Express Shipment Pick-Up and Delivery: Evaluating Airline Recovery Options

by

Jennifer C. Cheung

B.S., University of California at Berkeley (2002)

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Transportation

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2005

Author ....
Department of Civil and Environmental Engineering
May 13, 2005

Certified by ....
Cynthia Barnhart
Professor, Civil and Environmental Engineering and Engineering Systems
Thesis Supervisor

Accepted by ....
Andrew J. Whittle
Chairman, Departmental Committee for Graduate Students

THIS PAGE INTENTIONALLY LEFT BLANK

Express Shipment Pick-Up and Delivery: Evaluating Airline Recovery Options

by

Jennifer C. Cheung

Submitted to the Department of Civil and Environmental Engineering
on May 13, 2005, in partial fulfillment of the
requirements for the degree of
Master of Science in Transportation

Abstract

Irregular operations in the Express Shipment Service Delivery industry require real time solutions that can be implemented to determine routings for aircraft and time-sensitive commodities. During inclement weather, crew unavailability, and mechanical failures, operations personnel use various approaches to recover from disruptions, including rescheduling or canceling flight legs, diverting aircraft and commodities, or missing service all together. We present an optimization approach that can capture and evaluate the effects of different operating policies. Specifically, we compare and contrast three different strategies, namely: 1) *minimizing schedule delay*, 2) *minimizing the number of service failures*, and 3) *minimizing the combined cost of operations and service failures.* We provide proof of concept by implementing our optimization models and evaluating them using several representative scenarios and conducting computational experiments. The solutions, which are highly dependent on user-defined parameters, represent tradeoffs between costs of operations and service failures.

Thesis Supervisor: Cynthia Barnhart
Title: Professor, Civil and Environmental Engineering and Engineering Systems

THIS PAGE INTENTIONALLY LEFT BLANK

Acknowledgments

My gratitude and admiration for my advisor, Cindy Barnhart, has continuously increased with no bound during my graduate studies at MIT. I am most appreciative of her patience and encouragement during each stage of my research. I thank her for being much more than a research advisor to me, and for inspiring and helping me shape my personal and career goals.

I am grateful to the Operations Research team and the Contingency group at the United Parcel Service for the support they have provided me with completing this thesis. I would like to thank Keith Ware, Wynand Depuy, and Kathleen Duane for your interest and support of my research, as well as allowing me the opportunity learn and see so much of your company.

My thesis would not have been possible without the generous advice of Hai, Raphael, Mike, and Carl, who have each selflessly helped me with the technical (and tedious) aspects of my research. I have truly appreciated your time and suggestions. I also thank the incredible group of friends and classmates I have met during my short stay in Boston, whose support and camaraderie have made my graduate school experience as enjoyable and memorable as it has been.

Most of all, I would like to thank Mom, Baba, and Taylor for their unwavering love and faith in each of my endeavors.

THIS PAGE INTENTIONALLY LEFT BLANK

Contents

THIS PAGE INTENTIONALLY LEFT BLANK

List of Figures

THIS PAGE INTENTIONALLY LEFT BLANK

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1

Introduction

The U.S. commercial freight transportation market, which includes trucking, rail, air, water, and pipeline sectors, claimed approximately $702 billion of revenue in 2003 [28]. The largest transportation company in the United States is the United Parcel Service, which primarily handles ground parcels. FedEx Express is the largest player in the domestic air express sector, deriving $7.1 billion in the domestic overnight delivery market [10]. UPS generated 2003 revenues of $33.5 billion, $5.6 billion of which was from its domestic next-day air express service. This translates to delivering over 14 million packages and documents per day, over 2 million which are in the domestic air market. (UPS Fact Sheet [26]) Measured by 2003 volume, FedEx held a 41.5 percent of market share of the US domestic air express traffic, while UPS had 22.0 percent.

While the domestic package delivery service remains competitive, the industry has expanded to the international sector with the acquisition of aviation rights in other countries. In 2003, UPS saw a 128 percent increase in operating profit for its international segment, compared to adjusted 2002 results [25]. Over the past half decade, its operating profits in the international sector have grown by more than 200 percent. With the expansion of their international networks, competition among the leading carriers continues to drive the need for more efficient operations. Now a global package delivery industry, carriers are offering a broader range of services. UPS is the world's 11th largest airline in the world and the ninth largest domestically. Any development or improvement in operations has the potential to yield a decisive advantage for a carrier against its competitors.

Extensive research has been done on improving network design, especially in the service network. The goal of the planning problem is often to create schedules that maximize airline profitability. In recent years, we have seen an increased utilization of optimization techniques in the planning phase of the Express Shipment Delivery Service (ESDS) industry. This has also occurred in parallel with an increased acceptance and adoption of optimization based tools used during operations recovery in the passenger airline industry.

Kim [15] developed an algorithm and implementation of a solvable model for large scale transportation service network design problems. His approach results in annual operating cost savings of tens of millions of dollars, reduces the fleet size required, and decreases the time to develop operating plans. Armacost, et al. [5] created a system to optimize the design of service networks for delivering express packages at UPS. The system is accredited with saving the company over $87 million between 2000 and 2002. Their work has significantly influenced the planning and decision making process at UPS. It is now a central part of determining fleet assignments, aircraft routes, and package routings. Long term advantages of the tool include the potential savings in assessing aircraft acquisition and fleet composition.

The focus of our research is to develop and evaluate real-time recovery solutions for the air network of an ESDS carrier. While the optimization models implemented in designing airline schedules assume accurate departure and arrival times for all packages, planes, and crew members, day-to-day operations often do not progress as planned. Mechanical failures in aircraft, inclement weather, and insufficient capacity cause schedule perturbations that affect a carrier's ability to satisfy their delivery commitments. These schedule aberrations have the potential to significantly increase the carrier's costs on top of the costs of planning as the airline struggles to resolve the situation in real-time. Our objective is to evaluate the options the carrier has available during common schedule irregularities.

1.1 Description of Service Network and Operations at an ESDS Carrier

This section describes the different service types that an ESDS carrier offers, as well as the operations of its physical network.

1.1.1 Objectives

The package delivery industry may be segmented in different ways. Besides international and domestic markets, packages are differentiated by their time-sensitivity. *Expedited* delivery refers to packages that require overnight movement via aircraft. Next-Day Air (NDA) service at UPS is guaranteed overnight by 10:30am for the domestic market at a high premium to its customers. *Deferred* delivery refers to packages that require a delivery of two to three days and may or may not move via aircraft. Second Day Air (SDA) Delivery is guaranteed on-time for the day after next from the time of pick-up.

Within the two time-defined markets, additional segmentation exists in the form of Next-Day Air Early AM, which guarantees service by 8:00am and Next-Day Air Saver, which guarantees service by 3:00pm. For the *deferred* market, UPS also offers 2nd Day Air A.M., which guarantees on-time delivery by noon to destinations also served by the NDA network on the second business day. These services allow UPS to further prioritize its customers in the two time-defined markets.

1.1.2 Service Network

The network of an ESDS carrier is complex in its physical system and scheduling requirements. The movements of the carrier's transportation resources (i.e. aircraft) define its service network. The extremely time sensitive nature of the industry forces the carrier to aggressively set deadlines during its operations. We illustrate the constraints imposed on operations of one carrier's Next-Day Air (NDA) network.

Physical System

The resources for the *expedited* and *deferred* delivery sectors are not mutually exclusive. Aircraft and facilities operate during the day to service *deferred* delivery packages, while night time operations are dedicated solely for *expedited* packages. In addition, the international and domestic operations do not operate in isolation. International flights may feed into the NDA network, and take off as a domestic flight. The problem we research will only consider overnight air operations, with a limited international scope.

In the planning phase, the carrier must decide, based on projections of demand, which routes to service, which fleet types to place on those routes to best match demand, and how to schedule the flights to allow for sufficient time at hubs to make connections.

Express Shipment Delivery Operation

Consider the partial network shown in Figure 1-1. Customers introduce packages to the network at *ground centers*. Ground centers are the origin and destination points for the package commodities. The routing for the commodities is determined after a preliminary sort determined by the package's destination. From the *ground centers*, the packages enter the NDA network at *gateways* via small chartered feeder jets as well as the carrier's integrated ground transportation network (e.g., gateway a). At the gateways, the packages are loaded onto aircraft that are designated on a pickup route.

The carrier specifies *time windows* to meet *level-of-service* requirements. The earliest time an aircraft may depart from its origin gateway is designated as the *Earliest Pickup Time* (EPT). The pickup route may involve stopping at an intermediate gateway (e.g. gateway b) to pick up additional packages, but the termination of the pickup route is always at one of the hubs in the system. The aircraft must arrive at the hub in accordance to the hub's *sort start time*, which designates the latest that the packages can arrive at the hub from a pick-up route and still have the onboard packages sorted and loaded onto outbound aircraft on time.

At the hub, the packages are unloaded and sorted according to destination before being

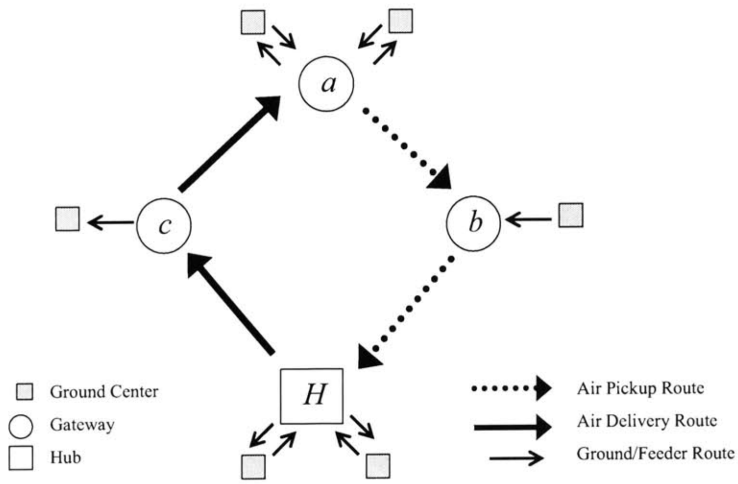

Figure 1-1: Express Shipment Delivery Service Network

reloaded onto outbound aircraft preparing for the night's delivery routes. In order for the packages to satisfy their time commitment, the aircraft designated for the delivery routes cannot depart the hub earlier than its *sort end time.* Sort end is the latest time that packages must be loaded onto aircraft delivery routes. The aircraft returns to the gateway locations, no later than the *Latest Delivery Time (LDT)*, where the packages are further sorted and delivered to the ground centers before making it to their final destination via the ground network.

In the event that UPS fails to attempt delivery within the time published to the customer, transportation charges will be credited or refunded to the customer. (United Parcel Service [27]) Given the tight time window that the overnight package delivery business operates within, as well as the high premiums charged, the NDA network dictates many of the carrier's resource and facility requirements.

Carrier Specific Resources

The NDA network consists of seven air hubs, six of which are regional, and one which is the main U.S. air hub. Regional air hubs are responsible for sorting packages that have an origin and destination somewhere within the same region. Packages destined for a location outside the region are sent to the main hub for processing.

UPS has over 250 aircraft, consisting of 9 fleet types, and charters an additional 300. The chartered aircraft are typically smaller corporate sized jets that feed the packages into the gateways. In the domestic NDA network, 90-100 aircraft operate during the overnight operations, serving approximately 100 domestic gateways. Each aircraft type has operating characteristics that determine the routes it can fly. The fleet type dictates the maximum flying range, effective speed, restrictions on the locations at which it can land, turn around time, and cargo capacity. The service network design problem seeks to minimize the cost of satisfying service requirement while covering all the demand, subject to capacity constraints of the routes and hub sort. In addition, particular gateways may limit the rate at which aircraft arrive and take off on its runways because of the capacity issues and noise restrictions.

In addition to aircraft that are assigned within the schedule, 21 spare aircraft, with a mix of fleet types, are allocated to the Contingency department. These aircraft are positioned throughout the NDA network and are entirely ready for deployment throughout the system should the need rise during the night's operations. For example, if a scheduled aircraft experiences a mechanical failure during operations, the Contingency department may decide to fly the nearest *hot spare* to the gateway to resume operations as scheduled. The carrier's planners continuously evaluate the location of these *hot spare* aircraft in the network to ensure that they are optimally placed to effectively respond to schedule disruptions. Boundary conditions designate the number of aircraft (including spare aircraft) of a certain fleet type required at a gateway at the end of that night's operations.

1.1.3 Irregular Operations

Schedules are generally not executed as planned. In the planning phase, flights are scheduled in a deterministic manner. This idealistic methodology assumes that planes arrive and depart as scheduled. The possibility of weather disruptions, aircraft mechanical failures, overflow of packages on a pick-up route, and delayed packages from a connecting flight are not incorporated in the scheduling decision. In the event that such perturbations do arise, the strategy to return the network to normal operations may involve any combination of rescheduling flight legs, rerouting aircraft, crew and packages. We will survey recent research in resolving airline schedule disruptions in the following chapter.

1.2 Robust Planning

The optimal schedules that planners devise are immediately futile when a schedule perturbation disrupts the plan, preventing it from operating on time. To resolve this situation, dispatchers may attempt to re-optimize tightly planned schedules after irregularities have occurred. Another approach, however, is to explicitly incorporate robustness during the planning phase of schedule design. Rather than only maximizing resource utilization in the schedule, robust schedule planning considers the possibility of schedule disruptions during the planning phase. By proactively addressing the possibility of cancellations and delays in the network before they occur, the negative effects of these disruptions may be alleviated than if the schedule had not been planned with robustness in mind. Incorporating *slack time* (time on the ground where resources are not utilized) might allow for more recovery options because operations personnel will have more time to react to a disruption. The addition of robustness in the schedule will not entirely eliminate the need for real-time recovery solutions in the event of a disruption, but perhaps the aftermath effects will not be as severe. The addition of slack time into the schedule from the initial phase may yield a plan that is not the most profitable if operated as planned, but the benefits may be great in the event of a disruption.

The difficulties in the robust airline schedule planning problem arise in defining the objectives. Robustness may be measured by a variety of metrics, such as minimizing expected delays or operating costs in the event of a disruption. One may assume worst case scenarios and evaluate the planned schedule's performance against a situation that may never happen, or it may target a certain level of service. The amount of robustness to incorporate into the schedule needs to be determined. This is challenging, however, because the cost of potential flight cancellations and propagation of delay compared to the cost of robust planning in the network is not readily apparent. There is a tradeoff between robustness and optimality in the schedule. *Planned* costs are almost always lower than *actual* costs, which include additional operational and crew costs as well as lost revenue. Another challenge with robustness is that incorporating stochasticity into optimization models adds a dimension of complexity to an already large computational problem.

We will discuss recent research on robust airline schedule planning and robust crew planning.

1.2.1 Literature Review-Robust Airline and Aircraft Scheduling

Rather than modeling the crew scheduling problem as deterministic, Yen and Birge [32] consider a stochastic crew scheduling model and devise a solution methodology for integrating disruptions in the evaluation of crew schedules. They use a two-stage stochastic integer programming model develop a branching algorithm to identify expensive flight connections and alternative solutions. The first stage is the crew scheduling problem and the second stage involves including penalty parameters for delays. Their model focuses on minimizing total expected crew costs and the disruption interactions between crew schedules when disruptions perpetuate to other planes in the network.

Ageeva [1] explores the tradeoff between optimality and robustness. Frequently, a number of unique solutions to the aircraft routing/scheduling problem have the same optimal cost. Ageeva creates multiple optimal solutions to the aircraft routing problem and suggests how to evaluate the solutions by comparing the flexibility of these optimal solutions to disruptions.

Schaefer et al. [19] suggest a stochastic extension to the deterministic crew scheduling problem. They provide a lower bound on the expected cost of any crew schedule in operations with disruptions. Their focus is on "push back" recovery heuristic, which delays the departure of each flight until all resources (crews and planes) are available. Using this recovery process, they use a Monte Carlo simulation to evaluate a crew schedule's performance in terms of expected crew pairing costs that more accurately reflect the cost of a pairing in operations with disruptions. They analyze crew schedules to offer insight on the types of crew pairings that perform well in the event of schedule disruptions.

Lan [16] presents two approaches for robust airline schedule planning to achieve minimum passenger disruptions. The first approach involves routing aircraft, and the second involves retiming flight departure times. The author examines the propagation of delay for a sequence of flight legs by rerouting aircraft by formulating the problem as a mixed integer programming problem with stochastically generated inputs. He also presents an algorithmic solution approach to minimize the number of passenger misconnection by re-timing the departure times of flight legs. He establishes time windows for each flight leg to depart within. For example, a 30-minute time window allows a flight leg to depart a maximum of 15 minutes before or after its scheduled departure time. Computational results with data from a major U.S. airline show that when this methodology is passenger delay can be reduced by 20

Kang and Clarke [14] propose a methodology that isolates the effects of disruptions. The authors derive a schedule that partitions flight legs into independent sub-networks, or layers, that isolate a set of itineraries to the impact of schedule perturbations. Models and algorithms are implemented to develop these schedules based on fleet assignment, schedule design, and aircraft maintenance routing considerations. The sub-networks are prioritized by the revenue obtained by operating the sub-network of flight legs. A high priority layer would generate more revenue, and thus be protected from capacity reductions, while less important layers would be the first to experience cancellations in itinerary. By simulating weather disruptions and comparing their degradable schedule to a traditional schedule, Kang and Clarke illustrate the significant effects of preferential ordering between passengers traveling in two layers. In addition, on-time performance measured by the flight legs in the degradable schedule is 5

Rosenberger, Johnson, and Nemhauser [18] present a robust fleet assignment model that emphasizes *hub isolation* and *short cycles.* Hub-and-spoke networks are highly sensitive to disruptions because of the concentrated level of activity at the hub. Hub connectivity is quantified by the number of flight legs in a route that begin and end at different hubs, with various stops at spoke gateways in between. Reducing hub connectivity *isolates* the hub and makes it less sensitive to disruptions at other hubs within the carrier's network. *Short cycles* are a sequence of a small number of flight legs that begin an end at the same station. When an aircraft requires unscheduled aircraft maintenance, for example, canceling a short cycle will limit the number of flight legs affected.

Their objective is to minimize cost subject to hub connectivity. The authors present the following notation in their robust fleet assignment model:

Sets

J : set of fleet types j

S : set of strings s, (sequences of flight legs commencing and ending at a hub, flown by the same aircraft

$\mathcal{X}$: set of feasible fleet assignment solutions

Parameters

h_{js} : hub connectivity metric for string $s \in S$ and fleet type $j \in J$, equals the number of flight legs in string s if s begins and ends at different hubs, 0 otherwise

c_{js} : cost of string s flown with fleet type j

ζ : threshold value of hub connectivity

Decision Variables

x_{js} : 1 if fleet type j is assigned to string s, 0 otherwise

$$Minimize \sum_{s \in S} \sum_{j \in J} c_{js} x_{js} \tag{1.2.1}$$

subject to:

$$\sum_{s \in S} \sum_{j \in J} h_{js} x_{js} \leq \zeta \qquad \forall x \in \mathcal{X} \tag{1.2.2}$$

In a related model, the authors maximize hub isolation subject to a maximum operations cost, ϱ. This model is formulated as:

$$Minimize \sum_{s \in S} \sum_{j \in J} c_{js} x_{js} \tag{1.2.3}$$

subject to:

$$\sum_{s \in S} \sum_{j \in J} c_{js} x_{js} \leq \varrho \qquad \forall x \in \mathcal{X} \tag{1.2.4}$$

Computational results indicate that solutions to their robust fleet assignment models are able to significantly reduce cancellations with only a slight increase in planned costs.

1.2.2 Literature Review-Robust Crew Models

Ehrgott and Ryan [9] develop a bicriteria optimization framework that has two separate objective functions, one for minimizing cost and another for maximizing robustness. The authors gauge the robustness of a crew schedule by the number of crews that change aircraft. In their model, they look for Pareto optimal solutions rather than optimal solutions. A Pareto optimal solution is one in which the achievement of two objectives simultaneously is infeasible. Their formulation minimizes the measure of non-robustness while staying below a pre-defined cost. By testing their approach on real airline data, Ehrgott and Ryan show that a small increase in cost can result in a considerable gain in robustness, which is measured by increased ground time between consecutive flights and a reduction in the number of aircraft changes. The number of aircraft changes is mostly minimized by increasing the number of overnights.

Chebalov and Klabjan [7] introduce the notion of *move-up crews*, which are crews that can be swapped with another. The swap is feasible if the crews have the same home base and the swapped pairings are feasible (i.e., the sequence of duties is feasible). In addition, the swap is only feasible if the two crews are scheduled to complete their respective pairings on the same day. This prevents the need to extend the length of a pairing or altering the subsequent schedule of a crew. The objective function is then to maximize the number of move-up crews, which increases the crew schedule's robustness by introducing the potential for swappable crews. To circumvent the problem of having extremely high crew costs for the sake of robustness, the authors add a constraint that ensures the crew cost is not too far from the traditional planned crew cost. Chebalov and Klabjan solve their model by utilizing a Lagrangian decomposition model. The methodology behind Lagrangian decomposition is to relax difficult constraints and subsequently add penalty terms to the objective function. The challenge with evaluating robustness in crew schedules is the need to evaluate one schedule against another and compare their relative performances against a simulation. An accurate simulation, however, must reflect realistic operations that model aircraft, crew, and passenger flow.

1.3 Thesis Contributions

In this research, our major contributions include:

- Reviewing various approaches to address airline schedule recovery.
- Developing an optimization model for the Express Shipment Schedule Recovery (**ESSR**) problem. We illustrate a representation of the NDA network that incor-

porates different recovery strategies in the event of a disruption. We take advantage of this representation to develop an optimization model that evaluates three objectives during recovery: 1) *minimizing schedule delay*, 2) *minimizing the number of service failures*, and 3) *minimizing the combined cost of operations and service failures.* The solution reflects a trade off between the cost of operations and service failures.

1.4 Thesis Overview

In Chapter 2, we provide an overview of the NDA Operations, with a comparison to the passenger airline industry. The objectives and costs of operations of both the package delivery and passenger airline industry are discussed. We review recent research in developing real-time recovery solutions. We also discuss UPS's current methodology for resolving schedule perturbations.

In Chapter 3, we describe the underlying time-space network used to represent the NDA operations. We take advantage of the representation to build our formulation for evaluating the tradeoffs between operations costs and missed package deliveries during the recovery period. We also illustrate the flexibility of our formulation to reflect different objectives during recovery by modifying the objective function.

In Chapter 4, we implement our modeling approach on computational experiments that are reflective of realistic schedule perturbations. We show that our formulation can evaluate different recovery strategies over a variety of metrics.

In the final chapter, we summarize the results and conclusions of our experimental results. We conclude with a discussion of the limitations of the proposed approach and suggest future research directions.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 2

Uncertainty in Air Operations

The previous chapter provided an overview of the NDA operations of an ESDS carrier. We review recent literature that addresses recovery in the aftermath of irregular operations and summarize current solution methodologies at UPS.

2.1 Examples of Disruptions

Clarke [8] summarizes the main categories of schedule perturbations at airports, as established by the Air Traffic Operating Management System (ATOMS) database system. They are:

- *Weather.* Wind, fog, thunderstorm, low cloud ceiling;
- *Equipment.* Air traffic radar/computer outage;
- *Runway.* Unavailable because of construction, surface repair, disabled aircraft;
- *Volume.* Aircraft movement rate exceeds capacity of the airport at a given time; and
- *Other.* Anything excluding runway, weather, volume, and equipment.

The study evaluated 20 hub airports for six large domestic carriers in the United States. The same disturbances that affect passenger airlines will affect an ESDS carrier. By studying the hub complexes of six of the largest passenger airline carriers in the United States, Clarke finds that the loss of capacity due to severe weather and traffic volume account for 93% of flight delays at airports.

During the month of March 2005, weather was the primary cause of delays for departures from U.S. origins in the NDA network for an ESDS carrier. Weather delays on departures, and their propagation throughout the night of operations accounted for approximately 35% of the carrier's delays that month.

2.2 Real-Time Recovery Solutions

Disruptions to a flight network and its economic implications have motivated researchers to develop optimization-based approaches for airline recovery. The focus, however, has been mainly on the passenger airline industry. Because airline schedules are carefully optimized and disruptions are very costly to airlines, the carrier's goal is often to return to the schedule as quickly as possible. Other objectives might be to minimize passenger delay, the cost of reserve crews and spare aircraft, or the number of passengers who miss their connections.

During recovery, operations personnel must consider a variety of physical constraints and regulations to develop a cost-effective solution quickly. Among these are *aircraft balance*, *station curfews*, *crew feasibility*, and *aircraft maintenance*. Aircraft balance refers to having the scheduled number of aircraft at each station at the end of the recovery period. This ensures that flights will operate as scheduled once normal operations are resumed. Station curfews prohibit arrivals or departures at gateways after certain hours, most often, to limit noise. The Federal Aviation Administration imposes complex regulations regarding pilot flying time, while contractual agreements with pilot unions further complex crew feasibility during operations. The FAA maintains rules for scheduled aircraft maintenance as well.

Filar, et al. [11] surveys recent literature in the area of recovery from schedule disruptions. Recovery options include adjustments in flight schedules, aircraft, and crew. Much of the research performed in developing potential decision support systems in the aftermath of irregular airline operations focus on the rescheduling of aircraft. Flight leg cancellations and delays are determined in this stage, as well as aircraft re-routing decisions. After the aircraft are rescheduled, crews are assigned to uncovered flights legs and repositioned accordingly. The final stage is to accommodate for passengers.

2.2.1 Literature Review-Aircraft Recovery

Early work by Teodorovic and Guberinic in 1984 [21] considered minimizing total passenger delay in the aftermath of a schedule disruption. They form a network of three planes and eight flights, in which nodes represent each available aircraft and the cost on arcs represent the sum of delays to each passenger. By utilizing branch and bound, they minimize the cost of aircraft routings to determine a new schedule plan for the airline fleet. Their problem was limited in scope and failed to consider realistic aspects such as aircraft balance, airport curfews and flight cancellations.

Teodorovic and Stojkovic [22] extended this work in 1990 by considering an airport *closing time*, which can be interpreted as the end of the recovery period, after which the regular schedule should be fully recovered. The defined recovery period may force certain flights to be cancelled. Their model had two objective functions: minimizing the total number of cancelled flights in the network and minimizing total passenger delay in the flight network. Minimizing the total number of cancelled flights in the network took precedence, and they used a dynamic programming heuristic that applied a sequential approach to assign as many flight legs to each aircraft as possible during recovery. When multiple optimal solutions yielded the same number of canceled flights, the second objective function of minimizing total passenger delay in the flight network was considered. The limitations of their work include the lack of consideration for crew scheduling, substitution of aircraft types, and ferrying (flying an aircraft without any passengers to a location where the aircraft can be utilized). This important omission resulted in solutions that would not have been feasible due to violations of crew rules.

Jarrah et al. [13] present two network flow models that seek to minimize costs due to aircraft shortages. Spare aircraft, as well as aircraft swaps (flight legs flown by aircraft not originally assigned to it) are allowed in the two models. They quantify cost by taking into account loss of revenue, the cost of passenger ill-will, the cost of missed connections in the propagation of delay, lost crew time, and the cost of delaying (or canceling) a flight. Their Delay Model identifies flight legs to postpone, while the Cancellation Model determines the flight legs to cancel. To assess the cost of delaying or canceling a flight, they consider the number of passengers on a flight, lost crew time, the number of passengers with a connection at a downline gateway, as well as the number of possible downline cancellations. In the case of their Delay Model, computational results indicate cost savings ranging between 20% and 90%, compared to the cost of implementing a schedule recovery process independent of decision support tools. The drawback of their approach is that delay and cancellation decisions are made sequentially rather than simultaneously.

Yan and Young [31] were the first to consider delay and cancellation simultaneously. In their objective to maximize the airline's profit, they consider the airline's revenue minus the cost of cancellation and/or delay. Limitations of their model include the lack of consideration for maintenance schedules, crew restrictions, and passenger connections.

Yan and Yang [30] develop a framework to handle schedule perturbations caused by aircraft malfunctions and are the first to consider ferrying aircraft. The objective of Yan and Yang's model is to minimize the period of disruption after an incident and to maintain the most profit given the disruption throughout the system. The framework is based on a basic schedule perturbation model that is based upon a time-space network. They develop flight, ground, position, and overnight arcs to model aircraft movements. By exploiting the structure of the time-space network, they develop four strategic models, each with incremental complexity. While all models permit the possibility of a flight cancellation, only two of the models (Models 2 and 4) allow ferrying, and only two of these (Models 3 and 4) allow flight delays. Models 1 and 2 are easily solved by the network simplex algorithm because the absence of flight delays gives rise to a pure network flow problem. In contrast, Models 3 and 4 result in a network flow problem with side constraints, and are classified as NP-hard. The authors solve these two models using Lagrangian relaxation and subgradient methods to find near optimal solutions. The scope of their research is limited to the operations of a single fleet and non-stop flights, which is justified for their work with a small Taiwanese airline carrier. They tested more than 500 scenarios, with the largest one consisting of 2,761 nodes and 42,262 arcs. In their case study, Models 1 and 2 (which are pure network flow problems) were solved to optimality within one minute, while Models 3 and 4 (network flow problems with side constraints) converged to 1% of the optimal solution within 5.5 minutes.

In [29], Yan and Tu extend [30] to include more than one fleet type (three in their case study), which allows for aircraft substitution in their model. Their work also allows for separate and distinct flight legs to be combined into one by eliminating all intermediate stops. They build upon the time-space network from [30], creating a multi-commodity flow problem that is again classified as NP-hard. These solutions once again required using Lagrangian relaxation and subgradient methods.

In [3], Argüello et al. study the airline schedule recovery problem in the event that aircraft are temporarily grounded or delayed. They establish a recovery period of a day to restore normal schedules, while implementing an interim schedule. The objective is to minimize lost revenues and aircraft operating costs. Each aircraft type is dealt with independently of others, and crew availability is ignored.

Argüello et al. present the following notation in their aircraft recovery model.

Sets

F : set of flights i

P : set of feasible aircraft routes j

Q : set of available aircraft k

S : set of stations s

Parameters

a_{ij} : equal to 1 if flight i is in aircraft route j; 0 otherwise

b_{tj} : equal to 1 if aircraft route j terminates at station t; 0 otherwise

c_i : cost of canceling flight i

d_j^k : cost for assignment of aircraft k to aircraft route j

h_t : number of aircraft required to terminate at station t

Decision Variables

x_j^k : assignment of aircraft k to aircraft route j

y_i : assignment of cancellation to flight i

$$Minimize \sum_{k \in Q} \sum_{j \in P} d_j^k x_j^k + \sum_{i \in F} c_i y_i \tag{2.2.1}$$

subject to:

$$\sum_{k \in Q} \sum_{j \in P} a_{ij} x_j^k + y_i = 1 \qquad \forall i \in F \tag{2.2.2}$$

$$\sum_{k \in Q} \sum_{j \in P} b_{ij} x_j^k = h_t \qquad \forall t \in S \tag{2.2.3}$$

$$\sum_{j \in P} x_j^k = 1 \qquad \forall k \in Q \tag{2.2.4}$$

$$x_j^k \in \{0, 1\} \qquad \forall j \in P, \forall k \in Q \tag{2.2.5}$$

$$y_i \in \{0, 1\} \qquad \forall i \in F \tag{2.2.6}$$

The cost d_j^k is equal to the sum of the delay costs observed by assigning aircraft k to aircraft route j. Note that $d_j^k = \infty$ when an aircraft route j begins at a station other than where aircraft k is located at the beginning of the recovery period. The objective function 2.2.1 minimizes the sum of delay costs associated with aircraft route assignments and flight cancellation costs. Constraints 2.2.2, 2.2.5 and 2.2.6 stipulate that all flights must be either assigned to an aircraft route or canceled. The aircraft balance constraints (2.2.3) ensure that aircraft are positioned according to schedule by the end of the recovery period. Constraints 2.2.4 assigns each aircraft to exactly one feasible aircraft route.

One of the difficulties in solving the problem is that the numnber of feasible aircraft routes explodes exponentially with the number of flights. Therefore, in solving the problem, the authors use a GRASP (greedy randomized adaptive search procedure) to generate feasible aircraft routings. The initial solution is found by canceling all flights in which the aircraft were grounded or delayed. Proceeding from the initial solution, neighboring solutions are generated by slightly modifying the incumbent solution and evaluating its cost. Only the best options are included in a restricted candidate list. This procedure is performed iteratively on the restricted candidate list until the CPU time limit is reached and the best solution is retained. The quality of the GRASP solutions is compared to the lower bound established by the time-band approximation scheme detailed in [2]. In over 90% of the computational tests, the best GRASP solution was within 10% of the lower bound. In addition, the GRASP and lower bounding procedure was completed in less than 15 CPU seconds.

Thengvall et al. [24] were the first to examine the irregular operations problem in a way that minimizes the deviations from the original aircraft routings. They allow flight delays and cancellations, but maintain a significant portion of the planned aircraft routings. Minimizing cancellations, delays and deviations are inherently conflicting objectives and doing so simultaneously reflects tradeoffs among them. Deviations are quantified by the number of unbroken flight paths and the number of swaps implemented.

They base their model on a simple time-space network that includes ground arcs and flight arcs to represent aircraft movement, and nodes to represent flight departures and arrivals. Their model *incorporates delays, discourages deviation in aircraft routings*, and *discourages interruption of a scheduled through-flight*. Through-flights are consecutive flights legs flown by the same aircraft that serve long-haul markets (with an intermediate stop). This methodology is described in further detail below:

- *Incorporating delays.* Flight copies are generated to represent delay options by creating a series of delay arcs separated by a determined amount of time (an hour, for example).

- *Discouraging deviation in aircraft routings.* Protection arcs are constructed to cover a number of continuous flight segments assigned to a single aircraft. Thus, one protection arc may encompass more than two separate flight legs. The authors assign an incentive value to the cost of the protection arc so that including the protection arc improves the objective function compared to flying the separate arcs with different aircraft. Thus, while the cost on each separate arc is equal to the revenue generated by that flight, *protection arcs* encompassing a series of successive flight legs with the same aircraft will be associated with a higher value than the sum of the two arcs separately. The assignment of an incentive to the protection arc discourages deviation from the original aircraft routing.

- *Consideration for through-flights.* The concern with *through-flights* is that passengers will not arrive at their destination should one of the flight legs composing the *through-flight* be canceled. The authors create an arc that begins at the flight's origin and ends at its termination, which represents one aircraft assigned to all legs of the *through-flight.* This encompassing flight arc carries the full revenue. However, by flying one leg of a *through-flight* on a different aircraft the revenue gained is discounted by the percentage of passengers originally scheduled to fly the entire *through-flight.*

Thengvall et al. use the following notation in their model:

Sets

Nodes

S	:	set of supply nodes s
T	:	set of termination nodes t
I	:	set of intermediate nodes i

Arcs

G	:	set of ground arcs g
F	:	set of flight arcs f
P	:	set of protection and through flight arcs p
$O(i)$	:	set of arcs originating at node i
$T(i)$	:	set of arcs terminating at node i
$F(\eta)$	:	set of arcs covering flight η; $F(\eta) \subset F \cup P$

Parameters

η	:	unique flight number representing each flight leg;
C_f	:	passenger revenue minus flight cost minus delay cost for flight arc f; if part of a through flight, revenues discounted by (1 - proportion of through-flight customers)
C_p	:	for protection arcs: if passenger revenues minus flight costs of all flights covered minus delay costs of all flights covered plus the appropriate incentive, or for through-flight arcs: passenger revenues minus flight costs of all flights covered
B_s	:	initial supply of aircraft at supply node s
B_t	:	number of aircraft required for end-of-period balance at termination node t
U_g	:	upper bound for ground arc g

Decision Variables

x_f : flow on flight arc f (binary)

y_p : flow on protection or through-flight arc p (binary)

z_g : flow on ground arc g (integer by construction)

$$Maximize \sum_{f \in F} C_f x_f + \sum_{p \in P} C_p y_p \tag{2.2.7}$$

subject to:

$$\sum_{g \in O(s)} z_g = B_s \qquad \forall s \in S \tag{2.2.8}$$

$$-\sum_{g \in T(t)} z_g = B_t \qquad \forall t \in T \tag{2.2.9}$$

$$\sum_{g \in O(i)} z_g - \sum_{g \in T(i)} z_g + \sum_{f \in O(i)} x_f - \sum_{f \in T(i)} x_f + \sum_{p \in O(i)} y_p - \sum_{p \in T(i)} y_p = 0 \qquad \forall i \in I \tag{2.2.10}$$

$$\sum_{f \in F(\eta)} x_f + \sum_{p \in F(\eta)} y_p \leq 1 \qquad \forall \eta \in N \tag{2.2.11}$$

$$x_f \in 0, 1 \qquad \forall (f) \in F \tag{2.2.12}$$

$$y_p \in 0, 1 \qquad \forall (p) \in P \tag{2.2.13}$$

$$0 \leq z_g \leq U_g \qquad \forall (g) \in G \tag{2.2.14}$$

The objective function 2.2.7 maximizes the profit under the new cost structure developed for the recovery period. The first term reflects the profit on the scheduled flight legs, while the second term captures the profit for the *protection* and *though-flight* arcs. Constraints 2.2.8 and 2.2.9 ensure aircraft balance at the beginning and end of the recovery period. Flow balance at intermediate nodes is maintained by Constraints 2.2.10. Constraints 2.2.11 ensure that each flight leg is covered by at most one arc that includes the flight leg. Flight arcs and protection arcs are modeled as binary variables (Constraints 2.2.12 and 2.2.13). Integrality will be guaranteed for the ground arcs based on the construction of the time-space network, and are hence only required to be nonnegative (Constraints 2.2.14).

While the objective of their mathematical model is to maximize "profit" under the modified cost structure, the authors note that the use of incentive costs creates an artificial measure of profit, but that this factor allows them to weigh the arcs to create a desirable solution that protects scheduled aircraft routing to some degree. The strength of the model lies in the flexibility in allowing the user to modify the number of delay options and the incentive cost of protecting a series of flight legs.

2.2.2 Literature Review-Crew Recovery

The objective of the crew rescheduling problem is to reassign the crews to the disrupted schedule at minimum cost. The problem of crew recovery is different than crew scheduling because of the desire to introduce as little disruption as possible to the *optimal* planned schedule.

The solution to the crew rescheduling problem must consider current partially flown pairings, reserve crew availability, monthly flown hours of the crews, and the crew's training to operate specific equipment types. A *pairing* is a sequence of flight legs assigned to a crew that satisfy complex legal restrictions set forth by the Federal Aviation Administration and union contract requirements. Pairings originate and terminate at the same airport where the crew is based. A pairing consists of *duties*, which are a sequence of flight legs followed by an overnight stay. Pairings may also include a flight leg called a *deadhead*. In the recovery problem, deadheads are used to reposition crews from the point of disruption to a later location where they are scheduled to cover a flight leg. Deadheading of crews is also used to return a stranded crew to its base. *Reserve crews* are crews that operate on call and are not assigned to a particular flight schedule. They are guaranteed a minimum guaranteed number of hours of pay even if they do not perform a duty. The availability of reserve crews is limited to only a few, if any, crews that are ready to be deployed during a normal day of operations.

Crew costs are second only to fuel, and is thus the source of potential cost savings in flight operations. Although recent studies have included the crew planning problem, literature on the crew recovery problem is much more limited. The IT requirements necessary to support the problem are extensive, requiring accurate information on crew locations, partially flown pairings, future schedules for each crew member, and contractual agreements with unions as well. Generating a solution in real-time that takes these factors into account poses another computational challenge.

Teodorovic and Stojkovic [23] were the first to consider crew planning decisions during the recovery process in 1995. The objective to minimize airline schedule disturbances remains unchanged, but the model takes into consideration the restrictions placed on crew pairings. Such restrictions include the length of time a crew member may be on duty as well as the number of take-offs and landings that can be performed during one work shift. Their approach is to resolve crew rotations first. They form a sequence of flights legs, called a route, that are feasible for the crew. These routes are sequentially assigned to aircraft, with each aircraft covering several routes. The approach assumes that each crew will remain with the same plane for the entire duration of their duty period. They employ the first-in, first-out (FIFO) principle to address the assignment of crew rotations. The first crew to arrive at an airport is assigned to the first flight leg departing that airport. Flight legs are assigned to the first crew by minimizing the total ground time spent by the crew. This, in turn, maximizes the crew's flying time during the duty period. Their algorithm schedules a new crew rotation prior to scheduling the aircraft rotation. The authors found that computing time was much longer when the process was reversed because the legality of the crew rotation would have to be checked in the second step. They note that designing the aircraft rotations first often resulted in plans that were infeasible due to a violation of crew working time.

Lettovsky, et al. [17] provide, in almost real time, a recovery plan for reassigning crews to restore a disrupted crew schedule, while protecting future crew assignments. They solve a set covering problem to obtain a minimum cost set of pairings that cover as many flight legs as possible with minimal impact on passengers when cancellations and delays are unavoidable. The authors limit deadheading of crews on each flight leg. They implement a heuristic solution that maintains the schedules for as many crews as possible. The benefit of this approach is that the optimal schedules generated from the crew planning phase are maintained for crews unaffected by the disruption. Their methodology is applied for medium sized disruptions, with acceptable running time.

Yu, et al. [33] present a crew recovery model as a set covering problem. Their mathematical programming model introduces penalties for deadheading, unassigned crews, and uncovered flights, while enforcing crew legalities and cover constraints. The objective is to minimize the sum of deadheading, cancellation, and modified crew schedule costs. Because their problem is highly complex, consisting of more than 4,000 pilots operating three to six flight segments per day, the number of possible flight sequences is not only large in size, but impractical to solve with conventional set-covering algorithms. They employ a heuristic-based search algorithm that generates or modifies a few pairings at a time. They sequentially test the performance of these pairings and incrementally modify and add additional pairings. The model was implemented on a domestic airline's fleet of Boeing 737, corresponding to disruptions affecting 1-40 flight legs. In each instance, the solution converged to 5% of optimality within eight minutes.

2.3 Organization and Structure of an ESDS Carrier

ESDS carriers sell not only service, but *on-time* delivery. Thus, the goal in the industry is not only to minimize the cost of operations, but to ensure that guaranteed service is satisfied. This goal is reflected during the planning phase as well as in the day of operations. This section describes the typical planning process for network planning and resolving disruptions during operations in the NDA network.

2.3.1 Network Planning

The objective of planning for the carrier's network is to route the planes and packages in the most cost-effective manner to satisfy customer demands. Long-range planners project two to ten years into the future to determine gateways and hub locations, and fleet acquisition plans to accommodate forecasted volume. With this in mind, network planners then develop the tactical plan for the current year. With the location and capacity of the hubs, fleet composition and fleet size fixed, they decide on hub assignments for O-D commodities, aircraft routes, and fleet assignment. Hub and gateway assignments are generally determined by models and algorithms that minimize total package-miles. The planners and analysts must decide the most efficient use of their resources in both the NDA and SDA network operations.

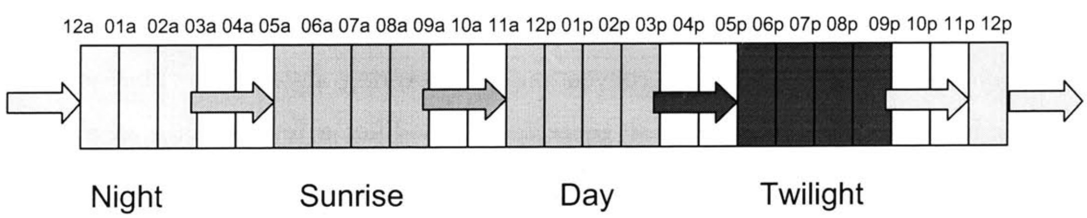

Figure 2-1: Sort Processing Bar Graph

Figure 2-1 illustrates the timeline of sorts that occur during a day of operations for the NDA and SDA operations. All pick-ups, regardless of service level, go first to a Twilight Sort for initial processing. This first sort is also the local sort at the origin ground center. The packages are then transported to gateways and held for transport to a processing hub. The packages are scheduled to arrive in time for the Night Sort, which occurs between 11pm and 3am. The Night Sort processes NDA packages (almost) exclusively, because of strict timing considerations for NDA service. The packages that come out of a Night Sort are further split into Sunrise cans that go out for delivery that same morning. The final stage of processing for a package, regardless of service level, is the Sunrise Sort. A Sunrise Sort does not typically feed any other sort because it is occupied with preparing packages for delivery upon completion. The Sunrise Sort at the destination gateway is the final stage of preparation for NDA packages before they are delivered to the customer via ground transportation or feeder jets.

The Day Sort is occupied with volume with the levels of priority other than NDA. The SDA packages processed in a Day Sort are those that were picked up in the previous business day, with a delivery date of the next Business Day. The Day Sort then feeds the Twilight Sort. The cycle of sorting continues, with each sort feeding the next. The exception is that Sunrise Sorts do not typically feed other sorts because the packages are loaded and sent out for their morning delivery to meet NDA service.

2.3.2 Contingency

While the network planners and analysts have optimized the plan that dictates the ideal routing of packages and planes during a night of NDA operations, the Contingency group is responsible for ensuring that day-to-day operations meet customer service requirements. This group monitors the positions of all flights and commodities during operations to guarantee that packages reach their required destinations (gateway, hub sort, final customer) at each stage of processing. Should a flight exception occur, they develop and implement plans of action, while communicating all actions and results to other departments.

The Contingency group consists of a team of approximately 10 employees, with one employee assigned to each region of the carrier's operations (i.e. North East, West Coast, international, etc.). Any irregularities in schedule that occur in that region will be handled by the specialist assigned to that region. Each specialist is trained with a deep understanding of the carrier's flights and how they are intertwined in the network. When they are notified of a disruption at a gateway, they have to evaluate all possible remedies and implement a solution. Their decision is highly dependent on the time they have to act. Because the carrier has 21 spare aircraft ready for deployment during every night of operations, the cost of having these crews and aircraft ready is already factored into the cost of operations. The primary goal echoed throughout the company's mission is to be an on-time airline and meet customer demands. To reflect this idea, the Contingency group is not hesitant to use any of the resources available to them to ensure timely delivery of packages.

Current Methodology for Resolving Disruptions

In the event of a schedule disruption, the carrier may use any combination of the following to resolve the perturbation.

- Flying a spare aircraft to rescue packages;
- Swapping aircraft;
- Rerouting a flight and/or packages;
- Delaying a flight; and/or
- Canceling a flight and missing service on packages

If the disruption involves packages stranded at a gateway en route to the hub (due to a mechanical failure of the inbound aircraft, for example), a common solution is to fly in one of the spare aircraft from one of the spare aircraft locations to rescue the packages and deliver them to the hub for processing. Depending on how much notice the Contingency group has of the perturbation, however, this might or might not be a feasible option. The constraint in this situation is having the packages arrive at the hub sort with sufficient time to sort the packages and load them on to their respective outbound flights. To determine this solution's feasibility, at the time of the disruption, one must calculate the earliest departure time of the spare aircraft from its location, the flying time to the disrupted gateway, the turnaround time required at the gateway for that spare aircraft, and the subsequent time of arrival at the hub for processing.

When an aircraft experiences a malfunction, the gateway issues a time estimate for the repair. Repairs may range from requiring a simple replacement of a part to unknown causes with an indefinite repair time. If the problem can be fixed with replacement parts that are in stock, for example, the scheduled aircraft might be ready before a spare can arrive, and launching a spare would not be necessary or beneficial in any sense. In situations that the spare can arrive sooner than the repair can be completed, the spare will be deployed. Even if the repair time is not entirely known, the spare aircraft might be set up to rescue the packages at the gateway while the maintenance crew simultaneously proceeds with repairs. If the scheduled aircraft is repaired while the spare is en route to the gateway, the crew of the spare aircraft will be informed in midair that the aircraft is no longer needed and the spare will turn around and return to its origin. The practice of concurrently repairing an aircraft and launching a spare aircraft underscores the carrier's dedication to on-time performance, guaranteeing the service that it sells.

Contingency operations need accurate information regarding the package count affected by the disruption and the region affected by the disruption. Because some regions have overlapping service territories, it might be possible to reroute the commodities to another hub. In that case, however, the outbound flights and operations from the other hub must be considered. Because the hub sort can to operate only to a maximum rate, the time that the packages arrive at the hub will then determine if the packages will be sorted in time to board their outbound flights. The time windows for the hub sort, however, may be flexible beyond those set in Figure 2-1 by extending the sort time.

Disruptions are evaluated on a case-by-case basis. If a large volume of high-value packages is arriving late into the hub, beyond the last inbound flight, the Contingency group member might decide to operate the sort beyond its schedule to salvage the packages. Bringing these packages into the hub late, however, is only worth it if they make it to their eventual destination on time. This can require the postponement of respective outbound flights. The planes waiting for the late incoming packages, however, have other commodities with other destinations onboard. By waiting for the late incoming packages, the dispatchers might be compromising the service of another commodity that would otherwise make it to its respective destination on time. Downline from the delivery to a gateway, packages will still need to be loaded onto scheduled ground transportation or smaller chartered jet service for transport to their destinations. Thus the essential element of the decision making process is to quantify the propagation of delay resulting from the initial disruption to identify the action that minimizes *total* package delay costs.

One option rarely used is canceling a flight. Only when absolutely necessary will the carrier not at least attempt to service the commodities at the gateway.

Aftermath

In addition to handling disruptions to the schedule during the day of operations that directly affect service requirements, Contingency is also responsible for resolving aircraft balance after rescuing packages. The spare aircraft involved in operations do not necessarily have to be returned to their original locations, but by the next night, the number of aircraft of each fleet type scheduled for operations must be present. To do this, Contingency may operate *ferry flights*. *Ferry flights* are non-revenue generating, empty aircraft that fly to locations that need the specific fleet type for the following night's operations. If two aircraft of the same fleet type are candidates for ferrying, Contingency will choose the closer one, taking advantage of the shorter flight and lower operating cost.

When a spare aircraft replaces a scheduled one for a particular flight leg, the spare might, in fact, have to fly all flight legs assigned to the disabled aircraft. While the spare aircraft might be a fine substitute for the original aircraft in operations and commodity transport, the maintenance schedule can be disrupted by the swap. Maintenance schedules specify the time and location for each aircraft. Once the substitution is made and the spare assumes the role of the scheduled flights, it can be weeks before it can return to its location and resume its position as a spare.

Shortcomings of the Process

The manual recovery process is dependent on a thorough understanding of the network. Decisions are made without any explicit regard to cost, although logical understandings of larger or older planes being more expensive to operate are assumed. The recovery resolutions are performed with little, if any, dependence on automated decision support systems. The actual cost of the implemented solution is not evaluated against other possible alternatives. Often the biggest constraint facing the Contingency group is time. With a tightly scheduled plan of flights in the network, they often have to scramble to come to a decision.

To summarize, the complexities of the decision making process, Contingency must consider the following factors when devising a solution to resolve a disruption:

- Total package count affected;
- Service territory of an alternate hub (for possible package and/or aircraft rerouting);
- Outbound aircraft capacity from hub;
- Sort times of each hub;
- Pilot flying time;
- Maintenance status of aircraft; and

Taking into account all the physical and regulatory constraints of the system without automated decision tools is a formidable task under the extreme time pressure imposed during operations. Members of the Contingency group often have had several years of experience within the carrier before advancing to the group. There, the team member starts off assigned to one of the smaller regions of the network with a less intense amount of traffic volume. They begin to develop insights as to how one event and its possible solutions have different downline propagation effects and consequences. Their decisions on a current situation must be feasible for aircraft, crew, and maintenance in both the present and future operations. Familiarity with specific aircraft and their scheduled positions in the future become intrinsic as they consider resolution actions in the present.

The criteria for evaluating a solution is varied and difficult to quantify. We can think of cost in terms of operations (crew costs and ferry flight costs); the number of service failures; the number of cancellations; or in delay time. Unfortunately, these goals are often conflicting with one another and one must subjectively determine the allowable tradeoff.

2.3.3 Comparison to Passenger Airlines

Bratu [6] provides an overview of passenger airline planning and operations. The most widely used delay metric to quantify a passenger airline's performance is the 15-minute-On-Time-Performance. Airlines focus heavily on this delay metric because it strongly influences passenger choice. For an ESDS carrier, the time of arrival for the flight is not nearly as important as the on-time delivery to the final customer.

Perhaps one of the largest differences in operations between the passenger airline and the ESDS industry is the resources available to each. While the ESDS carrier is equipped with 21 spare aircraft during the night, passenger airlines have few, if any, spare aircraft available for the same purpose due to its extreme cost. As a result, ferrying aircraft or usage of a spare aircraft in the passenger airline industry is not as common. In contrast, the high cost of spare aircraft and crew are assumed in the cost of operations in the ESDS industry.

Transporting passengers to their final destinations requires consideration of passenger perception. Not only does the passenger want to arrive on time, but they want to travel with the least number of connections possible to minimize their inconvenience. In contrast, an ESDS carrier does not have to consider commodity perception when providing their services. Thus, the carrier will not hesitate to reroute the commodities or use any other combination of resources in its integrated network to deliver the package on time. In the passenger airline industry, unsatisfactory customer service results in a cost of passenger good-will loss and some compensation costs, for example, for meals and lodging. For an

ESDS carrier, however, the carrier suffers from an immediate compensation to the customer of the price of service for the service failure as well as the loss of goodwill. These customers may be difficult to attract if they experience a service failure. The extreme time-sensitive nature of the industry requires quick decision-making, which if well-informed, should lead to improved service and reduced costs.

The objective prior to operations for both passenger airlines and ESDS carriers is to operate as close as possible to the planned schedule. This ensures feasibility for aircraft, crew, and maintenance schedules, while meeting service requirements for passengers and commodities. However, once a disruption occurs, one of the main objectives for operations personnel in the passenger airline industry is to return the schedule to its plan as soon as possible. In contrast, the goal for the ESDS carrier to meet service does not change even after a perturbation. This difference in approach is reflected in the resources and strategies that the two have assumed in the cost of operations.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 3

Optimization Based Recovery Models and Algorithms

The Express Shipment Schedule Recovery (**ESSR**) problem presented in the previous chapter is complex both in size and computational tractability. Thus, addressing the problem requires a practical solution that contains the size of the problem and can also be implemented in real time to allow for a fast solution.

The objective of the **ESSR** problem may involve a combination of minimizing the cost of operations, service failures, and minutes of delay. Depending on different disruption scenarios, the carrier may choose to focus on a different objective to resolve the particular disruption. Solution methods must be generic and robust enough to capture the various objectives.

In this chapter, we introduce the network structure and methodology for the **ESSR** problem. We first demonstrate how the underlying network is constructed and how feasible recovery options are generated. We then present the notation and mathematical formulation that captures the various objectives in the **ESSR** problem, depending only on the set of costs that the user inputs. As we will see, various proposed objective functions are possible to account for minimizing operations costs, missed service failures and delay. The result is an integer multi-commodity network flow model. Finally, we illustrate how the formulation strategy can be applied to realistic schedule perturbations and demonstrate its practical significance.

3.1 Network Representation

The development of formulations to address the **ESSR** problem is dependent on the representation of the underlying time-space network of the carrier. In this section, we present the construction of the time-space network. We demonstrate how this representation remains consistent for the various perturbations that the carrier may face during operations. Finally, we present a model that takes advantage of the developed network structure.

3.1.1 Basic Modeling Constructs

Consider the simple network shown in Figure 3-1. This reduced-sized network includes three gateways, a hub, three aircraft (one each of *type 1, type 2,* and *type 3*) and three package commodities (which are defined by an origin-destination pair). The figure is a snapshot of the schedule as planned. Thus, we have 100 packages from hub h^1 to gateway c, 200 packages from gateway b to gateway a, and 300 packages from gateway b to gateway c in this problem. We assume a capacity of 400 packages for each type of aircraft. Source nodes are created for each commodity at the point in time and location that they originate in the network. Intermediate nodes are created at each point that a commodity begins or terminates a movement. The sink nodes at the end of the period enforce proper aircraft balance and service requirements for commodities.

After the nodes are constructed, arcs are then created to connect the nodes. In the diagram, all flow is from left to right. In the situation where the schedule is executed as planned, we observe two types of arcs:

- The *scheduled flight arc*, connects two nodes at different times and locations, representint an aircraft movement, which may also correspond to a package movement. The arrival time at h^1 from b is the scheduled departure time at b plus the *Block time*

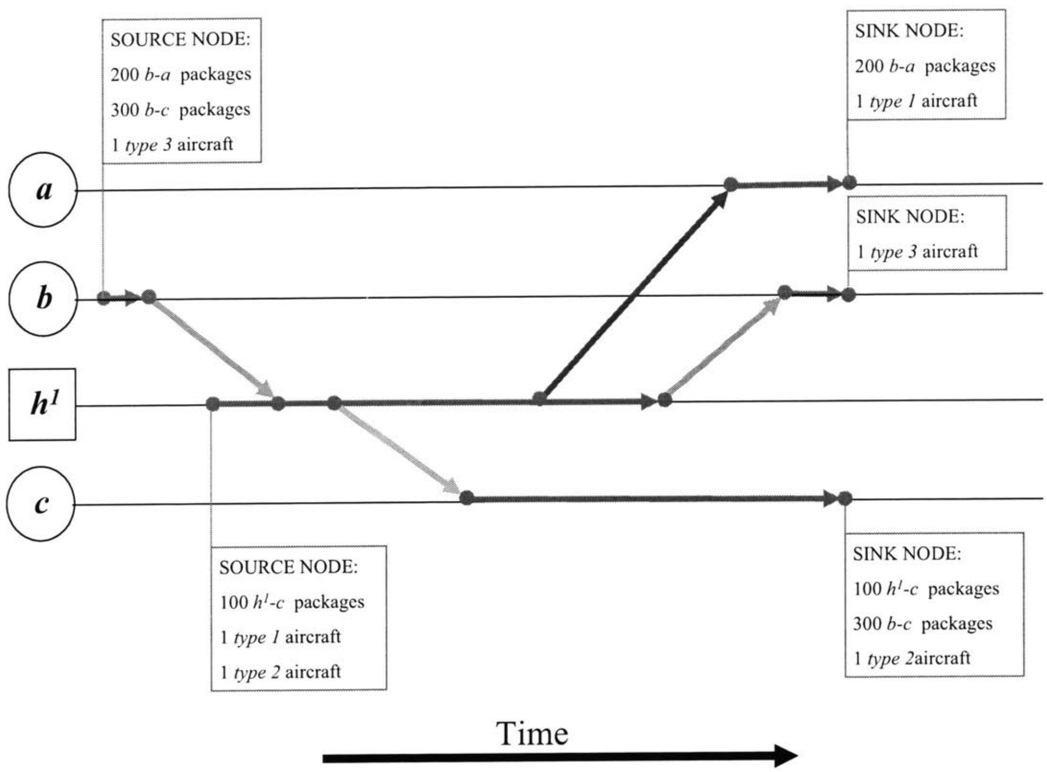

Figure 3-1: Time-Space Network for Scheduled Pick-Up and Delivery Flights

for the specific fleet type flying the arc. The *block time* is equal to flying time plus taxi time. The flow of aircraft on flight arcs is restricted to binary values. We place a node in the network at h^1 at its arrival time.

- A *ground arc*, connects two successive nodes at the same location. These arcs may represent minimum turnaround time after a plane has landed, sorting time at the hub, or just waiting time on the ground. Minimum turnaround time for a specific aircraft type is the minimum time required to unload and load shipments as well as refuel. In Figure 3-1, the arrows on the ground arcs are omitted for clarity.

3.1.2 Time-Windows and Level-of-Service

Armacost [4] and Shen [20] describe the time-windows and level-of-service requirements observed at an ESDS carrier. Referring back to the physical system described in Chapter 1, the carrier specifies an *Earliest Pickup Time from Center (EPTC)*, which is the earliest time that commodities may depart the ground center. The *EPTC* is determined to allow customers sufficient time to drop off their shipments at the ground center. Similarly, the *Latest Delivery Time to Center (LDTC)* specifies the latest time that packages can be delivered to their respective ground centers and still guarantee on-time delivery.

At a gateway, an aircraft cannot depart on a pickup route before a designated *Earliest (Gateway) Pickup Time (EPT)*. The *EPT* is determined by allowing sufficient time for all demand from the gateway to the hub to be sorted and loaded on departing aircraft. The *Latest (Gateway) Delivery Time (LDT)* is the time that the packages must arrive at the destination gateway in order to satisfy delivery service standards. The *EPT* and *LDT* for the gateway can be calculated by the *EPTC* and *LDTC* of the corresponding ground center and the transportation time between the gateway and the ground center.

Specific to the hubs within the network, timing requirements are specified by a *Latest Hub Arrival Time (LHAT)* and *Earliest Hub Departure Time (EHDT)*. The *LHAT*, which is the latest time that a pickup flight can arrive at the hub, is dictated by the hub sort time necessary to process the packages on the flight. The earliest time that a flight can depart a hub is noted as the *EHDT*, which takes into account sufficient time for the packages to be sorted and loaded onto the delivery flight before it can take-off.

The *sort start time* denotes the latest time that a pickup flight can arrive at the hub and still have its onboard packages make it through the sort and loaded onto delivery flights on time. The *sort end time* denotes the earliest time that a plane may depart from the hub for its destination gateway.

Example 1 In the absence of schedule perturbations, the aircraft at gateway b would transport the packages from b to hub h^1, where they would be sorted and reloaded onto delivery flights to gateways a and c. All of the packages and aircraft would observe the balance and demand constraints at each source, sink, and intermediate node and no additional costs beyond scheduled operational costs would be incurred.

Consider, however, the case that the aircraft scheduled to fly from gateway b to hub h^1 experiences a malfunction at gateway b, leaving the packages destined for gateways a and c stranded at gateway b.

We build the options for recovery by augmenting the network to introduce a spare aircraft that is assigned to gateway a. This is shown in Figure 3-2. After the disruption has occurred, the spare aircraft can possibly be deployed to fly from a to b to rescue some of the packages and continue to transport them to the hub. The delay of this flight leg beyond the arrival time of the originally scheduled flight from b to h^1 depends on the flying time from the spare location a to b, as well as the minimum *turnaround time* at b before the spare can proceed to h^1.

To model our recovery options, we construct three additional arc types:

- The *spare flight arc* models possible movements of spare aircraft. In this case, the spare aircraft at a can fly into gateway b, observe the minimum turnaround time specified for the aircraft type, continue on to the hub, and then return to its location at gateway a. The associate a cost with operating the spare, which is dominated by fuel costs. Because the carrier already assumes hot spares and hot crews in their cost of operations, operating the spare does not result in much incremental cost beyond

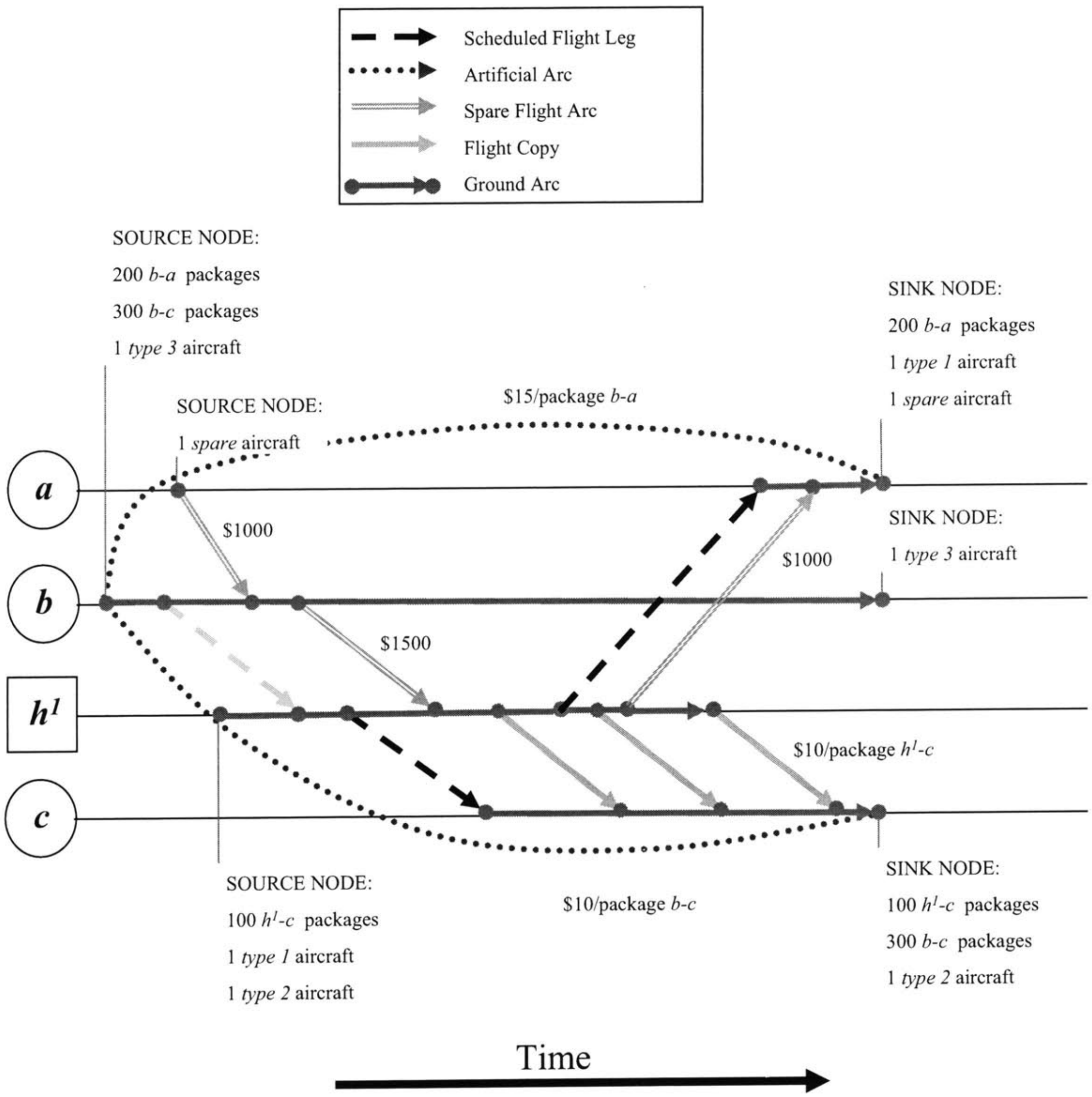

Figure 3-2: Time-Space Network of Recovery Options

the cost of fuel consumption.

- The *artificial arc* provides a direct connection between the package source node at gateway b to its sink node at gateways a and c. These arcs represent the possibility of a service failure, and incur the cost of missing service. This cost parameter must incorporate the other factors associated with missing service, such as the loss of customer goodwill. We assume that we can obtain a monetary value to capture these intangible cost on the arc. For the commodity that is associated with the artificial arc, we assign a cost per package to that arc, noting that a difference in cost per package on artificial arcs for different commodities represents a difference in value of the commodities that the carrier services.

- The *flight copy arcs* are generated at periodic intervals, after the originally scheduled departure, to represent the option of operating the flight leg at later times. In Figure 3-2, three delay options are shown for the flight leg from h^1 to c. These flight departure times are separated by an user-defined period of time, say 20 minutes. Flight copies and their costs are discussed in more detail in the next section.

3.1.3 Generating Flight Copies

In this example, the earliest that the spare aircraft can arrive is after the departure of the scheduled flight from b to c. It would be meaningless, however, for the spare aircraft to fly into the hub with packages from b if those packages missed their subsequent connection on to c. We generate flight copies to represent the possibility of delaying this outbound connecting flight at various times. In this example, we illustrate flight delays of 20-minute intervals and assume a linear hub-processing rate. Each subsequent flight copy can carry a larger fraction of the packages from b because of the additional processing time. This is illustrated by the constraints on the arcs in Figure 3-3. These flight copies illustrate the departure of a flight leg with only a fraction of the packages from b that have been sorted at the hub. The last flight copy in this example is constructed at the time when all the packages from gateway b have been processed and can be loaded onto the delivery route. It is also possible, however, that the sort will end before all the packages are sorted. In this case, the maximum number of packages that can flow on the arc will be the final number of packages processed in the sort. In all of the flight copies, $x \leq 1$.

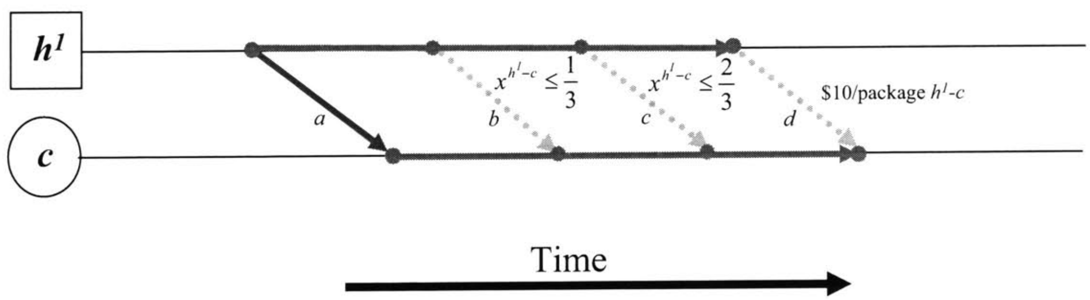

Figure 3-3: Incorporating Delay Options

While constructing the flight copies, we do not generate flight copies that arrive after the LDT of the destination gateway because packages on that aircraft could not be delivered on-time. Hence, flight legs arriving after the LDT are not economical. Instead, it is more efficient to operate an earlier flight copy and deliver even just a fraction of the packages on time, or to incur all the missed service costs but save the operating cost of the flight leg by canceling the leg.

Assume that packages $h^1 - c$ have an earlier LDT at c than packages $b - c$. As a result, all of the $h^1 - c$ will miss service for the last flight copy d, while the $b - c$ packages will still be deliverable on time. To capture the cost of missed service for $h^1 - c$ packages if flight copy d is selected, the cost for $h^1 - c$ commodity flow on flight copy d is set to the total cost of missed service for all $h^1 - c$ packages. Thus for the $h^1 - c$ commodities, certain flight copies are marked by the cost of missing service. Providing flight copy arcs allow trade-offs to be made between missed service and operating costs; and in the quantity and types of commodities to be delayed. In this example, at most one out of two commodities ($h^1 - c$ or $b - c$) can be fully delivered on time.

With the construction of these additional arcs, the number of nodes in the network correspondingly increases. We sort the nodes in increasing time and number them accordingly for identification purposes. Arcs are then identified by an originating node and an end node. Remaining ground arcs are constructed by sorting all the nodes in increasing number at a location and inserting arcs between consecutive nodes along the same location axis.

Example 2 While Figure 3-2 illustrates the case of a mechanical failure, the same recovery network is applicable to the situation of excess demand for the capacity of flight leg $b - h$. In this case, the originally scheduled flight leg (dashed arc) from $b - h^1$ operates as planned, but additional packages are left behind at gateway b. The options and costs in the network during the recovery period are identical for this problem. Operations personnel must decide if they will operate the spare aircraft to rescue the volume that did not fit on the original aircraft at gateway b. Subsequently, the flights departing hub h^1 have the same delay options for accommodating the additional packages arriving from b. There is a need, however, for an additional constraint to capture the limited outbound capacity on the delivery flights departing the hub. If the aircraft assigned to the flight leg from $h^1 - c$ can not accommodate the additional packages from gateway b, the option to fly the spare aircraft again from $h^1 - c$ should be considered as an additional recovery option. This option is feasible, however, only if aircraft maintenance schedules and balance constraints can be satisfied.

Example 3 We revisit the example once again to illustrate how to augment the network to incorporate the possibility of a flight diversion from its scheduled arrival hub into into another hub, h^2. In this example, we have a pick-up flight that includes two legs. A *type 3* aircraft is scheduled to originate at b, fly into gateway c to pick up additional packages before continuing into hub h^1 for sorting. Gateways b and c both have commodities with a gateway a destination. Consider, however, the case that an operational ban is enforced due to inclement weather, prohibiting aircraft from flying from b to c.

In Figure 3-4, we introduce another hub, h^2 in our network that is able to serve the

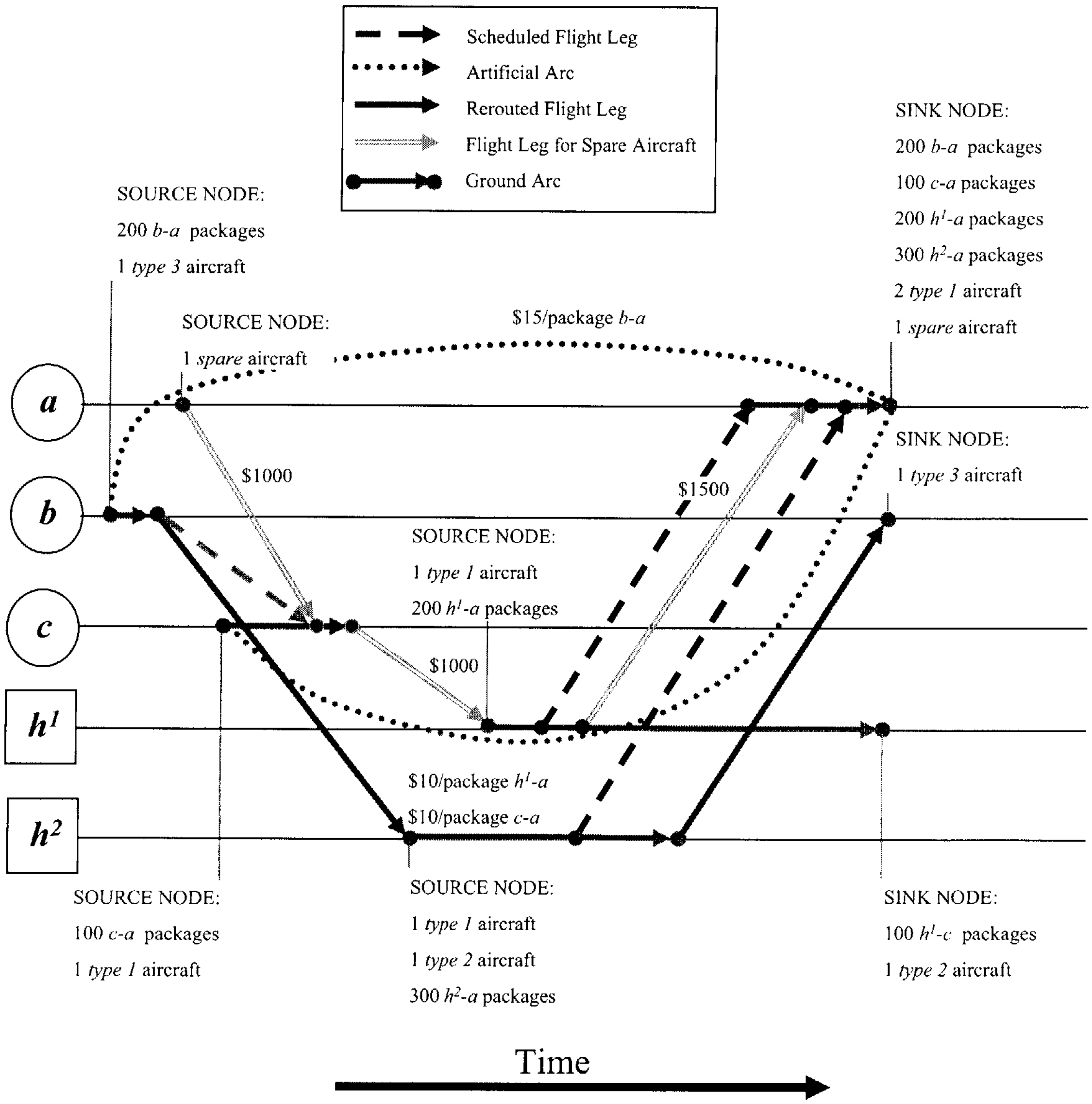

Figure 3-4: Time-Space Network with Rerouting

destinations for the package commodities that originate at b and c, namely gateway a. By inserting a flight arc from b to h^2, we represent a diverted flight. The package commodities from b will still have a means of reaching their destination. In this example, we assume that the diverted flight leg arrives at h^2 before any of the scheduled outbound departures. Hence, no flight copies are needed.

To account for the volume that is stranded at c, the spare from a can be deployed to rescue it and continue on to hub h^1. The cost on the arc would be that of operating the spare aircraft from a to h^1, but would save the cost of the c to h^1 commodities missing service. Flight copies for the flight leg from h^1 to a will also be generated at periodic intervals but are not shown in the figure for clarity. In this example tradeoffs exist among all the packages destined for gateway a. At both of the hubs, the inbound flights will arrive late, affecting the scheduled departure of flights for a. The successful service of the late commodity, however, might be at the expense of another commodity that could have been delivered on time if the delivery flight leg was not held.

Besides the tradeoff amongst the commodities involved, there exists an additional tradeoff between operations costs and package service. This is reflected in the operation of the spare aircraft to rescue packages at gateway c.

3.2 Multi-Commodity Flow Formulation

The essential idea in developing a strategy to address the **ESSR** problem is to develop the structure of the network to represent the set of recovery options and capture the cost of operations and possible service failures during the recovery period.

The objective is to flow the packages through the network from origin to destination while satisfying network capacity constraints and achieving *minimum cost*, with minimum cost defined to be measured against a variety of metrics. Three objectives that we consider are to minimize 1) the *combined cost of operations and service failures*; 2) the *number of minutes of aircraft delay*; and 3) the *number of service failures*.

Before presenting our **ESSR** model, we introduce the following notation:

Sets

K : set of commodities $k \in K$

F : set of fleet types F

A : set of arcs (i,j)

N : set of nodes

Parameters

d^k : number of packages of commodity k to be transported from origin $O(k)$ to destination $D(k)$

c^k_{ij} : cost to transport one unit of commodity k on arc (i,j)

c^f_{ij} : cost to operate an aircraft of fleet type f on arc (i,j)

u^f : capacity of an aircraft of fleet type f

Decision Variables

x^k_{ij} : fraction of d^k on arc $(i,j) \in A$, $k \in K$

y^f_{ij} : flow of fleet type f on arc $(i,j) \in A$

$$min \sum_{(ij)\in A} \sum_{k\in K} (c^k_{ij} d^k) x^k_{ij} + \sum_{(ij)\in A} \sum_{f\in F} c^f_{ij} y^f_{ij}$$

subject to:

$$\sum_{j\in N} x^k_{ij} - \sum_{j\in N} x^k_{ji} = \begin{cases} 1, & \text{if } i = O(k) \\ -1 & \text{if } i = D(k) \\ 0 & \text{otherwise} \end{cases} \qquad \forall i \in N, \forall k \in K \tag{3.2.1}$$

$$\sum_{j\in N} y^f_{ij} - \sum_{j\in N} y^f_{ji} = \begin{cases} 1, & \text{if } i = O(k) \\ -1 & \text{if } i = D(k) \\ 0 & \text{otherwise} \end{cases} \qquad \forall i \in N, \forall f \in F \tag{3.2.2}$$

$$\sum_{k \in K} d^k * x_{ij}^k \leq \sum_{f \in F} u^f * y_{ij}^f \qquad \forall (i,j) \in A \tag{3.2.3}$$

$$x_{ij}^k \leq \sum_{f \in F} y_{ij}^f \qquad \forall k \in K, (i,j) \in A \tag{3.2.4}$$

$$x_{ij}^k \geq 0, \qquad \forall (i,j) \in A, \forall k \in K \tag{3.2.5}$$

$$y_{ij}^f \in \{0,1\}, \qquad \forall (i,j) \in A, \forall f \in F \tag{3.2.6}$$

Constraints 3.2.1 and 3.2.2, respectively, ensure *conservation of flow* for each package commodity and aircraft type, enforcing that all commodities flow from their respective origins to their destinations. The *capacity* constraints (3.2.3) restrict the total flow of the package commodities on an arc (i,j) to at most the capacity assigned for arc (i,j). Because the *capacity* constraints (3.2.3) can lead to highly fractional solutions to the LP relaxation, we include the inequalities 3.2.4 to strengthen the LP relaxation of the formulation. Constraints 3.2.5 ensure that the flow of commodity k is nonnegative and constraints 3.2.6 enforce the flow of aircraft on an arc to be binary.

We include a set of constraints, with one constraint for each flight copy, to capture the increase in the number of packages that complete the sorting process at the hub as time increases (as illustrated in Figure 3-3). For example, we assume that sort processing rate is linear and that all late inbound packages can be sorted and loaded onto aircraft within 1.5 hours. If flight copies are in increments of 15 minutes, we include the following six constraints for the six consecutive flight copies for commodity k:

$$
\begin{aligned}
x^k_{copy_1} &\leq \frac{1}{6} \\
x^k_{copy_2} &\leq \frac{1}{3} \\
&\vdots \\
x^k_{copy_5} &\leq \frac{5}{6} \\
x^k_{copy_6} &\leq 1
\end{aligned}
\tag{3.2.7}
$$

Thus, the maximum number of packages that can be sorted, loaded and transported on the first flight copy (with a 15-minute delay) is limited to one-sixth of the total count for commodity k. Constraints 3.2.7 represent a linear distribution of the maximum number of packages that can be flown on each of the flight copies. Each subsequent flight copy can accommodate a proportionately larger number of commodities.

In the model, we must also limit the planes to scheduled flight arcs, flight copies, and ground arcs. Aircraft are not allowed to flow on artificial arcs. Based on the network representation presented, the artificial arc might provide a direct route in the network for the aircraft to flow from source node to sink node, but it does not represent a realistic flow in the network. The capacity parameter for aircraft is only valid on flight arcs that represent physical movements of aircraft. Thus, we include constraints such as 3.2.8 to prevent an aircraft of fleet type f from flowing on artificial arc (a, b).

$$
y^f_{ab} = 0 \qquad \forall (a,b) \in ArtificialArcs, \forall f \in F \tag{3.2.8}
$$

3.2.1 Hybrid Model- Minimize the Combined Cost of Operations and Service Failures

The objective to minimize the combined cost of operations and service failures is modeled in the formulation above. All of the package commodities must flow from their origins to their destinations, whether it be on a combination of flight and ground arcs or artificial arcs. We assign costs to the utilization of spare aircraft as well as to packages that do not satisfy their service requirements. This objective measures the worth of serving a commodity the cost of utilizing the resources to do so.

3.2.2 Aircraft Delay Model

Alternatively, during the recovery period, the carrier might concentrate solely on minimizing delay within the network. This objective is motivated by the carrier's desire to stay on schedule as much as possible, without delaying flights to wait for incoming packages at a hub. To capture this, the objective function in the above formulation is modified to:

$$Minimize \sum_{(ij)\in A} \sum_{f\in F} c_{ij}^f y_{ij}^f \tag{3.2.9}$$

with c_{ij}^f representing the number of minutes flight leg (ij) is behind schedule. Under this representation and approach, flight copies are the only arcs with a non-zero cost.

In addition, we introduce a dummy variable, y_{ij}^{dummy} for the disrupted flight leg (ij) to capture the cancellation of flight leg. Referring to Example 1, the flight leg from $b - h^1$ must be covered by either the spare aircraft or this dummy. If the spare aircraft covers the flight, the delay would be the offset from the original schedule. However, if the dummy variable covers the flight, the delay could be quantified by 1440 minutes (a full day) or even 0 minutes (if alternative routes exist for the packages that are supposed to be served by the canceled aircraft). Setting the delay to 1440 minutes, however, might discourage cancellation, although it is the common methodology used by passenger airlines to minimize aircraft delay. While it does not represent a late flight in the usual sense, this methodology allows us to incorporate a penalty when evaluating a cancellation in terms of delay. The dummy variable is binary, equal to 1 if arc (ij) is cancelled and 0 otherwise. For the disrupted flight arc (I, j), we include the following constraint in the model when we implement this objective function:

$$y_{ij}^{dummy} + y_{ij}^{spare} = 1 \tag{3.2.10}$$

After the "optimal" aircraft routing is determined to minimize the minutes of aircraft delay, we can then subsequently determine the package routings. Having the aircraft and the flight legs they will fly to minimize delay as an input, we can then run the model a second time with the objective function to minimize service failures. Without this objective function, there is no cost associated with the artificial arcs, and hence, there is no incentive to service packages.

3.2.3 Service Failure Model- Minimize the Number of Service Failures

In the case that the carrier is willing to take any necessary action to provide service for the package commodities, the carrier will take full advantage of its resources, regardless of the cost of operations. This particular objective can be considered as a diametrically opposed operating philosophy to that of minimizing aircraft delay. In this case, all means possible will be utilized to achieve service. The objective is formulated as:

$$Minimize \sum_{(ij)\in A} \sum_{k\in K} (c_{ij}^k d^k) x_{ij}^k \tag{3.2.11}$$

with $c_{ij}^k = 1$ for all arcs (i, j) for which commodity k misses service (on certain flight copies and artificial costs), and $c_{ij}^k = 0$ for all other arcs. Using this cost structure, the objective function represents a count of service failures within the recovery period.

3.3 Strengths of the Model

The strength of the model lies in its flexibility with respect to the recovery period, the objective function, and the parameters.

Recovery Period

The recovery period can begin and end at arbitrary points during the day. The user has the flexibility to define any preferred recovery period length. For example, the recovery period could be limited to a few hours to resolve the immediate implications of a disruption. The source and sink nodes for all aircraft and package commodities would be established and the "optimal" routing would be determined. We could just as easily consider a recovery period of a full day to evaluate delay propagation beyond the directly affected delivery flights. This case would include many more aircraft and recovery options, while requiring that the aircraft are positioned appropriately by the next day to resume scheduled operations.

Objective Function

As described earlier, the model allows for various user-defined objective functions. The objective during recovery may differ among disruptions, and the model presented allows the possibility of evaluating three objective functions and comparing the monetary cost, the number of service failures, and the minutes of delay resulting from each of these solutions.

Parameters

This modeling approach allows the user to generate solutions based on changing user preferences, namely: the cost of a service failure and the cost of operating a spare aircraft. By adjusting the cost of service failures on the artificial and delay arcs, one can provide strong incentives to ensure that certain shipments make service regardless of cost. Consider, for example, that the carrier has a contract with a valued customer who encourages the carrier to protect specific shipments. For example, a high volume vendor might exclusively use one ESDS carrier, and rely on it for timely delivery to its stores. The ESDS carrier might give higher priority to this vendor's shipments than to other vendors' shipments because the possible loss of goodwill from this high-volume vendor could pose a large decrease in long-term revenue for the carrier.

3.4 Computational Experience

The models were implemented in Java and solved using a OPL in conjunction with ILOG CPLEX 7.1 [12]. Please refer to Appendix B for the OPL code. CPLEX's MIP solver was called to solve the formulation. Computations were performed with a 1200 MHz workstation with 384 MB of RAM.

Chapter 4

Computational Experiments

To evaluate the recovery models and solution approach described in Chapter 3, we obtained data sets from a large ESDS carrier, from which we constructed disruption scenarios and then ran our models to generate recovery plans. The data involved the schedule of the NDA and SDA operations for the month of June 2004 for the entire fleet of 118 aircraft servicing 146 domestic and international destinations. The information provided for each flight leg includes an origin and destination city, a departure and arrival time, a unique flight number, and a specific aircraft assignment detailing the fleet type and aircraft identification number. Zulu time is used to report times. In addition, we have data that details the block times for every gateway pair that the carrier services, broken down by aircraft type.

Documentation of disruptions that occurred during the month of June 2004 was also provided by the ESDS carrier. We extract three scenarios representative of common disruptions, namely: *a mechanical failure en route to a hub*; *excess demand on a pick-up flight*; and *inclement weather*. We provide proof-of-concept by generating and evaluating solutions for recovery.

- *Mechanical Failure en route to a Hub.* We refer back to Example 1 from Chapter 3 to illustrate this situation. An aircraft scheduled for a pick-up route from a gateway to a hub experiences a mechanical failure. The packages at the gateway are stranded without means of reaching the hub for sorting. This disruption requires the Contingency group to evaluate the possibility of operating a spare aircraft to the disrupted gateway to pick-up the packages and deliver them to the hub for processing. Doing so would also require that affected outbound delivery flights from the hub would have to be delayed if the packages were to make their connections.

- *Excess Demand on Pick-Up Flight.* In this situation, a pick up flight could not transport all the demand at the gateway. While an initial estimate gauged that all of the volume could be transported on the pick-up flight from the corresponding gateway to the hub, it was determined only shortly before the scheduled departure of the pick-up flight that some demand could not be accommodated. This disruption results in the same options and considerations as Example 1, namely, whether the Contingency group will operate a spare aircraft to the disrupted gateway to rescue the stranded packages and deliver them to the hub for sorting, and whether outbound flights will be delayed to wait for late incoming packages.

- *Inclement Weather.* As discussed earlier, weather is the most common disruption during operations at an ESDS carrier. In one of the documented incidents, a lengthy operational ban due to thunderstorms and lightning prevented an aircraft from operating a pick-up route that included two flight legs prior to arriving at a regional hub for sorting (similar to Example 3 in the previous chapter). Packages at the second gateway had no means of reaching the regional hub for sorting unless a spare was deployed. Packages from the first gateway could be rerouted to the all-points hub. Flight delays for delivery flights out of both hubs were a possibility.

4.1 Underlying Assumptions

While we do not have complete information of the actual commodities on board aircraft at the moment of a disruption, we will assume three randomly assigned commodities and quantities to each origin gateway in the network, for illustrative purposes. For all commodities other than the ones at the disrupted gateway, we then created source nodes based on the commodity's arrival time into the hub from its origin gateway. Thus, we only consider one isolated disruption and assume that all other commodities and planes arrive at their locations as scheduled. We select a recovery period that includes immediately affected connecting flights and an immediate return of the spare aircraft to its assigned location after rescuing packages at the disrupted gateway. In setting up the network, we made the following assumptions and used the following parameters to characterize the network:

- *Sort time at hub.* When a plane arrives late into the hub for processing, we assume a sort time of 1.5 hours to process all commodities completely. Thus, only the delivery flights that have a scheduled departure time within 1.5 hours after the arrival of the late flight would have the possibility of postponement. We assume that the 1.5 hours is sufficient for unloading the packages from the late inbound flight, processing them at the hub and loading them onto outbound flights. In addition, not all of the delivery flights in the 1.5 hour window will necessarily be delayed. We need to identify the destinations of all the commodities on the late inbound flight, as the flights to the affected destinations that are also within the 1.5 hour window will be the only ones for which we will create flight copies.

- *Isolation from SDA network.* For the cases tested in the model, we omit SDA flights in the recovery period or solution. While the time windows for SDA flights might be blurred at the edges of the hub sort process, we ignore its possible interaction with the NDA network during disruptions.

- *Spare aircraft utilization.* Because the NDA operations include 21 spare aircraft throughout the network, the Contingency group has a choice of which spare aircraft to use when they decide that one is necessary. The choice is dependent on factors such as proximity, availability, and the potential for disruptions in other parts of the network during operations. In our approach, we assume that the spare closest in proximity is available and proceed with options based on that aircraft's block time, turn around time, and capacity.

- *Aircraft swaps.* We did not consider aircraft swaps (assigning an aircraft to a flight leg it was not originally assigned to) in our recovery approach.

- *Minimum Turn-Time.* The turnaround time for selected aircraft types is listed in Table 4.1. This is the amount of time that the aircraft must be on the ground.

Aircraft Type	Minimum Turn-Time
Boeing 727	+30
Boeing 757	+45
DC-8	1+00
Airbus A300	1+15
Boeing 747	2+00

Table 4.1: Minimum Turn-Time for Selected Fleet Types

- *Aircraft delays and cancellations.* In the Aircraft Delay Model defined in the previous chapter, we set the parameter for a flight cancellation equal to a full day, or 1440 minutes. This parameter reflects a carrier's views on canceling a flight and can be easily adjusted to reflect different philosophies.

- *Service failure costs.* A range of values from \$0 to \$20 is assigned to the cost of missing service to understand the sensitivity of the models to various delay costs.

4.2 Mechanical Failure en route to Hub

4.2.1 Observations of Recovery Process

In the actual situation from which this scenario was modeled, mechanics attempted to repair the aircraft while the Contingency group simultaneously launched to the disrupted gateway the closest spare aircraft with a hot spare crew onboard. The block time for the spare to reach the gateway was determined, as well as the time for the packages and spare to reach the hub for processing. Approximately 3,500 packages were stranded at the disrupted gateway when this mechanical failure occurred.

The Contingency group also had to determine which crew (scheduled or hot spare) to operate the flight leg from the gateway to the hub in order to minimize unproductive crew time. Contingency had to consider the one hour of pre-flight time dictated by crew regulations, which would have added more time to the recovery operation if the original crew operated the flight leg. By having the hot spare crew operate the flight leg to the hub, however, the process was determined to be more seamless.

As a result of Contingency's decision, 12 outbound flights from the regional hub were delayed for an average of 29 minutes. Approximately 1,000 packages missed service because of these delays.

4.2.2 Implementation of Modeling Approach

We directly implement the model on the reduced size network illustrated in Example 1 from Chapter 3. Please refer to Figure 3-2 for an illustration of the network. We construct the network by inserting nodes along each location axis at the appropriate times when aircraft and package commodities possibly arrive or depart a gateway. These times are determined according to the block times and minimum turn-times established by the carrier. In this example we allow one option of flight delay (one flight copy) that is able to service the late incoming commodity $(b - c)$ in its entirety. This option, however, is at the expense of missing service for the other commodity $(h^1 - c)$ with the same destination.

We evaluate the optimal strategies for each of the three objectives described earlier: minimizing the sum of operations costs and the cost of missed service, minimizing the number of minutes of delay, and minimizing the number of service failures. In the Aircraft Delay Model, we first determine the aircraft routing to minimize delay, and then follow with the Service Failure Model, as described in the previous chapter. Table 4.2.2 quantifies various metrics associated with each of these objective functions.

Even with this small network of flights, we can already see the complexity of the tradeoffs that occur during the recovery period. In this case, we chose parameters for plane capacity and package quantities where the spare had insufficient capacity to rescue all the volume

	Minimize Minutes of Delay and Subsequently Minimize Number of Service Failures	Minimize Number of Service Failures	Minimize Sum of Operations and Service Failure Costs
Service Failure Cost ($)	3,000	5,000	6,000
Operations Cost ($)	3,500	3,500	0
Total Cost ($)	6,500	8,500	6,000
Aircraft Delay (Minutes)	177	297	1,440
Number of Service Failures	300	200	500

Table 4.2: Results for Mechanical Failure Example

at the disrupted gateway. From the outset, we recognize that a complete rescue will not be possible, yet we still have to evaluate the worth of rescuing even a fraction of these commodities.

We see that the solution is to operate the spare aircraft in both instances of minimizing aircraft delay and number of service failures. The commodities serviced, however, will be different depending on the objective function considered. While the spare aircraft is flown to minimize total system-wide delay (rather than canceling the flight leg, and incurring a day of delay), subsequent flight legs are not delayed, resulting in service failures. The spare aircraft is flown to rescue packages, but the necessary measures downline in the schedule are not taken to ensure that those packages will make their subsequent connections. This should never occur in practice, and is a shortcoming of the model. We will discuss this in more detail later in this chapter.

As expected, the objective of minimizing service failures will take advantage of all the resources the carrier has available to its operations, regardless of cost. This is reflected in the operation of the spare aircraft and a higher aircraft delay than that observed by minimizing aircraft delay. By quantifying only the volume of packages that miss their service commitment, however, we have neglected to consider the relative worth of each commodity. While the objective of minimizing absolute service failures is minimized, we see that this model would have actually missed service on the highest yield commodities.

The resulting solution is dependent on the parameters associated with the cost of missing service and operations. With the costs assigned to the flight arcs in this example, minimizing the sum of operations and service failure costs results in missing service on all packages originating at the disrupted gateway. Even though the hybrid objective function results in the lowest monetary cost, we see that it results in the greatest aircraft delay (resulting from the parameter we assigned to the aircraft delay for flight cancellations) as well as the greatest number of service failures. If, instead, the carrier reweighted the cost parameters to reflect more "worth" of the commodities or a higher cost for missed service, the solution could have resulted in fewer service failures.

The objective that most closely matches the recovery plan chosen by the Contingency group is that of minimizing the number of service failures. The driving goal when they decided to launch the spare aircraft was to minimize the number of failures, without regard to cost of operations.

4.3 Excess Demand on Pick-Up Flight

4.3.1 Problem Description and Observations of Recovery Process

In this situation, volume arrived at a gateway from a nearby ground center that had suffered from a power outage. The power outage prevented the facility from making the initial sort splitting the volume between the all-points hub and regional hubs. As a result, all of the volume was routed to the all-points hub where the commodities could still be serviced. Over 1000 packages of NDA volume, however, were left behind at the gateway due to insufficient capacity from the gateway to the all-points hub. Hence, the Contingency group launched a nearby spare to recover the volume. After loading the packages, the aircraft departed for the all-points hub, one hour and 18 minutes after the originally scheduled flight had departed. The late arrival of the flight contributed to the all-points hub operating 69 flights late for an average of 13 minutes.

4.3.2 Implementation of Modeling Approach

By extracting the flight schedule that was affected by this flight disruption, we coded and tested our formulation for a reduced-scale model of the perturbation. We chose a recovery period that begins when the Contingency group receives notice that insufficient aircraft capacity exists. The recovery period we consider ends after the commodities have reached their destination gateways and aircraft have completed the flight legs that are affected by the commodities involved. The spare aircraft is assumed to fly a repositioning flight to its originating gateway after operating the flight leg in question. We build the network of nodes and arcs to include the source nodes of all planes and commodities, as well as sink nodes to reflect service and schedule commitments for packages and planes. The flights legs for the spare take into account block times and turnaround times necessary for the particular fleet types.

By incorporating the assumptions detailed above, we tested a representative model for a problem size of 629 rows and 850 columns with a corresponding matrix density of 0.0032. The problem has 2185 constraints. The number of rows corresponds to the number of nodes and the number of columns corresponds to the sum of the number of arcs (scheduled flight arcs, artificial arcs, and delay arcs) identified in the recovery period. We purposefully design the network so that the number of packages exceeds plane capacity. In addition, we create the example to have less capacity on a delivery flight than the inbound pick-up flight. Doing so will allow us to analyze our models' solutions involving the routing of planes and packages. We assign a range of costs for service failures from \$0 to \$20 and a plane capacity between 500 and 800 packages for the various aircraft types operating flight legs in the recovery period.

Table 4.3.2 provides a summary of the statistics resulting from running the model under the three objective functions. Under the same set of parameters for service failures and operations costs, we see that the spare aircraft was operated in each of the solutions for the three objective functions. The difference in total cost, however, was determined by the different service failures observed.

	Minimize Minutes of Delay and Subsequently Minimize Number of Service Failures	Minimize Number of Service Failures	Minimize Sum of Operations and Service Failure Costs
Service Failure Cost ($)	6,100	6,100	4,300
Operations Cost ($)	9,000	9,000	9,000
Total Cost ($)	15,100	15,100	13,300
Aircraft Delay (Minutes)	78	78	78
Number of Service Failures	550	550	550
Computation Time (CPU s)	2.54 + 1.95	1.96	2.32

Table 4.3: Results for Excess Demand Example

Multiple Optimal Solutions

While the Service Failure Model and the Hybrid Model both resulted in the same number of service failures, we see that the models chose different commodities to service, resulting in a different cost. As expected, the commodities chosen to miss service in the Hybrid Model had a lower cost than the ones that were serviced with the Service Failure Model. Thus, in this case, minimizing service failures results in multiple optimal solutions. Both solutions involved an identical aircraft routing through the constructed network. The solution returned by implementing the Hybrid model serviced all of the high yield packages in the network. In the Service Failure Model, however, we do not have any incentive to consider the worth of the different commodities, and thus this model served an equal number of high and low yield commodities. We could easily reweight the parameters for costs on the commodities in the Hybrid Model and observe a complete swap of the packages serviced.

The Aircraft Delay Model in this case also illustrates multiple optimal solutions. The amount of delay in this solution is the same as achieved by the Hybrid Model, but the Aircraft Delay Model performed similarly to the Service Failure Model by indiscriminately servicing the commodities. In the second step of minimizing the quantity of service failures, if we had instead minimized the *cost* of service failures, we would achieve the solution obtained from the Hybrid Model.

Thus, in this example, we can solve all three Models using their respective objective functions, and arrive at the same solution. This occurs because the lowest yield packages that are forced to miss service in the Hybrid Model coincide with the packages that cannot fit on the plane (regardless of cost). The solution would have missed service on a different set of packages if the late inbound packages incurred a higher cost, forcing the delayed outbound flight to forgo servicing packages from other origins already at the hub.

4.4 Inclement Weather

4.4.1 Problem Description and Observations of Recovery Process

A disruption scenario experienced by the carrier involved a lengthy operational ban for thunderstorms and lightning that prevented an aircraft from operating a pickup route that included two flight legs prior to arriving at a regional hub for sorting (similar to Example 3 in the previous chapter). Both of the pick-up gateways had approximately 1,500 packages destined for the regional hub. The Contingency group decided to launch the closest spare aircraft to rescue the volume from the second gateway and deliver it to the regional hub. Packages from the first gateway were rerouted to the all-points hub. Both of these actions required assessing the time of arrival that the late flights would reach the regional and all-points hubs for processing. Along with other disruptions that night, this situation contributed to six flights being delayed out of the regional hub for late packages from the second gateway, in addition to 73 flights being delayed out of the all-points hub.

4.4.2 Implementation of Modeling Approach

The model was coded and tested for a representative problem containing 698 rows and 966 columns with a corresponding matrix density of 0.0021 with 813 constraints. Flight copies were generated for delivery flights departing both hub locations. The recovery period includes directly impacted flights departing both hubs. The spare aircraft is also assumed to immediately operate a ferry flight back to its original location after it is deployed. Table

4.4.2 illustrates the results from implementing each of the objective functions.

	Minimize Minutes of Delay and Subsequently Minimize Number of Service Failures	Minimize Number of Service Failures	Minimize Sum of Operations and Service Failure Costs
Service Failure Cost ($)	3,900	3,900	5,900
Operations Cost ($)	6,000	6,000	0
Total Cost ($)	9,900	9,900	5,900
Aircraft Delay (Minutes)	23	23	1,560
Number of Service Failures	300	300	1,200
Computation Time (CPU s)	1.23 + 0.94	0.96	1.56

Table 4.4: Results for Inclement Weather Example

While the Excess Demand Example illustrates how the three models can generate the same solution, this Inclement Weather Example illustrates that quite different solutions are obtainable. In this experiment, the Hybrid Model did not deploy the spare aircraft during operations because missing service on the lowest yield packages could be accomplished without the use of an additional aircraft. This results in a relatively large number of service failures and aircraft delay (because of the cancellation of the flight).

The parameters chosen in this example result in the cost of service failures in the Hybrid Model ($5,900) being less than the cost of operating the spare ($6,000). If the Service failure cost had been slightly higher, we would have observed the deployment of the spare aircraft. The difference in the number of service failures associated with the solutions to the three models is quite large in this example, driven by the considerable differences in relative "worth" assigned to the commodities.

4.5 Evaluation

Important observations were made while evaluating the three Models:

- *Number of Service Failures.* For the Hybrid Model, the number of service failures increases when either the cost of missing service is low or the cost of operations is high.

- *Operation of Spare Aircraft.* An available spare will always be flown for the Service Failure Model. The objective function of minimizing service failures illustrates the use of all available resources during operations regardless of cost.

- *Delay Model.* Given a resulting delay of one day due to a flight cancellation, the Delay Model will not choose to cancel a flight because doing so is more costly than delaying that flight.

- The three objective functions can produce different or similar solutions. The low yield packages that are candidates for missing service in the Hybrid Model may coincide with packages that would miss service when minimizing the number of service failures.

The multicriteria objective function illustrates a tradeoff between operations costs and service failures. The solution obtained by this approach will be Pareto optimal, which does not allow for an improvement in service without an increase in cost.

Canceling a flight leg is often the easiest way to minimize delay in the passenger airline industry. Doing so reduces delay propagation in downline flights. This approach, however, is not practical in the ESDS industry because of the high premiums charged for not servicing NDA commodities. Passenger airlines are not required to provide passengers money-back guarantee if the passenger experiences a late arrival, but ESDS carriers are financially responsible for delivering on-time service. Hence, flight cancellations typically are not effective in meeting an ESDS carrier's goals.

In each of the observations of the recovery process, the Contingency group's approach is most consistent with the Service Failure Model. Cost savings in the form of operations costs are not a driving goal because the available spare aircraft are already assumed to be an integral part of operations and a sunk cost. The ESDS carrier's operating philosophy strongly reflects the desire to minimize missed service.

When the cost of the commodities on the aircraft is much higher than the cost of operations, the Hybrid Model is not very relevant. In situations where the disparity is not as high and the costs are more balanced, however, the Hybrid Model can be a useful tool to determine a recovery solution. The Hybrid Model can be beneficial to an ESDS carrier in situations where missed service is unavoidable and the carrier needs to determine the most cost effective manner to miss service with constrained resources. In many circumstances, however, the Service Failure Model is sufficient on its own during operations.

We should also note, however, that the Service Failure Model only considers the quantity of packages that miss service. A natural modification to this approach would be to consider the cost of the service failures, an easy modification to the model.

4.6 Shortcomings of our Approach

With the parameters chosen, aircraft delay minutes increase as cancellations are introduced to the solution. This misses situations in which flight cancellation and the utilization of spare aircraft are in fact beneficial to the overall solution. Another issue is that the The Delay Model will always operate spare aircraft because doing so is the least "expensive" option. The spare will be flown, but none of the outbound delivery flights from the hub will be held because doing so will increase delay. This solution should never occur in practice because of its extra operating costs are incurred by no service failures are averted.

Our model does not consider maintenance schedules or extended recovery periods beyond the immediate disruption. If we had included an extended recovery period, issues of delay propagation would become more apparent, and flight copies would have to be generated for every subsequently affected flight. The problem would grow dramatically in size and real-time solutions could be a challenge.

The IT requirements to support such a model are extensive. As we have seen, the model requires a number of few inputs, such as a flight schedule, block times, turn around times, spare aircraft availability and location, commodity service requirements, as well as costs. While the flight schedule and aircraft performance characteristics are readily available, real-time information of the commodities involved during recovery is not as accessible. While our model randomly generated three commodities and quantities for each origin in our proof-of-concept model, an ESDS carrier might have thousands of commodities aboard one aircraft en route to the hub for sorting. Thus, rather than solving a problem involving a large quantity of several commodities, we would have to solve a problem involving thousands of commodities in smaller quantities. The incorporation of such commodities in the network would drastically increase the size of the network with the addition of nodes and arcs of each type (ground, scheduled, and artificial).

A seamlessly integrated system that informs the user of the commodities currently on and assigned to an aircraft is necessary for the implementation of the model. An accurate reflection of the costs is also assumed in the model, which is difficult to obtain in real-time. If each package is a standard envelope/letter that is charged the NDA rate and we know the value of loss of goodwill, we can use these costs to in the model.

Aircraft capacity is more often measured by payload, or weight, rather than by the number of packages. The carrier transports commodities of many shapes, sizes, and weights. Thus, commodities (with the same origin and destination) might vary greatly in terms of premiums charged. Thus, our assumption of considering all commodities to have the same cost of missing service might not be accurate.

4.7 Conclusions

Solving realistic recovery problems for an ESDS carrier is very complex. The strategy we have presented uses a network flow model to model aircraft and commodities in the network during irregular operations. Our approach is a computationally attractive optimization model that yields feasible solutions to the experiments tested.

In theory, this model can be extended to include a larger scope of flight legs and commodities. The primary challenge, however, is the timely generation of the recovery network.. We address model improvements in the next chapter.

Chapter 5

Conclusions and Future Research Directions

5.1 Contributions

In this thesis, we analyze the **ESSR** problem for an ESDS carrier. Our contributions are as follows:

- We develop a formulation to address recovery options during irregular operations at an ESDS carrier. The proposed model not only includes options for delaying and canceling flights, but also incorporates the possibility of rerouting aircraft and package commodities. We provide proof of concept by implementing our model and performing computational experiments using representative realistic disruption scenarios for a large carrier. The time-sensitive nature of the ESDS industry requires real-time solution strategies addressing service requirements. The fast run-times of our approaches demonstrate their potential to provide decision support during schedule disruptions.

- We demonstrate the flexibility of our methodology by presenting three models to optimize various metrics during the **ESSR** recovery process. These three models are the 1) *Hybrid Model*, which minimizes the combined cost of operations and cost of service failures, the 2) *Aircraft Delay Model*, which minimizes aircraft schedule delay, and the 3) *Service Failure Model*, which minimizes the number of service failures.

- We analyze the effects of the different objective of each model. Our approach is of practical use in that it provides alternative solutions, each making different trade-offs between the cost of operations and service failures. Hence, the effects of different recovery policies and strategies can be evaluated with our approach. Moreover, our methodology provides users with flexibility to define the length of the recovery period, and different values for enabling decision makers to quantify the effects of different recovery strategies and operating policies.

5.2 Future Research Directions

This thesis provides a framework for automating approaches to recovery during irregular operations at an ESDS carrier. In developing the optimization model and approaches to evaluate recovery options for an ESDS carrier, we identified possible extensions to our approach and future areas of research.

5.2.1 Challenges to Implementation

While our model was evaluated using a simplified, reduced-size network, the IT requirements to support such a model are extensive for the entire NDA network at an ESDS carrier. Generating the network of nodes and arcs for all commodities and aircraft involved in the recovery period is computationally intensive. Ideally, this process should be automated. Another challenge in solving operational problems is to obtain real-time data of the physical system's characteristics. These include: costs and quantities of package commodities in the network, current and scheduled locations of aircraft during the recovery period, possible routing of aircraft according to each hub's service territory, remaining crew flying time, and remaining capacity on specific aircraft and routes.

5.2.2 Extensions to our Approach

Specific areas of future research include:

- *Considering the propagation of delay.* Our methodology can easily incorporate a longer recovery period taking into account the propagation of delay on flight legs downline from the disruption. These instances should be solved to further evaluate the performance of the model.

- *Incorporating aircraft swaps.* In the case that an aircraft requires only a minor repair, but another aircraft can be launched immediately, an aircraft swap might be a feasible solution. An aircraft swap is defined as assigning an aircraft to a flight leg to which it was not originally assigned. This problem includes reassigning aircraft of different fleet types.

- *Considering aircraft maintenance requirements and crew schedules.* Aircraft maintenance requirements and crew work schedules are developed to adhere to federal regulations. Operating aircraft (both scheduled and spare) in strict accordance to the schedule eliminates deviation from aircraft maintenance and crew schedules. Because schedules are rarely operated as planned, a more complete recovery approach would minimize the deviation from the maintenance and crew schedules as part of the recovery process.

- *Incorporating robustness into the planning phase.* As discussed in Chapter 1, the possibility of schedule disruptions can be considered explicitly during the planning phase of schedule design by adding slack, excess capacity, or other forms of flexibility and redundancy into the schedule. By explicitly addressing the possibility of delays and cancellations when generating the schedule, operations personnel might have to generate recovery options less frequently, or be able to provide cost effective recovery plans in less time because the number of recovery options are greater or more obvious.

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix A

Glossary

AIRCRAFT SWAP: assigning an aircraft to a flight leg it was not originally assigned to

BLOCK TIME: flying time plus taxi time

EHDT: Earliest Hub Departure Time; earliest time that an aircraft may depart from the hub on a delivery route

EPT: Earliest (Gateway) Pick-Up Time; the earliest time that an aircraft may depart from the gateway on a pick-up route

EPTC: Earliest Pick-Up Time from Center; earliest time that commodities may depart from the ground center

ESDS: Express Shipment Delivery Service

ESSR: Express Shipment Service Recovery

LDT: Latest (Gateway) Delivery Time; latest time that commodities can arrive at destination gateways

LDTC: Latest Delivery Time to Center; latest time that commodities can be delivered to the ground center

LHAT: Latest Hub Arrival Time; latest time that a pick-up flight can arrive at the hub

MINIMUM TURN AROUND TIME: Minimum time for an aircraft to stay on the ground; this includes the time to unload and load packages, as well as refuel

NDA: Next Day Air; expedited delivery service for commodities; these operations occur during the night

SDA: Second Day Air; 48-hour delivery service for commodities; these operations occur during the day

Appendix B

OPL Code

```
//*read in the arc labels*/
{string} planeArcs              < "C:/java/0795test/plane_arcs.txt";
{string} packageArcs            < "C:/java/0795test/package_arcs.txt";

/*read in the node labels*/
{string} planeNodes             < "C:/java/0795test/PlaneNodeNumbers.txt";
{string} packageNodes           < "C:/java/0795test/PackageNodeNumbers.txt";

/*read in the cost of each arc */
int Cplane[planeArcs] < "C:/java/0795test/plane_cost.txt";
int Cmisspack[packageArcs] < "C:/java/0795test/missedPackage_cost.txt";
int Cpackage[packageArcs]  < "C:/java/0795test/package_cost.txt";
//int Cminutes[planeArcs]  < "C:/java/0795test/planeMinutes_cost.txt";

/*read in the node-arc incidence matrix for the planes and packages*/
int Aa[planeNodes, planeArcs]     < "C:/java/0795test/plane_naMatrix.txt";
int Ab[packageNodes, packageArcs] < "C:/java/0795test/package_naMatrix.txt";
```

```
/*read in the demand matrix for each node */
int Da[planeNodes] < "C:/java/0795test/plane_demand.txt";
int Db[packageNodes] < "C:/java/0795test/package_demand.txt";

/*read in the package quantities*/
int Qb[packageArcs ] < "C:/java/0795test/package_quantity.txt";

/*read in the capacity of each arc */
int u[planeArcs] < "C:/java/0795test/plane_capacity.txt";

var int y[planeArcs] in 0..1;
var float+ x[packageArcs];

//HYBRID MODEL
minimize sum(j in packageArcs) Cpackage[j] * Qb[j] * x[j]
+ sum(t in planeArcs) Cplane[t] * y[t]

//SERVICE FAILURE MODEL
//minimize sum(j in packageArcs) Cmisspack[j] *Qb[j] * x[j]

//DELAY MODEL
//minimize sum(t in planeArcs) Cminutes[t] * y[t]

subject to{

//CONSERVATION OF FLOW CONSTRAINTS
forall(i in packageNodes)
sum (j in packageArcs) Ab[i,j]*x[j] = Db[i];
```

```
forall(k in planeNodes)
sum (t in planeArcs) Aa[k,t]*y[t] = Da[k];

//PLANE CAPACITY CONSTRAINT
forall (t in planeArcs)
sum (j in packageArcs) Qb[j]*x[j] <= sum{t in planeArcs} u[t]*y[t];

//BOUND
forall (j in packageArcs)
x[j] <= sum (t in planeArcs) y[t];

//PLANES CANNOT FLOW ON ARTIFICIAL ARCS
y["A30_SDF_SLC_14_35"] = 0;

forall(j in packageArcs) x[j] <= 1;
forall(t in planeArcs) y[t] <= 1;

};
display (t in planeArcs: y[t] >0) y[t];
display (j in packageArcs: x[j] >0) x[j];

display sum(j in packageArcs) Cpackage[j] * Qb[j] * x[j];
display sum(t in planeArcs) Cplane[t] * y[t];
display sum(j in packageArcs) Cmisspack[j] * Qb[j] * x[j];
display sum(t in planeArcs) Cminutes[t] * y[t];
```

THIS PAGE INTENTIONALLY LEFT BLANK

Bibliography

[1] Y. Ageeva. Approaches to incorporating robustness into airline scheduling. Master's thesis, Massachusetts Institute of Technology, August 2000.

[2] M.F. Argüello. *Framework for exact solutions and heuristics for approximate solutions to airlines' irregular operations control aircraft routing problem.* PhD thesis, University of Texas, Austin, 1997.

[3] M.F. Argüello, J.F. Bard, and G. Yu. A GRASP for aircraft routing in response to groundings and delays. *Journal of Combinatorial Optimization*, 5:211–228, 1997.

[4] A. Armacost. *Composite Variable Formulations for Express Shipment Service Network Design.* PhD thesis, Massachusetts Institute of Technology, 2000.

[5] A. Armacost, C. Barnhart, K. Ware, and A Wilson. Ups optimizes its air network. *Interfaces*, 34(1):15–25, January-February 2004.

[6] S. Bratu. *Airline Passenger On-Time Schedule Reliability: Analysis, Algorithms and Optimization Decision Models.* PhD thesis, Massachusetts Institute of Technology, 2003.

[7] S. Chebalov and D. Klabjan. Robust airline crew scheduling: move-up crews. In *Proceedings of the 2002 NSF Design, Service, and Manufacturing Grantees and Research Conference*, 2002.

[8] M. Clarke. Irregular airline operations: A review of the state-of-the-practice in airline operations control centers. Working Paper, International Center for Air Transportation, Massachusetts Institute of Technology, 1997.

[9] M. Ehrgott and D.M. Ryan. Constructing robust crew schedules with Bicriteria optimization. *Journal for Multi-Criteria Decision Analysis*, 11(3):139–150, May/June 2002.

[10] FedEx Annual Report, 2003.

[11] J.A. Filar, P. Manyem, and K. White. How airlines and airports recover from schedule perturbations: A survey. *Annals of Operations Research*, 108:315–333, 2001.

[12] ILOG CPLEX 7.1.

[13] A.I.Z. Jarrah, G. Yu, N. Krishnamurthy, and A. Rakshit. A decision support framework for airline flight cancellations and delays. *Transportation Science*, 27(3):266–280, 1993.

[14] L.S. Kang and J.P. Clarke. Degradable airline scheduling. Technical report, Global Airline Industry Program, Massachusetts Institute of Technology, 2002.

[15] D. Kim. *Large Scale Transportation Service Network Design: Models, Algorithms and Applications.* PhD thesis, Massachusetts Institute of Technology, 1997.

[16] S. Lan. *Planning for Robust Airline Operations: Optimizing Aircraft Routings and Flight Departure Times to Achieve Minimum Passenger Disruptions.* PhD thesis, Massachusetts Institute of Technology, 2003.

[17] L. Lettovsky, E.L. Johnson, and G.L. Nemhauser. Airline crew recovery. *Transportation Science*, 34(4):337–348, November 2000.

[18] J.M. Rosenberger, E.L. Johnson, and G.L. Nemhauser. A robust fleet-assignment model with hub isolation and short cycles. *Transportation Science*, 38(3):357–368, August 2004.

[19] A.J. Schaefer, E.L. Johnson, A.J. Kleywegt, and G.L. Nemhauser. Airline crew scheduling under uncertainty. Technical report, The Logistics Institute, Georgia Institute of Technology, 2001.

[20] S. Shen. *Logistics Service Network Design: Models, Algorithms, and Applications.* PhD thesis, Massachusetts Institute of Technology, 2004.

[21] D. Teodorovic and S. Guberinic. Optimal dispatching strategy on an airline network after a schedule perturbation. *European Journal of Operational Research*, 15(2):178–182, February 1984.

[22] D. Teodorovic and G. Stojkovic. Model for operational daily airline scheduling. *Transportation Planning and Technology*, 14(4):273–285, 1990.

[23] D. Teodorovic and G. Stojkovic. Model to reduce airline schedule disturbances. *Journal of Transportation Engineering*, 121(4):324–331, 1995.

[24] B.G. Thengvall, J.F. Bard, and G. Yu. Balancing user preferences for aircraft schedule recovery during irregular operations. *IIE Transactions*, 32:181–193, 2000.

[25] UPS Annual Report, 2003.

[26] UPS Fact Sheet. http://www.ups.com/content/us/en/about/facts/worldwide.html.

[27] UPS Rate and Service Guide, January 2005.

[28] A. West. Standard & Poor's Commercial Transportation Industry Survey, February 2005.

[29] S. Yan and Y.P. Tu. Multifleet routing and multistop flight scheduling for schedule perturbation. *European Journal of Operational Research*, 103:155–169, 1997.

[30] S. Yan and D.H. Yang. A decision support framework for handling schedule perturbation. *Transportation Research Part B*, 30(6):405–419, 1996.

[31] S. Yan and H. Young. A decision support framework for multi-fleet routing and multi-stop flight scheduling. *Transportation Research Part A*, 30(5):379–398, 1996.

[32] J.W. Yen and J.R. Birge. A stochastic programming approach to the airline crew scheduling problem. In *Proceedings of the 35th Annual Conference of the Operational Research Society of New Zealand*, pages 19–30, Wellington, New Zealand, December 2000.

[33] G. Yu, Argüello, M. Song, S. McCowan, and A. White. A new era for crew recovery at Continental Airlines. *Interfaces*, 33(1):5–22, 2003.