ANALYSIS AND EVALUATION OF EFFICIENT DISPATCHING RULES FOR MULTI-DEVICE HANDLING SYSTEMS WITH RANDOM MOVE REQUESTS

by

Chate Eamrungroj

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctoral of Philosophy
(Industrial and Operations Engineering)
in the University of Michigan
2017

Doctoral Committee:

Professor Yavuz A. Bozer, Chair
Associate Professor Mariel Lavieri
Associate Professor Amitabh Sinha
Professor Mark P. Van Oyen

Chate Eamrungroj

chateae@umich.edu

ORCID iD: 0000-0003-0532-9799

ACKNOWLEDGEMENT

First, I would like to express my sincere gratitude to my advisor, Professor Yavuz Bozer, for his immense knowledge, inspiration, patience, and continuous support throughout my Ph.D. study. Without his guidance, researching and writing this thesis would not have been possible. I would also like to thank the members of my committee, Professor Mariel Lavieri, Professor Amitabh Sinha and Professor Mark Van Oyen, for their insights and feedback. Their input and questions helped improve the dissertation both in substance and style. I would like to extend my special thanks to the UMHS staff, in particular Frank Krupansky (Materiel Services Director) and Lakita Pogue (Patient Transportation Manager), for providing me with the opportunity to study patient movement at the U-M hospital. They were gracious in answering my ever-expanding questions and giving me their valuable times.

I would like to thank my mother Sriwan, father Praphan, and sister Amy, for their love, encouragement, and financial support throughout my life and career. Finally, and most importantly, I would like to thank my wife Joanna for her unwavering love, patience, and for bringing joy to my life.

TABLE OF CONTENTS

CHAPTER

LIST OF TABLES

TABLE

LIST OF FIGURES

FIGURE

LIST OF APPENDICES

APPENDIX

LIST OF ABBREVIATIONS

AEQ	Already-Equipped
AGV	Automated Guided Vehicle
BSI	Bozer-Srinivasan Index
BSD	Busy State Dispatching
B-STTF	Bounded Shortest-Travel-Time-First
EQ	Equipment
EQMA	Equipment Marshalling Area
CF	Closest First
CL	Closet
DID	Device Initiated Dispatching
DS	Dispatching System
FCFS	First-Come-First-Served
GR	Gurney
GRMA	Gurney Marshalling Area
LB	Lower Bound
L/OF	Local, if not, Oldest First
MA	Marshalling Area
MHS	Material Handling System

Mod-FCFS	Modified First-Come-First-Served
MR	Move Request
M-M LB	Maxwell and Muckstadt Lower Bound
NF	Net Flow
OF	Oldest First
OL	System Overload
PM	Patient Mover
PP	Percentage Point
P-MR	Priority Move Request
R-MR	Regular Move Request
SID	Station Initiated Dispatching
SSUF	Shortest-Set-Up-First
STTF	Shortest-Travel-Time-First
UMHR	University of Michigan Hospital Rule
UMHS	University of Michigan Health System
UHB	University Hospital Building
UNS	System Unstable
WH	Wheelchair
WHMA	Wheelchair Marshalling Area

ABSTRACT

Three essays are presented concerned with device dispatching within the framework provided by trip-based material handling systems, which represent a wide range of systems such as lift trucks and unit load automated guided vehicles in industrial applications and patient movement systems in healthcare applications. The move requests (MRs) arrive according to a Poisson process, and each MR is served, one at a time, by one of the devices, according to the dispatching rule. In the first essay, a new analytic model is developed to estimate empty device travel with multiple devices operating under the modified first-come-first-served (Mod-FCFS) dispatching rule. The analytic technique used in the first essay is extended in the second essay to develop a new analytic model to estimate empty device travel with multiple devices operating under the shortest-travel-time-first (STTF) rule, which is a simple, well-known, and efficient rule but difficult to model analytically. To our knowledge, the analytic model in the second essay is the first model to explicitly estimate empty device travel under STTF. Using simulation, we show that both models perform well in estimating empty device travel and the expected device utilization. We also investigate the MR wait times under STTF, and propose a new rule, namely B-STTF, to avoid excessive wait times which are known to occur under STTF. The results show that B-STTF is as efficient as STTF while successfully avoiding excessive MR wait times.

In the third essay, we model intra-facility patient movement systems as a trip-based material handling system, where a “device” represents a patient mover (PM), and each “MR” represents a patient. Using simulation, we study the performance of the patient movement system as a function of the dispatching rule and the number/location of the equipment marshalling areas (that is, wheelchairs, gurneys, and closets). System performance is measured in terms of efficiency (i.e., reducing empty travel for the PMs and the expected MR wait times) and effectiveness (i.e., avoiding excessive wait times for the MRs). We observe that the FCFS dispatching rule is less efficient but it avoids excessive patient wait times since the MRs are served according to their order of arrival. In contrast, the STTF rule is efficient but some patients experience excessive wait times, which is detrimental. We thus present a new rule that strikes a balance between efficiency and effectiveness. We also analyze the impact of the number and location of the equipment marshalling areas. We observe that carefully planning the number and location of the marshalling areas based on usage by the PMs can improve the performance of the system as much as, if not more than, a more efficient dispatching rule.

CHAPTER 1

Overview

Material handling systems (MHSs) are responsible for delivering the right materials at the right place and at the right time. Although material handling in manufacturing is considered a non-value adding function, a well-designed MHS often reduces operating costs and improves productivity. A poorly designed MHS may disrupt workflows, increase work-in-progress inventories, and may even impact the safety and quality of the product made or service delivered. Many facilities including manufacturing plants, warehouses, distribution centers, and service facilities rely heavily on the performance of the MHS. As a result, material handling plays a significant role in many industries. A study by Research and Market, for example, estimated the 2015 North America MHS market to be over $21 billion (2016).

Although there are many types of MHSs, they can be categorized mainly into two types of systems; namely, trip-based MHSs or conveyor-based systems (Figure 1.1).

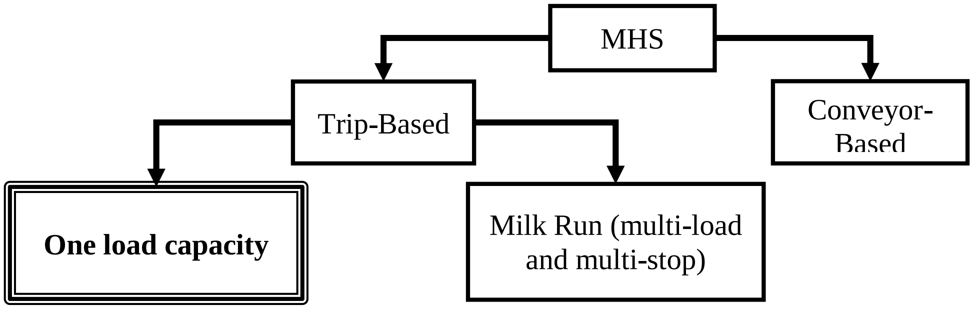


Figure 1.1: Types of MHSs

The dissertation consists of three essays. The first two essays are concerned with device-dispatching in single-load, trip-based MHSs, where one or more devices are dispatched by a centralized dispatching system to serve the move requests (MRs) one at a time. The MRs arrive one-at-a-time, according to a Poisson process, with a known origin and destination for each MR. Once assigned to a MR, the device travels empty from its current location to pick up the MR, and then travels full/loaded to the destination of the MR. Upon delivering the load (i.e., upon serving the MR), the device is assigned to the next MR, or it becomes idle if there are no MRs in the system. For a given layout and flow data, the efficiency of the MHS depends largely on the dispatching rule, which determines which MR to assign to an empty device, and vice versa. In most cases, a computer-based technology/system is employed to manage and dispatch the devices.

Trip-based MHSs of the above type can be used to model many handling systems in various applications. In industrial applications such as manufacturing and warehousing, devices such as lift trucks, unit-load automated guided vehicles (AGVs), and bridge cranes, to name a few, can be modeled as trip-based MHSs. In transportation applications such as taxicabs and Uber, the system can again be modeled as a trip-based MHS, with vehicle dispatching being one of the primary concerns, although moving passengers is technically not considered "material handling."

Designing a successful trip-based handling system depends on a number of factors, including the location of the stations, device routing and MR flow data. Given these factors, one is often concerned with determining the number and utilization of the devices. However, depending on the dispatching rule, analytic estimation of the empty trips (and device utilization) can range from straightforward to very difficult, and, simulation models are often used to estimate the performance of the system.

In the first essay, an iterative algorithm is used to develop a new analytic model to estimate empty device travel in a multi-device system operating under the modified first-come-first-served (Mod-FCFS) dispatching rule. In the second essay, the analytic technique from the first essay is extended to develop a new analytic model to estimate empty device travel in a multi-device system operating under the shortest-travel-time-first (STTF) dispatching rule, which is a well-known and efficient rule but difficult to model analytically. To our knowledge, the analytic model in the second essay is the first model to explicitly estimate the empty device trips under the STTF rule. Furthermore, the MR wait times under STTF is investigated, and a bound is imposed in order to avoid excessive MR wait times.

Generally speaking, in a healthcare setting, patient conveyance can be categorized into *inter-facility* patient transport and *intra-facility* patient movement. Inter-facility transports are typically performed by ambulance (and sometimes helicopter), where multiple patients can be transported at a time. Intra-facility patient movements are often performed by PMs, where the patients are moved one at a time, on either a wheelchair or a gurney. In a large hospital, patient movement is a non-trivial, time-sensitive operation that often takes place in a multi-floor facility, involving retrieving and/or depositing the wheelchair and/or gurney, while serving multi-priority patients. For example, the University of Michigan Health System (UMHS), recently renamed Michigan Medicine, is a group of interconnected, multi-floor buildings, with over 300,000 patient moves per year.

Trip-based MHSs can also be used for modeling patient movement or patient transport, where ambulances or patient movers (PMs) transport patients from one point to another. The third essay in fact is concerned with dispatching PMs for intra-facility patient movement in a hospital setting. Such systems have an additional requirement in that the PM must first acquire the proper equipment (wheelchair or gurney) before moving a patient. The purpose of the study is to develop insights and recommend improvements for the patient movement system by investigating alternative PM dispatching rules. Since equipment is a key factor, we also investigate the impact of the equipment marshalling areas on the performance of the patient movement system.

CHAPTER 2

Throughput Analysis of Multi-Device Trip-Based Material Handling Systems Operating under the Modified-FCFS Dispatching Rule

2.1 Introduction

In many manufacturing and transportation applications, the material handling system plays a significant role. While in many cases material handling itself is a non-value-adding function, a poorly-designed material handling system often results in missed deliveries, poor customer service, large work-in-process, and reduced productivity. It may also adversely affect quality and safety.

The material handling system we focus on in this study is a trip-based material handling system, which consists of one or more material handling devices, operating independently to serve move requests (MRs) (Srinivasan et al., 1994). A MR is a physical entity that has a known origin (or pick-up station) and a known destination (or deposit station). It may be a unit load in a manufacturing system or a pallet in a warehouse. Each MR arrives one at a time and waits at its pick-up station. In order to serve a MR, a device travels empty from its current location to the appropriate pick-up station, picks up the load and travels full (or loaded) to the appropriate deposit station, where the load is deposited, and the device becomes empty again, ready to serve the next MR. (As the above description suggests, we will use the terms MR and "load" interchangeably.)

The devices are assumed to be homogeneous, and each device serves only one MR at a time. In manufacturing, a wide range of material handling systems can be modeled as a trip-based handling system including lift trucks, bridge cranes, unit load automated guided vehicles (AGVs), and manual systems (where operators move one load at a time using dollies or similar equipment). Transportation services that can be modeled as a trip-based handling system include taxi service in a city and patient transportation in a hospital.

The successful design/operation of a trip-based handling system depends on a number of factors including the location of the pick-up and deposit stations, the travel path of the devices and device routing, the MR flow data, and the device dispatching rule, which is concerned with assigning a MR to a device and vice versa. At a basic level, for a given MR flow matrix, coupled with a given layout and device travel times, one is often concerned with determining the number and expected utilization of the devices. One may also focus on the expected waiting times of the MRs, assuming that a sufficient number of devices is provided.

Since the MRs must be served on a timely basis (that is, the devices are required to perform loaded trips), empty device travel may be significant, depending on the data, the layout, and the dispatching rule (see [Egbelu and Tanchoco, 1984], and [Koo and Jang, 2002], among others). As shown by these studies, an efficient dispatching rule can lower empty device travel, which often reduces the number of devices required and/or the expected MR waiting times.

Assuming the MRs arrive randomly and independently, the analytic estimation of the number of empty trips, and thus the number of devices required, can be straightforward or very challenging, depending on the dispatching rule. In fact, for most dispatching rules, the problem is, generally speaking, not tractable and, therefore, either analytic approximations (or bounds) are developed or simulation models are used to estimate the number of devices required.

The purpose of this study is to develop an analytic model to estimate empty device travel for the Mod-FCFS dispatching rule proposed earlier by Srinivasan et al. (1994). Since it is an approximate model, we will use simulation to conduct a more detailed analysis of the system and to obtain other results such as the expected MR waiting times. Our intent is to develop an analytic model that can be used to rapidly evaluate a number of alternative handling systems and to conduct "what if" analyses based on varying the flow data and/or the layout. We also shed light on device versus station initiated dispatching, which has not been treated fully/correctly in the literature.

The remainder of the chapter is organized as follows. In Section 2.2, pertinent dispatching rules and analytic models in the literature are reviewed, while in Section 2.3, the problem setting and the assumptions are described. The analytic model is presented in Section 2.4. In Section 2.5, the MOD FCFS dispatching rule and the analytic model are evaluated using simulation under different layouts and flow matrices. Lastly, the results are summarized in Section 2.6, where possible future research directions are also discussed.

2.2 Literature Review

The literature review is limited largely to those papers concerned with device dispatching and analytic modeling in trip-based material handling systems. An early paper that compares alternative dispatching rules is presented by Egbelu and Tanchoco (1984), who identify two types of dispatching decisions. When a device delivers a load and becomes empty, deciding which (unassigned) MR the device should serve next is defined as "device-initiated dispatching" (DID) since the decision is invoked whenever a device delivers a load and there's at least one unassigned MR in the system. If there are no unassigned MRs in the system, the device becomes idle at its last point of delivery. (There are a few papers concerned with where to "park" idle devices;

however, such are beyond the scope of our study. The interested reader may refer to [Egbelu 1993], among others, for further information.)

On the other hand, when a MR arrives, if there is at least one idle device, deciding which (idle) device to assign to the MR is defined as "station-initiated dispatching" (SID) since the decision is invoked whenever a MR arrives and finds one or more idle devices. If the MR finds all the devices busy, it will eventually be served when an (empty) device is assigned to it. That is, if a MR is not served under SID, it will be served under DID. A fully-defined dispatching policy needs to specify the rule used for both DID and SID. As we shall see later, understanding how often each decision, DID versus SID, is invoked is important in terms of assessing the significance and impact of each rule.

In (Egbelu and Tanchoco, 1984), multiple rules are compared by a simulation model for both DID and SID, where random MR, oldest MR, closest MR, and maximum outgoing queue size are considered for the former, while random device, closest idle device, and longest idle device are considered for the latter. Note that, first-come-first-served (FCFS) dispatching would generally mean that the oldest MR rule is used for DID, and the longest idle device rule is used for SID. Likewise, shortest-travel-time-first (STTF) dispatching would generally mean that the closest MR rule is used for DID, and the closest idle device rule is used for SID.

Although the study was presented for an automated storage/retrieval system (AS/RS), Chow (1986a) was among the first to present a general analytic model for a single device operating under the FCFS rule. The paper introduces the "kij triplet" concept, where the service time is modeled as the sum of two components; the first one is the empty device travel time from its current location (station k) to pick up a load at station i, and the second component is loaded device travel from station i to station j. The author shows that modeling the FCFS rule is straightforward since the next MR to be served is independent of the current location of the device. In a subsequent paper, Chow (1986b) evaluates alternative dispatching rules for an AS/RS via simulation. (Chow's method was also used by Johnson and Brandeau [1994] to develop an analytic model for a single-device AGV system.)

For DID, Srinivasan et al. (1994) present an analytic model for a modified version of the FCFS rule (i.e., Mod-FCFS), where, upon delivering a load, the empty device first checks its current location for an unassigned MR. If no such MR is found, the device serves the oldest unassigned MR in the system. Since the analytic model is based on a single device, no rule is required for SID; when a MR arrives, there can be at most one idle device. However, for the multi-device simulation model, the authors use the longest idle device rule for SID. Arguing that in most cases SID would be invoked more often, Koo and Jang (2002), on the other hand, study the longest idle device and the closest idle device rules for SID, while using a simple rule (FCFS) for DID.

Bozer and Yen (1996) propose two dispatching rules; the modified-STTF rule (Mod-STTF) and the bidding-based dynamic dispatching (B^2D^2) rule, which aim to outperform the STTF rule. Under the Mod-STTF rule, a device may be reassigned to another load while it is performing an empty trip. Under the B^2D^2 rule, a device may be assigned to multiple MRs, although they are still served one at a time. When a MR arrives, all the devices in the system place a "bid." The bid placed by a device is based on the remaining distance it must travel to serve all the MRs that have been assigned to it, plus the empty travel distance to the new MR. The new MR is assigned to the device with the lowest bid. The B^2D^2 rule is novel in that it does not wait for a device to become empty to make a decision. Both of the above rules outperform the STTF rule. However, both rules, to the best of our knowledge, are analytically intractable and they lack the simplicity and practical appeal of the STTF rule.

The above rules are all considered “centralized” in that a computer must keep track of the location of all the MRs and the devices. As an alternative, a decentralized rule, namely, the First-Encountered-First-Served (FEFS) rule was analyzed by Bartholdi and Platzman (1989) for a closed-loop AGV system, where the stations are arranged around a unidirectional loop. Under FEFS, once a device becomes empty, it continues to travel around the loop, searching for a MR. (Such systems are similar to polling systems.) Once the device moves a load, it resumes searching for a load from its current location. The authors show that the FEFS rule is an effective rule in a closed-loop system.

Bozer and Srinivasan (1991) further analyze the FEFS rule and define “mandatory” empty trips based on the net flow concept (which is explained in section 2.3). Nazzal and McGinnis (2008) propose an alternate method to model the FEFS rule and incorporate device blocking by extending a Markov chain model and computing the appropriate transition probabilities.

While the above papers focus primarily on dispatching, some papers focus on determining the minimum number of devices required. Egbelu (1987) presents four alternative simple formulas to determine the number of devices required. Empty trips are included but not as a function of the dispatching rule used. Mahadevan and Narendran (1990, 1993) present an analytic model for a flexible manufacturing system (FMS). The model incorporates the job routing flexibility of an FMS, and the probabilities of possible job sequences are used to compute the minimum number of devices.

Using the net flow concept, Maxwell and Muckstadt (1982) present an analytical lower bound for the number of empty trips required, regardless of the dispatching rule used. (The lower bound is explained briefly later in the chapter.) Subsequently, Malmborg (1991) tightened the bounds by taking the dispatching rules into consideration. The rules studied were random, closest and furthest load for DID, and random, closest and furthest device for SID. The rule with the least (most) estimated empty travel time was considered the lower (upper) bound.

Lastly, some studies model the material handling system as a queueing network in order to find the average waiting time of the MRs. Chow (1986a) approximates a single-device AS/RS as an M/G/1/FCFS queue. The service time distribution is obtained from the flow matrix. Tanchoco et al. (1987) and Wysk et al. (1987) present a model based on CAN-Q (Computerized Analysis of Network of Queues, see Solberg [1980]) to determine the number of devices needed. However, Can-Q's structure prevents the explicit consideration of the DID or SID rules. Curry et al. (2003) estimate the average waiting time for the MRs by approximating the handling system as a queueing model, using oldest MR (FCFS) for DID, and closest idle device for SID.

In conclusion, while numerous studies focus on device dispatching, very few develop an analytic model to explicitly estimate the empty trips, and those that do, assume either a single device and/or a simple rule such as FCFS. To our knowledge, our model is the first one to consider multiple devices while explicitly modeling empty device travel under a rule more efficient than FCFS. Furthermore, some studies argue that DID plays a bigger role than SID, or vice versa. We analyze the frequency of DID versus SID, and our model does not emphasize one over the other.

2.3 Problem Setting and Assumptions

The system is composed of a set of stations, where each station has a pick-up point and a deposit point. Each MR is defined by its origin station (pick-up point) and destination station (deposit point). When a MR arrives, it joins the queue at its origin station; it also joins a global queue, which maintains the order of arrival of all the MRs across the system. The MRs arrive one at a time according to an independent Poisson process with a known rate. Once a device is dispatched to pick up a MR, the MR is removed from the global queue. Subsequently, when the

MR is picked up by the device, it is delivered to the deposit point of its destination station. Possible congestion or blocking are not modeled explicitly but the travel time between stations is assumed to be exponentially distributed. For simplicity, the devices travel at the same speed whether empty or loaded.

We assume that the queue at each pick-up point and the global queue all have unlimited capacity. Once a load is delivered, service of the MR is completed and the load exits the system immediately. For simplicity, we assume that the travel distance between the pick-up and deposit points of the same station is negligible. The load pick-up/deposit times are also negligible (although extending the model for non-negligible pick-up/deposit times would be straightforward).

As explained in section 2.2, adopting the Mod-FCFS rule for DID, Srinivasan et al. (1994) present an analytic model for the case with a single device. To model a system with k devices, the authors propose to use the single-device model with a device that travels k times faster. Since the model is based on a single device, selecting an idle device when SID occurs is not a concern. Our model, on the other hand, is based explicitly on multiple devices. As such, for DID, we use the same rule as Srinivasan et al. (1994) (see section 2.2), and for SID we assume that when a MR arrives, it first checks the origin station for an idle device. If no idle device is found at the origin station, it checks the system for other idle devices. If one or more are found, the longest idle device is assigned to the MR. If none are found, the MR will be served through DID.

Note that our implementation of Mod-FCFS is logical and consistent because whether it’s DID or SID, the system first checks for an opportunity to assign an empty device to a “local MR” (DID), or assign a MR to a “local (idle) device” (SID). For brevity, we will refer to this rule as “L/OF-L/OF,” where the first position (DID) stands for “local (load); if not, oldest (load) first,” and the second position (SID) stands for “local (idle device); if not, oldest (idle device) first,” where obviously the "oldest (idle device)" corresponds to the longest idle device. The list of the dispatching rules is shown in Table 2.1.

Table 2.1: DID and SID of Dispatching Rules

Rule	DID	SID
FCFS	Oldest First (OF)	Oldest First (OF)
Srinivasan et al. (1994)	Local; if not, Oldest First (L/OF)	Oldest First (OF)
Mod-FCFS	Local; if not, Oldest First (L/OF)	Local; if not, Oldest First (L/OF)
STTF	Closest First (CF)	Closest First (CF)

2.4 The L/OF-L/OF (Mod-FCFS) Rule and the Analytic Model

In this section, we present a multi-device analytic model for the L/OF-L/OF rule. Since the loaded trips are given as data, our goal is to compute the empty trips, and ultimately the expected device utilization. We also investigate DID versus SID since, contrary to claims made in the literature, they both play a key role in estimating the performance of the system and the resulting empty trips.

The number of devices in the system is denoted by D, and the number of stations by S. The number of empty and loaded trips per hour are denoted as e_{ij} and f_{ij}, where e_{ij} represents the number of trips an *empty* device makes per hour from station i to station j to pick up a load at station j, and f_{ij} represents the number of trips a *loaded* device makes per hour from station i to station j $(i \neq j)$ to deliver a load at station j.

Assuming that a sufficient number of devices is provided, let α_f (< 1) and α_e (< 1) denote the proportion of time a device is traveling loaded and empty, respectively. (Recall that the devices are assumed to be homogeneous.) Letting t_{ij} denote the (loaded or empty) travel time in minutes from station i to station j, the expected device utilization $(\rho < 1)$ is computed as follows:

$$\rho = \alpha_f + \alpha_e = \frac{\sum_{i\in S}\sum_{j\in S}(t_{ij}f_{ij})}{60D} + \frac{\sum_{i\in S}\sum_{j\in S}(t_{ij}e_{ij})}{60D} \tag{2.1}$$

Computing the first term is straightforward since the f_{ij} values are given as data. (It is also straightforward to include possible load pick-up/deposit times by adjusting the t_{ij} values in the first term.) However, in order to compute the second term, the e_{ij} values are needed. Before we present the analytic model to estimate the empty trips (section 2.4.3), we first address three related issues in the following sections.

2.4.1 Net Flow

The net flow (NF) of station i is based on the rate at which loads are delivered at station i versus picked up from station i. As defined in (Bozer and Srinivasan, 1996), if Λ_i denotes the rate at which loads are delivered at station i, and λ_i denotes the rate at which loads are picked up from station i, then:

$$NF_i = \Lambda_i - \lambda_i = \sum\nolimits_{k\in S} f_{ki} - \sum\nolimits_{k\in S} f_{ik} \tag{2.2}$$

Furthermore, if we let $\Lambda_T = \sum_{i\in S}\Lambda_i$ and $\lambda_T = \sum_{i\in S}\lambda_i$, then $\Lambda_T = \lambda_T$ and $\sum_{i\in S} NF_i = 0$ since flow is conserved globally. However, individual stations may have positive or negative NF values depending on the data. As explained in (Maxwell and Muckstadt, 1982), stations with positive NF values “generate” empty devices, while stations with negative NF values “consume” empty devices. Stations with a zero NF value are “balanced” stations but a device that delivers a load to such a station may still depart empty.

As with other dispatching rules, the performance of L/OF-L/OF depends largely on the flow data and the layout of the stations. If the flow is highly unbalanced (that is, most or all of the stations have very negative or very positive NF values), it is less likely that, upon delivering a load (upon arriving), a device (a MR) will find a local MR (a local idle device). The more often a local load (or local idle device) is *not* found, the more the performance of L/OF-L/OF will approach that of OF-OF (that is, FCFS or "oldest first" for both DID and SID). If the flow is balanced (that is, most or all of the stations have zero or near-zero NF values), the device (the MR) is more likely to find a local load (a local idle device), and as a result, the L/OF-L/OF rule should perform better than the OF-OF rule.

We demonstrate the impact of NF through a simple, 4-station example with 3 devices. The system is simulated with unbalanced and balanced flow data, *although the total workload is fixed* at 28 loads/hr (see Appendix 2.A). Taking the closest-first (CF) rule as the baseline for each case, the ρ values and the percent increase in them for both cases are shown in Table 2.2. As expected, ρ increases considerably when the flow is unbalanced, and the performance of L/OF-L/OF improves when the flow is balanced. Past results have established that FCFS gives high ρ due to unnecessary empty travels while STTF is an efficient rule. And under light traffic, our empirical results suggest that all the rules listed in Table 2.2 show comparable performance. This is primarily because, under light traffic, there are very few (typically no more than one or two) MRs present when the device becomes empty. Consequently, a device is almost always dispatched to serve the same MR regardless of the dispatching rule in effect (Srinivasan et al. 1994).

Table 2.2: Expected Device Utilization in a 4-station Example

	OF-OF (FCFS)	**L/OF-L/OF (Mod-FCFS)**	**CF-CF (STTF)**
Unbalanced flow	0.894 (23%)	0.847 (17%)	0.727
Balanced flow	0.855 (33%)	0.711 (11%)	0.643

2.4.2 DID versus SID

As shown in section 2.4.3, DID and SID impacts the empty trips and the analytic model. Furthermore, there have been conflicting views in the literature on the significance of DID versus SID and how often the two are invoked. For example, Srinivasan et al. (1994) argue that SID “generally has little or no impact on the throughput capacity of the system since it is usually invoked very seldom” when the “minimum or near-minimum required number of devices are used” because “the probability of finding two or more idle devices in the system diminishes quite rapidly.” While we partly agree with the above statement, it is somewhat ambiguous since the near/minimum number of devices may correspond to a range of ρ values (such as $\rho \cong 0.98$ or $\rho \cong 0.90$), and it is not clear how seldom SID would be invoked for more realistic ρ values such as $\rho \cong 0.80$. Also, the number of devices would impact the validity of the above statement.

On the contrary, it is argued by Koo and Jang (2002) that SID would be invoked more frequently since the probability of invoking DID, that is, the probability of finding all k devices busy, which is claimed to be equal to ρ^k, decreases rapidly with the number of devices. Unfortunately, this argument is not valid since ρ^k does not correspond to the above probability. In general, if the MR finds one server busy, it increases the probability of finding another server busy (see [Nelson 1995, p. 322] for an elegant proof for the $M/G/k$ queue.) In fact, the probability of invoking DID is provided by the well-known Erlang C formula for the $M/M/k$ queue (Nelson 1995, 372), which is repeated below for D devices:

$$E_C = \frac{\frac{(\rho D)^D}{D!}}{\frac{(D)^D}{D!} + (1-\rho)\sum_{\ell=0}^{D-1}\frac{(\rho D)^\ell}{\ell!}} \tag{2.3}$$

For example, for $\rho = 0.90$ and $D = 10$, the probability of DID and SID is about 0.67 and 0.33, respectively. As suggested in (Srinivasan et al. 1994), DID is more dominant but both probabilities are non-negligible, and one cannot say that SID would be invoked “very seldom.” And certainly one cannot say that SID would be invoked more frequently (even though $D = 10$).

Table 2.3: DID vs SID, Layout 3, 7 Devices, Flow 1

		CF-CF			L/OF-L/OF			OF-OF		
		SIM	E_C	ρ^k	SIM	E_C	ρ^k	SIM	E_C	ρ^k
Case 1	ρ	0.794 ± 0.008			0.890 ± 0.007			0.977 ± 0.012		
	Pr (DID)	0.482 ± 0.014	0.473	0.199	0.684 ± 0.017	0.694	0.442	0.932 ± 0.033	0.932	0.850
	Pr (SID)	0.518 ± 0.014	0.527	0.801	0.316 ± 0.017	0.306	0.558	0.068 ± 0.033	0.068	0.150
	W_q	27.99 ± 1.68			66.38 ± 3.61			832.3 ± 393.4		
Case 2	ρ	0.685 ± 0.008			0.793 ± 0.007			0.848 ± 0.008		
	Pr (DID)	0.301 ± 0.010	0.277	0.071	0.472 ± 0.014	0.471	0.197	0.591 ± 0.020	0.592	0.315
	Pr (SID)	0.699 ± 0.010	0.723	0.929	0.528 ± 0.014	0.529	0.803	0.409 ± 0.020	0.408	0.685
	W_q	12.69 ± 0.46			29.67 ± 1.79			66.75 ± 5.28		

For our case, however, the Erlang C formula represents an approximation since the device service times (i.e., empty + loaded travel) are non-exponential, and furthermore, the sequence of service affects the service times. (Empty travel distances depend on the sequence in which the MRs are served.) Hence, to assess the performance of the Erlang C formula as an estimate for DID versus SID, a simulation experiment was conducted with one of the layouts used in our study (Layout 3, Appendix 2.B). (The details of the simulation are provided in section 2.5.) The results are shown in Table 2.3, where W_q is the overall wait time (in secs) averaged across all the MRs. Testing three dispatching rules, we observe that the Erlang C formula provides a reasonably accurate estimate for the probability of DID versus SID over a wide range of ρ values. We also observe that for $\rho \cong 0.98$, DID is invoked over 90% of the time. However, such a large value is unlikely to be suitable in practice since the expected waiting time is excessive relative to the travel times. (Ultimately, the cost of the devices must be weighed against the expected waiting times and customer expectations.) For more realistic ρ values that fall between approximately 0.70 and 0.90, the percent breakdown between DID-SID in Table 2.3 ranges from about 30%-70% to 70%-30%, which clearly indicates that in most systems, both DID and SID would be invoked with moderate frequency, and one does not strongly dominate the other. (As anticipated, using ρ^k to estimate the probability of DID leads to significant errors as shown in Table 2.3.)

2.4.3 Analytic Model to Estimate Empty Device Travel under L/OF-L/OF

Each empty trip occurs either as a result of DID (i.e., upon delivering a load, the device remains *busy* and is immediately assigned to a MR) or SID (i.e., a MR arrives and finds one or more *idle* devices). Thus, each empty trip, e_{ij}, occurs either from a busy (B) state or idle (I) state:

$$e_{ij} = e_{ij}^B + e_{ij}^I \tag{2.4}$$

Breaking down e_{ij} into two components as shown above is a key step in developing an analytic model to estimate the empty trips in the system.

Since flow is conserved, the number of empty trips per hour into station i must equal the number of loaded trips per hour out of station i. Likewise, the number of empty trips per hour out of station i must equal the number of loaded trips per hour into station i. In other words, for every loaded trip, there is a preceding empty trip. Note that, if a device that has just delivered a load at station i finds a load waiting there (i.e., a "local load"), it will perform an empty trip, e_{ii}, from the deposit point of i to the pick-up point of i, which we assumed to be negligible in travel time.

Assuming that device arrivals occur at random points in time, we have:

$e_{i\bullet}^B = \Lambda_i(E_C)$ number of DID-based empty trips per hour out of station i (2.5a)

$e_{i\bullet}^I = \Lambda_i(1 - E_C)$ number of SID-based empty trips per hour out of station i (2.5b)

$e_{\bullet i}^B = \lambda_i(E_C)$ number of DID-based empty trips per hour into station i (2.5c)

$e_{\bullet i}^I = \lambda_i(1 - E_C)$ number of SID-based empty trips per hour into station i (2.5d)

Node-to-node empty trips under DID and SID, denoted by e_{ij}^B and e_{ij}^I, respectively, are computed separately and are described next, where m (M) denotes the number of MRs in the global queue (in the system).

2.4.3.1 Device Initiated Empty Trips

DID occurs when a device delivers a load and finds $m \geq 1$. In order to estimate e_{ij}^{B}, we consider two possible cases: e_{ii}^{B} and e_{ij}^{B} $(i \neq j)$, where the first case represents the instance where the device finds a local load. Let q_i be the probability that the device finds a local load at i, given that $m \geq 1$. Note that q_i is a departure instance probability since it occurs when the device has just delivered a load. However, for systems with only negative and positive unit jumps, the equilibrium state distribution observed by departures is the same as that observed by arrivals (see, for example, [Cooper 1981, 186]). Since the MRs arrive according to a Poisson process, and such arrivals observe time averages (Cooper 1981, p. 77; Wolff 1982), the equilibrium state distribution observed by departures is the same as that observed by time-averaging.

Let $\mathbb{P}_M$ (p_m) denote the probability that there are M (m) MRs in the system (the global queue). Determining the equilibrium, time-average probabilities for the number of MRs in our system/queue is far from straightforward and it may not be analytically tractable. Therefore, as an approximation, we use the results from the $M/M/c$ queue (where $c = D$) to estimate $\mathbb{P}_M$ and p_m. That is, given ρ and D, we have (Kleinrock 1975, p. 102):

$$\mathbb{P}_M = \begin{cases} \mathbb{P}_0 \dfrac{(\rho D)^M}{M!}, & \text{for } 1 \leq M \leq D \\ \mathbb{P}_0 \dfrac{\rho^M D^D}{D!}, & \text{for } M \geq D \end{cases} \tag{2.6}$$

where,

$$\mathbb{P}_0 = \left[\sum_{\ell=0}^{D-1} \frac{(\rho D)^{\ell}}{\ell!} + \left(\frac{(\rho D)^D}{D}\right)\left(\frac{1}{1-\rho}\right)\right]^{-1} \tag{2.7}$$

Given that $m \geq 1$, we renormalize the above steady-state probabilities to obtain:

$$p'_m = \frac{\mathbb{P}_{m+D-1}}{1 - \sum_{\ell=0}^{D-1} \mathbb{P}_\ell}, \tag{2.8}$$

where $1 - \sum_{\ell=0}^{D-1} \mathbb{P}_\ell$ is the probability that the global queue content is non-zero.

Assuming independence of the global queue contents, and assuming that a MR, selected randomly from the global queue, is at station i with probability λ_i/λ_T, given a global queue with exactly m MRs ($m \geq 1$), the probability that none of them are at station i is equal to $\left(1 - \frac{\lambda_i}{\lambda_T}\right)^m$, and the probability that at least one of them is at station i is equal to $1 - \left(1 - \frac{\lambda_i}{\lambda_T}\right)^m$. Hence,

$$q_i = \sum_{m=1}^{L} (p'_m) \left[1 - \left(1 - \frac{\lambda_i}{\lambda_T}\right)^m\right], \tag{2.9}$$

where we treat L as a sufficiently large number, and

$$e_{ii}^B = (q_i)(e_{i\bullet}^B) \qquad \text{for } i \in S \tag{2.10}$$

With $m \geq 1$, if the device does not find a local load at station i, it serves the oldest MR in the system, say, at station j, which leads to the second case, i.e., e_{ij}^B $(i \neq j)$. Given that there is no local load at station i, we have:

$$e_{ij}^B = (1 - q_i)(e_{i\bullet}^B)\left(\frac{\lambda_j}{\lambda_T - \lambda_i}\right) \qquad \text{for } i, j \in S, i \neq j \tag{2.11}$$

2.4.3.2 Station Initiated Empty Trips

SID occurs when a MR arrives at j and finds one or more idle devices, i.e., $M \leq (D-1)$. The approach we use to estimate e_{ij}^I is similar to DID except that we view the system from the MR's perspective. We consider two possible cases as before: e_{jj}^I and e_{ij}^I $(i \neq j)$, where the first case represents the instance where the MR arriving at station j finds a local idle device at j. Let r_j be the probability that the MR finds a local idle device at j, given that $M \leq (D-1)$.

Let π_d denote the probability that the MR arriving at station j finds d devices idle ($d \leq D$). (Recall that this probability is the same obtained from time-averaging). Again, using the $M/M/c$ queue as an approximation, given ρ and D, we have

$$\pi_d = \mathbb{P}_{D-d}, \qquad \text{for } 1 \leq d \leq D$$

Given that $M \leq (D-1)$, we renormalize the above steady-state probabilities to obtain:

$$\pi'_d = \frac{\pi_d}{\sum_{M=0}^{D-1} \mathbb{P}_M} \tag{2.12}$$

where $\sum_{M=0}^{D-1} \mathbb{P}_M$ is the probability that there is at least one idle device in the system.

The probability that an idle device is at station j is equal to Λ_j/Λ_T. Therefore, given d idle devices, the probability that at least one of them is at station j is given by $1 - \left(1 - \frac{\Lambda_j}{\Lambda_T}\right)^d$. Hence,

$$r_j = \sum_{d=1}^{D} (\pi'_d)\left[1 - \left(1 - \frac{\Lambda_j}{\Lambda_T}\right)^d\right], \tag{2.13}$$

and

$$e_{jj}^I = (r_j)(e_{\bullet j}^I) \qquad \text{for } j \in S \tag{2.14}$$

Consider next the second case, i.e., e_{ij}^I $(i \neq j)$. When SID occurs but there is no local idle device at j, the MR is assigned to the longest idle device, which is at station i with probability $\Lambda_i/(\Lambda_T - \Lambda_j)$. That is,

$$e_{ij}^I = (1 - r_j)(e_{\bullet j}^I)\left(\frac{\Lambda_i}{\Lambda_T - \Lambda_j}\right) \qquad \text{for } i,j \in S, i \neq j \tag{2.15}$$

2.4.3.3 Empty Trips Rescaled

Conservation of flow dictates that we have:

$$\sum_{j \in S} e_{ij}^B = e_{i\bullet}^B \quad \text{for each } i \tag{2.16a}$$

$$\sum_{i \in S} e_{ij}^B = e_{\bullet j}^B \quad \text{for each } j \tag{2.16b}$$

$$\sum_{j \in S} e_{ij}^I = e_{i\bullet}^I \quad \text{for each } i \tag{2.16c}$$

$$\sum_{i \in S} e_{ij}^I = e_{\bullet j}^I \quad \text{for each } j \tag{2.16d}$$

Since e_{ij}^B and e_{ij}^I are estimated values, they do not necessarily satisfy equation (2.16). More specifically, because of how they were derived, the e_{ij}^B values satisfy (2.16a) (i.e., the sum of row i equals $e_{i\bullet}^B$) but they may not satisfy (2.16b) (i.e., the sum of column j may not equal $e_{\bullet j}^B$). Hence, we let

$$\delta_j = \frac{e_{\bullet j}^B}{\sum_{k \in S} e_{kj}^B} \qquad \text{for each } j \in S \tag{2.17}$$

and rescale the e_{ij}^B values in column j as follows:

$$e_{ij}^B \leftarrow (\delta_j)(e_{ij}^B) \qquad \text{for each } j \in S \tag{2.18}$$

As a result of rescaling, the *relative* values of the e_{ij}^B's in column j remain the same but their absolute values are adjusted so that equation (2.16b) is satisfied. However, when each column is rescaled, and we reconsider the rows, equation (2.16a) may no longer be satisfied for each row. Therefore, we let

$$\delta_i = \frac{e_{i\bullet}^B}{\sum_{k \in S} e_{ik}^B} \qquad \text{for each } i \in S \tag{2.19}$$

and rescale the e_{ij}^B values in row i as follows:

$$e_{ij}^B \leftarrow (\ \delta_i)(e_{ij}^B) \qquad \text{for each } i \in S \tag{2.20}$$

As a result of rescaling each row, equation (2.16a) is again satisfied for each row, but equation (2.16b) may not be satisfied for each column. Hence, we again rescale the columns, and we continue in this manner, alternating between rescaling the columns and the rows until both equations (2.16a) and (2.16b) are satisfied. (The δ_i and δ_j values decrease in each iteration and eventually approach one.)

The above procedure is repeated for the e_{ij}^I values until equations (2.16c) and (2.16d) are satisfied. The only difference is that, initially, equation (2.16d) is satisfied for each column, and we start the iterative process by rescaling each row.

Finally, since we now have the estimated values for both e_{ij}^B and e_{ij}^I, we can use equations (2.1) and (2.4) to compute the values of α_e and ρ.

2.4.4 Iterative Algorithm to Compute $\boldsymbol{\rho}$

In section 2.4.3, E_C is used to estimate the values of e_{ij}^B and e_{ij}^I. However, E_C is based on a given ρ. We, therefore, employ the following iterative scheme to estimate the values of e_{ij} and ρ for a user-specified number of devices (D):

1) Set $n = 1$.

2) Compute a lower bound on ρ, and set $\rho^{(n)}$ equal to the lower bound. If $\rho^{(n)} \geq 1$, **stop**; more devices are needed.

3) Using $\rho^{(n)}$, compute $E_C^{(n)}$ from equation (2.3), and estimate the values of $e_{ij}^{B(n)}$ and $e_{ij}^{I(n)}$ using the results in section 2.4.3.

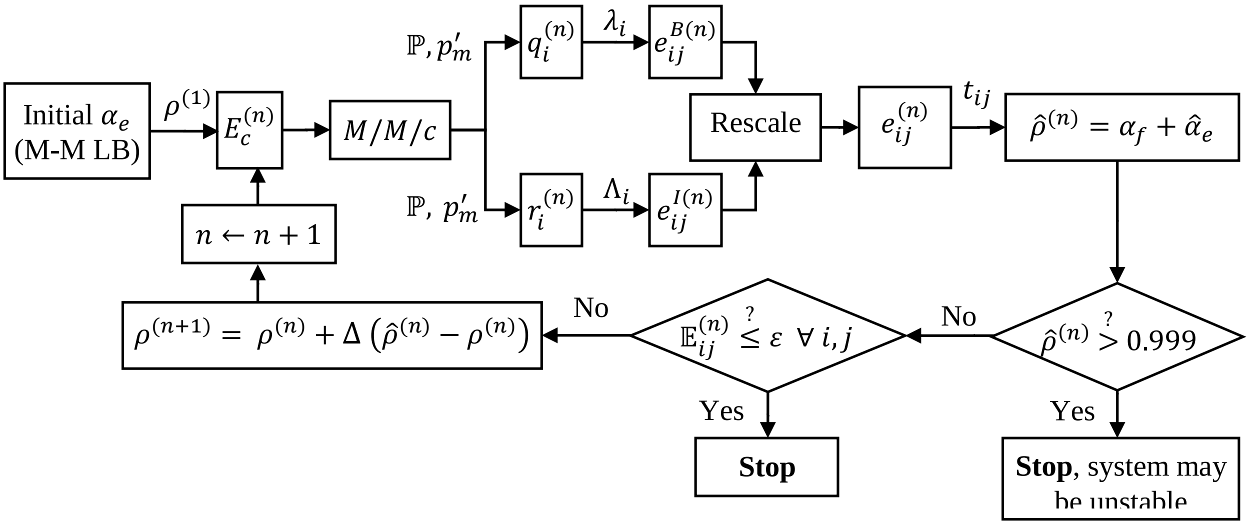
4) Set $e_{ij}^{(n)} = e_{ij}^{B(n)} + e_{ij}^{I(n)}$. Compute the new value of the expected device utilization, $\hat{\rho}^{(n)}$, using equation (2.1). If $\hat{\rho}^{(n)} > 1$, set $\hat{\rho}^{(n)} = 1$.

5) Set $\rho^{(n+1)} = \rho^{(n)} + \Delta\left(\hat{\rho}^{(n)} - \rho^{(n)}\right)$, where Δ is a sufficiently small step size.

6) If $\rho^{(n+1)}$ is approaching 1, say, $\rho^{(n+1)} > 0.999$, **stop**; the system may or may not be stable. More devices are needed to obtain a realistic ρ value.

7) Set $\mathbb{E}_{ij}^{(n)} = \left|e_{ij}^{(n)} - e_{ij}^{(n-1)}\right|$ and let ε be a sufficiently small number. If $\mathbb{E}_{ij}^{(n)} \leq \varepsilon \ \forall \, i, j$, **stop**; the algorithm has converged; the expected device utilization is equal to $\rho^{(n)}$. Otherwise, let $n \leftarrow n + 1$ and go to step 3.

The lower bound in step 2 is straightforward to compute, using the method presented by Maxwell and Muckstadt (1982). First, using equation (2.2), the net flow is computed for each station. Then, treating the stations with positive and negative net flows as supply and demand nodes, respectively, a transportation problem is solved to obtain the minimum possible e_{ij} values, which yields a lower bound on ρ. We abbreviate the lower bound as the M-M LB. The above iterative procedure is depicted in Figure 2.1.



Initial α_e (M-M LB)
$\rho^{(1)}$
$E_c^{(n)}$
$M/M/c$
$\mathbb{P}, p'_m$
$q_i^{(n)}$
λ_i
$e_{ij}^{B(n)}$
$\mathbb{P}, p'_m$
$r_i^{(n)}$
Λ_i
$e_{ij}^{I(n)}$
Rescale
$e_{ij}^{(n)}$
t_{ij}
$\hat{\rho}^{(n)} = \alpha_f + \hat{\alpha}_e$
$n \leftarrow n + 1$
$\rho^{(n+1)} = \rho^{(n)} + \Delta\left(\hat{\rho}^{(n)} - \rho^{(n)}\right)$
No
$\mathbb{E}_{ij}^{(n)} \overset{?}{\leq} \varepsilon \ \forall \, i, j$
No
$\hat{\rho}^{(n)} \overset{?}{>} 0.999$
Yes
Stop
Yes
Stop, system may be unstable

Figure 2.1: Iterative Procedure to Compute the Expected Device Utilization

2.5 Model Evaluation and Simulation Results

We next present simulation results to assess the performance of the above analytic model and to compare the performance of the L/OF-L/OF against to two other well-known dispatching rules; namely, OF-OF and CF-CF. We also compare L/OF-L/OF with L/OF-OF, the rule proposed by Srinivasan, et al. (1994) (refer to Table 2.1 for the list of dispatching rules). Our simulation model is based on the Tecnomatix Plant Simulation package (2014) by Siemens.

Three layouts, labeled LO1, LO2 and LO3, equipped with 3 or 7 devices, are used for the simulation experiment. Layout LO3 is taken from Srinivasan et al. (1994). In order to evaluate the analytic model across a range of device utilizations, the flow data are kept constant, but the device travel speed is adjusted in order to raise/lower the utilization. Since the flow data may impact the performance of the dispatching rules, the experiment is conducted with two sets of flow data: Flow 1 (nearly-balanced flows that yield small NF values for each station), and Flow 2 (unbalanced flows that yield a range of NF values for the stations). The complete data sets are shown in Appendix 2.B.

The simulation results are based on 10 replications, with 20,000 loaded trips per device per replication, following a warm-up period of 1,000 loaded trips. However, the simulation is terminated sooner if the system becomes “overloaded” (OL), that is, if the number of MRs in the global queue exceeds a pre-determined limit of ($300 \times$ the number of stations). While we cannot conclude that such systems are unstable, an excessive number of MRs in the global queue would not be acceptable in most applications. Last, the device travel speed selected for each scenario is such that L/OF-L/OF yields a target ρ value of approximately $0.90, 0.80,$ and 0.60, which corresponds to high, medium, and low device utilization, respectively. Once the device speed is selected, the same flow data are simulated with the other rules.

2.5.1 Performance of the L/OF-L/OF Rule

The simulation results comparing the performance of the rules are presented in Tables 2.4a and 2.4b for flow sets 1 and 2, respectively (refer to Table 2.1 for the rules). The estimated lower bound (M-M LB) is also included in the comparison. Since the α_f values do not change, the rules are compared on the basis of empty device travel (α_e) and the expected MR wait time $\left(W_q\right)$.

In general, L/OF-L/OF performs better than both OF-OF and L/OF-OF, that is, the rule has lower W_q, α_e and ρ values. Also, as expected, the performance gap between L/OF-L/OF and OF-OF increases when the flow is more balanced. However, the above gap is also impacted by the number of devices and their utilization. A larger number of devices improves the effectiveness of SID (since a MR is more likely to find an idle device that is local), and lower device utilization increases the proportion of SID. On the other hand, if there are a fewer number of devices, and the proportion of SID is large (due to low device utilization), L/OF-L/OF yields results comparable to OF-OF. Similarly, the performance gap between L/OF-L/OF and L/OF-OF is impacted by the number of devices and their utilization. The L/OF-L/OF and L/OF-OF results are more comparable when the proportion of DID is larger (higher utilization, and/or lower number of devices), and the performance gap increases as the proportion of SID increases (lower utilization and/or more devices).

Compared to CF-CF, L/OF-L/OF performs relatively well if the device utilization is high. This is because a device is more likely to find a local load (provided the flow is not highly unbalanced). Additionally, if there are fewer devices in the system, the MR is less likely to find a local idle device. Therefore, the performance gap between CF-CF and L/OF-L/OF is larger in systems with lower utilization and/or fewer devices. Overall, the results indicate that L/OF-L/OF is a reasonably efficient rule with a performance that stands between OF-OF and CF-CF. Figure 2.2a and 2.2b illustrate the α_e comparison at $\rho = 0.80$.

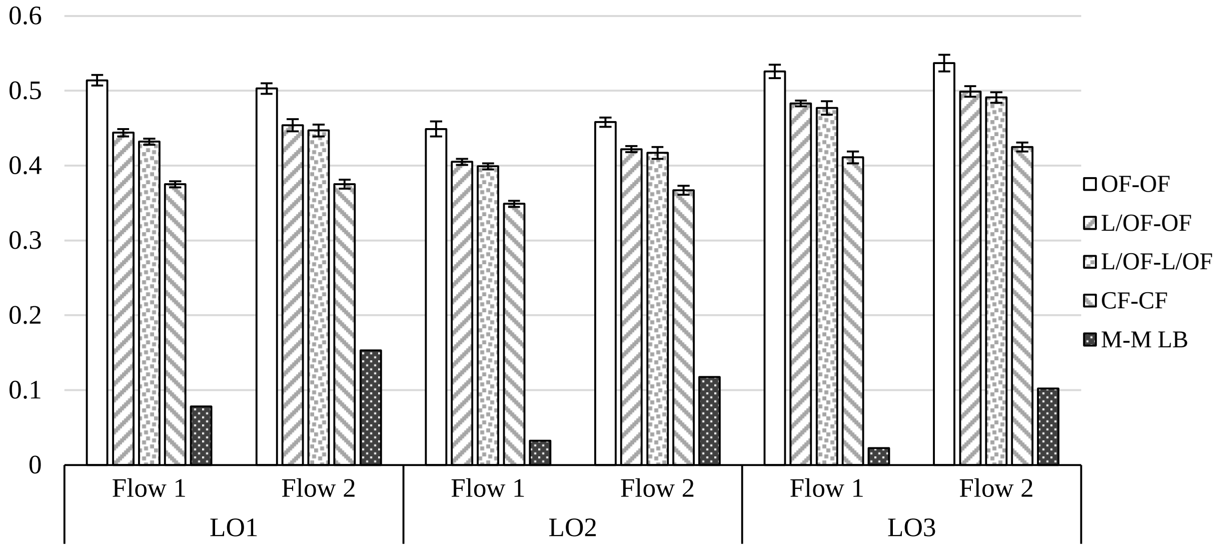


Figure 2.2a: Simulated α_e Values for Alternative Dispatching Rules at 3 Devices, $\rho = 0.80$

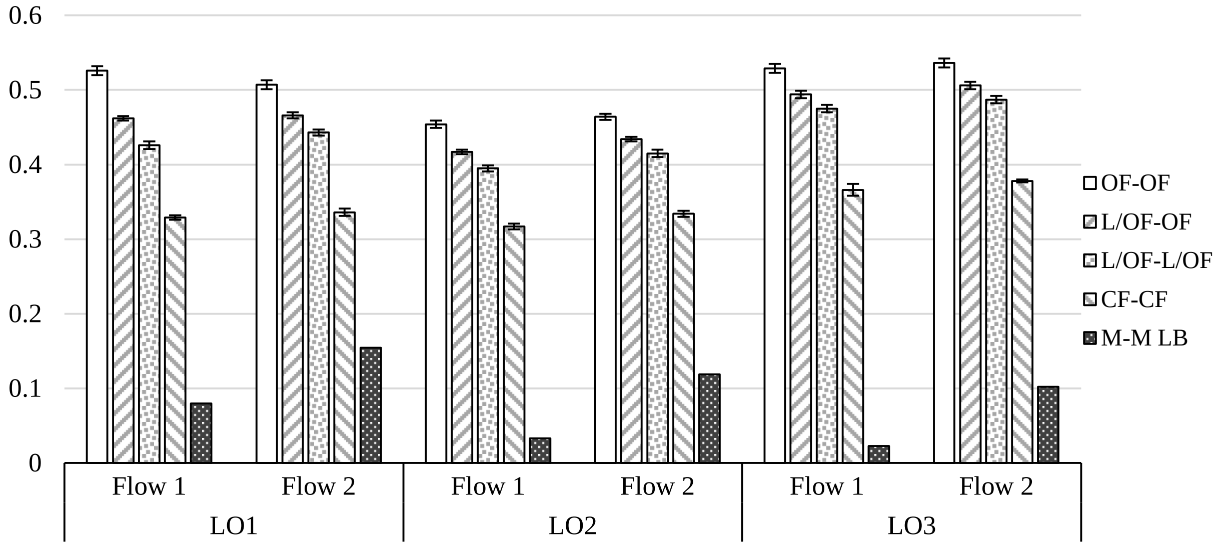


Figure 2.2b: Simulated α_e Values for Alternative Dispatching Rules at 7 Devices, $\rho = 0.80$

Table 2.4a: Simulation Comparison of the Rules with 3 Devices

			Flow 1			Flow 2		
		Target ρ =	**0.90**	**0.80**	**0.60**	**0.90**	**0.80**	**0.60**
LO1	α_e	**OF-OF**	OL	0.514 ± 0.007	0.373 ± 0.005	OL	0.526 ± 0.006	0.398 ± 0.005
		L/OF-OF	0.466 ± 0.006	0.444 ± 0.005	0.355 ± 0.005	0.477 ± 0.002	0.462 ± 0.003	0.385 ± 0.005
		L/OF-L/OF	0.460 ± 0.003	0.432 ± 0.004	0.336 ± 0.005	0.460 ± 0.004	0.426 ± 0.005	0.319 ± 0.004
		CF-CF	0.405 ± 0.004	0.375 ± 0.004	0.282 ± 0.003	0.382 ± 0.004	0.329 ± 0.003	0.215 ± 0.005
		M-M LB	0.094	0.078	0.057	0.093	0.080	0.060
	α_f	**L/OF-L/OF**	0.442 ± 0.006	0.369 ± 0.007	0.267 ± 0.005	0.441 ± 0.004	0.376 ± 0.004	0.285 ± 0.003
	ρ	**OF-OF**	OL	0.881 ± 0.010	0.640 ± 0.007	OL	0.902 ± 0.008	0.683 ± 0.006
		L/OF-OF	0.906 ± 0.007	0.812 ± 0.005	0.623 ± 0.006	0.918 ± 0.004	0.838 ± 0.005	0.670 ± 0.007
		L/OF-L/OF	0.901 ± 0.005	0.801 ± 0.005	0.603 ± 0.006	0.901 ± 0.004	0.803 ± 0.005	0.604 ± 0.006
		CF-CF	0.847 ± 0.006	0.743 ± 0.007	0.550 ± 0.004	0.823 ± 0.006	0.705 ± 0.006	0.500 ± 0.007
	W_q	**OF-OF**	OL	249.7 ± 15.2	31.1 ± 1.5	OL	288.1 ± 16.8	25.1 ± 0.6
		L/OF-OF	161.3 ± 3.1	78.7 ± 1.9	22.4 ± 0.6	147.6 ± 2.7	71.6 ± 1.0	17.0 ± 0.4
		L/OF-L/OF	158.1 ± 3.6	77.2 ± 1.2	21.3 ± 0.5	140.7 ± 3.0	64.0 ± 1.3	12.8 ± 0.4
		CF-CF	92.6 ± 2.3	48.4 ± 1.6	14.8 ± 0.3	72.6 ± 1.5	31.7 ± 0.7	5.5 ± 0.2
LO2	α_e	**OF-OF**	0.526 ± 0.007	0.449 ± 0.010	0.327 ± 0.004	OL	0.454 ± 0.005	0.341 ± 0.004
		L/OF-OF	0.430 ± 0.005	0.405 ± 0.004	0.318 ± 0.005	0.440 ± 0.005	0.417 ± 0.003	0.336 ± 0.003
		L/OF-L/OF	0.427 ± 0.004	0.399 ± 0.004	0.307 ± 0.006	0.429 ± 0.005	0.395 ± 0.004	0.295 ± 0.004
		CF-CF	0.376 ± 0.005	0.349 ± 0.004	0.266 ± 0.004	0.363 ± 0.005	0.317 ± 0.004	0.216 ± 0.004
		M-M LB	0.038	0.032	0.024	0.038	0.033	0.025
	α_f	**L/OF-L/OF**	0.470 ± 0.008	0.399 ± 0.005	0.292 ± 0.003	0.472 ± 0.005	0.405 ± 0.004	0.304 ± 0.005
	ρ	**OF-OF**	0.995 ± 0.012	0.849 ± 0.013	0.619 ± 0.008	OL	0.860 ± 0.008	0.645 ± 0.006
		L/OF-OF	0.900 ± 0.008	0.805 ± 0.009	0.609 ± 0.006	0.912 ± 0.006	0.823 ± 0.007	0.639 ± 0.006
		L/OF-L/OF	0.897 ± 0.009	0.798 ± 0.007	0.598 ± 0.007	0.902 ± 0.007	0.800 ± 0.007	0.599 ± 0.008
		CF-CF	0.847 ± 0.008	0.748 ± 0.007	0.558 ± 0.006	0.835 ± 0.008	0.723 ± 0.008	0.519 ± 0.007
	W_q	**OF-OF**	3638 ± 1497	95.2 ± 7.4	14.7 ± 0.4	OL	90.9 ± 4.6	9.4 ± 0.2
		L/OF-OF	95.4 ± 3.1	46.5 ± 1.3	11.9 ± 0.4	89.3 ± 1.8	39.1 ± 0.7	7.6 ± 0.2
		L/OF-L/OF	93.7 ± 3.7	45.4 ± 1.4	11.7 ± 0.4	86.9 ± 1.6	36.6 ± 0.9	6.4 ± 0.3
		CF-CF	54.9 ± 1.4	29.0 ± 0.6	8.6 ± 0.2	45.4 ± 1.1	19.6 ± 0.3	3.3 ± 0.1
LO3	α_e	**OF-OF**	0.617 ± 0.006	0.526 ± 0.009	0.384 ± 0.007	0.617 ± 0.005	0.529 ± 0.006	0.396 ± 0.003
		L/OF-OF	0.522 ± 0.003	0.483 ± 0.004	0.376 ± 0.005	0.531 ± 0.003	0.494 ± 0.005	0.391 ± 0.003
		L/OF-L/OF	0.521 ± 0.004	0.477 ± 0.009	0.365 ± 0.006	0.522 ± 0.004	0.475 ± 0.005	0.354 ± 0.004
		CF-CF	0.450 ± 0.006	0.411 ± 0.008	0.311 ± 0.007	0.429 ± 0.004	0.366 ± 0.008	0.240 ± 0.004
		M-M LB	0.027	0.023	0.017	0.027	0.023	0.017
	α_f	**L/OF-L/OF**	0.380 ± 0.007	0.323 ± 0.006	0.237 ± 0.005	0.380 ± 0.004	0.326 ± 0.003	0.245 ± 0.003
	ρ	**OF-OF**	0.996 ± 0.012	0.849 ± 0.013	0.621 ± 0.010	0.997 ± 0.008	0.856 ± 0.008	0.640 ± 0.006
		L/OF-OF	0.903 ± 0.007	0.805 ± 0.006	0.612 ± 0.006	0.912 ± 0.005	0.821 ± 0.007	0.636 ± 0.005
		L/OF-L/OF	0.901 ± 0.010	0.799 ± 0.013	0.602 ± 0.009	0.902 ± 0.006	0.801 ± 0.008	0.599 ± 0.006
		CF-CF	0.831 ± 0.011	0.734 ± 0.013	0.548 ± 0.011	0.808 ± 0.007	0.693 ± 0.011	0.485 ± 0.006
	W_q	**OF-OF**	5388 ± 4107	84.0 ± 7.7	12.7 ± 0.4	5972 ± 3469	72.6 ± 3.6	7.8 ± 0.2
		L/OF-OF	84.1 ± 2.4	40.6 ± 1.1	10.7 ± 0.2	77.1 ± 1.8	33.5 ± 0.6	6.4 ± 0.2
		L/OF-L/OF	83.2 ± 3.0	39.9 ± 1.0	10.5 ± 0.4	74.9 ± 1.2	31.5 ± 1.0	5.4 ± 0.1
		CF-CF	40.3 ± 1.2	22.2 ± 0.6	6.9 ± 0.2	30.8 ± 0.6	13.5 ± 0.4	2.1 ± 0.1
		DI-SI	O = Oldest	L = Local	C = Closest	F = First		

Table 2.4b: Simulation Comparison of the Rules with 7 Devices

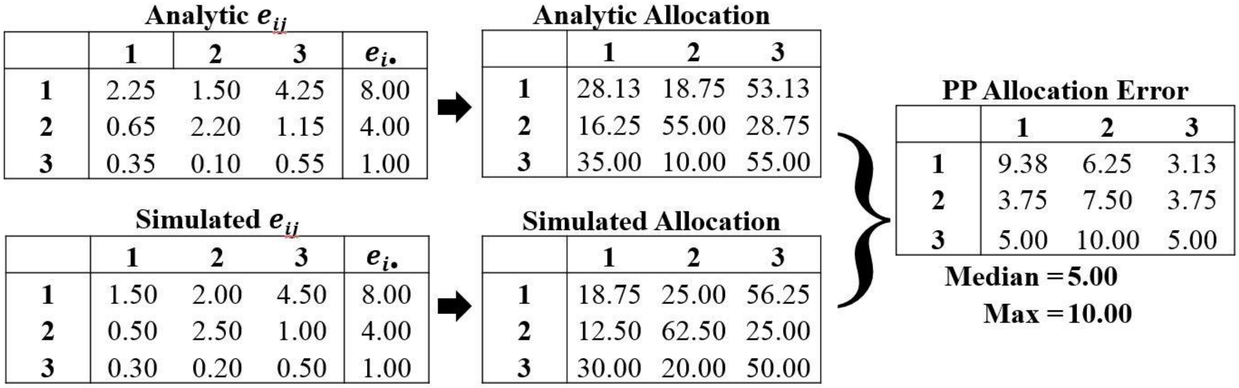
			Flow 1			**Flow 2**		
		Target ρ =	**0.90**	**0.80**	**0.60**	**0.90**	**0.80**	**0.60**
LO1	α_e	**OF-OF**	OL	0.526 ± 0.006	0.398 ± 0.005	OL	0.507 ± 0.006	0.382 ± 0.004
		L/OF-OF	0.477 ± 0.002	0.462 ± 0.003	0.385 ± 0.005	0.494 ± 0.003	0.466 ± 0.004	0.375 ± 0.003
		L/OF-L/OF	0.460 ± 0.004	0.426 ± 0.005	0.319 ± 0.004	0.482 ± 0.004	0.443 ± 0.004	0.333 ± 0.004
		CF-CF	0.382 ± 0.004	0.329 ± 0.003	0.215 ± 0.005	0.386 ± 0.004	0.336 ± 0.005	0.226 ± 0.004
		M-M LB	0.093	0.080	0.060	0.178	0.154	0.116
	α_f	**L/OF-L/OF**	0.441 ± 0.004	0.376 ± 0.004	0.285 ± 0.003	0.420 ± 0.005	0.363 ± 0.003	0.273 ± 0.002
	ρ	**OF-OF**	OL	0.902 ± 0.008	0.683 ± 0.006	OL	0.871 ± 0.007	0.655 ± 0.006
		L/OF-OF	0.918 ± 0.004	0.838 ± 0.005	0.670 ± 0.007	0.914 ± 0.005	0.830 ± 0.006	0.648 ± 0.005
		L/OF-L/OF	0.901 ± 0.004	0.803 ± 0.005	0.604 ± 0.006	0.902 ± 0.006	0.806 ± 0.006	0.607 ± 0.005
		CF-CF	0.823 ± 0.006	0.705 ± 0.006	0.500 ± 0.007	0.805 ± 0.006	0.699 ± 0.008	0.500 ± 0.007
	W_q	**OF-OF**	OL	288.1 ± 16.8	25.1 ± 0.6	OL	180.7 ± 10.3	18.9 ± 0.4
		L/OF-OF	147.6 ± 2.7	71.6 ± 1.0	17.0 ± 0.4	176.2 ± 4.1	77.1 ± 1.8	15.0 ± 0.5
		L/OF-L/OF	140.7 ± 3.0	64.0 ± 1.3	12.8 ± 0.4	170.3 ± 6.3	70.5 ± 2.1	12.3 ± 0.4
		CF-CF	72.6 ± 1.5	31.7 ± 0.7	5.5 ± 0.2	63.2 ± 1.2	29.5 ± 0.8	4.8 ± 0.2
LO2	α_e	**OF-OF**	OL	0.454 ± 0.005	0.341 ± 0.004	0.532 ± 0.006	0.464 ± 0.004	0.348 ± 0.004
		L/OF-OF	0.440 ± 0.005	0.417 ± 0.003	0.336 ± 0.003	0.463 ± 0.003	0.434 ± 0.003	0.342 ± 0.005
		L/OF-L/OF	0.429 ± 0.005	0.395 ± 0.004	0.295 ± 0.004	0.454 ± 0.004	0.415 ± 0.005	0.311 ± 0.003
		CF-CF	0.363 ± 0.005	0.317 ± 0.004	0.216 ± 0.004	0.383 ± 0.005	0.334 ± 0.004	0.225 ± 0.003
		M-M LB	0.038	0.033	0.025	0.137	0.119	0.089
	α_f	**L/OF-L/OF**	0.472 ± 0.005	0.405 ± 0.004	0.304 ± 0.005	0.443 ± 0.006	0.386 ± 0.005	0.290 ± 0.005
	ρ	**OF-OF**	OL	0.860 ± 0.008	0.645 ± 0.006	0.976 ± 0.012	0.849 ± 0.009	0.637 ± 0.007
		L/OF-OF	0.912 ± 0.006	0.823 ± 0.007	0.639 ± 0.006	0.906 ± 0.007	0.820 ± 0.007	0.632 ± 0.007
		L/OF-L/OF	0.902 ± 0.007	0.800 ± 0.007	0.599 ± 0.008	0.897 ± 0.008	0.801 ± 0.009	0.600 ± 0.008
		CF-CF	0.835 ± 0.008	0.723 ± 0.008	0.519 ± 0.007	0.826 ± 0.007	0.719 ± 0.007	0.514 ± 0.006
	W_q	**OF-OF**	OL	90.9 ± 4.6	9.4 ± 0.2	836.3 ± 202.7	73.2 ± 3.2	8.1 ± 0.4
		L/OF-OF	89.3 ± 1.8	39.1 ± 0.7	7.6 ± 0.2	85.5 ± 2.7	37.7 ± 1.1	6.8 ± 0.1
		L/OF-L/OF	86.9 ± 1.6	36.6 ± 0.9	6.4 ± 0.3	83.6 ± 1.7	35.5 ± 1.2	5.9 ± 0.2
		CF-CF	45.4 ± 1.1	19.6 ± 0.3	3.3 ± 0.1	40.6 ± 0.9	17.7 ± 0.4	2.8 ± 0.1
LO3	α_e	**OF-OF**	0.617 ± 0.005	0.529 ± 0.006	0.396 ± 0.003	0.626 ± 0.008	0.536 ± 0.006	0.402 ± 0.004
		L/OF-OF	0.531 ± 0.003	0.494 ± 0.005	0.391 ± 0.003	0.550 ± 0.004	0.506 ± 0.005	0.398 ± 0.005
		L/OF-L/OF	0.522 ± 0.004	0.475 ± 0.005	0.354 ± 0.004	0.541 ± 0.005	0.487 ± 0.005	0.366 ± 0.003
		CF-CF	0.429 ± 0.004	0.366 ± 0.008	0.240 ± 0.004	0.445 ± 0.004	0.378 ± 0.002	0.254 ± 0.002
		M-M LB	0.027	0.023	0.017	0.119	0.102	0.077
	α_f	**L/OF-L/OF**	0.380 ± 0.004	0.326 ± 0.003	0.245 ± 0.003	0.359 ± 0.003	0.308 ± 0.004	0.231 ± 0.003
	ρ	**OF-OF**	0.997 ± 0.008	0.856 ± 0.008	0.640 ± 0.006	0.985 ± 0.011	0.845 ± 0.008	0.633 ± 0.006
		L/OF-OF	0.912 ± 0.005	0.821 ± 0.007	0.636 ± 0.005	0.910 ± 0.006	0.815 ± 0.007	0.629 ± 0.006
		L/OF-L/OF	0.902 ± 0.006	0.801 ± 0.008	0.599 ± 0.006	0.901 ± 0.007	0.795 ± 0.008	0.597 ± 0.006
		CF-CF	0.808 ± 0.007	0.693 ± 0.011	0.485 ± 0.006	0.804 ± 0.007	0.686 ± 0.005	0.485 ± 0.004
	W_q	**OF-OF**	5972 ± 3469	72.6 ± 3.6	7.8 ± 0.2	1295 ± 432	66.7 ± 3.0	7.2 ± 0.2
		L/OF-OF	77.1 ± 1.8	33.5 ± 0.6	6.4 ± 0.2	79.3 ± 2.4	32.7 ± 0.9	6.0 ± 0.2
		L/OF-L/OF	74.9 ± 1.2	31.5 ± 1.0	5.4 ± 0.1	76.5 ± 1.8	30.8 ± 1.0	5.1 ± 0.1
		CF-CF	30.8 ± 0.6	13.5 ± 0.4	2.1 ± 0.1	30.1 ± 0.6	12.9 ± 0.4	2.0 ± 0.1
		DI-SI	O = Oldest	L = Local	C = Closest	F = First		

2.5.2 L/OF-L/OF Analytic Model Evaluation

The same target ρ values of $0.90, 0.80$ and 0.60 are used in assessing the performance of the L/OF-L/OF analytic model, where the analytic estimates of α_e and ρ are compared with the simulation results. (The analytic α_f values are straightforward to obtain.) The percentage errors are reported relative to the simulation results.

Since the accuracy of the analytic model depends largely on how well it estimates e_{ij}^{B} and e_{ij}^{I}, we will compare their values with the empty trips obtained from simulation. However, since some of the $e_{ij}^{B/I}$ values can be very small, reporting the difference as a percentage error would give misleading results. Likewise, comparing their absolute differences may also be misleading since some of $e_{ij}^{B/I}$ values are large. Furthermore, due to conservation of flow, and the rescaling we perform in section 2.4.3.3, the analytic model has no errors in determining the sum of the empty trips out of each row. Rather, errors may occur in how the empty trips in a given row are allocated across the stations.

Hence, we compare the analytic and simulation results by computing the absolute error in how the empty trips are *allocated* within each row on a percentage basis. To do so, for a given row i, we first compute the percentage allocation of the empty trips across each station j; that is, we compute $\left(e_{ij}/e_{i\bullet}\right)(100)$, and we compare it, using absolute values, to the percentage allocation obtained from simulation. Once all the allocation errors are computed, we report the median and the maximum error. A simple 3-station example is shown in Figure 2.3, where “PP” stands for percentage point. Note that 18.75% (25%) of the empty trips out of station 1 go to station 2 according to the analytic (simulation) model, which yields an absolute error of 6.25 PPs. To assess the analytic model, the PP allocation error is computed for all the empty trips, i.e., e_{ij}^{B}, e_{ij}^{I}, and e_{ij}.



Analytic e_{ij}

	1	2	3	$e_{i\bullet}$
1	2.25	1.50	4.25	8.00
2	0.65	2.20	1.15	4.00
3	0.35	0.10	0.55	1.00

Analytic Allocation

	1	2	3
1	28.13	18.75	53.13
2	16.25	55.00	28.75
3	35.00	10.00	55.00

Simulated e_{ij}

	1	2	3	$e_{i\bullet}$
1	1.50	2.00	4.50	8.00
2	0.50	2.50	1.00	4.00
3	0.30	0.20	0.50	1.00

Simulated Allocation

	1	2	3
1	18.75	25.00	56.25
2	12.50	62.50	25.00
3	30.00	20.00	50.00

PP Allocation Error

	1	2	3
1	9.38	6.25	3.13
2	3.75	7.50	3.75
3	5.00	10.00	5.00

Median = 5.00

Max = 10.00

Figure 2.3: A 3-Station Example to Show How the Allocation Error is Computed

Table 2.5a: Analytic Model Evaluation with 3 Devices

3 Devices		Flow 1							Flow 2						
	Target	Results				PP Allocation Error			Results				PP Allocation Error		
	ρ		ANA	95% CI SIM	Error %		Med	Max		ANA	95% CI SIM	Error %		Med	Max
LO1	0.90	α_e	0.427	(0.457, 0.463)	7.120	e_{ij}^B	0.90	10.78	α_e	0.465	(0.477, 0.489)	3.603	e_{ij}^B	0.48	9.12
		α_f	0.442	(0.436, 0.448)	-	e_{ij}^I	0.24	1.15	α_f	0.424	(0.418, 0.432)	-	e_{ij}^I	0.16	1.72
		ρ	0.869	(0.896, 0.906)	3.566	e_{ij}	0.59	7.28	ρ	0.890	(0.899, 0.917)	2.020	e_{ij}	0.34	5.95
	0.80	α_e	0.417	(0.428, 0.436)	3.597	e_{ij}^B	0.60	6.87	α_e	0.438	(0.439, 0.455)	2.087	e_{ij}^B	0.36	5.74
		α_f	0.369	(0.362, 0.376)	-	e_{ij}^I	0.13	0.65	α_f	0.360	(0.356, 0.366)	-	e_{ij}^I	0.13	1.26
		ρ	0.786	(0.796, 0.806)	1.871	e_{ij}	0.36	4.27	ρ	0.798	(0.796, 0.820)	1.219	e_{ij}	0.25	3.48
	0.60	α_e	0.335	(0.331, 0.341)	0.362	e_{ij}^B	0.18	1.90	α_e	0.340	(0.335, 0.347)	0.332	e_{ij}^B	0.20	2.09
		α_f	0.267	(0.262, 0.272)	-	e_{ij}^I	0.10	0.45	α_f	0.261	(0.258, 0.266)	-	e_{ij}^I	0.12	1.29
		ρ	0.602	(0.597, 0.609)	0.179	e_{ij}	0.09	0.74	ρ	0.601	(0.595, 0.611)	0.337	e_{ij}	0.10	0.79
LO2	0.90	α_e	0.400	(0.423, 0.431)	6.464	e_{ij}^B	0.54	9.51	α_e	0.433	(0.448, 0.458)	4.464	e_{ij}^B	0.45	11.67
		α_f	0.470	(0.462, 0.478)	-	e_{ij}^I	0.19	1.12	α_f	0.442	(0.434, 0.450)	-	e_{ij}^I	0.20	1.92
		ρ	0.870	(0.888, 0.906)	2.997	e_{ij}	0.37	6.79	ρ	0.874	(0.887, 0.903)	2.237	e_{ij}	0.31	8.46
	0.80	α_e	0.386	(0.395, 0.403)	3.143	e_{ij}^B	0.32	5.09	α_e	0.408	(0.409, 0.425)	1.987	e_{ij}^B	0.25	6.66
		α_f	0.400	(0.394, 0.404)	-	e_{ij}^I	0.15	0.69	α_f	0.381	(0.375, 0.387)	-	e_{ij}^I	0.13	1.52
		ρ	0.786	(0.791, 0.805)	1.491	e_{ij}	0.19	3.19	ρ	0.790	(0.786, 0.810)	1.059	e_{ij}	0.16	3.63
	0.60	α_e	0.305	(0.301, 0.313)	0.601	e_{ij}^B	0.17	1.75	α_e	0.319	(0.316, 0.326)	0.626	e_{ij}^B	0.16	3.28
		α_f	0.291	(0.289, 0.295)	-	e_{ij}^I	0.10	0.62	α_f	0.280	(0.276, 0.284)	-	e_{ij}^I	0.10	1.16
		ρ	0.596	(0.591, 0.605)	0.367	e_{ij}	0.10	0.74	ρ	0.599	(0.593, 0.609)	0.335	e_{ij}	0.09	1.68
LO3	0.90	α_e	0.493	(0.517, 0.525)	5.392	e_{ij}^B	0.35	7.97	α_e	0.518	(0.532, 0.544)	3.633	e_{ij}^B	0.28	7.42
		α_f	0.380	(0.373, 0.387)	-	e_{ij}^I	0.20	0.89	α_f	0.360	(0.354, 0.366)	-	e_{ij}^I	0.16	1.00
		ρ	0.873	(0.891, 0.911)	3.068	e_{ij}	0.24	5.61	ρ	0.878	(0.891, 0.905)	2.254	e_{ij}	0.19	5.02
	0.80	α_e	0.466	(0.468, 0.486)	2.268	e_{ij}^B	0.19	4.25	α_e	0.485	(0.484, 0.498)	1.312	e_{ij}^B	0.14	4.16
		α_f	0.323	(0.317, 0.329)	-	e_{ij}^I	0.14	0.93	α_f	0.308	(0.304, 0.314)	-	e_{ij}^I	0.12	0.90
		ρ	0.789	(0.786, 0.812)	1.264	e_{ij}	0.13	2.70	ρ	0.793	(0.790, 0.810)	0.851	e_{ij}	0.11	2.64
	0.60	α_e	0.364	(0.359, 0.371)	0.243	e_{ij}^B	0.16	1.57	α_e	0.373	(0.367, 0.379)	0.149	e_{ij}^B	0.14	1.70
		α_f	0.237	(0.232, 0.242)	-	e_{ij}^I	0.10	0.59	α_f	0.224	(0.220, 0.228)	-	e_{ij}^I	0.09	0.66
		ρ	0.601	(0.593, 0.611)	0.109	e_{ij}	0.08	0.54	ρ	0.597	(0.589, 0.605)	0.061	e_{ij}	0.08	0.56

Table 2.5b: Analytic Model Evaluation with 7 Devices

7 Devices		**Flow 1**							**Flow 2**						
	Target	**Results**				**PP Allocation Error**			**Results**				**PP Allocation Error**		
	ρ		**ANA**	**95% CI SIM**	**Error %**		**Med**	**Max**		**ANA**	**95% CI SIM**	**Error %**		**Med**	**Max**
LO1	**0.90**	α_e	0.430	(0.456, 0.464)	6.606	e_{ij}^B	0.93	10.61	α_e	0.467	(0.478, 0.486)	3.161	e_{ij}^B	0.53	9.31
		α_f	0.440	(0.437, 0.445)	-	e_{ij}^I	0.18	1.66	α_f	0.420	(0.415, 0.425)	-	e_{ij}^I	0.10	0.89
		ρ	0.870	(0.897, 0.905)	3.399	e_{ij}	0.53	6.17	ρ	0.886	(0.896, 0.908)	1.743	e_{ij}	0.32	5.05
	0.80	α_e	0.415	(0.421, 0.431)	2.681	e_{ij}^B	0.51	6.07	α_e	0.437	(0.439, 0.447)	1.367	e_{ij}^B	0.38	6.89
		α_f	0.377	(0.372, 0.380)	-	e_{ij}^I	0.10	0.85	α_f	0.364	(0.360, 0.366)	-	e_{ij}^I	0.21	2.47
		ρ	0.791	(0.798, 0.808)	1.397	e_{ij}	0.27	3.04	ρ	0.801	(0.800, 0.812)	0.687	e_{ij}	0.38	8.49
	0.60	α_e	0.320	(0.315, 0.323)	0.188	e_{ij}^B	0.10	0.59	α_e	0.334	(0.329, 0.337)	0.202	e_{ij}^B	0.15	1.51
		α_f	0.285	(0.282, 0.288)	-	e_{ij}^I	0.09	0.61	α_f	0.274	(0.271, 0.275)	-	e_{ij}^I	0.08	3.38
		ρ	0.605	(0.598, 0.610)	0.163	e_{ij}	0.08	0.53	ρ	0.607	(0.602, 0.612)	0.141	e_{ij}	0.07	2.35
LO2	**0.90**	α_e	0.403	(0.424, 0.434)	6.143	e_{ij}^B	0.57	9.37	α_e	0.436	(0.450, 0.458)	4.040	e_{ij}^B	0.46	11.41
		α_f	0.473	(0.467, 0.477)	-	e_{ij}^I	0.12	1.16	α_f	0.443	(0.437, 0.449)	-	e_{ij}^I	0.11	1.09
		ρ	0.876	(0.895, 0.909)	2.841	e_{ij}	0.33	5.81	ρ	0.879	(0.889, 0.905)	2.050	e_{ij}	0.29	6.65
	0.80	α_e	0.386	(0.391, 0.399)	2.223	e_{ij}^B	0.29	4.53	α_e	0.409	(0.410, 0.420)	1.461	e_{ij}^B	0.26	5.21
		α_f	0.406	(0.401, 0.409)	-	e_{ij}^I	0.07	0.44	α_f	0.386	(0.381, 0.391)	-	e_{ij}^I	0.06	0.56
		ρ	0.791	(0.793, 0.807)	1.061	e_{ij}	0.13	2.12	ρ	0.795	(0.792, 0.810)	0.766	e_{ij}	0.11	2.44
	0.60	α_e	0.295	(0.291, 0.299)	0.043	e_{ij}^B	0.15	1.14	α_e	0.311	(0.308, 0.314)	0.176	e_{ij}^B	0.13	2.74
		α_f	0.304	(0.299, 0.309)	-	e_{ij}^I	0.06	0.46	α_f	0.289	(0.285, 0.295)	-	e_{ij}^I	0.07	1.58
		ρ	0.599	(0.591, 0.607)	0.063	e_{ij}	0.05	0.34	ρ	0.600	(0.592, 0.608)	0.059	e_{ij}	0.06	0.93
LO3	**0.90**	α_e	0.496	(0.518, 0.526)	4.879	e_{ij}^B	0.34	7.94	α_e	0.521	(0.536, 0.546)	3.684	e_{ij}^B	0.29	8.60
		α_f	0.380	(0.376, 0.384)	-	e_{ij}^I	0.11	0.87	α_f	0.360	(0.356, 0.362)	-	e_{ij}^I	0.08	1.01
		ρ	0.877	(0.896, 0.908)	2.794	e_{ij}	0.20	4.75	ρ	0.881	(0.894, 0.908)	2.174	e_{ij}	0.17	5.07
	0.80	α_e	0.467	(0.470, 0.480)	1.747	e_{ij}^B	0.18	4.15	α_e	0.483	(0.482, 0.492)	0.996	e_{ij}^B	0.14	4.11
		α_f	0.327	(0.323, 0.329)	-	e_{ij}^I	0.07	0.45	α_f	0.308	(0.304, 0.312)	-	e_{ij}^I	0.07	0.68
		ρ	0.793	(0.793, 0.809)	1.021	e_{ij}	0.09	2.11	ρ	0.791	(0.787, 0.803)	0.567	e_{ij}	0.07	1.79
	0.60	α_e	0.355	(0.350, 0.358)	0.218	e_{ij}^B	0.12	0.95	α_e	0.366	(0.363, 0.369)	0.062	e_{ij}^B	0.11	1.19
		α_f	0.244	(0.242, 0.248)	-	e_{ij}^I	0.06	0.34	α_f	0.231	(0.228, 0.234)	-	e_{ij}^I	0.05	0.89
		ρ	0.599	(0.593, 0.605)	0.030	e_{ij}	0.05	0.35	ρ	0.597	(0.591, 0.603)	0.038	e_{ij}	0.05	0.62

Tables 2.5a and 2.5b show the results obtained from the analytic versus simulation model for flow sets 1 and 2, respectively. The PP allocation error for e_{ij}^B, e_{ij}^I, and e_{ij} are also shown. Overall, the analytic model performs quite well at $\rho = 0.80$ and 0.60, with small percentage errors for α_e and small median values for the allocation error. For $\rho = 0.90$, although the above errors increase slightly, the maximum error in α_e is less than about 7%, and the median allocation error is consistently less than 1 PP.

A closer look at Tables 2.5a and 2.5b indicate that the allocation error for DID $\left(e_{ij}^{B}\right)$ is higher than that of SID $\left(e_{ij}^{I}\right)$. To further investigate this difference, Tables 2.6a and 2.6b present examples of the PP allocation error for LO1 at $\rho = 0.90$, with 3 devices and 7 devices, respectively. However, instead of absolute errors, the positive (negative) values represent overestimation (underestimation) by the analytic model. The diagonal values in Tables 2.6a and 2.6b indicate that the model overestimates the probability of the device finding a local MR, and consequently it underestimates α_e. Therefore, a larger proportion of DID (which occurs at large ρ values) implies a larger error in estimating ρ.

Table 2.6a: PP Allocation Error for DID - LO1, 3 Devices, $\rho = 0.90$, Flow Set 1

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	7.008	-0.791	-0.828	-0.547	-0.674	-0.556	-0.888	-1.544	-0.363	-0.818
2	-1.026	7.172	-0.841	-0.561	-0.823	-0.619	-1.042	-1.171	-0.310	-0.779
3	-1.322	-0.898	9.927	-0.996	-0.707	-1.077	-1.281	-1.892	-0.584	-1.169
4	-1.510	-1.231	-1.107	10.779	-0.978	-0.857	-1.535	-1.906	-0.458	-1.196
5	-0.628	-0.613	-0.620	-0.280	5.952	-0.601	-1.020	-1.083	-0.353	-0.754
6	-1.613	-1.135	-1.455	-1.153	-0.839	10.234	-1.232	-1.466	-0.491	-0.852
7	-1.042	-0.638	-1.142	-0.696	-0.813	-0.456	6.723	-0.837	-0.322	-0.778
8	-1.213	-0.866	-1.140	-0.728	-0.667	-0.947	-0.904	7.443	-0.260	-0.720
9	-1.464	-1.128	-1.211	-0.901	-0.716	-0.907	-1.461	-0.994	9.508	-0.726
10	-0.944	-0.604	-1.000	-0.593	-0.552	-1.182	-0.755	-0.748	-0.419	6.797

Table 2.6b: PP Allocation Error for DID - LO1, 7 Devices, $\rho = 0.90$, Flow Set 1

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	7.150	-0.716	-0.820	-0.752	-0.719	-0.706	-0.979	-1.371	-0.421	-0.667
2	-0.990	7.346	-0.628	-0.761	-0.780	-0.814	-0.937	-1.229	-0.376	-0.832
3	-1.210	-1.035	9.715	-0.829	-0.885	-1.123	-1.367	-1.742	-0.547	-0.976
4	-1.342	-1.010	-1.280	10.640	-0.910	-1.056	-1.495	-1.923	-0.477	-1.147
5	-0.623	-0.563	-0.532	-0.481	5.859	-0.721	-0.822	-1.150	-0.210	-0.759
6	-1.606	-1.160	-1.514	-1.020	-0.842	10.299	-1.148	-1.594	-0.428	-0.986
7	-1.188	-0.876	-0.964	-0.670	-0.689	-0.770	7.254	-1.108	-0.195	-0.794
8	-1.313	-0.748	-1.142	-0.692	-0.688	-0.868	-1.000	7.694	-0.321	-0.923
9	-1.212	-0.947	-1.418	-1.125	-0.589	-0.980	-1.223	-0.876	9.368	-0.996
10	-1.093	-0.811	-0.909	-0.555	-0.529	-0.821	-0.736	-0.865	-0.290	6.610

We believe the above probability is overestimated because we used the $M/M/c$ model to estimate the probability distribution of the number of MRs in the global queue. Since the $M/M/c$ model is based on FCFS, it is less efficient than L/OF-L/OF, resulting in more MRs in the global queue *for the same ρ value.* (A simple numeric example is shown in Table 2.7.) Since the $M/M/c$ model overestimates the number of MRs in the global queue, the analytic model slightly overestimates the probability that a device finds a local MR. The above difference in the average queue length decreases, and consequently the analytic model performs better, if ρ is smaller or the flow is unbalanced (rendering L/OF-L/OF less effective).

Table 2.7: Average Number of MRs in the Global Queue at $\rho = 0.94$

ρ = 0.94	Flow 1	Flow 2
FCFS	13.3	12.9
Mod-FCFS	4.78	7.75

2.5.3 System Stability

Let S^+ and S^- denote the set of stations with positive and negative NFs, respectively. In deriving the M-M LB, Maxwell and Muckstadt (1982) argue that empty trips occur only from station i to station j, where $i \in S^+$ and $j \in S^-$. For example, if $\Lambda_i = 5$ and $\lambda_i = 2$, then $NF_i = +3$, meaning three empty trips/hr will originate at i. Behind their argument is the assumption that, *in the best case*, two loaded devices/hr arriving at station i will both find a local load and leave station i loaded, while the remaining inbound loaded devices (3/hour) will have to leave station i empty. Likewise, if $\Lambda_j = 1$ and $\lambda_j = 3$, then $NF_j = -2$, meaning, *in the best case*, one inbound loaded device/hr will find a local load and leave station j loaded, while two additional empty devices/hr must be dispatched to station j to pick-up the remaining loads. (As explained in section 4.4, solving a transportation problem yields the e_{ij} values.)

Using a logic similar to M-M LB, but allocating the empty devices proportionally (instead of solving a transportation problem), we arrive at a different result. For example, numbering the stations 1 through 5, if $S^+ = \{+3, +5\}$ and $S^- = \{-2, -2, -4\}$, then under L/OF-L/OF (which serves the oldest MR in the system whenever a local load is not found), the 3 empty devices/hr generated at station 1 will be allocated as follows: $e_{13} = (2/8)(3)$, $e_{14} = (2/8)(3)$ and $e_{15} = (4/8)(3)$. The same allocation applies to station 2 but with 5 empty devices/hr. Generalizing the above allocation, we obtain:

$$e_{ij} = (NF_i)\left(\frac{NF_j}{\sum_{j \in S^-} NF_j}\right) \text{ for } i \in S^+ \text{ and } j \in S^- \tag{2.21}$$

Interestingly, equation (2.21) is the same expression obtained by Bozer and Srinivasan (1991) for "mandatory empty trips" under FEFS. Hence, we name it the Bozer-Srinivasan Index (BSI).

We hypothesize that α_e obtained from the e_{ij} values in equation (2.21) serves as a stability condition for L/OF-L/OF. As $\rho \rightarrow 1$, the proportion of DID → 1, and the system begins to mimic the best case explained above, i.e., incoming loaded devices find a local load, with the remainder traveling empty to stations in set S^- as explained above. We cannot prove our hypothesis but we provide empirical evidence, although doing so is challenging since, without an exact analytic result, one cannot draw firm conclusions by computing an approximate ρ value or examining the global queue. Therefore, the results, although convincing, are subject to further investigation.

The experiment for the stability test is as follows; we first set the device travel speed such that the utilization is very high, but the system is still stable (case C1). More cases are created by gradually decreasing the device travel speed, thus increasing their utilization (up to case C7). The M-M LB, BSI, analytic, and simulation results are compared in each case. The experiment was conducted for LO1 and LO3, equipped with 7 devices, and with flow set 1 and 2.

Table 2.8a: Stability Test for LO1

		Flow 1				**Flow 2**			
		M-M LB	**BSI**	**ANA**	**SIM**	**M-M LB**	**BSI**	**ANA**	**SIM**
C1	α_e	0.143	0.177	0.304	0.325	0.206	0.366	0.467	0.487
	α_f	0.674	0.674	0.674	0.674	0.485	0.485	0.485	0.485
	ρ	0.817	0.851	0.977	0.999	0.690	0.851	0.952	0.972
C2	α_e	0.148	0.184	0.283	0.299	0.212	0.376	0.463	0.482
	α_f	0.700	0.700	0.700	0.700	0.499	0.499	0.499	0.499
	ρ	0.848	0.884	0.983	0.999	0.710	0.875	0.962	0.981
C3	α_e	0.154	0.191	0.260	0.271	0.224	0.399	0.451	0.464
	α_f	0.728	0.728	0.728	0.729	0.529	0.529	0.529	0.530
	ρ	0.882	0.919	0.988	1.000	0.753	0.928	0.980	0.993
C4	α_e	0.161	0.199	0.234	0.242	0.231	0.412	0.443	0.452
	α_f	0.758	0.758	0.758	0.758	0.545	0.545	0.545	0.545
	ρ	0.919	0.957	0.992	1.000	0.777	0.957	0.989	0.997
C5	α_e	0.168	0.208	0.205	0.208	0.239	0.425	0.433	0.437
	α_f	0.791	0.791	0.791	0.792	0.563	0.563	0.563	0.563
	ρ	0.959	0.999	0.996	1.000	0.802	0.988	0.996	1.000
C6	α_e	0.171	0.213	UNS?	OL	0.247	0.439	UNS?	OL
	α_f	0.808	0.808	UNS?	OL	0.582	0.582	UNS?	OL
	ρ	0.980	1.021	UNS?	OL	0.829	1.021	UNS?	OL
C7	α_e	0.175	0.218	UNS?	OL	0.255	0.454	UNS?	OL
	α_f	0.827	0.827	UNS?	OL	0.602	0.602	UNS?	OL
	ρ	1.002	1.044	UNS?	OL	0.857	1.056	UNS?	OL

Table 2.8b: Stability Test for LO3

		Flow 1				Flow 2			
		M-M LB	**BSI**	**ANA**	**SIM**	**M-M LB**	**BSI**	**ANA**	**SIM**
C1	α_e	0.052	0.096	0.246	0.254	0.179	0.314	0.448	0.460
	α_f	0.746	0.746	0.746	0.746	0.539	0.539	0.539	0.540
	ρ	0.799	0.842	0.992	1.000	0.718	0.853	0.987	1.000
C2	α_e	0.054	0.100	0.217	0.224	0.184	0.323	0.435	0.444
	α_f	0.776	0.776	0.776	0.776	0.555	0.555	0.555	0.556
	ρ	0.830	0.876	0.993	1.000	0.739	0.878	0.990	1.000
C3	α_e	0.057	0.104	0.187	0.192	0.189	0.332	0.421	0.428
	α_f	0.808	0.808	0.808	0.808	0.571	0.571	0.571	0.572
	ρ	0.865	0.913	0.995	1.000	0.761	0.904	0.993	1.000
C4	α_e	0.059	0.109	0.153	0.156	0.195	0.342	0.406	0.411
	α_f	0.843	0.843	0.843	0.844	0.588	0.588	0.588	0.589
	ρ	0.903	0.952	0.997	1.000	0.784	0.931	0.995	1.000
C5	α_e	0.062	0.114	0.116	0.118	0.201	0.353	0.390	0.393
	α_f	0.882	0.882	0.882	0.882	0.607	0.607	0.607	0.607
	ρ	0.944	0.996	0.998	1.000	0.808	0.960	0.997	1.000
C6	α_e	0.065	0.119	UNS?	OL	0.215	0.377	UNS?	OL
	α_f	0.924	0.924	UNS?	OL	0.647	0.647	UNS?	OL
	ρ	0.989	1.043	UNS?	OL	0.862	1.024	UNS?	OL
C7	α_e	0.068	0.125	UNS?	OL	0.222	0.390	UNS?	OL
	α_f	0.970	0.970	UNS?	OL	0.670	0.670	UNS?	OL
	ρ	1.038	1.095	UNS?	OL	0.892	1.059	UNS?	OL

Since we are testing for stability of each case, the simulation model will run for a total of one million loaded trips, with only one replication. Recall that the simulation model will terminate sooner if the system becomes overloaded (OL), and the analytic model cannot determine the system’s stability (UNS?) when $\rho > 0.999$ during the iterative algorithm in section 2.4.4.

The results in Table 2.8a and 2.8b show that the iterative algorithm for the analytic model converges when the simulation model indicates that the system is not OL, and vice versa. Furthermore, both BSI and the analytic model seem to correctly predict when the system is OL. (Of course, an OL system may be stable but it is virtually impossible to verify that through simulation.) Hence, our empirical results suggest that BSI may serve as an approximate stability indicator for L/OF-L/OF, and the iterative algorithm correctly returns a result of "UNS?" when BSI is close to one or larger.

2.6 Summary and Conclusions

The L/OF-L/OF (Mod-FCFS) rule, proposed by Srinivasan et al. (1994), is a simple dispatching rule that is reasonably efficient and in most cases performs better than OF-OF (FCFS). We present a new analytic model for the L/OF-L/OF rule to evaluate trip-based handling systems with multiple devices. Using an iterative algorithm, we estimate the station-to-station empty trips under DID and SID, which then yields the estimated expected device utilization.

Overall, the analytic model performs well over a range of ρ values; the empty trips for both DID (e_{ij}^B) and SID (e_{ij}^I) are estimated with reasonable accuracy. Although the empty device allocation errors for e_{ij}^B is larger than those of e_{ij}^I, the median allocation error is consistently less than one percentage point. Furthermore, the analytic model, combined with the BSI, is a good tool to predict the "stability" of the system, and it is consistent with the results obtained from simulation. In the process of developing the analytic model, we also resolved conflicting views in the literature concerning the significance/dominance of DID versus SID in trip-based handling systems. Using the Erlang-C equation as an approximation, we showed that for most ρ values, both DID and SID play a role in the performance of the system.

For future research, an analytic model to estimate the expected MR waiting times would be desirable. Although Kingman's formula (1961) can be used, the results are generally inaccurate for efficient dispatching rules that reduce empty travel. It would also be desirable to extend the analytic model to systems with multiple MR priorities or systems where each device carries two or more loads at-a-time. New performance measures may be needed to evaluate such systems. Finally, the iterative algorithms to rescale the empty trips (section 4.3.3) and to compute ρ (section 4.4) always converged (except when the latter was terminated when $\rho > 0.999$); however, it would be desirable to identify the conditions under which the two algorithms are guaranteed to converge.

2.7 Appendices

Appendix 2.A: 4-Station Problem

The following Figure and Tables show the layout, the flow data (MRs/hour), and the travel times (in mins) for the 4-station example problem.

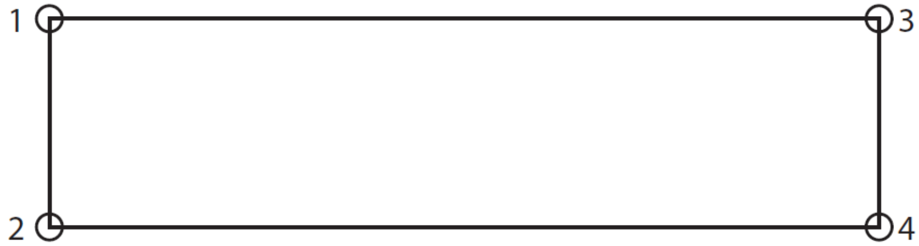


Figure 2.A1: Layout of 4 Station Example

Table 2.A1: Unbalanced Flow

	1	**2**	**3**	**4**	**NF**
1		12			-10
2	2				10
3					14
4			14		-14

Table 2.A2: Balanced Flow

	1	**2**	**3**	**4**	**NF**
1		7			0
2	7				0
3				7	0
4			7		0

Table 2.A3: Travel Times

	1	**2**	**3**	**4**
1	0	100	400	420
2	100	0	420	400
3	400	420	0	100
4	420	400	100	0

Appendix 2.B: Input Data for Model Evaluation

Figure 2.B1 illustrates the layout configuration for LO1. Flow 1, flow 2 (in MRs/hour) and travel distance matrices (in units) for LO1 are shown in Table 2.B1, Table 2.B2 and Table 2.B3, respectively. The flow and travel distance data for LO2 (see Figure 2.B2) are shown in Table 2.B4, Table 2.B5 and Table 2.B6. And the flow and travel distance data for LO3 (see Figure 2.B3) are shown in Table 2.B7, Table 2.B8 and Table 2.B9. Note that LO1 and LO3 have unidirectional paths, and LO2 has bidirectional paths. Lastly, Table 2.B10 shows the device travel speed (in units/min) of each experiment.

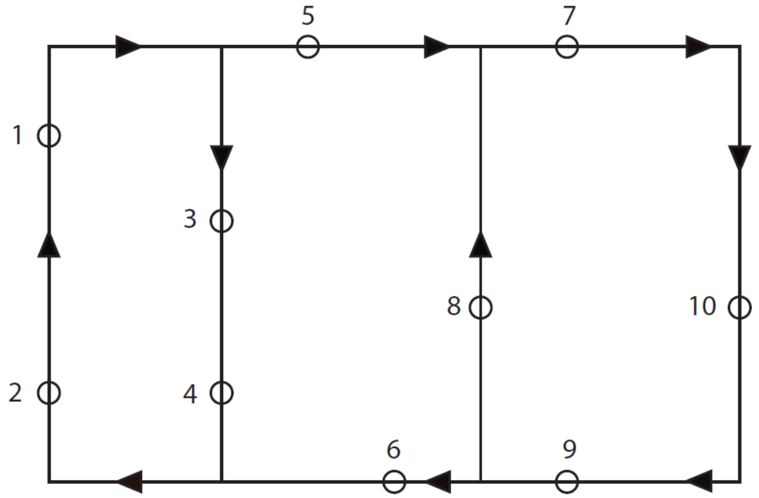


Figure 2.B1: Illustration of LO1 (Unidirectional paths)

Table 2.B1: Flow 1 for LO1

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Out**	**NF**
1			3	1	4		2				**10**	**1**
2	1		2	1			1	1			**6**	**2**
3	2	3		2	3						**10**	**-3**
4	3	1			1		2			1	**8**	**-3**
5								2	1	2	**5**	**4**
6	2	1	2	1			3				**9**	**-3**
7					1	1		4		2	**8**	**2**
8	1	2				2	2		2	4	**13**	**-2**
9						1		3			**4**	**-1**
10	2	1				2		1			**6**	**3**
In	**11**	**8**	**7**	**5**	**9**	**6**	**10**	**11**	**3**	**9**	**79**	

Table 2.B2: Flow 2 for LO1

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Out**	**NF**
1			4	2	3	1	3			6	**19**	**-15**
2			5	7							**12**	**-7**
3					1					1	**2**	**10**
4		1					1			1	**3**	**8**
5							1			2	**3**	**5**
6	4	1	2	2	3						**12**	**-7**
7			1			1		6		3	**11**	**-5**
8		1			1		1		1	2	**6**	**5**
9		2				2		4			**8**	**-7**
10						1		1			**2**	**13**
In	**4**	**5**	**12**	**11**	**8**	**5**	**6**	**11**	**1**	**15**	**78**	

Table 2.B3: Travel Distance Matrix for LO1

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	0	11	5	7	4	18	7	19	16	12
2	3	0	8	10	7	21	10	22	19	15
3	9	6	0	2	13	27	16	28	25	21
4	7	4	12	0	11	25	14	26	23	19
5	22	19	27	29	0	14	3	15	12	8
6	8	5	13	15	12	0	15	27	24	20
7	19	16	24	26	23	11	0	12	9	5
8	23	20	28	30	27	15	4	0	13	9
9	10	7	15	17	14	2	7	3	0	12
10	14	11	19	21	18	6	11	7	4	0

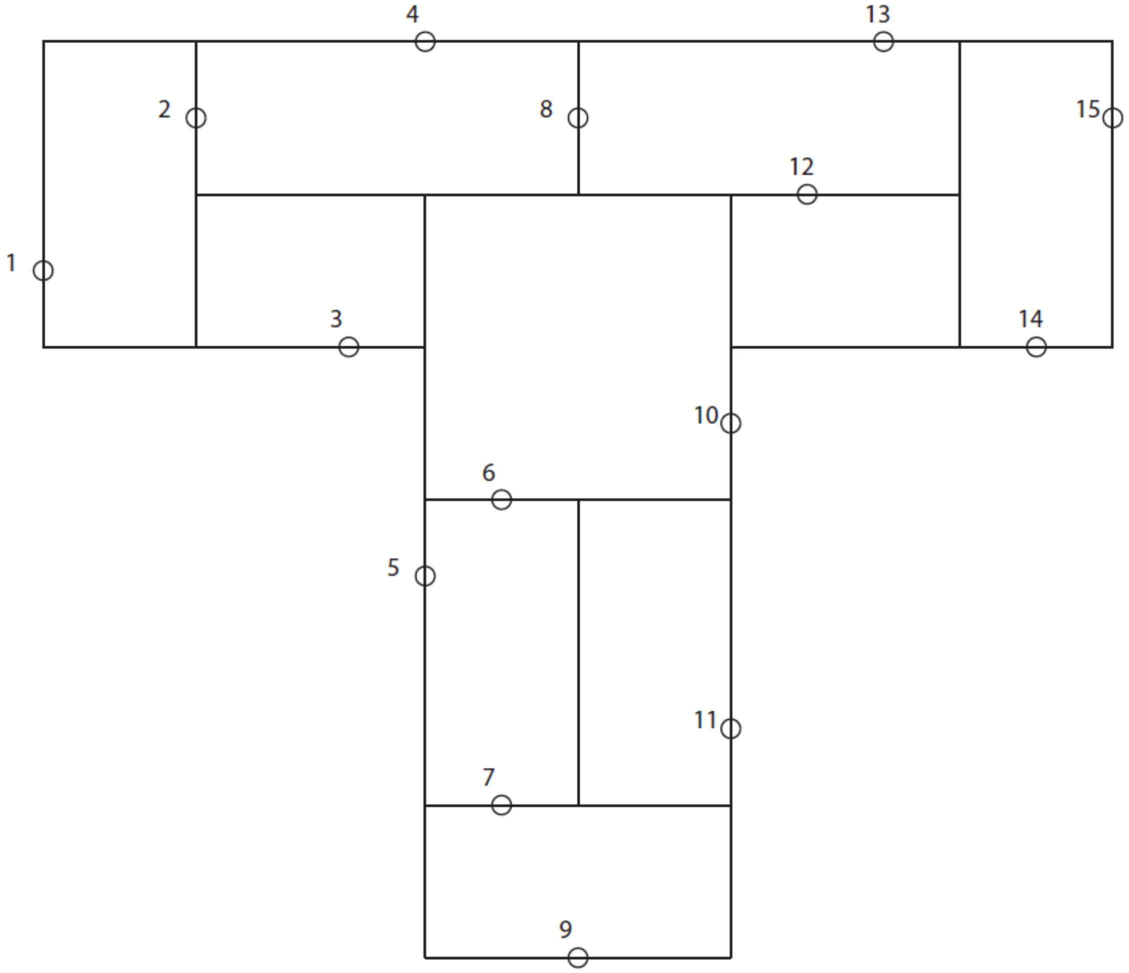


Figure 2.B2: Illustration of LO2 (Bidirectional paths)

Table 2.B4: Flow 1 for LO2

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Out	NF
1		1		3			1					1	1			**7**	**2**
2			1		2	1		2		1		2				**9**	**2**
3	2	2				3	1	2			1			1		**12**	**-3**
4		2					1	2					1		2	**8**	**2**
5	2			1		2		1	3			1				**10**	**2**
6	2	1	1		3					2		1			2	**12**	**0**
7		2		2	2				3		1				1	**11**	**0**
8	3	1	2		1							2	1			**10**	**0**
9			2		1		2						1			**6**	**3**
10		2		2		1					2			1		**8**	**-3**
11					3				3						1	**7**	**-1**
12			2			2	1				2		1	2		**10**	**0**
13			1				2	1		1					3	**8**	**-3**
14				2		3				1					2	**8**	**-3**
15							3	2				3		1		**9**	**2**
In	**9**	**11**	**9**	**10**	**12**	**12**	**11**	**10**	**9**	**5**	**6**	**10**	**5**	**5**	**11**	**135**	

Table 2.B5: Flow 2 for LO2

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Out	NF
1		3	2	2		3	2	1		1			1			**15**	**-9**
2			3		2	1		2				1				**9**	**-4**
3	2					1	2	2			2	2				**11**	**0**
4	1							3				1				**5**	**4**
5			3	1			2	3	1		1	2		1		**14**	**-10**
6													2			**2**	**12**
7						2					1					**3**	**8**
8	1											2	2		1	**6**	**15**
9			1			2	3			2	2	1		1		**12**	**-11**
10				1	2	1	2	3			2	2	1	3	1	**18**	**-14**
11						2		2				2	1	1		**8**	**0**
12				2				1					1		2	**6**	**13**
13	2	1	2	2				2				1		1	2	**13**	**0**
14												3	2		1	**6**	**3**
15		1		1		2		2		1		2	3	2		**14**	**-7**
In	**6**	**5**	**11**	**9**	**4**	**14**	**11**	**21**	**1**	**4**	**8**	**19**	**13**	**9**	**7**	**142**	

Table 2.B6: Travel Distance Matrix for LO2

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
1	0	6	5	8	10	9	13	11	16	13	15	13	14	18	18
2	6	0	5	4	10	9	13	7	16	11	15	9	10	14	14
3	5	5	0	9	5	4	8	6	11	8	10	8	11	13	15
4	8	4	9	0	12	11	15	3	18	9	13	7	6	12	10
5	10	10	5	12	0	3	3	9	6	7	7	11	14	12	16
6	9	9	4	11	3	0	6	8	9	4	6	8	13	9	13
7	13	13	8	15	3	6	0	12	5	8	4	12	17	13	17
8	11	7	6	3	9	8	12	0	15	6	10	4	5	9	9
9	16	16	11	18	6	9	5	15	0	9	5	13	18	14	18
10	13	11	8	9	7	4	8	6	9	0	4	4	9	5	9
11	15	15	10	13	7	6	4	10	5	4	0	8	13	9	13
12	13	9	8	7	11	8	12	4	13	4	8	0	5	5	7
13	14	10	11	6	14	13	17	5	18	9	13	5	0	6	4
14	18	14	13	12	12	9	13	9	14	5	9	5	6	0	4
15	18	14	15	10	16	13	17	9	18	9	13	7	4	4	0

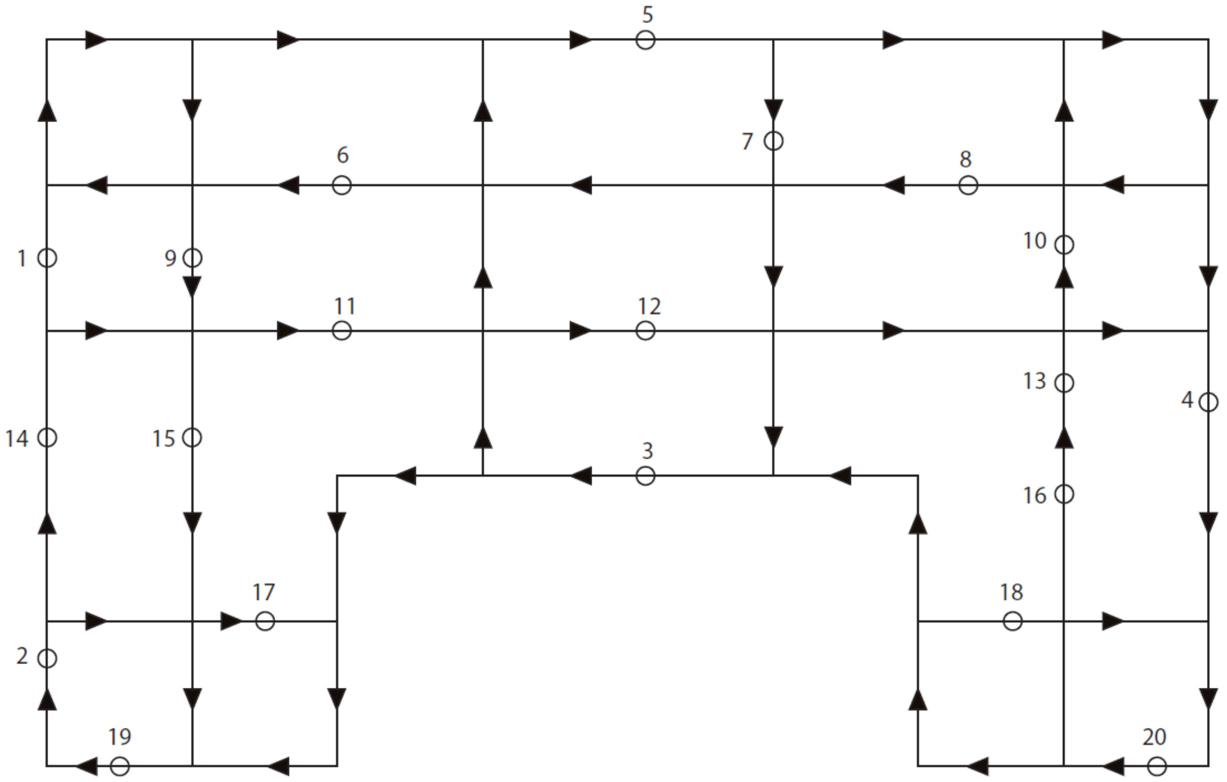


Figure 2.B3: Illustration of LO3 (Unidirectional paths)

Table 2.B7: Flow 1 for LO3

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Out	NF
1			1		1	2	3		1												**8**	**0**
2	1									1	1	3		2			2				**10**	**0**
3		1		1	1	2				3		1							2		**11**	**0**
4						1							3			2		3		4	**13**	**0**
5				2		2	2			2					1						**9**	**0**
6		2			1				3						2		1				**9**	**1**
7		1	3	2					2												**8**	**-1**
8			2			1			1						2		1		1		**8**	**0**
9	2				1		2										2		3		**10**	**0**
10			1						2		2				2					1	**8**	**0**
11				2	2	1				1		2									**8**	**-3**
12				2				3		1										2	**8**	**3**
13			1	2				1								1				2	**7**	**-1**
14	2				1				1			2									**6**	**-2**
15		3												2							**5**	**2**
16				1				3													**4**	**1**
17	1	2																			**3**	**3**
18			1					1					3			2					**7**	**-1**
19	2	1			2	1					2	1									**9**	**-3**
20			2	1								2						3			**8**	**1**
In	**8**	**10**	**11**	**13**	**9**	**10**	**7**	**8**	**10**	**8**	**5**	**11**	**6**	**4**	**7**	**5**	**6**	**6**	**6**	**9**	**159**	

Table 2.B8: Flow 2 for LO3

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Out	NF
1			1				2		3		4				3		2				**15**	**-9**
2										1	2	3		2			1				**9**	**-3**
3																					**0**	**10**
4																		2		1	**3**	**8**
5			1	2		2	3	1				1									**10**	**-10**
6							1		2						4		1		3		**11**	**-5**
7			3												3		2		1		**9**	**0**
8	1						1								1		1		1		**5**	**9**
9	1	1		2							2						3		2		**11**	**-3**
10			1	2		2		2	1		1						1			1	**11**	**0**
11				2			2		1	2		3									**10**	**0**
12			2	3						3									1	3	**12**	**-5**
13								3												2	**5**	**3**
14																					**0**	**7**
15	2	2												3			2		3		**12**	**0**
16						2		1	1	2			3		1						**10**	**-10**
17		2						3						2					2		**9**	**4**
18			1					2					3							1	**7**	**-2**
19	2	1									1										**4**	**9**
20			1					2		3			2					3			**11**	**-3**
In	**6**	**6**	**10**	**11**	**0**	**6**	**9**	**14**	**8**	**11**	**10**	**7**	**8**	**7**	**12**	**0**	**13**	**5**	**13**	**8**	**164**	

Table 2.B9: Travel Distance Matrix for LO3

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	0	46	52	60	28	42	36	61	20	55	28	38	91	54	26	87	35	91	40	74
2	14	0	46	54	42	36	50	55	34	49	22	32	85	8	20	81	9	85	14	68
3	48	34	0	38	26	20	34	39	28	33	36	16	69	42	34	65	43	69	28	52
4	90	76	42	0	68	62	76	41	70	35	78	58	31	84	76	27	85	31	70	14
5	72	58	24	32	0	24	8	33	32	27	40	40	63	66	38	59	47	63	52	46
6	48	34	40	48	36	0	44	49	8	43	16	26	79	42	14	75	23	79	28	62
7	64	50	16	24	22	16	0	25	24	19	32	32	55	58	30	51	39	55	44	38
8	69	55	21	29	27	21	35	0	29	24	37	37	60	63	35	56	44	60	49	43
9	40	26	32	40	28	22	36	41	0	35	8	18	71	34	6	67	15	71	20	54
10	75	61	27	25	33	27	41	6	35	0	43	43	56	69	41	52	50	56	55	39
11	62	48	24	32	20	14	28	33	22	27	0	10	63	56	28	59	37	63	42	46
12	62	48	14	22	40	34	48	23	42	17	50	0	53	56	48	49	57	53	42	36
13	79	65	31	9	37	31	45	10	39	4	47	47	0	73	45	36	54	40	59	23
14	6	32	38	46	34	28	42	47	26	41	14	24	77	0	12	73	21	77	26	60
15	34	20	66	74	62	56	70	75	54	69	42	52	105	28	0	101	9	105	14	88
16	83	69	35	13	41	35	49	14	43	8	51	51	4	77	49	0	58	44	63	27
17	35	21	67	75	63	57	71	76	55	70	43	53	106	29	41	102	0	106	15	89
18	89	75	41	19	47	41	55	20	49	14	57	57	10	83	55	6	64	0	69	13
19	20	6	52	60	48	42	56	61	40	55	28	38	91	14	26	87	15	91	0	74
20	76	62	28	26	54	48	62	27	56	21	64	44	17	70	62	13	71	17	56	0

Table 2.B10: Device Travel Speed of each Experiment, in Distance Units/Minute

		3 Devices			**7 Devices**		
	Target *ρ*	**0.90**	**0.80**	**0.60**	**0.90**	**0.80**	**0.60**
Flow 1	**LO1**	9.60	11.50	15.90	4.13	4.83	6.38
	LO2	12.67	14.90	20.45	5.40	6.30	8.40
	LO3	59.50	70.00	95.50	25.50	29.70	39.70
Flow 2	**LO1**	9.60	11.30	15.60	4.16	4.80	6.38
	LO2	13.20	15.30	20.80	5.64	6.48	8.64
	LO3	63.00	73.50	101.00	27.00	31.50	42.00

CHAPTER 3

Analysis of the Shortest-Travel-Time-First Dispatching Rule with Multiple Devices

3.1 Introduction

Designed for delivering the right material, at the right place, and at the right time, material handling systems play a significant role in manufacturing and service operations. Although material handling is considered a non-value adding function in manufacturing, a well-designed material handling system often increases productivity and on-time deliveries while reducing operating costs and work-in-process inventories.

The material handling system we address is a *trip-based* material handling system, where one or multiple devices serve move requests (MRs) one at-a-time (Srinivasan et al., 1994). A MR is a unit load, which is transported by a device from its origin (pick-up station) to its destination (deposit station). The devices are homogeneous and operate independently. The MRs arrive one at-a-time and wait at their origin. Once a device is assigned to a MR, it performs an *empty trip* from its current location to the pick-up station of the MR. After picking up the MR, the device performs a *loaded trip* and delivers the load at the appropriate deposit station, where it becomes available to serve the next MR. (The terms MR and "load" are used interchangeably.) A wide range of systems such as overhead cranes, lift trucks, and unit load automated guided vehicles (AGVs) can be modeled as trip-based handling systems.

Several elements play a role in the design of a material handling system, including the location of the pick-up/deposit stations (i.e., the layout), the flow network and the routing of the devices, the MR flow data, and the dispatching rule, which assigns each MR to a device, and vice versa. Given these elements, one is frequently concerned with determining the number of devices required and their expected utilization as well as the expected MR waiting times.

The loaded travel times are impacted by the layout and the flow network, which are assumed to be given in our study. The empty travel times, which are unproductive, are also impacted by the layout/flow network but they depend on the dispatching rule as well. Among numerous rules presented in the literature, the shortest-travel-time-first (STTF) rule is a simple, yet efficient dispatching rule, which seeks to reduce empty device travel by assigning an available device to the closest MR, and vice versa. It is used as a benchmark in the literature, and often found in commercial applications as well. Material handling vendors such as Savant Automation (2016) and Frog AGV Systems (2016), and fleet management providers such as Telogis (2016), employ the STTF rule. It is also frequently used for online taxi dispatching (Jung et al., 2013).

Although it is a common rule, an analytic model to estimate empty device travel under STTF is not available in the literature. The primary challenge is that the dispatching decision depends on the (changing) location of a device relative to the MRs in the system. Furthermore, for a single-device system, Larson and Odoni (2007, p. 516) state that successive service times (empty plus loaded travel) are not independent because after serving a MR with a long service time, the device is more likely to find multiple MRs with shorter empty travel times. Thus, the STTF rule is unlikely to be analytically tractable. In fact, according to Larson and Odoni (2007), systems using STTF are often modeled via simulation.

It has also been remarked in some studies that, under STTF, depending on the flow and the layout, some MRs may experience excessive wait times (see [Bozer and Yen, 1996], and [de Koster et al., 2004], among others). This is primarily because, with STTF, a MR is likely to experience multiple "slips" (Larson, 1987). (A "slip" occurs each time MR_y is served before MR_x, although MR_x arrived before MR_y. By definition, there are no slips in a global queue with FCFS.)

In this chapter we extend the analytical technique presented in Chapter 2 to develop an analytic model to estimate the empty device trips, and ultimately, the expected device utilization under the STTF rule. To the best of our knowledge, our model is the first single- or multi-device analytic model for the STTF rule, while accounting for both DID and SID. Since the model is an approximate one, we do not propose it as a substitute for simulation. Rather, it can be used to rapidly evaluate alternative handling systems and determine the number of devices required, while performing "what if" analyses for the layout or the flow data. Simulation can be used to perform a detailed analysis of the system, including the expected MR waiting times and possible congestion/blocking. Additionally, we perform simulation analysis to investigate the MR wait times under STTF, and we propose a bound to limit excessive MR wait times.

In the next section, the analytic model for the STTF rule is presented. In section 3.3, the analytic model is evaluated using simulation. In section 3.4, the MR wait times under STTF is investigated. Lastly, the summary results and possible future research directions are discussed in section 3.5.

3.2 Analytic Model to Estimate Device Empty Travel under STTF

In this section, we present a multi-device analytic model for the STTF rule. The objective of the model is to estimate the empty trips and the expected device utilization, given the layout and data for loaded trips. Given that a sufficient number of devices is provided, that is, the system is stable, the expected device utilization can be computed as shown earlier in section 2.4.

Under the STTF rule, DID assigns the closest MR to the device, and SID assigns the closest idle device to the MR (see Table 2.1). Hence, the empty trips that occur as a result of DID and SID, denoted by e_{ij}^{B} and e_{ij}^{I}, respectively, are derived separately, that is,

$$e_{ij} = e_{ij}^{B} + e_{ij}^{I}\ , \tag{3.1}$$

and the probability of invoking DID and SID are estimated using the Erlang C equation (as shown in section 2.4.2).

3.2.1 Device-Initiated Empty Trips

Let m (M) denote the number of MRs in the global queue (in the system). DID occurs when a device delivers a load at station i and finds $m \geq 1$. In order to estimate e_{ij}^{B}, we need to compute the probability that the empty device, currently at station i, is dispatched to station j.

In order to locate a MR, the dispatch system (DS) searches the stations in order of proximity, i.e., from closest to farthest station. Given that the device is currently at station i, the sequence of stations is sorted as $\Omega_{(1)}, \Omega_{(2)}, .., \Omega_{(S)}$, where $\Omega_{(1)}$ is the station closest to station i, and $\Omega_{(S)}$ is the station farthest from station i. By definition, the DS first checks station $\Omega_{(1)}$ for a MR, and if one is found, the search ends and the MR is assigned to the device. If none is found, then the DS searches the next closest station ($\Omega_{(2)}$), and so on, until a MR is located. (Recall that $m \geq 1$.)

Let $q_{(j)}$ be the probability that the device is assigned to a MR at station j. (The parenthesis denotes that the stations are sorted based on proximity.) Let $\mathbb{P}_M$ (p_m) denote the probability that there are M (m) MRs in the system (the global queue). Since determining the time-average probabilities for the number of MRs in the system is difficult, and is possibly analytically intractable, we estimate $\mathbb{P}_M$ and p_m by using the results for $M/M/c$ queue (where $c = D$). Given ρ and D, we have (Kleinrock, 1975, p.102):

$$\mathbb{P}_M = \begin{cases} \mathbb{P}_0 \dfrac{(\rho D)^M}{M!}, & \text{for } 1 \leq M \leq D \\ \mathbb{P}_0 \dfrac{\rho^M D^D}{D!}, & \text{for } M \ \geq D \end{cases} \tag{3.2}$$

where

$$\mathbb{P}_0 = \left[\sum_{\ell=0}^{D-1} \frac{(\rho D)^\ell}{\ell!} + \frac{(\rho D)^D}{D} \frac{1}{1-\rho} \right]^{-1} \tag{3.3}$$

Given that $m \geq 1$, we re-normalize the above steady-state probabilities to obtain:

$$p'_m = \frac{p_m}{1 - \sum_{\ell=0}^{D-1} P_\ell} \tag{3.4}$$

where $1 - \sum_{\ell=0}^{D-1} P_\ell$ is the probability that there is at least one MR in the global queue.

As stated earlier, the DS first checks station $\Omega_{(1)}$. Assuming independence among the contents of the global queue, the probability that a randomly selected MR is located at $\Omega_{(1)}$ is $\lambda_{(1)}/\lambda_T$, where $\lambda_{(1)}$ denotes the total flow out of station $\Omega_{(1)}$. Hence, given a global queue with exactly m MRs ($m \geq 1$), the probability that none of them are located at station $\Omega_{(1)}$ is equal to $\left(1 - \frac{\lambda_{(1)}}{\lambda_T}\right)^m$. Therefore, the probability that the DS finds a MR at station $\Omega_{(1)}$ is given by:

$$q_{(1)} = \sum_{m=1}^{L} (p'_m) \left[1 - \left(1 - \frac{\lambda_{(1)}}{\lambda_T} \right)^m \right] \tag{3.5}$$

where L is a sufficiently large number.

If the DS does not find a MR at station $\Omega_{(1)}$, it checks station $\Omega_{(2)}$. Given m MRs in the global queue, the probability that the DS does not find a MR at station $\Omega_{(1)}$ is equal to $\left(1-\frac{\lambda_{(1)}}{\lambda_T}\right)^m$. Therefore, the probability that the DS finds a MR at station $\Omega_{(2)}$ is given by:

$$q_{(2)} = \sum_{m=1}^{L} (p'_m) \left[\left(1 - \left(1 - \frac{\lambda_{(2)}}{\lambda_T - \lambda_{(1)}}\right)^m\right) \left(1 - \frac{\lambda_{(1)}}{\lambda_T}\right)^m \right] \tag{3.6}$$

The probability that the DS fails to find a MR at station $\Omega_{(2)}$ (after failing to find a MR at station $\Omega_{(1)}$), given m MRs in the global queue, is equal to $\left(1-\frac{\lambda_{(2)}}{\lambda_T-\lambda_{(1)}}\right)^m \left(1-\frac{\lambda_{(1)}}{\lambda_T}\right)^m$. Therefore, the probability that the DS finds a MR at station $\Omega_{(3)}$ is given by:

$$q_{(3)} = \sum_{m=1}^{L} (p'_m) \left[\left(1 - \left(1 - \frac{\lambda_{(3)}}{\lambda_T - \left(\lambda_{(1)} + \lambda_{(2)}\right)}\right)^m\right) \left(1 - \frac{\lambda_{(2)}}{\lambda_T - \lambda_{(1)}}\right)^m \left(1 - \frac{\lambda_{(1)}}{\lambda_T}\right)^m \right] \tag{3.7}$$

The above search can be described by defining “events” as follows. Let event A denote the case where there is at least one MR at station $\Omega_{(1)}$, event B the case where there is at least one MR at station $\Omega_{(2)}$, and event C the case where there is least one MR at station $\Omega_{(3)}$, and so on. The probability that a device is dispatched to a station (i) is given by:

$$\begin{aligned} q_{(1)} &= P(A) \\ q_{(2)} &= P(B/A')P(A') \\ q_{(3)} &= P(C/B'/A')\, P(B'/A')P(A') \end{aligned} \tag{3.8}$$

Given the above probabilities, it is straightforward to convert $q_{(j)}$ to q_{ij}, i.e., the probability that the device at station i is assigned to a MR at station j. Hence, we obtain:

$$e_{ij}^{B} = \left(q_{ij}\right)\left(e_{i\bullet}^{B}\right) \qquad \text{for } i,j \in S \tag{3.9}$$

where $e^B_{i\bullet}$, the number of DID-based empty trips per hour out of station i, is estimated as shown earlier in sections 2.4.2 and 2.4.3.

3.2.2 Station-Initiated Empty Trips

The approach to estimate e^I_{ij} is similar to that of e^B_{ij}, except that the dispatching decision is viewed from the MR's perspective. Recall that SID occurs when a MR arrives at station j and finds one or more devices idle, i.e., $M \leq (D-1)$. By definition, the DS assigns the closest idle device to the MR at station j. To do so, the DS searches for an idle device, starting with the station closest to station j. Let $r_{(i)}$ be probability that the DS locates an idle device at station i, given that $M \leq (D-1)$. (The parenthesis denotes that the stations are sorted based on proximity.) Note that $r_{(i)}$ is an arrival instance probability but the arriving MRs observe the same equilibrium state distribution as time-averaging (Wolff, 1982).

Let π_d denote the probability that there are d idle devices when a MR arrives at station j ($d \leq D$). Again, using the $M/M/c$ queue as an approximation, we have:

$$\pi_d = \mathbb{P}_{D-d}, \qquad \text{for } 1 \leq d \leq D$$

Given that $M \leq (D-1)$, the above steady-state probabilities are normalized as follows:

$$\pi'_d = \frac{\pi_d}{\sum_{M=0}^{D-1} P_M} \tag{3.10}$$

where $\sum_{M=0}^{D-1} P_M$ is the probability that there is at least one idle device in the system.

Given that a MR just arrived at station j, let $\Omega_{(1)}, .., \Omega_{(S)}$ denote the set of stations sorted by proximity. The probability that an idle device is located at station $\Omega_{(1)}$ is $\Lambda_{(1)}/\Lambda_T$, where $\Lambda_{(1)}$ denotes the total flow into station $\Omega_{(1)}$. Therefore, given d idle devices, the probability that at least one of them is at station $\Omega_{(1)}$ is $1-\left(1-\frac{\Lambda_{(1)}}{\Lambda_T}\right)^d$. Hence,

$$r_{(1)} = \sum_{d=1}^{D} (\pi'_d) \left[1 - \left(1 - \frac{\Lambda_{(1)}}{\Lambda_T} \right)^d \right] \tag{3.11}$$

We can see that the above equation is the same as equation (3.5), but with a change in the variables. Similarly, $r_{(2)}$ is computed by making the same changes to equation (3.6):

$$r_{(2)} = \sum_{d=1}^{D} (\pi'_d) \left[\left(1 - \left(1 - \frac{\Lambda_{(2)}}{\Lambda_T - \Lambda_{(1)}} \right)^d \right) \left(1 - \frac{\Lambda_{(1)}}{\Lambda_T} \right)^d \right] \tag{3.12}$$

and $r_{(3)}$ is computed by making the same changes to equation (3.7). The sequence of events shown in equation (3.8) are applied to determine $r_{(i)}$ for each $i \in S$. As before, the $r_{(i)}$ values are converted to r_{ij}, i.e., the probability that the idle device at station i is assigned to the MR at station j, and we obtain:

$$e^I_{ij} = (r_{ij})(e^I_{\bullet j}) \qquad \text{for } i, j \in S \tag{3.13}$$

where $e^I_{i\,\bullet}$, the number of SID-based empty trips per hour out of station i, is estimated as shown earlier in sections 2.4.2 and 2.4.3.

3.2.3 Rescaling the Empty Trips

Due to how e^B_{ij} and e^I_{ij} are derived, the estimated values may need to be rescaled in order to satisfy conservation of flow. We utilize the same empty trip rescaling algorithm shown in section 2.4.3.3. Once the algorithm converges, it yields the estimated values of both e^B_{ij} and e^I_{ij}, which are used in equations (2.1) and (3.1) to compute the values of α_e and ρ.

3.2.4 Iterative Algorithm to Compute ρ

In the previous sections, a given ρ value is used to estimate e^B_{ij} and e^I_{ij}, which can then be used to estimate a ρ value. Hence, our analytic model cannot be solved in closed form, and we employ the same iterative scheme shown in section 2.4.4 to estimate the values of e_{ij} and ρ. The iterative procedure proceeds as follows (also depicted in Figure 3.1):

1) Set $n = 1$.
2) Set the initial $\rho^{(n)}$ equal to the lower bound (computed by solving a transportation problem presented in [Maxwell and Muckstadt, 1982]). If $\rho^{(n)} \geq 1$, **stop**; the system is unstable, and more devices are needed.
3) Using $\rho^{(n)}$, estimate the values of $e_{ij}^{B(n)}$ and $e_{ij}^{I(n)}$.
4) Using $e_{ij}^{B(n)}$ and $e_{ij}^{I(n)}$, compute the new expected device utilization, $\hat{\rho}^{(n)}$. If $\hat{\rho}^{(n)} > 1$, set $\hat{\rho}^{(n)} = 1$.
5) Set the next iteration, $\rho^{(n+1)} = \rho^{(n)} + \Delta\left(\hat{\rho}^{(n)} - \rho^{(n)}\right)$, where Δ is a sufficiently small step size.
6) If $\rho^{(n+1)}$ is approaching 1, i.e., $\rho^{(n+1)} > 0.999$, **stop**; the system may or may not be stable, and more devices are needed to obtain a realistic ρ value.
7) Set $\mathbb{E}_{ij}^{(n)} = \left|e_{ij}^{(n)} - e_{ij}^{(n-1)}\right|$, and let ε be a sufficiently small value. If $\mathbb{E}_{ij}^{(n)} \leq \varepsilon \ \forall\, i, j$, **stop**. The algorithm has converged, and the expected device utilization equals $\rho^{(n)}$. Otherwise, let $n \leftarrow n + 1$ and go to step 3.

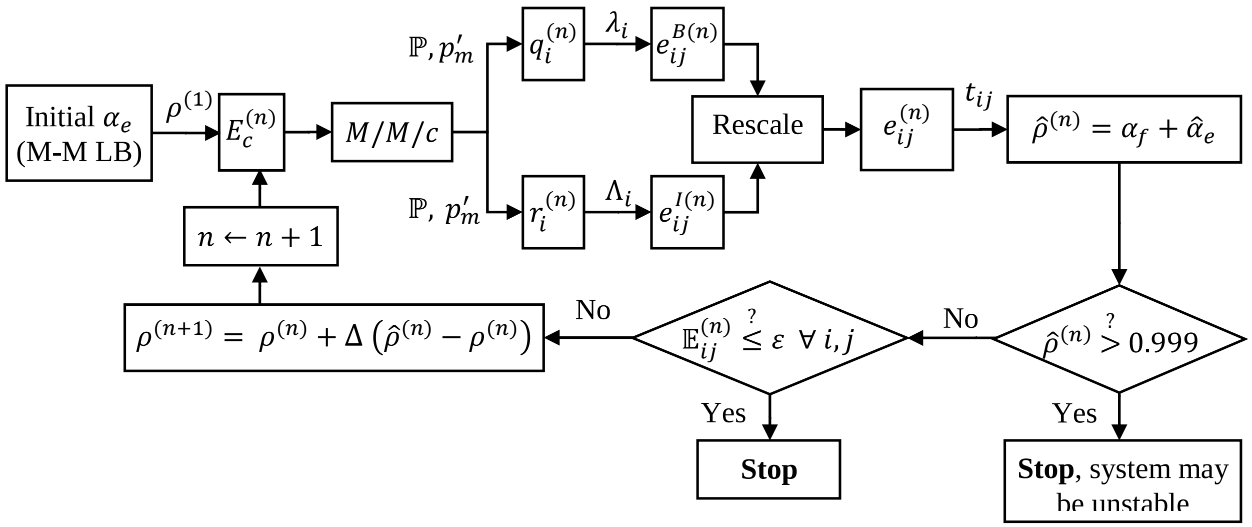

Initial α_e (M-M LB)
$\rho^{(1)}$
$E_c^{(n)}$
$M/M/c$
$\mathbb{P}, p'_m$
$q_i^{(n)}$
λ_i
$e_{ij}^{B(n)}$
$\mathbb{P}, p'_m$
$r_i^{(n)}$
Λ_i
$e_{ij}^{I(n)}$
Rescale
$e_{ij}^{(n)}$
t_{ij}
$\hat{\rho}^{(n)} = \alpha_f + \hat{\alpha}_e$
$n \leftarrow n+1$
$\rho^{(n+1)} = \rho^{(n)} + \Delta\left(\hat{\rho}^{(n)} - \rho^{(n)}\right)$
No
$\mathbb{E}_{ij}^{(n)} \overset{?}{\leq} \varepsilon \ \forall\, i,j$
No
$\hat{\rho}^{(n)} \overset{?}{>} 0.999$
Yes
Stop
Yes
Stop, system may be unstable

Figure 3.1: Iterative Procedure to Compute the Expected Device Utilization

3.3 Simulation Results and Model Evaluation

3.3.1 STTF Analytic Model Evaluation

In this section, test problems and simulation are used to evaluate the performance of the STTF analytic model. The experiment is based on the same three layouts and two flow sets used in Chapter 2. (The three layouts and the data sets are shown in Appendix 2.B.) The device speed is adjusted in each case to generate target ρ values of approximately 0.9 (high), 0.8 (medium) and 0.6 (low). (See Table 3.1 for the device speed used for each case.) As in Chapter 2, the simulation results are based on 10 replications, with 20,000 loaded trips per device per replication, and a warm-up period of 1,000 loaded trips. The Tecnomatix Plant Simulation package (2014) is used for the simulation model.

In order to evaluate the analytic model, the analytic estimates of α_e and ρ are compared with the simulation results, and the difference is reported as a percentage error. The estimated empty trips, e_{ij}^B, e_{ij}^I, and e_{ij}, are also compared with the empty trips obtained from simulation using the percentage point (PP) allocation error explained in section 2.5.2.

Table 3.1: Device Travel Speed for each Case, in Distance Units/Minute

		3 Devices			**7 Devices**		
	Target ρ	**0.90**	**0.80**	**0.60**	**0.90**	**0.80**	**0.60**
Flow 1	**LO1**	8.6	10.4	14.4	3.64	4.25	5.53
	LO2	11.5	13.8	19	4.85	5.65	7.45
	LO3	52	63	87	22	26	33.5
Flow 2	**LO1**	8.5	10.2	14.1	3.54	4.16	5.45
	LO2	11.8	14	19.3	4.95	5.8	7.5
	LO3	55	66	92	23.4	27.3	35.5

Tables 3.2a and 3.2b show the results obtained with flow sets 1 and 2, respectively. Overall, the analytic model performs well. The median PP allocation error is consistently less than 1.5 PP. At $\rho = 0.80$ and 0.60, the percentage errors for α_e and ρ, and the median values for the PP allocation error are small. At $\rho = 0.90$, the errors increase slightly, with a maximum error of about 12% in α_e and about 5% in ρ.

The results also indicate that, at medium and high utilization levels, the PP allocation errors for DID $\left(e_{ij}^B\right)$ are higher than those of SID $\left(e_{ij}^I\right)$. Table 3.3a presents the allocation error for e_{ij}^B at $\rho = 0.90$, where the positive (negative) values represent overestimation (underestimation) by the analytic model. We observe that the model overestimates the probability that the device finds a MR at the first station in the sequence (i.e., the diagonal values), and consequently it underestimates α_e. Therefore, systems with medium or high levels of utilization, which have a larger proportion in DID, underestimate α_e.

Table 3.2a: Analytic Model Evaluation with 3 Devices

3 Devices		**Flow 1**							**Flow 2**						
	Target	**Results**				**PP Allocation Error**			**Results**				**PP Allocation Error**		
	ρ		**ANA**	**95% CI SIM**	**Error %**		**Med**	**Max**		**ANA**	**95% CI SIM**	**Error %**		**Med**	**Max**
LO1	**0.90**	α_e	0.360	(0.404, 0.412)	11.676	e_{ij}^B	1.37	14.38	α_e	0.380	(0.410, 0.420)	8.461	e_{ij}^B	1.19	14.19
		α_f	0.494	(0.488, 0.498)	-	e_{ij}^I	0.48	2.85	α_f	0.479	(0.472, 0.490)	-	e_{ij}^I	0.39	2.45
		ρ	0.854	(0.896, 0.906)	5.203	e_{ij}	0.82	9.41	ρ	0.859	(0.886, 0.904)	4.087	e_{ij}	0.76	9.00
	0.80	α_e	0.363	(0.390, 0.398)	7.686	e_{ij}^B	1.11	9.59	α_e	0.373	(0.395, 0.401)	6.308	e_{ij}^B	0.96	10.38
		α_f	0.408	(0.403, 0.413)	-	e_{ij}^I	0.38	1.65	α_f	0.399	(0.394, 0.406)	-	e_{ij}^I	0.26	1.24
		ρ	0.771	(0.796, 0.806)	3.720	e_{ij}	0.72	5.68	ρ	0.772	(0.790, 0.806)	3.218	e_{ij}	0.56	5.68
	0.60	α_e	0.302	(0.308, 0.314)	2.846	e_{ij}^B	0.65	3.12	α_e	0.305	(0.310, 0.316)	2.485	e_{ij}^B	0.38	3.57
		α_f	0.295	(0.290, 0.300)	-	e_{ij}^I	0.22	1.12	α_f	0.289	(0.283, 0.295)	-	e_{ij}^I	0.20	1.44
		ρ	0.597	(0.601, 0.611)	1.526	e_{ij}	0.31	1.70	ρ	0.594	(0.594, 0.610)	1.369	e_{ij}	0.25	1.23
LO2	**0.90**	α_e	0.338	(0.373, 0.383)	10.810	e_{ij}^B	0.89	13.44	α_e	0.371	(0.406, 0.414)	9.329	e_{ij}^B	0.97	16.80
		α_f	0.518	(0.510, 0.526)	-	e_{ij}^I	0.32	1.46	α_f	0.494	(0.487, 0.503)	-	e_{ij}^I	0.24	3.50
		ρ	0.856	(0.888, 0.906)	4.536	e_{ij}	0.57	9.26	ρ	0.866	(0.896, 0.912)	4.255	e_{ij}	0.72	12.45
	0.80	α_e	0.339	(0.361, 0.369)	7.214	e_{ij}^B	0.80	7.41	α_e	0.363	(0.384, 0.394)	6.566	e_{ij}^B	0.83	10.43
		α_f	0.432	(0.426, 0.438)	-	e_{ij}^I	0.19	1.12	α_f	0.417	(0.407, 0.427)	-	e_{ij}^I	0.21	2.39
		ρ	0.770	(0.790, 0.804)	3.292	e_{ij}	0.45	4.20	ρ	0.780	(0.792, 0.818)	3.181	e_{ij}	0.52	6.46
	0.60	α_e	0.279	(0.283, 0.289)	2.552	e_{ij}^B	0.42	2.75	α_e	0.292	(0.293, 0.305)	2.075	e_{ij}^B	0.47	3.75
		α_f	0.314	(0.311, 0.317)	-	e_{ij}^I	0.15	1.04	α_f	0.302	(0.297, 0.307)	-	e_{ij}^I	0.13	1.76
		ρ	0.592	(0.594, 0.606)	1.261	e_{ij}	0.20	1.49	ρ	0.595	(0.590, 0.610)	0.976	e_{ij}	0.22	2.32
LO3	**0.90**	α_e	0.415	(0.460, 0.468)	10.428	e_{ij}^B	0.72	12.15	α_e	0.442	(0.482, 0.492)	9.239	e_{ij}^B	0.69	14.78
		α_f	0.435	(0.425, 0.445)	-	e_{ij}^I	0.29	1.84	α_f	0.412	(0.407, 0.415)	-	e_{ij}^I	0.22	1.89
		ρ	0.850	(0.887, 0.911)	5.373	e_{ij}	0.45	7.67	ρ	0.854	(0.893, 0.903)	4.895	e_{ij}	0.41	9.68
	0.80	α_e	0.408	(0.431, 0.445)	7.010	e_{ij}^B	0.59	6.94	α_e	0.427	(0.451, 0.459)	6.288	e_{ij}^B	0.58	8.13
		α_f	0.359	(0.352, 0.368)	-	e_{ij}^I	0.20	1.67	α_f	0.343	(0.338, 0.348)	-	e_{ij}^I	0.15	1.08
		ρ	0.767	(0.785, 0.811)	3.905	e_{ij}	0.36	3.72	ρ	0.770	(0.792, 0.806)	3.570	e_{ij}	0.34	4.41
	0.60	α_e	0.332	(0.336, 0.348)	2.761	e_{ij}^B	0.36	2.19	α_e	0.341	(0.342, 0.356)	2.138	e_{ij}^B	0.27	3.17
		α_f	0.260	(0.254, 0.266)	-	e_{ij}^I	0.16	1.25	α_f	0.246	(0.243, 0.249)	-	e_{ij}^I	0.14	1.27
		ρ	0.592	(0.591, 0.613)	1.543	e_{ij}	0.20	1.47	ρ	0.588	(0.586, 0.604)	1.146	e_{ij}	0.15	1.44

Table 3.2b: Analytic Model Evaluation with 7 Devices

7 Devices		Flow 1							Flow 2						
	Target	Results				PP Allocation Error			Results				PP Allocation Error		
	ρ		ANA	95% CI SIM	Error %		Med	Max		ANA	95% CI SIM	Error %		Med	Max
LO1	0.90	α_e	0.354	(0.395, 0.407)	11.779	e_{ij}^B	1.20	12.51	α_e	0.375	(0.411, 0.415)	9.141	e_{ij}^B	0.94	13.00
		α_f	0.500	(0.496, 0.504)	-	e_{ij}^I	0.83	4.86	α_f	0.493	(0.488, 0.496)	-	e_{ij}^I	0.64	3.55
		ρ	0.853	(0.895, 0.907)	5.270	e_{ij}	0.83	7.83	ρ	0.868	(0.899, 0.909)	4.020	e_{ij}	0.64	7.49
	0.80	α_e	0.343	(0.370, 0.380)	8.461	e_{ij}^B	0.83	7.26	α_e	0.359	(0.382, 0.390)	7.085	e_{ij}^B	0.70	8.83
		α_f	0.428	(0.423, 0.433)	-	e_{ij}^I	0.64	3.26	α_f	0.420	(0.415, 0.423)	-	e_{ij}^I	0.64	3.06
		ρ	0.771	(0.793, 0.813)	3.925	e_{ij}	0.62	4.72	ρ	0.778	(0.799, 0.811)	3.323	e_{ij}	0.59	4.79
	0.60	α_e	0.244	(0.264, 0.278)	9.943	e_{ij}^B	0.22	1.92	α_e	0.261	(0.283, 0.289)	8.768	e_{ij}^B	0.17	1.81
		α_f	0.329	(0.327, 0.331)	-	e_{ij}^I	0.76	4.08	α_f	0.320	(0.316, 0.324)	-	e_{ij}^I	0.74	4.04
		ρ	0.573	(0.592, 0.608)	4.515	e_{ij}	0.76	4.25	ρ	0.581	(0.600, 0.612)	4.086	e_{ij}	0.70	3.84
LO2	0.90	α_e	0.334	(0.371, 0.379)	10.754	e_{ij}^B	0.86	12.79	α_e	0.368	(0.405, 0.409)	9.552	e_{ij}^B	0.90	16.79
		α_f	0.527	(0.522, 0.532)	-	e_{ij}^I	0.60	2.53	α_f	0.505	(0.498, 0.512)	-	e_{ij}^I	0.48	4.42
		ρ	0.861	(0.896, 0.906)	4.457	e_{ij}	0.52	7.70	ρ	0.873	(0.905, 0.919)	4.248	e_{ij}	0.62	11.43
	0.80	α_e	0.327	(0.349, 0.357)	7.436	e_{ij}^B	0.62	6.56	α_e	0.351	(0.371, 0.377)	6.300	e_{ij}^B	0.69	8.42
		α_f	0.452	(0.446, 0.458)	-	e_{ij}^I	0.45	1.95	α_f	0.431	(0.428, 0.434)	-	e_{ij}^I	0.32	3.70
		ρ	0.779	(0.796, 0.814)	3.233	e_{ij}	0.48	3.49	ρ	0.782	(0.799, 0.811)	2.897	e_{ij}	0.48	5.05
	0.60	α_e	0.239	(0.253, 0.261)	6.845	e_{ij}^B	0.18	1.27	α_e	0.262	(0.270, 0.280)	4.704	e_{ij}^B	0.27	2.15
		α_f	0.343	(0.339, 0.345)	-	e_{ij}^I	0.51	2.65	α_f	0.333	(0.329, 0.337)	-	e_{ij}^I	0.33	3.46
		ρ	0.582	(0.593, 0.605)	2.862	e_{ij}	0.50	2.65	ρ	0.595	(0.599, 0.617)	2.100	e_{ij}	0.34	3.40
LO3	0.90	α_e	0.411	(0.457, 0.461)	10.344	e_{ij}^B	0.63	11.08	α_e	0.436	(0.478, 0.484)	9.228	e_{ij}^B	0.60	13.10
		α_f	0.441	(0.436, 0.446)	-	e_{ij}^I	0.48	3.28	α_f	0.415	(0.410, 0.418)	-	e_{ij}^I	0.43	3.50
		ρ	0.852	(0.893, 0.905)	5.241	e_{ij}	0.41	5.89	ρ	0.851	(0.890, 0.900)	4.876	e_{ij}	0.38	6.98
	0.80	α_e	0.391	(0.415, 0.427)	7.201	e_{ij}^B	0.45	5.05	α_e	0.411	(0.437, 0.445)	6.846	e_{ij}^B	0.40	6.61
		α_f	0.373	(0.368, 0.376)	-	e_{ij}^I	0.37	2.22	α_f	0.356	(0.353, 0.359)	-	e_{ij}^I	0.34	3.18
		ρ	0.764	(0.786, 0.802)	3.731	e_{ij}	0.38	3.08	ρ	0.767	(0.792, 0.802)	3.815	e_{ij}	0.35	3.85
	0.60	α_e	0.288	(0.306, 0.316)	7.372	e_{ij}^B	0.16	1.50	α_e	0.299	(0.319, 0.329)	7.658	e_{ij}^B	0.15	2.37
		α_f	0.290	(0.287, 0.291)	-	e_{ij}^I	0.42	2.81	α_f	0.274	(0.270, 0.276)	-	e_{ij}^I	0.41	4.99
		ρ	0.578	(0.594, 0.606)	3.719	e_{ij}	0.41	3.06	ρ	0.572	(0.589, 0.605)	4.131	e_{ij}	0.43	4.87

We believe the overestimation of the diagonal values of e_{ij}^B stems from the use of the $M/M/c$ model to estimate the probability distribution of the number of MRs in the global queue. The $M/M/c$ model assumes that the MRs are served on FCFS basis. However, STTF is more efficient than FCFS, resulting in a lower number of MRs in the global queue for the same ρ value. (Table 3.4 presents an example of the probability distribution of the number of MRs in the global queue.) Since the $M/M/c$ model overestimates the number of MRs in the global queue, the analytic model overestimates the diagonal e_{ij}^{B} values.

Table 3.3a: PP Allocation Error for DID in LO1 (7 devices, $\rho = 0.90$, and Flow 1)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	8.764	-0.414	-0.210	-2.382	-1.383	-0.637	-1.631	-0.768	-0.468	-0.870
2	1.069	8.572	-1.438	-2.331	-1.831	-0.528	-1.457	-0.760	-0.412	-0.883
3	-1.221	-0.439	7.546	0.720	-0.863	-0.933	-1.617	-1.315	-0.675	-1.203
4	-0.937	0.902	-0.569	8.312	-1.433	-1.073	-1.913	-1.421	-0.658	-1.210
5	-1.115	-0.528	-0.805	-0.110	8.758	-1.854	0.610	-2.489	-0.929	-1.537
6	-3.428	-1.665	-1.467	-0.846	-1.595	10.890	-0.677	-0.661	-0.116	-0.435
7	-1.700	-1.192	-0.464	-0.006	-0.723	-1.556	8.890	-3.512	-0.001	0.264
8	-2.415	-1.811	-1.110	-0.368	-0.743	-1.950	-1.688	12.513	-0.872	-1.556
9	-2.198	-2.032	-1.213	-0.239	-0.885	4.272	-1.883	-2.822	7.531	-0.530
10	-2.005	-1.639	-0.782	-0.126	-0.693	0.199	-1.180	-3.090	1.891	7.426

Table 3.3b: PP Allocation Error for SID in LO1 (7 devices, $\rho = 0.60$, and Flow 1)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	2.239	0.012	1.447	0.753	0.036	-0.731	-0.599	-1.828	-0.536	-0.795
2	1.860	1.503	0.911	0.759	-0.004	-0.751	-1.038	-1.921	-0.498	-0.821
3	-0.406	0.250	2.840	1.787	-0.467	-0.863	-0.938	-1.317	-0.357	-0.530
4	0.487	1.595	0.071	2.793	-0.257	-0.572	-1.128	-1.819	-0.379	-0.791
5	-1.406	-0.989	-1.443	-1.154	2.301	-0.047	1.353	0.436	0.272	0.677
6	0.545	1.191	0.249	0.174	-0.316	1.330	-0.831	-1.427	-0.390	-0.525
7	-1.401	-0.919	-1.480	-1.206	-0.284	0.564	1.514	1.183	0.723	1.306
8	-1.002	-0.646	-1.086	-0.881	-0.614	-0.645	0.816	4.079	-0.082	0.060
9	-0.196	0.280	-0.427	-0.417	-0.150	1.817	-1.516	-0.123	0.563	0.168
10	-0.697	-0.524	-0.893	-0.975	-0.277	1.317	-0.298	0.524	0.871	0.951

On the other hand, the PP allocation error for SID is slightly higher than that of DID in systems with a large number of devices and low utilization ($\rho = 0.60$). We investigate this error by examining the PP allocation error in e_{ij}^{I} for $\rho = 0.60$ (see Table 3.3b). The model overestimates the probability that the MR finds an idle device at the first station (i.e., the diagonal values), which results in underestimating α_e.

We believe the above slight overestimation in the diagonal values of e_{ij}^{I} is caused by the Erlang C formula to estimate the proportion of DID versus SID. Table 3.5 shows that the Erlang C formula overestimates the proportion of SID at low ρ values and a large number of devices. This results in overestimating the number of idle devices, which causes the model to overestimate the diagonal e_{ij}^{I} values.

Table 3.4: Probability Distribution of the Number of MRs in the Global Queue

	p_m (for LO1, 7 Devices, Flow 1, and ρ = 0.90)		
m	$M/M/c$	FCFS (SIM)	STTF (SIM)
0	0.352	0.353	0.385
1	0.072	0.068	0.099
2	0.065	0.061	0.095
3	0.058	0.055	0.085
4	0.052	0.049	0.074
5	0.047	0.044	0.062
6	0.043	0.040	0.050
7	0.038	0.035	0.040
8	0.034	0.031	0.030
9	0.031	0.028	0.023
10	0.028	0.025	0.017
…	…	…	…
E(m)	**6.480**	**6.120**	**2.872**

Table 3.5: DID vs SID for Layout 3 and Flow 1 under STTF

		ρ = 0.90		ρ = 0.80		ρ = 0.60	
		SIM	E_C	SIM	E_C	SIM	E_C
3 devices	ρ	0.901 ± 0.005		0.801 ± 0.005		0.606 ± 0.005	
	Pr (DID)	0.810 ± 0.010	0.819	0.645 ± 0.008	0.649	0.369 ± 0.011	0.362
	Pr (SID)	0.190 ± 0.010	0.181	0.355 ± 0.008	0.351	0.631 ± 0.011	0.638
7 devices	ρ	0.899 ± 0.006		0.794 ± 0.008		0.600 ± 0.006	
	Pr (DID)	0.708 ± 0.010	0.718	0.482 ± 0.014	0.473	0.193 ± 0.007	0.165
	Pr (SID)	0.292 ± 0.010	0.282	0.518 ± 0.014	0.527	0.807 ± 0.007	0.835

3.3.2 Stability Tests

Next we evaluate how well the analytic model performs in determining whether or not a system is stable. The experiment for the stability test is the same as in section 2.5.3. We first set the device speed such that the utilization is high but the system is clearly stable (case C1), and then generate more cases by decreasing the device speed (up to case C6). The M-M LB, analytic, and simulation results are shown for each case. The experiment is conducted for LO1 and LO3, with 7 devices and flow sets 1 and 2. The simulation model is run for only one replication with one million loaded trips, but it is terminated sooner if the system is overloaded.

The results in Tables 3.6a and 3.6b show that in most cases, the analytic model converges when the simulation result suggests that the system is stable (i.e., it is not overloaded). Likewise, when the simulation model suggests that the system is overloaded, the iterative algorithm is forced to stop since $\rho > 0.999$ (UNS?). Although there is one case where the iterative algorithm converges whereas the simulation indicates that the system is overloaded, the ρ value that the algorithm converges to is very high, indicating that the system may be borderline unstable. Overall, our empirical results indicate that the analytic model yields results that are similar to the simulation model in terms of determining the stability of the system.

Table 3.6a: Stability Test for LO1

		Flow 1			**Flow 2**		
		M-M LB	**ANA**	**SIM**	**M-M LB**	**ANA**	**SIM**
C1	α_e	0.15126	0.24793	0.28258	0.27425	0.32578	0.34638
	α_f	0.71335	0.71335	0.71451	0.64638	0.64638	0.64692
	ϱ	0.86461	0.96128	0.99709	0.92063	0.97216	0.99331
C2	α_e	0.15429	0.23744	0.26914	0.27942	0.31913	0.33665
	α_f	0.72762	0.72762	0.72878	0.65858	0.65858	0.65916
	ϱ	0.88190	0.96506	0.99792	0.93801	0.97771	0.99581
C3	α_e	0.15743	0.22625	0.25695	0.28480	0.31182	0.32637
	α_f	0.74247	0.74247	0.74229	0.67125	0.67125	0.67164
	ϱ	0.89990	0.96871	0.99924	0.95604	0.98306	0.99801
C4	α_e	0.16071	0.21431	0.24168	0.29038	0.30370	0.31566
	α_f	0.75794	0.75794	0.75822	0.68441	0.68441	0.68427
	ϱ	0.91865	0.97225	0.99991	0.97479	0.98811	0.99992
C5	α_e	0.16770	0.18804	0.21057	0.29619	UNS?	OL
	α_f	0.79089	0.79089	0.78943	0.69810	UNS?	OL
	ϱ	0.95859	0.97893	1.00000	0.99429	UNS?	OL
C6	α_e	0.17532	UNS?	OL	0.30224	UNS?	OL
	α_f	0.82684	UNS?	OL	0.71234	UNS?	OL
	ϱ	1.00216	UNS?	OL	1.01458	UNS?	OL

Table 3.6b: Stability Test for LO3

		Flow 1			**Flow 2**		
		M-M LB	**ANA**	**SIM**	**M-M LB**	**ANA**	**SIM**
C1	α_e	0.05238	0.23142	0.25380	0.55483	0.55483	0.55663
	α_f	0.74615	0.74615	0.74600	0.18395	0.39264	0.43581
	ϱ	0.79853	0.97758	0.99980	0.73878	0.94747	0.99244
C2	α_e	0.05448	0.20584	0.22356	0.57115	0.57115	0.57225
	α_f	0.77600	0.77600	0.77639	0.18936	0.38280	0.42300
	ϱ	0.83048	0.98184	0.99995	0.76050	0.95395	0.99525
C3	α_e	0.05675	0.17753	0.19258	0.62642	0.62642	0.62131
	α_f	0.80833	0.80833	0.80741	0.20768	0.34453	0.37868
	ϱ	0.86508	0.98586	0.99999	0.83410	0.97095	1.00000
C4	α_e	0.05921	0.14616	0.15785	0.66962	0.66962	0.66837
	α_f	0.84348	0.84348	0.84215	0.22200	0.31150	0.33163
	ϱ	0.90269	0.98964	1.00000	0.89163	0.98112	1.00000
C5	α_e	0.06190	0.11133	0.11948	0.71922	0.71922	OL
	α_f	0.88182	0.88182	0.88052	0.23845	0.27091	OL
	ϱ	0.94372	0.99315	1.00000	0.95767	0.99013	OL
C6	α_e	0.06810	UNS?	OL	0.80913	UNS?	OL
	α_f	0.97000	UNS?	OL	0.26825	UNS?	OL
	ϱ	1.03810	UNS?	OL	1.07738	UNS?	OL

3.4 The MR Wait Times under STTF

The STTF is an efficient rule that reduces α_e and ρ, which also decreases the average MR wait time. However, as mentioned previously, under STTF, some MRs are penalized and may experience excessive wait times. To address this concern, de Koster et al. (2004) proposed using STTF with a threshold such that a MR that has been waiting longer than the threshold, is given a higher priority. The authors propose to set the threshold equal to k times the expected MR wait time. Although the authors demonstrate that this method decreases the maximum MR wait time, an initial simulation run is required in order to determine the expected MR wait time, and further experimentation is needed to determine the appropriate k value. (The authors suggest $k = 4$ or $k = 5$).

Also, it is not clear which MR is served if there are two or more MRs above the threshold. If the oldest MR is served (i.e., FCFS), it may further burden the system, especially if it is a busy system with high device utilization and multiple MRs in the global queue with long wait times.

3.4.1 Bounded STTF

Instead of setting a threshold which is a multiple of the expected MR wait time, we propose to use the number of MRs served as the threshold. Since we can estimate the expected service time per MR using the analytic model we developed for STTF, no simulation results are needed to set the value of the threshold. (Note that there are no analytic models to estimate the expected MR wait times under STTF.) The proposed rule, namely, the bounded-STTF rule is denoted by B-STTF. The bound, denoted by β, acts as the threshold; that is, if β MRs have been served since MR x joined the global queue, then MR x is the next MR to be served. If there are two or more MRs that reach the limit, the closest of those MRs is the next MR to be served.

Since the threshold is based on the number of MRs served, the maximum wait time a MR would experience under B-STTF can be *approximated* by $(\beta)(ST)/D$, where ST is the expected service time. For example, in a 2-device system with $ST = 100$ seconds, if we set $\beta = 10$, then we would expect that, on average, the limit is reached when a MR has been waiting for more than 500 seconds. Since the expected service time can be computed with the analytic model presented in this chapter, and the user can compute the desired β value, we believe our approach is more practical and requires minimal experimentation.

3.4.2 Performance of the B-STTF Rule

We evaluate the performance of the B-STTF rule through simulation experiments. It is important to note that STTF may not always lead to excessive MR wait times, as it largely depends on the layout and the flow data. In our case, flow set 2 in LO2 and LO3 indicated excessive MR wait times since the maximum MR wait time under STTF was significantly larger than the maximum MR wait time under FCFS (for the same ρ value). Hence, we used LO2 and flow set 2 with 3 devices, and LO3 and flow set 2 with 7 devices to test the B-STTF rule.

Since STTF is more efficient that FCFS, comparing their performances directly can be misleading, as the ρ value under FCFS is often much higher (or the system is unstable). Therefore, we included in the comparison a case where the device speed is increased such that the ρ value under FCFS is the same as the ρ value under STTF. Instead of reporting only the average and the maximum MR wait time, we also include the average wait time of the MRs that are in the top 0.5%, 1% and 5% of the MR wait times. Although the user would select the appropriate β value, for demonstration purposes we used $\beta = 14$ and $\beta = 28$ for both layouts. We also report the percentage of MRs that reach the limit, and the average and maximum number of "slips" (Larson, 1987). Using Kingman's formula (1961), it is straightforward to analytically estimate the average MR wait time (W_q) under FCFS. Using Kingman's formula and the results shown in this chapter for STTF, we can also analytically estimate W_q under STTF (although it may not be an accurate estimate since, under STTF, the sequence of service impacts the empty travel portion of the service times).

The results, presented in Table 3.7a and 3.7b, show that, in general, imposing the threshold has no significant impact on the ρ values, while the maximum wait time and the average wait time of the top 0.5% decrease dramatically, indicating that B-STTF works quite well. Also, there is a small decrease in the average wait time of the top 1%, and little to no significant decrease in the average wait time of the top 5%, indicating that excessive wait times occur mostly at the upper 1% tail and higher.

In busy systems (i.e., high ρ value), W_q slightly increases when the limit is set at a smaller value. This is because the MRs in a busy system are more likely to experience excessive wait times, resulting in a larger number of MRs reaching the limit. Thus, in a busy system, B-STTF is very effective; it reduces the maximum wait time as well as the average wait time for the top 0.5% and 1% (and to a lesser extent, the top 5%). If the β value is set too low, more MRs are likely to reach the threshold. Since the closest of such MRs is served, the B-STTF rule is likely to perform similarly to the regular, unbounded STTF rule as the β value decreases.

Table 3.7a: B-STTF for LO2, Flow Set 2 and 3 Devices

			FCFS	FCFS (inc. speed)	B-STTF (14 β)	B-STTF (28 β)	STTF
90	SIM	W_q	UNS	166.14 ± 38.72	94.39 ± 7.82	84.84 ± 5.96	82.86 ± 5.00
		W_q of top 5%	UNS	762.0 ± 278.7	497.8 ± 51.5	510.3 ± 41.9	527.6 ± 48.2
		W_q of top 1%	UNS	1,024.9 ± 413.0	682.9 ± 65.3	739.24 ± 72.7	930.6 ± 146.2
		W_q of top 0.5%	UNS	1,116.3 ± 480.8	761.9 ± 151.0	804.5 ± 166.4	1,132.7 ± 214.7
		Max Wait Time	UNS	1,417.0 ± 532.6	1,062.7 ± 240.9	1,075.4 ± 269.1	2,702.5 ± 1,584.4
		Avg Slip	UNS	-	1.82 ± 0.14	1.98 ± 0.14	2.06 ± 0.16
		Max Slip	UNS	-	16.70 ± 1.09	29.00 ± 1.34	112.70 ± 62.89
		Service Time	UNS	68.61 ± 0.30	69.05 ± 0.45	68.96 ± 0.43	68.81 ± 0.43
		α_e	UNS	0.491 ± 0.008	0.412 ± 0.003	0.411 ± 0.006	0.410 ± 0.003
		ρ	UNS	0.901 ± 0.011	0.907 ± 0.008	0.906 ± 0.009	0.904 ± 0.006
		DID (%)	UNS	81.78 ± 2.15	82.39 ± 1.53	82.13 ± 1.30	81.89 ± 1.12
		SID (%)	UNS	18.22 ± 2.15	17.61 ± 1.53	17.87 ± 1.30	18.11 ± 1.12
		Limit Reached (%)	UNS	-	5.377 ± 1.156	0.952 ± 0.261	-
	ANA	W_q	UNS	164.24	-	-	116.15
		Service Time	UNS	68.50	-	-	65.80
		α_e	UNS	0.491	-	-	0.371
		ρ	UNS	0.901	-	-	0.865
80	SIM	W_q	195.53 ± 442.28	57.15 ± 6.95	40.65 ± 1.97	39.56 ± 2.56	39.37 ± 2.15
		W_q of top 5%	848.3 ± 220.4	312.9 ± 42.4	265.7 ± 10.8	263.8 ± 13.4	265.4 ± 14.1
		W_q of top 1%	1,162.2 ± 419.6	454.42 ± 85.7	390.5 ± 21.7	432.6 ± 24.3	442.1 ± 28.5
		W_q of top 0.5%	1,259.9 ± 463.3	510.6 ± 110.2	432.9± 29.9	507.2 ± 32.8	525.3± 39.7
		Max Wait Time	1,534.9 ± 3,471.9	680.0 ± 149.7	634.4 ± 157.4	801.0 ± 163.1	1,335.1 ± 505.9
		Avg Slip	-	-	1.11 ± 0.06	1.15 ± 0.09	1.16 ± 0.07
		Max Slip	-	-	15.70 ± 2.15	28.80 ± 0.95	59.30 ± 17.65
		Service Time	69.65 ± 0.65	60.93 ± 0.61	61.33 ± 0.32	61.29 ± 0.29	61.30 ± 0.54
		α_e	0.498 ± 0.010	0.436 ± 0.007	0.389 ± 0.004	0.388 ± 0.004	0.389 ± 0.005
		ρ	0.915 ± 0.016	0.801 ± 0.012	0.806 ± 0.010	0.805 ± 0.009	0.805 ± 0.012
		DID (%)	84.30 ± 2.97	64.67 ± 1.89	65.15 ± 1.56	65.08 ± 1.83	64.97 ± 1.57
		SID (%)	15.70 ± 2.97	35.33 ± 1.89	34.85 ± 1.56	34.92 ± 1.83	35.03 ± 1.57
		Limit Reached (%)	-	-	1.166 ± 0.170	0.111 ± 0.046	-
	ANA	W_q	206.03	58.67	-	-	51.54
		ST	69.72	60.89	-	-	59.30
		α_e	0.500	0.437	-	-	0.363
		ρ	0.917	0.801	-	-	0.780
60	SIM	W_q	19.21 ± 1.36	12.55 ± 0.74	10.75 ± 0.63	10.56 ± 0.72	10.53 ± 0.73
		W_q of top 5%	136.5 ± 11.3	94.5 ± 5.0	98.6 ± 5.2	97.0 ± 5.6	96.6 ± 4.9
		W_q of top 1%	203.9 ± 20.7	144.0 ± 12.3	167.8 ± 12.5	164.1 ± 15.4	163.1 ± 12.3
		W_q of top 0.5%	231.2 ± 29.7	165.6 ± 20.2	199.0 ± 16.4	194.5 ± 23.5	193.3 ± 19.4
		Max Wait Time	370.5 ± 121.5	278.3 ± 101.3	379.4 ± 71.4	426.6 ± 191.1	430.4 ± 204.7
		Avg Slip	-	-	0.54 ± 0.04	0.53 ± 0.04	0.53 ± 0.04
		Max Slip	-	-	14.70 ± 1.53	21.40 ± 8.68	21.50 ± 8.94
		Service Time	50.60 ± 0.27	45.68 ± 0.38	45.77 ± 0.45	45.69 ± 0.44	45.69 ± 0.44
		α_e	0.362 ± 0.005	0.327 ± 0.006	0.299 ± 0.006	0.299 ± 0.005	0.299 ± 0.005
		ρ	0.665 ± 0.008	0.600 ± 0.009	0.601 ± 0.009	0.600 ± 0.009	0.600 ± 0.009
		DID (%)	44.06 ± 1.31	35.21 ± 1.01	36.02 ± 1.40	35.78 ± 1.36	35.76 ± 1.38
		SID (%)	55.94 ± 1.31	64.79 ± 1.01	63.98 ± 1.40	64.22 ± 1.36	64.24 ± 1.38
		Limit Reached (%)	-	-	0.061 ± 0.041	0.001 ± 0.007	-
	ANA	W_q	20.65	12.98	-	-	12.76
		Service Time	50.58	45.68	-	-	45.22
		α_e	0.363	0.328	-	-	0.292
		ρ	0.665	0.601	-	-	0.595

Table 3.7b: B-STTF for LO3, Flow Set 2 and 7 Devices

			FCFS	FCFS (inc. speed)	B-STTF (14 β)	B-STTF (28 β)	STTF
90	SIM	W_q	UNS	129.26 ± 29.69	68.55 ± 4.16	62.03 ± 3.16	60.97 ± 2.97
		W_q of top 5%	UNS	661.5 ± 226.4	394.0 ± 32.0	399.2 ± 28.9	404.0 ± 33.8
		W_q of top 1%	UNS	936.6 ± 417.0	527.6 ± 67.7	596.8 ± 58.2	685.1 ± 69.2
		W_q of top 0.5%	UNS	1,045.1 ± 426.5	577.9 ± 88.0	654.7 ± 92.5	823.9 ± 100.3
		Max Wait Time	UNS	1,512.6 ± 550.9	883.7 ± 190.8	967.7 ± 216.6	2,235.4 ± 642.3
		Avg Slip	UNS	-	1.85 ± 0.11	1.99 ± 0.12	2.04 ± 0.14
		Max Slip	UNS	-	17.50 ± 2.44	29.30 ± 1.09	107.80 ± 42.13
		Service Time	UNS	138.00 ± 0.74	138.26 ± 0.97	137.80 ± 1.12	137.58 ± 0.92
		α_e	UNS	0.570 ± 0.007	0.485 ± 0.004	0.482 ± 0.003	0.480 ± 0.003
		ρ	UNS	0.898 ± 0.010	0.900 ± 0.007	0.897 ± 0.005	0.895 ± 0.006
		DID (%)	UNS	71.41 ± 2.95	71.27 ± 1.79	70.33 ± 1.23	70.15 ± 1.36
		SID (%)	UNS	28.59 ± 2.95	28.73 ± 1.79	29.67 ± 1.23	29.85 ± 1.36
		Limit Reached (%)	UNS	-	4.457 ± 0.506	0.715 ± 0.101	-
	ANA	W_q	UNS	130.55	-	-	77.54
		Service Time	UNS	138.07	-	-	130.82
		α_e	UNS	0.571	-	-	0.436
		ρ	UNS	0.899	-	-	0.851
80	SIM	W_q	856.32 ± 386.14	40.55 ± 5.09	29.75 ± 2.19	28.43 ± 1.78	28.55 ± 1.71
		W_q of top 5%	3,480.3 ± 1,403.4	269.9 ± 62.3	232.5 ± 22.5	230.6 ± 18.4	228.9 ± 10.6
		W_q of top 1%	4,072.5 ± 1,549.0	386.5 ± 109.0	345.9 ± 31.6	384.2 ± 35.3	387.2 ± 21.3
		W_q of top 0.5%	4,241.3 ± 1,574.3	428.5 ± 128.3	383.4 ± 40.2	450.0 ± 43.7	462.1 ± 36.0
		Max Wait Time	4,614.1 ± 1,794.8	680.3 ± 274.4	612.9 ± 120.1	787.4 ± 128.8	1325.4 ± 572.8
		Avg Slip	-	-	1.28 ± 0.10	1.32 ± 0.11	1.33 ± 0.10
		Max Slip	-	-	16.70 ± 1.86	29.20 ± 1.43	64.80 ± 40.62
		Service Time	149.71 ± 0.51	123.08 ± 0.87	122.76 ± 0.47	122.40 ± 0.55	122.47 ± 0.59
		α_e	0.619 ± 0.008	0.509 ± 0.008	0.443 ± 0.004	0.441 ± 0.005	0.441 ± 0.004
		ρ	0.974 ± 0.011	0.801 ± 0.011	0.799 ± 0.007	0.797 ± 0.008	0.797 ± 0.006
		DID (%)	92.39 ± 3.46	48.80 ± 2.46	49.24 ± 1.69	48.76 ± 1.72	48.91 ± 1.54
		SID (%)	7.61 ± 3.46	51.20 ± 2.45	50.76 ± 1.69	51.24 ± 1.72	51.09 ± 1.54
		Limit Reached (%)	-	-	1.199 ± 0.237	0.121 ± 0.049	-
	ANA	W_q	710.94	42.08	-	-	31.74
		ST	149.70	123.10	-	-	117.80
		α_e	0.619	0.509	-	-	0.411
		ρ	0.974	0.801	-	-	0.767
60	SIM	W_q	23.44 ± 2.94	6.00 ± 0.61	6.24 ± 0.62	6.13 ± 0.57	6.09 ± 0.72
		W_q of top 5%	184.5 ± 22.5	67.5 ± 7.7	83.7 ± 8.8	83.2 ± 9.5	82.8 ± 11.5
		W_q of top 1%	276.9 ± 42.8	115.9 ± 20.8	159.4 ± 18.1	160.2 ± 19.9	159.8 ± 22.3
		W_q of top 0.5%	312.9 ± 57.0	136.2 ± 31.1	193.5 ± 20.9	196.7 ± 26.9	195.9 ± 28.7
		Max Wait Time	521.7 ± 137.5	264.4 ± 95.1	409.4 ± 78.8	512.0 ± 130.9	559.1 ± 221.6
		Avg Slip	-	-	0.72 ± 0.08	0.72 ± 0.07	0.72 ± 0.09
		Max Slip	-	-	15.40 ± 1.17	27.00 ± 3.84	33.10 ± 10.63
		Service Time	115.04 ± 0.80	92.12 ± 0.66	91.87 ± 0.49	91.69 ± 0.79	91.58 ± 0.70
		α_e	0.475 ± 0.007	0.381 ± 0.006	0.325 ± 0.005	0.323 ± 0.007	0.323 ± 0.008
		ρ	0.749 ± 0.010	0.599 ± 0.008	0.598 ± 0.009	0.597 ± 0.010	0.596 ± 0.010
		DID (%)	38.44 ± 2.04	16.43 ± 1.21	19.05 ± 1.10	18.84 ± 1.27	18.84 ± 1.39
		SID (%)	61.56 ± 2.04	83.57 ± 1.21	80.95 ± 1.09	81.16 ± 1.27	81.16 ± 1.39
		Limit Reached (%)	-	-	0.085 ± 0.033	0.004 ± 0.006	-
	ANA	W_q	25.53	6.55	-	-	5.34
		Service Time	115.12	92.11	-	-	87.94
		α_e	0.476	0.381	-	-	0.299
		ρ	0.749	0.599	-	-	0.572

On the other hand, in systems with low ρ values, the majority of the MRs are served on the basis of SID, and the MRs rarely experience excessive wait times. As a result, the limit is reached very seldom, and the rule has only a small impact on the maximum wait time and the average of the top 0.5%, 1% and 5% of the MR wait times. In systems with medium ρ values, B-STTF is still effective, albeit less than it is in busy systems. Last, if the β value is set too large, the limit is almost never reached, and the rule will have only a minimal effect, regardless of the device utilization.

3.5 Summary and Conclusions

We extended the technique in Chapter 2 to develop an analytic model for multi-device STTF systems. The model estimates the station-to-station empty trips under DID and SID, which are consequently used to estimate the expected device utilization. To our knowledge, the model presented in this Chapter is the first analytic model that explicitly approximates empty device travel under the STTF rule with multiple devices.

The analytic model performs well in estimating the empty trips, with a median allocation error consistently less than 1.5 percentage points. Although the errors in high-utilization cases are higher than those of medium- and low-utilization cases, the maximum error in α_e is less than about 12%, and the error in ρ is less than about 5%. Additionally, we empirically show that the analytic model performs as well as a simulation model in determining whether or not a system is stable.

We also presented the bounded-STTF rule in order to avoid excessive MR wait times. The user may determine the bound (the β value) based on the expected service time estimates obtained from the analytic model. Although the appropriate β value must be determined by the user, using $\beta = 14$ and $\beta = 28$ as an example, we showed that B-STTF is an effective rule. The bound not only reduces the maximum MR wait time, but it also decreases the average wait time for the top 0.5% and top 1% of the MRs. In cases where the STTF rule does not indicate excessive MR wait times, B-STTF is less effective as one would expect but it does not harm the overall performance of the system.

Multiple directions can be considered for future research. First, it would be desirable to estimate the expected wait time analytically. Kingman's formula (1961) overestimates the expected wait time for efficient dispatching rules. Second, one can investigate alternative bounds for the bounded STTF rule based on the remaining queue space at the MR origin station. (Egbelu and Tanchoco [1984] previously explored the *minimum remaining outgoing queue space rule*, using simulation but it was used as a dispatching rule instead of a bound). Third, the STTF rule can be extended to consider multiple MR priorities. And lastly, instead of random MR arrivals, there may be cases where there is a time window specified for the arrival of each MR. In such cases, it would be desirable to develop a look-ahead dispatching rule and an analytic model to evaluate its performance.

CHAPTER 4

Analysis of Patient Mover Dispatching and Equipment Marshalling Areas: A Simulation Study at the University of Michigan Hospital

4.1 Introduction

Patient movement is an essential function in hospitals. For a variety of reasons, inpatients or outpatients are moved in a hospital from one point to another, such as moving a patient from the ER to a short-stay bed, or moving a patient from their room to a clinic/department (and back at a later time), and so on. (We use the terms clinic and department interchangeably.) Delays in moving patients not only impact the physicians and the staff but they also disrupt the workflow and schedule in the clinics, and they may ultimately impact the quality of care. Even a seemingly innocuous delay in moving a patient being discharged, for example, can delay bed availability, inconvenience the patient and their family, and create congestion in the lobby. (Special cases, such as moving critically ill patients and infection control are beyond the scope of our study.)

While moving patients may not be a time-consuming or resource-intensive task in a small- to medium-sized hospital, it is a major function in large hospitals, which typically have many departments and 500 to 1,000 beds or more (Table 4.1, [Becker’s Hospital Review, 2016]). Also, most large hospitals are multi-floor facilities since they are often located in or near urban areas with limited land. As a result, the distance/time for many moves can be significant, involving both horizontal and vertical (elevator) travel. The number of moves is significant as well. In 2014, nearly 35 million patients were admitted to registered hospitals in the U.S. (AHA, 2016). Per Mongrain (2016), “if each of these patients was (moved) only to their room on admission, to and from one test, and then from their room to the exit, there would be 140 million (patient moves/year).” The University of Michigan Health System (UMHS), for example, which was recently renamed Michigan Medicine, is a group of interconnected, multi-floor buildings in the medical campus, with over 600,000 square meters of indoor space, and over 300,000 patient moves/year.

Table 4.1: Examples of Large Hospitals with Approximately 500 to 1,000 beds

FACILITY	**No. of beds**
Stanford Health Care, Stanford Hospital, Palo Alto, CA	613
Texas Children's Hospital in Houston, TX	650
Rush University Medical Center in Chicago, IL	664
The Mayo Clinic Hospital, Methodist Campus, Rochester, MN	794
Baylor St. Luke's Medical Center, Houston, TX	850
Northwestern Memorial Hospital, Chicago, IL	894
Duke University Hospital, Durham, NC	938
Thomas Jefferson University Hospitals, Philadelphia, PA	951
University of Michigan Medical Center, Ann Arbor, MI	1,059
Beaumont Hospital, Royal Oak, MI	1,070
Virginia Commonwealth University Medical Center, Richmond, VA	1,125
University of Alabama Hospital, Birmingham, AL	1,157
The Mayo Clinic Hospital, Saint Mary’s Campus, Rochester, MN	1,265
The Cleveland Clinic, Main campus, Cleveland, OH	1,400

Proper equipment (EQ) and sanitation are important factors. Each patient is moved by a patient mover (PM), one at a time, either in a wheelchair (WH) or on a gurney (GR), which are staged in equipment marshalling areas (EQMAs) located throughout the hospital. (Certain patients may need additional, specialized equipment. Such equipment are not managed by the PMs and they are beyond the scope of our study.) The equipment required for each move is communicated to the PM. Also, as explained later, certain moves requests (MRs) may have higher priority. The

EQ requirements and the priority of the MRs impact the dispatching decisions and ultimately the performance of the system.

Given the volume of the patient moves, large hospitals (such as UMHS) employ full-time PMs and they use various methods (localized or computer-based/centralized technologies) to dispatch and manage the PMs (Schittekat and Nordlander, 2012). The purpose of this study is to analyze the performance of centralized patient movement systems, and to identify possible improvements by examining alternative PM dispatching rules and measuring their impact on the efficiency of the PMs as well as the MR wait times. Since proper EQ is a key factor, we also investigate the impact of the EQMAs.

After our literature review, in section 4.3 we describe the problem setting and the assumptions for the study. In section 4.4, we present the patient movement process, including EQ considerations. In sections 4.5 and 4.6, we use simulation to analyze alternative dispatching rules and propose a new rule, assuming equal-priority and non-equal-priority moves, respectively. In section 4.7, we investigate the impact of the EQMAs and suggest changes to improve the performance of the system. In section 4.8 we summarize our conclusions and present possible directions for future research. The UMHS hospital is used as a real-world application to demonstrate the concepts and the results. However, our analysis and insights apply to virtually any large hospital that uses dedicated PMs and centralized dispatching.

4.2 Literature Review

There are two types of patient conveyance in healthcare; *inter-facility* patient transport, and *intra-facility* patient movement. Inter-facility transport is performed usually by ambulance, and multiple patients may be transported at one time. (In some cases fixed-wing aircraft or helicopters are used.) Intra-facility moves, on the other hand, are typically performed by PMs, and only one patient is moved at a time, using WHs or GRs. The literature review focuses on studies concerned with intra-facility patient moves. However, as two related topics, we also include in our review inter-facility patient transport as well as material handling dispatching in manufacturing settings.

An early paper by Schall (1988) suggests that a centralized patient movement system with dedicated staff (PMs) can reduce the overall number of staff in a hospital and increase productivity, while maintaining the quality of service. Chen at al. (2005) describe the challenges associated with, and the steps needed to develop, a successful centralized PM dispatching system. The authors stress that effective communications between the clinics, the dispatch system (DS), and the PMs must be established properly, and they recommend using an automated DS. (In centralized systems, the DS is usually computer-based, and it keeps track of all the MRs and PMs in the system.)

As part of a case study, Dershin and Schaik (1993) propose a scheduling and staffing model for the PMs to accommodate fluctuations in demand during the day. The authors develop staggered schedules where the PMs start and end their work shifts at different times in order to match the staff size to the demand level. They also compare a one-way and two-way communication system (between the DS and the PMs). In another case study, conducted at the Vancouver General Hospital, Odegaard et al. (2007a, 2007b) focus on determining the optimal number of PMs based on a staggered schedule. They also investigate the impact of centralized versus decentralized dispatching of the PMs. Turan et al. (2011), on the other hand, model the PM routing and scheduling problem as a static dial-a-ride problem (DARP), where the MR arrivals are known in advance. They report a tractable model for medium-sized hospitals.

Fiegl and Pontow (2009) study patient moving in a hospital in Austria. They develop an online scheduling algorithm, assuming the MR arrivals are known in advance. The average weighted flow time is used as the objective function, where the flow time of a MR is defined as the elapsed time from its arrival to its completion, and the weight reflects the importance of the MR. Schittekat and Nordlander (2012) also propose a scheduling algorithm to assign the MRs to the PMs, with an objective to minimize the MR wait time and the PM idle time. However, the details and performance of the algorithm are not presented.

The DARP was also applied to inter-facility patient transport problems. Beaudy et al. (2010) model the system as a dynamic DARP, where the MRs arrive any time during the day, and each MR has a desired pick-up and drop-off time. Multiple MRs may be assigned to a device (i.e., an ambulance) at one time, and the device may transport multiple patients simultaneously. The authors develop a two-phase heuristic to solve the problem, with an objective of minimizing the operating costs. In the first phase, new MRs, as they arrive, are assigned to an ambulance, adding each MR to the ambulance's list of assignments. In the second phase, a tabu search (Glover, 1989) is used to sequence the pick-up and drop-off points on the list of each ambulance and determine their routes. The heuristic method was applied in a hospital complex in Germany with 100 buildings and a fleet of 11 ambulances operating over a road network of 15 km. Hanne et al. (2009) also model the ambulance-based patient transport problem as a dynamic DARP. The authors present Opti-TRANS, a transportation planning system, designed to support all phases of patient transport, including MR arrival and ambulance routing/scheduling.

Kergosien et al. (2011) study the transport of patients via ambulance within a hospital network in France. They define three types of transport (classic, contagious and medical monitoring) in three types of vehicles. Since certain types of transport can only be performed in a certain type of vehicle, the ambulance crews switch vehicles as needed. The authors use tabu search to solve the scheduling problem with an objective to minimize the total transportation cost.

Device dispatching has also been studied in material handling systems; mostly for Automated Guided Vehicles (AGVs) used in manufacturing. Egbelu and Tanchoco (1984) describe two types of dispatching decisions. When a device delivers a load and becomes empty, selecting the next MR to serve is defined as "device-initiated dispatching" (DID). If there are no MRs in the system, the device becomes idle at its last point of delivery. Conversely, when a MR arrives, if it finds one or more devices idle, selecting an idle device for the MR is defined as "station-initiated dispatching" (SID). If all the devices are busy, the MR is served later when an empty device is assigned to it under DID. A fully-defined dispatching rule must identify the rule used for both DID and SID (see section 2.2).

Using simulation, Egbelu and Tanchoco (1984) compare the performance of various dispatching rules. For DID, they consider the random MR rule, the oldest MR rule (FCFS), the closest MR rule, and the maximum-outgoing-queue-size rule. For SID, they consider the random idle device rule, the closest idle device rule, and the longest-idle device rule (FCFS).

Srinivasan et al. (1994) define "trip-based material handling systems," which consist of one or more devices, operating independently to serve the MRs one at a time. Such systems cover AGV systems and others such as lift trucks and cranes. The authors present an analytic model for a single device operating under the Mod-FCFS rule. Under Mod-FCFS, after delivering a load, the device first checks the current station for a MR. If no MRs are found at its current location, the device serves the oldest MR in the system. Using simulation, the authors show that, depending on the flow data, Mod-FCFS performs better than FCFS but not as good as the shortest-travel-time-first (STTF) rule.

The STTF rule, which employs the closest MR rule for DID, and the closest idle device rule for SID, is a well-known rule used in industrial applications. Material handling system providers such as Savant Automation (2016) and Frog AGV System (2016) use STTF. It is also the most prevalent rule for online taxi dispatching (Jung, 2013). As a result, some studies seek to develop dispatching rules that outperform STTF. For example, Bozer and Yen (1996) propose two alternative rules; namely, the Mod-STTF rule and bidding-based dynamic dispatching. Under the former, a device can be reassigned to another MR during its empty trip. Under the latter, when a MR arrives, each device places a bid based on its current list of MR assignments, and the new MR is either assigned to the device with the lowest bid or it is again offered for bidding at a later time. Hwang and Kim (1998) also propose a bidding-based dispatching rule, where the empty travel time of the device to the MR, and the queue length at the origin and destination of the MR, are included in the bid. Although the above bidding-based rules generally outperform STTF, they are considerably more complicated and they lack the simplicity of STTF.

Using simulation and three real-world settings, de Koster et al. (2004) compare multiple dispatching rules, including STTF, STTF with a time threshold, and Mod-FCFS. The STTF rule with a time threshold is intended to address a weakness of the ordinary STTF rule, where some of the MRs, depending on the flow and the layout, may experience long wait times—also known as "orphaning;" see, for example, Bozer and Yen (1996). To avoid long waits, a MR is given a higher priority when its wait time exceeds a user-specified threshold. The authors show that a well-picked threshold reduces the maximum wait time, with minimal impact on the overall performance of STTF as measured by the expected wait time for all the MRs. They also show that using pre-arrival information for the MRs leads to significant improvements.

In short, a relatively small number of studies address the intra-facility patient movement problem in hospitals. A majority of these studies either focus on the number of the PMs and their (staggered) schedules, or they model the problem as a DARP with known MR arrivals. Although there are numerous studies concerned with device dispatching in material handling systems, the MRs in such systems have no EQ switching requirements and no EQMAs. Also, most of these studies assume the MRs have equal priority (except those that may dispatch devices based on the due dates of the MRs). The intra-building PM dispatching problem we focus on is unique in that it explicitly considers EQ switching, and it takes place in a multi-floor facility with elevators, which is very common. We also explicitly consider MRs with higher priority and we investigate the impact of the EQMAs.

4.3 Problem Setting and Assumptions

In this section we describe the patient movement process and the problem setting. We use UMHS as the application facility. For simplicity, we consider only one building, i.e., the University Hospital building (UHB), which is a multi-floor facility with multiple departments, and patient pick-up and drop-off points in each department. For practical reasons, we do not define each patient bed as a pick-up/drop-off point (since there are hundreds of beds). Instead, we cluster the patient beds and the rooms into departments.

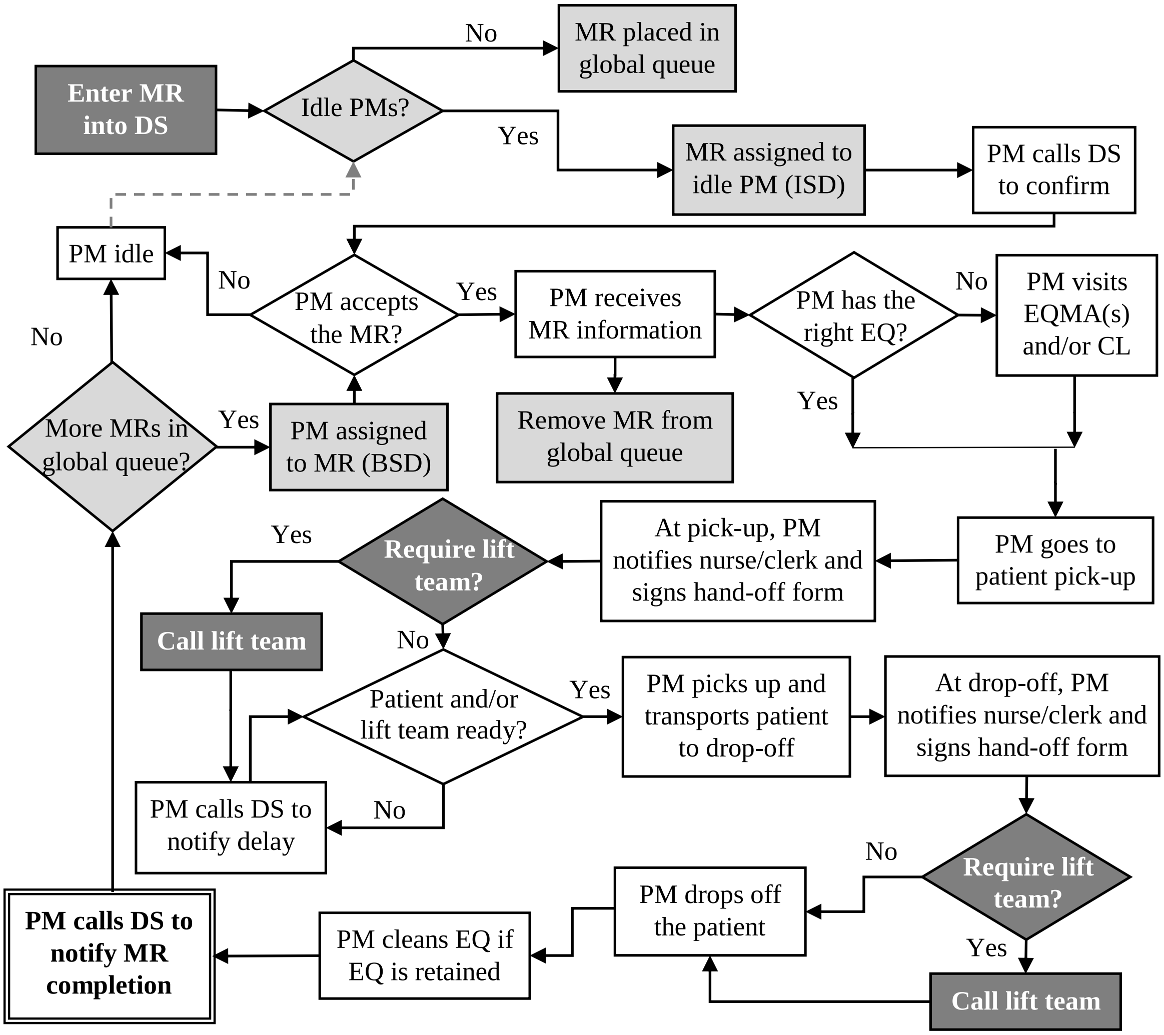
As the need arises, a nurse or clerk enters the MR into the DS, specifying the patient's origin (pick-up point), destination (drop-off point), and the EQ requirement (WH or GR). The MR is automatically placed in a global queue, which maintains the order of arrival of all the MRs. Ideally, the patient needs to be moved (and must be ready to be moved) as soon as the MR is placed in the global queue. However, some of the requests are "appointment MRs," which represent patients who have predetermined appointments. The PMs are responsible for dropping off such patients at their designated department before their appointment time. Since appointment MRs are made well ahead of time, they do not join the global queue immediately. Instead, each appointment MR is automatically entered into the global queue by the DS x minutes before the patient's appointment time. (At UMHS, $x = 30$ minutes.) Once an appointment MR is entered into the global queue, it is treated in the same manner as other MRs. Since look-ahead scheduling strategies are beyond the scope of our study, we assume that all the MRs, including appointment MRs, arrive randomly and one at a time according to an independent Poisson process, and they join the global queue. We also assume that the global queue has an unlimited capacity.

Not all the MRs have equal priority. However, to generalize our results, we present both the equal- and non-equal-priority cases. In the case of UMHS, the priority assigned to an MR is based on the patient's destination, which is consistent with a hospital setting. For example, a patient who needs to be moved to the ICU has priority over a patient who needs to be moved back to his/her hospital bed after a clinical visit.

To serve an MR, the PM must have the appropriate EQ on-hand before picking up the patient. If a WH (or GR) is needed, it is retrieved from one of the wheelchair (or GR) marshalling areas, abbreviated as WHMA (or GRMA). Additionally, a GR needs to be first dressed with clean sheets, which are retrieved from a closet (CL). We assume that WHs/GRs (and clean sheets) are always available in the EQMAs (and CLs). The replenishment of the EQMAs (which is performed by another team) is beyond the scope of our study. Upon retrieving a WH or GR (and visiting the CL in the latter case), the PM travels to the patient pick-up point. The time taken for the PM to retrieve the EQ and travel to the patient pick-up point is defined as the "set-up time." Subsequently, the patient is picked up, moved to his/her destination, and then dropped off, completing the service of the MR. Once service is completed, the PM is available to serve another MR.

If there are no MRs in the global queue, we assume the PM remains at the point of delivery until he/she is assigned to the next MR. There are studies in material handling concerned with "parking" idle devices in designated points; see Egbelu (1993), among others. However, "parking" idle PMs would pose further complications—such as determining the number and locations of the parking point(s), how to allocate the idle PMs among multiple parking points, and what to do if a MR arrives while a PM is traveling towards a parking point. Furthermore, two or more idle PMs congregating at the same point may cause unintended delays. Our assumption above for idle PMs is consistent with UMHS, and ultimately, too many idle PMs imply that there is surplus capacity in the system.

Provided the patient has been prepared and is ready for pick-up, at time of pick-up, the nurse/clerk is notified, the appropriate paperwork is completed, and the patient is transferred onto the EQ. (If the patient is not ready for some reason, the PM notifies the DS and waits; he/she is not permitted to leave and serve another MR.) At the point of patient drop-off, the nurse/clerk is notified, the patient is transferred off the EQ, and the EQ is cleaned by PM (assuming the PM retains the EQ). Since the time needed for the above steps may vary, the patient pick-up and drop-off times are assumed to be exponentially distributed. Furthermore, in some cases, including UMHS, a separate "lift team" is employed to transfer the patient onto/off of the EQ during pick-up or drop-off. We assume that, if needed, the lift team is ready when the PM arrives, and the lift time is included in the patient pick-up and drop-off time. (The above process for moving patients is summarized in Figure 4.1.)



Enter MR into DS
Idle PMs?
No
MR placed in global queue
Yes
MR assigned to idle PM (ISD)
PM calls DS to confirm
PM idle
No
PM accepts the MR?
Yes
PM receives MR information
PM has the right EQ?
No
PM visits EQMA(s) and/or CL
No
More MRs in global queue?
Yes
PM assigned to MR (BSD)
Remove MR from global queue
Yes
Yes
Require lift team?
At pick-up, PM notifies nurse/clerk and signs hand-off form
PM goes to patient pick-up
Call lift team
No
Patient and/or lift team ready?
Yes
PM picks up and transports patient to drop-off
At drop-off, PM notifies nurse/clerk and signs hand-off form
PM calls DS to notify delay
No
No
Require lift team?
PM drops off the patient
PM calls DS to notify MR completion
PM cleans EQ if EQ is retained
Yes
Call lift team

Figure 4.1: Flowchart of the Patient Move Process
(White box – PM, Grey box – DS, Dark Grey box – Nurse/Clerk)

With some MRs, the patient may already be equipped (AEQ) when the MR is placed; i.e., the patient is already in a WH or on a GR. In such cases, the PM does not retrieve any EQ to serve the MR but he/she needs to return to the EQMA any EQ he/she may have from the previous MR. In other cases, the EQ used for moving a patient must stay with the patient when he/she is dropped off. Hence, as shown in Table 4.2, after dropping off a patient, the PM may either have no EQ, or have a WH or a GR, and the next MR may either require a WH or a GR, or the patient is AEQ.

Table 4.2: EQ Status of the PM and EQ Required by the Next MR

EQ status of PM	EQ required by the next MR
1. No EQ	A. WH-AEQ
2. WH	B. GR-AEQ
3. GR	C. WH
	D. GR

Each MR waits in the global queue until a PM is dispatched. (The MR departs from the queue as soon as a PM is dispatched.) However, if the wait time of a MR exceeds a predefined limit, action is taken in order to avoid excessive wait times and negative consequences for the patient or the hospital. If a MR reaches the time limit, it is removed from the global queue, and the patient is moved by a staff member instead of a PM. At UMHS, the limit is 45 minutes, and the patient is moved by the patient transport supervisor.

The trips performed by the PMs may consist of both horizontal and vertical travel (on an elevator). In order to account for possible congestion in the aisles, and the fact that human PMs are manually moving human patients, the horizontal travel times are assumed to be exponentially distributed. The PM travel speed depends on the EQ being used, i.e., a PM with no EQ travels faster than one with a WH, and a PM with a WH travels faster than one with a GR. For vertical travel, we assume that the elevators used by the PMs are dedicated elevators for staff use only. The elevators transport one PM at a time, and the vertical travel speed is assumed to be constant. Also, the wait time for the elevator is assumed to be exponentially distributed. (We did not explicitly simulate the elevators since at UMHS staff other than the PMs use the same elevators and there were no data available on elevator usage by other staff. Therefore, we simulated the wait time of a PM for an elevator instead of simulating the movements of the elevator itself.)

The layout used for the study is based on the UHB; a 10-floor building (floors B2, B1, and 1 through 9, with no floor 3) that houses 37 departments. There are two staff elevators; one at the east side and one at the west side of the building. Floor 9 is accessed only by the west elevator, since the east elevator does not reach the top floor. There are 5 WHMAs, 4 GRMAs and 11 CLs in the UHB. The MAs are located generally next to an elevator. A WHMA is located next to the east elevator on floors 4, 5, and 8, and next to the west elevator on floor 9. A GRMA is located next to the east elevator on floors 6 and 8, and next to the west elevator on floors 4 and 6. One additional WHMA is located in the lobby. The CLs are located on the east and west side of the building on floors 4 through 9. (See Appendix 4.E.)

The patient flow data were obtained from the UMHS database for the first quarter of 2015. Since crew sizing is out of the scope of our study, we do not explicitly model the fluctuation of MR arrivals during different time periods of the day, and instead we model the MR arrivals as a Poisson process with a constant rate. The number of PMs in the system is also assumed to be constant (i.e., no staggered scheduling). The data for the travel distance between the departments, the wait time at the elevators, the dwell time at an EQMA (or CL), and the patient pick-up/drop off times were obtained through observations at the UMHS. The complete data sets are shown in Appendices 4.A through 4.D.

The DS communicates with the PMs via devices such as tablets, house phones, or pagers, and keeps track of the status and location of each PM and MR. Although various technologies exist to implement such a DS, their details are beyond the scope of our study. The interested reader may refer to MediNav (2016) and others that provide indoor navigation/wayfinding and location-based tracking, reporting and analytics.

4.4 PM Set-Up Time

The patient movement system can be modeled as a trip-based material handling system (Srinivasan et al., 1994), where the PM travels “empty” to pick up the patient, and then travels “full” to move the patient. The empty travel time corresponds to the set-up time (as described in the previous section). Assuming that a sufficient number of PMs are provided, we let α_{wp} (< 1), α_{pd} (< 1), and α_{su} (< 1) denote the proportion of time a PM is traveling with a patient, performing patient pick-up/drop off, and traveling empty to perform a set-up trip, respectively. By definition, the expected utilization of a PM, ρ, is equal to $\alpha_{wp} + \alpha_{pd} + \alpha_{su}$.

The set-up trip depends on two factors—the EQ status of the PM, and the EQ required by the next MR. From Table 4.2, there are 12 possible cases (i.e., 1-A, 1-B, 1-C, etc.), and each case involves one or multiple legs (up to four). For example, case 2-C is a one-leg set-up trip, where the PM (located in department i) has a WH, and the MR (located in department j) requires a WH. Hence, the PM travels directly from the drop-off point in department i to the pick-up point in department j. (The PM wipes the WH between the trips.) Case 1-C, on the other hand, is a set-up trip with two legs. The PM has no EQ on hand but the next MR requires a WH. Hence, the PM visits a WHMA to retrieve a WH, and then travels to the patient pick-up point, resulting in ($i \rightarrow WHMA \rightarrow j$) for the set-up trip. An example of a set-up trip with three legs is case 3-C. The PM has a GR but the next MR requires a WH. Hence, the PM travels from department i to a GRMA to deposit the GR, then travels to a WHMA to retrieve a WH before traveling to the patient pick-up point, resulting in ($i \rightarrow GRMA \rightarrow WHMA \rightarrow j$) for the set-up trip. The only case with four legs is case 2-D, where the MR requires a GR but the PM has a WH on hand, which results in ($i \rightarrow WHMA \rightarrow GRMA \rightarrow CL \rightarrow j$).

Since there are multiple EQMAs, one must determine which MA the PM will visit. For example, in the four-leg set-up trip, one must determine which WHMA, GRMA and CL the PM will visit. Instead of solving a traveling salesman problem to determine the optimum route from i to j, we assume that the PM travels to the nearest EQMA or CL from his/her current location (also known as the "nearest neighbor" policy). Moreover, if two locations are on different floors, we assume the PM selects the elevator that minimizes the horizontal travel time between the two locations.

The main function of the DS is to assign a PM to a MR and vice versa. As described earlier, in manufacturing systems this is known as either DID or SID. In our case, it would not be appropriate to refer to the PMs as devices. Therefore, we refer to DID as "busy state dispatching" (BSD) since the PM remains busy after dropping off the previous patient. Likewise, we refer to SID as "idle state dispatching" (ISD) since one or more PM(s) are idle when a MR arrives.

In the next two sections, using simulation, we study the performance of alternative dispatching rules with equal and non-equal MR priorities. The simulation model is based on the Tecnomatix Plant Simulation package (2014).

4.5 Dispatching Rule Comparison, Equal-Priority MRs

Three dispatching rules are considered for the study; namely, FCFS, the University of Michigan Hospital Rule (UMHR), and shortest-set-up-first (SSUF). Recall that under FCFS, the oldest MR in the global queue is assigned to the PM when BSD occurs, and the longest idle PM is assigned to the MR when ISD occurs. Although FCFS is a simple, analytically tractable rule, it is generally less efficient, and is used only as a benchmark. UMHR is based on the dispatching rule currently used at the UMHS. Under UMHR, a numerical value is used to select a MR when BSD occurs. The value is based on the proximity of the PM to the MR. More specifically, if the PM and the MR are in the same department, the proximity value is 16. If the two are not in the same department but are on the same floor, the value is 12. If the two are not on the same floor, but are in the same half of the building, the value is 10. (Floors B2-2 comprise the bottom half, and floors 4-9 comprise the top half.) Finally, if the two are not in the same half of the building, the value is 2. Once the value is determined for each MR, the PM is assigned to the MR with the highest value. For ISD, UMHR assigns the longest idle PM to the MR (same as FCFS). Under the third rule, i.e., SSUF, the MR with the shortest set-up time is assigned to the PM when BSD occurs, and the idle PM with the shortest set-up time is assigned to the MR when ISD occurs. (Note that SSUF is essentially the STTF rule, where the total empty travel time is computed to determine the set-up time.)

The simulation results are based on 10 replications, with 200,000 patient moves per replication, following a warm-up period of 1,000 patient moves. The number of PMs is selected to obtain a reasonable expected utilization value ($\rho \cong 0.80$). Recall that, if a MR waits more than 45 minutes, it is served by the supervisor instead of a PM. In the simulation model, all such MRs are automatically removed from the global queue and from the system; the supervisor is not part of the model.

Table 4.3: Comparison of Dispatching Rules with Equal-Priority MRs

	FCFS	**UMHR**	**SSUF**	**SSUF**
# PM	**8**	**8**	**8**	**7**
$\mathbf{W_q}$	4.01 ± 0.35	2.67 ± 0.18	1.98 ± 0.10	4.36 ± 0.18
SUT	6.84 ± 0.02	6.42 ± 0.03	5.57 ± 0.02	5.51 ± 0.03
$\mathbf{W_q}$ **+ SUT**	10.85 ± 0.36	9.09 ± 0.15	7.55 ± 0.10	9.87 ± 0.16
α_{wp}	0.273 ± 0.002	0.273 ± 0.002	0.273 ± 0.002	0.275 ± 0.002
α_{pd}	0.248 ± 0.002	0.248 ± 0.002	0.248 ± 0.002	0.309 ± 0.002
α_{su}	0.302 ± 0.002	0.283 ± 0.001	0.245 ± 0.002	0.281 ± 0.002
ρ	0.823 ± 0.006	0.804 ± 0.005	0.766 ± 0.005	0.865 ± 0.005
# BSD	98,493 ± 3,442	88,383 ± 2,438	75,354 ± 2233	121,748 ± 2263
# ISD	101,507 ± 3445	11,1617 ± 2,440	124,646 ± 2237	78,252 ± 2266
# MR removed	51.6 ± 33.8	83.0 ± 24.8	292.4 ± 59.0	1,936.6 ± 179.7

The results to compare the three dispatching rules are presented in Table 4.3, where W_q denotes the expected MR wait time in the global queue, and SUT denotes the average set-up time per MR served (in mins). The sum of the two, i.e., $W_q + SUT$, represents the average time a patient waits at his/her point of origin until the PM arrives. The number of occurrences of BSD and ISD are also reported as well as the number of MRs that are removed from the system since they reach the 45-minute limit. The performance of each rule is measured in terms of its efficiency (as measured by W_q and SUT) and effectiveness (as measured by the number of MRs removed from the system).

With 8 PMs, all three rules yield reasonable ρ values (close to 0.80 or below). The SSUF rule performs best in terms of efficiency, and it yields the smallest average patient wait time at the origin. However, it also has the largest number of MRs removed, since it has a tendency to delay serving MRs with remote origins and/or long set-up times. (The STTF rule may delay serving some MRs; see, for example [de Koster et al., 2004] and [Bozer and Yen, 1996]). In contrast, FCFS is the least efficient rule but it has the smallest number of MRs removed. UMHR is more efficient than FCFS but less efficient than SSUF since the proximity value considers the locations of the PM and the MRs but it does not take into account the EQ requirements of the MRs, which may have a significant impact on the empty travel of the PM. Furthermore, when ISD occurs, UMHR selects the longest idle PM just like the FCFS rule. (The W_q values shown in Table 4.3 do not include the wait time of the MRs that were removed from the system. Since the number of such MRs is generally very small, i.e., less than 300 out of 200,000, their impact on W_q would be minimal.)

Given a ρ value of 0.766 under SSUF, one might consider using 7 PMs instead of 8 PMs. We observe in Table 4.3 that, with 7 PMs, both ρ and W_q increase somewhat (as expected), with virtually no impact on SUT. However, there is a significant increase in the number of MRs removed, which suggests that 7 PMs would not be acceptable, although an expected wait time of 4.36 minutes would be considered reasonable in most cases.

SSUF performs well primarily because it includes proximity as well as EQ requirements. For example, a PM with a WH is more likely to be assigned to a MR that requires a WH, provided the MR is not "too far" compared to the other MRs. Given 8 PMs, Table 4.4 shows, for each rule, the proportion of times the PM switches EQ based on the 12 possible cases shown earlier in Table 4.2. We observe in Table 4.4 that indeed less EQ switching occurs under SSUF.

Table 4.4: Proportion of EQ Switching for Each Dispatching Rule

FCFS					UMHR					SSUF				
	WH-AEQ	GR-AEQ	WH	GR		WH-AEQ	GR-AEQ	WH	GR		WH-AEQ	GR-AEQ	WH	GR
No EQ	0.037	0.218	0.074	0.140	No EQ	0.041	0.245	0.064	0.119	No EQ	0.047	0.276	0.059	0.087
WH	0.012	0.074	0.025	0.047	WH	0.013	0.073	0.025	0.046	WH	0.012	0.067	0.055	0.024
GR	0.029	0.173	0.059	0.111	GR	0.024	0.146	0.069	0.133	GR	0.020	0.122	0.043	0.187

Given the impact of EQ switching, one might consider using dedicated PMs by EQ type; that is, WH-PMs would serve only those MRs that require a WH, and GR-PMs would serve only those MRs that require a GR. Such practices are common in manufacturing systems where a lift truck operator, for example, would only move palletized loads, and he/she would not switch devices to move another type of load. However, there are some drawbacks to dedicating the PMs by EQ type. A WH-PM may remain idle even if there are many GR-MRs waiting to be served (and vice versa), and crew sizing may be challenging since variations in MR demand and EQ requirements may be difficult to accommodate, resulting in unequal workloads. From an administrative point of view, having an idle PM while one or more MRs are waiting because the "EQ type did not match" does not seem appropriate for a hospital setting. Furthermore, we note that the SSUF rule explicitly considers the impact of EQ switching, and it avoids EQ switching if it increases the set-up time of the PM relative to other MRs in the global queue.

4.6 Dispatching Rule Comparison, Unequal-Priority MRs

We first compared the rules and showed their efficiency and effectiveness with equal-priority MRs (see previous section) because studying a system with unequal-priority MRs introduces, for BSD, the additional complexity of selecting a MR with a lower priority but shorter set-up time versus a MR of higher priority but longer set-up time.

Although some systems may be set up with multiple levels of priority, we will follow the UMHS model and assume that the MRs have only two levels of priority (which does not eliminate but somewhat simplifies the above selection between two MRs of unequal-priority). That is, we assume there are "priority MRs" (P-MRs) and "regular MRs" (R-MRs). The priority of a MR is determined by the patient's destination as shown in Appendix 4.A.

Using the same layout (i.e., the UHB), we compare the three dispatching rules from section 4.5 and two additional rules. The FCFS and SSUF rules remain unchanged. For BSD, in addition to the proximity value described in the previous section, the UMHR includes a priority value, where a P-MR and a R-MR are assigned a value of 16 and 10, respectively. Furthermore, for every

10 minutes (15 minutes) a P-MR (R-MR) waits in the global queue, its priority value is increased by 2 points. When BSD occurs, the PM is assigned to the MR with the largest value as determined by the sum of the proximity value and priority value. (ISD for the UMHR remains unchanged.)

The two additional dispatching rules, namely, SSUF-2 and SSUF-3, are based on SSUF but the MR priority is part of the dispatching decision. When ISD occurs, by definition there is only one MR in the system, and the priority is irrelevant since there is at least one idle PM. Hence, under both SSUF-2 and SSUF-3, the idle PM with the shortest set-up time is assigned to the MR. However, when BSD occurs, SSUF-2 and SSUF-3 consider both the priority and the set-up time of all the MRs. The decision-making process for both rules are shown in Figures 4.2 and 4.3.

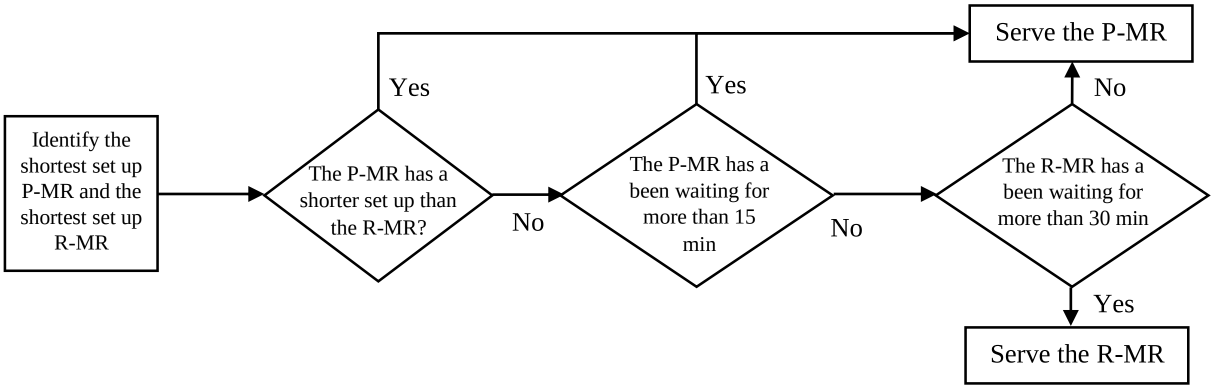

Identify the shortest set up P-MR and the shortest set up R-MR
The P-MR has a shorter set up than the R-MR?
Yes
No
The P-MR has a been waiting for more than 15 min
Yes
No
The R-MR has a been waiting for more than 30 min
No
Yes
Serve the P-MR
Serve the R-MR

Figure 4.2: SSUF-2 Decision Making for BSD

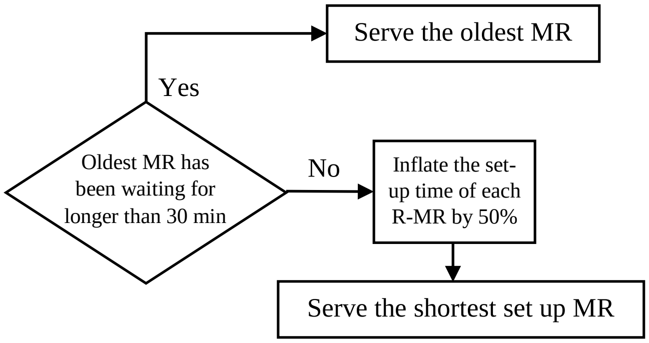

Oldest MR has been waiting for longer than 30 min
Yes
Serve the oldest MR
No
Inflate the set-up time of each R-MR by 50%
Serve the shortest set up MR

Figure 4.3: SSUF-3 Decision Making for BSD

The results of the comparison are shown in Table 4.5, where W_q, SUT, and $W_q + SUT$ are shown for P-MRs and R-MRs. We observe that the proportion of P-MRs and R-MRs are about equal. With 8 PMs, all the dispatching rules yield a reasonable ρ value, with SSUF being the most efficient, and FCFS being the least efficient. We also observe that, in general, the P-MRs have larger SUT values than the R-MRs (which is likely due to the flow data). As a result, SSUF is more likely to serve a R-MR before a P-MR as evidenced by a lower W_q for R-MRs, and a smaller number of R-MRs removed from the system. Obviously, this is an undesirable result for SSUF. Under UMHR, on the other hand, a P-MR is more likely to be served before a R-MR, provided that the P-MR is not located "too far" from the PM. (Since FCFS serves the MRs according to their order of arrival, regardless of priority, both types of MRs have equal W_q values and the same number of MRs are removed from the system.)

While SSUF-2 is relatively efficient, its major drawback is the large number of R-MRs removed, as the rule gives preference to the P-MRs at the expense of the R-MRs. (Among the five rules, SSUF-2 has the lowest W_q value for the P-MRs.) In a sense, SSUF-2 demonstrates the negative consequences of neglecting some of the R-MRs in favor of the P-MRs. In contrast, SSUF-3 immediately serves any MR that has been waiting longer than 30 minutes, and as a result, the number of MRs removed is virtually zero. As intended, inflating the set-up times of the R-MRs makes the rule more likely to serve P-MRs but not to the extent of sacrificing efficiency, which results in SSUF-3 being comparable to SSUF in efficiency.

Table 4.5: Comparison of Dispatching Rules with Unequal-Priority MRs

	FCFS	**UMHR**	**SSUF-2**	**SSUF-3**	**SSUF**	**SSUF-3**
# PM	**8**	**8**	**8**	**8**	**8**	**7**
W_q	4.01 ± 0.35	2.61 ± 0.16	2.40 ± 0.16	2.16 ± 0.15	1.98 ± 0.10	6.55 ± 0.55
P-MR W_q	4.01 ± 0.37	2.36 ± 0.15	1.25 ± 0.06	1.92 ± 0.12	2.32 ± 0.13	5.69 ± 0.48
R-MR W_q	4.00 ± 0.34	2.88 ± 0.19	3.70 ± 0.27	2.44 ± 0.19	1.61 ± 0.08	7.53 ± 0.63
SUT	6.84 ± 0.02	6.45 ± 0.03	5.83 ± 0.02	5.57 ± 0.02	5.57 ± 0.02	5.60 ± 0.03
P-MR SUT	8.05 ± 0.03	7.75 ± 0.03	6.97 ± 0.03	6.73 ± 0.03	6.64 ± 0.03	6.82 ± 0.04
R-MR SUT	5.49 ± 0.03	4.98 ± 0.04	4.55 ± 0.03	4.27 ± 0.03	4.38 ± 0.03	4.24 ± 0.04
Wq + SUT	10.85 ± 0.36	9.05 ± 0.14	8.23 ± 0.16	7.73 ± 0.15	7.55 ± 0.10	12.16 ± 0.55
P-MR [Wq + SUT]	12.06 ± 0.39	10.11 ± 0.14	8.22 ± 0.08	8.64 ± 0.12	8.96 ± 0.12	12.51 ± 0.48
R-MR [Wq + SUT]	9.49 ± 0.35	7.87 ± 0.16	8.25 ± 0.26	6.70 ± 0.20	5.99 ± 0.08	11.77 ± 0.64
α_{wp}	0.273 ± 0.002	0.273 ± 0.002	0.272 ± 0.002	0.273 ± 0.002	0.273 ± 0.002	0.312 ± 0.003
α_{pd}	0.248 ± 0.002	0.248 ± 0.002	0.247 ± 0.002	0.248 ± 0.002	0.248 ± 0.002	0.283 ± 0.002
α_{su}	0.302 ± 0.002	0.284 ± 0.001	0.256 ± 0.002	0.245 ± 0.002	0.245 ± 0.002	0.282 ± 0.002
ρ	0.823 ± 0.006	0.804 ± 0.005	0.776 ± 0.005	0.767 ± 0.006	0.766 ± 0.005	0.876 ± 0.006
# BSD	98,493 ± 3,442	88,222 ± 2,519	80,310 ± 2,413	75,999 ± 2,576	75,354 ± 2,233	128,807 ± 3,023
# ISD	101,507 ± 3,445	111,778 ± 2,520	119,690 ± 2,417	124,001 ± 2,577	124,646 ± 2,237	71,193 ± 3,025
# P-MRs served	105,513 ± 620	105,654 ± 541	105,881 ± 474	105,518 ± 574	105,369 ± 588	105,519 ± 443
# R-MRs served	94,487 ± 618	94,346 ± 542	94,119 ± 472	94,482 ± 574	94,631 ± 589	94,481 ± 443
# MRs removed	51.6 ± 28.5	312.1 ± 62.7	693.6 ± 106.1	6.0 ± 6.6	292.4 ± 52.3	453.9 ± 118.0
# P-MRs removed	26.4 ± 22.4	25.2 ± 14.5	3.7 ± 3.9	2.7 ± 5.4	233.0 ± 49.1	231.6 ± 83.0
# R-MRs removed	25.2 ± 17.7	286.9 ± 60.9	689.9 ± 106.0	3.3 ± 3.7	59.4 ± 18.1	222.3 ± 83.8

The results indicate that SSUF-3 has the best performance; it is comparable in efficiency to SSUF while also being the most effective, i.e., it has the least number of MRs removed. It outperforms UMHR in terms of both efficiency and effectiveness. Furthermore, sensitivity tests performed on the 30-minute threshold we used for the SSUF-3 rule indicate that varying the threshold between 25 and 35 minutes has minimal impact on the performance of the rule. Although SSUF-2 and SSUF are both efficient rules, due to the high number of MRs removed, we conclude that they are less desirable than SSUF-3.

Since SSUF-3 is the best performing rule, we again consider reducing the number of PMs. We observe the same results as in section 4.5; i.e., reducing the number of PMs from 8 to 7 slightly increases the ρ value but it significantly increases W_q and the number of MRs removed.

4.7 Impact of the Configuration of the EQMAs

The set-up time for a MR is influenced not only by the dispatching rule but also the configuration of the EQMAs, i.e., the number and locations of the EQMAs, including the CLs. A carefully-designed EQMA configuration would help reduce the set-up time for the PMs and improve the performance of the system under virtually any dispatching rule. In this section, we propose a method to improve the EQMA configuration and we measure its impact on the performance of the system. A key question to consider is whether or not a better EQMA configuration improves the performance of the system as much as a better dispatching rule improves it. And if so, how one might obtain a better EQMA configuration.

4.7.1 High-Density EQMA Configuration

In the high-density configuration, we maximize the number of EQMAs by placing one WHMA and one GRMA next to each elevator on each floor (while keeping one WHMA in the lobby), which results in 20 WHMAs and 19 GRMAs. Eight *additional* CLs are added to the current configuration bringing the total CLs to 19. (The additional CLs are placed next to each elevator on floors B2 to 2.) No other changes are made; that is, the same elevators are used, and the dispatching rule is the UMHR since we would like to show how much the performance of the system can be improved by a better EQMA configuration without changing the dispatching rule.

The results shown in Table 4.6 indicate that UMHR used with the high-density EQMA configuration outperforms SSUF-3 (the best rule in section 4.6) used with the current EQMA configuration. While this is a significant finding, a large number of EQMAs (and CLs) require more space and is likely to adversely impact EQ availability (and the availability of clean sheets) at each EQMA (and CL). Even if more EQ and clean sheets are provided to avoid potential shortages, replenishment of the EQMAs and the CLs will be more time-consuming and laborintensive since more locations are involved. (If EQ availability and/or the replenishment process deteriorates due the larger number of EQMAs and CLs, any gains expected in the set-up time may diminish since the PMs would be forced to check multiple MAs and/or CLs to locate what they need.)

Hence, although the high-density configuration yields a valuable benchmark, there is strong incentive to look for a configuration that improves the performance of the system by making minimal changes to the current EQMA configuration. In the next section we show that such a configuration can indeed be obtained by identifying the most-visited EQMAs and CLs in the high-density EQMA configuration, which leads us to a "usage-based EQMA configuration."

4.7.2 Usage-based EQMA Configuration

From the high-density EQMA configuration, we select the 5 most-visited WHMAs, and the 5 most-visited GRMAs. Twelve CLs are selected, 11 of which are the same CLs in the current configuration, with an additional CL located on floor B1 at the west elevator (see Appendix 4.E).

As expected, the results in Table 4.6 show that the high-density configuration yields the largest improvement in the efficiency of UMHR. *However, the usage-based configuration also improves the efficiency of UMHR, yielding comparable results to SSUF-3 under the current EQMA configuration.* Although both the high-density configuration and the usage-based configuration significantly reduce the number of MRs removed from the system under UMHR, SSUF-3 under the current EQMA configuration is still more effective.

Lastly, we consider the "best case" scenario, where a better dispatching rule (SSUF-3) is combined with a usage-based EQMA configuration. Since we expect the performance of the system to improve significantly, we reduce the number of PMs from 8 to 7. By comparing the results of the best case scenario (shown in Table 4.6, rightmost column) with the current hospital scenario (Table 4.6, leftmost column), we observe that the best case has a higher W_Q (due to one less PM) but a lower SUT (because it is more efficient). As a result, the best case has a slightly larger expected patient wait time at the origin ($W_q + SUT$), meaning it is less efficient than the current hospital scenario. However, the best case has a smaller number of MRs removed, thus it is more effective. Thus, the results indicate that using a better dispatching rule in conjunction with a better (usage-based) EQMA configuration can help reduce the number of PMs without a significant impact on the performance of the system.

Table 4.6: Impact of EQMA Configuration on Dispatching Rules

	UMHR			**STTF-3**	
EQM Layout	**Current**	**High Density**	**Usage-based**	**Current**	**Usage-based**
# PM	**8**	**8**	**8**	**8**	**7**
W_q	2.61 ± 0.16	1.72 ± 0.12	2.31 ± 0.14	2.16 ± 0.15	5.80 ± 0.50
P-MR W_q	2.36 ± 0.15	1.56 ± 0.11	2.09 ± 0.13	1.92 ± 0.12	4.94 ± 0.44
R-MR W_q	2.88 ± 0.19	1.91 ± 0.14	2.57 ± 0.16	2.44 ± 0.19	6.77 ± 0.57
SUT	6.45 ± 0.03	5.33 ± 0.02	6.10 ± 0.03	5.57 ± 0.02	5.36 ± 0.02
P-MR SUT	7.75 ± 0.03	6.10 ± 0.03	7.21 ± 0.03	6.73 ± 0.03	6.42 ± 0.03
R-MR SUT	4.98 ± 0.04	4.47 ± 0.03	4.86 ± 0.03	4.27 ± 0.03	4.17 ± 0.03
Wq + SUT	9.05 ± 0.14	7.06 ± 0.11	8.42 ± 0.12	7.73 ± 0.15	11.16 ± 0.50
P-MR [Wq + SUT]	10.11 ± 0.14	7.66 ± 0.10	9.30 ± 0.11	8.64 ± 0.12	11.36 ± 0.45
R-MR [Wq + SUT]	7.87 ± 0.16	6.38 ± 0.14	7.43 ± 0.15	6.70 ± 0.20	10.93 ± 0.56
α_{wp}	0.273 ± 0.002	0.273 ± 0.002	0.273 ± 0.002	0.273 ± 0.002	0.312 ± 0.003
α_{pd}	0.248 ± 0.002	0.248 ± 0.002	0.248 ± 0.002	0.248 ± 0.002	0.283 ± 0.002
α_{su}	0.284 ± 0.001	0.235 ± 0.001	0.269 ± 0.002	0.245 ± 0.002	0.269 ± 0.002
ρ	0.804 ± 0.005	0.756 ± 0.005	0.789 ± 0.005	0.767 ± 0.006	0.865 ± 0.006
# BSD	88,222 ± 2,519	69,494 ± 2,441	82,171 ± 2,309	75,999 ± 2,576	123204 ± 3138
# ISD	111,778 ± 2,520	130507 ± 2,446	117,839 ± 2,308	124,001 ± 2,577	76802 ± 3139
# P-MR served	105,654 ± 541	105,561 ± 569	105,662 ± 553	105,518 ± 574	105498 ± 634
# R-MR served	94,346 ± 542	94,439 ± 567	94,338 ± 551	94,482 ± 574	94509 ± 634
# MR removed	312.1 ± 62.7	116.6 ± 33.1	244.9 ± 34.6	6.0 ± 6.6	254.1 ± 68.6
# P-MR removed	25.2 ± 14.5	6.4 ± 6.8	21.2 ± 10.9	2.7 ± 5.4	127.4 ± 44.0
# R-MR removed	286.9 ± 60.9	110.2 ± 32.3	223.7 ± 32.8	3.3 ± 3.7	126.7 ± 52.7

It is important to note that the simulation results in this study are based on 8 PMs. This does not reflect the actual number of PMs at UMHS since our layout and flow data are limited to only one building in the UMHS complex and we do not stagger the start/end times of the PMs. In relation to the latter, our study can be interpreted as representing one time period with 8 PMs, where the time period may correspond to a low, medium, or high demand period for the hospital. Since the relative performance of the dispatching rules are unlikely to vary greatly with the number of PMs (provided that an adequate number of PMs is provided), the results of our study do not depend specifically on which time period is selected. (Some of the results may change if too few or too many PMs are provided; however, in order to reduce cost, the hospital adjusts the number of PMs to match low/high demand periods.) The UMHS data for the complete complex indicate that up to 24 PMs may be engaged during peak periods (i.e., early afternoon hours), and as few as 3 PMs may be engaged during slow periods (i.e., early nighttime hours).

4.8 Summary and Conclusions

We study the intra-building patient movement problem as a trip-based handling system, where each move consists of a set-up trip that depends on the EQ requirements, followed by a "full trip" by which the PM moves the patient. Depending on the EQ status of the PM, and the EQ required by the next MR, the set-up trip may involve one or multiple legs. As a result, the performance of the system is influenced by how the PMs are dispatched as well as the number and locations of EQMAs and CLs. Using UMHS as the application setting, we conduct multiple simulation experiments to analyze the impact of various dispatching rules (including the one used at the hospital) and alternative EQMA configurations on the performance of the patient movement system, with equal and unequal MR priorities.

The performance of each rule and EQMA configuration is measured based on its efficiency (patient wait times and PM set-up times) and effectiveness (number of MRs that are not served within a user-defined time limit). The FCFS rule is the least efficient rule, but due to its nature, it avoids excessive patient wait times. In contrast, SSUF is the most efficient rule, but some patients experience long wait times and they reach the time limit. The rule used at the hospital (UMHR) stands between FCFS and SSUF both in terms of efficiency and effectiveness. We also present and analyze a new dispatching rule; namely, SSUF-3, which is comparable to SSUF in terms of efficiency but at the same time avoids long patient wait times.

In addition, we study the impact of the configuration of the EQMA. We observe that improving the EQMA configuration through careful planning, based on the usage of the MAs, increases the performance of the system as much as, if not more than, a more efficient dispatching rule. By using an improved EQMA configuration together with a more efficient dispatching rule, we show that the number of PMs may be reduced without a significant performance loss in the system. Our results indicate that designing a patient movement system should put equal emphasis on the dispatching rule and the configuration of the EQMAs.

For future research, it would be desirable to extend the study by treating the "appointment MRs" in a different manner. Two possibilities may be considered. In the first case, one can use the appointment information to develop a "look ahead" dispatching rule. In the second case, the dispatching rule is not changed but one may search for the best x value to use for appointment MRs. (Recall in section 4.3 that each appointment MR is automatically entered into the global queue x minutes before the appointment time and that at UMHS, $x = 30$ minutes.) Also, the performance measure for the system may be extended to include the tardiness or earliness of such MRs and the number of appointments missed. Another direction for future research is to develop models to study the replenishment of the MAs (which is performed by a separate team at UMHS) and the availability of EQ (and clean sheets) at the MAs (and the CLs) as well as the number, location, and availability of the elevators.

4.9 Appendices

Appendix 4.A: UMB Departments

Table 4.A1 shows a list of departments at UHB, including the department floors and the department priority as a destination. Note that there are no floor 3 at the UHB.

Table 4.A1: Department Information

Dep Number	Dep Name	Floor	Destination Priority	Dep Number	Dep Name	Floor	Destination Priority
1	MRI	B2	P-MR	**20**	4-D Beds	4	R-MR
2	Oncology	B2	P-MR	**21**	5-A Beds	5	R-MR
3	AMOU	B1	P-MR	**22**	5-B Beds	5	R-MR
4	CT Rooms	B1	P-MR	**23**	5-C Beds	5	R-MR
5	ER	B1	P-MR	**24**	5-D Beds	5	R-MR
6	IR	B1	P-MR	**25**	6-A Beds	6	R-MR
7	Pulmonary	B1	P-MR	**26**	6-B Beds	6	R-MR
8	Ultrasound	B1	P-MR	**27**	6-C Beds	6	R-MR
9	Burn Center	1	P-MR	**28**	6-D Beds	6	R-MR
10	Lobby	1	P-MR	**29**	7-A Beds	7	R-MR
11	OR	1	P-MR	**30**	7-B Beds	7	R-MR
12	PACU	1	P-MR	**31**	7-C Beds	7	R-MR
13	Short stay Beds	1	P-MR	**32**	Dialysis	7	P-MR
14	Apheresis	2	P-MR	**33**	8-A Beds	8	R-MR
15	MPU	2	P-MR	**34**	8-B Beds	8	R-MR
16	Oral Surgery	2	P-MR	**35**	8-C Beds	8	R-MR
17	4-A Beds	4	R-MR	**36**	8-D Beds	8	R-MR
18	4-B Beds	4	R-MR	**37**	9 Beds	9	R-MR
19	4-C Beds	4	R-MR				

The horizontal travel distances (in m) from the department to east/west elevator (recall that floor 9 only has access to west elevator) are shown in Table 4.A2, and the horizontal travel distances between departments that are located on the same floor are shown in Table 4.A3. Some departments have non-zero distances between its pick-up and drop off points. Furthermore, PM horizontal travel speeds depends on the EQ. That is, the PM travels at 0.8 m/s when there is no EQ on-hand, at 0.6 m/s with a WH, and at 0.5 m/s with a GR.

Table 4.A2: Distance from Department to Elevator

Dep	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
East Elevator	60	80	50	25	40	50	130	75	80	110	90	60	70	200	95	220	40	40	70	135	40	40	70	135	40	40	70	135	40	40	70	135	40	40	70	135	-
West Elevator	175	200	150	75	140	50	30	25	135	155	110	80	155	25	240	40	130	80	35	45	130	80	35	45	130	80	35	45	130	80	35	45	130	80	35	45	40

Table 4.A3: Horizontal Distance Between Departments on the Same Floor

		B2		B1						1					2			4				5				6				7				8				9
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
B2	1	0	140																																			
	2	140	0																																			
B1	3			0	75	45	100	170	125																													
	4			75	0	65	30	105	70																													
	5			45	65	0	90	170	115																													
	6			100	30	90	0	80	30																													
	7			170	105	170	80	0	55																													
	8			125	70	115	30	55	0																													
1	9									0	100	160	130	45																								
	10									100	0	180	150	90																								
	11									160	180	0	30	160																								
	12									130	150	30	0	130																								
	13									45	90	160	130	0																								
2	14														0	220	40																					
	15														220	0	240																					
	16														40	240	0																					
4	17																	20	55	100	160																	
	18																	55	20	45	105																	
	19																	100	45	20	60																	
	20																	160	105	60	20																	
5	21																					20	55	100	160													
	22																					55	20	45	105													
	23																					100	45	20	60													
	24																					160	105	60	20													
6	25																									20	55	100	160									
	26																									55	20	45	105									
	27																									100	45	20	60									
	28																									160	105	60	20									
7	29																													20	55	100	160					
	30																													55	20	45	105					
	31																													100	45	20	60					
	32																													160	105	60	0					
8	33																																	20	55	100	160	
	34																																	55	20	45	105	
	35																																	100	45	20	60	
	36																																	160	105	60	20	
9	37																																					35

Appendix 4.B: Dwell Times and Pick-up/Drop off Times

Table 4.B1 shows the average dwell time, in secs, at EQMAs (and CLs) to retrieve/deposit EQ (and clean sheets). The patient pick-up/drop off times, in secs, are shown in Table 4.B2. Recall that the patient can be AEQ for the pick-up, and the EQ may need to stay with patient at drop off. The dwell times at the EQMAs (and CLs) and the patient pick-up/drop off times are assumed to be exponentially distributed.

Table 4.B1: Dwell Times at EQMAs and CLs

	WHMA	GRMA	CL
Dwell Time	11	17	28

Table 4.B2: Patient Pick-Up and Drop Off Times for Each EQ Case

Patient Pick-Up Time	**WH-AEQ**	**GR-AEQ**	**WH**	**GR**
	52	154	156	240
Patient Drop Off Time	**WH (stays w/ patient)**	**GR (stays w/ patient)**	**WH (PM retains WH)**	**GR (PM retains GR)**
	98	117	147	240

Appendix 4.C: Elevator Times

The east and west elevator is assumed to have constant vertical speed, both at 10 seconds/floor. The time taken for a PM to enter and exit an elevator are assumed to be exponentially distributed, with an average of 21 seconds and 10 seconds, respectively. The time waiting for an elevator is also assumed to be exponentially distributed, where the average wait time at the east elevator is 124 seconds, and the average wait time at the west elevator is 92 seconds.

Appendix 4.D: Patient Flow Data

The following tables show the patient flow (of the entire first quarter of 2015) for each EQ case.

Table 4.D1: WH-AEQ Patient; WH Retained by PM at Drop-Off

		B2		B1						1					2			4				5				6				7				8				9	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	out
B2	1										42							54	12		18	6	24	6			12	6			36	12			18			12	258
	2										36								12			6				18		6		6				48					132
B1	3										12																12	6			6			12	6				54
	4										132			6				30	12	6	6		6	12	6	12	12	12		6	12	24		36	24	24	12	18	408
	5																		12			30	30	18		6	72	24		6		6		24	42	12			282
	6										24																			6		12			6				48
	7										12									36		6	12				12	48		60	36	138		36	42	6			444
	8										150			18				18	12	24		30		48		12		12		12	24	24		6	30				420
1	9																		12			66		6		12										6			102
	10																																						0
	11																																						0
	12																					6		6															12
	13																																						0
2	14																	78	36	6			96	48			12					6		42	6	6			336
	15																	6				6																	12
	16													12				6	36	36		6	24				36	36		6	30	96			36	30			390
4	17																																						0
	18																																						0
	19																																						0
	20																																						0
5	21																																						0
	22																																						0
	23																																						0
	24																																						0
6	25																																						0
	26																																						0
	27																																						0
	28																																						0
7	29																																						0
	30																																						0
	31																																						0
	32										78												6			6	18					6		6					120
8	33																																						0
	34																																						0
	35																																						0
	36																																						0
9	37																																						0
	in	0	0	0	0	0	0	0	0	0	486	0	0	36	0	0	0	192	144	108	24	162	198	144	6	66	186	150	0	102	144	324	0	210	210	84	12	30	3018

Table 4.D2: WH-AEQ Patient; WH Stays with Patient at Drop-Off

		B2		B1						1					2			4				5				6				7				8				9	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	out
B2	1			60		12																																	72
	2			6	6	18																																	30
B1	3	42	18		48	6	6	6	12								18																42						198
	4		6	42					6	6																													60
	5		6																																				6
	6		6	6																													6						18
	7				6																																		6
	8			24											6																		6						36
1	9				6		12																																18
	10																																						0
	11																																						0
	12																																						0
	13																																						0
2	14				6																												6						12
	15								6																														6
	16		6	12																																			18
4	17																																						0
	18																																						0
	19																																						0
	20																																						0
5	21																																						0
	22																																						0
	23																																						0
	24																																						0
6	25																																						0
	26																																						0
	27																																						0
	28																																						0
7	29																																						0
	30																																						0
	31																																						0
	32			18		6			6																														30
8	33																																						0
	34																																						0
	35																																						0
	36																																						0
9	37																																						0
	in	42	42	168	72	42	18	6	30	6	0	0	0	0	6	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60	0	0	0	0	0	510

Table 4.D3: WH Patient; WH Retained by PM at Drop-Off

		B2		B1						1					2			4				5				6				7				8				9	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	out
B2	1																																						0
	2																																						0
B1	3																																						0
	4																																						0
	5																																						0
	6																																						0
	7																																						0
	8																																						0
1	9																																						0
	10																	6		12	18		6	36	12		6	30	18	6	12	24		66	6	6	6	6	276
	11																																						0
	12																																						0
	13																		18																				18
2	14																																						0
	15																																						0
	16																																						0
4	17										24																												24
	18										12																												12
	19										12																												12
	20										66							12																					78
5	21										462								6							12									6				486
	22										234																			6				6					246
	23										192																												192
	24										24								12			12		6												12			66
6	25										18																									6			24
	26										210																									24			234
	27										438																			6						6			450
	28										18												60				54	48		30	6	6		36	42	6			306
7	29										360																												360
	30										240																							6		24			270
	31										396																												396
	32																																						0
8	33										240															6													246
	34										96																					6							102
	35										234																24	6			24								288
	36										42											18	6	30		6						6			6	6			120
9	37																																						0
	in	0	0	0	0	0	0	0	0	0	3318	0	0	0	0	0	0	18	36	12	18	30	72	72	12	24	84	84	18	48	42	42	0	114	60	90	6	6	4206

Table 4.D4: WH Patient; WH Stays with Patient at Drop-Off

		B2		B1						1					2			4				5				6				7				8				9	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	out
B2	1																																						0
	2																																						0
B1	3																																						0
	4																																						0
	5																																						0
	6																																						0
	7																																						0
	8																																						0
1	9																																						0
	10	36	36		54	18		24	30							6																	6						210
	11																																						0
	12																																						0
	13				12				6								12																						30
2	14																																						0
	15																																						0
	16																																						0
4	17	60			36				6						96		6																						204
	18	6	12		12				6						54		36																6						132
	19					6		48	24						6		48																						132
	20	18			6																																		24
5	21	6	12			18	6	12	24	6							6																						90
	22	18	6		6		6	42	6						120		36																						240
	23	6			30	12		6	6						66	6	6																6						144
	24				6																																		6
6	25	6	30		12				6																								6						60
	26	6			6	6		42							24		54																18						156
	27	6	12		12	18		120	6								48																						222
	28																																						0
7	29	6			12		6	132	6						12		12																						186
	30	36			30			102	18								30																18						234
	31	12			18			246	18						6		108																						408
	32																																						0
8	33	6	60		24			60	6						66																		12						234
	34	12	6		24	6		108	24						12		42																6						240
	35	6			24			6							12		24																						72
	36				12	6																																	18
9	37	18			18																																		36
	in	264	174	0	354	90	18	948	192	6	0	0	0	0	474	12	468	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	78	0	0	0	0	0	3078

Table 4.D5: GR-AEQ Patient; GR Retained by PM at Drop-Off

		B2		B1						1					2			4				5				6				7				8				9	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	out
B2	1																	218	42	6	16	40	99	30	2	14	98	64	2	51	58	34		75	61	22	3	13	948
	2										8							11	8			8	10		5	18	21	38	2	29	18			101	32				309
B1	3																	5	2			2	13				14	11		3	3	2			3	5			63
	4													10				338	106	74	38	130	238	158	14	62	205	162	13	136	138	133		186	182	157	14	10	2504
	5													3				184	67	19		242	584	138			693	549		178	110	6		346	294	309	2		3724
	6													2				21	16	8		6	45	14	2	5	37	32	2	29	67	34		30	38	21	3		412
	7																			2		2	2					3			2	6							17
	8										10			59				370	310	1027	19	584	664	730	19	170	558	611	10	414	510	518		547	531	418	58	21	8158
1	9										8							2	2			29				2	2									2			47
	10																																						0
	11																	2				2																	4
	12													8				2	10		2	29	3	10			2							6		13		3	88
	13																																						0
2	14										3							5	2				2	5								2		3	2	2			26
	15																		2	2		2	18	10			14	5		6	2				2	3			66
	16																						2				2	2				2							8
4	17																																						0
	18																																						0
	19																																						0
	20																																						0
5	21																																						0
	22																																						0
	23																																						0
	24																																						0
6	25																																						0
	26																																						0
	27																																						0
	28																																						0
7	29																																						0
	30																																						0
	31																																						0
	32										53								2				14	27		3	11	5		3	13	6		10	6	6	2		161
8	33																																						0
	34																																						0
	35																																						0
	36																																						0
9	37																																						0
	in	0	0	0	0	0	0	0	0	0	82	0	0	82	0	0	0	1158	569	1138	75	1076	1694	1122	42	274	1657	1482	29	849	921	743	0	1304	1151	958	82	47	16535

Table 4.D6: GR-AEQ Patient; GR Stays with Patient at Drop-Off

		B2		B1						1					2			4				5				6				7				8				9	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	out
B2	1			51	5				5	5		2												2				2							5	2			79
	2			2	3		2	3	6																								2	6					24
B1	3	54			141		48		269				3						2								3						34		2				556
	4		2	112				2	19	13			8			2	2		2				2	2	3		2			2			5		2				180
	5			2	2		2			2																		2											10
	6		2	43	2	2				18			2			5											5	2		3	3		6						93
	7				2				3							2																							7
	8	6	3	314	21					50			3		5	3			2			5	6	5			3	3		2	10	3	16	8	3				471
1	9	2			19		21		51			2	11									2				5							27				2		142
	10																																						0
	11	2																																					2
	12			3	3				3	8								3	6			11	2	18							3					3			63
	13												2									3																	5
2	14																		3														2						5
	15			2					5																								2						9
	16			2	2																																		4
4	17								2				6																				21						29
	18				2								6		2																		51						61
	19																																35						35
	20												3																				2						5
5	21								3			2	19						5	2						3			2				14						50
	22				3								3													2							106						114
	23	2							6			2	35													3							115			2			165
	24				3								6									2														2			13
6	25																																45						45
	26				2		2						5									2											168						179
	27								2				3																2				122	2					131
	28										2												6		3		2						2		3		16		34
7	29				2		3						3																				106						114
	30		2		2		3		2		2		3																				192						206
	31				2																					2							107						111
	32	2	2	38	13		10		80	26			3		3	2		13	46	35	2	14	99	107		40	165	114	2	90	194	90		43	192	37			1462
8	33		5									2	2						2				2					3			2		53						71
	34	3			2				3			2	2																				203						215
	35	2											5																				37						44
	36												2											3															5
9	37																																						0
	in	73	16	569	231	2	91	5	459	122	4	12	135	0	10	14	2	16	68	37	2	39	117	137	6	55	180	126	6	97	212	93	1473	59	207	46	18	0	4739

Table 4.D7: GR Patient; GR Retained by PM at Drop-Off

		B2		B1						1					2			4				5				6				7				8				9	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	out
B2	1																																						0
	2																																						0
B1	3																																						0
	4																																						0
	5																																						0
	6																																						0
	7																																						0
	8																																						0
1	9																																						0
	10																						2																2
	11																																						0
	12																																						0
	13																		2																	3			5
2	14																																						0
	15																																						0
	16																																						0
4	17										16																												16
	18										2																												2
	19																																						0
	20										14																												14
5	21										5																2									2			9
	22										8																												8
	23										5																									2			7
	24										27											2					2							2	2	2			37
6	25																																						0
	26										3																												3
	27										5																										2		7
	28										98								2			2	32	5			29	38		18	11			16	27		5		283
7	29										16																							2					18
	30										32											2												2					36
	31										16												2			2													20
	32																																						0
8	33										13													2			2	2		2	2				2	2			27
	34										8																												8
	35										3																				2			2					7
	36										11								2				2	2										2					19
9	37																																						0
	in	0	0	0	0	0	0	0	0	0	282	0	0	0	0	0	0	0	6	0	0	6	38	9	0	2	35	40	0	20	15	0	0	26	31	11	7	0	528

Table 4.D8: GR Patient; GR Stays with Patient at Drop-Off

		B2		B1						1					2			4				5				6				7				8				9	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	out
B2	1																																						0
	2																																						0
B1	3																																						0
	4																																						0
	5																																						0
	6																																						0
	7																																						0
	8																																						0
1	9																																						0
	10	2																																					2
	11																																						0
	12																																						0
	13				26		2		14																														42
2	14																																						0
	15																																						0
	16																																						0
4	17	221	13		432		22		283						5																								976
	18	46	10		181		21		262	2			5																				2						529
	19	8			202		10		928							2																	2						1152
	20	19			40		2		18																														79
5	21	37	8		186	2	5		542	3			3			2																							788
	22	102	13		366		51		542				3																				22						1099
	23	34			373		19	2	536				2		2																		26						994
	24	3	6		26		2		13																														50
6	25	16	19		107		5		126								2																5						280
	26	102	21		325	5	50		472																								21						996
	27	61	48		298	2	38		493				5																				6						951
	28	3	2		21		8		6																														40
7	29	54	34		232	3	37	2	330				27																				6						725
	30	61	22		216		72		450				3																				29						853
	31	38			202		46		443						2		2																13						746
	32																																						0
8	33	74	115		291	2	42		462				3		2																		8						999
	34	69	30		294		42		426				2																				24						887
	35	24	3		274		30		320																								10						661
	36	3			27		2	2	40																								2						76
9	37	14			10				21				198																										243
	in	991	344	0	4129	14	506	6	6727	5	0	0	251	0	11	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	176	0	0	0	0	0	13168

Appendix 4.E: UHB Layout

The following figures illustrates the layout of each floor, including the main walkways, elevators, and the location of each department at the UHB. The current EQMAs and usage-based EQMAs configuration are also shown.

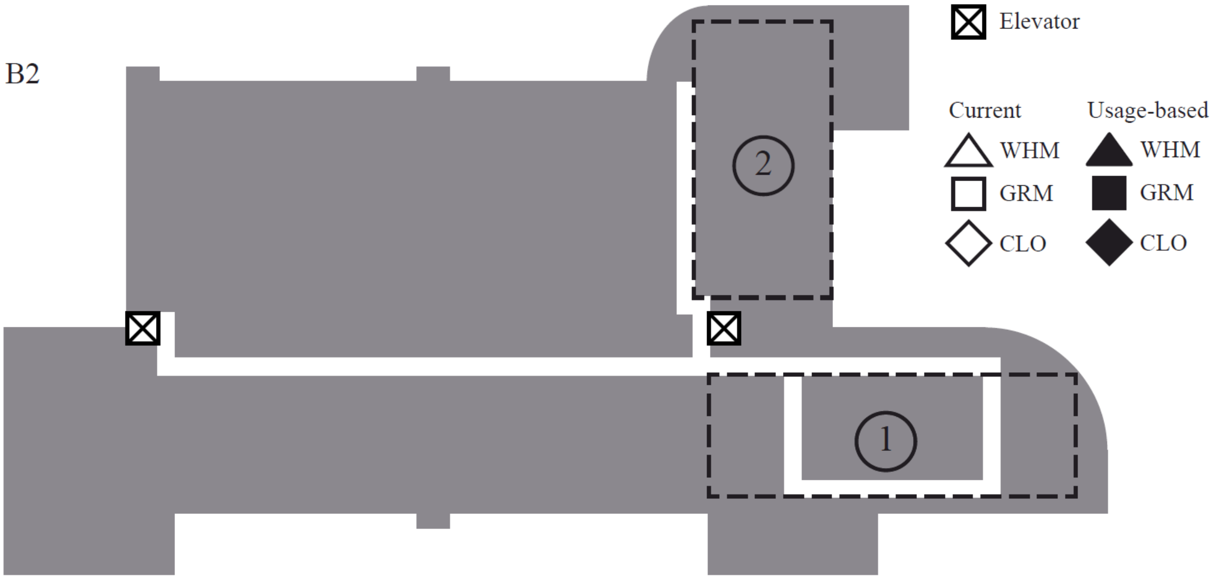


Figure 4.E1: Floor B2 Layout

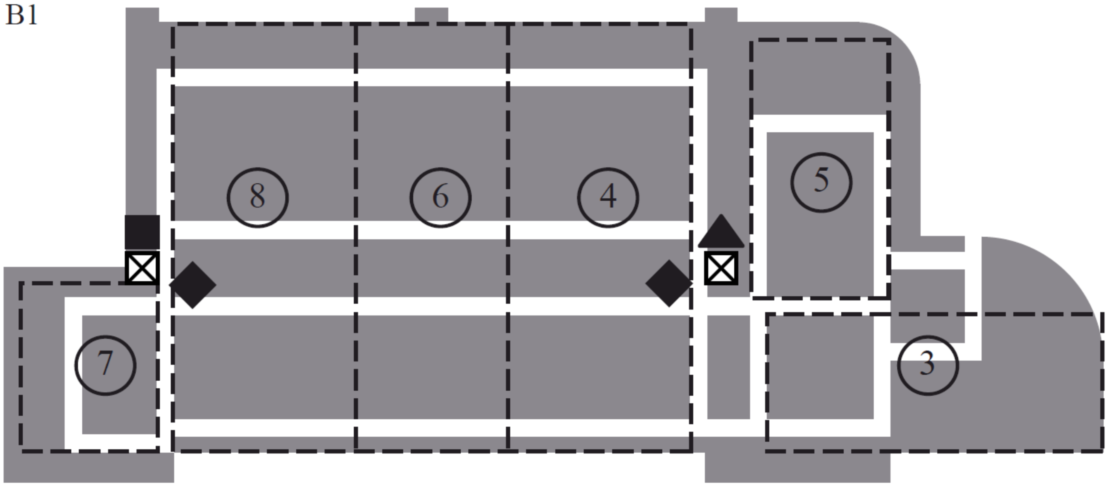


Figure 4.E2: Floor B1 Layout

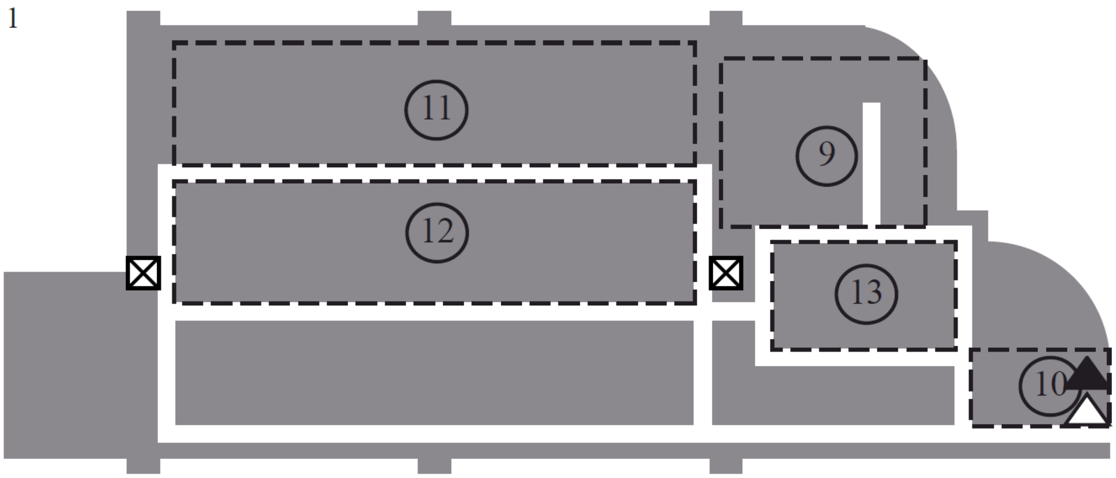


Figure 4.E3: Floor 1 Layout

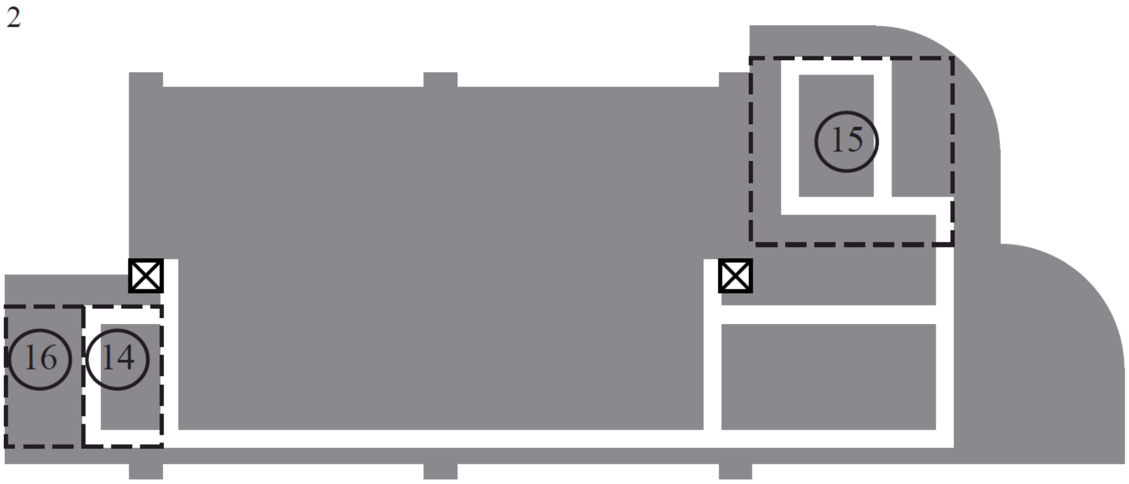


Figure 4.E4: Floor 2 Layout

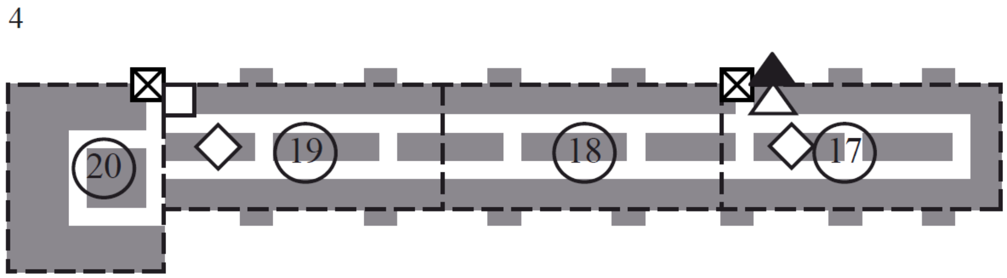


Figure 4.E5: Floor 4 Layout

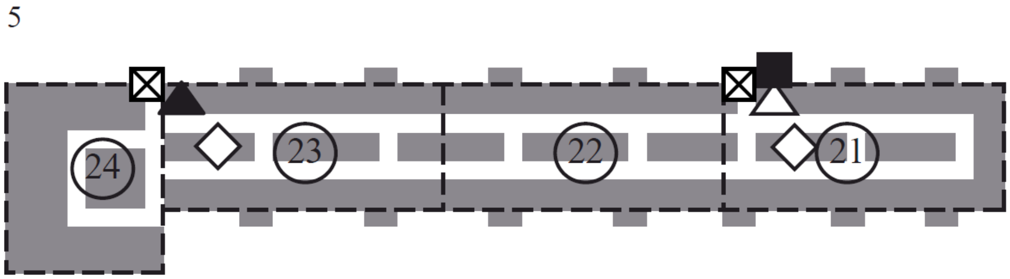


Figure 4.E6: Floor 5 Layout

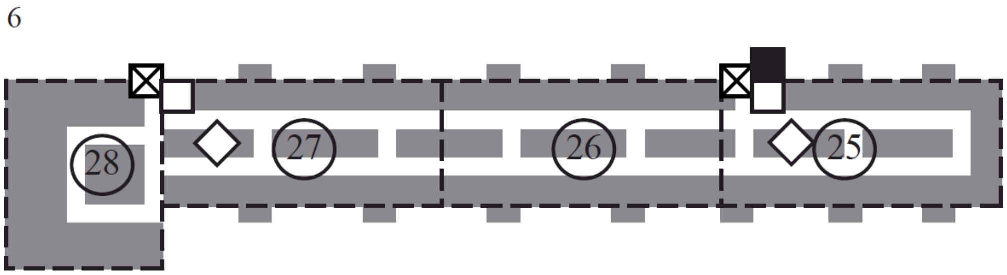


Figure 4.E7: Floor 6 Layout

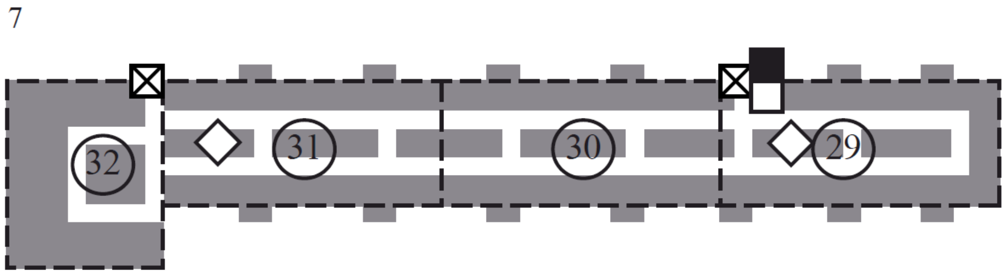


Figure 4.E8: Floor 7 Layout

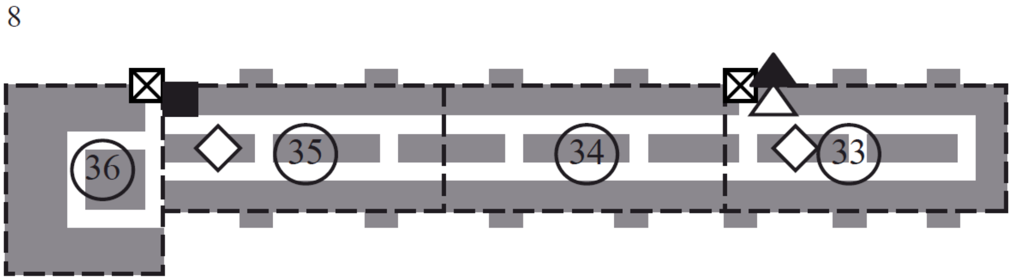


Figure 4.E9: Floor 8 Layout

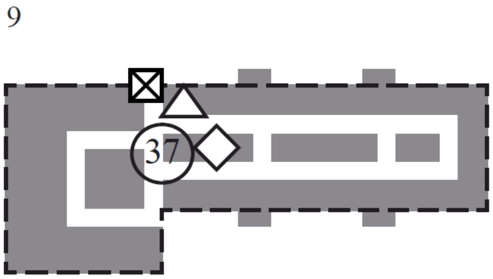


Figure 4.E10: Floor 9 Layout

CHAPTER 5

Overall Summary and Conclusions

In trip-based MHSs, MRs arrive at the system, one at a time, according to a Poisson process, and are served, one at a time, by one of the devices. Depending on the dispatching rule, the analytic evaluation of a trip-based MHS can be very difficult. As a result, such systems are often analyzed through simulation models. We present analytic models for a trip-based MHS in the first two essays of this dissertation. In the third essay, we model an intra-building patient movement system as a trip-based MHS.

In the first essay, we assess the performance of the Mod-FCFS rule against other well-known dispatching rules. Simulation results suggest that Mod-FCFS is a reasonably efficient rule that stands between FCFS and STTF. A new analytic model is developed to estimate empty device travel and the expected device utilization for a multi-device system operating under Mod-FCFS. The analytic technique used in the first essay is extended in the second essay to develop a new analytic model to estimate empty device travel for a multi-device system operating under STTF. Both models are approximate models. However, using simulation to test multiple systems, we show that both of the above analytic models perform well over a wide range of parameter values. We also show that both models can be used to assess the stability of a system.

The STTF rule is a simple and efficient rule but as pointed out in the literature, depending on the layout and the flow data, STTF may lead to excessive MR wait times. Using simulation, we investigated the MR wait times under STTF, and we introduce a new rule, namely, bounded-STTF (B-STTF), by imposing a bound that is based on the number of MRs that are served while another MR is waiting in the global queue. If a MR reaches the above bound, it is selected as the next MR to be served. If two or more MRs have reached the bound, the closest of those MRs is served. The result from the STTF analytic model (i.e., the expected service time per MR under STTF) is used as a guideline to determine the appropriate value of the bound. Using simulation, we show that B-STTF reduces not only the maximum MR wait time, but also the average wait time of the top 0.5% and 1% of the MRs, while achieving results comparable to STTF in terms of efficiency (i.e., the expected wait time across all the MRs and the utilization of the devices).

The performance of a trip-based MHS also depends on the rules used for DID and SID. However, there are conflicting views in the literature on the significance and the frequency of occurrence of DID versus SID. We demonstrate that in most cases both DID and SID are invoked frequently, and one does not strongly dominate the other. In fact, in a well-designed system with a reasonable ρ value ($\rho \cong 0.80$), DID and SID are invoked approximately in equal proportions.

Since the expected MR wait time depends largely on the efficiency of the dispatching rule, estimating the expected wait time is often not straightforward for efficient rules. In terms of future research, it would be desirable to extend the analytic models to estimate the expected MR waiting time. The dispatching rule, and the analytic model, may also be extended to multi-load systems where a device can serve 2 or more loads at a time. Systems with multi-priority MRs may also be of interest.

The third essay is concerned with patient movement systems in large hospitals, which is modeled as a trip-based handling system, where the empty travel time corresponds to the set-up time by the PM. The set-up time involves one or multiple trips (up to 4 legs), depending on the

EQ status of the PM, and the EQ required by the next MR. To study the above system, we conduct a simulation experiment to analyze the impact of alternative dispatching rules and the number and locations of EQMAs, using UMHS hospital building as the application setting.

The performance measure of the system is based on efficiency (expected MR wait times and PM utilization) and effectiveness (excessive MR wait times). We observe that FCFS is a less efficient but effective rule. In contrast, SSUF is more efficient but it is less effective since some of the MRs experience long wait times. We therefore present and evaluate a new dispatching rule that is comparable in efficiency to SSUF while avoiding excessive MR wait times. Furthermore, we observe that a carefully-planned EQMA configuration based on usage may improve the performance of the system as much as, if not more than, a more efficient dispatching rule, which suggests that the dispatching rule and the EQMA configuration both play an important role in the design and operation of a patient movement system.

For future research, it would be desirable to extend the model to treat appointment MRs differently than we have. (Appointment MRs have known arrival times since they are intended for patients with known appointment times.) The performance measure may include tardiness or earliness of such MRs, and the number of appointments missed. Furthermore, it would be worthwhile to develop models to study the replenishment of the EQMAs and/or the operation of the elevators to improve the overall performance of the patient movement system.

BIBLIOGRAPHY

American Hospital Association, 2016. Fast Facts on US Hospitals. www.aha.org/research/rc/stat-studies/fast-facts2016.shtml, accessed on December 2016.

Bartholdi III, J.J., and Platzman, L.K., 1989. Decentralized Control of Automated Guided Vehicles on a Simple Loop. *IIE Transactions*, 21(1), 76-81.

Beaudy, A., Laporte, G., Melo, T., and Nickel. S., 2010. Dynamic Transportation of Patients in Hospitals. *OR Spectrum*, 32(1), 77-107.

Becker's Hospital Review, 2016. 100 Great Hospitals in America 2016. www.beckershospitalreview.com/lists/100-great-hospitals-in-america-2016.html, accessed on December 2016.

Bozer, Y.A., and Srinivasan, M.M., 1991. Tandem Configurations for Automated Guided Vehicle Systems and the Analysis of Single Vehicle Loops. *IIE Transactions*, 23(1), 72-82.

Bozer, Y.A., and Yen, C.K., 1996. Intelligent Dispatching Rules for Trip-Based Material Handling Systems. *Journal of Manufacturing Systems*, 15(4), 226-239.

Chen, L., Gerschman, M., Odegaard, F., Puterman, D.K., Puterman, M.L., and Quee, R., 2005. Designing an Efficient Hospital Porter System [Online Case Study]. *Healthcare Quarterly.*

Chow, W.M. 1986., Design for Line Flexibility." *IIE Transactions*, 18(1), 95-103.

Chow, W.M. 1986., An Analysis of Automated Storage and Retrieval Systems in Manufacturing Assembly Lines, *IIE Transactions*, 18(2), 204-214.

Cooper, R.B. 1981., *Introduction to Queueing Theory, 2nd edition*. New Elsevier North Holland, New York, NY.

Curry, G.L., Peters, B.A., and Lee, M., 2003. Queueing Network Model for a Class of Material-Handling Systems. *International Journal of Production Research*, 41(16), 3901-3920.

de Koster, R., Le-Anh, T., and van der Meer, T.R., 2004. Testing and Classifying Vehicle Dispatching Rules in Three Real-World Settings. *Journal of Operations Management*, 22(4), 369-386.

Dershin, H., and Schaik, M.S., 1993. Quality Improvement for a Hospital Patient Transport System. *Hospital & Health Services Administration*, 38(1), 111-119.

Egbelu, P.J., 1987. The Use of Non-Simulation Approaches in Estimating Vehicle Requirements in an Automated Guided Vehicle Based Transport System. *Material Flow*, 4(1), 17-32.

Egbelu, P.J., 1993. Positioning of Automated Guided Vehicles in a Loop Layout to Improve Response Time. *European Journal of Operation Research*, 71(1), 32-44.

Egbelu, P.J., and Tanchoco, J.M.A., 1984. Characterization of Automatic Guided Vehicle Dispatching Rules. *International Journal of Production Research*, 22(3), 359-374.

Fiegl, C., and Pontow, C., 2009. Online Scheduling of Pick-up and Delivery Tasks in Hospitals. *Journal of Biomedical Informatics*, 42(4), 624-632.

Frog AGV Systems, 2016. AGV Knowledge Center. www.frog.nl/oplossingen/AGV_Kennis_Instituut, accessed on December 2016.

Glover, F.W., 1989. Tabu Search-Part I. *ORSA Journal of Computing*, 1:3, 190-206.

Hanne, T., Melo, T., and Nickel, S., 2009. Bringing Robustness to Patient Flow Management Through Optimized Patient Transports in Hospitals. *Interfaces*, 39, 241-255.

Hwang, H. and Kim, S.W., 1998. Development of Dispatching Rules for Automated Guided Vehicle Systems. *Journal of Manufacturing Systems,* 17(2), 137-143.

Johnson, M.E., and Brandeau, M.L., 1994. An Analytic Model for Design and Analysis of Single-Vehicle Asynchronous Material Handling Systems. *Transportation Science,* 28(4), 337-353.

Jung, J., Jayakrishnan, R., and Park, J.Y., 2013. Design and Modeling of Real-Time Shared Taxi Dispatch Algorithms. *Transportation Research Board Conference*, January. Washington D.C.

Kergosien,Y., Lente, C., Piton, D., and Billaut, J.C., 2011. A Tabu Search Heuristic for the Dynamic Transportation of Patients Between Care Units. *European Journal of Operational Research*, 214(2), 442-452.

Kingman, J.F.C. 1961., The Single Server Queue in Heavy Traffic. *Mathematical Proceedings of the Cambridge Philosophical Society*. 57(4), 902-904.

Kleinrock, L., 1975. *Queueing Systems, Volume 1: Theory*. Wiley-Interscience, New York, NY.

Koo, P.H., and Jang, J., 2002. Vehicle Travel Time Models for AGV Systems Under Various Dispatching Rules. *The International Journal of Flexible Manufacturing Systems*, 14(3), 249-261.

Larson, R.C., 1987. Perspectives in Queues: Social Justice and the Psychology of Queueing. *Operations Research*, 35(6), 895-905.

Larson, R.C., and Odoni, A.R., 2007. *Urban Operations Research*. 2nd edition. Dynamic Ideas, Belmont, MA.

Mahadevan, B., and Narendran, T.T., 1990. Design of an Automated Guided Vehicle-based Material Handling System for a Flexible Manufacturing System. *International Journal of Production Research*, 28(9), 1611-1622.

Mahadevan, B., and Narendran, T.T., 1993. Estimation of Number of AGVs for an FMS: An Analytic Model. *International Journal of Production Research*, 31(7), 1655-1670.

Malmborg, C.J., 1991. Tightened Analytical Bounds on the Impact of Dispatching Rules in Automated Guided Vehicle Systems. *Applied Mathematical Modelling* 15(6), 305-311.

Maxwell, W.L., and Muckstadt, J.A., 1982. Design of Automatic Guided Vehicle Systems. *IIE Transactions*, 14(2), 114-124.

MediNav, 2016. Connexient, www.connexient.com/medinav-2, accessed on December 2016.

Mongrain, C.H., 2016. Patient Transport – Expert Article on Injuries in Healthcare Facilities. *Robson Forensic*, www.robsonforensic.com/articles/patient-transport-expert-witness, accessed on December 2016.

Nazzal, D., and McGinnis, L.F., 2008. Throughput Performance Analysis for Closed-Loop Vehicle Based Material Handling System. *IIE Transactions,* 40(11), 1097-1106.

Nelson, R. 1995. *Probability, Stochastic Processes, and Queueing Theory*. Springer-Verlag, New York, NY.

Odegaard, F., Chen, L., Quee, R., and Puterman, M.L., 2007. Improving the Efficiency of Hospital Porter Services, Part 1: Study Objectives and Results. *Journal of Healthcare Quality*, 29(1), 4-11.

Odegaard, F., Chen, L., Quee, R., and Puterman, M.L., 2007. Improving the Efficiency of Hospital Porter Services, Part 2: Schedule Optimization and Simulation Model. *Journal of Healthcare Quality*, 29(1), 12-18.

Research and Market, 2016. North America Material Handling Market - Forecasts from 2016 to 2021. www.researchandmarkets.com/reports/3841886/north-america-material-handling-market#pos-13, accessed on December 2016.

Savant Automation, 2016. AGV Basics. Available at: www.agvsystems.com/agvs-basics/basics-agvs, accessed on December 2016.

Schall, R.T., 1988. Increased Productivity Through a Centralized Transportation System. *Hospital Materiel Management Quarterly*, 9(4), 77-81.

Schittekat, P., and Nordlander, T.E., 2012. Optimized Patient Transport. *International Conference on Applied Operational Research*, July, Bangkok, Thailand.

Solberg, J.J., 1980. *CAN-Q User's Guide: Optimal Planning of Computerized Manufacturing Systems*. West Lafayette, Indiana: School of Industrial Engineering, Purdue University.

Srinivasan, M.M., Bozer, Y.A., and Cho, M., 1994. Trip-Based Material Handling Systems: Throughput Capacity Analysis. *IIE Transactions*, 26(1), 70-89.

Tanchoco, J. M.A., Egbelu, P.J., and Taghaboni, F., 1987. Determination of the Total Number of Vehicles in an AGV-based Material Transport System. *Material Flow*, 4, 33-51.

Tecnomatix Plant Simulation, 2014. Siemens Product Lifecycle Management Software, Inc. www.plm.automation.siemens.com/en_us, accessed on December 2016.

Telogis, 2016. Fleet Management Software. www.telogis.com/solutions/fleet, accessed December 2016.

Turan, B., Schmid, V., and Doerner, K.F., 2011. Models for Intra-Hospital Patient Routinug. *3rd IEEE International Symposium on Logistics and Industrial Informatics*, 51-60.

Wolff, RW., 1982. Poisson Arrival Sees Time Average. *Operations Research*, 30(2), 223-231.

Wysk, R.A., Egbelu, PJ., Zhou, C., and Ghosh, B.K., 1987. Use of Spread Sheet Analysis for Evaluating AGV Systems. *Material Flow*, 4, 53-64.