Tactical Management of Freight Transportation Services by Rail: Evaluation of Yard Performances

Marin V. Marinov and José M. Viegas
Instituto Superior Tecnico at Technical University of Lisbon,
Dept. of Civil Engineering and Architecture,
Transport Infrastructure, Systems and Policy Group
Av. Rovisco Pais, 1049-001 Lisbon, Portugal, e-mail: marinov@civil.ist.utl.pt

Abstract
When discrepancies between planning and operations occur, the final product of the service provided easily becomes of poor quality. Motivated by a similar situation, over the last 5 years, we have studied transportation services with freight trains by the Portuguese Rail Freight Operator – “CP Carga”. We have unveiled a problematic cycle caused by multiple inadequacies involving tactical management and operation. At the bottom of this problem lies the incomplete methodology for fulfilment of both annual and weekly planning processes seen in disregarding the maximum processing capabilities of the yards. We believe that one possible way to break the problematic cycle is to provide reliable tools for evaluation of yard performances that could be used at both annual and weekly planning. More specifically, a two steps approach for evaluating yard performances at tactical management level is discussed. The suggested approach involves analytical modelling with G/G/m queues and Event-based simulations with SIMUL 8 and is implemented for the purposes of “CP Carga”. The main concept employed is that in planning the yard operation, one should consider these facilities as Pull Production Systems that explicitly control the “Work in Process”, i.e., the amount of trains to be processed by the yard in question. Thus, seamless operations with freight trains are experienced.

Keywords
Rail Freight Transportation, Tactical Planning, Rail Yard

1 Introduction

Based on the Anthony’s classification (Anthony [1]), there are three classical management decision-making levels, as follows: strategic, tactical and operational. In the context of freight transportation by rail, the three classical management levels have been introduced by many, e.g. Assad [2]; Crainic et al. [6]; Crainic and Roy [7]; Crainic and Laporte [8]; Crainic [9]; Gualda and Murgel [11]; Pachl and White [31], Marinov [26], and therefore the following definitions can be specified:

1) The strategic level is related to long term vision and involves decisions for setting overall goals and targets, overall level and types of resources available, redesign and reconstruction of the physical rail network, relocation of rail facilities, building and demolishing rail infrastructure, acquisition of new resources that are of big dimension to the company, etc.

2) The tactical level deals with medium term planning. At this level the transportation plans are prepared basing upon adopted production schemes in operation.
At this level capacity research, congestion analysis as well as evaluation of systems’ performances are generally conducted.

3) The operational level is for short term planning (i.e., over the same day) and delivery of service. This level is dedicated to how the transportation plans are “day-to-day” implemented in fulfilling freight transportation service by rail. This level incorporates daily empty car distribution, daily locomotive assignment problem, daily crew scheduling, daily timetable setting, and daily dispatching.

Over the last 5 years, we have studied transportation services with freight trains by the Portuguese Rail Freight Operator – “CP Carga”. In this paper, we report about our motivation, objectives and methodology, some of our findings, main concepts employed as well as approaches applied as the scope of analysis involves mainly tactical management and evaluation of yard performances.

It should be noted that a part of this work was presented in “2nd International Seminar on Railway Operations Modelling and Analysis, “RailHannover 2007”, Leibniz University of Hannover, Germany (Marinov and Viegas [25]). In that work we have concentrated on analyzing freight train movement, and explored the behaviour of four flat-shunted rail yards in terms of a network (we called these facilities “formation yards”) with the purpose of evaluating the current production scheme of a rail freight operator and further trying to improve the system performances by identifying alternative ways to serve the demand origins and destinations in the rail network under study. The concept there presented for examining freight train movement in a rail network was based on traffic intensities, i.e., not on detailed freight train schedules.

Since then the scope of this work has been significantly extended. The simulation models there presented have been refined. A two steps approach for evaluating yard performances consisting of analytical modelling with G/G/m queues and Event-based simulations with SIMUL 8, has been developed. G/G/m queues have been used to provide initial evaluations. The simulations has confirmed the results obtained by the G/G/m queues and further allowed us to study yard performances in a greater detail by including factors such as behaviour of yard personnel e.g. that cannot be studied precisely by G/G/m queues. Thus, two simulation modelling methodologies at Micro and Macro level have been developed and implemented. Micro level deals with a single yard performances. Macro level deals with yards performing in a network where we explicitly considered the scheduled freight train operation in a network with the purpose of analyzing the effect of both scheduled movement of freight trains and schedule deviations of freight trains on the rail yards’ performances.

In brief, we have seen that for tactical management purposes, the planners should specify upper bounds that plausibly indicate the maximum processing capabilities of the yards in their networks and in order to experience a seamless operation these upper bounds must not be violated, otherwise the yards are (over)saturated, the yard personnel encounters difficulties to serve trains which starts causing delays, the fluidity is unsettled and the freight transportation service provided by rail becomes of poor quality.

The rest of this paper is organized, as follows: motivation for developing this work comes next followed by a brief description and additional explanations on the objectives of the study. Next, the adapted methodology, which is Systems Approach, is presented in short followed by a detailed discussion of the application of the Systems Approach in rail freight yards and network. We elaborate briefly on the formulation of the problem,
analytical modelling concepts as emphasis is placed on G/G/m queues and two simulation modelling methodologies, where we look at both a single yard performances and yards performing in a network. Evaluation of yard performances comes next, where we provide an interesting discussion on obtained results. We wrapped up with conclusions and F.R..

2 Motivation

2.1 EU Policy in the Rail Sector

Freight transportation services constitute a major impelling factor for most economic and social activities in the world. Simply put, the freight transportation systems allow the commodities to move as they provide the required links between suppliers and customers.

Environmental impacts of freight transportation services, however, are a matter of concern mainly in relation with road transport. Even if significant progress in engines and fuels has been made, driven by tougher legislation, and leading to big reductions of emissions, the local impacts on road vehicles (noise, particles and vibrations) are very resented by the population. It is also well known that concentration of emissions in mountainous areas leads to higher impacts, and so justifies special measures to contain them (Viegas [44]). Reduction of these impacts and hence guaranteeing environmental sustainability is an imperative more than merely adequate policy. Realizing these facts the European Commission launched strategies for modal split from road to rail, looking at the rail freight transport as environmental-friendly transport mode.

However, as stated by many “…for many decades, railways in most of Europe have been seen as a problem. They have steadily lost market shares, falling from 10% to 6% of passenger kilometres and 20% to 8% of freight tonne kilometres over 30 years. They also require high and increasing levels of subsidy; … less than half of the total costs of rail transport in Europe are borne directly by passenger and freight customers” (Nash and Rivera-Trujillo [29]). In response to this situation, a set of “remedy actions” has been launched by the European Commission in order to turn the trends and revitalize the European railway sector by creating an integrated, efficient, competitive and safe railway area as well as setting up a network dedicated to freight services.

The European Commission developed a new European Union railway policy based on encouraging the competition in the railway market by implementing vertical disintegration in the sector. More precisely, vertical disintegration in terms of European Union Railways means: separation of railway infrastructure from operation, where further opening of the railway market for entry of new railway operators (also called “undertakings”) has been expected. Moreover, every Railway Operator must possess an operating certificate and must pay fees for infrastructure use (“access fees”). This new policy has been underpinned by a number of regulations, which have stipulated and framed the pace of the railway structural and legislative reform in Europe. We shall not present a detailed overview of European Union legislation in the rail sector because this discussion is not new and can be found in the official site of the European Commission, (i.e., http://ec.europa.eu/transport/rail/index_en.html, consulted on October, 28, 2008). Instead, in the following section we report what was set out so far as officially announced by the European Commission, i.e…:

- developing a common approach to rail safety with the objective of gradually integrating the national safety systems
- bolstering the measures of interoperability in order to operate trans-frontier services and cut costs on the high-speed network
- setting up an effective steering body – the European Railway Agency - responsible for safety and interoperability
- extending and speeding up opening of the rail freight market in order to open up the national freight markets
- joining the Intergovernmental Organization for International Carriage by Rail (OICR)

Furthermore, in the forthcoming future other goals as well as lines of actions are expected, such as:
- ensuring high-quality rail services (goal)
- improving the environmental performance of rail freight services (goal)
- improving rail passengers' rights (goal)
- removing barriers to entry into the rail freight market (line of action)
- gradually setting up a dedicated rail freight network (line of action)
- progressively opening up the market in passenger services by rail (line of action)

Either way, currently we are witnessing that most European railway operators (i.e., undertakings) cannot yet serve their clients well and cover their full costs. Therefore, they are considered as “still-ineffective” organizations. It seems that little progress has been made thus far through the new policy launched by the European Commission and its implementation into practice. As addressed by Viegas [44] “the general behaviour has been that the railway sector has preferred to protest and resist change but done very little homework to improve its efficiency and quality of service”. It seems that other ranges of problems are hidden that have been received a little attention so far. It seems that much attention is paid to close the political gaps in the sector while not enough attention is paid to railway operations and performance of railway facilities.

2.2 Freight Transportation by Rail

Within the context of railway freight transportation, there are (at least potentially) many clients with different needs that want to transport different quantities and classes of freights, from many different demand origins to many different demand destinations. Not every consignment corresponds to a full (block) train. A single freight car or a block of freight cars does not move usually on one freight train directly to its demand destination. Instead, the freight car moves on various freight trains. The freight cars shift the freight trains at the rail yards, which are indispensable part of the rail freight operation. Consequently, the railway freight operators perform “network-based” businesses that consist of consecutive operating processes executed in different places over the railway network. These operating processes are inter-connected and inter-dependent. They employ many operations. For execution of each operation static and dynamic resources are involved. The quality of provided service strongly depends on the execution of each operation. A single resource is missing, one operation fails and the entire process deteriorates. Therefore, there may be many reasons for providing a rail freight transportation service of poor quality.

2.3 Rail Freight Operator under Study – CP Carga

The rail freight operator under study is CP Carga, the Portuguese Railway Freight Operator (“CP - Comboios de Portugal”). The operating processes with freight trains at CP Carga are fulfilled in three stages, as follows:
- Annual Planning
Weekly Planning

Annual Planning
In the early beginning of every year the clients should provide to the CP Carga the so-called “annual input”, which consists of the type of freight to be transported, the planned tonnage to be transported and the demand origins/destinations. On the basis of this input the planning department elaborates the supply production scheme in order to satisfy the annual customer demand. The supply scheme is a manually made proposal of how the customer demand would be served during the forthcoming year. Usually, the supply scheme is prepared taking into account either a recent historical review or the actual transportation plan. Then, some fixed timetables for freight trains are stipulated, which are further adjusted at the weekly planning.

What is observed is that the annual freight train plan is constructed on the basis of an exaggerated tonnage that should be transported during the forthcoming year. Thus, more freight trains are yearly planned and some of them with excessive tonnage. Moreover, during the annual planning process the maximum capacity of the yards is not considered. An explanation provided for this is that in the beginning of the year the client declares more tonnage for transportation than what he really intends to do. In terms of CP Carga the cancellation by the client of planned freight train has no costs for him, so this permits more flexibility to the client. The planning personnel are used to say: “we are lucky if the cancelled planned freight trains come up to 30% per week…the client requests three freight trains per week but he finally does only two…there are periods in which the cancelled planned freight trains adds up to 60% per week!”

Weekly Planning
The clients communicate with the Commercial department of CP Carga. The commercial department consists of four commercial sections (CSs), namely:

- Section responsible for containers
- Section responsible for construction materials (i.e., sand, cement, etc.)
- Section responsible for agro-industrial materials (i.e., cereals, manure, wood, etc.)
- Section responsible for special materials (i.e., automobiles, fuels, iron ore, coal etc.)

Those commercial sections have a key role in the planning process because they set up with the client the freight, the tonnage, the demand origins/destinations, the times for loading and unloading and the frequency (i.e., how many freight cars and/or trains per day). On the basis of this information the CS sets and sends a price for the transportation service. This process is negotiable and ends when both sides accept the price. Then, the CSs provide the afore-mentioned information to the planning department to elaborate the actual production scheme. Next, the production scheme is sent to the operation department for implementation. There are cases when the client requires transportation service from the morning to be fulfilled in the afternoon. It is possible and normally executed but what is preferred by CP Carga is a week in advance.

Every week, until Thursday, 14:00, the commercial department reports to the planning department the demand of freight cars for the next week, i.e., from Sunday to Saturday. The weekly planning process is totally performed manually. The planner tries to satisfy the weekly required demands for transportation according to the yearly planned freight trains and available slots in the actual timetable in a case of extra freight trains. The planner must announce the planned freight trains that should be cancelled and the new-
planned extra freight trains that should be run. This information has to be introduced in the system for information exchange and production control of CP called “TrainOffice” until Thursday, 18:00 (i.e., the same day). In the weekly planning process methodology the maximum capabilities of the yards are not considered explicitly. There is no instrument for it. Therefore, some suggestions for freight train movement are infeasible with respect to the maximum yards capacities.

The weekly planned freight trains may be further reduced (numbers and weight) by the planner responsible for the locomotive scheduling process. The planner is used to say: “there are not enough available locomotives”! According to the available locomotive fleet considering the characteristics of the railway network the planner must suggest a locomotive production scheme for fulfilment of the planned freight trains that should run during the next week. He must introduce the locomotive scheme in “TrainOffice” until Friday, 18:00 (i.e., on the next day) announcing either the cancelled planned freight trains and/or the reduced planned freight trains. Consequently, in this phase of planning it is known that there will be freight cars (or freight car blocks) that will not be transported and will have to be left behind, usually in the yards. The consequences of this situation are not considered.

Every Day Operation
The operation staff is responsible for service execution and resolves all daily problems occurred in real time. The operational personnel are used to say: “We manage conflicts”!

It is observed that the freight operation by rail is executed with little concern given to what was weekly planned. The yard personnel are used to say: “the superiors consider the shunting to be executed for 20 - 30 minutes, we cannot perform it because only the break test takes about 20 minutes...moreover, they planned too many freight cars to stay in the yard, there is not enough space, we need lines to execute our work, therefore in every opportune case regardless of the plan we send freight cars away to ensure space for the incoming freight trains!” In response to this situation the planning personnel are used to say: “we planned well what was required by the commercial department but the operation did not execute it as we planned!” Thus, both sides have a low confidence in each other.

This awkward situation between the planning and operation of course contributes to low utilization of the moving assets and low efficiency in providing the freight transportation service which further generates a significant increase of average costs in long term. From the customer viewpoint, it contributes to unreliable service seen in infeasible contracts, unfulfilled expectations and finally customer dissatisfaction and CP Carga cannot build a reputation as a reliable provider of freight transportation.

Consequently, there is a problematic cycle caused by multiple inadequacies involving tactical management and operation. At the bottom of this problem lies the incomplete methodology for fulfilment of both annual and weekly planning processes seen in disregarding the maximum processing capabilities of the yards. We believe that one possible way to break the problematic cycle is to provide reliable tools for evaluation of yard performances that could be used at both annual and weekly planning. The annual and weekly planning specifies the tactical management level of CP Carga. And, this is where this study takes place.

3 Objectives and Explanations

The main objective of the conducted study is to provide reliable tools for analysis and evaluation of the performance capabilities of rail yards (both performing individually and
in a network) using an appropriate approach. These tools are envisaged to be used at tactical planning in specifying upper bonds (in number of freight trains) that replicate the true processing capabilities of the yards under study. When these bounds are not violated, problems are not expected to occur. The objectives thus specified require further explanations, as follows:

The subject of this research is the railway facility “YARD”, the main function of which in short is to produce freight trains. The yard has a limited capacity specified by a limited number of static and dynamic resources involved. Therefore, the yard is thought as a pull production system with limited capacity that explicitly limits the amount of work in process (Cheng and Podolsky [3]; Hopp and Spearman [16], [16]); and in studying yards the systems approach is used.

The yard is an indispensable element of railway freight transportation system. This system is a complex, hierarchical and dynamic entity in which no single element exists in isolation. Therefore, a significant gain cannot be achieved by only confining to yard relative tasks and problems. Instead, a wider spectrum of tasks and problems relative to the railway freight transportation are examined.

There are three classical decision-making levels: strategic, tactical and operational. This study falls within the “TACTICAL” decision-making level because the tactical management of the railway freight operator in question appears to fulfil its tasks using incomplete methodology seen in disregarding the processing capabilities of its yards. However, recalling the previous paragraph, the yard does not exist in isolation, but is a part of integrate system. Therefore, for the objective of this discussion railway freight transportation tasks and problems encountered at tactical management level are studied. Where it is of interest, railway freight transportation tasks and problems encountered at strategic and operational management levels are also considered.

The yard performances are subordinated to the production scheme in operation. There are two basic production schemes: improvised and scheduled (White [45]). The railway operator under study tends towards scheduled operation but this operation easily turns into improvised due to the fact that operation unit hardly ever fulfils the schedules. Therefore, “improvised vs. scheduled patterns” are examined.

The real system cannot be used for experimental purposes. Instead, realistic meaningful MODELs which adequately replicate the yard performances (and yards performing in a network) are needed. There are physical models, conceptual models as well as mathematical models. For the objectives of this discussion conceptual and mathematical models are applied. The models are developed by using an appropriate theory.

After the models are provided, the real experiment begins. The provided models are the potential tools for analysis and evaluation of yard performances. However, they can be considered as RELIABLE TOOLS only after having proven their high level of reliability. Therefore, for the objectives of this discussion the most models provided are validated and tested.

4 Methodology

4.1 Systems Approach

The methodology used in accomplishing the main objectives of this research falls within a larger scheme known as the systems approach. Systems approach methodology is appropriate because the end objective of the approach is to decide the best way to operate
a system (Pask [32]; Churchman [4]; de Neufville and Stafford [10]; Hall [12]). The systems approach end objective is achieved with the following four steps:

- Formulation
- Modelling
- Evaluation
- Decision

Formulation
The purpose of the formulation step is to specify the characteristics and specificities of the situation at hand and than isolate and define a problem.

Modelling
The purpose of modelling step is to develop a reliable replication of the real system that can be used to provide a better understanding of how it operates.

Evaluation
After the model is created, the next step is to use it to evaluate the system performances through adequate measures of system performance (MOP).

Decision
By considering the information obtained from the evaluation step to make adequate decisions according to the company’s objectives. The decision making (i.e., selecting the best alternative) should be simple if one has adopted a simple objective and performance measure, however, it can be very complicated if there are multiple dimensions involved.

5 Application of the Systems Approach

5.1 Formulation

In the early stages of this study, based on real data collected, preliminary analyses of CP Carga performances have been conducted. Indices such as: characteristics of freight trains, turnaround of freight car, daily run of freight car, commercial speed, and schedule deviations of freight trains have been examined. Generally speaking, the following results have been obtained:

The cargo weight per freight train comes up to 450 tons on average. The gross weight per freight train adds to approximately 790 tons. The average length of CP freight train is approximately 390 metres. The number of freight car groups per freight train is 4, and the number of freight cars per freight car group is 6 on average.

The turnaround of freight car for the conditions of CP adds up to 4.4 days on average (i.e., 107 hours). During one turnaround the freight car is loaded 1.3 days (approximately 31 hours); and stands empty 3.1 days (approximately 76 hours). The average whole run of a freight car cycle is approximately 529.6 kms, where the loaded run of freight car is approximately 266 kms and the empty run of freight car is approximately 263.6 kms.

On average, the freight car stands loaded in dispatch stations by approximately 3 hours; it stands loaded in yards by approximately 18 hours; it stands loaded in terminal stations by approximately 4 hours; and the running time in loaded state is about 6 hours.

On average, the freight car stands empty in dispatch stations by 26 hours; it stands
empty in **yards by approximately 24 hours**; it stands empty in terminal stations by 22 hours; and the running time in empty state is approximately 4 hours.

For the conditions of CP the **commercial speed is approximately 16.7 km/h**, whereas the EU average commercial speed mentioned in the White Paper of 2001 is 18 km/h (COM 2001).

The **average starting delay** of CP freight trains from the **yards is 40 minutes** which still increases by another 10 to 15 minutes during the run until the terminal stations (or the next yard), some 100 to 120 kms away.

From these rough estimates obtained at this early stage, those that most attracted our attention were a low commercial speed, a significant average starting delay from the yards, a significant average dwell time of freight car (loaded and empty) in its turnaround. In all these estimates, the yards come into play.

Next, a set of interviews, desk-top analyses and frequent observations of the system (i.e., CP Carga) in operation have been conducted. The problems at the yards have become much more apparent and our intuition that the yards encounter difficulties to fulfil what is planned has been confirmed.

### 5.2 Modelling

In the literature yard performances at tactical management level have been studied by:

- **Deterministic analytical methods**
- **Queueing Methods**
- **Simulations**

**Deterministic analytical methods**

Deterministic analytical methods have been introduced e.g. by Konstantinov [24]; Tasev [40]; Skalov [38]; Tasev and Karagyozov [41]; Raikov [35]; Pachl [30]. These methods are mostly used to roughly determine the absolute minimum number of yard tracks and consist of a simple equation, such as:

\[
 n_{\text{tracks}, \min} = \frac{t_{\text{dwell, average}} \times n_{\text{mins}}}{t_{\text{average}}}
\]

where,

- \( n_{\text{tracks}, \min} \) - minimum number of tracks
- \( t_{\text{dwell, average}} \) - average dwell time in min per freight train
- \( \left( \frac{n_{\text{mins}}}{t} \right)_{\text{average}} \) - average traffic flow, freight trains per period

The deterministic analytical methods are simple and easy to implement. However, they do not estimate operation factors. At the yards not only static resources are concentrated but also dynamic (i.e., yard personnel). The dynamic resources are also limited and operate within the limited capacity imposed by the static resources. The performance of dynamic resources controls the level of operating efficiency of the entire facility. Disorder and disorganisation among the dynamic resources cause the queues to materialize, the transportation product to linger as well as the service to delay. Finally, little is produced and huge unjustifiable costs are accumulated.
Further, simply put, the dwell (or “throughput”) time of freight train consists of waiting times and operating times. Those times cannot be estimated precisely by deterministic methods because they do not consider the variability in the arrival and service patterns. Furthermore, the deterministic methods do not identify the “bottlenecks” over the throughput line of the yards. Therefore, they should not be implemented for analysing the performance capabilities of the yards at tactical planning level.

Queueing Methods

The railway freight network fluidity affects the overall system efficiency. The railway freight operation is a complex interdependent cyclic one and to be efficient the network cycles have to be fulfilled as fast as possible. This requires comprehensive knowledge on the insight of freight operation and operational relationships between its segments. Here, queueing theory has been used to provide a means for establishing operational relationships. In early studies (Hein [14]; Peterson [33], [34]; Martland [27]; Turnquist and Daskin [42]; Tasev and Karagyosov [41]) queueing methods have been used in predicting yard throughput average time in relation to a yard’s layout in order to understand its capacity limitations and processing capabilities.

The first step of the modelling process is to clearly define the yard subsystems in order for meaningful information to be obtained on the behaviour of each subsystem. The main reason for such an analysis is to isolate the main sources of delay in a yard. So the term cut was used, meaning the string of freight cars with the same final destination passing through a yard. It was assumed that a given cut encounters delays as follows: (a) Reception and Inspection Delays; (b) Classification or Sorting Delays; (c) Connection Delays; (d) Train Assembly Delays; and (e) Outward Inspection Delays (Assad [2]). Peterson (e.g. Peterson [33], [34]) in his systematic analyses concludes that (a) Reception and Inspection Delays and (e) Outward Inspection Delays can be modelled realistically by fixed service times (i.e., D - deterministic). All other yard operations exhibit congestion effect and should be modelled by general service-time distributions (i.e., G - general), however, it was found that exponential service time distributions worked reasonably well. Thus was founded the concept for directly analysing yard behaviour by a limited class of queueing systems (QS) that are said to operate in steady state. For this limited class of QS the probability distribution of the queue length is time independent and the arrival process is stationary. It is a limited class because analytical exact formulae are not available for all queueing system characteristics such as average queuing times, average queue lengths, average time in system, and the probability of overloading the system. If an exact formula does not exist, approximations for computation of the measures of subsystem performance would be used.

Another class of queueing models is a network of queueing systems. This issue is not yet resolved in a satisfactory way. Models of this type are difficult to solve, very complex and complicated for practical purposes. In the literature three “products” limited in application are found. These are:
- Open Queueing Network
- Closed (circuit or cyclic) Queueing Network
- Queueing Network Analyser

The product of closed queueing network is used when the number of customers is a relatively small known number. Otherwise, the system is treated as an open queueing network product dealing with infinite number of customers. There is a classical exponential queueing network product of Jackson (Jackson [17], [18]) which can be used
only if the arrival process is described by Poisson distribution, the service times are exponential, buffer sizes approach infinity. If one of these conditions does not hold then the Jackson network is inappropriate. In this situation, the queueing network analyser (QNA) may be used (Whitt [46]). In order to define the customer routing through the network the QNA solves traffic equations, then decomposes the network into single G/G/m queues and treats them independently. The formulas for G/G/m queues are not exact, but employ approximations and therefore the results may not always face realism, however.

In general, the queueing models quickly provide insights on the expected system behaviour (van Dijk [43]). However, queueing networks that have general distributions are difficult to solve, operate generally with approximations and cannot handle non-stationary behaviours. They are much more restrictive than simulation.

Simulation
The results obtained by simulation are susceptible to random fluctuations and simulation captures non-stationary behaviours. However, as stated by Hall [12] "...simulation is dangerous...there is even a tendency to forget that simulation models require empirical data ... - data that must be obtained through observation”.

Conceptually, simulation should take place at the end of analysis. It is an effective tool for evaluating “What-if” alternatives, tactical approaches, production schemes, design changes and capacity expansions. The strength of simulation model lies in its capability to capture a large amount of processes, decisions and details.

To make the best use of simulation, when analysing the complex yard behaviour (or any other complex system) one better operates with a specific simulation tool created for this particular purpose, as in Germany (Pachl and White [31]) or the software package VIRTUOS (Klima [22], [23]; Kavicka [21]).

When no specific yard simulation software is available or appropriate, one needs to analyse, choose and adapt an existing simulation tool for this purpose. A simulation language that has been used in examining terminal behaviour is General Purpose Simulation System (GPSS). GPSS is a process-oriented simulation language that combines sequence of events into single subroutines called blocks (Nadel and Rover [28], Karagyozov [19]; Katchaunov et al., [20]; Razmov [36]). We are also aware of a class project on hump yard simulation performed by Harrod [13]. The subject of this independent project is Queensgate yard - one of North America’s largest hump switching yards. Harrod has studied Queensgate yard performances by using Arena Simulation Tool. Arena employs an object-oriented design for entirely graphical model development. Simulation analysts place graphical objects—called modules—on a layout in order to define system components such as machines, operators, and material handling devices. The core technology of Arena is the SIMAN Simulation Language (Takus and Profozich [39]). After creating a simulation model graphically, Arena automatically generates the underlying SIMAN model used to perform simulation runs. As far as our knowledge goes, SLAM II Simulation Language was also used in studying yard performances.

For our purposes the yards under study are decomposed into areas, such as:

- **Arrival Yard (AY)** - a number of tracks within the yard limits where freight trains arrive
- **Shunting Zone (SZ)** - a number of tracks within the yard limits where freight cars are shunted
• **Departure Yard (DY)** – a number of tracks within the yard limits from where freight trains depart
• **Workshop (WS)** – a number of tracks within the yard limits where the locomotives undergo maintenance and stand while waiting for their new assignment
• **Car Yard (CY)** - a number of tracks within the yard limits where freight cars not in current use are stored. This is also called “Rip Tracks”. It should be noted that if one deals with a medium or a small term flat shunted yard, rip tracks may not be considered, and so this area might not be of interest.

In other words, in order to model yard performances we employ the decomposition approach, which allows us to divide the system under study into segments considering that all segments belong to one complete system. Thus, one is able to analyse and evaluate the performances of each system segment separately as well as the performance of the whole system without neglecting the level of influence between the system segments and the global impact within the system limits.

After having decomposed the yards into areas, a two step approach is suggested. The first step is considered as Preliminary Study and employs modelling with G/G/m queues. The second step employs Event-Based Simulations using SIMUL8, enlarges the span of the evaluation, verifies the results obtained by the G/G/m queues and is considered as Basic Study. We believe that such an approach provides a greater precision and trust in the conducted experiment.

### Modelling with G/G/m Queues

Let us decompose the yard under study into Gi yard areas/subsystems (, where i = 4 , in our case). Then, for tactical management purposes the yard subsystems can be analytically modelled with G/G/m queues, as follows:

- **Arrival yard (AY) say, subsystem Gi**
  - λGi,in – Arrival rate, i.e., Number of freight trains to arrive in Arrival Yard per hour
  - TGi,s – Service time at Arrival Yard per freight train on average
  - Tsf,Gi – Time for further processing at Arrival Yard
  - TGi,s = TGi,s * Tperiod/period Tsf,Gi
  - Service time at arrival yard considering the time for further processing

- **Gi** - Service rate, i.e., Number of freight trains to be served per hour - μGi = 1/TGi,s

- S(m)Gi – Servers, i.e., Number of yard crews fulfilling the operations at AY

- Absolute utilization of Gi - ρGi = \[ \frac{\lambda_{Gi}}{\mu_{Gi}} \rightarrow because \rightarrow S(m)_{Gi} = 1 \]

- S(mmin)Gi – Minimum number of servers required in AY can be estimated by satisfying the following condition: \[ X_{Gi} = \rho_{Gi} = \frac{\lambda_{Gi}}{\mu_{Gi} * S(m_{(\gamma)})_{Gi}} < 1 \]

- Class of queueing system for Gi G/G/1
- Measures of Gi performances:
- Expected freight trains in queue of $G_1$ - $L_{q,G1}$ - computed by Allen Cunnen approximation, e.g.
- Freight trains in Arrival Yard - $L_{s,G1}$ - computed by Little’s formulas
- Time in the queue of Arrival Yard - $W_{q,G1}$ - computed by Little’s formulas
- Time in Arrival Yard - $W_{s,G1}$ - computed by Little’s formulas
- $M_{tracks,G1}$ - Minimum number of tracks required in Arrival Yard can be estimated by
  satisfying the following condition: $\frac{\lambda_{G1,in} * W_{s,G1}}{M_{tracks,G1}} < 1$

**Shunting Zone** (SZ) say, subsystem $G_2$

An important argument for the modelling accuracy has to be expressed. The principal difference of yards depends on whether or not the breaking down and making up of freight trains are physically separated within the yard as well as how many shunting crews are in operation at the time. If these two processes are physically separated and executed by independent shunting crews, then they are independent of each other and can be modelled separately. However, if these two processes are not physically separated and are executed by the same shunting crew, then they become strongly dependent of each other and must be jointly modelled. If, there is one shunting crew in operation at the time performing both breaking down and making up of freight trains, so then, these two processes are dependent of each other and are jointly modelled, as follows:

- $\lambda_{G2,in}$ - Arrival rate, i.e., Number of freight trains for shunting per hour
- $T_{s,G2}$ - Service time for shunting per freight train on average
- $T_{d,G2}$ - Time for further processing for shunting
  
  $T_{d,G2} = T_{d,G2} * \frac{T_{period}}{T_{period} - T_{f,G2}}$ - Service time for shunting considering the time for further processing

- $\mu_{G2}$ - Service rate, i.e., Number of freight trains to be broken down and made up per hour
  
  $\mu_{G2} = \frac{1}{T_{s,G2}}$

- $S(m)_{G2}$ - Servers i.e., Number of shunting crews fulfilling the shunting process

- Absolute utilization of $G_2$ - $\rho_{G2} = \left( \frac{\lambda_{G2}}{\mu_{G2}} \rightarrow because \rightarrow S(m)_{G2} = 1 \right.$

- $S(m_{min})_{G2}$ - Minimum number of servers required in SZ of yard can be estimated by
  satisfying the following condition: $\chi_{G2} = \rho_{G2} = \frac{\lambda_{G2}}{\mu_{G2} * S(m_{(1)})_{G2}} < 1$

- Class of queueing system for $G_2$ G/G/1
- Measures of $G_2$ performances:

  - Expected freight trains in the queue for shunting - $L_{q,G2}$ - computed by Allen Cunnen approximation, e.g.
  - Freight trains under shunting - $L_{s,G2}$ - computed by Little’s formulas
  - Time in the queue for shunting - $W_{q,G2}$ - computed by Little’s formulas
  - Time for the process of shunting - $W_{s,G2}$ - computed by Little’s formulas
- \( M_{\text{tracks},G2} \) - Minimum number of tracks required for shunting process in the yard can be estimated by satisfying the following condition: 
\[
\frac{\lambda_{G2,in} \cdot W_{G2}}{M_{\text{tracks}}(G2)} < 1
\]

**Locomotive Workshop (WS),** say subsystem G3.

As a matter of principle this task is a locomotive scheduling problem, which is beyond the scope of this study. However, the train-engine operations in departure yard cannot be performed before the road locomotive is put on the composition. Therefore, for the objectives of this discussion the average time for waiting a road locomotive is assumed as given and its value is obtained through observations and timing. Furthermore, during the analytical modelling this time is explicitly considered as a waiting time of already assembled compositions. Nevertheless, scheduling road locomotives to assembled compositions in yards is a very interesting topic and deserves careful examination. This issue is one of the most difficult tasks of railway freight transportation and its solution depends on many factors, such as: the size of the available road locomotive fleet, type of traction over the railway network, locomotive haulage distances, cycles of maintenance, etc. Locomotive scheduling problem should be always on the agenda.

**Departure yard (DY),** say subsystem G4

- \( \lambda_{G4,in} = \lambda_{G4,out} \) - Arrival rate equals the departure rate of the yard, i.e., Number of freight trains to depart from the yard per hour
- \( T_{s,G4} \) - Service time at departure yard per freight train on average
- \( T_{G4} \) - Time for further processing at departure yard

\[
T_{s,G4}^* = T_{s,G4} \cdot \frac{T_{\text{period}}}{T_{\text{period}} - T_{G4}} \quad \text{Service time at G4 considering the time for further processing}
\]

- \( \mu_{G4} \) - Service rate, i.e., Number of freight trains to be served at departure yard per hour

- \( S(m)_{G4} \) - Servers, i.e., Number of crews fulfilling the operations at departure yard, where \( m = 1 \) (assumed)

- Absolute utilization of G4 - \( \rho_{G4} = \frac{\lambda_{G4}}{\mu_{G4}} \rightarrow because \rightarrow S(m)_{G4} = 1 \)

- \( S(m_{\text{min}})_{G4} \) - Minimum number of servers required in DY can be estimated by satisfying the following condition: 
\[
\lambda_{G4} = \rho_{G4} \cdot S(m_{\text{min}})_{G4} < \frac{\lambda_{G4}}{\mu_{G4}}
\]

- Class of queueing system for G4: G/G/1
- Measures of G4 performances:
  - Expected freight trains in the queue of departure yard - \( L_{q,G4} \) - computed by Allen Cunnen approximation, e.g.
  - Freight trains in departure yard - \( L_{x,G4} \) - computed by Little’s formulas
  - Time in the queue of departure yard - \( W_{q,G4} \) - computed by Little’s formulas
  - Time in G4 - \( W_{x,G4} \) - computed by Little’s formulas
- Minimum number of tracks required in departure yard can be estimated by satisfying the following condition:

\[ \frac{\lambda_{G4,in} \cdot \delta \cdot W_{s,G4}}{M_{\text{tracks}(?)G4}} < 1 \]

**Network Measures of Yard Performance:**

- Expected freight trains in queue in yard, say \( L_{q,Yard} \):

\[ L_{q,Yard} = \sum_{i=1}^{n} L_{q,Gi} \]

- Freight trains in yard, say \( L_{s,Yard} \):

\[ L_{s,Yard} = \sum_{i=1}^{n} L_{s,Gi} \]

- Time in queue in yard, say \( W_{q,Yard} \):

\[ W_{q,Yard} = \sum_{i=1}^{n} W_{q,Gi} \]

- Time in yard, say \( W_{s,Yard} \):

\[ W_{s,Yard} = \sum_{i=1}^{n} W_{s,Gi} \]

**Simulation Modelling of Yard Performances using SIMUL 8**

Two simulation modelling methodologies for analysing yard performances have been developed and implemented in terms of CP Carga (i.e., the Portuguese Rail Freight Operator) using SIMUL 8. SIMUL 8 is a computer package for discrete-event simulations that allows us to create visual animation models of a wider range of queueing systems by drawing functional objects and attributes directly on the program screen (Shalliker and Ricketts [37]). When the system being examined is properly modelled a simulation experiment can be undertaken. The flow moving over the queueing network is shown by animation so that the appropriateness of the created simulation models can be more easily assessed. When the structure of the model is confirmed a number of trials are run. Measures of performance, such as: utilization rates, throughput time, time in queue, etc., are described statistically. SIMUL 8 operates with Building Blocks, such as:

- **Work Entry Point** – generator of work items, which we use to generate the inbound trains.
- **Work Centres** – system servers, which we use to replicate the yard subsystems, such as: cutting road locomotive from the train composition, classifying freight cars, inspections, breaking down train, making up train, train-engine operations and the like. Every work centre is characterized with a number of inbound trains, service pattern and a number of outbound trains. The averages of the service times of the work centres in our models are obtained through observations; which was imposed by the fact that CP Carga does not maintain such a record of data.
- **Storage Areas** - found as buffers or queues, which we use to replicate the limited capacity of the yard subsystems in a limited number of tracks.
- **Work Exit Point** – this is to indicate where the outbound trains leave the simulated yard and the service is declared as “finished”.
- **Resources** - could be a human, a machine, an operator, a brigade, an employee, an order, etc. or just a single signal. So the resources are items in the simulation model which are necessary for fulfillment of the service at the work centres. For our purposes, we use resources in order to study the utilization levels of the yard.
personnel (i.e., classification man, inspection man, shunting crew)

In addition, in order to replicate the arrival patterns at the yards, we employed “Time-Dependent Distributions”, which is subordinated to quasi-steady state regimes of work. There the arrival rate \( \lambda(t) \) is not a fixed value, but a function of time and yields the expected number of trains to arrive over a certain time interval. It should be noted that in conducting the experiment of “Improvised vs. Scheduled” operation in terms of yards-performing-in-network, the improvised operation was modelled by time-dependent distributions with exponential inter-arrivals, and the scheduled operation was modelled by time-dependent distributions with uniform inter-arrivals.

Recalling the two simulation modelling methodologies, we look at both a single yard performances and yards performing in a network. Thus, two levels of scrutiny are specified, namely:

- **Micro Level** - deals with simulation modelling of a single yard performance
- **Macro Level** - deals with simulation modelling of yards performing in a network

We shall not discuss in great detail these two simulation modelling methodologies, because they were discussed elsewhere. The interested reader is encouraged to consult Marinov [26]. Instead, we raise the concepts employed, as follows:

**Micro Level**
The employed concept for micro level simulation is to follow the throughput line of the yard under examination. The yard in question is decomposed into yard areas, as mentioned above. Within each yard area a number of consecutive operations are considered as follows:

- **Within the Arrival Yard considered are:** Receiving train; Cutting Locomotive by Shunting man; Classifying Freight Cars by Classification man; Technical Inspection by Inspection man; Further Operation in AY involving the arrival of shunting crew and preparations for Breaking Down.
- **Within the Shunting Zone considered are:** Breaking Down by Shunting Crew and Making Up by Shunting Crew
- **Within Departure Yard considered are:** Inspection for departure by inspection man, Train-engine operations (i.e., arriving of the road locomotive, putting the locomotive on and full brake test) and outbound trains leaving the yard.
- **Within Locomotive Workshop considered are:** inspection to verify/put the locomotive is in technical readiness for its next assignment
- **Within Car Yard considered is that** freight train composition stand while waiting next assignment

All the afore-exposed operations are replicated by individual Work Centres. The limited number of tracks of each yard area, are replicated by Storage Areas with finite capacity equal to the number of tracks in the given area.

**Macro Level**
The Micro Level methodology is extended to a network of yards and one is thus able to develop simulation models for evaluation of yard performances in a network. By this exercise, the global impact of the complex freight train operation both within the yards and between the yards is captured. The employed concept for macro level simulation is to replicate the complex freight train movement in a network, where the entire network in
question is decomposed/partitioned into segments, such as: Dispatch and Terminal yards (i.e., yards where loading and unloading is fulfilled); Railway Stations; Railway Lines, Rail Yards (i.e., shunting yards where our focus is on); there may be Rail Crossings as well. In order to replicate plausibly the characteristics of each segment of the network under study, a comprehensive data is collected and analysed. Then, the limited capacity of each segment is modelled by Storage area(s) and each operation by Work Centre(s) depending on the particular case.

Implementing the foregoing methodologies one creates event simulation models on the basis of which is able to analyse and evaluate the operating processes with freight trains within the yards and between the yards involved in the network being studied. Next, many simulation trials can be run; the system behaviour can be examined under many different conditions. One is able to obtain estimates for every yard subsystem performances separately. Then, the performances of similar yards’ subsystems may be compared through meaningful measures. Further evaluations of the behaviour of yards in a network can be obtained through aggregate measures of yard performances (Marinov and Viegas [25]). One is able to observe the total number of freight trains entered in the yards in the network; the total number of freight train processed by the yards in the network; the total average queueing time in the yards in the network per freight train per the period of the experiment. Thus, a lot of meaningful information is obtained; and exactly this information has to be the foundation for an adequate tactical management, not groundless discussions.

5.3 Evaluation

The suggested two steps approach for studying and evaluating yard performances has been applied in terms of three flat-shunted rail freight yards that are operated by the Portuguese Railway Freight Operator – CP Carga. The yards are: Gaia, Pampilihosa and Entroncamento. In terms of yards-performing-in-network, our simulation model consists of the three closely-studied yards, eight railway stations (i.e, Gaia, Ovar, Mogofores, Pampilihosa, Souselas, Alfarellos, Soure and Entroncamento), seven lines with sections and five dispatch/terminal yards (i.e., Ovar, Mogofores, Souselas, Alfarellos, Soure). Because of space limitations we shall not preset a detailed discussion of the technical and physical characteristics of these facilities. The interested reader is encouraged to consult Marinov [26]. Instead, in the following section we present some of the interesting results obtained during the conducted experiments involving both G/Gm queues and Event-based simulations.

In Figure 1, one observes the effect of increasing freight train arrivals on Gaia Yard Subsystems’ utilization. These results are estimated by G/Gm queue and demonstrate that with the increase of inbound freight trains the utilisation of the yard subsystems increase linearly.

The results shown in Figure 2, below depict the effect of increasing freight trains arrivals on the utilization rates of the Gaia yard personnel (i.e., indicated as Floating Resources where considered are: Classification man, Shunting Crew and Inspection man). These results are obtained through Simulation and confirm theoretically the fact that with the increase of inbound freight trains the utilisation of the yard personnel increases linearly.

With the increase of the number of inbound freight trains the queues within the yard limits are also on the increase, however.
Figure 1. Effect of Increasing Freight Train Arrivals on Gaia Subsystems’ Utilization estimated by G/G/m queues

Figure 2. Effect of Increasing Freight Trains Arrivals on Utilization Rates of Floating Resources, Gaia Experience, estimates obtained through Simulation

Figure 3 and Figure 4 below analyse the effect on Total Time in Queue and Average Throughput Time in Gaia Yard and Entroncamento Yard due to Increases in Freight Train Arrivals. These results are estimated by G/G/m queues. Next, in Figure 5, one observes a similar curve, where increases in Queueing Times of Gaia Shunting Zone with Increases in Freight Train Arrivals are shown. These results are obtained through Simulation. Note that the shape of the curves are quite the same demonstrating that with the increase of the number of inbound freight trains the time in the queue and the time in the yard increase. The important point is that after a given amount of inbound freight trains the curves start indicating significant jumps, which is quite visible in Figure 5. When significant jumps start to occur, it means that the yard under study reaches its maximum processing capability and is on the border of being oversaturated. In such situations operating problems are expected to occur and the yard cannot perform a seamless operation, which
will have a negative on the entire network and the final product will be of poor quality. Therefore, right here the concept of *Pull Production System*, which is one that explicitly limits the amount of work in process that can be in the system (Cheng and Podolsky [3]; Hopp and Spearman [15], [16]). Needless to say, the number of freight trains in the yard at any time cannot exceed the true processing capability of the yard. Consequently, there is no point to keep sending freight trains to be served by a yard over a certain period if the yard is already saturated. The number of inbound freight trains should be controlled up to a defined upper bound instead of pushing everywhere in the yard throughput line. Largely, this will ease the yard production control and will further guarantee a seamless freight train operation being executed by the yard.

![Figure 3. “Effect” on Total Time in Queue and Average Throughput Time in Gaia due to Increases in Freight Train Arrivals, estimated by G/G/m queues](image)

Figure 3. “Effect” on Total Time in Queue and Average Throughput Time in Gaia due to Increases in Freight Train Arrivals, estimated by G/G/m queues

![Figure 4. “Effect” on Total Time in Queue and Average Throughput Time in Entroncamento due to Increases in Freight Train Arrivals, estimated by G/G/m queues](image)

Figure 4. “Effect” on Total Time in Queue and Average Throughput Time in Entroncamento due to Increases in Freight Train Arrivals, estimated by G/G/m queues
Figure 5. Increases in Queueing Times of Gaia Shunting Zone with Increases in Freight Train Arrivals, obtained through Simulation

Figure 6. Aggregate Queueing Times, Arrival Yards and Shunting Zones, “Insignificant Deviation vs. Significant Deviation”, results obtained by Simulation

We have also studied “Improvised vs. Scheduled” Operations by CP Carga in terms of yards performing in a network. Figure 6, above, was drawn during this experiment, providing a comparison in Aggregate Queueing Times. We have examined the system behaviour when insignificant deviations from the schedules occur and when significant deviations from the schedules occur. The operation with freight trains that is characterized with insignificant schedule deviations has shown to be the better tactical strategy throughout the conducted experiment. For further details, the interested reader is encouraged to consult Marinov [26].
5.4 Decision

The decision making should be explicitly focused on the accomplishment of the scheduled freight train operation without violating the performance capabilities of the yards in the network, which will lead to continued improvements.

6 Conclusions

The main conclusion that can be drown from our research is that for tactical management purposes and seamless freight train operation, the yards should be thought as Pull Production Systems, meaning one keeps the yard workload in a number of freight trains (just) below an upper bound in which the yard queues start to apparently build up by controlling the number of the inbound freight trains to be processed by the yard in question (i.e., the work in process “WIP”). Thus, the freight trains will move through the yard subsystems “unimpeded” and seamless operation will be experienced.

7 Further Research

An issue outlined to be explored is how the operating processes with freight trains at yards will be fulfilled according to the new policy being geared in Europe towards competition on the railway market by the implementation of “vertical disintegration” in the sector, meaning separation of infrastructure from operation, and lowering the barriers of entering new railway operators into the Railway European Market. Let us be reminded that it is interpreted as: the European Railway Bodies now are supposed to consist of Infrastructure Manager and Railway Operators. In the case of Railway Freight Transportation this stands as “One Infrastructure Manager” and “Two or More Railway Freight Operators” with specific obligations to share the same infrastructure in providing their network-based businesses. Due to the nature of railway freight transportation the yards play a very important role for effective-good-quality freight transportation service and in gearing the new EU policies in the sector yards’ function cannot be ignored. However, the yards suffer a limited number of static and dynamic resources. The questions of improving operating efficiency and providing a service of good quality by the European railway freight operators become quite challenging when the same tracks at yards must be used by two or more railway freight operators. These circumstances open a range for further research.

References

[28] Nadel R. and Rover E., The Use of a Computer Simulation Model for Classification
23

[29] Nash C. and Rivera-Trujillo C., Rail Regulatory Reform in Europe – Principles and Practice, STELLA Focus Group 5 synthesis meeting, Athens, 2004
[31] Pachl J. and White T., Efficiency through Integrated Planning and Operation, International Heavy Haul Association, Virginia Beach, 2003
[34] Petersen E., Railyard Modeling: Part II. The Effect of Yard Facilities on Congestion, Transportation Science, Vol. 11, 1977
[37] Shalliker J. and Rickets Ch., An Introduction to SIMUL 8, School of Mathematics and Statistics, University of Plymouth, 1997
[38] Skalov K., Development and Reconstruction of Rail Stations and Junctions, Journal of Transport, Moscow, Russia (written in Russian), 1972
[42] Turnquist M. and Daskin M., Queuing Models of Classification and Delay in Railyards, Transportation Science, Vol. 16., 1982
[43] van Dijk N (2000) On Hybrid Combination of Queuing and Simulation, University of Amsterdam/Incontrol Business Engineers, the 2000 Winter Simulation Conference, the Netherlands

23