SIMULATION FOR THE EVALUATION OF OPTIMISED OPERATIONS POLICIES IN A CONTAINER TERMINAL

A. E. Rizzoli, L. M. Gambardella, M. Zaffalon, M. Mastrolilli
Istituto Dalle Molle di Studi sull'Intelligenza Artificiale
Corso Elvezia 36, Lugano, Switzerland
Email: { andrea , luca, zaffalon, monaldo}@idsia.ch

Keywords: intermodal terminal management, simulation and optimisation.

ABSTRACT

A simulation model must be used to evaluate the impact of new operations policies, not only to validate the policies, but also as a tool to convince the decision-makers of the potential advantages in adopting the proposed enhancements in the management. Terminal resource allocation policies and ship loading/unloading policies are obtained by means of Operations Research techniques for the case study of La Spezia Container Terminal (LSCT); a simulation model of the terminal is designed, implemented and validated. The simulation model is used to test the policies and to assess their robustness in front of the inherent stochasticity of the real world.

INTRODUCTION

The management of an intermodal container terminal is a complex task, which involves a great number of decisions to be taken at different levels, from strategic development down to the single move of a container. Most terminals rely on information management systems, often interfaced with automated data gathering devices, in order to take informed decisions. Almost as often, the sheer amount of the information makes nearly impossible for the human operators to see the "big picture", that is, the terminal in its complexity, considering the multiple interactions of the various concurrent processes, such as yard planning, resource allocation, ship loading and unloading. It is clear that Operations Research techniques could bring the advantage of making a better use of the available information and, consequently, to increase the overall performance. This is not an easy task, the mathematical models describing the terminal processes must be studied and designed with great care, often with a conflicting objective in mind: the model must include all the significant characteristics and, at the same time, it must be simple to be computationally solvable. During the model design phase we are obliged to make simplifying assumptions, which eventually may reveal themselves as correct, but which are often unacceptable from the point of view of the decision makers, since they might conflict with their personal experience. It is pointless to design a model in which the decision-makers have no trust, unless you convince them to trust it. Simulation is the tool that can be used to build this trust.

We have studied the intermodal container terminal of La Spezia, operated by Contship SpA, and in four years we have investigated different processes such as the container flow to and from the terminal, by land and sea, the yard storage allocation policies, the resource allocation policies, the load and unload scheduling policies. With the aid of the ISO 9001 documentation provided by LSCT we have tracked down the current decisional processes and we have noted that the terminal decision makers have adopted a distributed decision making scheme, centered on the ship loading and unloading process. This management choice has enabled the port to reduce the size of the problem, allocating decision making resources to each ship which "competes" for the accomplishment of their task: processing the whole ship in due time. While this approach guarantees the best service for the customers (the shipping companies) thus making La Spezia an "attractive" destination, often it neglects the possibility of synergies in the use of the yard resources. The historical data we have processed supported by the terminal operators' experience, shows that often there are resource conflicts on the yard cranes, which access the storage area of the terminal. It may happen that, despite the yard planners tend to organise the yard in order to de-couple the effect of two ships accessing the same storage area, two yard cranes must concurrently access the same area. For this purpose we have designed resource allocation algorithms and loading and unloading operations scheduling algorithms which will be described in the section titled “Synthesis of management policies”.

The models employed in the synthesis of the optimised policies are deterministic and include a set of assumptions which made possible the applicability of known algorithm for their fast and efficient solution (even if we had to devise an enhancement to classical job-shop scheduling theory). Even if these models were calibrated and validated before their employ, they do not intend to represent all the terminal behaviour, but to limit their applicability to the phenomena they want to describe. To assess the impact of the optimal policies obtained from these models, we have designed a micro simulation model of the container terminal. It is based on stochastic discrete-event simulation and it can be used to perform a trace-driven simulation, using the historical data sets, which are continuously collected during terminal operations. The terminal model is able to operate to replicate the current "distributed" operations control policy, but also to incorporate the optimised policies, to compare their effect with the current management practices. In section “Simulation” we describe the design and the implementation of the discrete event simulation model.

The model has been calibrated on one week of terminal operations, while another week has been used to validate it. A panel of terminal experts has positively judged the validation results. The model has been used to show to the panel the expected impact of the proposed policies and how the terminal performance may increase while the operation costs decrease, thanks to a more coordinated usage of the terminal resources. The model has also been employed to test the robustness of the policies, in front of unexpected events and of variable service times. These results are provided in subsection “calibration and validation” and they are discussed in the final “Discussion” section.

THE PROBLEM AND THE CASE STUDY SITE

Containers arrive at La Spezia Container Terminal (LSCT) by train, vessel or truck and are stored in the terminal yard. Containers then leave the terminal by the same transport means to reach their next destinations. The flow of containers is composed of an *import flow*, i.e. containers unloaded from ships, to be either transshipped or directed to the final destinations by trucks and trains, and an *export flow*, i.e. containers loaded on ships leaving the terminal.

In the LSCT, containers are stacked up to the fifth level on the yard by rail-mounted cranes (*yard cranes*) which unload trucks and trains. This stack height is quite unusual and is due to the lack of space on the yard. LSCT is a terminal with a high traffic on a small yard and therefore the management of space is a critical issue. *Quay cranes* unload vessels and place containers on *shuttle trucks*, which move them to storage locations in the yard. Loading a vessel is a similar process, where the shuttle receives the container from the yard cranes and moves it to the proper quay.

The problems of the LSCT terminal are related to the current workflow management procedures. All the work processes relative to a ship are planned, supervised and managed by a single *ship planner*, who does not know the decisions made by other ship planners, which are working on other ships present at the same time in the terminal.

After an analysis phase, we identified three areas where mathematical models, optimisation and simulation could be useful: dynamic allocation of storage areas on the yard, allocation of the resources required to perform the ship loading and unloading (L/U) operations, and finally the scheduling L/U operations. It is easy to show that these management actions are interdependent, a modification of the policy used to decide where to place a container on the yard has a direct impact on the resource (crane) which will move it. In the same way, if the set of resources assigned to unload and load a ship varies, the optimal sequence of L/U operations will change.

For simplicity, we decided to analyse the three problems separately, since their time scales are different: decisions regarding yard space planning have an horizon of some days, while resource allocation is made on an horizon of 24 hours and L/U operations are scheduled before each work shift, that is, every six hours.

In this paper we present the results relative to resource allocation and L/U operations scheduling, since we haven’t yet tested the effectiveness of the yard allocation policies.

SYNTHESIS OF MANAGEMENT POLICIES

In this section we summarize the work done in the synthesis of the resource allocation and scheduling policies. We also outline how the two policies can be used together to take operational decisions in the terminal.

Resource Allocation

The role of the resource allocation (RA) module is to determine the best allocation of resources for vessel loading and unloading operations, with the objective of maximizing the profit, given by the difference between income and expenses (Zaffalon and Gambardella, 1998, Zaffalon *et al*. 1998). The income is a term proportional to the number of moved container, whereas the expenses are a linear function of the allocated resources. Since more resources produce a greater movement capacity, it is clear that the RA problem corresponds to find the right balance between moving containers while saving resource costs.

As far as RA is concerned, the terminal can be interpreted as a mechanism which routes the container flows from their sources to the proper destinations. This view models the terminal as a network of flow (Papadimitriou and Steiglitz, 1982). In such a network, the transport capacities of the arcs are functions of the number of resources that are allocated: hence, the focus is on dimensioning arc capacities to maximise the profit. The latter problem is a particular case of the so-called network design (Magnanti and Wong, 1984, Ahuja *et al.*, 1993). In the terminal case, the graph of the network is also extended along the different work shifts (1 shift = 6 hours) in order to represent time (thus allowing the allocations to be computed over all the period under study).

The resulting model is a complex mixed-integer linear program, whose solution is demanded to the branch & bound (Papadimitriou and Steiglitz, 1982) capabilities of Cplex 5 (Ilog, 1997). The problem is solved in an approximate way, since getting optimality is a time-consuming task in our experience, because the complexity of the problem produces a huge search tree. Notwithstanding, a good solution is generally found quickly (few minutes on a Pentium 133 MHz, 32 Mbytes of RAM); experimental results based on real cases of the LSCT show that the best integer value found is usually *close* to the continuous bound.

Scheduling o L/U Operations

The RA problem is necessarily a high-level problem, since the resources must be allocated at least 24 hours (up to a week) in advance, when the data describing the ships and the yard state are only partially known. When the ships arrive to the terminal, the load/unload process is based on the resources previously allocated and on the updated state of the terminal. In particular, on the basis of the latter information, the scheduling module must build the Loading and Unloading List (LUL) for all the ships at the terminal. The LUL is a document that precisely specifies the origin and the destination of every container to be moved to and from the moored ships. As in the RA problem, the focus is on the proper way of sharing resources in order to produce a LUL that services the ships within their (given) deadlines.

The difference with the RA problem is that available resources are needed as input to the LUL module. Hence, it does not take into account economic factors, since the expenses are already sustained at the time of the allocation of resources. Rather, it must fill the gap between the two different views of the terminal (high and low-level).

Our scheduling module is based on a Flexible Job-Shop (FJS) model, which is extended in order to take into account set-up times of the resources and to dynamically treat the evolution of the terminal. It is heuristically solved by local search, according to an original approach (Mastrolilli and Gambardella, 1998). In the FJS problem *n* jobs are processed to completion on a set of unrelated machines. Each job consists of a sequence of ordered operations where each operation has to be executed on a machine out of a set of machines. The processing time depends on the chosen machine. Each machine is always available and it is allowed to process one operation at a time without preemption. The objective is to define the operation sequence on each machine in order to minimize the maximal completion time of all operations (makespan).

In our case a job is the sequence of loading and unloading operations performed by a quay crane. An operation is a container move. From an analysis of the real world data, the yard cranes appeared as the bottleneck machines for the problem, thus suggesting to model them as machines in a FJS based model. The goal is to maximize their performances.

Hence, the problem is solved finding which yard crane moves which container (routing problem) and when (sequencing problem) in order to minimise the maximum completion time of all operations. The processing times of each machine are assumed deterministic. We have tuned and validated our model using real world data. Our optimisation procedure is able to improve the yard crane performances of about 31%. The computational effort is very low, about 1 minute on a 266 MHz Pentium for each shift. The latter feature makes it possible to use the proposed procedure as a reactive scheduler in order to manage unexpected events

SIMULATION

The simulator is used as a test bench to evaluate the management policies produced by the optimisation modules (Gambardella *et al.*, 1998). The performances of various resource allocation policies, used in conjunction with computer generated L/U lists, are compared with real world data. The indicator used to compare the solutions is the net profit of the terminal operations during the simulation horizon. This indicator takes into account the cost of cranes and operators during the various work shifts, the penalty to be paid to the shipping company if the ship departure is delayed and the income generated by each container loaded and unloaded from a ship. Besides these economical indicators, the simulator allows to assess the resource utilization and the measure important congestion indicators such as the average queue length of operations on the terminal cranes. The simulation model has been implemented using MODSIM III (Caci, 1997).

The model

The simulation model tries to replicate the terminal activities and it is based on the principle that external events generate responses by the simulation agents, which in turn operate on simulation components. The responses of simulation agents are determined according to the policies that can either be generated by the optimisation modules or by a representation of the experience of terminal operators.

External events are: trucks arriving at terminal gate; ships arriving at terminal pier; trains arriving at terminal. The *arrival generator* is a part of the simulation module that generates these arrivals. Ship and train arrivals are read from a database, since they are known in advance, while truck arrivals are generated according to statistical distributions.

When a ship, train or truck enters the terminal, it has a list of containers (or just one, in the truck's case) which is imported and a list of containers to be exported. The yard and ship planners use these lists.

The *ship planner* is a simulation agent dedicated to organise the loading and unloading operations of a ship. The ship planner performs the following tasks:

1. Allocate the quay cranes work shifts needed to load and unload the ship, given the ship import and export list. This task can be performed either using the resource allocation optimisation module or by entering the resource allocation strategy decided by the human operator.

2. Compute the bay plan. Resource allocation assigns to each quay crane a set of bays to work on. In general, unloading occurs before loading, and these two activities must respect the ship structural stability constraints, these constraints result in the work order of the bays.

3. Ask the yard planner to assign destinations on the yard to the containers to be unloaded. These containers are unloaded in order as stowed. The unloaded containers will be stored in sub-regions of the yard areas, named *import areas*. The size and location of import areas is a decision variable.

4. Communicates to the yard planner the containers to be loaded. This list is ordered by a set of constraints which imposes a sequence to be respected in stowing containers aboard according to their size, weight, port of destination, and to a series of distinctive characteristics such as hazard class, kind of good transported, etc.

5. Put the quay cranes to work according to the plan previously determined. Supervise loading and unloading operations, collect statistics and evaluate performance.

The yard planner uses the lists of import and export containers (see items 3 and 4 in the previous numbered list) to build the schedule solving the job-shop problem associated with yard crane operations.

The task of the *yard planner* simulation agent is to organise the container allocation on the yard in order to maximise yard crane performance, avoid crane deadlocks (when two cranes try to work on the same yard area), and minimise the time to access containers during storage and retrieval. In detail, its tasks are as follows:

1. Allocate the yard cranes work shifts, given the list of containers to be loaded and unloaded by all the ships and trains that are present or are due to arrive. Again, the resource allocation module can provide these decisions.

2. Organise the yard space according to a given policy.

3. Solve the job-shop scheduling problem, using the available data on trains, trucks and ships to be loaded and unloaded. The result is the work list for each yard crane (the ordered list of containers to be moved). These work lists are computed using the reactive scheduling algorithm implemented in the optimisation module. This centralised policy can be replaced by a distributed policy generated by local rules used by crane agents. This latter behaviour describes the current decentralized management approach.

Besides this high level management performed by the ship and the yard planners, there are the local management decisions taken by "less intelligent" simulation agents such as cranes and shuttle trucks.

Quay cranes start to work when the ship planner assigns them a list of containers to be loaded and unloaded. They stop working when they have finished to process their lists. Quay cranes move containers to and from shuttle trucks running between the quays and the yard cranes. When the quay crane unloads a container, it asks the yard planner which yard crane is assigned to it; the truck will therefore travel to the yard area where that yard crane is working.

Yard cranes pick up and put down containers on the yard. They have a queue of jobs to be performed. A job is a container movement, either picking it up from a truck and placing it on the yard or vice versa, and even temporary moves to unpack stacked containers are jobs. As we have seen before, this queue of jobs (the work list) can be automatically optimised by a job-shop scheduling, or can be managed by local rules, which try to emulate the behaviour of the human operator. Yard cranes are also provided with tie-breaking mechanisms to avoid deadlock: it can happen that, given the randomness associated with the time a crane moves a containers, the job queues push the cranes towards a conflict, such as trying to move two containers which are stored in the same bay in the same time. The yard cranes can acknowledge this potential deadlock and reassign one of the container moves to contiguous crane (this is a sub optimal solution, but avoids computing again the whole job-shop problem).

Calibration and Validation

Before assessing the validity of the resource and L/U policies, the simulator must be calibrated and validated in order to verify its capability of reproducing the real terminal behaviour (Banks *et al.* 1996).

The calibration and validation data were extracted from a period of two weeks, from 5/11/1998 to 5/24/98. The data describe the activity of LSCT in great detail, in such a way that every container movement can be tracked down. Notice that in the calibration and validation phase, the simulation module evolves using the same inputs and policies used by the terminal management over the calibration and validation periods. For this reason, the resource allocation and the L/U policies are the ones adopted by terminal managers during the periods when calibration and validation data were collected.

We focused our attention on a *trace-driven* simulation for the purpose of generating a test environment for the management policies.

After performing a set of experiments, we obtained for the quay cranes an average calibration error of 13.6% and a

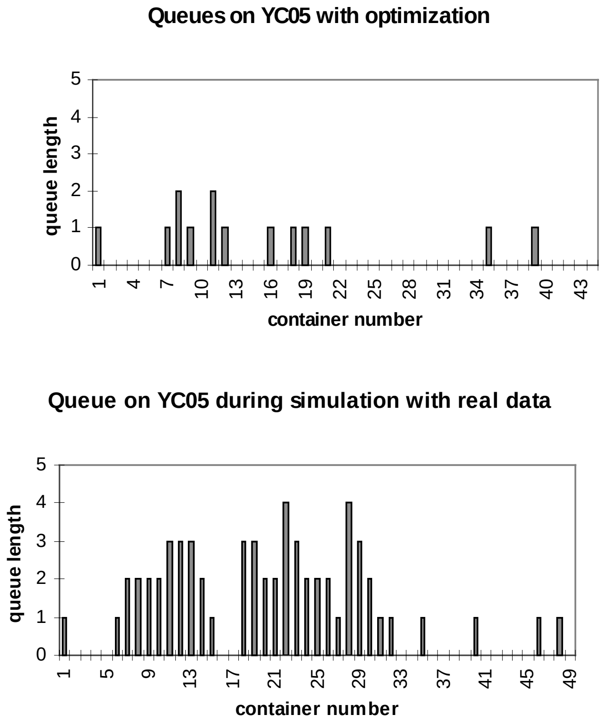

Figure 1. Queues of containers under the yard cranes.

validation error of 14.2 %. For the yard cranes, the results are slightly better, 6% in calibration and 13% in validation. These results deserve a few words of comment, since they seem susceptible to improvements. An important remark is that our main objective is to obtain a model of the terminal to test alternative scheduling policies. For this purpose, the model, as a first approximation, does not handle unexpected and rare events such as crane breakdowns, and inexperienced operators. Unfortunately, such events happen in the real world terminal, and this is evident, for instance, in shift 1 on crane Q6, where the crane performance is well below average, despite the terminal is working at half capacity since this is a shift when trucks and trains do not arrive. A more detailed analysis of the real world data helps to discover that Q6 moves fewer containers than the average since it stops working for some periods during the work shift.

We are interested in designing a simulation model that represents the "average" behaviour of the terminal, where the only bottlenecks can be caused by resource competitions on the yard, since we are designing scheduling policies for a terminal where the cranes are operating. Scheduling in front of unexpected events, such as breakdowns, would be the subject of a research into reactive scheduling techniques.

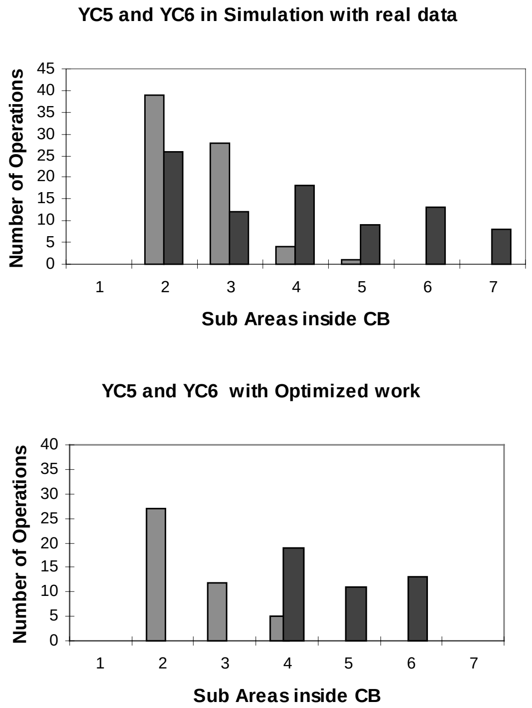

Figure 2. Areas of activity of the yard cranes.

DISCUSSION

We have performed the cyclic process of "allocate schedule simulate" described in the previous section in order to discover if the computer generated management policies were feasible once implemented in the low-detail simulation model. Moreover, we were interested in discovering how the terminal would have reacted to the implementation of such policies, especially for what it concerned the behaviour of the yard cranes.

The computer generated RA policies allowed an average saving of 30%. This means that fewer resources were used to perform the same tasks. It is important to notice that the computer generated RA does not impose a quay crane intensity (number of container moved per hour) higher than the one normally attained by the real world terminal. The RA module uses the real world crane intensities as a parameter to fix the arc capacities. Since the improvement is not to be found in an increase of the quay crane intensities, it is instead due to a more rational use of night shifts (which are more expensive) on one hand and by a better organization of the yard crane work on the other. In fact, it turns out that optimised LUL allow a more efficient use of yard cranes, thus increasing their intensity. This is possible since the yard crane "conflicts" (i.e. when two yard cranes try to access two containers which are very close at the same time) are dramatically reduced. The reason for the reduction is shown in Figure 2: without optimisation, yard cranes YC05 and YC06 tend to cover the whole area CB, while, with optimisation, YC05 covers only the left hand side of CB and YC06 the right hand side, thus virtually eliminating the possibility of conflicts. Another desirable side effect of optimisation is the average reduction of queue lengths under yard cranes: a better balance in the crane usage reduces the possibility of having lengthy queues, as it is shown in Figure 1.

In conclusion, simulation results show that the application of computer generated management policies could improve the terminal performance, making possible the allocation of fewer resources, thanks to a better usage of the yard cranes.

ACKNOWLEDGEMENT

This research is supported by Swiss UFES project 97.0315 under the European Project PLATFORM.

REFERENCES

Ahuja, R., T. Magnanti, J. Orlin. 1993. "Network Flows", Prentice Hall, Upper Saddle River, NJ.

Banks, J., J.S. Carson, B.L. Nelson. 1996. "Discrete-Event System Simulation, 2nd edition". Prentice Hall, Upper Saddle River, NJ.

CACI. 1997. "MODSIM III User's Manual". CACI Products Co. La Jolla, CA.

Gambardella, L.M., A.E. Rizzoli, M. Zaffalon. 1998. "Simulation and planning of an intermodal container terminal". *Simulation,* no. 71, vol. 2: 107–116.

Ilog. 1997. "Cplex 6 User's Manual", http://www.cplex.com

Magnanti, T.L., R.T. Wong. 1984. "Network design and transportation planning: models and algorithms." *Transportation Science*, no. 18, vol. 1: 1–56.

Mastrolilli, M., L.M. Gambardella. 1998. "New neighborhood function for the flexible job-shop problem." Technical Report 5-98. IDSIA, Lugano, CH.

Papadimitriou, H., K. Steiglitz. 1982. "Combinatorial optimization: algorithms and complexity". Prentice-Hall, Upper Saddle River, NJ.

Zaffalon, M., L.M. Gambardella. 1998. "A network design approach to the allocation of resources in an intermodal terminal." Technical Report 10-98. IDSIA, Lugano, CH.

Zaffalon, M., A.E. Rizzoli, L.M. Gambardella, M. Mastrolilli. 1998. "Resource allocation and scheduling of operations in an intermodal terminal". In Proceedings of the 10th ESS (Nottingham, UK, Oct 26-28). SCS International, Delft, NL.