DSS (Decision Support Systems) for Operations in a Container Shipping Terminal

Katta G. Murty[1], Jiyin Liu[2], Yat-Wah Wan[2], Chuqian Zhang[2],
Maria C. L. Tsang[2], and Richard Linn[3]

[1] Author for correspondence, Department of IOE
University of Michigan, Ann Arbor, Michigan 48109-2117, USA
Phone: 734-763-3513; fax: 734-764-3451; e-mail: katta_murty@umich.edu
webpages: http://www-personal.engin.umich.edu/~murty/
http://www.guthy.com/murty/index.html
[2] Department of IEEM, Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong, China
[3] Department of ISE, Florida International University
Miami, Florida, USA

29 December 2000

Abstract

We describe the variety of decisions to be made in the course of daily operations at a container shipping terminal. All these decisions are inter-related. The ultimate goal is to make all these decisions so as to: minimize time taken to process the vessels, minimize the resources used to handle the workload, minimize the wait time of customer trucks, minimize the congestion on the roads inside the terminal, and finally to make the best possible use of the storage space available. Given the scale and complexity of these decisions, to further enhance the operational efficiency of container shipping terminals, it is essential to bring in decision support tools. This paper reports on a DSS that we are developing for this purpose. We discuss the mathematical models and the algorithms being used in designing the DSS, and some of the results.

Key words: Container shipping terminals, port time (turnaround time) of vessels, quay crane rate, congestion, routing, reshuffling, storage space allocation, crane deployment, truck allocation to cranes, truck hiring plans .

1 Introduction

Most of the cargo transported in ocean-going vessels around the world today can be classified into two types:

- bulk shipping of huge quantities of commodities like crude oil, coal, ores, grains, etc. which are shipped using specialized vessels called bulk carriers
- containerized shipping in which a variety of goods are packed into standard size steel containers that are shipped on vessels.

In this paper we will focus on containerized shipping. A **container shipping terminal** (or **terminal** in short) in a port is the place where sea-going vessels dock on a berth and unload inbound containers (empty or with cargo) and pick up outbound containers. The terminals have large storage yards for the temporary storage of these containers (Figure 1). [The pictures in Figures 1 — 4 are taken from the website of Modern Terminals Limited (http://www.mtl.com.hk/) for examples.]

Figure 1. The container storage yard in a terminal with containers stored in columns or stacks of 4 to 6 containers one on top of the other. RTGC (Rubber Tyred Gantry Cranes), that help stack containers and retrieve them, can also be seen among the containers. The storage yard is divided into rectangular areas called blocks.

Containers are steel boxes of dimensions (all measurements are in feet) $20 \times 8 \times 8.5$ or $20 \times 8 \times 9.5$ (called **TEU** meaning 20 ft. equivalent units), or $40 \times 8 \times 8.5$ or $40 \times 8 \times 9.5$ (called **FEU** meaning 40 ft. equivalent units), or specialized slightly larger size boxes (for example refrigerated containers for cargo that must be kept at specified cold temperatures during transit). In the past the TEU used to be the most common container, but now-a-days the FEU is beginning to dominate the scene.

The storage yard in a terminal is usually divided into rectangular shaped regions called **blocks**. Each block has seven rows of spaces, six of which are used for storing containers in **stacks** or **columns**, and the seventh reserved for truck passing. Each row typically consists of 26 stacks with TEUs stored lengthwise side by side. If the same row is used for storing FEUs, the number of stacks in it will be 13.

In each stack, containers are stored one on top of the other. The placing of a container in a stack, or its retrieval from the stack, are carried out by a huge crane called **RTGC (Rubber Tyred Gantry Crane)** that moves on rubber tyres. The RTGC stands on two rows of tyres that contain the seven rows of spaces between them (Figure 2). The bridge (top arm) of the RTGC has a spreader (container picking unit) that can travel across the width of the block between rows 1 to 7 on that bridge. The RTGC can move on its tyres along the length of the block. With these two motions, the RTGC can position its spreader to pick up or place down a container in any stack of the block, or on top of a truck in the truck passing row.

The height of an RTGC determines the height of each stack (i.e., the number of containers that can be stored vertically in a stack). Older models of RTGC are 5-height RTGC. It can only store 4 containers in a stack, the 5th height is needed for container movement. Newer models are 6-height RTGCs, they can store 5 containers in a stack, and use the 6th height for container movement.

Some blocks are served by fixed **RMGC (Rail Mounted Gantry Cranes)** with 13 rows of spaces between their legs and higher storage height (6-height). The RMGC are fixed to a block,

Figure 2. An RTGC sits across the width of a block with the 7 rows of spaces between its two legs. 6 of these rows are used for container storage, the 7th (rightmost in the figure, a truck parked in it can be seen) for truck passing.

but the RTGC which move on rubber tyred wheels can be transferred from block to block offering greater flexibility. Because of this flexibility for movement, now-a-days the RTGC is the most commonly used equipment for storage yard operations. The RTGC and RMGC are both called **YC (Yard Cranes)**.

Customers bring outbound containers to the terminal, and take away inbound containers from the terminal, on their own trucks, which are called **XTs (External Trucks)**. Within the terminal itself containers are moved using trucks known as **ITs (Internal Trucks** or **Stevedoring Tractors)**.

The unloading of containers from a vessel, or the loading of containers into a vessel, are carried out by huge cranes called **QC (Quay Cranes)** (Figures 3, 4).

This paper describes the work being carried out to develop DSS (computerized decision support systems) for optimizing the operations of container shipping terminals in Hong Kong.

Hong Kong is the busiest container port in the world. Container shipping is the lifeblood of Hong Kong's economy. Being the principal entry port in Southern China region, a region having a strong economic growth, the volume of container transport through Hong Kong has been increasing by 10% yearly since 1986. The throughput in 1999 totaled 16.1 million TEUs, and is estimated to reach 32.8 million TEUs by 2016 (Report of Hong Kong Port Development Board, 1998). The intensity of container traffic in Hong Kong is estimated to be seven times that of New York, and the cramped space in Hong Kong yards makes it very challenging to provide high quality service.

The container port in Hong Kong is currently run by four terminal operators, HIT, MTL, Sea-Land, and HIT-Cosco. The daily operations in these companies are extremely complex. For example, one of these companies operates 10 berths that handle about 100 vessels per week from more than 30 major shipping lines. The company moves about 10,000 containers/day at vessel side. On average about 10,000 trucks pass through their gates daily for bringing outbound containers or picking up inbound containers. Their container storage yard is divided into 70 blocks (each can store about 600 containers) served by RTGC, and 12 other blocks (each can store about 1300 containers) served by RMGC. Their storage yard is typically 70% full. Normally they have at least one vessel being loaded and unloaded in their terminal all the time. They use more than 30 QCs, 100 RTGCs, 20 RMGCs, and 160 ITs for their work, and hire additional ITs on a daily basis as needed. The terminal has about 10 miles of roads which sometimes get congested with the traffic of XTs and ITs.

Given the scale and complexity of their operations, to further enhance their operations, it is

Figure 3. A QC, and truck lanes underneath it for the parking of ITs serving it. A vessel docked on the berth can be seen ahead. The QC unloads inbound containers from the vessel and places it on an IT which takes it to the storage yard for storage until it is picked up by the consignee. ITs also bring outbound containers from storage yard for the QC to load into the vessel. On the left is a portion of the storage yard with stored containers.

Figure 4. A QC unloading inbound containers from a vessel. ITs serving the QC are queued up under the QC.

essential to bring in computerized DSS tools. This paper reports on the DSS that is under development to help these companies.

2 The Yard Operations

The Functions of a Container Terminal

A continuer terminal serves as an interface between ocean and land transportation. Its main functions are:

- Receive outbound containers from shippers for loading into vessels, and unload inbound containers from vessels for picking up by consignees.
- Temporary storage of containers between ocean and land transportation.

The Flows of Outbound and Inbound Containers

Outbound containers being brought in by customer XTs enter the terminal through the **TG (Terminal Gatehouse)** where the container and its documentation are checked. The TG then

Arrival & storage					Retrieval & loading			
Cust.	$\rightarrow$ XT	TG	$\rightarrow$ XT	SY	$\rightarrow$ IT	QS	$\rightarrow$ QC	Vessel
		Document Inspection Stor. space assignment		YC				

Figure 5. Flow of outbound containers. Cust. = Customer, SY = Storage Yard, QS = Quayside. Underneath each location or operation, we list the equipment that handles the container there.

Retrieval & pickup					Unloading, storage			
Cust.	$\leftarrow$ XT	TG	$\leftarrow$ XT	SY	$\leftarrow$ IT	QS	$\leftarrow$ QC	Vessel
				YC				

Figure 6. Flow of inbound containers.

instructs the XT to proceed to the storage block where the container will be stored until the vessel into which it will be loaded arrives. The RTGC working at that block removes the container from the XT and puts it in its storage position. When the time to load comes, the RTGC removes the container from the stored position, puts it on an IT, which takes it to a QC for loading into the vessel. So the flow of outbound containers can be represented as in Figure 5.

The flow of inbound containers has the reverse features as indicated in Figure 6.

How is a Containerized Vessel Organized ?

Inside a vessel containers are stacked one on top of the other. The vessel is divided along its length into several storage areas known as **hatches** or **holds** or **bays**. The number of hatches in a vessel may be as high as 15 depending on its size. Some of the big vessels may carry over 7000 TEUs. A vessel may call on 5 to 10 ports in a sailing. For quick unloading, loading, they usually assign containers to hatches according to ports of call. When the vessel is docked on the berth, normally 3 to 4 QCs work on it simultaneously, each QC working on a separate hatch. The unloading, loading sequence of containers is usually determined by a special algorithm to make sure that the docked vessel's balance is not affected while it is on the berth.

RTGC Operations

The RTGC are very expensive equipment whose proper utilization is very critical to the efficiency of container handling operations in the storage yard.

The RTGC working in a block move containers from ITs or XTs and put them in their storage locations, or retrieve stored containers and put them on ITs or XTs. ITs and XTs which arrive at this block to deliver or pick up a container, queue up in the truck passing lane of the block until an RTGC working in that block can serve them. Thus if RTGC are not efficient in their work, there may be congestion of trucks on the road near its block. Also, if the RTGC holds up the ITs serving a QC working on a vessel, the QC may have to wait, resulting in a delay in unloading, or loading the vessel.

The container retrieval operations of an RTGC are of two types:

- A **productive move**, if the container is moved directly from its storage location to an IT or XT waiting to pick it up. For example to retrieve the top container A from the stack of four stored containers shown in Figure 7 will be a productive move.
- An **unproductive** or **reshuffling move**, if the movement of this container, say A, becomes necessary to retrieve another container stored underneath it in the same stack. In this reshuffling move container A is placed in another stack in the same bay, and may be put back in the original stack after the desired container is retrieved. For example to retrieve container C in Figure 7, containers A, B stored above it have to be moved first in reshuffling moves.

A
B
C
D

Figure 7. Four containers stored in a stack.

The number of reshuffling moves depends a great deal on the strategy used for allocating storage spaces to arriving containers.

RTGC movement from one block to another is slow. If two adjacent blocks share a width line (as B_1, B_2 in Figure 8) the RTGC can move among them straight on its tyres without any turning motion. To move from a block to an adjacent one sharing a length line with it (as B_1 and B_3 in Figure 8), the RTGC has to come to the road on one end of the block, make a 90^0 turn of its wheels, then move on the road parallel to the width line to the correct position for the adjacent block, make a 90^0 turn back of the wheels, and enter that block. These 90^0 turns take extra time, and also obstruct traffic on the road for all that time.

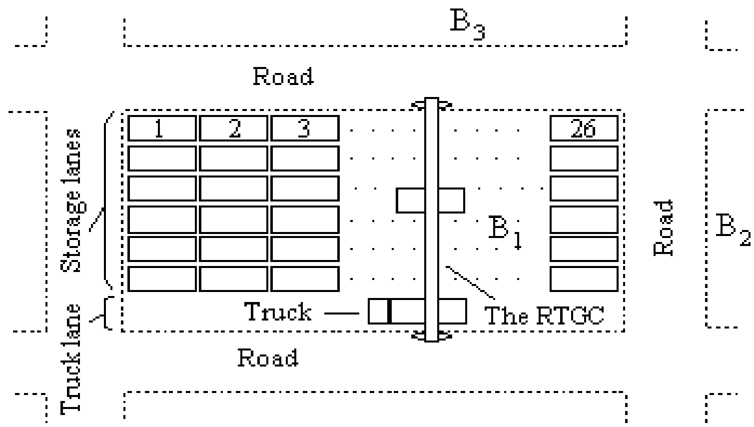

Figure 8. Top view of a block B_1 being served by an RTGC. Adjacent block B_2 shares a width line with B_1; adjacent block B_3 shares a length line with B_1.

Key Performance Measures of a Container Terminal

Container terminals work under multiple operational objectives. The most critical performance measure for rating the terminals is the **ship turnaround time** (also called **port time of the ship**) which is the average time the terminal takes to unload, load a docked vessel. This important objective function is to be minimized.

Closely related to the ship turnaround time, another important measure is the **QC rate**, which is the terminal's throughput measure during a period, given by

$$\text{QC rate} = \frac{\text{no. of containers unloaded, loaded}}{\text{no. of hours all QCs worked}}$$

which is to be maximized.

With so many trucks operating, the roads in the terminal get congested. Congestion slows the trucks from carrying out their operation of transporting containers from one location to another. This has an undesirable effect on truck utilization, and more importantly on the time taken to process the vessels. Hence another important measure of performance is **congestion on the roads** in the terminal, which is to be minimized.

A Quantitative Measure of Congestion

Congestion is an easily recognized intuitive concept, but in order to optimize it, a quantitative measure for it in terms of the decision variables that can be controlled in terminal operations, has to be developed. We will now describe a measure for congestion developed in Murty[1997].

We represent the gate, the berths, the blocks, and all road intersections by nodes in a network. A road segment is a portion of a road joining two such nodes. If the road is two-way, each direction of it is considered a separate road segment. Thus each road segment can be traveled in only one direction, and can be represented by a directed arc joining the two nodes on it. This leads to the representation of the road system inside the terminal as a directed network which we denote by G.

The most common vehicles using the road network G are trucks transporting containers. We will measure flows on arcs in G in units of trucks traveling through those arcs per unit time. Let s denote the number of arcs in G, and for $k = 1$ to s, let

f_k = flow (number of trucks traveling/unit time) on kth arc.

As the number of containers handled by the terminal increases, flows on the arcs in G are expected to increase. We are interested in a measure of congestion in the whole road network G, rather than on particular arcs. For a given volume of containers handled, we can expect the best traffic situation to prevail if the flows along the various road segments are as close to each other as possible. So, a measure of congestion in the road system inside the terminal is $\theta - \mu$, where

$$\theta = \text{Max}\ \{f_k : k = 1, \ldots, s\},$$
$$\mu = \text{Min}\ \{f_k : k = 1, \ldots, s\}.$$

We will use $\theta - \mu$ as the measure of congestion to minimize.

Other Important Measures of Performance to Optimize

Some of the other measures of performance to be optimized are given below.

- Average waiting time of XTs that come to the terminal to deliver outbound containers or pickup inbound containers, to be minimized.
- Average waiting time of the ITs in queues at the QC and RTGC waiting to be serviced, to be minimized.
- Waiting time of the QC waiting for an IT, to be minimized.
- Volume of reshufflement in the storage yard, to be minimized.
- Total number of ITs used in the various shifts each day, to be minimized.

Optimizing all these performance measures requires good resource allocation decisions, i.e., allocation and scheduling of berths, QC to docked vessels, storage spaces to containers, RTGC deployment among blocks, IT allocation to QCs, and allocation of routes to traveling XTs and ITs, so as to achieve a smooth and orderly flow of containers between the gatehouse, the storage yard, and the quayside.

3 Decisions to be Made in Daily Operations

Container terminals work round the clock, every day. They work three 8-hour shifts each day. each shift is divided into two 4-hour periods. The terminals find it very convenient to organize their planning for work for each 4-hour period separately. Hence we take our **planning horizon** for decision making to be a 4-hour period. A day is divided into six planning periods: 00:00 − 4:00, 4:00 − 8:00, 8:00 − 12:00, 12:00 − 16:00, 16:00 − 20:00, 20:00 − 24:00.

There are a variety of decisions to be made each day. These are described below.

D1. Allocation of berths to arriving vessels: Allocate berths to arriving vessels so as to minimize ships waiting time, port cost, and the dissatisfaction felt by vessel's captains about berthing order; and to maximize utilization of berths and QCs.

D2. Allocation of QCs to docked vessels: Develop a plan to allocate and schedule QCs to work on docked vessels. This decision influences the turnaround time of vessels, and the throughput rate of the terminal.

D3. Appointment times to XTs: Ideally, all customers book the time to deliver outbound containers, or to come to pick up their inbound containers, by calling beforehand and taking appointments. Develop simple online decision rules to generate these appointment times when they call, to help to minimize XT waiting time, and congestion in the road network.

D4. Route trucks: Develop an algorithm to route XTs and ITs inside the terminal to minimize congestion on the roads.

D5. Storage space assignment: Assign storage spaces to arriving containers to minimize reshuffling volume and congestion on the roads.

D6. RTGC deployment: Determine how many RTGC should work in each block, and when to move an RTGC from one block to another. This decision influences the port time of vessels, and the waiting times of QC, XT, and IT.

D7. IT allocation to QCs: Determine how many ITs to allocate to each working QC to minimize the vessel turnaround time, and the waiting times of IT and QC.

D8. Optimal IT hiring plans: Estimate IT requirements in each half-hour interval of the day, and develop a plan to hire ITs to meet these requirements, to minimize the total number of ITs hired each day, and to maximize IT utilization.

D9. Resource acquisition: Determine whether to acquire expensive equipment like the RTGC, QC, in order to improve performance and resource utilization.

These decisions are inter-related. But the very large number of decisions involved, the multi-objective nature of the problem, the uncertainty, and the complexity of decisions, makes it impossible to derive answers to all the decisions by solving a single mathematical model. The only practical way to obtain reasonable answers to these decision problems is to study each of them separately in a hierarchical fashion. We found the technique called hierarchical decomposition with substitute objective functions for each stage developed in Murty and Djang[1999] very useful in handling these decision making problems.

The development of DSS to make all these decisions optimally is an ongoing effort. So far we have investigated strategies for making decisions D4 to D8, because these decisions involve resource planning issues in the storage yard, which are issues of primary interest to the terminals in Hong Kong. Next we will describe the algorithms being implemented in the DSS for making these decisions.

4 Storage Space Assignment and the Routing of Trucks in the Storage Yard

Since storage spaces are the destinations for truck travel within the terminal, the two decisions of storage space assignment and the routing of trucks are highly inter-related, and we study them together.

For storage space allocation decision, it is convenient to group outbound containers to be loaded into the same bay of the same vessel for unloading eventually at the same port (all of which have the same time for retrieval from storage at the terminal) into a set of containers called a **consignment**. In the same way inbound containers being unloaded from a vessel that will be picked up by the same customer or trucking company are grouped as a consignment. The important property of a consignment is that all the containers in it have the same retrieval time from storage, and when that time comes, the order in which they are retrieved is unimportant, i.e., there will be no reshuffling moves in retrieving them even if they are stacked one above the other.

The planning period for these decisions is a 4-hour period (the beginning or the ending half of a shift) as discussed in Section 3. The input data needed for the models consists of: the road network G discussed in Section 3, the appointment time database (the list of customers who have been given appointments during the planning horizon and the outbound containers they are expected to bring in, or the inbound containers in storage they are expected to pick up), the data on the retrieval times for the containers in storage (from the vessel arrival time data and the loading order for outbound containers, appointment time data for XTs picking up inbound containers), and other data elements listed below.

$b =$ Number of berths.

$m =$ Number of blocks in the yard.

$k_i =$ Container storage spaces in block i, $i = 1$ to m.

$F_i =$ **Fill-Ratio** in block i = (number of containers in storage in block i at that time)/k_i, $i = 1$ to m. This is obtained by updating the inventory of containers in storage.

$ES_i =$ Number of outbound containers stored in block i to be retrieved during this period for loading into a vessel, $i = 1$ to m.

$ES_i^d =$ Number of containers among ES_i that have to go to berth d, $d = 1$ to b.

$IS_i =$ Number of inbound containers in storage in block i expected to be retrieved during this period for pickup by customers, $i = 1$ to m.

$IS =$ $\sum_{i=1}^{m} IS_i$ = total number of inbound containers to be picked up by customers during this period.

$IM =$ Total number of inbound containers expected to be unloaded from vessels during this period.

$IM^d =$ Number of inbound containers to be unloaded from vessels docked at berth d, $d = 1$ to b. We have $\sum_{d=1}^{b} IM^d = IM$.

$EX =$ Number of outbound containers expected to be brought in for storage during this period.

Storage Space Allocation Decision

Each of the $IM + EX$ new containers arriving at the terminal during this period has to be assigned a specific storage space in the yard for storing it. This assignment policy could have an impact on congestion in the road network, and on the total volume of reshuffling, both of which have to be minimized.

There are a large number of container storage spaces in the yard, some of which are already occupied by containers in storage at the beginning of this period. Determining a specific open storage space for each of the newly arriving $IM + EX$ containers leads to a huge mathematical model which will be impractical to solve. So, in Murty[1997], this storage space assignment decision has been divided into two stages.

Stage 1: Block Assignment: Determine how many of the containers among the arriving $EX + IM$ will be stored in block i; these lead to the following decision variables for this stage.

EX_i [IM_i] = Total number of outbound [inbound] containers arriving at the terminal in this period to be stored in block i, $i = 1$ to m.

Stage 2: Storage Position Assignment: For each container arriving at block i for storage, determine the optimal available position in the block for storing it, to minimize the incidence of reshuffling that may arise while retrieving them later.

A Multicommodity Network Flow Approach for Stage 1 Storage Problem (Block Assignment)

Given EX, IM^d, ES_i^d, IS_i, the problem of determining the values of the decision variables EX_i, IM_i for $i = 1$ to m, to minimize congestion on the road network has been formulated in Murty[1997] as a multicommodity network flow problem on the network G, that can be solved efficiently as a linear program using a commercial software package like CPLEX. To use this approach, one such linear program has to be solved in each period to obtain the storage block assignment for the next period. However, the container terminals thought that this approach is too mathematical for practical implementation in the DSS.

Fill-Ratio Equalization Approach for Stage 1 Storage Problem (Block Assignment)

This alternate approach developed in Murty[1997] for block assignment is based on the following hypothesis.

Hypothesis: Traffic congestion in the terminal will be at its minimum (i.e., traffic volume will tend to be equally distributed on all the road segments in the terminal) if the fill-ratios of all the blocks are maintained to be nearly equal.

Applying this hypothesis leads to the following very simple procedure for determining the Stage 1 decision variables EX_i, IM_i for $i = 1$ to m; one that has found acceptance from the container terminals. We describe this procedure as it is applied in a period, say period t, to compute the values of EX_i, IM_i for $i = 1$ to m, during the next period, period $t + 1$. So, here

$EX_i(t + 1), IM_i(t + 1)$	refer to the computed values of the decision variables corresponding to period $t + 1$,
EX, IM	refer to the values of these data elements for period $t + 1$,
$EX_i(t), IM_i(t)$	refer to the values of the decision variables corresponding to the current period t,
F_i	refers to the fill-ratio of block i at the beginning of the current period t,
ES_i, IS_i	refer to the values of these data elements for the current period t.

Procedure for Stage 1 (Block Assignment) for Period $t + 1$

Step 1. Compute target fill-ratio $\bar{F}$: Set this as

$$\bar{F} = \frac{EX + IM + \sum_{i=1}^{m}(F_i k_i - ES_i - IS_i + EX_i(t) + IM_i(t))}{\sum_{i=1}^{m} k_i}.$$

Step 2. Rearrange blocks in increasing order of fill-ratio at the end of the current period t: For $i = 1$ to m, the expected fill-ratio in block i at the end of the current period t is $(F_i k_i - ES_i - IS_i + EX_i(t) + IM_i(t))/k_i$. Rearrange the blocks in increasing order of this quantity (i.e., after rearrangement this quantity will be increasing with i).

Step 3. Determining the value of $EX_i(t + 1)$: Considering arriving outbound containers first, allocate the total number of arriving outbound containers, EX, to blocks in the order $i = 1$ to m, bringing the fill-ratio of each block up to $\bar{F}$. For this, define α = number of outbound containers remaining to allocate. Set $\alpha = EX$ initially.

Do for $i = 1$ to m until α becomes 0: set $IM_i(t+1) = 0, EX_i(t+1) = \max\{\bar{F}k_i - (F_i k_i - ES_i - IS_i + EX_i(t) + IM_i(t)), 0\}$. $\alpha = \alpha - EX_i(t + 1)$.

Let p be the last value of i at which this iteration stops.

Step 4. Determining values of $IM_i(t+1)$: Now begin to allocate the total number of inbound containers, IM, to blocks in the order $p + 1$ to m; again bringing the fill-ratio of each block up to $\bar{F}$. Define β = number of inbound containers remaining to allocate. Set $\beta = IM$ initially.

Do for $i = p + 1$ to m until β becomes 0: set $EX_i(t + 1) = 0, IM_i(t + 1) = \max\{\bar{F}k_i - (F_i k_i - ES_i - IS_i + EX_i(t) + IM_i(t)), 0\}$. $\beta = \beta - IM_i(t + 1)$.

This provides the broad outline of the procedure for determining $EX_i(t+1), IM_i(t+1)$. Container terminals do make heuristic modifications in this broad procedure according to their preferences, and any special circumstances. A convenient feature of this procedure is that each block either receives outbound containers, or inbound containers, for storage during a period, but not both. This seems to help increasing the efficiency of RTGC operations.

Notice that the above procedure only determines the values of the decision variables EX_i, IM_i, $i = 1$ to m; it does not assign each arriving container specifically to a storage block. Once the period begins, arriving containers are assigned for storage to blocks by the following online procedure.

Online Procedure for Assigning Containers to Storage Blocks During a Period

For outbound containers: The blocks are arranged in decreasing order of EX_i for the period (i.e., after rearrangement, EX_i decreases as i increases), and this information is left with the TG (terminal gatehouse). As outbound containers arrive during the period, TG assigns each to the first block in the order $i = 1$ to m that is not closed. While assigning containers for storage to block i, the TG continues assigning containers for storage to block i until the number assigned reaches EX_i, and then closes it.

For inbound containers: Procedure same as for outbound containers with the exception of using IM_i in place of EX_i.

Containers in a consignment (group of containers as described earlier) tend to arrive sequentially at the same time. it is convenient to assign all containers in a consignment to the same block for storage as far as possible, as this can help to keep reshuffling volume low. The terminals also make other heuristic modifications to this online block assignment procedure according to their preferences.

Allocating Storage Spaces to Containers Within Blocks, the Stage 2 Problem

Since we cannot control the order in which containers arrive for storage in a block, or hold arrived containers before storage, only online algorithms assigning spaces to containers as they arrive in the block, are suitable for implementation.

The assignment of a storage space to a container arriving in this block mainly impacts the total reshuffling in this block. So, these assignments have to be made with the objective of minimizing the total reshuffling in this block.

A block consists of a number of stacks for storing containers. Some of these stacks may already have some containers stored in them, others may be empty.

For each arriving container, we have an estimate of the time when it will be retrieved from storage. For example, for outbound containers, the retrieval time is the time when the loading of the vessel into which they are to be loaded is expected to begin. For inbound containers, the retrieval time is the expected time at which the customer truck to pick it up will arrive at the terminal.

Define the **reshuffle index** of an arriving container A for a stack in the block that has space for storing it, to be

$$\gamma = \begin{cases} 0 & \text{if the stack is empty when } A \text{ arrives for storage,} \\ & \text{number of stored containers in the stack with retrieval time before} \\ & \text{that of } A\text{, otherwise.} \end{cases}$$

The reshuffle index measures the number of times that container A may have to be reshuffled if it is put in storage in this stack at this time.

Each stack can be thought of as a bin with space to hold 5 (or 4, depending on the height of the RTGC serving this block) containers. Then the problem of storing the containers assigned to this block can be viewed as a bin packing problem, with the objective function of minimizing the sum of reshuffle indices of all the containers in their storage positions. For this problem, the well known **best fit approach** provides a very nice online approach that gives excellent results in practice. Using this leads to the following online procedure.

Online Procedure for Assigning Storage Positions to Containers Arriving for Storage in a Block

When the container arrives, find the stack in the block with an empty space, that has the smallest reshuffle index for it among all such stacks, breaking ties in favor of the stack with the maximum number of containers stored already. Store the container in the top position of that stack.

This procedure is well accepted by terminals.

Several other strategies for this decision have been proposed and analyzed in Zhang[2000].

5 Optimal Deployment of RTGC Among the Blocks

The RTGC is a critical resource whose performance in the storage yard influences the waiting times of the XT, IT, and QC; and in turn the port time of vessels. As the workload in the various storage blocks changes over time, deployment of RTGCs among storage blocks in order to provide more RTGCs to blocks with heavier workloads is an extremely important issue in terminal operations management.

The planning horizon for the RTGC deployment problem is also a 4-hour period (the beginning or ending half of a shift) as discussed in Section 3.

Because of the limited size of a block, there can be at most two RTGCs working in a block at any time.

Since RTGC movement is slow, and this movement results in not only the loss of productive time of the moving RTGC, but also obstructs traffic on the roads between the blocks; the terminals have adopted the policy that an RTGC can move from one block to another at most once in each period. Such a move can occur at any time within the period.

As a period is 4 hours long, the capacity of an RTGC in a period can be stated as 240 RTGC minutes. From observations in the storage yard we determine

$\tau =$ the average time in minutes that it takes an RTGC to either stack a container in a column, or retrieve the stored top container from a column.

$\rho =$ reshuffling frequency per productive move. This is the number of containers reshuffled on an average per container retrieved from storage; or the average number of unproductive reshuffling moves per one productive move.

Given these, and the expected number of containers to be stacked, and retrieved, from a block during a period, it is possible to measure the RTGC workload in that block in that period in terms of RTGC minutes. The RTGC deployment models use this workload data for deployment decisions.

It is very important that the RTGC workload in each block in a period be finished in that period itself, otherwise this may result in excessive XT waiting time, or delay vessel departure times. Unfinished RTGC workload will be carried over to the next period, this will have to be finished during the first part of the next period. We will call the unfinished RTGC workload in a block in a period as **delayed RTGC workload**. The aim of RTGC deployment is to minimize the total delayed RTGC workload in each period.

Besides the data elements τ, ρ mentioned above, RTGC deployment models for a period need the following data.

$m =$	Number of blocks in the yard.
$n =$	Total number of working RTGC in the yard.
$r_i =$	Number of RTGC stationed in block i at the end of the previous period, $i = 1$ to m.
$c_{ip} =$	Time in minutes that it takes an RTGC to move from block i to block p; for $i \neq p = 1$ to m. c_{ip} is set to ∞ or a large positive number if the terminal wants to forbid the movement of RTGC from block i to block p for any reason.
$ES_i =$	Number of outbound containers stored in block i to be retrieved during this period for loading into a vessel, $i = 1$ to m.
$IS_i =$	Number of inbound containers in storage in block i expected to be retrieved during this period for pickup by customers, $i = 1$ to m.
$EX_i[IM_i] =$	Total number of outbound [inbound] containers arriving in the terminal in this period, to be stored in block i, $i = 1$ to m, computed as in Section 4.

From this data, we can compute the expected RTGC workload in block i in this period, measured in RTGC minutes, to be

$$w_i = (ES_i + IS_i)(1 + \rho)\tau + (EX_i + IM_i)\tau.$$

However, the thing that makes the RTGC deployment problem hard is the following important fact.

Fact 1 about RTGC workload: If all the work of an RTGC in a period is available to be carried out at the beginning of the period, then the RTGC can quickly finish it and move. Unfortunately, the RTGC workload in a period is not all available at the beginning of the period. The need for RTGC occurs only when the trucks arrive in its block for its service. Thus the work to be carried out by RTGCs becomes available in a spread out fashion over the duration of the period as trucks arrive.

In Murty[1997] the RTGC deployment problem for a single period, for minimizing the total delayed workload, has been modeled as an MIP (Mixed Integer Programming problem) that can be solved quite efficiently by commercial software packages like CPLEX. Zhang[2000], Zhang, Wan, Liu, Linn[2000] modeled the same problem for 6 periods (a whole day) as a much larger MIP, and developed a special approach for solving it efficiently based on Lagrangian relaxation.

One shortcoming of all these models is that they ignore Fact 1 mentioned above. Also, the terminals thought that these approaches are too mathematically intense for practical implementation, and have asked for simpler approaches to be developed. We will now describe one such approach.

If there is full time work for an RTGC in the block in which it is stationed at the beginning of this period, that RTGC should continue to stay in the same block during the entire period. So, we first classify all RTGCs in their current positions at the beginning of this period into

eligible RTGC	those which have some free time after completing all their workload in this period in their current block
noneligible RTGC	those for which there is full-time workload in this period in their current block.

Unfortunately, it is not a simple matter of comparing the workload w_i in block i with the available $240r_i$ RTGC minutes from the r_i RTGC stationed in this block at the beginning of this period, to determine whether some of these RTGC can be considered eligible, because of Fact 1. For this reason Crane Supervisors make decisions on which RTGC are eligible to move, based on the following considerations.

(a) If a block has two RTGC stationed in it at the beginning of the period, and after time t in the period, if the expected arrival of trucks in the block is low enough that only one RTGC can handle them, then one of these RTGC is eligible to move from this block at that time.

(b) If there are no trucks expected to arrive at a block after some time t in this period, then all the RTGC in this block are eligible to move from this block at that time.

(c) An RTGC can move easily between a pair of adjacent blocks that share a common width line. If (B_1, B_2) is such a pair, and if the expected arrival pattern of trucks in these two blocks during a period is such that k (= 1, or 2, or 3) RTGC in these two blocks together can handles this workload by moving between these two blocks freely, then any RTGC in these two blocks put together beyond k is eligible to move at the beginning of this period itself.

Crane supervisors use many such considerations to determine which crane is eligible to move from its initial block in this period, and the time at which it can move.

Also, a block j in a period is classified as a **sink block** if $r_j = 1$ or 0, and the expected workload in it exceeds the capacity of these RTGC; and the expected number of additional RTGC it needs in this period, δ_j is

$\delta_j = 1,$	if $r_j = 1$, because there can be at most two RTGC in a block at any time,
$\delta_j = 1$, or 2	if $r_j = 0$, determined based on estimated workload in that block.

Let p be the number of eligible RTGC, number them as RTGC 1 to p. Let q be the number of sink blocks, number them as blocks 1 to q. Define the decision variable x_{ij} for $i = 1$ to p, $j = 1$ to q by

$$x_{ij} = \begin{cases} 1, & \text{if RTGC } i \text{ moves to block } j, \\ 0, & \text{otherwise.} \end{cases}$$

If c_{ij} is the minutes of time it takes RTGC i to move to block j, then $\sum_{i=1}^{p} \sum_{j=1}^{q} c_{ij} x_{ij}$ measures the RTGC productive time lost in the moves. This is like a substitute objective function for the total delayed RTGC workload. Thus the following transportation problem

$$\begin{aligned} \text{Minimize} \quad & \sum_{i=1}^{p} \sum_{j=1}^{q} c_{ij} x_{ij}, \\ \text{subject to} \quad & \sum_{j=1}^{q} x_{ij} \leq 1, \quad i = 1, \ldots, p, \\ & \sum_{i=1}^{p} x_{ij} = \delta_j, \quad j = 1, \ldots, q, \\ & x_{ij} = 0 \text{ or } 1. \end{aligned}$$

is a suitable single period model for RTGC deployment in that period. This transportation problem is easily solved to optimality. But the strategy of using the more easily computed Vogel solution for it as a near optimum is simpler to implement by the terminals.

6 Optimal Allocation of ITs to QCs

When a QC is unloading containers from a docked vessel, a certain number of ITs are allotted to it. These ITs queue up under the QC. The QC unloads a container from the vessel, puts it on the IT at the head of the queue. That IT then takes the container to the storage block where it is to be stored, and queues up in the truck passing lane of the block for the YC (i.e., RTGC or RMGC) to take the container from its top and store it. Then the IT returns to the queue under the QC, and the same process repeats until the unloading activity is completed.

While the QC is loading containers into the vessel, the same process goes on in reverse with the ITs bringing containers from the storage block to the QC. Let

n_a = number of ITs allocated to the QC.

If n_a is too small, the QC idle time waiting for the IT may be high. If n_a is too large, then the idle time of the ITs waiting in the queues at the QC and in the storage block may be high. In order to optimize the twin objectives of minimizing the QC waiting time, and the total number of IT hours used to unload and load a vessel, it is necessary to find the optimum value for n_a, which may depend on the position of the QC, and the location of the storage block to which the ITs have to travel. The input data needed to make this decision are the probability distributions of the following random variables.

x_a, y_a = Travel times of an IT from the QC position to the storage block, and back.
z_a = QC processing time for a single container.
w_a = YC serving time for an IT in the storage block.

The QC on the dock, the YC in the storage block, and the ITs traveling back and forth between them, form a queueing network. The mathematical analysis of this queueing network has been carried out in Murty, Tsang, and Wan[1996] under the assumption that all these distributions are negative exponential. However, observations made on these random variables indicate that negative exponential is not a good fit for their distributions, they seem to follow normal distributions. Analysis of the queueing network under the normality assumption by mathematical techniques is very hard. So, we are estimating the optimal values of the decision variables n_a for every (QC position on the dock, Storage block) pair by running simulation experiments. As an example, we provide the results from two such experiments. All times are in seconds, and the entries in the tables shown are:

t_Q = estimated QC idle time in seconds,
t_I = estimated idle time of all ITs in seconds,
cost = cost of QC, IT idle times under the assumption that one second of QC [IT] idle time has cost 1000 units [5 units],
processing duration = time in seconds to process all the containers.

Experiment 1: No. containers = 30

n_a	t_Q	t_I	Cost	Processing duration
1	8344.14	0	8,344,140	11942
2	2519.72	183.56	2,520,636	6137
3	695.82	682.06	699,222	4293
4	223.97	2887.67	238,403	3821
5	220.88	6202.49	251,893	3810
6	221.67	9382.46	268,581	3820
7	224.54	12636.88	287,726	3820
8	216.86	15715.20	295,434	3822

Experiment 2: No. containers = 40

n_a	t_Q	t_I	Cost	Processing duration
1	11177.62	0	11,177,621	15940
2	3357.39	228.97	3,358,539	8153
3	21.07	19.90	21,167	5663
4	5.64	90.29	6089	5020
5	5.54	207.08	6575	5014
6	5.48	314.94	7058	5017
7	5.47	425.65	7594	5023
8	5.50	527.29	8140	5024

In both experiments it can be seen that $n_a = 4$ ITs allocated to the QC yielded the best cost estimate, and very nearly the best processing duration for all the containers.

The optimum number of ITs to allocate to QCs are being determined in this way using simulation.

For each (QC position on the dock, storage block) pair, and the number of containers to be processed involving this pair, the simulation experiment with the optimum value of n_a, not only gives the estimate expected time in seconds that this job will take, but also the variance of this time. These variance estimates are used for making decisions on optimum IT hiring plans over the day discussed in the next section.

7 IT Hiring Over the Day

Terminals work round the clock every day, and the demand for ITs varies over the day, depending on the schedule of the vessels arriving or leaving that day. Here we discuss how to prepare a profile of how many ITs are needed to help the QCs while the are working on vessels, in half-hour intervals. Using this information, we develop a model to determine the optimal number of ITs to hire at various times of the day to meet the requirements for them, while minimizing the total number of ITs hired during the day.

We divide the day into 48 half-hour intervals: 8:00 − 8:30, 8:30 − 9:00, etc. These half-hour intervals are numbered 1 to 48 in this order.

The most common plan for ITs in Hong Kong is to work for four hours, take an hour's meal break, and then work for three more hours. The model for determining the IT hiring plan is based on the assumption that this is the work plan for all the ITs hired. It is easy to modify the model to accommodate other work plans.

The input data for these decisions are: vessel arrival and departure times during the day, vessel berthing information, the unloading and loading sequence for each of these vessels, storage block assignments for unloaded containers, and the storage blocks from which containers to be loaded come from.

Once a vessel is docked on the berth, its processing consists of a series of jobs. Each **job** is associated with the following information:

> hatch of the vessel (this determines the position of the QC that will work on this job), storage block to which containers in this job will go to or come from, and the number of containers in this job.

Thus we can view each job as a triple (QC position, storage block, no. of containers). Jobs associated with the same QC position are carried out sequentially in the unloading loading sequence; while those associated with different QC positions may be carried out concurrently depending on the number of QCs that will work on this vessel.

From the decisions discussed in Section 6, we know how many ITs are required for each job, the expected duration μ of the job in seconds, and its variance σ^2. For planning purposes, a safe practice is to allow $\mu + \lambda\sigma$ seconds for completing the job, where λ is given some value between 0.5 to 1 depending on the terminal's preference. From the vessel's schedule, and unloading, loading sequence, we know which of these jobs will be carried out sequentially, and which will be carried out concurrently. From all this information, we obtain a profile of how many ITs are needed for the processing of this vessel over time, during the processing interval of this vessel.

Adding up this information from all the vessels expected to be docked during the day, we get a profile of the number of ITs needed over time during the whole day. We then take the IT requirement in any half-hour interval to be the maximum requirement at any point of time in this interval. Let

$d_t =$ requirement for ITs (i.e., number of ITs needed to help QCs working on docked vessels) during the tth half-hour interval of the day, $t = 1$ to 48.

We assume that ITs can begin work at the terminal at the beginning of any of the 48 half-hour intervals. Define the decision variable

$x_t =$ number of ITs beginning work at the beginning of the tth half hour interval of the day, $t = 1$ to 48.

Since the workplan is for ITs to work for four hours, take a one hour meal break, then work for three more hours, the ITs which will be working in the tth half-hour interval are those that started working at the beginning of the half-hour intervals in the set

$S_t = \{t - 15 \text{ to } t - 10,\ t - 7 \text{ to } t\}.$

Depending on the value of t, some of the intervals in the set S_t might belong to the previous day. This set S_t depends on the work plan for ITs used at the terminal. If different workplans are used, define S_t to be

$S_t =$ set of half-hour intervals satisfying the property that all ITs which started working at the beginning of one of these intervals will be working during the tth half-hour interval.

Then the model for determining the decision variables x_t to meet IT requirements using the minimum number of ITs, is the following integer program

$$\begin{array}{rrcll} \text{Minimize} & \displaystyle\sum_{t=1}^{48} x_t, & & & \\ \text{subject to} & \displaystyle\sum_{p \in S_t} x_p & \geq & d_t, & t = 1, \ldots, 48, \\ & x_t & \geq & 0, & \text{and integer for all } t. \end{array}$$

This integer program can be solved quite efficiently using a commercial software package like CPLEX. Also, since the terminal uses ITs for many other activities like reshuffling/restacking of stored containers etc., we can also take the solution obtained by rounding up the optimum solution of the LP relaxation of this model, to implement, and assign any spare time of ITs to these other activities.

8 Conclusion

The daily operations of a container shipping terminal are complex and involve a variety of decisions to be made. We described these in detail, and discussed the strategies being used to design a computerized DSS for making some of these decisions optimally, as part of an ongoing effort to keep the terminals in Hong Kong as the most efficient in the shipping industry.

9 References

[1] Katta G. Murty, Maria C. L. Tsang, and Yat-Wah W. Wan (1996), "A Newsboy Type Model for Scheduling Stevedoring Tractors At Container Terminals in Hong Kong", Technical Report, Dept. of IEEM, Hong Kong University of Science and Technology.
[2] Katta G. Murty (1997), Storage Yard Planning in Container Shipping Terminals", Technical Report, Dept. of IEEM, Hong Kong University of Science and Technology.
[3] Katta G. Murty, Philipp A. Djang (1999), "The US Army National Guard's Mobile Simulators Location and Routing Problem", *Operations Research*, 47(1999)175-182.
[4] Chuqian Zhang (2000), "Resource Planning in Container Storage Yard", Ph.D. thesis, Dept. of IEEM, Hong Kong University of Science and Technology.
[5] Chuqian Zhang, Yat-Wah W. Wan, Jiyin Liu, and Richard Linn (2000), "Dynamic Crane Deployment in Container Storage Yards", Technical Report, Dept. of IEEM, Hong Kong University of Science and Technology.