain period of time without collecting any metrics. This feature is crucial in processes where the simulation should end a day with a certain work-in-process (WIP) and continue its production in the next day. However, considering that the store is always going to vacate after each simulated day, the warm-up period should be set to zero (*cf.* Figure 35 in Appendix).

Regarding the simulation data collection period, it was decided to run a simulation for one year, since Inline®'s collected data and ERSE's monitoring of KPIs has an annual periodicity, hence easing the comparison of simulated results with real ones.

Finally, to set a desirable number of runs, the trial calculator of Simul8 was used to determine the necessary number of runs to achieve a precision of 95% (*i.e.,* confidence limits of the simulation results). After running this computation, it was suggested by Simul8 to run 4 trials (*cf.* Figure 49 in Appendix).

Consequently, embedding these last variables, the simulation resulted in a forecasted AWT of 8:13 minutes and a QI of 92.3%.

6.3.2 Model's Verification and Accuracy Assessment

In this subsection, tests will be performed to verify and to validate the model, guaranteeing that the model is mimicking the real environment prior to continue to the next steps of the proposed framework.

6.3.2.1 Verification of the model

Each component was tested individually by running the simulation on slow-speed and by creating sub-systems where closer analyses of the simulated behaviour were performed. To guarantee that every component was tested, it was chosen to simply follow the journey of several individual entities (customers) and verify that *events* were being conducted as planned (*cf.* subsection 4.2.2.1).

Thus, firstly, in the *input source*, the five arrival blocks were confirmed to be: generating entities with the correct label, by creating dummy queues and verifying their contents; and to be following the appropriate arrival distributions, by verifying if the number of arrivals per ticket-type, in each time-slot, matched the respective probability distribution. It was also verified if, during the day, queues were formed between the entrance and the printing ticket activity. After all, not being part of the scope and being modelled only to ease future works, entities should move instantaneously to their respective queues and to pass through that activity with a fixed time of zero seconds.

Then, in the Queue area, each one of the five *queues* was verified to be: sending to service its oldest client (*i.e.,* highest WT) by testing the calling mechanism throughout an entire simulated day; and to not be leaving any unserved clients at the end of the day. Moreover, it was also tested that the queues were both infinite, by isolating them and running the simulation, and populated with their respective ticket-type by verifying that clients were being routed to their respective queue after the printing activity according to the "L_TR" label.

Finally, in the *service mechanism*, both activities and servers were verified to be mimicking the current system of Bragança's store:

- Concerning the activities, the routing of the entities between their queues and their service-stations was tested guaranteeing that each service station called every ticket-type with the FIFO discipline. Moreover, knowing that the official Inline® system was also going to be tested, it was also ascertained that each station was able to follow a certain queue discipline (*cf.* Figure 51 in Appendix) either by calling only a specific subset of ticket-types, or by calling clients following a different ruling (LIFS, or Priority-levels). Finally, the enforcement of the modelled AST distributions was verified, by comparing the average time in-system and the AWT (*i.e.,* the AST per ticket) with the respective distribution of each ticket-type.
- Concerning the servers, it was tested if they were in fact taking their lunch-breaks at the scheduled time by not being available to call any entity during that hour. Still, knowing that they cannot leave the store in the middle of a service, the scenario where a customer was called one minute before the server's lunch-break was tested to see if, in fact, the server would remain until the end of its service (prior to take his lunch-break). Finally, it was also verified that the server was not always available to call a client following the efficiency ratio by running an entire simulation with only one server and with no-lunch time, so that his absence would only be justifiable by an efficiency loss.

6.3.2.2 Validation of the model

After performing every single test, the model's behaviour was verified to mimic the real system. Therefore, its estimated performance in subsection 6.3.1 was compared with the actual performance of Bragança's store in 2016 (*cf.* Table 5). The simulated AWT was 1.63% below the real 2016's AWT and the simulated QI was 3.6% above the real QI. Being aware that both results convey a high accuracy rate (*ca. 97.5%)*, especially considering the usual 95% benchmark in most simulations (Sargent, 2013), the model was considered to be valid and accurate, while being also acknowledged a slight tendency to underestimate the queuing effect.

6.4 Simulation of different solutions

In this Section, several scenarios are going to be tested after having verified the accuracy and the validity of the model. Whenever required, it will be presented an explanation of the performed changes between different scenarios, as some of the previous steps of the framework will need to be revised throughout this testing, either due to the addition of new components, or to change of some parameters.

6.4.1 Design of new systems and Assessment of their performance

In this subsection several solutions based on either best-practices, or on developed concepts, will be tested and compared. The goal is to find at least one scenario where an improvement of the KPIs is foreseeable and still be applicable to Bragança's store.

6.4.1.1 Developed Solutions based on the Inline® system

Considering the outcomes of the performed data analyses on subsection 6.2.1, several best-practices were identified in Guarda's store. Since this store proved to be comparable with Bragança's, the motivation to adopt these BPs arose. Reiterating the findings, Guarda applies a queuing discipline that prioritizes fast services but is flexible enough to call a non-prioritized ticket whenever it is waiting for long periods of time.

Nevertheless, before starting to redesign the system, the current Inline® prioritization system had to be simulated, to objectively reject this option. After all, so far it is only known that it is the apparent consensual opinion of store managers is to not rely on Inline® and that performances improved since they stopped using the system (*cf.* subsection 6.1.3.2). Yet, there has not been actual objective confirmation on whether the system decreases the stores' performances. Thus, after setting each *activity block* according Inline®'s prioritization system in place (*cf.* Table 12), by deleting some of the routes (since not every station call the same amount of ticket-types) and setting Inline®'s priority-levels discipline (*cf.* Figure 50 in Appendix), the system was simulated.

As seen in Table 16, the results of simulation B support the staff feedback that the system was performing worse when they followed Inline®'s automatic mode. According to this simulation, the AWT would raise by 6% and the QI would decrease 1.1%, which would be a devastating result to a store that is already the third worst performing store of EDP (*cf.* Table 4).

Table 16: Comparison of the current (A) system with the simulated scenarios based on the Inline® system

Scenario	AWT	St.Dev. WT	CV	QI	%Var. AWT	%Var. QI	AWT 'A' Ticket	%Var. AWT('A' Ticket)
A) Current Strategy in-place / 90%	08:13	0.31	55.2	92.30%	-	-	08:29	-
B) Ruled by Inline Bragança / 90%	08:42	2.03	336.6	91.30%	+6.00%	-1.10%	07:20	-14%
C) Ruled by Inline Guarda / 90%	09:38	1.97	295.2	91.00%	+17.20%	-1.40%	08:19	-2%
D) Prioritizing E,C,B,A,D / 90%	07:58	4.85	877.1	92.10%	-3.00%	-0.30%	12:10	+43%
E) Prioritizing expired E,C,B,A,D / 90%	07:50	4.06	745.3	91.00%	-4.60%	-1.40%	10:33	+24%

Three scenarios were also simulated (with again 90% of efficiency level) to guarantee that a decrease in performance whenever it a strict prioritization scenario was followed was an actual trend. Therefore, simulation C tested the performance of the system when following Inline® rules of Guarda. As expected the result was even worse, what could prove that there was an effort by Inline®'s team to tailor each solution to each store, even if it proved unfruitful. Also, in a final effort to test the AWT reduction potential of this queuing discipline, two priority rules were tested that prioritized the most sought tickets in-store first (E>C>B>A>D, as described in Table 7). Simulation D was able to reduce the AWT, but it deteriorated the QI by 0.3% and the AWT of 'A' tickets by 43%. Moreover, it increased the standard deviation of WT (σ_{WT}) among different ticket-types from 0.31 to 4.85 (*i.e.,* 16 times more). In simulation E, trying to mitigate this increase of σ_{WT}, a dominant rule was set to call any ticket that was approaching the 7 minutes of wait (close to EDP's goal) and that otherwise followed the ticket-type prioritization. This scenario's goal of mitigating the σ_{WT} proved to be accomplished by a decrease of the CV and of the AWT. However, the σ_{WT} was still higher than desired and the QI continued to decrease.

6.4.1.2 Developed Solutions based on Guarda's Example

Having confirmed that this system required a more dynamic solution, it was decided to proceed by testing several scenarios implementing Guarda's speed-track concept (combination of fast- and slow-track).

To mimic Guarda's solution, while four service-stations in-store were operating in pure FIFO strategy, the fifth station had to be allowed to call firstly the subset of BCE ticket-types (previously conducted to a *fast-track,* FT) in a FIFO discipline, and only then the remaining ticket-types (arriving from a *slow-track*).

Therefore, this scenario required a deeper redesign of the system. This time, the creation of two new triage *activity blocks* connected to one of the service-stations was required (*cf.* Figure 52 in Appendix). Two important settings were defined for both new *activities*: firstly, they had a zero-duration activity, guaranteeing that an entity would cross the *fast-track* instantaneously towards the service-station; and secondly, they were only allowed to collect an entity from its queue when the destination station was available (*i.e.,* empty), avoiding the situation where an entity would be retrieved from its queue before the server became available (inaccurately reducing the AWT). As the remaining stations would continue to call every ticket-type following FIFO and service-times are based on ticket-types (and not servers), the speed-track system could be connected to any service-station without influencing the results.

After successfully implementing and verifying the system, the simulation could proceed. Simulating Guarda's solution (scenario F in Table 17), results showed a 2% decrease of the AWT that prompted a QI increase of 0.6%. The AWT of 'A' tickets also decreased considerably (-13%). The only problem with this solution, was the increase of the CV that almost doubled to 101.4. This would trigger the store manager, who would become doubtful especially considering his testimony where he prioritized the perceived level of satisfaction in-store (*cf.* subsection 6.1.3.2).

Table 17: Comparison of the current (A) system with the simulated scenarios based on Guarda's strategy

Scenario	AWT	St.Dev. WT	CV	QI	%Var. AWT	%Var. QI	AWT 'A' Ticket	%Var. AWT('A' Ticket)
A) Current Strategy in-place / 90%	08:13	0.31	55.2	92.3%	-	-	08:29	-
F) Mimicking Guarda's solution / 90%	08:03	0.57	101.4	92.8%	-2.0%	+0.6%	07:21	-13%
G) FIFO: 1 FT: B,C,E 4,4,2 min / 90%	08:00	0.33	59.4	93.3%	-2.7%	+1.1%	08:19	-2%
H) FIFO: 1 FT: B,C,E -,-,2 min./ 90%	08:06	0.37	66.3	93.3%	-1.4%	+1.1%	08:23	-1%
I) FIFO: 1 FT: E / 90%	08:03	0.86	153.3	93.2%	-2.0%	+1.0%	08:59	+6%
J) FIFO: tables 'AD' & counters 'BCE' / 90%	07:54	2.13	387.5	95.7%	-3.9%	+3.7%	10:59	+29%
K) FIFO: 1 dedicated counter E / 90%	09:37	1.31	195.6	90.1%	+17.1%	-2.4%	10:41	+26%

Thus, developing Guarda's concept, a new scenario G was tested granting *fast-track* priority to BCE tickets that would be triggered whenever their WTs would be higher than a specific value. To compute these time-triggers, it was decided that WTs would desirably have in consideration the weight of each ticket-type (in terms of demand) and their AST. As presented in Table 18, a ratio was defined for each ticket that not only prioritizes the ticket-types that represent more hours of wait (column 4), but also considers their expected service-duration (column 3). In this way, the tickets which presently endure a higher weighted AWT (column 5) and that, at the same time, will have shorter service-durations (column 6) will be set to wait lower AWTs on the long term. It was decided to set the trigger of 'A' tickets in 2 minutes (approximately half of its AST), hence, considering the suggested ratio, BC tickets will be set to 4 minutes.

Table 18: Definition of the AWT Triggers for Scenarios G, H, L to P, and T to W

Ticket Types	**(1) AWT**	**(2) Demand**	**(3) AST**	**(4) =(1)*(2)**	**(5) =%(4)**	**(6) =(5)/(3)**	**Suggested Ratio [max(6)] / (6)**
A	08:38	4989	09:49	718	15%	21.91	4
B	08:06	6237	06:49	841	18%	37.01	2
C	08:29	8283	06:47	1172	24%	51.70	2
D	08:24	2565	06:59	359	7%	15.43	6
E	08:31	12107	05:34	1717	36%	92.31	1
Grand Total	-	-	-	4807	100%	-	-
Maximum Values	-	-	-	-	-	92.31	-

Thanks to this systematic tailored trigger-setting, scenario G was able to simultaneously reduce the expected AWT (-2.7%) and increase its QI (+1.1%). Moreover, the σ_{WT} (0.33) would approximately reach the current observed levels in Bragança (0.31).

Thus, this concept of FTs with time-triggers had a good performance. Still, other solutions were also tested to guarantee that a shared fast-track among BCE tickets would lead to better performances. Having in mind the weight of 'E' tickets in the demand-mix (*cf.* Table 18), four alternative solutions arose.

Firstly, scenario H tested the impact of the previously defined triggers in the *fast-track* by setting a trigger in 'E', but not in the remaining 2 tickets-types. This way, the *fast-track* would collect entities following a pure FIFO among BCE queues, unless there was an 'E' triggered ticket. The result proved to harm the system's performance when compared with the scenario H, even though it was still leading to better KPIs than the current solution.

Secondly, scenario I simulated a system where only 'E' tickets would be collected by the fast-track, being all the other types routed to the *slow-track*. Even though it increased the KPIs when confronted with the current performance, it also increased the AWT of 'A' queues (+6%), as 'BC' tickets were not being prioritized, hence being served in a broader range of stations, and increasing the σ_{WT} (almost 3 times more than in scenario A).

Thirdly, scenario J fitted the stores' system to the ideal conceptual design of EDP branding where clients seeking slow services would be seated in tables, and the rest would be served in counters (*cf.* subsection 3.2.3). Since Bragança has 2 tables and 3 counters (*cf.* Figure 18) this increased considerably the AWT of

‘A’ tickets (+29%) and the σ_{WT} that was now almost seven times the current one (scenario A). However, the stores’ AWT improved (-3.9%), and so did the QI (+3.7%).

Finally, in scenario K, a new concept of dedicating one of the counters to ‘E’ tickets (which have 35% of the demand, *cf.* Table 7) was tested. The performance of the system decreased notably, with an increase of the AWT of 17.1% and a decrease of the QI of 2.4%.

Overall, analysing these results some takeaways may be drawn. Firstly, it was verified the premise that the system truly urges a flexible solution. Whenever flexibility was reduced by dedicating service-station (scenarios I and K), or by not implementing time-triggers (scenarios F and H), the performance would deteriorate. Secondly, it was clear that the triggering rationale was leading to better comprehensive performances (*i.e.,* scenario G improved the two KPIs, AWT and QI, without a significant increase of the σ_{WT}). And, lastly, it should be noticed that the scenarios where stations called every ticket-type were leading to lower CVs, which is important when seeking the store manager support.

6.4.1.3 Developed Solutions improving Guarda’s Example

After concluding that the *fast-track* concept based on Guarda’s solution was also applicable in this system and the WT triggering had a positive effect, a set of simulations described in Table 19 focused on understanding until when would it be beneficial to create more *fast-tracks*.

Table 19: Comparison of the current (A) system with the simulated scenarios developing Guarda’s strategy

Scenario	AWT	St.Dev. WT	CV	QI	%Var. AWT	%Var. QI	AWT 'A' Ticket	%Var. AWT('A' Ticket)
A) Current Strategy in-place / 90%	08:13	0.31	55.2	92.3%	-	-	08:29	-
L) PD+FIFO: 2 FTs: B,C,E 4,4,2min / 90%	07:38	0.34	64.1	94.3%	-7.2%	+2.1%	07:52	-7%
M) PD+FIFO: 2 FTs: B,C,E -,-,2min / 90%	07:49	0.31	57.7	94.2%	-4.9%	+2.0%	08:09	-4%
N) PD+FIFO: 3 FTs: B,C,E 4,4,2min / 90%	07:29	0.35	67.2	94.7%	-9.0%	+2.6%	07:52	-7%
O) PD+FIFO: 4 FTs: B,C,E 4,4,2min / 90%	07:26	0.34	65.8	95.1%	-9.6%	+3.0%	07:46	-9%
P) PD+FIFO: 5 FTs: B,C,E 4,4,2min / 90%	06:56	0.22	46.6	95.4%	-15.6%	+3.4%	07:03	-17%

A new model with two *fast-tracks* was designed with the addition of a second pair of *activity blocks* (one *fast-* and another *slow-track*) to a second station to meet that goal. This second station to be connected was based on the first one, since it was sought a station occupied by a server with a different lunch-break shift (*cf.* Figure 53 in Appendix).

Simulating this solution, scenario L achieves an even better performance than scenario G decreasing the AWT by 7.2% and improving the QI by 2.1%. Once again, in scenario M, it was confirmed that the triggering rationale led to better performances, since whenever they were removed (*cf.* explanation of scenario H) higher AWTs would be achieved.

Thus, considering that strategy L led to the best performance so far, from scenario N to P, the number of service-stations that were calling clients with a prior *speed-track* triage was progressively increased. The last one (scenario P, with 5 *fast-tracks*) achieved the best performance so far. With this configuration, the system was able to achieve a 3.4% increase of the QI prompted by not only a record low AWT (6:56 minutes) which implies a decrease of 15.6%, but also a significantly lower σ_{WT}.

6.4.1.4 Developed Solutions testing new trigger sets

Considering scenario P, the trigger set was tested, since the easiest way of justifying that set to a staff member was to show by comparison that a specific set would lead to better results than others, hence avoiding getting into mathematical details.

Anticipating that the fact of tailoring the trigger to the store's demand-mix would be easily understood, scenarios Q and R tested how the system would react by decreasing the trigger of 'C' tickets (2nd highest demand after 'E' tickets), or by increasing the trigger of 'B' tickets (lowest demand among the ticket-types in the *fast-track*), respectively. In Table 20, both performances prove to achieve worse performances than scenario P. The latter scenario will decrease the AWT by 15.2% which is smaller than P's reduction, and the former doubles the σ_{WT} crucial for attain the support of the store manager.

In scenario S, a different approach was taken. Considering the computed weights in column 6 of Table 18, 'E' tickets were over-prioritized and 'B' tickets under-prioritized. This solution is also forecasted to decrease the system's performance with a lower decrease of AWT and higher CV.

Overall, even though these scenarios are all leading to better results than the strategy currently in-place, the previously defined time-trigger based on the store's demand-mix proves to lead to better results.

Table 20: Comparison of the current (A) system with the simulated scenarios developing Guarda's strategy

Scenario	AWT	St.Dev. WT	CV	QI	%Var. AWT	%Var. QI	AWT 'A' Ticket	%Var. AWT('A' Ticket)
A) Current Strategy in-place / 90%	08:13	0.31	55.2	92.3%	-	-	08:29	-
Q) PD+FIFO: 5 FTs: B,C,E 4,3,2min / 90%	06:55	0.40	84.2	95.5%	-15.9%	+3.4%	07:19	-14%
R) PD+FIFO: 5 FTs: B,C,E 5,4,2min / 90%	06:58	0.35	71.6	95.4%	-15.2%	+3.4%	07:15	-15%
S) PD+FIFO: 5 FTs: B,C,E 5,4,1min / 90%	07:01	0.46	95.3	95.4%	-14.6%	+3.4%	07:11	-15%

6.4.1.5 Developed Solution with service-practices implementation

After simulating 19 scenarios and finding one that was clearly improving every dimension in-scope (scenario P), it was decided to proceed by simulating the impact of the identified best-practices in subsection 6.1.3.1.

To do that, it was assumed that after standardizing the service processes each ticket AST would follow a normal distribution with the parameters represented on Table 10. As represented in Table 21, a significant reduction of the AST and its standard deviation (σ_{ST}) was expected. After modelling, implementing them in Simul8, and updating the label-based distribution, the store's performance would improve in every parameter (*cf.* Table 22): the QI would reach 96% (an increase of 4%) and the AWT would decrease by 53% (to 3:52 minutes). Moreover, the σ_{WT} would decrease by 29% and the AWT of 'A' tickets 51%.

Table 21: Estimated Impact of the BPs adoption on Bragança's store

Ticket Type	AST (min.)	σ_{ST} (min.)	% Var. AST (%)	% Var. σ_{ST}	Original AST (min.)	Original σ_{ST} (min.)
A	08:03	03:31	-20%	-69%	10:01	11:23
B	05:33	01:59	-20%	-78%	06:57	08:55
C	05:55	03:31	-15%	-62%	06:56	09:13
D	07:38	03:34	+7%	-58%	07:07	08:35
E	03:56	02:17	-31%	-72%	05:41	08:16

Table 22: Comparison of the current (A) system with the simulated scenario applying the identified BPs

Scenario	AWT	St.Dev. WT	CV	QI	%Var. AWT	%Var. QI	AWT 'A' Ticket	%Var. AWT('A' Ticket)
A) Current Strategy in-place // 90%	08:13	0.31	55.2	92.3%	-	-	08:29	-
T) PD+FIFO: 5 FTs: B,C,E 4,4,2min +BPs / 90%	03:52	0.22	82.3	96.0%	-53.0%	+4.0%	04:10	-51%

6.4.1.6 Sensitivity Analysis on the developed Solution with service-practices implementation

Regarding the last proposed feature of the Framework, the final two scenarios focused on performing a sensitivity analysis concerning the number of server in-system. As the store had only 5 servers and any proposal would imply a complete redesign of the store layout, analysing scenarios where more than one server would be hired (or dismissed) would not be realistic, especially considering that its expected traffic (*cf.* Table 5) is already 5% lower than the average demand of the remaining stores with that same layout (*cf.* Figure 6). Therefore, to prove that the feature could produce interesting results, the sensitivity analysis was performed by hiring, or dismissing one server. Both scenarios are depicted in Table 23.

Table 23: Comparison of the current (A) system with the simulated scenario upon the hiring and dismissal of one server in scenario T

Scenario	AWT	St.Dev. WT	CV	QI	%Var. AWT	%Var. QI	AWT 'A' Ticket	%Var. AWT('A' Ticket)
A) Current Strategy in-place // 90%	08:13	0.31	55.2	92.3%	-	-	08:29	-
U) PD+FIFO: 4 FTs: B,C,E 4,4,2min / 90% w/ BPs but 4 workers	07:53	0.19	35.4	95.5%	-3.9%	+3.5%	07:57	-6%
V) PD+FIFO: 6 FTs: B,C,E 4,4,2min / 90% w/ BPs and 6 workers	02:40	0.21	114.4	98.0%	-67.6%	+6.2%	01:29	-83%
W) PD+FIFO: 5 FTs: B,C,E 4,4,2min / 90% w/ BPs and 6 workers	01:59	0.20	145.2	97.4%	-75.8%	+5.5%	01:54	-79%

On the one hand, after implementing the *fast-track* and the BPs among servers, even if a server would be dismissed, the stores' KPIs would be better than the ones from the current strategy. In scenario U, it is shown that in that case AWT would still decrease 3.9%, increasing the QI 3.5%.

On the other hand, if EDP decided to hire a new assistant the question of whether to implement the FT concept, or not (following a pure FIFO strategy), arose. The former case was simulated in scenario V, and the latter in W. Comparing both, scenario W has a bigger impact on the AWT KPI (-75.8% *vs.* -67.6%). This indicates that a plateau may have been reached and that the five previously existing sets of *speed-tracks* (pair of *fast-* and *slow-tracks*) are already absorbing most of the BCE tickets. However, scenario V, proves to foster a more robust solution by leading to a better QI (+6.2% *vs.* 5.5%) and a lower CV (114.4 *vs.* 145.2).

6.4.2 Final proposal

Considering every simulated scenario, a final proposal should be drafted in the last step of the Framework.

Among all the different tested queuing priorities (scenarios B to P), scenario P is the one that led to the best performance with an estimated impact of 01:17 minutes reduction of the AWT (*cf.* Figure 20) and a 3.1% increase of Bragança's store Quality Index (*cf.* Figure 21).

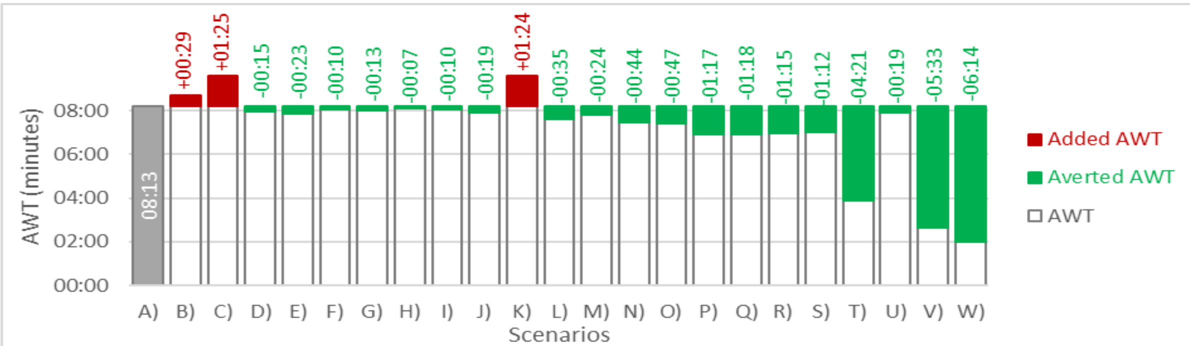

Figure 20: Performance of each simulated scenario according to its AWT

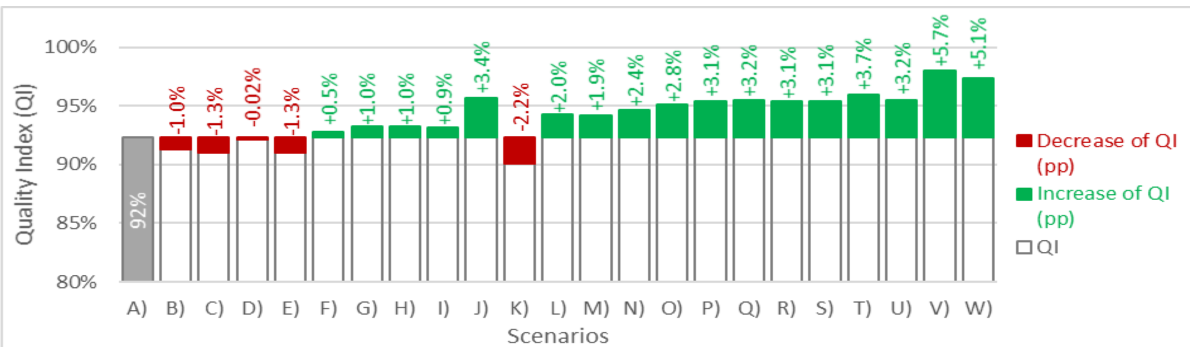

Figure 21: Performance of each simulated scenario according to its QI

Moreover, from these, P was also the only scenario capable of reducing the system's CV (*cf.* Figure 22). This accomplishment is crucial even though it is not one of the established KPIs in EDP. After all, it is known that the entire set of measures undertaken by the store manager in 2015 (*cf.* subsection 6.1.3.2) were aimed at reducing the AWT disparity among Bragança's clients. Thus, any solution that not only reduces that disparity, but also improves the store's KPIs will be certainly welcomed.

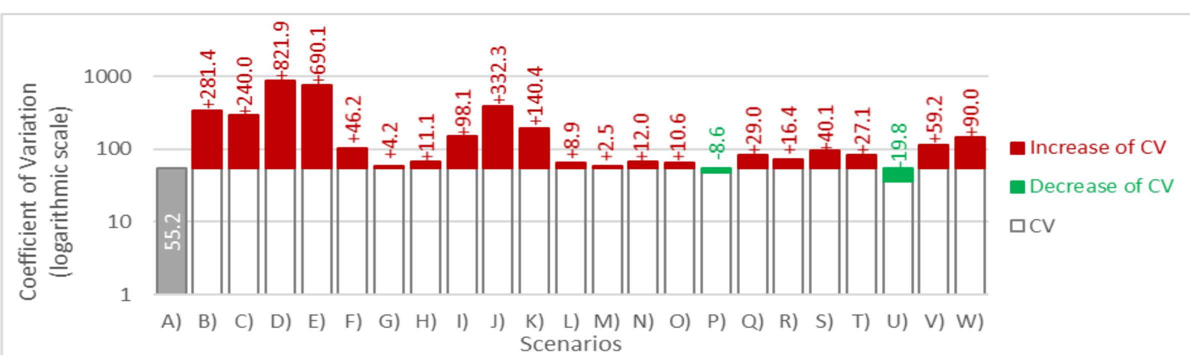

Figure 22: Performance of each simulated scenario according to its CV

Considering that it is clear that this dynamic queue discipline based on time-triggers is well dimensioned for the intendent objective (finding an implementable solution that increases the system's performance) and all the other different sets of time-triggers (scenarios Q to S) are leading to an increase of the CV (*cf.* Figure 22), the implementation of scenario P is proposed, if it is decided not to carry on with the initiative of sharing the identified BPs and training the servers accordingly.

On the other hand, if best-practices were to be implemented, scenario T would be proposed for implementation instead. With well-trained servers (*i.e.,* following the defined SBPs), Bragança's AWT would decrease 4:21 minutes and the system's QI would increase 3.7%. Even though it would increase the store's CV, the new QI would yield a benchmark position as the 5th top-ranking EDP store (*cf.* Table 4).

Finally, concerning the hiring, or dismissal, of one server, the recommendation would be scenario V, as it is assessed to provide a more robust (*i.e.,* less variable) solution than W, even though incurring into slightly higher AWT. This way, the store manager would be expected to be more receptive of this new concept. However, a final proposal cannot be done without conducting further analysis. In fact, EDP's unitary improvement valorisation of its KPIs (*i.e.,* the associated financial benefits of increasing the QI by 1%, or decreasing the AWT by 1 minute) must be calculated to assess the worthiness of the involved financial investment. Even though it was inferred, EDP has shown unable to quantify these benefits at the moment. Still, it was shared the annual cost of each employee (*ca.* 28 000€) (Martins, 2017).

For the sake of argument, Table 24 explains that if, on average, a contractual client ('A' ticket) is willing to spend 1.83€ more per each avoided minute of waiting, EDP will profit from hiring this new server.

Table 24: Assessment of the required benefit to break even the investment of hiring a new server

	(1) AWT 'A' Ticket (min.)	**(2) AWT decrease (min.)**	**(3) Number of 'A' clients**	**(4) Annual Cost of one FTE (€)**	**(5) =(4)/(3) (€/client)**	**(6) =(5)/(2) (€/client/minute)**
Scenario T)	04:10					
Scenario V)	01:29	-02:41				
2016's Data			5 700	28 000€		
Premise					4.91€	1.83€

This hypothesis takes into consideration that the annual number of clients seeking 'A' tickets is about 5 700 (*cf.* Table 7 and Table 5) and that they are all seeking to make a new contract with EDP, or adding a new product to the existing contract (Martins, 2017). Moreover, it considers that the sixth server would further reduce 2:41 minutes of the AWT (column 2) as yielded in scenario V when compared with T's (4:10 minutes). Knowing the previously stated annual cost (column 4) of one server (FTE), each client 'A' would have to pay 4.91€ (column 5) per visit, which, divided by the avoided AWT, yields the previously mention value of 1.83€.

Following this same rationale, the hiring of a sixth server would even be profitable to EDP, if for each reduced minute of AWT, EDP's 'A' customers granted a unitary valorisation greater than 1.83€.

6.5 Chapter’s Conclusions

Throughout this Chapter the several steps of the Framework have been successfully implemented and, as predicted during its ideation (Chapter 5), they have resulted in three robust proposals that will not only increase EDP’s performance, but will also address the specific concerns of Bragança’s store manager.

This framework fostered a process that incites to first plan, then to analyse, and only afterwards to act (*i.e.,* simulate). This proved to be very useful considering that each simulation model, even with just a few conceptual adaptations, may take hours to be built, tested and verified as having a good mimicking behaviour.

Following this mindset, this Chapter began by defining the model’s scope upon several data analyses. Collected data from 2016’s Inline® was compiled and analysed to find a suitable pair of comparable stores. To do so, it was concluded that a clustering technique had to be implemented based on the stores demand and their layout. After identifying two compatible stores (Guarda and Bragança), their main components were analysed and compared. Thanks to a quantitative analysis, stores proved to be comparable and service best-practices were already identified at this early stage. Moreover, relying on Lean’s philosophy, a qualitative analysis was also performed unveiling queue management best-practices by meeting with both stores’ managers. Overall, from these analyses, crucial data was collected to ensure that the framework would lead to a realistic and sharable set of best-practices.

At that moment, with the model’s scope defined, the design of the model took place. Every component property (*i.e.,* parameter) was defined having in mind the available features in Simul8. Assumptions were also documented for future-works. Once everything was fully characterised, and prior to the simulation of new scenarios, the model was tested by running the current state of Bragança’s store. As results indicated a good accuracy (*ca.* 97.5%) the model was considered ready to be implemented and to estimate the potential of the suggested improvements for the system in-scope.

The simulation phase encompassed the testing of 23 scenarios. They were based on either the current Inline® system, or on the identified queue management BP solution (Guarda’s). As the adoption of best-practices proved to lead to better results, Guarda’s concept (the speed-track) was progressively developed until every service station was calling through a triage system that would prioritize certain time-triggered ticket-types. The time-triggering setting also proved to be a robust solution (improving both KPIs), as several time-triggers were then tested and fail to lead to better results. Thus, among this step (scenarios B to S), scenario P was proposed for implementation. The next simulation step assessed the improvement potential upon the adoption of the identified service best-practices. If store employees were trained accordingly, the store would further increase its performance, and then scenario T was also recommended. Still, the scenario with the utmost performance was obtained by simulating scenario V that combined both sets of best-practices (i.e., SBP and queue management BP) with the hiring of a sixth employee. If implemented, the store’s QI would raise to 98.0%, which on its own would mean that Bragança store would be top-ranked among EDP stores in 3rd place (*cf.* Table 4). Consequently, its implementation was also recommendable.

As the proposed scenarios P, T, V encompass the cumulative adoption of new strategies, only scenario V is presented Figure 23. Observing the Simul8 model, it is clear that the improvement opportunities scope focused on the Service Mechanism, as the Queue subsystem cannot be interfered with (in the market's regulatory frame) and the Input Source was out of scope (*cf.* subsection *3.2.5*). Moreover, Resources' parameters (besides the number of servers) remained unchanged as well, which constitutes a further improvement opportunity to be developed in future-works.

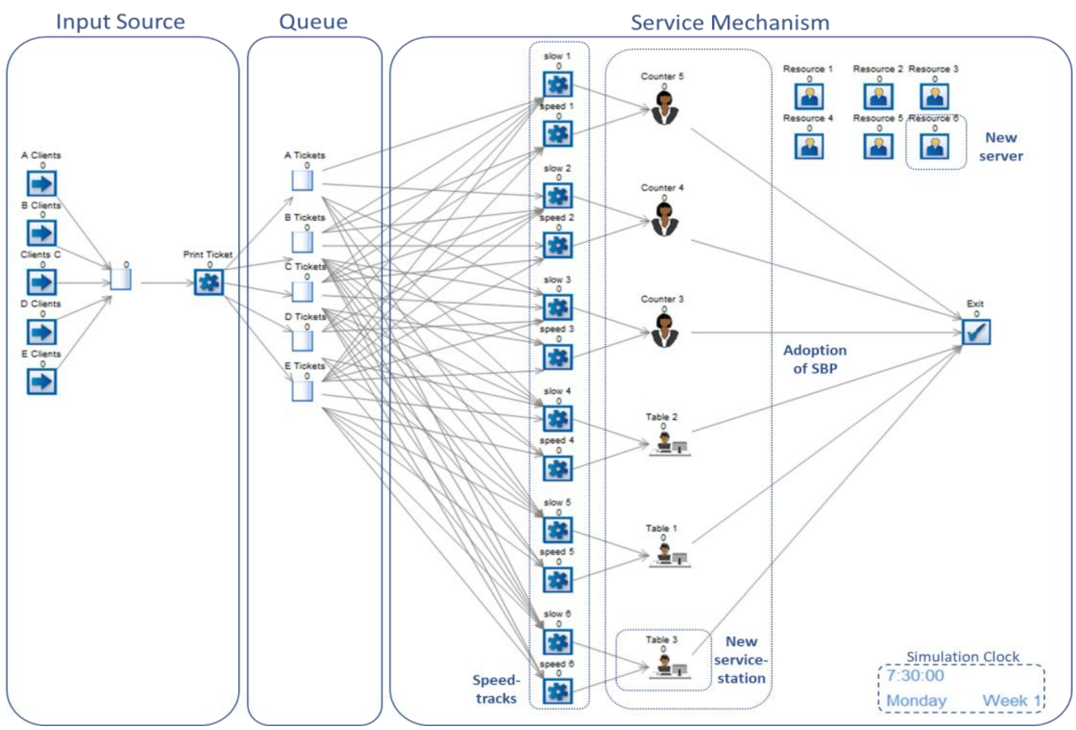

Figure 23: Simul8 Model of Bragança's store (proposed future state – solution V)

This choice of scenario V was largely influenced by Bragança's store manager focus on reducing the variability of the AWTs. After all, if the focus would be to solely reduce the AWT, the recommendation would be scenario W. In fact, since the latter option leads to the biggest improvement of the AWT, in Figure 24, its estimated impact (-76%) is decomposed into the three initiatives.

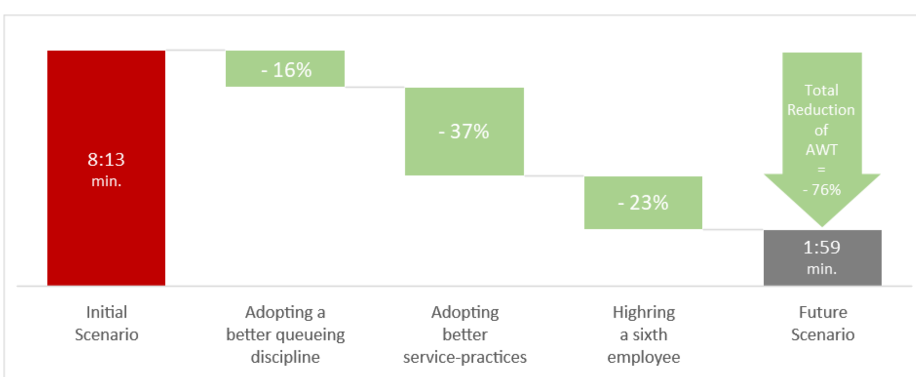

Figure 24: Estimated impact of the Framework in Bragança's AWT

This was done by ignoring possible cumulative symbiotic gains, since between each testing stage only one parameter was being changed (*i.e.,* first the queue discipline, then the AST distribution, and finally the increased service-capacity). As presented in Figure 24, spreading SBPs among employees is the most impactful initiative reducing almost 40% of Bragança's AWT. Even so, the simple adoption of a better Queuing discipline (with the *speed-tracks*) can reduce the AWT by 16% and, if automatized, it will require no effort from the employees (*i.e.,* potentiating an immediate payback). Finally, the hiring of a new employee would prove to reduce the remaining percentage (23%). The design of scenario W may be observed in Figure 25, where a similar model to the one of Figure 23 was created with the difference of having the sixth service-station calling by a pure FIFO strategy.

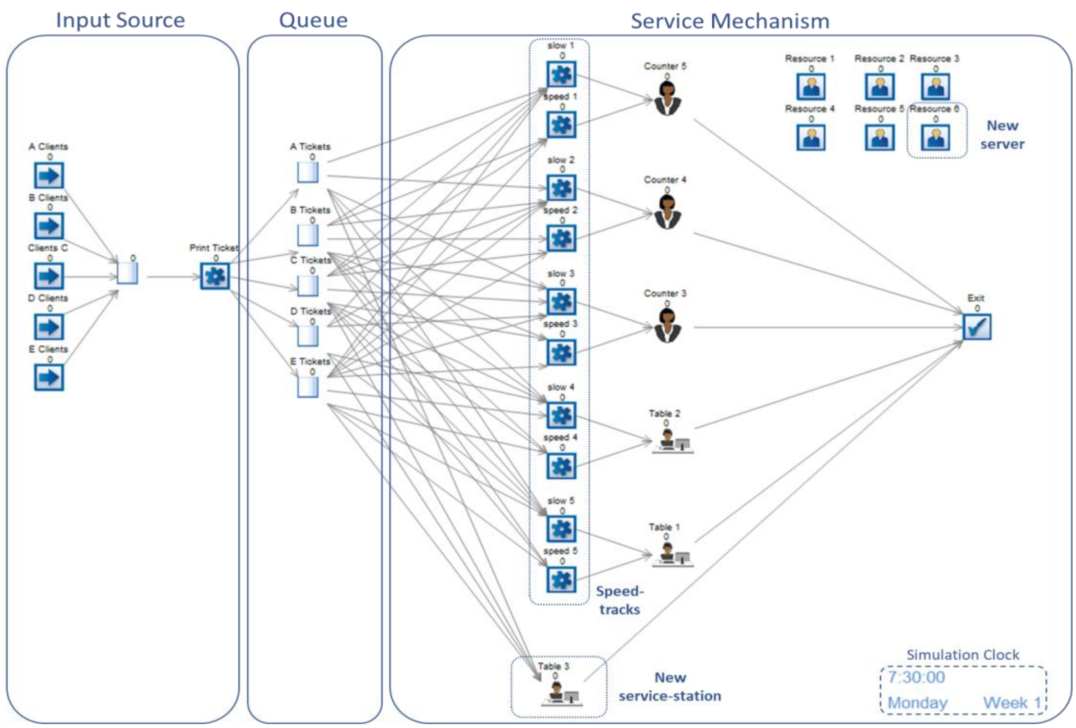

Figure 25: Simul8 Model of Bragança's store (simulation W)

Overall, whether they are, or not, applied in the sixth service-station, the *speed-tracks* have an unquestionable impact on the system's improvement. In fact, this concept of the *speed-tracks* resorting to *time-triggers* has already been mentioned in some works where it is even stated that relying solely on vague KPIs like the QI may not communicate useful information, as it just yields the number of customers who have waited less than a specific time, whereas methodologies such as the one used in this Dissertation aim at identifying, or at least estimating, the acceptable waiting time per ticket-type, and use it to call customers who are about to reach that limit (Legros, 2016).

Therefore, by adopting this triage solution a customer will be called to a station, according not only to his ticket-type, but also to his WT. Being more flexible than the system in-place, it eliminates the five rigid and distinct queues per ticket-type model and, in its place, introduces *fast-* and *slow-tracks* so that, at any moment, the station may call the most urgent ticket (*i.e.,* almost reaching, or having passed the time-trigger). In fact, their implementation has been able to keep under control the stores AWT as proved by scenarios G, H and L to W.

Moreover, further analysing the results, it is striking the existing improvement opportunities, that, at the same time, are hampered by the difficult trade-off caused by the store's manager concern in reducing the number of complaints (*cf.* subsection 6.1.3.2). After all, the system in-place is well designed to his intentions being difficult to achieve a lower CV. Still, two solutions (P and U) were able to further reduce it. It is important to notice that, even though it is not in EDP's strategy, this careful analysis of the CV will be crucial when implementing the solution de facto.

It is also interesting to realise the uncovered improvement potential that does not imply any significant investment. In fact, just by changing among several queuing disciplines, the QI would increase as much as 3.4% (scenario J) and reduce the AWT by 15.9% (scenario Q). Similarly, the adoption of services best-practice that only needs to set a training-day for the employees could decrease the AWT by more than half (*i.e.,* 4:21 minutes in scenario T). And even the most expensive solution could be paid-off by the impact it has on the AWT of the 'A' tickets (*cf.* subsection *6.4.2*).

Concluding this Chapter, Table 25 summarises the three proposed scenarios with their respective estimated impacts that have previously been discussed.

Table 25: Recommended solutions for adoption

Recommended Solutions	AWT	%Var. AWT	QI	%Var. QI	CV	Required Investment
(1) Adoption of BP (queuing discipline)	06:56	-15.6%	95.4%	+3.4%	46.6	N/A
(2) Adoption of BP (services) + (1)	03:52	-53.0%	96.0%	+4.0%	82.3	Low (1-day training)
(3) Hiring a sixth server + (2)	02:40	-67.6%	98.0%	+6.2%	114.4	28 000€/year

The next Chapter will conclude this Dissertation by summarizing the takeaways of this Dissertation in terms of literature research and feasibility of the proposed Framework.

7 Dissertation Conclusions

This Dissertation's last Chapter will summarize the work convened in the previous six chapters and reflect on the main learnings that can be taken by reading them. Thus, it will begin by outlining the theoretical notions that proved to be crucial when implementing the proposed Framework. A focused critique of the Framework's implementability will follow, stressing its main features and results, together with unveiled future work opportunities. Finally, an assessment of the Dissertation's impact on EDP's stores system will be shared, predicting what would be its impact if implemented on the entire EDP system.

7.1 Theoretical Takeaways

After describing EDP's store system status and its two main challenges (regulatory frame and competition), it was clear that in order to succeed, EDP would have to increase its stores operational performance. This goal would not only increase the customers' satisfaction-levels, but also guarantee that the new solution would maintain a strict fulfilment of the regulations of the Portuguese Energy sector.

The system was described and characterised by Queuing Theory, having in mind Newell's (2013) remark on the academia's excessive concern on designing complex models that often lack pragmatism when they seek an optimal solution. Each one of its components was studied and classified according to the literature findings. To guarantee that, at each step, the most appropriate classification was chosen to each component, a broad review of the several interpretations and problem-dimensions was convened and discussed considering a total of 109 works throughout the process.

When analysing the potential impact of Queuing Theory on the system on-hands by comparing with existing studies, it was concluded that proceeding with mathematical modelling the current queue system (M/M/C/k/PD) would lead to unmanageable results according to Brandwajn & Begin (2017). However, Simulation would not only be able to model it considering most of its components' specificities, but would also lead to better decision-making process by testing and assessing several scenarios (Pidd, 2004).

Following these recommendations, a second literature review began examining a particular simulation approach – Discrete Event Simulation (DES). DES models were specially developed to mimic dynamic and stochastic systems whose variables change at discrete moments and may be analysed quantitatively (Banks *et al.*, 2005). This fits EDP stores, as the system variables (*e.g.,* number of customers entering the store) are not only varying randomly along the day, but are being recorded by Inline® system.

At that moment, knowing how to model and to assess the system's performance according to a specific scenario, the framework was lacking objective and proved academic concepts on how to uncover current best-practices, raising the implementability of those scenarios as they were already being implemented in comparable stores. Lean became acknowledged as being able to foster processes that: decrease the systems variability (Liker & Morgan, 2006), crucial considering KPIs like the QI; increase employees' motivation to accept change, important since it would be the store staff who ultimately was going to implement the solution; and to continuously improve the current standards, which would guarantee a sustainable long-term improvement of EDP's performance (Baril *et al.*, 2016).

Overall, this sequential research of the three areas proved to have a symbiotic effect. In retrospective, this structure has followed a LAMBDA cycle (Ward & Sobek II, 2014), without any premeditated intention. In fact, it began by describing the system (Look); secondly, it sought to understand it through Queuing Theory (Ask); thirdly, it explained how to model a DES system (Model); fourthly, it focused on how to foster the employees' involvement and brainstorming (Discuss); and finally, it enabled to convene all the steps in a tailored framework that will be examined next (Act).

7.2 Implementation Results of the Framework and proposed Future Work

The developed Framework proved its feasibility by successfully achieving the aspired goals of this Dissertation (section 1.2). Being supported by a broad literature review previously convened, every system component was carefully characterised in light of those findings. Despite encompassing some assumptions, the computed parameters of the model have been able to lead to an accuracy rate of about 97.5% when simulating the real scenario of 2016, with approximately half (*ca.* 2.5%) of the usual 5% error-margin benchmark in most simulations (Sargent, 2013).

In addition to accurately modelling the current system, the Framework discovered best-practices thanks to quantitative and qualitative analyses. The latter was used to understand the different queuing disciplines in-place and EDP's corporate view on the identified differences, whereas the former was used to identify the best service practices among the two stores environments. After confirming that both stores were comparable by clustering the data and choosing two similar stores, the identified best-practices were fundamental to develop the simulated scenarios. Having studied the simulation results, three solutions were proposed anticipating three different scenarios of financial budget allocations for the project.

The first recommendable solution requires no investment and can be promptly implemented, as it solely consists on changing the queuing discipline. Thanks to the inclusion of the *speed-track* priority discipline with tailored time-triggers for each prioritised ticket, this initiative is estimated to increase the system's QI by 3.1%, and to decrease its AWT by 1:17 minutes while still being able to reduce the coefficient of vacation among different ticket-types (-16%). The solution proved to lead to robust results and only if it were to be programmed into the Inline® system, would it imply an investment. However, considering that nowadays the store is already calling manually, there is no indication that there will be a need for developing the system.

The second proposal would only require small investment (one-day training), since the most critical part (SBP identification) has been already conducted. Thus, by organizing a short session where all the employees of both stores would gather, map and share the identified service best-practices, the stores' QI would further increase 0.7% making Bragança EDP's store with the 5th highest QI in the country. Furthermore, it would decrease the AWT by an extra of 3:04 minutes.

The third recommendation would include the hiring of one employee, and, therefore, it would require the biggest investment among the three recommendations (*cf.* Table 25). Leading to the biggest increase of QI (+5.7%) and the second biggest decrease of AWT (-5:33 minutes), this addition of a new server also has the particularity of having asserted what appears to be the turning point of the *speed-track* solution. In fact, if the sixth service-station resort to a pure FIFO strategy (scenario W), the AWT would decrease more than when following a speed-track solution (scenario V). Yet, as latter option provides a more robust performance with a higher QI and a smaller CV, its implementation is the one proposed in this analysis.

In the two last budget scenarios, it is must be stressed that the store manager would need to be involved at every step of the implementation process, so he would understand the improvement potential. After all, they will both slightly increase the Bragança's AWT CV, which is not aligned with his current personal strategy that he tailored to the store's environment (*cf.* subsection *6.1.3.2*).

Overall, the Framework has fit to the intended purposes and enable a robust analysis of the system and of its performance stimulants. The four sequential stages were followed leading to a focused process that yielded on impactful initiatives to be adopted by EDP.

Throughout this Framework, some opportunities for future work were also uncovered. In fact, they can even be categorized as: model improvements, implementation of the designed VSM, and assessment of the frameworks' implementation profitability.

Regarding the model, some further considerations can be made to further widen its improvement scope.

- Firstly, it would be interesting to analyse the impact of the cancelled tickets on the system's performance and the corresponding increase of the server's idleness (*i.e.,* decrease of his efficiency), since whenever a server calls a customer who has already left the store, he must wait awhile before cancelling the ticket. This could even be further analysed by analysing the different existing reneging behaviours in-system (*cf.* subsection 4.1.3.2) and its causes, adapting the previously mentioned time-triggers towards the minimization of abandonment ratio (*i.e.,* percentage of clients who have exited the store before being called due to excessive waiting, or even due to the number of people in-system).
- Secondly, if Inline® would begin to record individual NPS for a specific amount of time, not only would it be possible to perform Platz & Østerdal's (2017) suggestion of assessing FIFO's impact on the customers' welfare, but it would also allow a comparison of the different simulated queuing disciplines' impact on the stores NPS (*i.e.,* the addition of a new KPI to the model).
- Thirdly, the model's scope could be broaden, unveiling more improvement opportunities through: the inclusion of the before-opening and after-closing periods, together with their encompassed activities, to improve the efficiency level of the employees; the modelling of the printing activity and its queue, to better describe the entire customer journey in-store; the rethinking of the demand distribution, by creating appointment-only services, or by advertising the less congested periods; and the reallocation of the lunch-breaks, to ascertain its potential impact on the AWT during lunch-time.

Concerning the VSM technique that was detailed on the methodology, but not considered in the framework's scope due to its time-requirements, future work should focus on the improvement of each service process, identifying the NVA activities they comprise. Considering that the adoption SBPs (following the same SOPs) would decrease the AWT by 37%, it is admissible to foresee that the design of more efficient processes would further improve the store's KPIs and their variability (Wang & Disney, 2016).

Finally, about the assessment of this Framework's financial impact, further inquiries could be made assessing the actual valorisation that EDP clients grant to a decrease of their AWT, and its estimated impact on EDP's ability to sell services, hence increasing its revenues (Cayirli & Veral, 2003; Lakshmi & Iyer, 2013).

To conclude the appraisal of this work it was decided to also implement this proposed framework in Guarda's store, as the quantitative analyses of the services best-practices unveiled a potential improvement in that store too (*cf.* subsection 6.1.3.1). Both the same steps and assumptions, as the ones detailed in Chapter 6, were followed during this inquiry leading to the creation of the store's model (Figure 26) with an accuracy rate of approximately 97%. The results estimate a decrease of the AWT of 13.3% (-30 seconds) and an increase of Guarda's QI of 1.5% (where only 0.01% of the customers would have to endure waiting-times greater than 20 minutes).

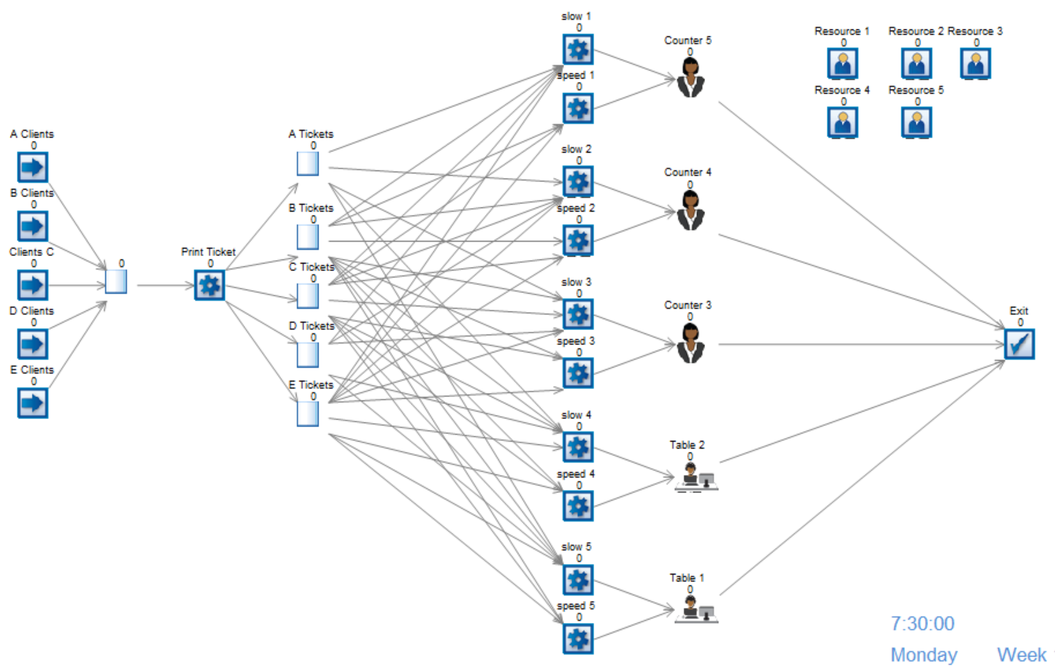

Figure 26: Simul8 Model of Guarda's store (with adoption of SBPs and four new speed-tracks)

This last effort of seeking to implement the Framework and validate both its worthiness and its applicability to any EDP store (or similar system), proved to be successful. Indeed, the proposed Framework was able to assess the improvement potential of a store. Despite being a top-performing store (*cf.* Table 4 where Guarda is ranked 1st among EDP stores) it still had some margin to improve and to potentially become even better with lower AWTs and a higher QI. After all, following Lean's advocated principle, the improvement of every organization should be continuous, not settling for the *status-quo* (*cf.* section 4.3).

7.3 Impact of this Dissertation

This Dissertation confirmed the existence of improvement opportunities in EDP stores system. Having an annual demand of more than 2 million visitors, any performance improvement will have a real impact on EDP's customers satisfaction levels and their leaning to acquire services.

Throughout this work, solutions to decrease the customers' perceived time in-store proved to be successful by reducing the AWT (Legros, 2016). Furthermore, they were able to increase the stores QI, a KPI that is monitored by the Energy Market regulator – ERSE.

Overall, customers will become more satisfied, as the unintended consequences of queuing will be mitigated (Mann, 1970). In fact, the impact of the uncovered solutions is forecasted to be witnessed on the short-term (short payback), since they will simultaneously improve the customers' journey and the store's performance. Thus, the NPS of any store that follows this framework is expected to raise after improving their metrics (as witnessed in Guarda's example, *cf.* subsection 6.1.3.2). This may convert itself to more sales, as the customer's willingness to pay increases with lower queuing (Li & Hensher, 2011), hence generating a higher annual revenue in EDP stores, where, on average, one residential customer (classified as a BTN client) represents a yearly profit after taxes of 50€ (Brandão, 2005).

Considering that the Dissertation focused on concept-proving its Framework on one store (Bragança's), but still assessed the same impact on Guarda's, an overall improvement of EDP's stores system is expected to be reached when this methodology begins to be implemented on the remaining 39 stores. This extension is expected to be feasible, since EDP's management team has been involved in the project and is already adopting some of its findings (scenario P in Bragança's store, for now). Following Lean's Principles (*cf.* subsection *4.3.2.5*), EDP has been participating in both formal and informal meetings, where future-state ambitions, or constraints, were shared. Being part of its process design, having understood and agreed with its solutions, and being confronted with its simulated results, EDP trusts the recommendations and is expected to implement them in the near-future, especially, as most of these solutions have a critical impact on all the KPIs, but do not imply any significant financial, nor time, investment. Whilst the two first solutions (*cf.* Table 25) convene the adoption of existing best-practices, only the third suggests hiring a new server. However, even this one may constitute a profitable investment upon analysing the impact of decreasing waiting-times in EDP's ability to sell. Even so, it is plausible that this last suggestion is not implemented due to the difficulty on proving its profitability. This should not be discouraging, though, as from inception the sensitivity analysis of hiring, or dismissing someone, was conceived as being merely advisory, aiming at forecasting what would be the financial effort required to meet a specific AWT target.

To conclude, it should be stated how satisfactory it was to develop this topic, and how motivating it is to see that this Dissertation is, not only, feasible, but implementable. Especially when it will clearly impact the experience of millions of customers, who annually visit EDP stores, by reducing their average waiting time.

Bibliography

Al Hamadi, H. M., Sangeetha, N., & Sivakumar, B. (2015). Optimal control of service parameter for a perishable inventory system maintained at service facility with impatient customers. *Annals of Operations Research, 233*(1), 3-23. doi:10.1007/s10479-014-1627-1

Al Hanbali, A., Alvarez, E. M., & van der Heijden, M. C. (2015). Approximations for the waiting-time distribution in an M/PH/c priority queue. *OR Spectrum, 37*(2), 529-552. doi:10.1007/s00291-015-0388-9

Almási, B., Roszik, J., & Sztrik, J. (2005). Homogeneous finite-source retrial queues with server subject to breakdowns and repairs. *Mathematical and Computer Modelling, 42*(5-6), 673-682. doi:10.1016/j.mcm.2004.02.046

Angel, E., Bampis, E., & Pascual, F. (2008). How good are SPT schedules for fair optimality criteria. *Annals of Operations Research, 159*(1), 53-64. doi:10.1007/s10479-007-0267-0

Argon, N. T., Deng, C., & Kulkarni, V. G. (2016). Optimal control of a single server in a finite-population queueing network. *Queueing Systems*, 1-24. doi:10.1007/s11134-016-9507-9

Asmussen, S. (2008). *Applied probability and queues* (Vol. 51). Springer Science & Business Media. ISBN:978-0-387-21525-9

Baines, T. S., Asch, R., Hadfield, L., Mason, J. P., Fletcher, S., & Kay, J. M. (2005). Towards a theoretical framework for human performance modelling within manufacturing systems design. *Simulation Modelling Practice and Theory, 13*(6), 486–504. doi:10.1016/j.simpat.2005.01.003

Balci, O. (2003). Verification, validation, and certification of modeling and simulation applications: verification, validation, and certification of modeling and simulation applications. *Proceedings of the 35th conference on Winter simulation: driving innovation* (pp. 150-158). Winter Simulation Conference. doi:10.1109/WSC.2003.1261418

Banks, J., CARSON II, J. S., & Barry, L. (2005). *Discrete-event system simulation* (4th ed.). Pearson. ISBN:978-0136062127

Baril, C., Gascon, V., Miller, J., & Côté, N. (2016). Use of a discrete-event simulation in a Kaizen event: A case study in healthcare. *European Journal of Operational Research, 249*(1), 327-339. doi:https://doi.org/10.1016/j.ejor.2015.08.036

Baumann, H., & Sandmann, W. (2017). Multi-server tandem queue with Markovian arrival process, phase-type service times, and finite buffers. *European Journal of Operational Research, 256*(1), 187-195. doi:10.1016/j.ejor.2016.07.035

Beckers, D., Wijnhoven, F., & Amrit, C. (2016). *Reducing waste in administrative services with lean principles.* AIS. Retrieved from http://essay.utwente.nl/68460

Bhattacharjee, P., & Ray, P. K. (2014). Patient flow modelling and performance analysis of healthcare delivery processes in hospitals: A review and reflections. *Computers & Industrial Engineering, 78*, 299-312. doi:10.1016/j.cie.2014.04.016

Bicheno, J., & Holweg, M. (2000). *The lean toolbox* (Vol. IV). Buckingham: PICSIE books. ISBN:978-0954124458

Boots, N. K., & Tijms, H. (1999). A multiserver queueing system with impatient customers. *Management Science, 45*(8), 444-448. doi:10.1287/mnsc.45.3.444

Bortolotti, T., & Romano, P. (2012). Lean first, then automate: a framework for process improvement in pure service companies. A case study. *Production Planning & Control, 23*(7), 513-522. doi:10.1080/09537287.2011.640040

Braga, L. (21 de August de 2017). Bragança's store - Day-to-day Operations. (C. G. Almeida de Frias, Entrevistador)

Brandão, N. (2005). *Impact on profitability of the Energy Market regulation.* Lisbon: EDP. Obtido em 9 de September de 2017, de http://www.erse.pt/pt/consultaspublicas/historico/documents/cp_09/09_3/sess%C3%A3op%C3%BAblicanbv19052005.pdf

Brandwajn, A., & Begin, T. (2017). Multi-server preemptive priority queue with general arrivals and service times. *Inria-Research Centre Grenoble–Rhône-Alpes*. Retrieved from https://hal.inria.fr/hal-01515328/document

Breinbjerg, J., Sebald, A., & Østerdal, L. P. (2016). Strategic behavior and social outcomes in a bottleneck queue: experimental evidence. *Review of Economic Design, 20*(3), 207-236. doi:10.1007/s10058-016-0190-4

Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S., & Zhao, L. (2005). Statistical analysis of a telephone call center: A queueing-science perspective. *Journal of the American statistical association, 100*(469), 36-50. doi:10.1198/016214504000001808

Cardoen, B., Demeulemeester, E., & Beliën, J. (2010). Operating room planning and scheduling: A literature review. *European Journal of Operational Research, 201*(3), 921-932. doi:10.1016/j.ejor.2009.04.011

Caulkins, J. P. (2010). Might randomization in queue discipline be useful when waiting cost is a concave function of waiting time? *Socio-Economic Planning Sciences, 44*(1), 19-24. doi:10.1016/j.seps.2009.04.001

Cayirli, T., & Veral, E. (2003). Outpatient scheduling in health care: a review of literature. *Production and operations management, 12*(4), 519-549. doi:10.1111/j.1937-5956.2003.tb00218.x

Chiang, W. C., & Urban, T. L. (2006). The stochastic U-line balancing problem: A heuristic procedure. *European Journal of Operational Research, 175*(3), 1767-1781. doi:10.1016/j.ejor.2004.10.031

Creemers, S., & Lambrecht, M. (2010). Queueing models for appointment-driven systems. *Annals of Operations Research, 178*(1), 155-172. doi:10.1007/s10479-009-0646-9

Daduna, H., & Schassberger, R. (1981). A discrete-time round-robin queue with Bernoulli input and general arithmetic service time distributions. *Acta Informatica, 15*(3), 251-263. doi:10.1007/BF00289264

Daellenbach, H., McNickle, D., & Dye, S. (2012). *Management science: decision-making through systems thinking.* Palgrave Macmillan. ISBN:978-1403941749

Detty, R. B. (2000). Quantifying benefits of conversion to lean manufacturing with discrete event simulation: a case study. *International Journal of Production Research, 38*(2), 429-445. doi:10.1080/002075400189509

EDP. (2015). *Energy with intelligence - Dados Ibéricos.* Lisboa: EDP - Energias de Portugal, SA.

EDP. (2016). *Annual Report.* Lisboa: Energias de Portugal, S.A. Retrieved April 12, 2017, from http://www.edp.pt/en/Investidores/publicacoes/relatorioecontas/2016/Company%20reports%202016/Annual%20Report%202016.pdf

EDP. (2017). *Energy Sources*. Obtido em 11 de March de 2017, de EDP: https://energia.edp.pt/particulares/apoio-cliente/origem-energia/

Engström, T., Jonsson, D., Medbo, L., & Engström, T. J. (2006). Production model discourse and experiences from the Swedish automotive industry . *International Journal of Operations & Production Management, 16*(2), 141–158. doi:10.1108/01443579610109893

Erlang, A. K. (1918). Solutions of some problems in the theory of probabilities of some significance in automatic telephone exchanges. *The Post Office Electrical Engineers' Journal, 10*, 189-197.

ERSE. (2017). *Comercializadores de Energia*. Retrieved April 23, 2017, from http://www.erse.pt/pt/gasnatural/agentesdosector/comercializadores/Paginas/Residenciais.aspx

ERSE. (2017). *ERSE Mission*. Retrieved April 23, 2013, from http://www.erse.pt/pt/aerse/missao/Paginas/default.aspx

European Commission. (2016). *Clean Energy For All Europeans.* Brussels. Retrieved from http://eur-lex.europa.eu/resource.html?uri=cellar:fa6ea15b-b7b0-11e6-9e3c-01aa75ed71a1.0001.02/DOC_1&format=PDF

Fisher, M. L. (1981). The Lagrangian relaxation method for solving integer programming problems. *Management Science, 27*(1), 1-18. doi:10.1287/mnsc.1040.0263

Foster, E., Hosking, M., & Ziyaa, S. (2010). A Spoonful of Math Helps the Medicine Go Down: An Illustration of How Healthcare can Benefit from Mathematical Modeling and Analysis. *BMC Medical Research Methodology, 10*(1). doi:10.1186/1471-2288-10-60

Gautam, N. (2008). *Queueing theory, Operations Research and Management Science Handbook.* London: CRC Press, Taylor and Francis Group. ISBN:978-1439806586

Gómez-Corral, A. (2002). Analysis of a single-server retrial queue with quasi-random input and nonpreemptive priority. *Computers & Mathematics with Applications, 43*(6), 767-782. doi:10.1016/S0898-1221(01)00320-0

Gonçalves, H. (24 de August de 2017). Guarda's store - Day-to-day Operations. (C. G. Almeida de Frias, Entrevistador)

Grishechkin, S. A. (1991). Branching processes and queueing systems with repeated orders or with random discipline. *Theory of Probability & Its Applications,, 35*(1), 38-53. doi:10.1137/1135004

Gross, D., & Harris, C. M. (1998). *Fundamentals of queueing theory.* ISBN:978-0471791270

Gutacker, N. S. (2016). Waiting time prioritisation: Evidence from England. *Social Science & Medicine*, 140-151. doi:10.1016/j.socscimed.2016.05.007

Harchol-Balter, M., Osogami, T., Scheller-Wolf, A., & Wierman, A. (2005). Multi-server queueing systems with multiple priority classes. *Queueing Systems, 51*(3-4), 331-360. doi:10.1007/s11134-005-2898-7

Hernández, G. (2001). Discrete model for fragmentation with random stopping. *Physica A: Statistical Mechanics and its Applications, 300*(1), 13-24. doi:10.1016/S0378-4371(01)00343-0

Hillier, F. S., & Boling, R. W. (1967). Finite queues in series with exponential or Erlang service times—a numerical approach. *Operations Research, 15*(2), 286-303. doi:10.1287/opre.15.2.286

Hillier, F. S., & Lieberman, G. J. (2010). *Introduction To Operations Research* (9th ed.). McGraw-Hill. ISBN:978-0-07-337629-5

Hines, P., Martins, A. L., & Beale, J. (2008). Testing the boundaries of lean thinking: observations from the legal public sector. *Public money and management*, 35-40. doi:http://dx.doi.org/10.1111/j.1467-9302.2008.00616.x

IEA. (2016). *World Energy Outlook 2016.* International Energy Agency. ISBN:978-92-64-26495-3

Imai, M. (2012). *Gemba Kaizen.* McGraw-Hill Education. ISBN:978-0071790352

IMF. (2016). *The World Economic Outlook.* International Monetary Fund.

Jaiswal, N. K. (1968). *Priority queues* (Vol. 50). New York: Academic Press. ISBN:978-0-08-095558-2

Joel, Z. L., Louis, N. W., & Chuan, T. S. (2000). Discrete–event simulation of queuing systems. *Proceedings of the Sixth Youth Science Conference*, 1-5. doi:10.1.1.118.118

Keiningham, T. L., Aksoy, L., Cooil, B., Andreassen, T. W., & Williams, L. (2008). A holistic examination of Net Promoter. *Journal of Database Marketing & Customer Strategy Management, 2*, 79-90. doi:10.1057/dbm.2008.4

Kendall, D. G. (1953). Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. *The Annals of Mathematical Statistics*, 338-354. doi:10.1214/aoms/1177728975

Khinchin, A. Y., Andrews, D. M., & Quenouille, M. H. (2013). *Mathematical methods in the theory of queuing.* Courier Corporation. ISBN:978-0486490960

KICG. (2017a). *About Us: Kaizen Institute Consulting Group*. Retrieved March 10, 2017, from Kaizen Institute Consulting Group: https://www.kaizen.com/about-us/kaizen-institute.html

KICG. (2017b). *KBS Introduction.* Porto: Kaizen Institute Consulting Group.

KICG. (2017c). *Foundations & Awareness.* Porto: Kaizen Institute Consulting Group.

KICG. (2017d). *Growth Model.* Porto: Kaizen Institute Consulting Group.

KICG. (2017e). *Q.C.D. Model.* Porto: Kaizen Institute Consulting Group.

KICG. (2017f). *Motivation Model.* Porto: Kaizen Institute Consulting Group.

Lakshmi, C., & Iyer, S. A. (2013). Application of queueing theory in health care: A literature review. *Operations research for health care, 2*(1), 25-39. doi:10.1016/j.orhc.2013.03.002

Lam, S. S. (1977). Queuing networks with population size constraints. *IBM Journal of Research and Development, 21*(4), 370-378. doi:10.1147/rd.214.0370

Langaris, C. (1999). Markovian polling systems with mixed service disciplines and retrial customers. *Sociedad de Estadística e Investigación Operativa, 7*(2), 305-322. doi:10.1007/BF02564729

Lariviere, M. A., & Van Mieghem, J. A. (2004). Strategically seeking service: How competition can generate Poisson arrivals. *Manufacturing & Service Operations Management, 6*(1), 23-40. doi:10.1287/msom.1030.0030

Law Decree nr. 455/2013, 455 (Diário da República: 2ª Série November 29, 2013). Retrieved from https://dre.pt/

Law Decree nr. 58/2016, 165 (Diário da República: 1ª Série August 29, 2016). Retrieved from https://dre.pt/

Law Decree nr. 97/2002, 86 (Diário da República: 1ª Série-A April 12, 2002). Retrieved from https://dre.pt/

Legros, B. (2016). Unintended consequences of optimizing a queue discipline for a service level defined by a percentile of the waiting time. *Operations Research Letters, 44*(6), 839-845. doi:10.1016/j.orl.2016.10.011

Lehaney, B., Clarke, S. A., & Paul, R. J. (1999). A case of an intervention in an outpatients department. *Journal of the Operational Research Society, 50*(9), 877-891. doi:10.1057/palgrave.jors.2600796

Li, Z., & Hensher, D. A. (2011). Crowding and public transport: A review of willingness to pay evidence and its relevance in project appraisal. *Transport Policy, 18*(6), 880-887. doi:10.1016/j.tranpol.2011.06.003

Liker, J. K., & Morgan, J. M. (2006). The Toyota way in services: the case of lean product development. *The Academy of Management Perspectives, 20*(2), 5-20. doi:10.5465/AMP.2006.20591002

Likhanov, N., Tsybakov, B., & Georganas, N. D. (1995). Analysis of an ATM buffer with self-similar (" fractal") input traffic. *IEEE, 3*, 985-992. doi:10.1109/INFCOM.1995.515974

Maister, D. H. (1984). *The psychology of waiting lines.* Harvard Business School. Retrieved from http://www.columbia.edu/~ww2040/4615S13/Psychology_of_Waiting_Lines.pdf

Mann, L. (1970). The social psychology of waiting lines. *American Scientist, 58*(4), 390–398. doi:10.2307/27829163

Martins, B. (1 de September de 2017). EDP's corporate view on the two different management strategies. (C. G. Almeida de Frias, Entrevistador)

Mason, S., Baines, T., Kay, J. M., & Ladbrook, J. (2005). Improving the design process for factories: Modeling human performance variation. *Journal of Manufacturing Systems, 24*(1), 47–54. doi:10.1016/S0278-6125(05)80006-8

Masud, A. M., & Ravindran, A. R. (2008). *Operations Research and Management Science Handbook.* Taylor & Francis Group. ISBN:978-0849397219

McDonald, T., Van Aken, E. M., & Rentes, A. F. (2002). Utilising simulation to enhance value stream mapping: a manufacturing case application. *International Journal of Logistics, 5*(2), 213-232. doi:10.1080/13675560210148696

McGregor, D., & Cain, M. (2004). *An Introduction to SIMUL8.* University of Canterbury. Retrieved from http://www.mang.canterbury.ac.nz/courseinfo/msci/msci480/mcgregor_cain2004/webpage/simul8.pdf

Melton, T. (2005). The Benefits of Lean Manufacturing. *Chemical Engineering Research And Design, 83(6)*, 662-673. doi:http://dx.doi.org/10.1205/cherd.04351

Montgomery, D., & Woodall, W. (2008). An Overview of Six Sigma. *International Statistical Review, 76*(3), 329-346. doi:http://dx.doi.org/10.1111/j.1751-5823.2008.00061.x

Morozov, E., & Phung-Duc, T. (2017). Stability analysis of a multiclass retrial system with classical retrial policy. *Performance Evaluation*. doi:http://dx.doi.org/10.1016/j.peva.2017.03.003

Mostafa, S., Dumrak, J., & Soltan, H. (2015). Lean maintenance roadmap. *Procedia Manufacturing*(2), 434-444. doi:10.1016/j.promfg.2015.07.076

Nah, F. F. (2004). A study on tolerable waiting time: how long are web users willing to wait? *Behaviour & Information Technology, 23*(3), 153-163. doi:10.1080/01449290410001669914

Ndukwe, H. C., Omale, S., & Opanuga, O. O. (2011). Reducing queues in a Nigerian hospital pharmacy. *Pharmacy and Pharmacology, 5*(8), 1020-1026. doi:10.5897/AJPP11.015

Newell, C. (2013). *Applications of queueing theory* (Vol. IV). Springer Science & Business Media. ISBN:978-94-009-5972-9

Nosek, R. A., & Wilson, J. P. (2001). Queuing theory and customer satisfaction: a Review of terminology, trends, and applications to pharmacy practice. *Hospital pharmacy, 36*(3), 275-279.

Nova, A. M. (2008). *Apontamentos de Simulação.* Departamento de Engenharia e Gestão.

Ohno, T. (1998). *Toyota Production System: Beyond Large-Scale Production* (1st ed.). Tokyo: Diamond, Inc. doi:0-915299-14-3

Omogbai, O., & Salonitis, K. (2016). A lean assessment tool based on systems dynamics. *49th CIRP Conference on Manufacturing Systems. 50*, pp. 106-111. Elsevier B.V. doi:10.1016/j.procir.2016.11.034

Pardo, M. J., & De la Fuente, D. (2008). Optimal selection of the service rate for a finite input source fuzzy queuing system. *Fuzzy Sets and Systems, 159*(3), 325-342. doi:10.1016/j.fss.2007.05.014

Pidd, M. (2004). *Computer simulation in management science* (5th ed.). Chichester: John Wiley & Sons, Ltd. ISBN:978-0470092309

Piercy, N., & Rich, N. (2009). Lean transformation in the pure service environment: the case of the call service centre. *International Journal of Operations & Production Management, 29*(1), 54-76. doi:10.1108/01443570910925361

Platts. (2016). *Global Energy Company Rankings.* Retrieved April 12, 2017, from https://top250.platts.com

Platz, T. T., & Østerdal, L. P. (2017). The curse of the first-in–first-out queue discipline. *Games and Economic Behavior, 104*, 165-176. doi:10.1016/j.geb.2017.03.004

Pourvaziri, H., & Pierreval, H. (2017). Dynamic facility layout problem based on open queuing network theory. *European Journal of Operational Research, 259*(2), 538-553. doi:10.1016/j.ejor.2016.11.011

Ribeiro, D., Fontes, E., & Martins, D. (2017, March 21). (C. G. Almeida de Frias, Interviewer)

Rivera, L., & Chen, F. F. (2007). Measuring the impact of Lean tools on the cost–time investment of a product using cost–time profiles. *Robotics and Computer-Integrated Manufacturing, 23*(6), 684-689. doi:10.1016/j.rcim.2007.02.013

Rossetti, M. D. (2015). *Simulation modeling and Arena.* John Wiley & Sons. ISBN:978-1-118-60791-6

Rubinstein, R. Y., & Kroese, D. P. (2016). *Simulation and the Monte Carlo method.* John Wiley & Sons. ISBN:978-0-470-17794-5

Sachidananda, M., Erkoyuncu, J., Steenstra, D., & Michalska, S. (2016). Discrete Event Simulation Modelling for Dynamic Decision Making in Biopharmaceutical Manufacturing. *Procedia CIRP, 49*, 39-44. doi:10.1016/j.procir.2015.07.026

Sakuma, Y. (2010). Asymptotic behavior for MAP/PH/c queue with shortest queue discipline and jockeying. *Operations Research Letters, 38*(1), 7-10. doi:10.1016/j.orl.2009.09.011

Sargent, R. G. (2013). Verification and validation of simulation models. *Journal of Simulation, 7*(1), 12–24. doi:10.1057/jos.2012.20

Schriber, T. J., Brunner, D. T., & Smith, J. S. (2012). How discrete-event simulation software works and why it matters. *Winter Simulation Conference* (p. 3). Proceedings of the Winter Simulation Conference. doi:10.1109/WSC.2012.6465274

Shi, S. (2016). Customer relationship and sales. *Journal of Economic Theory, 166*, 483-516. doi:10.2139/ssrn.1829582

Simul8. (2017). *About us: Simul8*. Retrieved June 7, 2017, from Simul8: https://www.simul8.com/about/

Smith, G., Poteat-Godwin, A., Harrison, L. M., & Randolph, G. D. (2012). Applying Lean principles and Kaizen rapid improvement events in public health practice. *Journal of Public Health Management and Practice, 18*(1), 52-54. doi:10.1097/PHH.0b013e31823f57c0

Strang, K. D. (2012). Importance of verifying queue model assumptions before planning with simulation software. *European Journal of Operational Research, 218*(2), 493-504. doi:10.1016/j.ejor.2011.10.054

Sundar, R., Balaji, A., & Kumar, R. S. (2014). A Review on Lean Manufacturing Implementation Techniques. *Procedia Engineering, 97*, 1875-1885. doi:10.1016/j.proeng.2014.12.341

Sztrik, J. (2001). Finite-source queueing systems and their applications. *Formal Methods in Computing, 1*, 7-10. Retrieved from http://irh.inf.unideb.hu/user/jsztrik/research/fsqreview.pdf

Tako, A. A., & Robinson, S. (2010). Model development in discrete-event simulation and system dynamics: An empirical study of expert modellers. *European Journal of Operational Research, 207*(2), 784–794. doi:10.1016/j.ejor.2010.05.011

Tarabia, A. M. (2008). Analysis of two queues in parallel with jockeying and restricted capacities. *Applied Mathematical Modelling, 32*(5), 802-810. doi:10.1016/j.apm.2007.02.014

Tensator Group. (2017). *About Us*. Retrieved April 25, 2017, from Tensator: http://www.tensatorgroup.com/about-us/

The Wind Power. (2017). *Wind Energy Market Intelligence.* Retrieved from http://www.thewindpower.net/operators_en.php

Tomás, E. A. (2017, April 5). EDP Stores - Queue Management and the Inline System. (C. G. Almeida de Frias, Interviewer)

Toriizuka, T. (2001). Application of performance shaping factor (PSF) for work improvement in industrial plant maintenance tasks. *International Journal of Industrial Ergonomics, 28*(3-4), 225–236. doi:10.1016/S0169-8141(01)00036-1

UN. (2017). *The Paris Agreement.* Retrieved from http://unfccc.int/paris_agreement/items/9485.php

Wang, J., Baron, O., & Scheller-Wolf, A. (2015). M/M/c queue with two priority classes. *Operations Research, 63*(3), 733-749. doi:10.1287/opre.2015.1375

Wang, X., & Disney, S. M. (2016). The bullwhip effect: Progress, trends and directions. *European Journal of Operational Research, 250*, 691–701. doi:https://doi.org/10.1016/j.ejor.2015.07.022

Ward, A. C., & Sobek II, D. K. (2014). *Lean product and process development.* Lean Enterprise Institute. ISBN:978-1934109137

White, K. P., & Ingalls, R. G. (2015). Introduction to Simulation. *Winter Simulation Conference (WSC)*, 1741-1755. doi:10.1109/WSC.2009.5429315

Womack, J. P., & Jones, D. T. (2010). *Lean thinking: banish waste and create wealth in your corporation.* Simon and Schuster. ISBN:978-0743249270

Xu, X., Lian, Z., Li, X., & Guo, P. (2016). A Hotelling queue model with probabilistic service. *Operations Research Letters, 44*(5), 592-597. doi:10.1016/j.orl.2016.07.003

Yang, T., & Li, H. (1995). On the steady-state queue size distribution of the discrete-time Geo/G/1 queue with repeated customers. *Queueing systems,, 21*(1), 199-215. doi:10.1007/BF01158581

Yue, D., Zhang, Y., & Yue, W. (2006). Optimal performance analysis of an M/M/1/N queue system with balking, reneging and server vacation. *International Journal of Pure and Applied Mathematics, 28*(1), 101-115.

Zehrer, A., & Raich, F. (2016). The impact of perceived crowding on customer satisfaction. *Journal of Hospitality and Tourism Management, 29*, 88-98. doi:10.1016/j.jhtm.2016.06.007

Zenios, S. A. (1999). Modeling the transplant waiting list: A queueing model with reneging. *Queueing systems, 31*(3), 239-251. doi:10.1023/A:1019162331525

Zhao, Y., & Grassmann, W. K. (1995). Queueing analysis of a jockeying model. *Operations research, 43*(3), 520-529. doi:10.1287/opre.43.3.520

Zhou, W. H., & Wang, A. H. (2008). Discrete-time queue with Bernoulli bursty source arrival and generally distributed service times. *Applied Mathematical Modelling, 32*(11), 2233-2240. doi:10.1016/j.apm.2007.07.014

Appendix

Inline® system

- Enables accurate queue recording and management;

Distributed network system

- Assures the local system viability even when the central server is down;

Inline® Displays

- Connects the Inline® to the TV screens through which the client is called;

Inline® client

- Connects any PC/tablet to the system so that the server can call the next client;

Inline® business analyser

- A tool that allows to intuitively perform data analysis and presentations through pivot tables.

Figure 27: Inline® System Components

Figure 28: Ticket Dispenser

KAIZEN™
INSTITUTE

Figure 29: KICG's logo (KICG, 2017a)

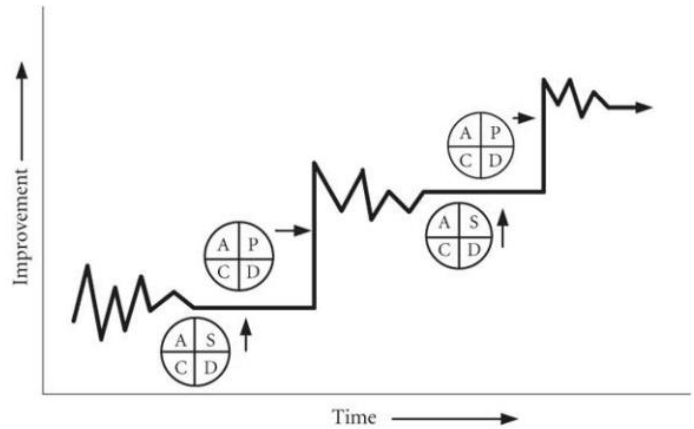

Figure 30: The impact of SDCA/PDCA cycles on the overall improvement of an organization throughout time, retrieved from Imai (2012)

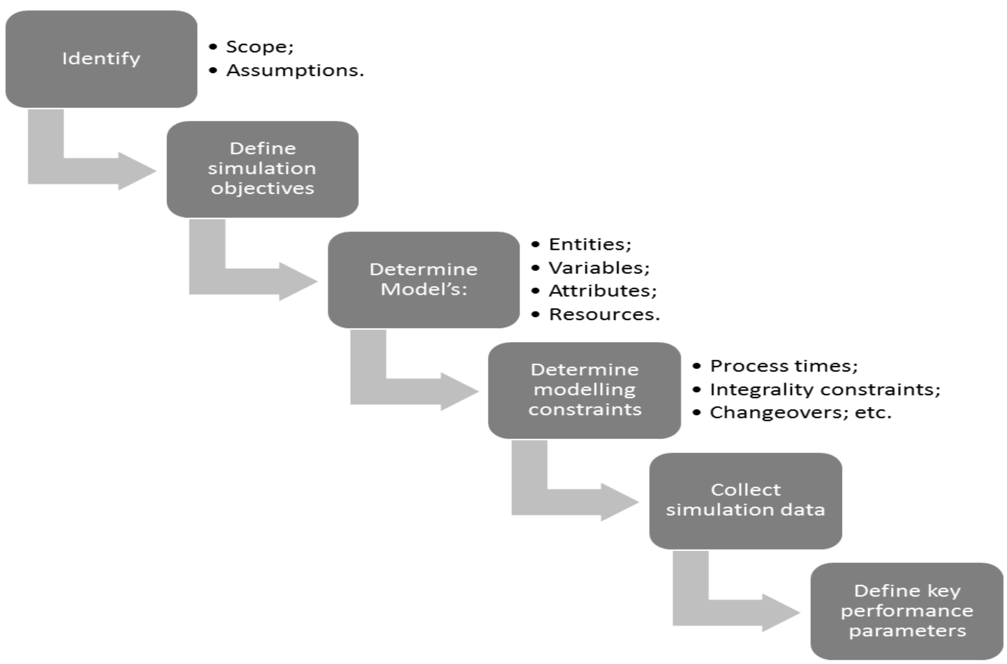

Figure 31: Methodology of a Discrete Event Simulation (DES), adapted from Sachidananda et al. (2016)

Back-office

- Process Name
- Category
- Responsible
- Process description
- Frequency
- Time of day it occurs
- Average Execution Time
- Occupation Factor (0-4)
- Tendency for Errors (0-1)
- Estimated Process Relevance (0-4)
- Opportunities for Improvement

Front-office

- Process Name
- Category
- Process Description
- Average Execution Time
- Opportunities for Improvement

Figure 32: Audit Form of the Undergoing Processes on the Stores

Figure 33: Guarda's store (source: Google Images)

Figure 34: Bragança's store (source: Google Images)

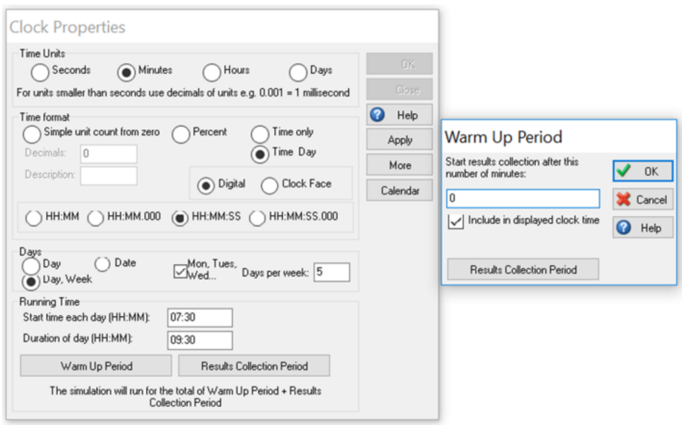

Figure 35: Definition of the Simulation Clock and the Warm Up Period in Simul8

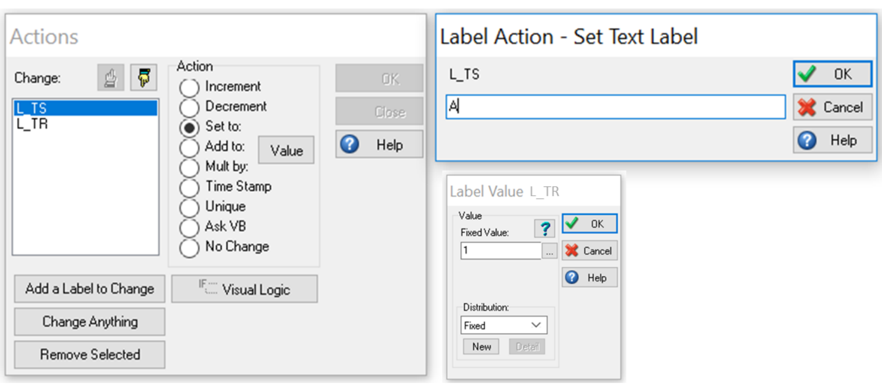

Figure 36: Definition of the L_TS and L_TR labels in start-block 'A' in Simul8

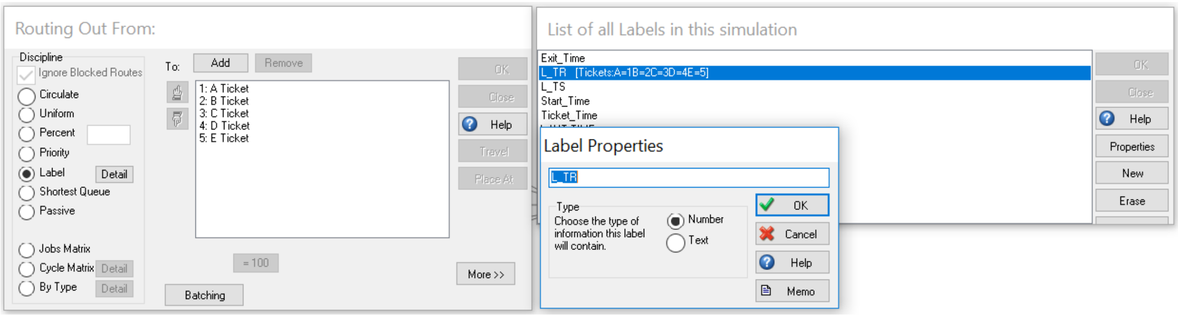

Figure 37: Setting the Print Ticket activity routing-out properties according to the embedded value of L_TR in Simul8

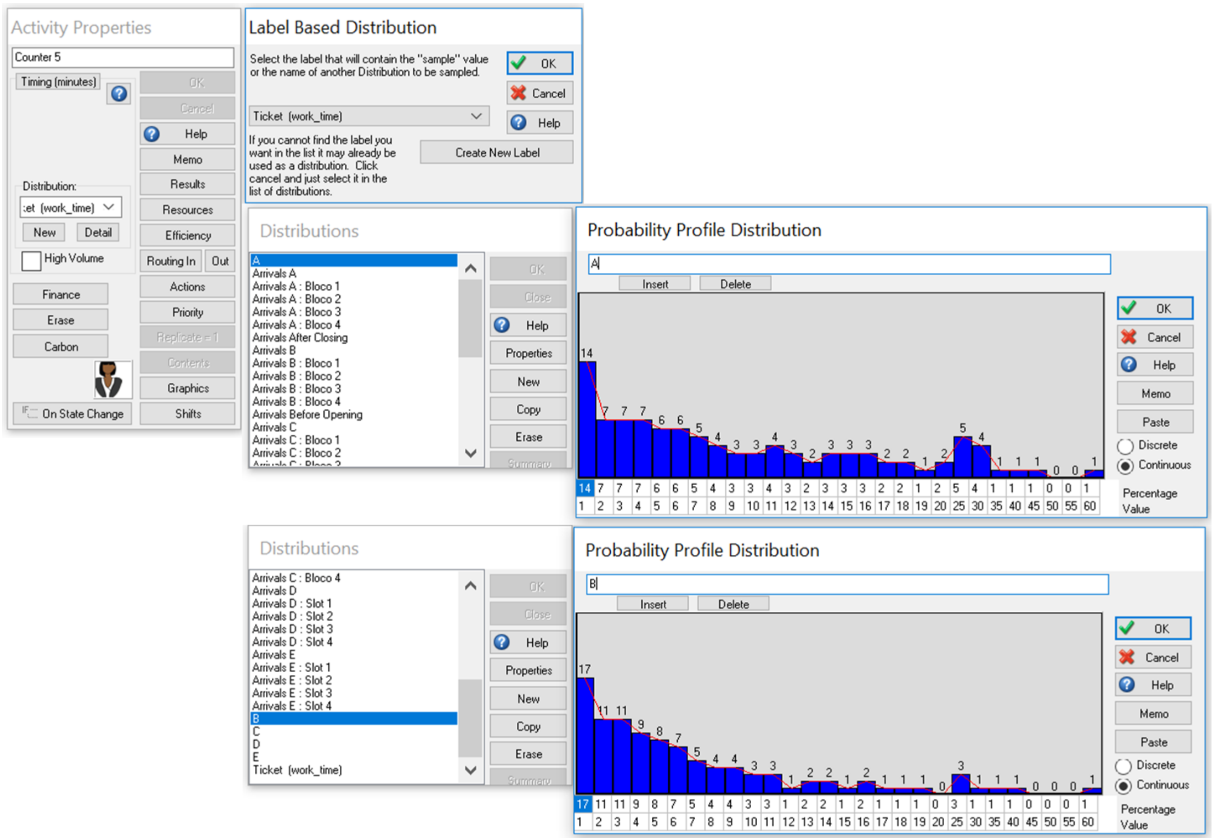

Figure 38: Setting label-based distributions in each service-station to model its service-times in Simul8

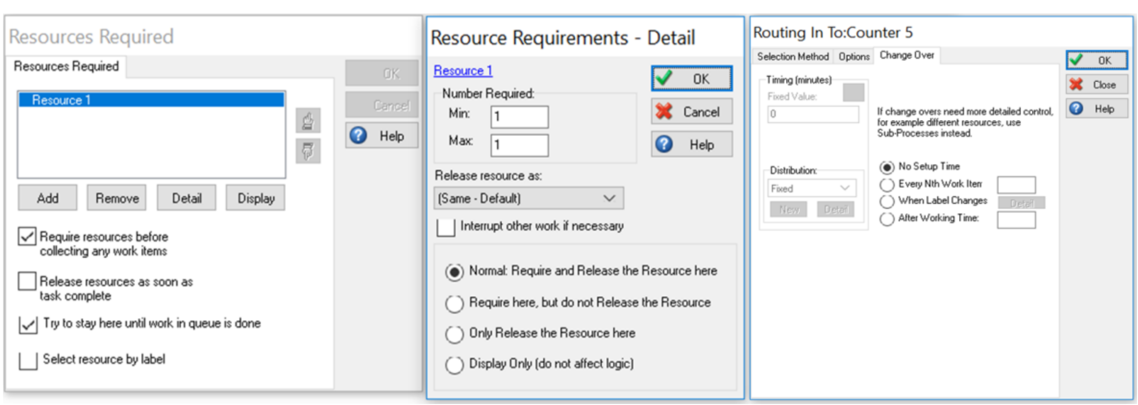

Figure 39: Set of rules that was programmed in Simul8 to mimic the real services

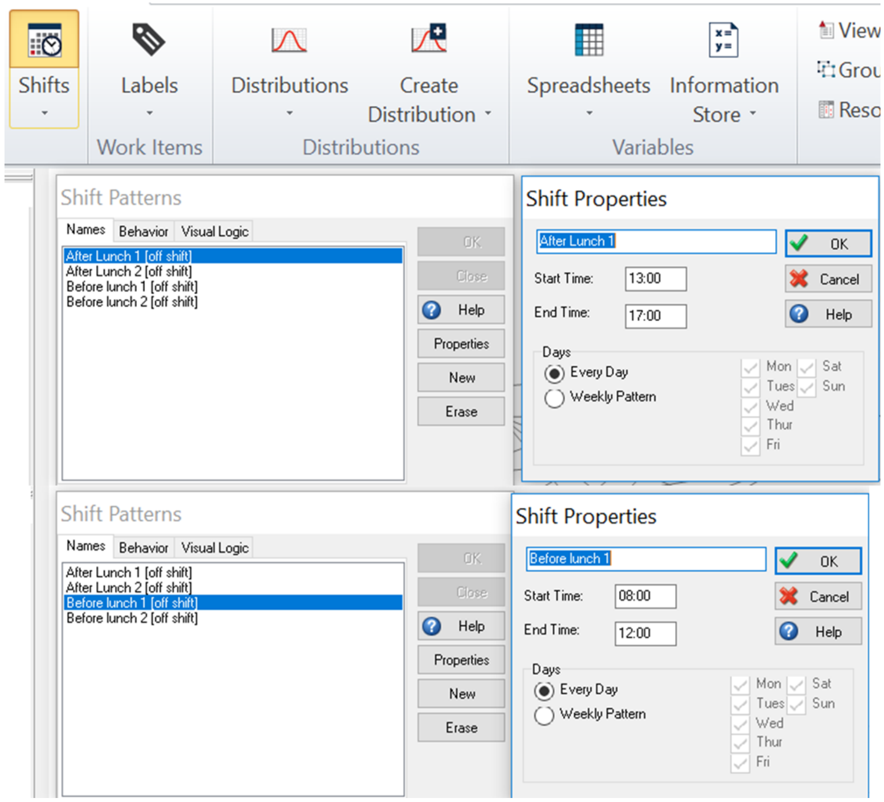

Figure 40: Shifts definition in Simul8

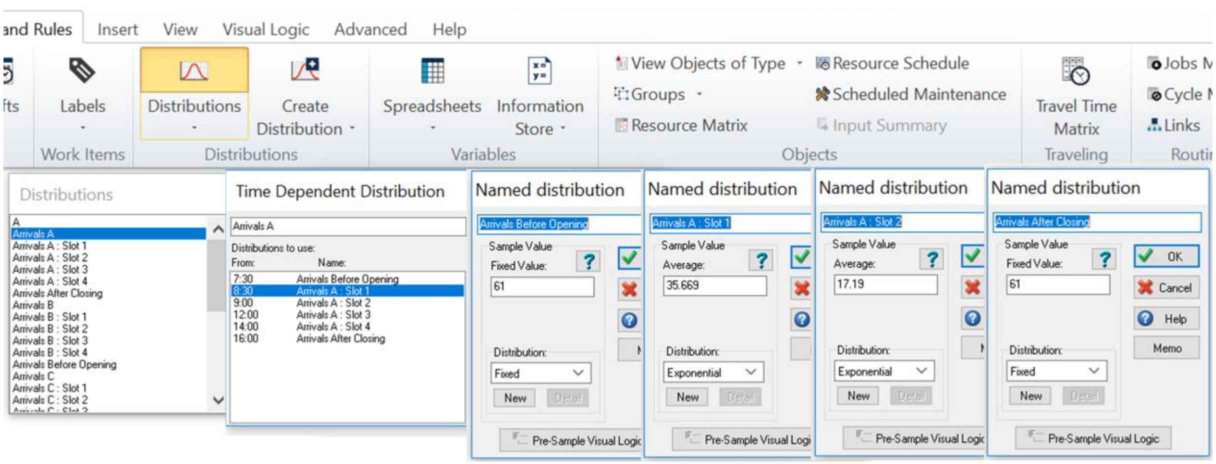

Figure 41: Arrivals characterization in Simul8: creation of a Time Dependent Distributions per ticket-type

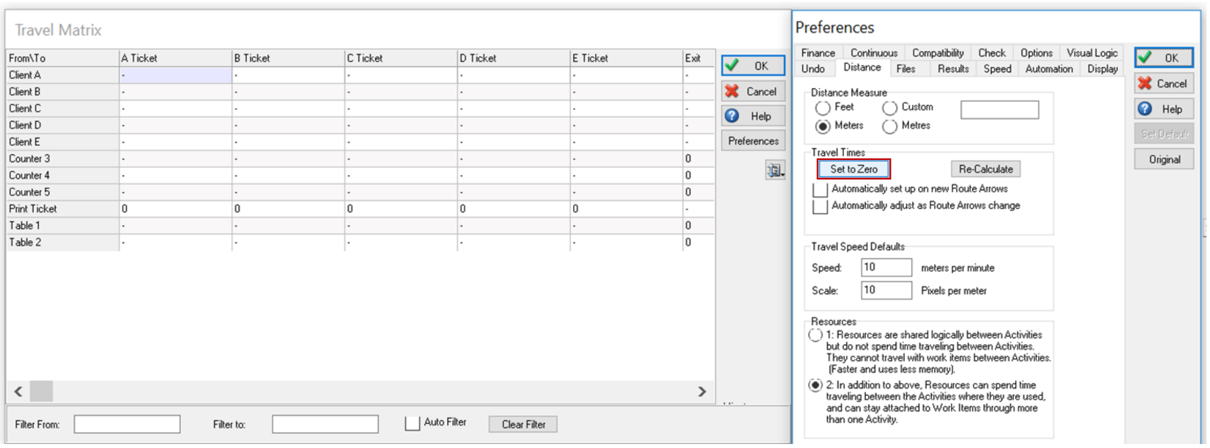

Figure 42: Travel Matrix in Simul8: setting all travel times to zero

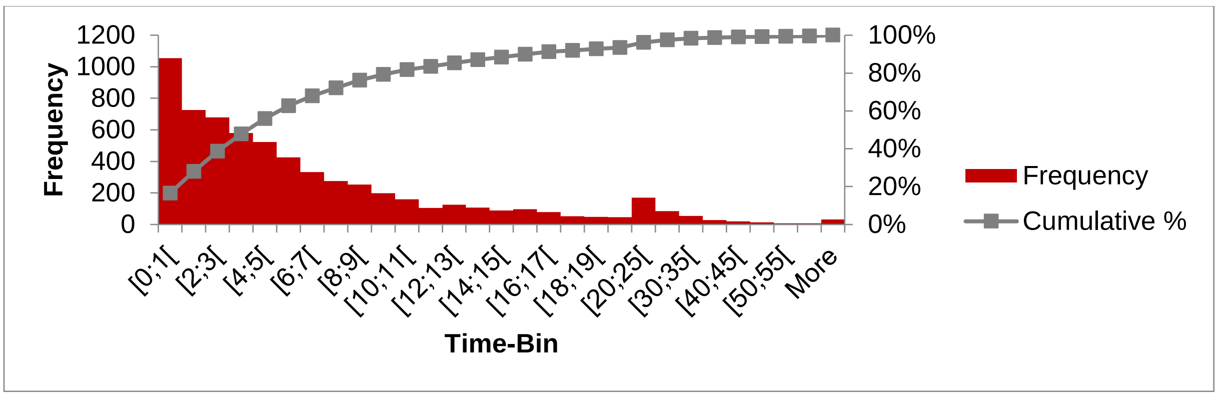

Figure 43: Histogram of the service times of tickets 'B'

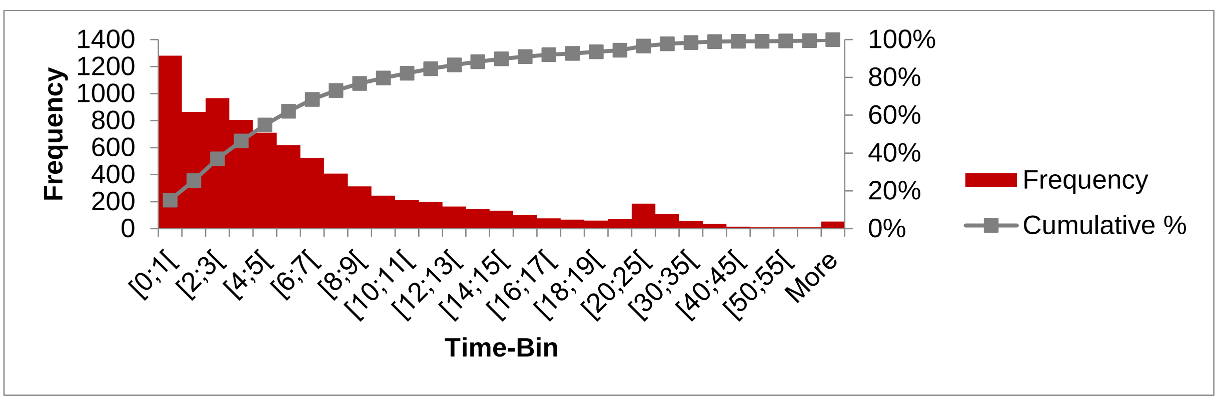

Figure 44: Histogram of the service times of tickets 'C'

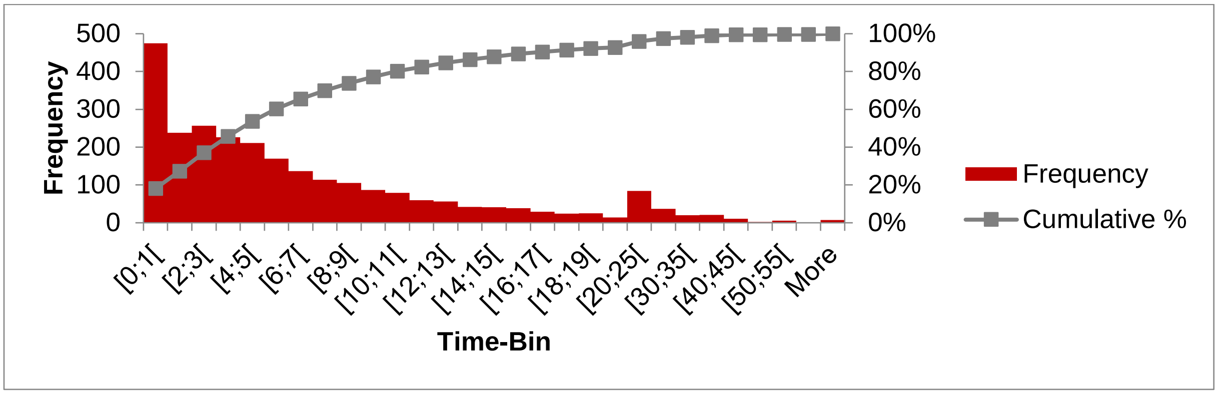

Figure 45: Histogram of the service times of tickets 'D'

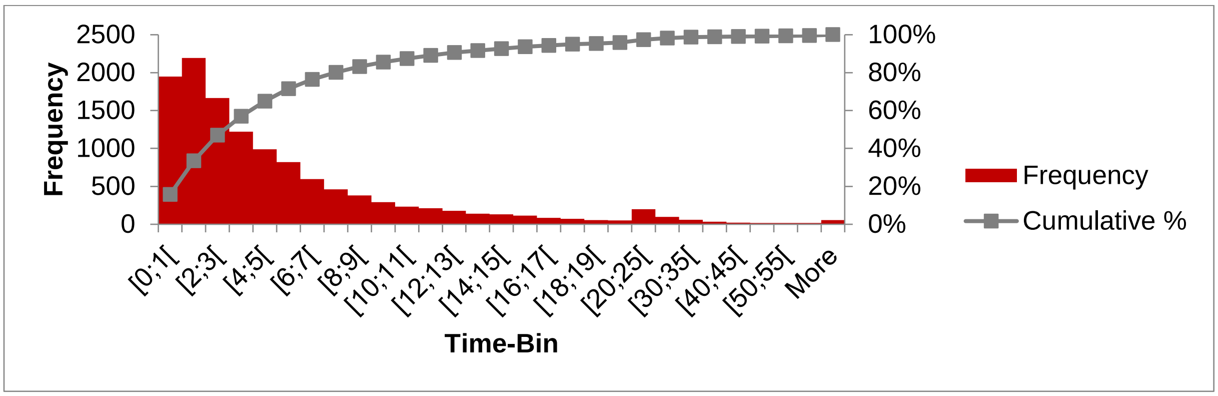

Figure 46: Histogram of the service times of tickets 'E'

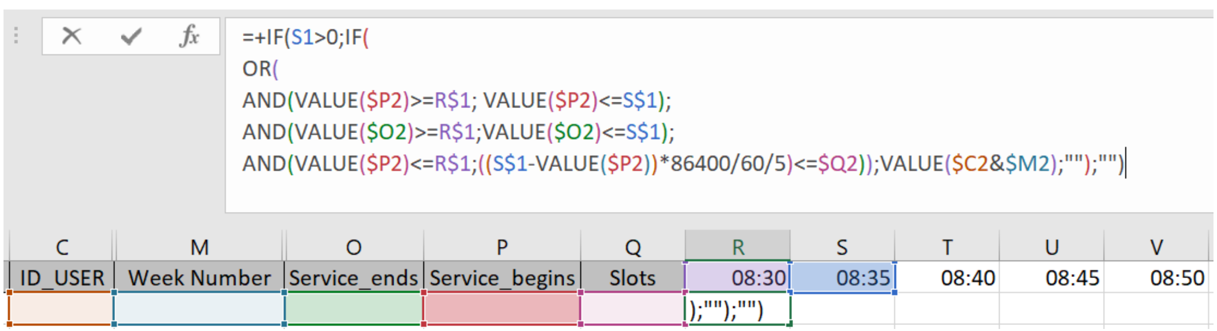

Figure 47: Developed algorithm to compute efficiency levels

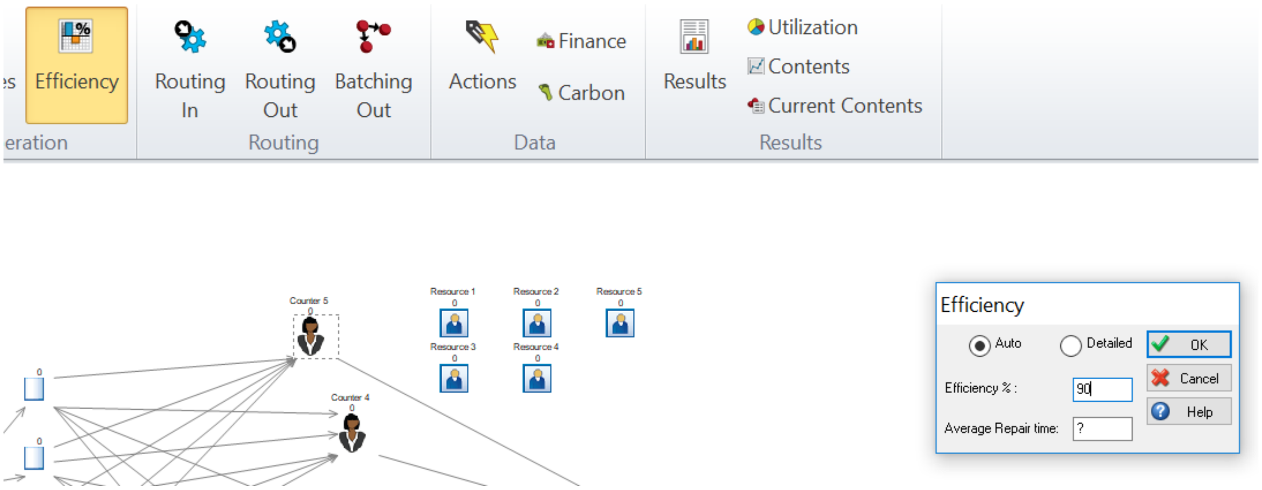

Figure 48: Definition of servers' efficiency level in Simul8

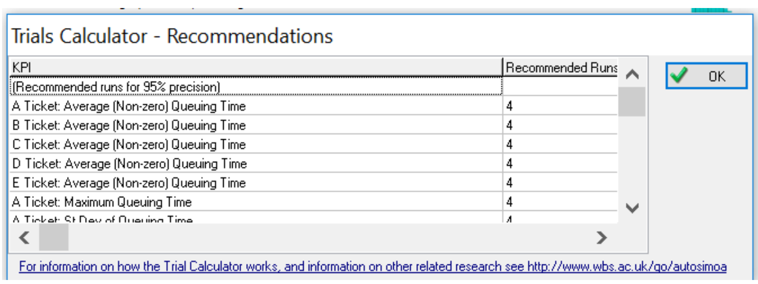

Figure 49: Trials Calculator for 95% precision in Simul8

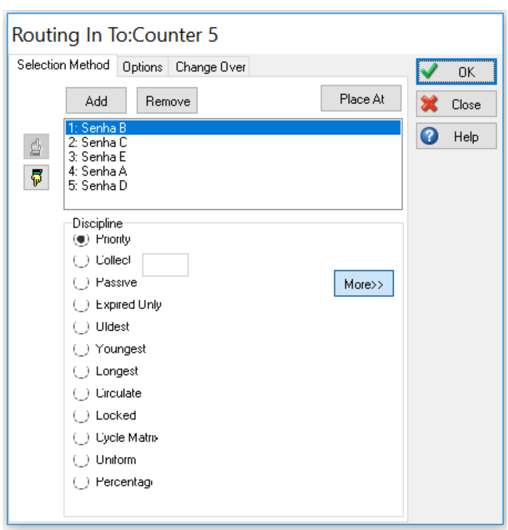

Figure 50: Setting the queuing discipline of Counter 5 to match Inline®'s prioritization model in Simul8

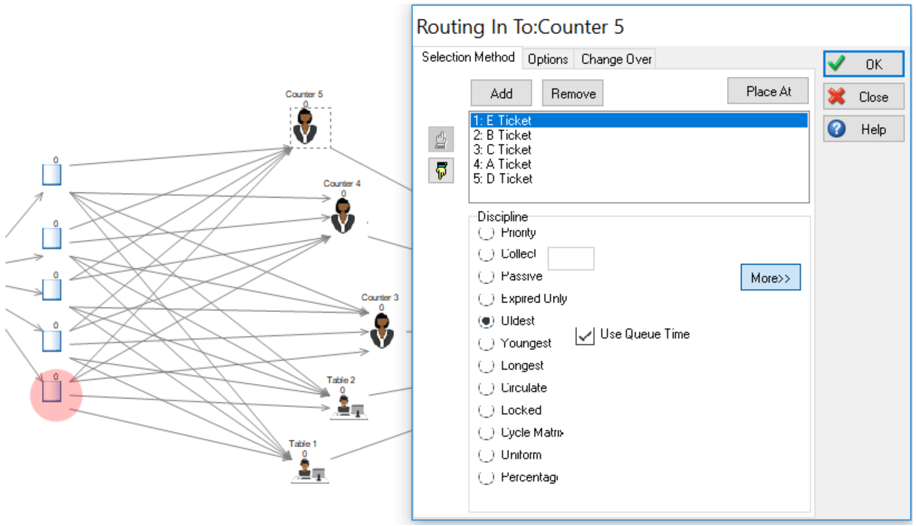

Figure 51: Setting the queuing discipline of Counter 5 to FIFO in Simul8

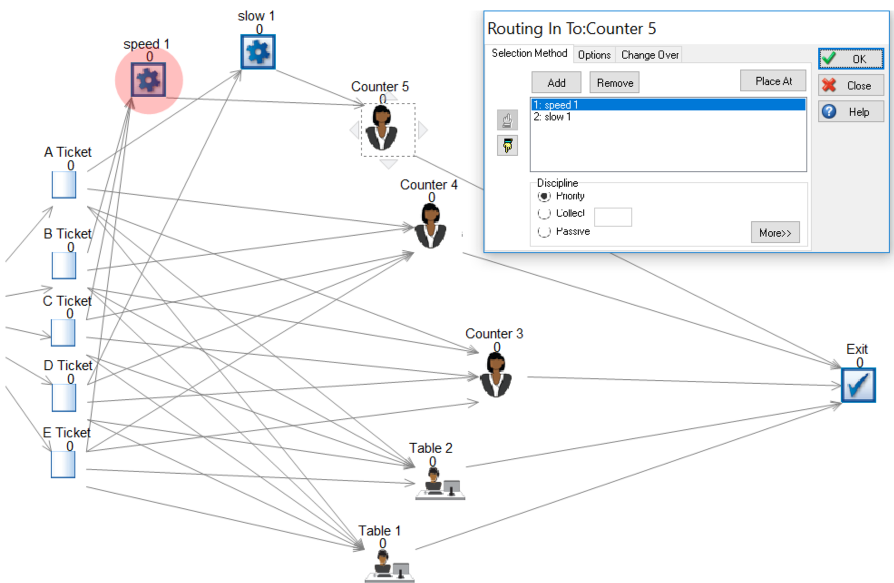

Figure 52: Introducing the speed-track priority discipline in one service-station in Simul8

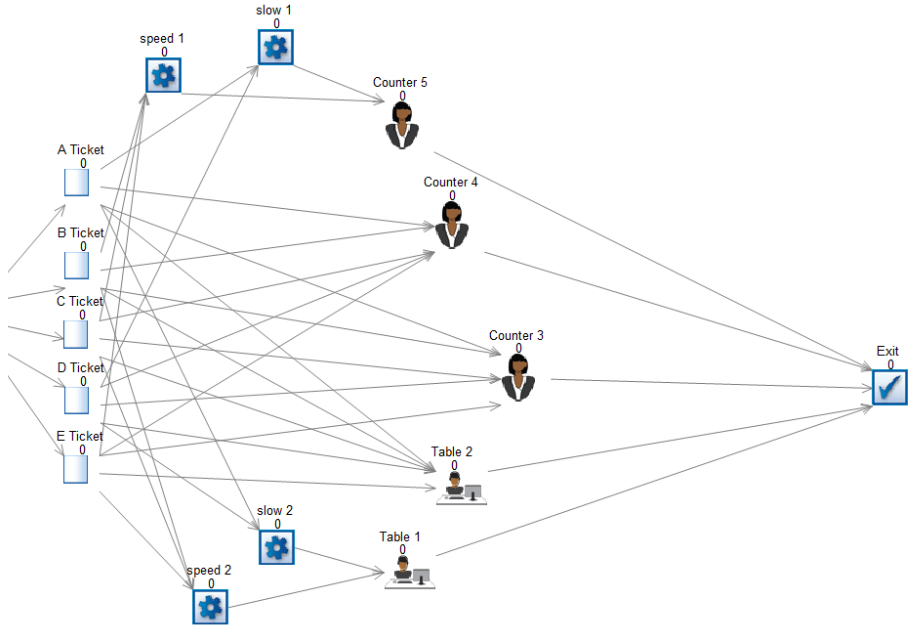

Figure 53: Introducing the speed-track priority discipline in two service-stations with different shifts in Simul8

Place des Doyens, 1 bte L2.01.01, 1348 Louvain-la-Neuve, Belgique www.uclouvain.be/lsm