Interfaces
Vol. 35, No. 1, January–February 2005, pp. 61–75
ISSN 0092-2102 | EISSN 1526-551X | 05 | 3501 | 0061

DOI 10.1287/inte.1040.0120


# Hongkong International Terminals Gains Elastic Capacity Using a Data-Intensive Decision-Support System

Katta G. Murty
Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan 48109-2117, murty@engin.umich.edu

Yat-wah Wan
Institute of Global Operations Strategy and Logistics Management, National Dong Hwa University, Hualien, Taiwan, ywan@mail.ndhu.edu.tw

Jiyin Liu
Business School, Loughborough University, Loughborough, Leicestershire LE11 3TU, United Kingdom, j.y.liu@lboro.ac.uk

Mitchell M. Tseng
Department of Industrial Engineering and Engineering Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China, tseng@ust.hk

Edmond Leung, Kam-Keung Lai
Hongkong International Terminals, Terminal 4, Container Port Road South, Kwai Chung, Hong Kong, China {leung.edmond@hit.com.hk, lai.kamkeung@hit.com.hk}

Herman W. C. Chiu
Hong Kong Productivity Council, HKPC Building, 78 Tat Chee Avenue, Kowloon, Hong Kong, China, hermanc@hkpc.org

As the flagship of Hutchison Port Holdings (HPH), Hongkong International Terminals (HIT) is the busiest container terminal on the planet. HIT receives over 10,000 trucks and 15 vessels a day, about six million twenty-foot equivalent units (TEUs) a year. HIT makes hundreds of operational decisions a minute. HIT's terminal management system, the productivity plus program (3P), optimizes resources throughout the container yard using operations research/management science (OR/MS) techniques and algorithms. It manages such interrelated decisions as how to route container trucks in the yard, where to store arriving containers, how many quay cranes to use for each vessel, how many trucks to assign to each crane, how many yard cranes to assign to each container storage block, and when to schedule incoming trucks for container pickup. As the number of container terminals in Asia grows, competition has become price driven and service driven. HIT realized its future rests not only with moving boxes but with mastering the associated information. This meant developing a decision-support system (DSS) to provide superior and differentiated services by generating optimal decisions, one that is very robust under uncertain arrival times of trucks and vessels. In its 10 years of operation, the implementation of the DSS through 3P has helped HIT to become the world's most efficient and flexible terminal operator. HIT alone saves US$100 million per year. By optimizing internal truck use at its sister terminals, the HPH group saves an additional US$54 million per year.

*Key words*: transportation: freight, materials handling; facilities: equipment planning, capacity expansion.

Hongkong International Terminals (HIT) was established with a single terminal in Hong Kong in 1969. Since its formation, it has become the flagship operation of the world's largest container-terminal developer and operator, Hutchison Port Holdings (HPH), which operates 187 berths in 17 countries. In containerized shipping, goods are packed in steel boxes (20-foot or 40-foot in length). Terminals calculate throughput based on twenty-foot equivalent units (TEUs); a 40-foot container is counted as two TEUs. In 2002, HPH's estimated market share of containerized trade stood at 13.3 percent or 35.8 million

TEUs. Lined up end to end, these containers would span the circumference of the globe more than five times.

In 2003, Hong Kong terminal operators collectively handled over 20 million TEUs, making Hong Kong the world's busiest port. As the largest privately owned terminal operator, HIT, along with its joint venture, China Ocean Shipping Group Company Pacific (COSCO Pacific), handles approximately one third of the territory's throughput. According to government statistics, port, logistics, and associated support services account for approximately 20 percent of Hong Kong's economy, and maintaining the port's premiere status is essential to the territory's future.

## The Business Environment in Hong Kong

HIT reached a turning point in the early 1990s. At the time, China's manufacturing boom was creating a remarkable demand for terminal services in southern China. The surge of exports from the region gave HIT a golden opportunity. Whichever terminal could offer the capacity would capitalize on this expanding market. However, with land scarce in Hong Kong, HIT could not simply expand its terminals to meet the growing demand. It had to scrutinize and restructure its operations to maximize its capacity to maintain its lead in the industry.

HIT also faced growing competition. Several new ports were opening along China's southern coast. With favorable labor costs and government concessions, these ports were poised to offer HIT's customers much cheaper prices than HIT could. HIT realized it had to adopt a customer-focused philosophy and provide premium services through higher productivity and efficiency.

According to Eric Ip, managing director of HIT, HIT was facing a dire situation of losing market share in a growing market. There are only 40 major shipping lines in the world, and they all take their decision of where to call very seriously. Once customers decide to leave, there is very little HIT can do to win them back.

To strengthen its role as southern China's cargo hub and secure a major share of the growing volume, HIT had to reestablish itself as the industry's benchmark for efficiency and productivity.

Entering crisis mode in early 1995, HIT engaged outside consultants to improve its decision making in daily operations. At the end of that year, HIT's managers judged the results to be mediocre. HIT then decided to take the bull by the horns and rely on its own management and technical resources to rearrange the terminal layout, to introduce higher yard cranes that could stack more containers vertically, and to reengineer its processes, using intensive computerization with advanced technological applications.

Understanding that a new decision-support system (DSS) for daily operations would form the core of this program to enhance productivity and capacity, HIT met with the industrial engineering and engineering management (IEEM) department at the Hong Kong University of Science and Technology (HKUST); Katta Murty, a visiting professor at the time, brought together a team to develop the foundation of the DSS.

## The Unusual Operating Problems in Hong Kong

Hong Kong is the principal entry port to southern China. Two striking features of container terminals in Hong Kong are (1) the massive volume of cargo throughput, and (2) the very limited industrial land available for development. The volume of containers transported through Hong Kong has increased by a yearly average of 11 percent since 1988, with the throughput in 2003 totaling over 20 million TEUs. Hong Kong is consistently ranked the world's busiest port.

The section of the terminal used for the temporary storage of containers between land and sea transportation is called the storage yard (or container yard). Depending on a terminal's mode of operations, containers can be stored either on the trailer on which they arrived or stacked one on top of the other. The storage yard is divided into rectangular regions called *storage blocks* (or *blocks*).

Inbound containers arriving at the terminal via container ships are called *import containers*. These import containers are unloaded from the vessels and kept in the storage yard until their owners pick them up.

Outbound containers brought into the terminal by external trucks to be loaded into vessels for export are called *export containers*. They are stored in the container yard until their designated departure vessel arrives at the terminal. *External trucks* (tractors) are independent trucks of other companies that bring export containers into the terminal and collect import containers from the terminal. The main access through which external trucks enter and leave the terminal is called the *terminal gate*.

One of the most direct indicators of container traffic in a terminal is the number of containers handled per acre (or per hectare) of storage yard space in a given period of time. Because space is so limited in Hong Kong, this indicator is about seven times that of Long Beach's and nine times that of Hamburg's. Needless to say, the space constraint poses unique challenges to Hong Kong terminal operators in delivering high-quality service.

Stacking containers one or two high and separating the storage of import, export, and empty containers are normal practices at container terminals in other parts of the world. Hong Kong terminal operators need to squeeze as much capacity as possible out of the limited land they have. This means mixing import and export containers and stacking containers five or six containers high.

Creating additional yard capacity by stacking containers so high causes unproductive or nonrevenue-earning reshuffling. Retrieving a bottom container from a stack of six consumes time and resources. To minimize reshuffling, HIT needed specialized algorithms to determine the optimal location for storing each container.

## Operations Flow

The huge cranes that unload import containers from and load export containers onto vessels are called *quay* (or *shore*) *cranes*. The operations of the quay crane are supported by the *internal trucks* (*tractors*) which are trucks used by the terminal to shuttle containers between the quay and the storage yard. Quay cranes unload import containers from vessels and load them directly onto the internal trucks lined up under them, one container per each internal truck, which then take them to the storage yard for temporary storage. In the same way, internal trucks bring export containers from the storage yard and line up under the quay crane, which picks them up from the internal trucks

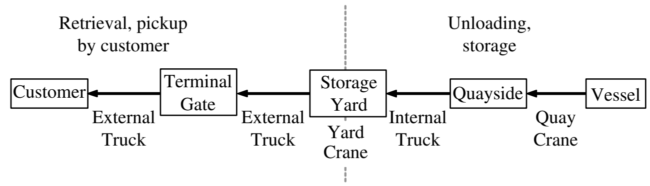


**Figure 1: Import containers take a number of short trips in traveling from container vessels to their eventual recipients, with specialized equipment handling each trip. HIT has a total storage capacity of 76,000 TEUs–227 acres of container storage yard (the equivalent of 111 football stadiums), divided into 95 storage blocks, each with a capacity of over 800 TEUs.**

and loads them directly into the hatch of the vessel. A *hatch* is the cover of a section of a vessel (or opening in the deck) through which cargo is loaded, or unloaded. The handling rate of a quay crane is a key performance indicator of a terminal.

Two continuous streams of containers move through the terminal over time: import or inbound containers arriving on container vessels (Figure 1), and export or outbound containers delivered by external trucks (Figure 2). Containers are steel boxes of dimensions (all measurements are in feet) $20 \times 8 \times$ (8.5 or 9.5) (*20-foot containers*), or $40 \times 8 \times$ (8.5 or 9.5) (40-foot containers). Quay cranes unload the import containers onto internal trucks, which take them to storage blocks. Each storage block is served by one or more large cranes (*yard cranes*) stationed there. The yard cranes unload the import containers from the internal trucks and put them into storage. They remain in storage until the yard cranes retrieve them and transfer them to external trucks for delivery to their ultimate recipients. Export containers follow a reverse sequence: they arrive on external trucks that

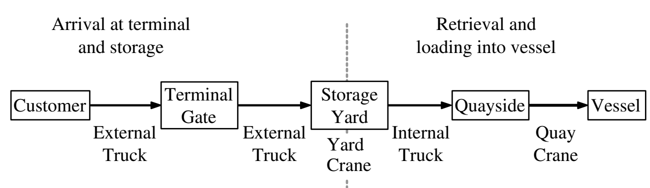


**Figure 2: Arriving export containers have their documentation checked and inspected at the terminal gate. The gate operator assigns each container to a storage block and instructs the driver to proceed to that block. The container takes a number of short trips until its eventual loading into its container vessel, with specialized equipment handling each trip.**

go through the terminal gate to storage blocks, are transferred to storage by yard cranes, and are held in storage until their designated vessels arrive. They are then retrieved from storage, taken by internal trucks to the berths, and loaded by quay cranes onto their vessels.

*Terminal road network* refers either to the road system within the terminal or to a directed network representation of it with blocks, berths, terminal gate, and road intersections as nodes, and road segments joining pairs of these as arcs.

There are many types of yard cranes in use in the industry. Most of the yard cranes in use at HIT are rubber-tired gantry cranes which can move from one block to another and thus offer great flexibility. HIT's equipment includes 10 berths totaling 9,797 feet of quay length (Terminals 4, 6, and 7 of the Hong Kong Port), 33 quay cranes, 122 yard cranes, and over 200 internal trucks. HIT serves over 40 shipping lines and handles over 125 vessel calls per week. HIT's workforce remained fairly stable throughout our project.

## The Main Performance Measures Shipping Lines Use to Rate Container Terminals

The most important performance measure to a shipping line in rating a terminal is the vessel turnaround time (the average time the terminal takes to unload and load a docked vessel). From a customer's perspective, this measure is a cost measure that should be minimized.

Another closely related measure that shipping lines use in choosing a terminal is the average quay crane rate, the quay cranes' throughput during a period, (total number of containers unloaded or loaded)/(total number of hours the quay crane operated). For a shipping line, this is a profit measure that should be maximized. Shipping lines judge container terminals largely based on vessel turnaround time and quay crane rate. All the decision problems we studied during this project contributed to improving these measures.

In addition, terminals want to optimize several performance measures, most of which concern equipment or labor utilization. We dealt with traffic congestion on the terminal road network, the number of internal trucks and drivers, the time yard cranes spend traveling between blocks, and the turnaround time of external trucks that come to the terminal.

## HIT's Planning and Operation Problems

Traffic flow in a container terminal is akin to the circulation of blood in a human being: life depends on it. Congestion on the terminal road network damages productivity and service quality.

In the 1990s, congestion was slowing the flow of container trucks at HIT. Internal trucks were getting stuck in traffic, and quay cranes were waiting for trucks. The productivity of waiting equipment comes to a halt, which reduces its handling rate and increases vessel turnaround time, the factor that is most critical to optimize. Clearly, congestion was the Gordian knot we needed to untie before any other improvements could benefit HIT.

Increasing the number of internal trucks serving each quay crane would be counterproductive, because increasing the trucks on the road would increase the congestion. HIT needed to reduce the number of internal trucks but maintain or even increase the throughput of the quay cranes.

Many researchers have developed OR/MS applications for container terminals. Murty et al. (2005) discussed the technical aspects of our work and our list of references. Bish et al. (2001) and Zhang et al. (2002, 2003) also discussed OR/MS research on container terminals. However, we did not find any publications on research about alleviating traffic congestion on terminal road networks, even though it is a common problem.

The terminal road network is virtually a closed system, with the terminal gate as the only access point for external trucks. The only vehicles operating on the terminal road network are container trucks and the occasional yard crane moving from one block to another. Increasing the number of vehicles operating on this finite system naturally increases traffic congestion. We realized that reducing traffic congestion required a two-pronged approach: reducing the number of vehicles operating on the network and routing them optimally.

**Reducing the Number of Vehicles Operating on the Terminal Road Network**

Reducing the number of internal trucks operating on the terminal road network affects the performance of the quay cranes they serve. We wanted to reduce the number of internal trucks, while furnishing equivalent (or better) service to the quay cranes. We also wanted to reduce the number of external trucks operating in the yard during the peak hours of 8:00 AM to 5:00 PM, and the time spent by yard cranes moving between blocks.

**Routing the Vehicles Optimally**

We also wanted to route internal and external trucks optimally. Making optimal choices for the various decisions in daily operation is a large and complex problem because it involves a dynamic terminal system that may have 10,000 trucks passing through the terminal gate and 15 vessels arriving each day, 80,000 TEUs of constantly accessed storage space, stochastic arrivals and departures of containers, and many complex objectives to optimize simultaneously.

Some features of the terminal's container storage and retrieval workload are important:

—The distribution of work over time is variable, because of uncertainties in weather conditions, road and sea traffic conditions, and other factors.

—It is impossible to control the order in which containers arrive or their times of arrival.

—Work must be carried out as it arrives without delay.

## Optimization Objectives and Project Organization—The Initial Charter for the Project Team

In 1995, one of the authors, Katta Murty, was a visiting professor in IEEM at HKUST. HIT invited Murty and several IEEM faculty members to work on developing an information technology and OR/MS-based DSS for the company's daily operations. In addition, graduate students in the IEEM department investigated specific issues related to this project. The initial charter for the project team was to focus on five critical resource-allocation problems (Zhang 2000, Meersmans and Dekker 2001, Vis and De Koster 2003):

D1: To route container trucks and allocate storage spaces to arriving containers to minimize traffic congestion on the terminal road network and to reduce reshuffling.

D2: To optimize the allocation of internal trucks to quay cranes to minimize the waiting times for the cranes and the trucks, and to minimize the number of internal trucks operating in the yard to reduce traffic congestion.

D3: To develop a procedure for estimating requirements for internal trucks and to develop an optimum hiring policy for drivers.

D4: To deploy yard cranes to minimize their travel time between blocks.

D5: To allocate appointment times to external trucks to minimize their turnaround time, to level off the workload at the terminal, and to reduce the number of external trucks in the yard during peak times to reduce traffic congestion.

For the DSS to succeed, the information infrastructure in the terminal must produce all the data the OR algorithms require. HIT did not have this infrastructure; it built the necessary information technology and communication infrastructure between 1995 to 1999.

## Decision Problem D1: Routing Container Trucks and Allocating Storage Spaces to Arriving Containers

Our main focus in problem D1 was to route container trucks optimally to minimize congestion. The policy for allocating storage space to arriving containers directly influences the routing of container trucks and is therefore an important factor in controlling congestion.

HIT employees work three eight-hour shifts a day. HIT organized work into six four-hour sections (half shifts) as the planning periods for the problem: 00:00 to 4:00, 4:00 to 8:00, 8:00 to 12:00, 12:00 to 16:00, 16:00 to 20:00, and 20:00 to 24:00.

In considering storage-space allocation for arriving containers, we realized that the traditional integer-programming models used (Bish et al. 2001, Zhang et al. 2003) would be impractical and inappropriate because the data on occupied storage spaces changes constantly before the model can be entered into the computer.

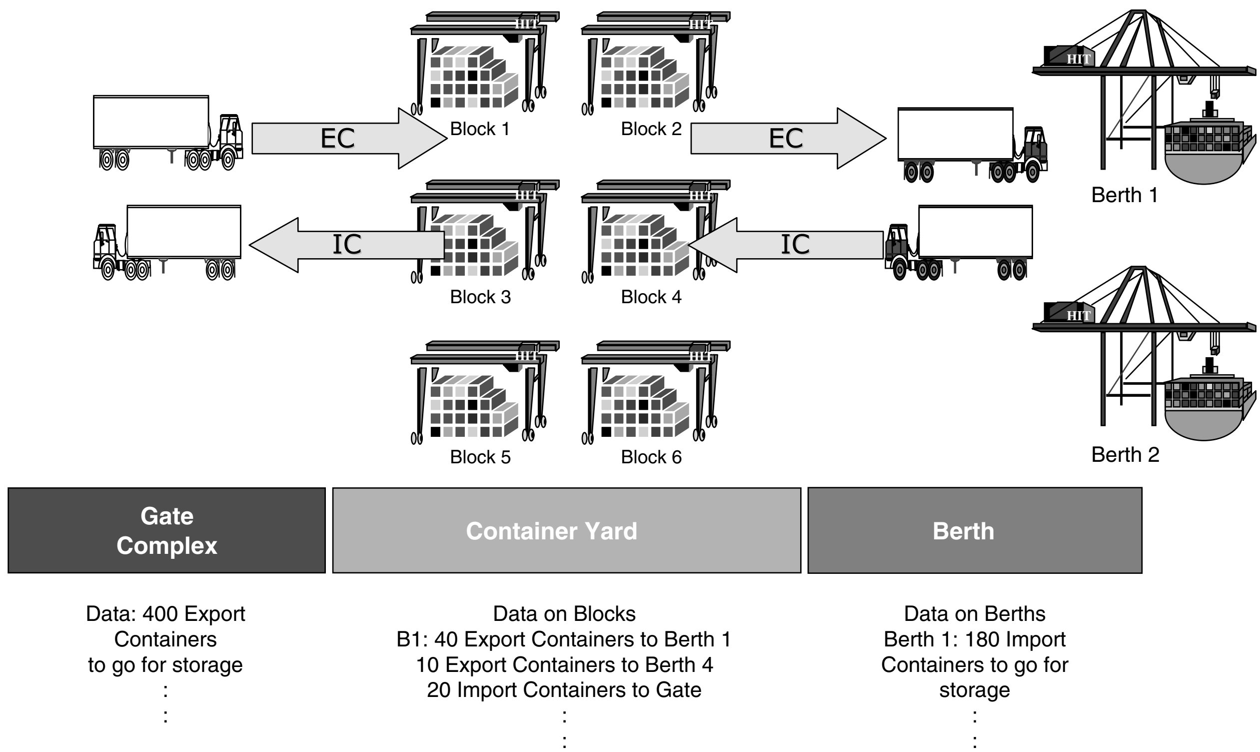


**Figure 3: The model based on the batch-processing strategy relies on data on the number of container trucks expected to flow between various origin, destination pairs in a planning period. EC and IC denote export and import containers, respectively.**

Another commonly used approach based on a batch-processing strategy for handling all the containers expected to arrive and leave at each node in a period (Figure 3) leads to a multicommodity flow problem, a large-scale linear program with many variables and thousands of constraints. We realized that the solution from this type of model would work well only if the workload were distributed uniformly over time. However, the stochastic features of the problem make the distributions highly nonuniform. Also, implementing the output from this model would be difficult because truck drivers resent being told which routes to take.

We therefore had to develop dynamic, real-time approaches to storing containers and routing trucks on the terminal road network to distribute the traffic uniformly along the various road segments continuously over time to alleviate congestion.

We defined the fill ratio in block $i$ at time $t$, $f_i(t) =$ (number of containers in storage in block $i$ at time $t$)/(total storage spaces in block $i$). We observed that this fill ratio strongly influences the workload rate at that block, that is, the containers being moved in and out of the block per minute. So, as Murty (1997) proposed and Murty et al. (2005) discussed more fully, maintaining the fill ratios of all the blocks nearly equal will ensure that the workload rates at all the blocks (and therefore the volumes of truck traffic in their neighborhoods) will be nearly equal, thus ensuring equal distribution of traffic on the terminal roads and minimizing congestion. This leads to a substitute-objective-function technique for controlling congestion, similar to that developed by Murty and Djang (1999) for other routing problems. For the planning period, we define the following:

$x_i =$ container quota for block $i =$ number of containers arriving in this period to be dispatched for storage to block $i$, a decision variable.

$a_i =$ number of stored containers that will remain in block $i$ at the end of this period if no additional containers are stored there during this period, a data element.

$N =$ number of new containers expected to arrive at the terminal in this period for storage, a data element.

$B, A =$ total number of blocks in the storage yard, the number of storage positions in each block.

$w_i(t) =$ number of container trucks waiting in block $i$ to be handled by the yard crane there at time $t$.

$x_i^R(t) =$ container quota remaining for block $i$ at time $t$ in the period
$= x_i -$ (number of new containers sent to block $i$ for storage up to time $t$ in this period).

Our approach breaks down into three separate policies: the fill-ratio-equalization policy, the arriving-container-dispatching policy, and the storage-position-assignment policy.

## The Fill-Ratio-Equalization Policy

The fill-ratio-equalization policy determines the decision variables $x_i$ for this period to make sure that the fill ratios in all the blocks are as nearly equal as possible at one time during the period, namely, the end of the period. The fill ratio in the whole yard at the end of this period will be $F = (N + \sum_i a_i)/(A \times B)$. If the fill ratios in all the blocks at the end of this period are all equal, they will all be equal to $F$. Thus, this policy determines $x_i$ to guarantee that the fill ratio in each block will be as close to $F$ as possible, by the least sum of absolute deviations measure, by the end of this period. This leads to a small linear-programming (LP) model with only 100 constraints that has very special structure. We can obtain the optimum solution of this LP by using a simple combinatorial scheme (Table 1):

(1) Rearrange blocks so that $a_i$ increases with $i$.
(2) Determine $x_i$ in increasing order of $i$ using

$$x_1 = \text{minimum}\{N, A \times F - a_1\},$$

and for $i \geq 2$,

$$x_i = \text{minimum}\left\{\text{maximum}\{0, A \times F - a_i\}, N - \sum_{r=1}^{i-1} x_r\right\}.$$

| $i$ | $a_i$ | $x_i$ | $a_i + x_i$ |
|---|---|---|---|
| 1 | 100 | 300 | 400 |
| 2 | 120 | 280 | 400 |
| 3 | 150 | 250 | 400 |
| 4 | 300 | 100 | 400 |
| 5 | 325 | 75 | 400 |
| 6 | 350 | 35 | 385 |
| 7 | 375 | 0 | 375 |
| 8 | 400 | 0 | 400 |
| 9 | 450 | 0 | 450 |
| Total | 2,570 | 1,040 | |

**Table 1: This table shows the computation of $x_i$ in a nine-block example for a period with 1,040 new containers arriving for storage and $a_i$ previously stored containers remaining in storage at the end of the period in block $i$.**

### Numerical Example

Suppose that there are $B = 9$ blocks in the terminal, each with $A = 600$ storage spaces. Suppose that we expect $N = 1{,}040$ containers to arrive for storage during the planning period. $a_i$ is already increasing with $i$ (Table 1). The fill ratio over the whole yard at the end of this period will be $F = (\sum a_i + 1{,}040)/(9 \times 600) = (2{,}570 + 1{,}040)/(5{,}400) \approx 0.67$, and $A \times F \approx 400$.

Our combinatorial scheme determines $x_i$ in the order $i = 1, 2, \ldots$ to bring $a_i + x_i$ to 400 until all 1,040 new containers are allotted to blocks for storage. Under this policy, $a_i + x_i$ will be the number of stored containers in block $i$ at the end of this period. From Table 1, we can verify that the values of all $a_i + x_i$ are nearly equal in this numerical example.

This policy determines only the quota number of containers for each block, not which containers will be stored in each block. That is determined by the dispatching policy.

## The Arriving-Container-Dispatching Policy

Regardless of how we determine the container quota numbers $x_i$, if we send a consecutive sequence of arriving container trucks to the same block in a short time interval, we will create congestion at that block. To avoid this possibility, we must ensure that the yard crane in that block has enough time to unload a truck we send there before we send another.

Hence, this policy dispatches each truck arriving (at the terminal gate and at each berth) at time point $t$ in the period to a block $i$ satisfying $w_i(t) = \min\{w_j(t)\colon j$ satisfying $x_j^R(t) > 0\}$, that is, a block with a remaining positive quota that has the smallest number of trucks waiting in it.

To implement this policy, we must continuously monitor how many trucks are waiting to be handled by the yard crane at each block. The module that does this is one of the critical modules we developed in the 3P project.

### The Storage-Position-Assignment Policy

The storage-position-assignment policy assigns storage positions to containers arriving at a block to minimize reshuffling. With thousands of storage-assignment decisions to make each day, we use an index called the reshuffle index. For a container $C$ arriving in a block, we define the reshuffle index of a stack that has storage space for it to be the number of containers stored in the stack already, with estimated retrieval times before that of $C$. This index measures the number of times container $C$ will have to be reshuffled if it is stored in that stack at this time.

We use the best-fit algorithm; it assigns $C$ for storage in the top position of a stack with an empty space that has the smallest reshuffle index. This is analogous to the best-fit algorithm for the bin-packing problem, which has been shown to produce solutions very close to the optimum.

By using this strategy, HIT reduced the number of reshuffling moves per productive move of the yard crane and increased yard-crane productivity. However, it did this when it had annual throughput of less than four million TEUs and was using yard cranes that could store at most four containers per stack. With increasing annual throughput, HIT introduced many new yard cranes that could stack five containers. Storing more containers per stack exponentially increases the reshuffling frequency.

### Innovations in the Work on Decision Problem D1

We introduced many innovations in working on decision problem D1.

—We were the first and only group to study traffic congestion in container terminals using OR/MS techniques and produce an effective solution. To implement our solution, HIT made a huge investment in information technology.

—We tried two LP models to solve this problem. The first, based on a standard approach, led to a model with thousands of constraints, but its output was poor, because the model did not take into account the wide fluctuations in workload over time. The second LP model, the one we implemented, is a highly innovative model based on a substitute-objective-function technique. This model had only about 100 constraints, and we can obtain its optimum solution by using a simple combinatorial scheme that takes only a few seconds to compute. It is highly effective.

—We are the first group to relate the container-stacking problem in a block to the classical bin-packing problem in combinatorial optimization and develop a best-fit-type algorithm for it.

Our work shows that OR/MS techniques have to be applied intelligently to produce good results, and that OR/MS plus information technology plus intelligent application makes a powerful combination for solving practical problems.

### Frequency of Decision Making

HIT solves the LP model to determine the container quota numbers for the 95 blocks once in each four-hour period, in only a few seconds.

## Decision Problem D2: Optimizing the Allocation of Internal Trucks to Quay Cranes

The focus of problem D2 is to optimize the allocation of internal trucks to quay cranes, to minimize the number of internal trucks operating in the yard (part of our strategy to control congestion), and to maximize the utilization of internal trucks. When we started our work in 1995, HIT was assigning eight internal trucks to each working quay crane.

The quay crane on the dock, the yard cranes in the block and the internal trucks traveling back and forth form a queuing network. We ran many simulations of this queuing network, which showed conclusively that four internal trucks per quay crane attain the minimum penalty (for the idle time of the quay crane and the waiting time of internal trucks in queues) and very nearly the shortest processing duration for a given number of containers to be unloaded from or loaded into the ship hatch (Table 2).

$h$ = number of containers processed = 30

| | | Process duration (minutes) | |
|---|---|---|---|
| $n$ | Total penalty | Average | Standard deviation |
| 1 | 3,977 | 131 | 5.6 |
| 2 | 1,324 | 77 | 3.3 |
| 3 | 734 | 65 | 2.2 |
| 4 | 664 | 62 | 1.9 |
| 5 | 700 | 62 | 1.9 |
| 6 | 752 | 61 | 1.8 |
| 7 | 811 | 61 | 1.8 |
| 8 | 880 | 61 | 1.9 |

$n$ = number of internal truck allotted/quay crane

**Table 2: In our results from a simulation run, the penalty is a weighted loss function for the idle time of quay cranes and the waiting time of internal trucks in queues. Process duration is the time for the quay crane to unload or load $h$ containers.**

Following the simulation results, the company has cut the number of internal trucks assigned to each quay crane to four. HIT has also done away with the traditional practice of having a separate group of internal trucks serve each quay crane. Instead, the terminal maintains a single pool of internal trucks to serve all the quay cranes, with the control tower dispatching each internal truck returning from the storage yard to the quay crane with the smallest number of internal trucks waiting under it. The control tower ensures that the number of internal trucks in the pool is about four times the number of working quay cranes.

Realizing the advantages of a common pool of internal trucks, many of HIT's competitors have also adopted this practice.

### Innovations in the Work on Decision Problem D2

We introduced many innovations in working on decision problem D2.

—We were the first group to recognize the importance of reducing the number of trucks in the system to reduce congestion. By reducing the number of internal trucks per quay crane from eight to four, the company saves costs for trucks and drivers and reduces traffic congestion in the terminal.

—The internal-truck pool system to serve all the quay cranes is an innovation that has been adopted worldwide.

## Decision Problem D3: Developing a Procedure for Estimating Hourly Requirements for Internal Trucks and an Optimal Hiring Policy for Drivers

Let $h$ denote the workload in a hatch (the total number of containers to be unloaded from and loaded into a hatch) of a docked vessel. From our simulations and using linear regression, we estimated that the expected time (in minutes) and the standard deviation of the time for processing a hatch with workload $h$ are $\mu(h) = 8.28 + 1.79h$ and $\sigma(h) = 1.31 + 0.19h$, respectively.

For planning purposes, the company decided to allow $\mu(h) + \sigma(h)$ minutes for processing a hatch when the workload in it is $h$.

A few days before its arrival, each vessel informs HIT of its expected arrival time and the workload in each hatch. The company then determines when to begin work on the vessel, how many quay cranes to allocate to it, and the order in which the quay cranes will work on the hatches. From this information, the procedure decides a planned processing duration for each hatch and when work on each will begin and end. Allowing four internal trucks per working quay crane, we can then determine the profile of internal truck requirements for processing the vessel over time (Table 3).

For this vessel, quay cranes 1, 2, and 3 will take respectively 350, 362, and 336 minutes to finish their jobs on the vessel. Hence, from 6:00 AM to 10:36 AM, all three quay cranes will be working on this vessel, creating a requirement for 12 internal trucks. Similarly, from 10:36 AM to 10:50 AM, the two cranes working will require eight internal trucks, and from 10:50 AM to 11:02 AM, the one crane still working will require only four internal trucks for this vessel.

Adding up the requirements for all of the vessels for any given day provides the profile of internal trucks needed. In each hour interval, we take the internal truck requirement to be the maximum required at any time within that interval. We thus determine the internal truck requirement profile for the day in hourly intervals.

| Hatch | Quay crane | Workload $h$ | Planned processing duration (minutes) |
|---|---|---|---|
| 1 | 1 | 32 | 68 |
| 2 | 1 | 26 | 57 |
| 3 | 1 | 36 | 75 |
| 4 | 1 | 32 | 68 |
| 5 | 1 | 40 | 82 |
| 6 | 2 | 56 | 111 |
| 7 | 2 | 39 | 81 |
| 8 | 2 | 27 | 59 |
| 9 | 2 | 56 | 111 |
| 10 | 3 | 25 | 55 |
| 11 | 3 | 35 | 73 |
| 12 | 3 | 46 | 93 |
| 13 | 3 | 58 | 115 |

**Table 3: This table shows the information for a single vessel. Quay cranes 1, 2, and 3 were allotted to work on a vessel with 13 hatches, beginning at 5:00 AM. Each quay crane works on the hatches listed for it in top to bottom order. The computed planned processing duration for each hatch is shown.**

### The Optimal Hiring Policy for Drivers of Internal Trucks

At HIT, drivers work eight-hour shifts, including a one-hour meal break. The meal-break plan the drivers follow determines the timing of their meal breaks. The most commonly used plan set by the company is to have a meal break four hours after beginning work.

As of 2004, HIT uses only three shifts for drivers (the morning shift 8:00 AM to 4:00 PM, the evening shift 4:00 PM to 12:00 midnight, and the night shift 12:00 midnight to 8:00 AM). We investigated whether we could reduce the number of internal truck drivers needed by using 24 different shifts, one beginning every hour on the hour (appendix). The model we used is a pure integer program, but we noticed that rounding the optimum solution of its LP relaxation almost always yields a plan very close to the integer optimum; therefore we use this much faster solution in day-to-day operations.

We found that introducing shifts beginning every hour for internal truck drivers would reduce the number of drivers needed by about 15 percent. However, the current practice of drivers working only three shifts per day is a decades-old established practice. HIT is negotiating with the contractors that provide the drivers to gradually introduce the practice of drivers coming to work at more than three times per day.

## Decision Problem D4: Deploying Yard Cranes to Blocks

The type of yard crane most commonly used in the Hong Kong port is the rubber-tired gantry crane, which can be moved from block to block. A typical block has seven rows (or lanes) of spaces, six of which are used for storing containers in stacks and the seventh is reserved for truck passing. Each row typically consists of over 20 20-foot container stacks stored lengthwise end to end. For storing a 40-foot container stack, two 20-foot stack spaces are used.

The rubber-tired gantry crane stands on two rows of tires and spans the seven rows of spaces of the block between the tires. The bridge (top arm) of the crane has a spreader (container picking unit) that can travel across the width of the block between rows 1 to 7.

The rubber-tired gantry crane can move on its tires along the length of the block. With these two motions, it 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.

Three of these cranes can work in a block simultaneously. HIT has 95 blocks and only 122 yard cranes. Because of uneven workload distribution, often one block will be served by three cranes while other blocks remain unattended. The company moves the cranes, from low-workload blocks to high-workload blocks. It reviews the workload distribution in the blocks in periods ranging from one to four hours and decides whether to move the cranes between blocks. However, typically the company moves each rubber-tired gantry crane at most once in any four-hour period because they move over the roads very slowly.

Because it moves slowly along the road, the rubber-tired gantry crane is unproductive and an obstruction to traffic. The goal of decision problem D4 is to minimize crane travel time on the road. We modeled and solved this problem as a transportation problem, which provides quicker response than analytically more involved methods, such as Zhang et al.'s (2002). HIT gets two benefits: an increase in the productivity of the cranes from minimizing their unproductive travel time and a reduction in the congestion they cause while traveling on the roads.

## Decision Problem D5: Setting Appointment Times for External Trucks

Our goals in solving decision problem D5 were to develop a system to distribute the workload of handling external trucks evenly over time, to minimize the turnaround time of external trucks, and to reduce the number of external trucks on the terminal road network during peak hours. HIT established advanced booking in 1997 (requiring external trucks to call the terminal to make appointments).

Most of the external trucks bringing export containers to the terminal come from mainland China. These drivers cannot control the time they spend at the border crossing. Hence, HIT does not require external trucks bringing export containers into the terminal to book appointments in advance. It requires only external trucks coming to the terminal for pickups to make appointments in advance and only for the terminal peak time of 8:00 AM to 5:00 PM.

The booking system currently implemented at HIT is a straightforward quota system. In every 30-minute time slot, each block has a certain preset quota for container pickups. The main challenge in designing this system was to determine the optimal value for this quota that minimizes a combined penalty for yard-crane idle time and for the fraction of time during which the queue of trucks waiting at the block for service from the yard crane is too large. We determined the optimal value for the quota using a simulation model.

The workload for the yard cranes in a block can be divided into three parts:

*Part* 1. Removing import containers from internal trucks and storing them, and retrieving stored export containers and placing them on internal trucks for loading into vessels.

*Part* 2. Removing export containers from external trucks for storage.

*Part* 3. Retrieving stored import containers and placing them on external trucks picking them up.

The workload in a block from Parts 1 and 2 depends on the vessel arrival and departure schedules, which the terminal cannot control. But we estimated the distribution of workload from Parts 1 and 2 from past data. The only way HIT can influence the workload in a block in any 30-minute interval is by controlling the number of external trucks visiting it to pick up import containers in Part 3. Of the external trucks that made appointments to come in a 30-minute interval, a fraction may not show up; we estimated this fraction from past data. From past records, we also estimated the time it takes a yard crane to serve a truck. All these estimates are inputs for the simulation model.

HIT's target is to keep the number of trucks waiting for service at a container block at six or fewer. For different values of the quota, the simulation runs obtained estimates of (1) a yard crane's average idle time, and (2) the fraction of time during which more than six trucks wait to be served by the yard cranes in a block. The optimal value for the quota should minimize a composite penalty for these two measures.

The booking system is an automated telephone-based system. The drivers of external trucks wanting to pick up containers phone to request times and are given appointments based on the remaining quotas for the blocks in which their containers are stored. Thus, the earlier they phone, the more time slots they can choose from. This system will soon be available online as well.

If an external truck arrives for a pick-up without advanced booking, it must go to a booking center to make an appointment. However, an external truck bringing an export container into the terminal can pick up a container on the way back without booking in advance.

HIT was the first container terminal to establish such an appointment system. It reduces the number of external trucks in the terminal during peak hours and their turnaround time.

Determining the quota for the number of appointments in each half-hour interval is a one-time decision.

## HIT Launches 3P

HIT implemented all our decision-making procedures in its productivity plus program (3P) for daily operations. Beyond the challenges that prompted our development of the DSS, we faced many obstacles in developing and implementing the DSS in 3P and replacing the legacy system. HIT had to work very

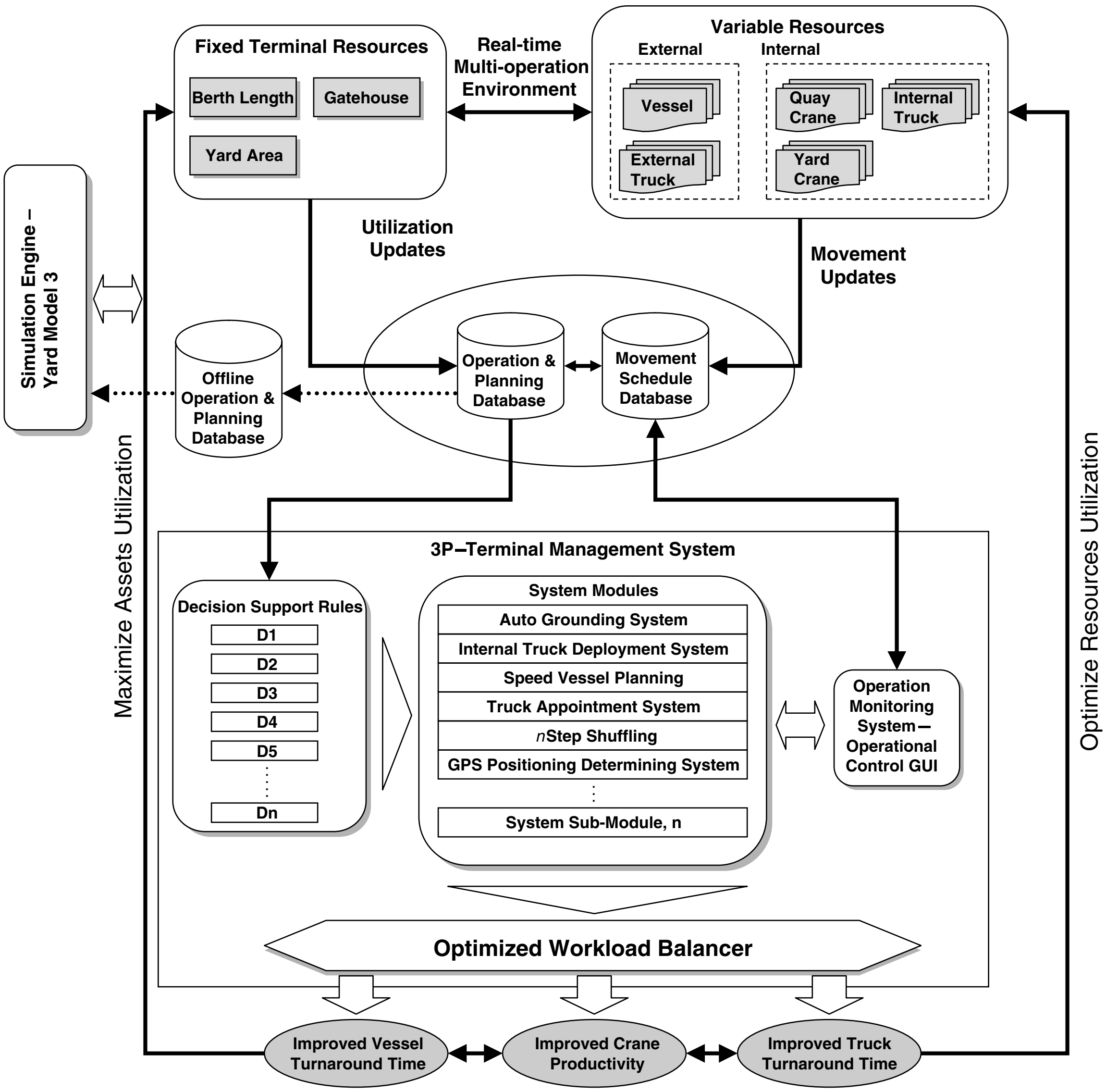


**Figure 4: 3P is the core of terminal operations and is closely connected to other internal and external systems, such as the resource-management system, billing, and the customer-plus system (a business-to-business interface). 3P has a three-tier client/server architecture based on the Tuxedo transaction monitor and the Oracle relational-database-management system; it also features multiple levels of resilience, replicated systems, and fully integrated subsystems. The system runs on HP9000 series UNIX with 10 CPUs. 3P is currently under major revamp to migrate to a Java (J2EE)-based environment. The new system, named nGen (the next generation terminal-management system), will further enhance operating capability and customers' connectivity.**

hard to retrain its existing staff and to gain broad support for 3P to realize its full potential.

3P is a terminal operating system, developed internally to provide end-to-end, real-time monitoring and optimization of container movements within the terminal. By translating the key findings and decision-support algorithms from our OR/MS research into intelligent business rules, 3P continuously monitors terminal conditions and provides optimized decision support for HIT.

3P boosted HIT's productivity. It also earned HIT the 1997 Computerworld Smithsonian Award intended to identify and honor those who use information technology to benefit mankind. HIT was the first Asian winner in the transportation category.

3P consists of a number of functional modules that implement the various decision-support algorithms. These modules perform functions ranging from simulation and planning to execution, monitoring, and performance analysis.

## Overall Impact Summary

HIT maintains a comprehensive record of all transactions and performance indicators at the terminal. With its extensive database, the company can thoroughly analyze operational enhancements' effects on productivity and efficiency.

Since HIT initiated the project, its annual throughput has grown from about four million TEUs in 1995 to about six million TEUs in 2002 (Figure 5).

By implementing the DSS and the resulting terminal-management system (3P), HIT gained the ability to handle this nearly 50 percent increase in annual throughput without increasing its staff, real estate, or equipment. As the terminal was already working at capacity when the project began, this increase in throughput demonstrates how effective the DSS and the associated productivity-enhancement programs have been.

## Benefits to HIT, HPH, and Beyond

No one can better assess the benefits of the DSS than HIT's managing director, Eric Ip, who summarized them in this way "Ports are the gateways in global distribution networks and play a significant role in completing the global supply chain. Regional competition amongst terminal operators is fierce. Every day players in the industry rack their brains to improve efficiency and service levels to surpass their competitors. It is not an exaggeration to say that efficiency is a matter of business survival. As terminal operations consist of many interrelated and complex decisions, achieving and beating targeted terminal efficiency requires us to make a series of right decisions. A single wrong decision, just like any weak link in the supply chain, may create havoc throughout the entire chain.

"3P, with its inherent intelligence, has provided HIT the capacity of almost five additional berths, at a mere fraction of the cost. Higher efficiency has provided HIT greater flexibility, not only in developing new services for the customers, but also further raising the bar for competition. With the valuable reputation of HIT as a catch-up port, customers can count on us to make up for any lost time on their schedules. As HIT's customers are the world's leading shipping lines, its efficiency has enabled HIT to become the gold standard of how any premier port should perform in the world. HIT has never been better positioned to expand its services to other parts of the world. As such, we are continuously and systematically transferring important systems and know-how to our sister HPH ports.

"In addition, the environmental impact of the DSS is noteworthy. HIT has effectively avoided the construction of resource-intensive container berths. At HIT alone, the number of diesel-burning internal trucks has been cut in half. On a grander scale, the increased productivity of the terminal as a whole is passed on throughout the supply chain. Container ships are more productive; truckers are more productive."

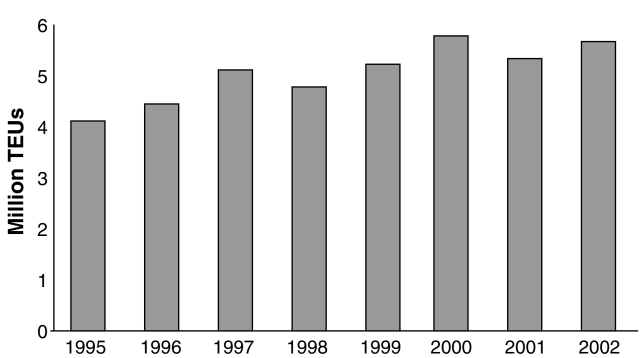


**Figure 5: The annual throughput (all quayside moves) at HIT has increased from about four million TEUs (20-foot equivalent units) in 1995 to almost six million in 2002.**

## Performance Benefits

HIT improved vessel movement: After 3P embedded the optimized decisions in daily operations, HIT saw a 30 percent improvement in vessel turnaround time and a 47 percent improvement in vessel operating rate (the total number of container moves per hour handled by all the quay cranes serving the same vessel).

HIT improved internal truck utilization: After implementing the DSS with 3P, HIT reduced the number of internal trucks it uses. With the use of a common pool of internal trucks, the company deploys four, rather than eight, trucks for each quay crane, a 50 percent improvement made without sacrificing productivity.

HIT reduced traffic congestion: Despite the 50 percent increase in throughput, the terminal is less congested. With optimized dispatch, routing, and allocation, the cycle time for internal trucks has dropped by 16 percent. The turnaround time for external trucks delivering or collecting containers has fallen by 30 percent from 60 minutes to 40 minutes.

HIT improved crane productivity: With optimized dispatch of internal trucks, allocation of yard storage space, and deployment of yard cranes, HIT increased the productivity of its yard cranes by 15 to 20 percent and increased the quay crane rate by 45 percent.

HIT cut the cost of handling containers: By optimizing its use of resources, HIT cut the cost of handling each TEU by more than 35 percent. HIT handled greater throughput at a lower rate without expanding yard space (Figure 6).

Recognizing HIT's superior performance in operations productivity, the Hong Kong Productivity Council awarded HIT the Hong Kong Award for Industry—Productivity Category in 2003.

## Financial and Economic Benefits

The savings from using fewer internal trucks per quay crane and the pooling system for internal trucks implemented at HIT in 1998 translate into a direct savings of US$6 million per year from just the reduction in the number of internal trucks used. Since 1998, HPH has implemented the system at four other large

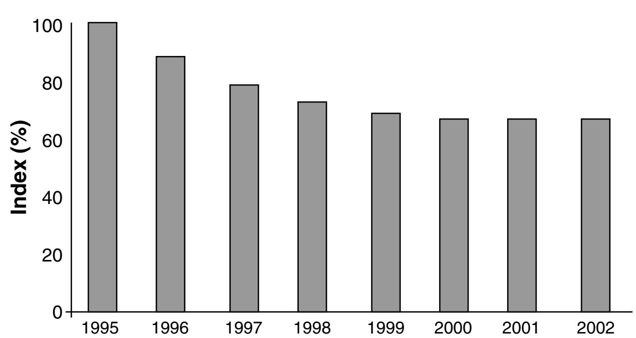


**Figure 6: The yearly average handling cost per TEU (20-foot equivalent unit) dropped between 1995 and 2002.**

terminals. Between 2004 and 2009, the HPH Group plans to implement the system at 25 additional container terminals within its global network of ports to save US$60 million per year.

—By reducing unit handling costs by 35 percent, HIT has reduced its costs for handling nearly six million TEUs each year.

—HIT has avoided capital investments that would have been necessary to handle its current throughput volume. In 1995, HIT had reached its annual handling capacity of four million TEUs. Without 3P, it could not have handled its current throughput without investing roughly US$333 million in the equivalent of at least another two-berth facility. Using a 20 percent cost of capital, it avoided capital expenditures of US$66 million annually.

HIT's customers also saved money. The time vessels spend docked at ports is not money-making time for shipping lines, so they would like to minimize it. As of 2003, HIT was handling 125 vessels per week on average. With the improvement in waterside efficiency, vessel turnaround time fell from 13 to nine hours, saving HIT's customer shipping lines US$65 million annually, based on the prevailing vessel charging rate and operating costs. Customer satisfaction breeds loyalty, which has strengthened HIT's reputation as the catch-up port among customers who have lost time at other ports in their service loops.

## Conclusion

Our team developed innovative operations research techniques and designed a DSS based on an information system (3P) for making five important decisions efficiently. HIT has implemented our solutions to four of these decision problems and obtained major benefits. The solution we developed for optimizing internal-truck driver shifts awaits implementation pending union negotiations.

HIT has achieved results that were much greater than originally expected. With almost a 50 percent improvement in capacity with no increase in land, we are well on our way to achieving elastic capacity. The accolades from the Smithsonian and other esteemed organizations attest to how HIT has created value and differentiated the company from the other firms even by established organizations.

As the center of excellence for HPH's 35 ports, HIT actively researches and replicates its best practices at its sister ports around the world. HIT's unified internal truck pool, for example, has been adopted in many places with quantifiable savings. It illustrates how our elastic capacity can multiply. Elastic capacity is crucial to HPH's business; it obviates the construction of costly ports and their potential environmental impact. It also enables the terminal to provide value-added services previously not possible.

Our work is far from finished. HPH needs to improve continuously while standardizing all of the group's ports with the same terminal-operating system. HIT is actively pursuing the application of operations research methods in many other areas. For example, by matching each discharge from a vessel to the yard with a corresponding load from the yard to a vessel and making the trucks continuously carry containers in both directions, we could double a terminal's productivity.

The elasticity of capacity and productivity means that there is no telling where the outer limits lie. The only certainty is that improved optimization can provide better daily decision making and continue to push the performance envelope for academic research, for HPH, and for the industry as a whole.

## Appendix. Model to Determine the Optimal Policy for Hiring Internal-Truck Drivers

Let $d_k =$ number of internal trucks required to help quay cranes during the $k$th hour interval of the day, $k = 1$ to 24. Assume that internal-truck drivers can begin work on the hour, any hour of the day. Drivers who begin work on the $j$th hour constitute the $j$th hour shift.

A driver who is available for work during the 4:00 AM to 5:00 PM interval may be from the 9:00 PM shift of the previous day, or a later shift, depending on the meal-break plan followed. Similarly, for the $k$th hour interval of the day, let $S_k$ denote the set of shifts of the drivers who will be working during the $k$th hour interval. The set $S_k$ depends on the meal-break plan in effect, and for some $k$, some of the shifts in $S_k$ may belong to the previous day. The decision variables for the model are:

$y_j =$ number of internal-truck drivers working in the $j$th hour shift.

Then, the model for determining the $y = (y_j)$ to meet the internal-truck requirements using the minimum number of drivers is the integer program

$$\begin{aligned} \text{minimize} \quad & \sum y_j \\ \text{subject to} \quad & \sum_{j \in S_k} y_j \geq d_k \quad \text{for all } k, \\ & y_j \geq 0 \quad \text{for all } j. \end{aligned}$$

It is convenient to implement the rounded solution of the LP relaxation of this integer program, because we found that it is very close to the true optimum of this integer program most of the time.

### References

Bish, E. K., T. Y. Leong, C. L. Li, J. W. C. Ng, D. Simchi-Levi. 2001. Analysis of a new vehicle scheduling and location problem. *Naval Res. Logist.* **48**(5) 363–385.

Meersmans, P. J. M., R. Dekker. 2001. Operations research supports container handling. Report EI 2001-22, Econometric Institute, Erasmus University, Rotterdam, The Netherlands.

Murty, K. G. 1997. Storage yard planning in container shipping terminals. Technical report, Department of Industrial Engineering and Engineering Management, Hong Kong University of Science and Technology, Kowloon, Hong Kong, China.

Murty, K. G., P. A. Djang. 1999. The US Army National Guard's mobile simulators location and routing problem. *Oper. Res.* **47**(2) 175–182.

Murty, K. G., J. Liu, Y.-w. Wan, R. Linn. 2005. DSS (decision support systems) for operations in a container terminal. *Decision Support Systems* **39**(3) 309–332.

Vis, I. F. A., R. De Koster. 2003. Transshipment of containers at a container terminal: An overview. *Eur. J. Oper. Res.* **147**(1) 1–16.

Zhang, C. 2000. Resource planning in container storage yards. PhD dissertation, Department of Industrial Engineering and Engineering Management, Hong Kong University of Science and Technology, Kowloon, Hong Kong, China.

Zhang, C., Y.-w. Wan, J. Liu, R. Linn. 2002. Dynamic crane deployment in container storage yards. *Transportation Res.: Part B Methodological* **36**(6) 537–555.

Zhang, C., J. Liu, Y.-w. Wan, K. G. Murty, R. J. Linn. 2003. Storage space allocation in container terminals. *Transportation Res.: Part B Methodological* **37**(10) 883–903.