The Influence of Network Effects and Yield Management on Airline Fleet Assignment Decisions

by

András Farkas
Diploma Eng., Technical University of Budapest, Hungary (1989)
S.M., Massachusetts Institute of Technology, Cambridge, MA (1993)

Submitted to the Department of Civil and Environmental Engineering in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY in Transportation Systems at the MASSACHUSETTS INSTITUTE OF TECHNOLOGY February 1996

© Massachusetts Institute of Technology 1996. All rights reserved.

Author

Department of Civil and Environmental Engineering
October 27, 1995

Certified by

Peter P. Belohaba
Assistant Professor of Aeronautics and Astronautics
Thesis Supervisor

Accepted by

Joseph Sussman
Chairman, Departmental Committee of Graduate Students
The Influence of Network Effects and Yield Management on Airline Fleet Assignment Decisions

by

András Farkas

Submitted to the Department of Civil and Environmental Engineering on October 27, 1995 in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Transportation Systems

Abstract

The fleet assignment problem is to find the optimal assignment of available aircraft types to a flight schedule (fixed flights and departure times) such that expected profits are maximized. The correct estimation of spill (or passenger demand turned away) and spill costs (or non realized revenues), is an integral part of the determination of optimal aircraft capacities in the airline fleet assignment process. The purpose of this dissertation is to demonstrate the extent to which network effects (leg-dependence effects) and airline yield management practices can affect the correct estimation of both spill and spill costs and influence the fleet assignment decisions. In light of this, many of the fundamental issues related to incorporating network and yield management effects into the fleet assignment decision process are addressed. We show the limitations of the state-of-the-practice methods and present new approaches for incorporating network and yield management effects into the fleet assignment decision process.

We address the possible implications of network effects on the fleet assignment solution. Starting from the deterministic demand case, we expand the analysis to more representative stochastic demand cases. We show that due to the censoring effects in leg-dependent networks, traditional approaches that use leg-independent unconstrained demand densities as a basis for spill estimation can overestimate both loads (expected revenues) and spill (spill costs). We also show that the overestimation from leg-dependence effects is not uniform across the flight legs or at different demand to capacity ratios. By analyzing network effects in different demand and network situations, we identify the important characteristics that influence leg-dependence effects on spill estimates in a network.

It is demonstrated that yield management systems, through their booking limits, influence both the total number of passengers spilled and their fare class mix, which in turn affects the total spill costs of given aircraft capacities assigned to the flight legs. On the basis of the numerical examples presented, the direction and magnitude of the bias in spill estimates from the currently used approaches (vertical and horizontal aggregation bias) are evaluated. The recursive spill models introduced in the dissertation provide more accurate approaches for estimating spill cost for a single flight leg.

The dissertation also reveals that different yield management approaches can influence the origin-destination passenger flows in the network in different ways and that under different yield management approaches, the optimal fleet assignment solutions may also be different. Through typical airline examples, it is demonstrated that ignoring yield management and/or network effects in the spill cost estimation may lead to sub-optimal fleet assignment solutions.

With this in mind, several estimation approaches and optimization models were developed. It is formally shown that under typical fare structures, fleet assignment formulations constitute a non-convex null in the feasible set. This dissertation presents a new iterative Nested embedded Simulation Fleet Assignment approach (NSFA), that under certain conditions overcomes the problem of non-convexity. The algorithm leverages on the existing fleet assignment solution techniques and solution methodologies, enhanced by an embedded Monte Carlo based simulation.

Thesis Supervisor: Professor Peter P. Belobaba, Assistant Professor of Aeronautics and Astronautics
Acknowledgment

I would like to express my greatest gratitude to Professor Peter Belobaba, for his advise, supervision, and support during the course of this research. His knowledge and experience in the field contributed in large measures to the success of this dissertation. I would like to thank him for introducing and guiding me through the wonderful area of air transportation. Most of all, however, I would like to thank him for his unforgettable friendship, understanding, and full support during the good and also during the hardest times of my stay at M.I.T.

I express my deep sense of gratitude to my doctoral committee members, Professor Cynthia Barnhart and Professor Bob Simpson, for their guidance, advice, and encouragement. I have learned a great deal from them both during the numerous committee meetings and from their classes. I would also like to thank Professor Amedeo Odoni, Professor Haris Koutsopoulos, and other professors at M.I.T. for their support and guidance.

I am also indebted to Matthew Berge and Craig Hoppersad, of Boeing Co., for being interested in and supportive of my research. Thank you for the very useful discussions, great ideas, and worthwhile inputs.

I would be failing in my duty if I did not mention my appreciation to Delta Airlines and KLM Royal Dutch Airlines for their partial financial support during my stay at M.I.T.

I would like to wish to thank all my friends at M.I.T., whose friendship made my stay more enjoyable and the tough times more bearable. First of all I would like to thank my eternal friend Amalia “Milly” Polydoropoulou for her warm friendship and support during the good and hard times. Many thanks also to Adriana Bernardino, Antulio Richetta, Rabi Mishalani, Rajesh Shenoi, Gábor Hetyei, András Benczur, Zoltán Szabó, Dinesh Gopinath, Chris Caplice, Daeki Kim, Kalidas Ashok, Francisco Jauffred, John Bowman, Yang Qi, Jai-kue Park, John Wilson, Jiang Chang, Naomi and Tamar Gendler, the Shiftan family, all my friends and fellow students, too numerous to list. I had great fun with you. I am also thankful to my fellow ice-hockey players, with whom I spent great times on the ice.

I would like to dedicate this dissertation to my dearest Father and Mother, whom I will love and miss forever, and whom I have to thank for everything:

Drága Édesapám és Édesanyám! Ezt a dolgozatot nektek ajánlom, örök hálával, köszönettel és emlékül.
Köszönök Nektek mindent Drága Szüleim!
# Contents

1. INTRODUCTION .......................................................................................................................... 12
1.1 Goal and Structure of the Dissertation ...................................................................................... 14

2. THE AIRLINE CAPACITY PLANNING PROCESS ................................................................. 18
2.1 The Traditional Fleet Assignment Approach (Leg-Independence) ......................................... 18
2.2 Network Effects (Leg-dependence) ............................................................................................ 20
2.3 Supply of Fare Products and Yield Management Systems ...................................................... 22
  2.3.1 Conventional Yield Management (YM) Practice ............................................................... 23
  2.3.2 The OD seat inventory control problem ............................................................................. 24
2.4 The Relationship of Fleet Assignment and Yield Management ............................................. 25
2.5 The Fundamental Dichotomy of Supply and Demand ............................................................. 28
2.6 Basic Assumptions and Terminology ......................................................................................... 29
2.7 Summary ................................................................................................................................ 31

3. PREVIOUS WORK AND CURRENT PRACTICES .................................................................. 32
3.1 Yield Management Optimization Algorithms .......................................................................... 32
3.2 Yield Management System Approaches .................................................................................... 35
  3.2.1 Conventional Fare Class (Yield) Based Control (FC) ....................................................... 35
  3.2.2 Value Class Control (VC) .................................................................................................. 37
  3.2.3 Displacement Costs Bid Price OD Heuristics (BD) .......................................................... 38
3.3 Fleet Assignment Decision Models ......................................................................................... 40
3.4 Overview of Spill Models .......................................................................................................... 45
3.5 Summary ................................................................................................................................ 50

4. NETWORK EFFECTS AND LEG-DEPENDENCE IN AIRLINE NETWORKS .................. 51
4.1 Network Effects (Leg-dependence) in Deterministic Networks ............................................. 53
  4.1.1 Traditional Fleet Assignment Models: Leg-Independent Approach ................................. 53
  4.1.2 Spill Calculation Considering Leg-Dependence ................................................................. 56
  4.1.3 Fleeting Solutions Using Leg-independent and Leg-dependent Estimates .................... 60
4.2 Network Effects and Leg-dependence (Stochastic Demand Case) ......................................... 63
  4.2.1 Calculating the Achievable Traffic Curve .......................................................................... 63
  4.2.2 Calculating Spill and Spill Costs ....................................................................................... 71
4.3 The Interpretation of Spill in Networks .................................................................................... 79
4.4 Summary ................................................................................................................................ 80

5. ATTRIBUTES INFLUENCING LEG-DEPENDENT EFFECTS ........................................... 82
5.1 Basis for Leg-dependence: Connecting Passenger Ratio ....................................................... 82
5.2 Amount of Spill ......................................................................................................................... 83
5.3 Booking Patterns of Origin-Destination Demands .................................................................. 82
5.4 Network Connectivity -- Origin-Destination Demand Distribution ....................................... 84
  5.4.1 Dispersion of Censoring Effects ....................................................................................... 85
  5.4.2 Concentration of Censoring Effects .................................................................................. 89
5.5 The Direction of Leg-Dependence Effect Propagation .......................................................... 92
5.6 Boundaries of Leg-dependence Effect Propagation .............................................................. 93
  5.6.1 Leg-dependence Effect Propagation in Case of Stochastic Demand .............................. 97
5.7 Summary ................................................................................................................................ 100

6. APPROACHES FOR CALCULATING LEG-DEPENDENT LOAD AND SPILL ESTIMATES IN NETWORKS ........................................................................................................... 101
6.1 DETERMINISTIC DEMAND CASE................................................................................. 101
6.2 STOCHASTIC DEMAND CASE.................................................................................. 103
   6.2.1 Boeing's P-Flow Model....................................................................................... 103
   6.2.2 The SMCS Model -- Calculating Load and Spill by Monte Carlo Simulation .. 105
6.3 SUMMARY.................................................................................................................. 110

7. YIELD MANAGEMENT IMPACTS ON AIRLINE SPILL ESTIMATION .................. 112
   7.1 CURRENT PRACTICE IN SPILL ESTIMATION FOR FLEET ASSIGNMENT .......... 112
   7.2 A NEW APPROACH FOR SPILL ESTIMATION UNDER YIELD MANAGEMENT .... 114
   7.3 VERTICAL AGGREGATION BIAS.......................................................................... 119
      7.3.1 Evaluation of Vertical Aggregation Bias......................................................... 120
   7.4 HORIZONTAL AGGREGATION BIAS...................................................................... 131
      7.4.1 Evaluation of Horizontal Aggregation Bias.................................................. 131
   7.5 EXAMPLE OF HORIZONTAL AND VERTICAL AGGREGATION COMBINED ......... 137
   7.6 THE INFLUENCE OF VARIOUS YIELD MANAGEMENT SYSTEMS ON SPILL AND SPILL COSTS IN NETWORKS 141
      7.6.1 The Modeled YM Approaches....................................................................... 141
      7.6.2 Leg-independent Spill Estimates Under Different YM Systems.................. 146
      7.6.3 Comparison of Leg-Dependent Estimates..................................................... 152
   7.7 ESTIMATING LEG-DEPENDENT SPILL COSTS IN NETWORKS FOR A LONGER TIME PERIOD .... 161
   7.8 FLEET ASSIGNMENT SOLUTIONS UNDER DIFFERENT SPILL ESTIMATION APPROACHES 163
      7.8.1 Evaluation of the Different Spill Cost Estimates on Fleeting Solutions in Leg-independent Networks ..................................................................... 163
      7.8.2 Evaluation of the Different Spill Cost Estimates on the Fleeting Solutions in Leg-dependent Networks........................................................................... 169
   7.9 CONCLUSION -- RECOMMENDED SPILL AND SPILL COST ESTIMATION APPROACH ...................................................... 173

8. FLEET ASSIGNMENT CONSIDERING LEG-DEPENDENCE AND YIELD MANAGEMENT EFFECTS ................. 176
   8.1 FLEET ASSIGNMENT WITH OD FLOW CONTROL ............................................. 176
      8.1.1 Combined Passenger Flow and Fleet Assignment Models.......................... 177
      8.1.2 Column Generation Based Combined Passenger Flow and Fleet Assignment Models 183
      8.1.3 Applicability and Relevance Issues............................................................... 190
   8.2 FLEET ASSIGNMENT WITH NO OD FLOW CONTROL ....................................... 191
      8.2.1 Nested Algorithm with Embedded Simulation.............................................. 202
   8.3 SUMMARY............................................................................................................. 215

9. CONCLUSIONS.............................................................................................................. 217
   9.1 SUMMARY OF RESEARCH FINDINGS AND CONTRIBUTIONS ......................... 217
   9.2 FURTHER RESEARCH DIRECTIONS ................................................................... 221

APPENDIX...................................................................................................................... 223

BIBLIOGRAPHY.............................................................................................................. 224
List of Figures

Figure 2.1: Simple example for portraying network (leg-dependence) effects ........................................21
Figure 2.2: Recursive relationship of Fleet Assignment and Yield Management decisions via the airline
passenger network .........................................................................................................................25
Figure 2.3: Inherently assumed “split” relationship with no (or lagged) feed-back effect between the
Fleet Assignment and Yield Management decisions in a decision cycle in the traditional fleet
assignment approaches .............................................................................................................27
Figure 2.4: The underlying flight network and the two additional network layers .................................28
Figure 3.1: Nested booking limits EMR curve and protection levels on a single flight leg ....................34
Figure 4.2: Airline network with OD demands represented as local and connecting ..........................54
Figure 4.3: Three-leg network example with differing connecting OD fare values ............................59
Figure 4.4: Censored probability distribution -- Capacity limit censoring only one OD demand .........64
Figure 4.5: Network example with probabilistic OD demands ..........................................................64
Figure 4.6: Achievable OD traffic densities on leg 1 for the case when capacity limits censor local and
connecting demands (simulated curves) ......................................................................................67
Figure 4.7: Achievable OD traffic density for the case when capacity limits censor local and connecting
demands -- reduced standard deviation for local demand (capacity on leg 2 =100 and local std=5) ......68
Figure 4.8: Achievable traffic densities on leg 1 (Ach_Traffic) and the leg-independent (Unconstr.
Dem.) leg demand ......................................................................................................................70
Figure 4.9: Calculation of local spill on leg 1 -- (l1 and x is the unconstrained local and connecting OD
demand, and w is the connecting achievable OD traffic on leg 1) ................................................74
Figure 5.1: Booking curves (Booking patterns) of a later booking (a) and earlier booking (b) OD
demand .........................................................................................................................................84
Figure 5.2: Example for dispersion of the leg-dependence effects (Flight leg 1 spills) ............................85
Figure 5.3: Simulated achievable traffic densities of a feeding leg and the unconstrained connecting
demand densities .........................................................................................................................88
Figure 5.4: Example of concentrating leg-dependent effects (legs 1.n spill and leg 1 has large
capacity) ........................................................................................................................................89
Figure 5.5: Simulated achievable traffic densities and the leg-independent total unconstrained leg
demand densities for different, n, number of feeding OD demands (feeding leg DF=0.9) ..............91
Figure 5.6: Network example for propagation analysis ......................................................................94
Figure 5.7: Flight legs scheduled in the same time (leg 1 and 3) are dependent on each other due to
overlapping OD demands on leg 2 ............................................................................................97
Figure 5.8: Network example for boundaries of effect propagation -- stochastic demand .................98
Figure 6.1: PFLOW algorithm Program Flow ....................................................................................104
Figure 6.2: Presentation of the basic simulation process ..................................................................106
Figure 6.3: Probability density function ..........................................................................................106
Figure 6.4: Presentation of the multiple-fare class multiple-booking period simulation process ......108
Figure 7.1: Spill estimates for Method 1 and Method 2 (M2) for different discount ratios d .............123
Figure 7.2: Method 2 Estimates of Spill Fares at various discount ratios d .......................................124
Figure 7.3: Differences in Spill Cost Estimates Relative to Method 2 .............................................125
Figure 7.4: Expected spill estimates for Data Example 2 ..................................................................127
Figure 7.5: Spill differences for Method 3 vs. Method 1 .................................................................127
Figure 7.6: Differences in spill cost estimates Method 3 minus Method 1 ..........................................128
Figure 7.7: Spill estimate curves for two individual days (spill_1, spill_2), their average (avg_spill), and
the spill estimate curve based of the combined demand density (CombinedSpill) .......................133
Figure 7.8: Differences in spill estimates (disaggregated-combined) (avg_spill-CombinedSpill) ......134
Figure 7.9: Spill differences for the 5 day time interval data (disaggregated-combined); standard
deviation equals the square root of means .................................................................................136
Figure 7.10: Spill differences for the 5 day time interval data (disaggregated-combined); standard deviation equals twice the square root of means ................................................................. 136
Figure 7.11: Network example 3L with a short and a medium haul leg feeding the long haul leg .......................... 147
Figure 7.12: Leg-independent spill estimates in the E-H network .............................................................. 149
Figure 7.13: Leg-independent spill fares in the E-H network ................................................................. 150
Figure 7.14: Leg-independent spill fares in the E-C network ................................................................. 150
Figure 7.15: Leg-independent spill costs in the E-H network .............................................................. 151
Figure 7.16: Leg-independent spill costs in the E-C network ................................................................. 151
Figure 7.17: Leg-independent and leg-dependent spill estimates for leg 3 as a function of capacity under the FC approach (Cap1=105 seats, Cap2=145 seats) ................................................................. 153
Figure 7.18: Leg-independent and leg-dependent spill estimates for leg 3 as a function of capacity under the VC system (Cap1 = 105 seats, Cap2 = 145 seats) ................................................................. 154
Figure 7.19: Comparison of leg-dependent spill estimates as a function of the employed YM system .......... 155
Figure 7.20: Comparison of leg-dependent spill cost estimates as a function of the employed YM system ................................................................. 155
Figure 7.21: Leg-dependent spill estimates in the E-H network .............................................................. 156
Figure 7.22: Leg-dependent spill estimates in E-C network ................................................................. 157
Figure 7.23: Difference (Leg-independent - Leg-dependent) between spill estimates in the E-H network ................................................................. 157
Figure 7.24: Difference (Leg-independent - Leg-dependent) between spill estimates in the E-C network .............................................................................. 158
Figure 7.25: Leg-dependent spill fares in the E-H network ................................................................. 159
Figure 7.26: Leg-dependent spill fares in the E-C network ................................................................. 159
Figure 7.27: Leg-dependent spill cost in the E-H network ................................................................. 160
Figure 7.28: Leg-dependent spill cost in the E-C network ................................................................. 160
Figure 7.29: Spill costs, operating costs and the total costs for a flight leg .................................................. 164
Figure 7.30: Example A -- Leg-independent two-leg example network ................................................................. 165
Figure 7.31: Spill cost, operating cost, and total cost estimates for leg 1 (Cost 1 referring to method VH1_C2 and Cost 2 referring to method VH3) ................................................................. 166
Figure 7.32: Spill cost, operating cost, and total cost estimates for leg 2 (Cost 1 referring to method VH1_C2 and Cost 2 referring to method VH3) ................................................................. 166
Figure 7.33: Example B -- Two-leg leg-dependent network example with local and connecting OD demands ................................................................................................................................. 169
Figure 7.34: Example C -- 4 leg hub network with leg-dependent flight legs .................................................. 171
Figure 8.1: Time-space network for a fleet type f ....................................................................................... 178
Figure 8.2: The schematic diagram of the constraint matrix in DYMFAFM .................................................. 180
Figure 8.3: Convex revenue curve with piece-wise linearization ............................................................... 181
Figure 8.4: The schematic diagram of the constraint matrix for column generation based DYMFAFMCG ................................................................................................................................. 185
Figure 8.5: Two-leg example for the simplified capacity assignment model .................................................. 192
Figure 8.6: Portraying the different directions of the objection functions in the two regions (Case 1 corresponds to the region above and Case 2 to the region below the "cut-line") .................................................. 196
Figure 8.7: Rotation of objective function and region of Case 2 parallel to the direction of the objective function of Case 1 .................................................................................................................. 197
Figure 8.8: Description of the objective function deviations at the three different fare structures ................................................................. 200
Figure 8.9: The stepwise description of the NSFA algorithm .................................................................................. 205
Figure 8.10: Two sub-networks (A and B) defined in a time-space network .................................................. 208
Figure 8.11: The replacement of sub-network A by m source and sink arcs .................................................. 209
Figure 8.12: Schematic Diagram of the Min-Cost Flow formulation in Step 4 .................................................. 213
List of Tables

Table 4.1: Leg-independent spill and spill cost estimates for each fleet type and flight leg ...............55
Table 4.2: Leg-dependent spill and spill costs on each leg at different fleeting combinations .............58
Table 4.3: Operating costs of the different fleet types...........................................................................60
Table 4.4: Total cost for fleeting combinations considering leg-independent spill cost estimates ......60
Table 4.5: Total cost for a fleeting combinations considering leg-dependent spill cost estimates .......61
Table 6.1: Run time performances for different leg-dependent, multiple fare class, multiple period network examples..............................................................................................................110
Table 7.1: A two fare class example showing that the total spill is a function of the fare ratios (i.e., optimal booking limits)..................................................................................................................120
Table 7.2: Data Example 1 and fares at various d discount ratios .............................................................122
Table 7.3: Differences of Spill Estimates (Method 2- Method 1)................................................................123
Table 7.4: Data Example 2 fare class means and fare values at different discount ratios .................126
Table 7.5: Data Example 3 ......................................................................................................................129
Table 7.6: Spill estimates and differences ...............................................................................................130
Table 7.7: Spill Cost estimates and differences ........................................................................................130
Table 7.8: Data for two days and the mean and standard deviation of the "combined" density ..........133
Table 7.9: Spill differences varying mean of day 2 (day 1 mean demand = 120 pax) .........................135
Table 7.10: Data example 4: five day data example with the corresponding "combined" density parameters ........................................................................................................................................135
Table 7.11: Data example 5: five day multiple fare class example with the corresponding combined density parameters ........................................................................................................................................137
Table 7.12: Spill estimates and differences for the vertical and horizontal aggregation example ......139
Table 7.13: Spill cost estimates for the vertical and horizontal aggregation example .........................140
Table 7.14: Differences in spill cost estimates for the vertical and horizontal aggregation example ...140
Table 7.15: ODF demand and fare values for the 5 OD and 5 fare classes -- 3L network example ....147
Table 7.16: Booking class demand, revenue, and revenue discount ratio parameters after aggregation in FC approach -- 3L network ........................................................................................................148
Table 7.17: Demand, revenue and discount ratio parameters for value classes on each leg in the VC approach-- 3L network ........................................................................................................................................148
Table 7.18: Comparison of leg-independent estimates in the VC and FC system ................................149
Table 7.19: Comparison of leg-independent and leg-dependent spill and spill cost estimates in the FC system -- 3L network ..............................................................................................................153
Table 7.20: Comparison of the leg-dependent and leg-independent spill estimates VC approach -- 3L network ........................................................................................................................................154
Table 7.21: Operating Cost, Spill Cost and Total Cost for leg 1 and 2 in Example A .......................166
Table 7.22: Total costs of assigning a fleet type and fleeting costs using two different approaches ....167
Table 7.23: Fleeting cost for the best fleeting using spill cost estimates of Method 1B and Method 3 and their differences in the 30 leg E-H network ........................................................................168
Table 7.24: ODF demand means and standard deviations in Example B ........................................170
Table 7.25: Leg-dependent and leg-independent spill cost estimates for leg 1 and leg 2 for two fleet types ........................................................................................................................................170
Table 7.26: Spill costs and their differences for different fleet type assignment to both flight legs ....170
Table 7.27: The four different fleeting combinations and associated fleeting cost for Example C ......171
Table 7.28: Spill cost and total costs (spill cost + operating cost) for different fleeting combinations in Example C assuming FC yield management approach ........................................................172
Table 7.29: Spill cost and total costs (spill cost + operating cost) for different fleeting combinations in Example C assuming VC yield management approach ........................................................173
1. Introduction

A passenger calls an airline reservation agent to book the cheapest seat on the 8 am flight from Boston to Seattle via the airline’s connecting hub. Curiously, on certain days the cheapest available fare is $300, whereas on other days it is substantially higher. This time the passenger is lucky: he is quoted a deeply discounted fare, and happily purchases the ticket. On the first leg of the journey, he is on a wide-body B767 aircraft with plenty of empty seats on the plane. Following a conversation with a fellow passenger, our passenger realizes that he paid approximately $500 less than his neighbor for a very similar ticket. When they arrive in the hub city, the two passengers, along with most other people arriving on other flights, rush through the airport to their connections. This time they find themselves on a much smaller A320 aircraft full of passengers. In amazement, the passenger gazes out the window and sees many different types of aircraft -- from the relatively tiny turbo-props to the behemoth wide-bodies -- pulling back from the gates, lining up to fly passengers to different destinations. Our friend presses his head against the window, reflects on the $500 he saved on his ticket, wonders why he needed to charge to a different type of aircraft, and while looking at the seemingly confusing mixture of different aircraft types that are scheduled leave to different destinations, he asks silently: Amidst this chaos, do airlines really know what they are doing here?

The above question is often asked by air travelers. What may appear to be a chaotic allocation of aircraft to the flight schedule and unreasonable fares are, theoretically at least, well-planned outcomes of yield management (or as it has more recently been called, Revenue Management or Perishable Asset Revenue Management) and fleet assignment decision systems, which are sophisticated decision support tools to increase airline profits. Although the results may not always seem rational to the passenger, yield management and fleet assignment decisions are critical components of airline efforts to maximize profits by maximizing expected passenger revenues and minimizing operating costs.

Since the deregulation of the airline industry in 1978, the observed trend is that load factors (the ratio between the traffic carried and the capacity supplied) have increased as a result of higher demand and consolidated passenger flows via hub networks. As a respond to the overcapacity that characterized the US airline industry during the early 90s, continued rationalization of route networks and reduction (or at least constancy) of offered capacities, coupled with increased demand for air travel, has prompted predictions that load factors will further increase. Higher load factors mean that on days when demand is high, an increasing number of passengers will be turned away (or spilled).

Airlines use yield management systems to maximize the expected revenues as constrained by total seat capacity on each flight leg. The basic idea behind yield management (YM) is to make seats that are expected to go unsold available for lower fare passengers who would otherwise not travel, while at the
same time ensuring that those lower-fare designated seats are not sold to passengers willing to pay more. Effective airline yield management practice comprises two distinct components [10]. *Differential pricing* is the practice of offering a variety of service products differentiated by service amenities and travel restrictions at different price levels. *Seat inventory control* is the practice of determining the number of seats available for each fare product, i.e., limiting the seats sold for early booking low-fare passengers and protecting seats for later-booking high-fare passengers. Specifically, it is the practice of determining the number of seats on a flight leg to be made available to a particular fare product [13].

Although yield management systems are widely employed throughout the airline industry, there is a wide disparity in levels of sophistication of use depending on the airline. Some airlines practice "conventional yield management," in which the number of seats made available for sale are managed by "fare type" on each flight. Other airlines practice "Origin-Destination (OD) seat inventory control yield management," in which seat availability is differentiated by passenger itinerary and revenue contribution to a network of connecting flights. The result of the yield management practice is that by limiting the number of seats sold to certain origin-destination (OD) fare class passengers, the actual OD passenger flows are affected and influenced in the network. Depending on actual demand characteristics, fare structures, and yield management approaches used (conventional or OD seat inventory control), OD passenger flows will be affected differently.

Before the airline decides how many seats to allocate to different passengers, it must first decide which aircraft type (fleet type) it should assign to each flight leg. This decision process is called *fleet assignment*. The *Fleet Assignment problem* is defined as the search for the optimal assignment of available fleet/aircraft types to a fixed flight schedule (fixed flights and departure times) such that expected profits are maximized. The inputs for fleet assignment decisions are the operating costs associated with flying the different fleet types on the particular flight legs and the expected revenues (or lost revenues/ *spill costs*) that can be realized (or encountered) by assigning a certain fleet type to a flight leg. Tradeoffs are made between a bigger aircraft that has higher operating costs but spills less demand (i.e., more accommodated passengers and lower spill costs) and therefore assures higher expected revenues (or less lost revenue potentials), versus a smaller aircraft that has lower operating costs but spills more passengers (higher spill cost) and therefore results in lower realized revenues (or higher lost revenue potential).

The correct estimation of spill, or passenger demand turned away, is integral to determining optimal aircraft capacities in the airline fleet assignment process. Despite having made advances in the solution methodology of the large-scale fleet assignment optimization problem, airlines have ignored the need to incorporate into spill estimation the effects of the yield management practices that have been widely implemented. The fleet assignment and yield management decision processes have been treated as
effectively independent decision processes; accordingly, the interrelationships between the two systems have been disregarded. In order to maximize expected revenues, yield management systems use flight leg capacity limits as inputs (determined by the actual fleet assignment) to allocate seat availability for the different OD passenger flows in the network. This influences the spill from the flights operated in the network. Fleet assignment decisions, in addition to factoring in operating costs, are directly affected by the expected number of passengers turned away, or spill, associated with a fleet assignment. Consequently, the input to one system is affected by the output decisions of the other system. Disregarding this feedback effect, therefore, may lead to suboptimal decisions.

Network effects (or leg-dependence) have also been left out of spill estimations and fleet assignment decisions. In this dissertation we will refer to the term network effects or leg-dependence to describe the interactions of the OD passenger flows as a function of the prevailing capacity limits in the network. Specifically, leg-dependence refers to the phenomenon that spill on a flight leg can be influenced not only by the seat capacity and demand conditions associated with the given flight leg, but also by the capacity and demand conditions on the other legs of the airline network. Using a simple example, if a multiple leg itinerary OD passenger is spilled on the first flight leg of his or her path, then this passenger cannot appear on the other flight legs of the path. Therefore, the achievable traffic (or the potential demand) on flight legs is a function of the seat (and aircraft) capacity limits of all flight legs of the path. This phenomenon affects both spill estimates and fleet assignment solutions.

In this dissertation we attempt to fill the gap in the spill estimation and fleet assignment research and to identify demand and network conditions when network effects and the effects of the yield management practices significantly influence spill estimation. The research also attempts to demonstrate the importance of these effects and how incorporating these effects in the decisions process can lead to different and more profitable fleet assignment solutions. As we demonstrate throughout the dissertation, the influence of network and yield management effects on spill estimates can be very substantial. Based on typical network scenarios, it is demonstrated that incorporating these effects into the spill estimates can result in different fleet assignment solutions, which in turn can provide substantial expected cost savings in longer run operations. Accordingly, the benefits of pursuing accurate estimation methods for spill and incorporating them into the fleet assignment decision are quite clear.

1.1 Goal and Structure of the dissertation

The goal of the dissertation is to address many of the fundamental issues related to incorporating network and yield management effects into the fleet assignment decision process. The thesis tries to identify the important demand and network characteristics and attributes under which network and yield management effects influence spill estimates and to evaluate the significance of these effects in typical stochastic airline situations. Besides identifying the significance of the above effects, it is important to develop approaches
that capture the effects in the spill estimates. For this reason, various methods and approaches are outlined in the dissertation that can incorporate the network and yield management effects in the spill estimation. Finally, the dissertation attempts to analyze the modeling complications and problems that arise when we try to incorporate network effects into fleet assignment mathematical programming formulations, and to propose approaches that heuristically overcome such complications.

This thesis limits its focus to two major effects -- network and yield management effects -- which have been mostly ignored in previous works. Consequently, various other important effects and practices, such as overbooking practices, path selection for recaptured (reaccommodated) spilled passengers on other flights or itineraries, etc., were not modeled. The two basic assumptions of the dissertation that demand is specified and available for each ODF (origin-destination fare class) and departure time with a fixed itinerary, and once a passenger is spilled, he or she is lost to the airline. It was believed that limiting the analysis to network and yield management effects creates a rich enough topic to explore and research for this dissertation. Therefore, the goal of the dissertation is to fill in a gap in this area, to determine some important implications, and at the same time to lay down some of the basics for further research and development in this area. As one of the initial efforts to organize, analyze, and address these two effects, this research opens new directions and dimensions in the analysis while taking an important step towards a deeper understanding of the important issues that relates to the fleet assignment decision process.

The dissertation is divided into nine chapters. Chapter 2 serves as a formal introduction to the fleet assignment problem and the current yield management practices. The chapter also addresses the underlying relationships and the interactions of the two practices. The phenomenon of network effects in airline networks is also defined. Finally, the chapter summarizes the basic definitions and assumptions used throughout the dissertation.

Chapter 3 reviews the most important mathematical models, formulations, and current practices in the areas of yield management, fleet assignment and spill estimation. The chapter also addresses the very limited literature that examines only deterministic network effects in airline networks. The goal of the chapter is to review the most important models and algorithms published in these fields and to serve as a reference to many of the models comparatively used throughout the analyses.

Because one of the objectives of this dissertation is to address the issue of network effects, Chapter 4 begins with an introduction of fundamentals of the network (leg-dependence) effects. First addressing the less complicated deterministic demand cases, the chapter shows the implications of network effects on the fleet assignment solution. The second part of the chapter addresses network effects assuming that demand is stochastic (which is a more appropriate representation of airline demand). Formulae are developed to calculate achievable traffic (or "potential demand") densities, and analytic formulae are derived to
calculate spill and spill costs for a small stochastic network example. The rigorous and complex results imply that for larger examples, analytic methods may be too cumbersome.

Chapter 5 is devoted to the identification of the likely factors and important demand and network characteristics that influence the existence and magnitude of leg-dependence effects in airline networks. Among other factors, the percentage of connecting passengers, the amount of spill, the booking patterns of OD demands, network connectivity issues (dispersions versus concentration of leg-dependence effects), the direction and boundaries of leg-dependence effect propagation in stochastic networks are addressed.

Chapter 6 starts with a review of two previously developed algorithms to calculate leg-dependent spill estimates in networks. After addressing the limitations of using these algorithms in typical airline networks, a new Monte Carlo simulation-based approach that overcomes the limitations is outlined. The approach, in addition to capturing network effects in the spill and spill cost estimates, can take into consideration the effects of the yield management decisions. The outlined simulation approach is one of the main tools of analysis in the following chapters and a main building block in the algorithm presented in Chapter 8.

The purpose of Chapter 7 is to demonstrate the nature and extent of yield management impacts on airline spill estimation. Horizontal and vertical aggregation biases are defined and the limitations of the state-of-the-practice methods as a consequence of these biases are shown. A new recursive approach is presented to allow more accurate spill and spill cost estimation for a flight leg. The second part of the chapter addresses the effects of the different yield management approaches with respect to their influence on the OD passenger flows in a network. It is shown that different yield management approaches affect spill costs differently, which in turn have an influence on the fleet assignment solutions. Numerical examples are present throughout the chapter to illustrate the extent to which the outcomes of the different estimation approaches differ. The typical network examples at the end of the chapter demonstrate that the differences in the spill and spill cost estimates are large enough to affect the optimal fleet assignment solutions. The examples also imply that ignoring network and leg-dependence effects can result in sub-optimal fleet assignment, which in turn can result in substantial additional costs or unrealized profit potentials.

Based on the findings and analysis of the previous chapters, Chapter 8 tries to analyze the modeling complications and problems that arise when we attempt to incorporate network effects in fleet assignment mathematical programming formulations. The chapter outlines some fleet assignment formulations, which under different model assumptions calculate leg-dependent fleet assignment solutions. The models in the first section assume that the airline is able to control passenger flows on the OD level. Combined Passenger Flow and Fleet Assignment models, for both the deterministic and stochastic demand case, are proposed. To overcome the expected large size of the problems, column generation-based formulations
are presented. The second part of the chapter addresses the more realistic case when the airline is assumed not to be able to control passenger flows on OD level. We show that, unfortunately, when network effects are considered in the fleet assignment formulations under this assumption, then the feasible set of the problem is not convex. This implies that under the most common airline practices today, most of the mathematical programming algorithms cannot be directly used to solve the $l_2$-based fleet assignment problem and take yield management into account. At the end of the chapter, an iterative Nested Algorithm with Embedded Simulation is proposed that overcomes the problem of non-convexity and it leverages on the state-of-the-art existing large scale fleet assignment solution methodologies.

Chapter 9 concludes this dissertation, summarizing the research findings and the contributions. Finally, future research directions building on the research presented in the dissertation are outlined.
2. The Airline Capacity Planning Process

This chapter presents the two fundamental airline capacity planning processes: fleet assignment and yield management. Fleet assignment is concerned with assigning aircraft types to scheduled flights and is a mid-term planning process. Using the fleet assignment as an input, yield management fine tunes the capacity allocation process by assigning seat availability for different fare class demands. The chapter shows the interrelationship of these processes and suggests that yield management optimization effects should be considered in fleet assignment decisions. The chapter also introduces the issue of leg-dependence and indicates that traditional fleet assignment approaches assume leg-independence between flight legs. This issue raises serious concerns about the accuracy of traditional fleet assignment models.

2.1 The Traditional Fleet Assignment Approach (Leg-Independence)

In the mid- to short-term, one major decision that an airline must make is deciding which aircraft to operate on each flight and how many seats to allocate to the different fare classes. The latter decision process is the function of yield management systems. The former is part of the fleet assignment models. In this dissertation we will refer to the following as the fleet assignment problem:

- **The Fleet Assignment** problem is to find the optimal assignment of available fleet/aircraft types to a fixed flight schedule (fixed flights and departure times) such that expected profits are maximized.\(^1\)

As airlines became more interested in the short term aspect of schedule optimization, which has also been more relevant for the new market environment, various models have been developed to solve the fleet assignment problem.

The fleet assignment problem is traditionally formulated and solved as a mathematical linear integer program -- usually as a multi-commodity flow formulation \([1],[24]\). The models are solved either by maximizing the expected profits or minimizing the total costs over all fleet/aircraft types and flight legs, subject to various aircraft operating and flow constraints: The flow balance constraints assure that a fleet type is assigned to a flight leg at a station only if it is actually present at that airport; the availability of

---

\(^1\) Earlier developed frequency planning models (then called as fleet assignment or aircraft assignment models, but we will refer to them as frequency planning models) tried to find the optimal set of routes to be served, the frequencies to be offered, and the aircraft types used for each flight leg \([32],[33]\). Those models by definition assumed that airlines can freely adjust their schedule to the model recommendations: That is, they assumed that airlines can freely choose routes and change the flight frequencies. It was also assumed that demand is a function of the frequencies offered. In this dissertation we refer to the fleet assignment problem as defined above, to the problem of assigning fleet types to a fixed schedule (with frequencies and departure times given). It is also assumed that demand is given and not a function of the fleeting decisions.
equipment constraints assure that no more aircraft are assigned at any time than the available maximum number of that type of aircraft; the cover constraints assure that each flight leg is covered by at least one fleet type; and the integrality constraints assure that exactly one fleet type is assigned to a flight leg.

Both the revenue-maximization and cost-minimization formulations produce the same optimal fleet assignment solution. The solution to the linear program gives the optimal fleet type assignment to each flight leg of the network. (We will often refer to the fleet type assignments as the fleeting during the dissertation). Because the solution always maintains a balance of equipment, the aircraft cycle rotations for each aircraft tail-number can be easily constructed.

The two types of objective functions can be generalized by the following:

$$\max \sum_{f \in \text{Fleet}} \sum_{l \in \text{Leg}} (\text{Rev}_{f,l} - \text{Op \_ Cost}_{f,l}) X_{f,l} \quad \text{or} \quad \min \sum_{f \in \text{Fleet}} \sum_{l \in \text{Leg}} \text{cost}_{f,l} X_{f,l}$$

s.t.

Flow balance,

Equipment availability,

$$X_{f,l} \in \{0,1\} \quad \forall l \in \text{Leg}, f \in \text{Fleet}$$

where $\text{Rev}_{f,l}$ is the expected revenue, $\text{Op \_ Cost}_{f,l}$ is the operating cost, and $\text{cost}_{f,l}$ is the total cost associated with assigning fleet type $f$ to leg $l$, and $X_{f,l}$ is the binary variable taking value 1 if fleet type $f$ is assigned to flight leg $l$, and 0 otherwise. The objective coefficient $\text{Rev}_{f,l}$ is estimated as the product of the expected load on leg $l$ assuming fleet type $f$ is assigned (see equation 3.25 in Section 3.4) and the estimated average fare of the accommodated passengers. The $\text{Op \_ Cost}_{f,l}$ is the operating cost of an $f$ type aircraft on leg $l$.

In the second objective function the $\text{cost}_{f,l}$ objective coefficient incorporates the operating costs and the spill costs. The spill cost is the opportunity cost, or estimated lost revenues, due to the insufficient capacities (seat availability). For each fleet type $f$ and flight leg $l$, a separate $\text{cost}_{f,l}$ objective coefficient is estimated.

Spill is the expected number of passengers that cannot be accommodated due to insufficient seat capacity. Given the forecasted demand probability function and the capacity for a single flight leg, spill can be calculated as:

$$\text{Spill} = \int_{i \in \text{Cap}} (i - \text{Cap}) f(i) \, di,$$  

(2.2)

\footnote{The actual definition of the decision variable $X$ and the used objective coefficients may vary.}
where \( f(i) \) represents the probability density function (pdf) of passenger demand on the leg, and \( \text{Cap} \) represents seat capacity. The spill cost can be calculated as the product of spill and an estimated fare of the spilled passengers, the \text{spill fare}.

\[
\text{Spill Cost} = \text{Spill} \times \text{Spill Fare}
\]

From the above linear programming formulations, equation (2.1), we can observe that for each decision variable \( X_{fl} \), a constant objective coefficient is estimated. For example, in the cost minimization formulation, one single objective coefficient \( cost_{fl} \) was associated with each decision variable. This implies that on a flight leg \( l \) the cost associated with an assignment of fleet type \( f \), depends only on the leg demand and the capacity of \( f \). Alternatively, we can say that if fleet type \( f \) is assigned to flight leg \( l \), the spill cost associated with this decision is constant and assumed to be equal to \( cost_{fl} \). This value is independent from the other decision variables or from the capacity assignments on other flight legs. This shows the following characteristics about the traditional fleet assignment approach:

1. It is a \text{leg-based} formulation, and
2. It assumes \text{leg-independence} in the demand representation.

The second characteristic highlights the fact that in these formulations it is assumed that a capacity limit on a flight leg \( l \) would affect the demand and thus the revenue or spill cost estimates of that leg only, and would have no effect on the objective coefficients of other legs. This assumption is very restrictive; it assumes no interaction and no network effects in the passenger flows among the flight legs. It also inherently assumes that if fleet type \( f \) is assigned to flight leg \( l \), then regardless of what type of aircraft and capacities are assigned to the other legs, (i.e. independent of the values of \( X_{fl} \) on legs other than \( l \)), the demand, the spill, and the spill cost are constant.

In this dissertation we will show that since the leg-independent assumption is an incorrect representation of the actual passenger flows in networks, the solution of the leg-independent fleet assignment formulations raise serious concerns about “optimality” in terms of airline profit maximization.

### 2.2 Network Effects (Leg-dependence)

We will use the term \text{network effects} or alternatively the term \text{leg-dependence} to describe the interactions of the passenger OD flows as a function of the prevailing capacity limits in the network. Particularly, network effects describe the phenomenon that \text{demand-related} load and spill estimates of each flight strongly depend on the \text{supply-related} (seat) capacity assignment decisions on the other flight legs as well. Thus,
• *Leg-dependence* refers to the phenomenon that a leg-based demand estimate can be influenced not only by the capacity and demand conditions of the actual leg, but also by the capacity and demand conditions on the other legs of an airline network.

We will see that connecting (multiple leg) OD passenger flows tie flight legs together and give rise to network effects. Figure 2.1 shows this situation. It is obvious that if a capacity limit of flight leg 1 constrains the connecting OD passenger flow of the network, i.e., the AC demand is spilled on leg 1, then the spilled demand cannot appear on leg 2 either. This means that the achievable traffic (or "potential demand") conditions are modified on leg 2, and are a function of the capacity limits on leg 1. This also implies that the load and spill estimates on leg 2 become a function of both leg capacity limits.

![Diagram of flight legs](image)

*Figure 2.1: Simple example for portraying network (leg-dependence) effects*

This strong correlation between the flight legs imply that leg-dependence should be considered in the calculations of load or spill estimates. We will address the traditional spill and spill cost estimation, and the traditional fleet assignment approach in light of the above, and will show that ignoring leg-dependence can lead to sub-optimal fleeting solutions.

According to our knowledge, only one published paper, by Phillips et. al. [29], recognizes the issue of network effects in airline networks. The paper recognizes that leg-dependent effects (or "cross-leg" effects as they refer to it) may influence the estimation of the "consistent" OD passenger flows. To a limited extent the paper mentions that inconsistent passenger flows may affect fleet assignment decisions as well; however, the fleet assignment problem in this light is not fully addressed. The paper also mentions implications on yield management decisions; however, the issue that yield management decisions may affect the calculation of the passenger flows of the network is not addressed either. In addition, the paper considers only deterministic demand cases, despite that typically airline demand is stochastic. Uncertainty (as the paper also admits and as we will see throughout the dissertation) complicates the calculation of the leg-dependent passenger flows to a large extent, and the basic ideas of the deterministic approach cannot be directly extended to the stochastic case.
In this dissertation we will enhance the basic observations and conclusions presented in [29], and will model the problem in the more complex stochastic environment. We will analyze leg-dependence which can affect the calculation of passenger flows and consequently will affect the fleet assignment decision process. Additionally, the stochastic effects of yield management optimization will be also included in the analysis.

2.3 Supply of Fare Products and Yield Management Systems

Once the flight schedule and the fleet types that fly the particular flight legs are decided, i.e., the capacity decisions has been made, the airline usually decides what fare products it offers and how many seats should be available at each fare. Based on the "willingness to pay" principle, the airline first identifies distinct segments of demand. These segments identify the separate groups of potential travelers who willing to pay higher fares for services. Allocating the passengers to different segments according to the passengers’ willingness to pay and charging different prices for the same services is called price discrimination. On the other hand, airlines need to differentiate between fare products. The concept of product differentiation has been applied to the different service options that are associated with the fare products. These include service amenities and various restrictions. Hence for an airline firm, differential pricing practices include the simultaneous use of price discrimination and product differentiation principles.

Airlines indeed apply differential pricing practices, and in each origin-destination (OD) market offer a number of different fare products for travel. A published OD fare product (ODF) in a market embodies the price for the trip for a given level of service amenities and restrictions. There may be a specific path or itinerary associated with an ODF, which includes the particular flight legs and departure times. The different service amenity levels and restrictions are part of the product differentiation principle. They are designed in such a way that the passengers who are willing to pay a higher fare should be prevented from buying the lower fare, more restricted fare products.

The aircraft is usually divided physically into different cabins, offering different levels of on-board services. A typical international configuration include the First Class, Business Class, and Economy Class cabins. Although only three different type of on-board services are offered, airlines offer more than three fare products. Particularly, in the Economy Class in a given OD market, airlines offer some 7-9 (or sometimes even more) fare products. Although the level of on-board services are the same for these fare products, the overall service level may differ. Differentiation is achieved by a complex structure of restrictions. The requirements of advance purchase, Saturday night stay over, round-trip purchase, non-refundability, etc., all serve as product differentiating restrictions. Thus, instead of using positive service differentiation (like higher service amenities in the higher priced cabins), airlines use service restrictions in the Coach (or Economy) cabin as a tool for (negative) product differentiation.
In practice, airlines group the published fare products into *fare classes*. A set of service amenities and restrictions and an average fare of the included fare products are associated with each fare class. The fare class structure of the airline fare products is what embodies airline's differential pricing strategies.

2.3.1 Conventional Yield Management (YM) Practice
The basic idea behind *yield management* (YM) (or, as it has more recently been called, *Revenue Management* or *Perishable Asset Revenue Management*) is to make available for lower revenue passengers only those seats that are expected to go unsold to higher-revenue passengers, while at the same time to ensure that lower-revenue fare products are not purchased by those passengers who are willing to pay more. Thus, airline yield management practice consist of two distinct components [9],[10].

*Differential pricing*, which was described above, and *seat inventory control*. Seat inventory control is the practice of limiting the number of available seats for early booking low-fare passengers while protecting seats for later-booking high-fare passengers. In particular, it is the practice of determining the number of seats on a flight leg to be made available to a particular fare product [13]. Yield management practice evolved from defining and controlling only two fare classes: "full fare" and "economy" fares of the economy (coach) cabin. The process of YM has by now progressed substantially beyond two fare class allocation of seats, and has recognized the potential revenue benefits of increased segmentation and control.

Traditionally, airlines classified the large number of ODFs into a limited number of fare classes (or booking classes) on each flight leg. The classification was based on yield (or fare/mile), so that the highest-valued booking class on the flight leg was associated with the Y full coach fare, and the lowest valued booking class with the most deeply discounted fare products\(^3\). In this yield-based booking class approach, yield management systems need to forecast the expected demand at departure on each leg for each booking class, and use the time-varying demand forecasts as inputs for the seat allocation algorithms [27]. The seat allocation algorithms determine the *protection levels or booking limits* (BL), that need to be applied to the booking classes for each flight. Many yield management architectures use booking limits for *nested* booking class structures, in which booking limits of higher classes include all seats made available for lower classes. This nested structure guarantees that a higher-fare request is always accepted as long as there remains an unsold seat for a lower-fare class passenger. In a nested booking class structure, the booking limit for a higher class will always be greater than or equal to the booking limit of a lower class, i.e.,

\[ BL_c \geq BL_{c+1} \]

where \(BL_c\) is the booking limit for class \(c\), and the highest-valued class is \(I\).

\(^3\) YM approaches with different categorization will be presented and discusses in Chapters 3 and 7.
Yield management algorithms determine the booking limits for each flight leg departure and communicate them to the computer reservation system (CRS). (Much work has been published on the topic of optimization algorithms for airline yield management. A detailed description of the developed yield management optimization algorithms is presented in Chapter 3.) For a passenger wishing to purchase a ticket for booking class $c$ on an itinerary $i$, on each flight leg $l$ of the itinerary, the number of seats available to class $c$ has to be greater than zero. The booking limit in booking class $c$ for itinerary $i$ is given by:

$$BL_{lc} = \min[BL_{cl}, \forall l \in i].$$ (2.6)

To summarize, conventional YM practice involves the classification of the fare products into booking classes according to their fare type or yield, with the goal of protecting the availability of seats for the higher valued fare class passengers. Forecasting and booking limit optimization are usually performed at the booking class level, for each flight leg independently.

2.3.2 The OD seat inventory control problem

The conventional leg-based YM approach has some obvious shortcomings if the airline’s objective is to maximize total revenues over the network. Maximizing revenues over flight legs individually does not guarantee that total revenues are maximized [43].

One of the most important shortcomings of the conventional fare, or yield-based, approach is its inability to recognize that it may be in the airline’s interest to accept a lower-yield connecting passenger who may travel on multiple flight legs producing higher revenue rather than the high-yield (or, even “full fare”.) local passenger who travels only a single leg. It is possible that the total network revenue contribution is higher for the connecting discount-class passenger whose total fare is higher (even though his yield is lower), than for the single-leg full-fare passenger (even though his yield is higher). The above situation may happen when a bottle-neck flight leg is presented in the network and different ODF passenger itineraries compete for the scarce seats on the leg.

Under certain network conditions, on the other hand, the optimal solution would promote the acceptance of the local passengers along a multiple leg path over the connecting passenger traveling the same path. For example, this is the case when instead of a bottle-neck leg, seats are scarce along all legs of a path. In this case, the probability that the seats will be sold to higher yield local passengers may be relatively high. Further, it is typical that in a given fare class the sum of the local fares along the path is greater than the fare associated with the connecting fare along the path. Therefore, in this case the airline’s interest is to sell the seats to local passengers, if the expected network revenues from the local passengers are greater than the single connecting fare.
In short, revenue maximization over a network of connecting flight legs requires a combination of two capabilities [13]. First, there must exist a capability to give preference in terms of seat availability to high revenue connecting (long-haul) passenger, regardless of the ODF’s yield. Second, in certain situation the system has to prevent these same high revenue long-haul connecting passengers from taking seats away from two (or more) high-yield local passengers, each of whom occupy seats only on one leg but whose combined summed revenues can be higher than that of the connecting passenger. (For a detailed introduction of some of the heuristic approaches to the OD seat inventory problem refer to Chapter 3).

2.4 Relationship Between Fleet Assignment and Yield Management

Fleet assignment and yield management decisions influence each other significantly. Figure 2.2 shows the underlying relationship and the interactions of the two processes. The figure shows that the inputs to the fleet assignment model are the aircraft operating costs, and the expected load, revenue, spill, and spill cost estimates on each leg, which are a function of the expected achievable traffic ("potential demand" conditions) in the network. The output of the fleet assignment process are the capacity decisions, which are also inputs to the yield management algorithms. The booking limits set by the yield management algorithms, which are also a function of the ODF demands and the underlying flight network, influence the ODF traffic flows in the network. Finally, the influenced ODF traffic flows determine the achievable traffic conditions, and thus the expected revenues or spill costs on each leg of the network. Thus, the yield management system feeds back to the fleet assignment decisions via the underlying flight network.

Figure 2.2: Recursive relationship of Fleet Assignment and Yield Management decisions via the airline passenger network
The fleet assignment optimization process is one step in the sequence of the airline schedule development process. Its objective is to assign the available fleet types to the network in such a way that the expected profit is maximized, (or alternatively total operating and spill costs are minimized,) subject to the various network operational constraints. The fleet assignment process can be characterized as a strategic, medium-term planning optimization. Although fleet assignment decisions can be altered later [15], decisions about fleeting are made up to few months before operations begin. The lead time may be required for reasons such as providing enough time for crew assignment and crew bidding procedure, allowing for rotation and maintenance planning, and entering the capacity information in the computer reservation systems for early reservations. The output of the fleet assignment process has effects on the actual revenue and traffic that can be realized and, at the same time, on the yield management optimization process.

Given the flight leg capacities set by the fleet assignment model, a yield management system tries to control the availability of seats to passengers on different ODF itineraries of the network so that the expected revenues are maximized. Over the booking period, it dynamically updates the booking limits set for each fare class so that the fixed leg-based seat capacities can be allocated optimally. Since yield management incorporates the changes in demands and the stochastic dynamics of the booking process, and because it tries to readjust or compensate for the inaccurate longer term flight leg capacity decisions, it can be characterized as a tactical dynamic optimization process operating with stochastic inputs.

The fleet assignment process attempts to allocate capacities given some forecast for expected passenger demand flows on each leg, while the yield management system tries to allocate the passenger OD demand flows, given the capacities. Since the output of one system is the input for the other, the two optimization processes will influence each other. The dynamics of this interaction depend on the timing of these decision processes. Passengers start booking seats after the fleet assignment decisions are set. Therefore, yield management algorithms start working and influencing the passenger flows later. In spite of this, the yield management system has an effect on the fleet assignment process.

Here we define two terminology that will be used throughout the dissertation. The unconstrained leg demand is the sum of all unconstrained ODF demands traversing the flight leg. The achievable (leg) traffic on leg $l$ represents the maximum “potential ODF demand” that can be expected on leg $l$ taking into consideration the prevailing capacity limits on legs other than $l$. Assuming that connecting ODF itineraries traversing leg $l$ can be constrained or spilled on other legs, the achievable traffic on a leg $l$ is often less than the unconstrained leg demand even if no capacity limit constraints the flow on leg $l$. The achievable traffic is influenced by the yield management systems and the interactions of the different ODF demands in competing for the available seat capacities along the flight legs. We will refer to these complex interactions of the ODF demands in the network as network or leg-dependence effects. The network rhombus in Figure 2.2 symbolically represents the modeling of the network effects.
Figure 2.3 shows the assumed relationship of fleet assignment and yield management processes in traditional fleet assignment models. The passenger network in the traditional approach can be interpreted as being "split". It is modeled as if the passenger traffic flow network that the yield management systems affect and the passenger traffic flow network that is modeled in the traditional fleet assignment models were independent from each other in a decision cycle. According to our knowledge, none of the developed fleet assignment models incorporate network effects, which are captured in the achievable traffic estimates. The basis of the load and spill estimates are unconstrained leg demand forecasts. Consequently, as we will show in the literature review, fleet assignment models assume "leg-independence" during their optimization. Particularly, it is assumed the revenue and spill cost estimates on a leg are only a function of the capacity limit assigned to the particular leg. The effects of the yield management systems in most of the developed and used models are not captured or considered. (Only one non-published internal report [40] incorporates the effects of the yield management systems by a rule-of-thumb approach.). The lack of an assumed feed-back effect from the yield management systems in a decision cycle is represented in Figure 2.3 by the "split" passenger network rhombus.

Figure 2.3: Inherently assumed "split" relationship with no (or lagged) feed-back effect between the Fleet Assignment and Yield Management decisions in a decision cycle in the traditional fleet assignment approaches

Next, we will address one of the most important characteristics of the economics of airline and many other transportation networks, one which significantly complicates the incorporation of the passenger flows and effects of the yield management optimization to the fleet assignment decisions.
2.5 The Fundamental Dichotomy of Supply and Demand

Supply in air transportation is defined and provided on a leg basis: On the other hand, demand is always defined in terms of an OD market. Because of the nature of the airline networks, especially hub networks, OD itineraries may include more than one flight leg in their paths. Therefore, flight leg based supply decisions do not directly overlap with the OD based demand. This fundamental dichotomy of supply and demand is what makes transportation so different from other industries and makes the analysis and decision process in airline networks complex. It is important that this fundamental characteristic, with its consequences, are incorporated in the airline planning models.

Thinking about the relationship of the supply and demand, we can imagine that two independent networks are defined on the premise of the schedule of flight services: the network of aircraft flows and the network of passenger flows (see Figure 2.4). The "commodities" in these two networks flow independently, but to a certain degree and at a certain time, on each leg they are tied to each other directly. There is a fundamental difference between the two networks. The supply-based aircraft flow network directly corresponds to the underlying schedule of flight services, while the demand-based network of passenger flows corresponds to the schedule only indirectly. The traffic (part of demand) that can be observed on a flight leg is a composition of the OD based passenger flows along their itineraries.

![Diagram of network of ODF passenger flows, underlying flight schedule, and network of aircraft flows](image)

Figure 2.4: The underlying flight network and the two additional network layers

Fleet assignment decisions must be made on a flight-level basis, that is, aircraft types need to be assigned to flight legs. In order to make the assignment feasible with respect to the number of available aircraft and the flow of equipment, the decision process involves consideration of various bounding network conditions: In the time-space network, the flow of aircraft and fleet types must satisfy the various flow balance, cover, and size constraints, etc. (Refer to Section 3.3 for a detailed discussion of the fleet assignment formulation.) The proposed fleet assignment models exploit the structure of the aircraft flow
network and often formulate the problem as a multicommodity integer programming network flow problem [24].

Demand arises on an OD pair; that is, passengers want to travel from their origin to their destination within a particular time window [26]. Often in airline networks, no direct non-stop flight is available in the OD market. Therefore, passengers may to choose a path with multiple flight legs that will connect their origin to their destination. Choosing the selected path from the available set of possible paths is a complex stochastic decision process. The passenger's decision is an ODF: a particular itinerary with a given path between an origin and destination in a particular service option. The demand for all ODFs constitutes another network, the network of passenger flows, where the various ODF passenger flows compete for the available capacities on the network of flight legs.

Given the nature of the possible passenger itineraries, ODF demand can be local or connecting (multiple leg). Local demand refers to the case when the demand's path includes only one flight leg, that is, the demand is accommodated on a non-stop flight. Connecting demand refers to the case when the itinerary path includes more than one flight leg. (Although it is possible that a multiple leg demand does not need to connect to an other flight, in this dissertation we will refer to multiple leg demands as connecting demand.)

In special cases, when all traffic is local demand, supply and demand decisions are defined at the same leg level and congruent. In this case, both the supply based aircraft flow network and the OD demand based passenger flow network directly correspond to the underlying schedule of flight services. In this case, the fundamental dichotomy of supply and demand is not important with respect to the fleet assignment problem.

However, in a typical hub-and-spoke network, it is the fundamental nature of operations that certain OD demand needs to be or chooses to be accommodated on multiple-leg paths and various OD traffic, local and connecting, share and compete for available leg capacities. Consequently, in a hub-and-spoke network, supply decisions remain to be defined on leg levels, but demand decisions are defined on OD levels. In this case, the networks of aircraft flows and passenger flows are not congruent.

2.6 Basic Assumptions and Terminology

In this section we will summarize the basic assumptions and terminology that we will use throughout this dissertation.

The first basic assumption is that demand is specified and given for each ODF and departure time with a fixed itinerary. It is assumed that a passenger with a confirmed booking will take the flight. Therefore, no-shows and cancellations were not considered and modeled.
The second basic assumption is that once a passenger is spilled, he or she is lost to the airline. Therefore, we do not consider the recapture of a spilled the passenger on any other flights or other itinerary paths.

We limit our focus to two major effects -- network and yield management effects -- which were mostly ignored in previous works. Consequently, various other important problems and practices, such as overbooking practices, path selection for recaptured (reaccommodated) spilled passengers on other flights or itineraries, upgrades, were not modeled.

About the OD demand we will assume the following:

- **Demand** is specified for each ODF (origin-destination fare class) and departure time with a fixed itinerary of flight legs in a deterministic or stochastic form. We do not consider multiple path choices for an ODF demand. (Different itineraries between an OD market for a fare class, are specified as independent ODF demands.) Local demand or local ODF refers to the case when the ODF's path includes only one flight leg, that is, the demand is accommodated on a non-stop flight. Connecting (or multiple leg) demand or ODF refers to the case when the demand's itinerary path includes more than one flight leg. Demand can be given in a one-period or multiple-period form. In the latter the overall booking period is subdivided into multiple, mutually exclusive, booking intervals, and demand is given for each interval in incremental form. This representation captures the associated *booking pattern* (or booking curve) of the ODF. The more such intervals are defined, the better the representation of the booking pattern can be. In the one-period representation, demand is given for the whole booking period, thus no specific information is given about the actual ODF booking pattern.

- **Independence of demand**: We will always assume that ODF demands are independent from each other. We assume zero correlation among the demands for different OD pairs, and also among the different fare class demands of an OD pair.

- **Deterministic demand**: For the deterministic demand we assume that demand books over the booking period with a constant booking rate. For example, if \( D \) is the total demand and \( T \) represents the length of the booking period, then the booking rate is constant over \( T \) and equals \( D/T \).

- **Stochastic demand**: Stochastic demand is given by a Gaussian (Normal) probability function with its parameters: mean and variance. A number of earlier studies have shown that airline demand can be correctly represented by a Normal distribution [10], [27], [36], [38]. In this dissertation, when stochastic demand is considered for a booking period or a booking interval, then stochasticity refers to the total demand for that booking period or interval. Therefore, we assume that the cumulative static total demand for the booking period or interval is Gaussian. However, we assume that the booking rate over the booking period or interval will be *uniform*. Thus if random variable \( D \) represents the total demand for the ODF and \( T \) is the length of the booking period or the booking interval, then the booking rate is assumed to be constant and equals \( D/T \).
• **Unconstrained ODF demand**: The given ODF demand with a fixed itinerary (or path), not affected by any capacity constraints.

• **Censored or constrained ODF traffic**: The ODF demand is constrained or censored by a capacity limit (aircraft capacity or booking limit). Thus, due to the capacity limit on flight legs along the ODF’s path, only a part of the demand can be accommodated.

• **Unconstrained leg demand (or leg demand)**: The sum of all unconstrained ODF demands traversing the flight leg.

• **Achievable ODF Traffic**: The maximum ODF demand that can be accommodated on a flight leg. This is often less than the unconstrained ODF demand, taking into account the constraining effects of other flight legs of the ODF’s path. For example, if the first leg of a two-leg itinerary ODF censors the demand, then the achievable ODF traffic on the second leg (i.e., the maximum ODF traffic that can be accommodated on the second leg if there are no capacity limit set for the ODF on that leg), is less than the unconstrained demand. If the ODF is not censored on any flight legs, then the achievable ODF traffic is equal to the unconstrained ODF traffic or demand. We will refer to the new constrained “demand” condition on the second leg as “achievable ODF traffic”.

• **Achievable (Leg) Traffic on leg l**: The sum of all achievable ODF traffics that traverse leg l, assuming that the capacity limit set on leg l does not censor any demands. Since local ODF’s are not censored, the achievable leg traffic is the sum of the local unconstrained ODF demands and the connecting achievable ODF traffics. Taking into consideration the prevailing capacity limits on legs other than l, the achievable (leg) traffic represents the maximum traffic (or “demand”) that can be expected on leg l. Thus, the achievable traffic represents the “potential demand” for leg l.

### 2.7 Summary

This chapter described the two capacity planning processes. The traditional fleet assignment formulations and yield management practices were addressed. The issue of network effects, or leg-dependence, was defined. Given that traditional fleet assignment models assume leg-independence, serious concerns were raised about the correctness of the formulations. It was also indicated that fleet assignment and yield management decisions mutually affect each other. This implies that besides leg-dependent effects, the effects of yield management decisions should also considered in fleet assignment.

The dissertation is intended to enhance the basic deterministic results of the only published paper [29] dealing with the issue of network effects. Besides the deterministic demand cases, leg-dependence will be examined under stochastic demand conditions as well. Further, the fleet assignment models and formulations will be investigated in light of leg-dependence. In addition, the effects of yield management optimization on the fleet assignment solution will be addressed.
3. Previous Work and Current Practices

In this chapter we will review the major models in areas of yield management (YM) optimization, fleet assignment, and spill estimation. The goal of the overview is to review the most important models and algorithms that were developed in these fields, and introduce those methods and algorithms that will be referred to later during the rest of the dissertation.

In the first section we review some basic yield management optimization algorithms and different yield management architectures in which these algorithms are used. The presented YM algorithms differ in the way they try to maximize revenues on a network level. Later, in Chapter 7, we will refer to these approaches in our analysis. In the second part of the chapter we review the recently developed and most commonly used state-of-the-art fleet assignment models. In the last part of the chapter, we review the limited literature on spill estimation.

3.1 Yield Management Optimization Algorithms

Littlewood [28] in 1972 proposed that for a simple two class single flight problem, revenue could maximized based on the expected marginal revenue concept. Recognizing that stochastic airline demand should be modeled as a probabilistic distribution; the probability of selling, $S$, number of seats on a flight leg is:

$$P[i \geq S] = \overline{P}[S] = \int_5^\infty f(i)di, \quad (3.1)$$

where $f(i)$ is the probability density function of, $i$, number of requests for seats. The expected marginal revenue for the $S$th seat is then:

$$EMR(S) = fare * \overline{P}[S], \quad (3.2)$$

where $fare$ is the average fare level for the related fare class.

Then, assuming that lower fare classes book before higher fare classes, for the two fare class problem expected revenues are maximized if seats for higher-valued class passengers (class 1) are protected as long as the revenue of the next marginal seat, $S_1$, is higher than the fare ($fare_2$) associated with the lower-valued class (class 2):

$$EMR(S_1) = fare_1 * \overline{P}[S_1] \geq fare_2. \quad (3.3)$$

Similar results to the above were formulated by Bathia and Parekh [16] and Richter [30], although utilizing different approaches.
While the above methodologies lead to optimal seat allocation for the nested two-fare class problem; the airlines' practice of offering multiple fare classes on each flight leg, led to the need for an approach that deals with the multi-class environment and nested booking classes. Belobaba [10] in 1987 proposed a heuristic solution to the nested multiple fare class problem on a single flight leg, based on what he defined as the Expected Marginal Seat Revenue (EMSR) method. In this method the multiple class protection levels, the number of seats that are protected for a higher fare class over a lower fare class, determined by the expected marginal revenue approach originated by Littlewood. In this method (which he later named as EMSRα) the seat allocations for any pair of fare classes are taken in isolation and thus it disregards the fact that in practice fare classes are interrelated because they are nested sequentially within each other. In 1992, Belobaba [11] modified the EMSRα approach to take into consideration the joint nested probability distribution of demand and to generate joint protection levels for higher fare classes relative to lower fare classes. In this approach (referred to as EMSRβ) the protection level for the highest valued fare class, $\Pi_1$, is defined by Littlewood's formula such that:

$$EMR(\Pi_1) \equiv fare_1 \cdot \bar{P}_1[\Pi_1] = fare_2.$$  \hspace{1cm} (3.4)

In the general case, however, for class $n$, the Gaussian joint probability distribution for classes 1 to $n$ are calculated as:

$$\bar{X}_{1,n} = \sum_{i=1}^{n} \bar{X}_i,$$  \hspace{1cm} (3.5)

$$\hat{\sigma}_{1,n} = \sqrt{\sum_{i=1}^{n} \hat{\sigma}_i^2},$$  \hspace{1cm} (3.6)

$$fare_{1,n} = \sum_{i=1}^{n} fare_i \bar{X}_i \bar{X}_{1,n}.$$  \hspace{1cm} (3.7)

Based on the above joint probability distribution parameters and the weighted average fare value for the $n$ classes, the joint protection level, $\Pi_n$, is calculated as:

$$EMR_{1,n}(\Pi_n) \equiv fare_{1,n} \cdot \bar{P}_{1,n}[\Pi_n] = fare_{n+1}.$$  \hspace{1cm} (3.8)

The nested booking limit for class $n+1$ is determined as follows:

$$BL_{n+1} = Cap - \Pi_n.$$  \hspace{1cm} (3.9)

Curry's optimal booking limit (OBL) approach addresses the problem of calculating optimal booking limits for more than two nested fare classes (the same problem was independently solved by Brumelle and McGill [18], Wollmer [44], and Robinson [31]). Similar to Belobaba's algorithm, it is based on the Expected Marginal Seat Revenue approach: A seat is protected for nested classes 1, 2, and $c-1$, till the
combined expected marginal seat revenue (EMSR) of these classes is bigger than the fare value of class $c$:

$$EMSR_{1,c-1}(\Pi_c) \geq fare_c.$$  

Figure 3.1 shows the relationships between $EMR(S)$, $fare_n$, $Cap$ and the nested booking limit $BL_n$. The figure also shows the combined protection levels for classes 1 to $n-1$.

![Figure 3.1: Nested booking limits EMR curve and protection levels on a single flight leg](image)

The difference between Belobaba's and Curry's algorithm is the way the combined nested expected marginal revenue is calculated. Belobaba uses a weighted, computationally easy formulation, while Curry uses a complex formulation utilizing convolution integrals. The protection level for the highest fare classes are identical to Littlewood's rule. The protection levels for other classes can be found by recursively solving the optimality conditions:

$$SL_i[\Pi_{i-1}, \Pi_i] = fare_{i+1}$$

$$SL_i[\Pi_{i-1}, \Pi_i] = fare_i \int_{n_i-n_{i-1}}^{n_i} f_i(r_i) dr_i + \int_{0}^{\Pi_i} dr_i f_i(r_i) SL_{i-1}[\Pi_{i-2}, \Pi_i - r_i]$$  \hspace{1cm} (3.10)

where $\Pi_0=0$; and $SL_i$ is the slope of the combined expected revenue function of fare classes 1 to $i$. The optimal booking limit for each class, $BL_i$, is then calculated as in was given in (3.9).

Extensive testing by various researchers [18], [31], showed that in realistic airline environments the difference in the expected revenues between Belobaba's heuristics and Curry's OBL method is very small. It shows the robustness of the computationally very advantageous EMSRb method. Consequently, through this dissertation, in the different steps of the modeled YM approaches we calculated leg-based nested booking limits according to the EMSRb algorithm.
For additional models and studies in the area of yield management optimization, the reader is referred to the works of Williamson [43], Weatherford [42].

3.2 Yield Management System Approaches

In this section we will introduce three different yield management approaches that characterize airline practices. The three approaches represent three different steps in the evolution of yield management systems: Conventional fare class (yield) based control; Value class based control; and Bid Price OD heuristics. Each system represents an additional step in the direction toward the OD control of passenger flows. Although the fare class (yield) based system has obvious shortcomings if the airline’s objective is to maximize total revenues over a network of flights, it is still part of most airlines’ yield management practices. Therefore, including this approach into our analysis seems to be up-to-date. Although there exist other yield management approaches, we limited our analysis only to these three systems in this dissertation. Nevertheless, we believe that these approaches are good representations of the currently used yield management practices.

3.2.1 Conventional Fare Class (Yield) Based Control (FC)

Conventional fare class based control (FC) is still a common form of yield management approach used by airlines. In the FC system ODF itineraries are assigned to booking classes, according to their fare classes or fare class designators. That is, irrespective of the ODF itinerary’s actual fare value, all full fare “Y-class” ODF’s are aggregated to the top booking class, and for example, all M class discounted ODF itineraries are aggregated also to the corresponding M booking class. The aggregation of the fare classes hence is based strictly on fare class types, and yield.

On any flight leg, a booking class can be regarded as an aggregation of different OD fare products. Given the fare class booking classes, nested leg-based optimization techniques (EMSRb or OBL) are applied to each flight leg independently to calculate nested protection levels and booking limits for each booking class. Since logic dictates that higher valued passenger bookings should always be accepted when seats originally allocated to lower valued fare classes are not filled, and because it has been shown that the nested structure of inventory classes generate higher revenues than partitioned or distinct structures [10], these optimization systems calculate nested protection levels and nested booking limits.

One of the most critical issues in the fare class based control system is calculating the revenue values associated with each booking class and used as input to optimization models. In the fare class based

---

1 We will use the word revenue for the aggregated fare associated with the booking classes, and will use fare for the ODF original fares.
control system the revenue value associated with a booking class commonly is the weighted average fare
of the different ODF total fare values weighted by the mean ODF demands.

Some systems use prorated fare values for the weighted fare calculations while other systems use non-
prorated fare values giving a "flavor of OD mix" in each class. In our examples we will use non-prorated
fare values in calculating the revenue value calculations associated with the booking classes. Using the
total fare value instead of the prorated fare, information about the possible network revenue is introduced
into the booking class revenue calculation.

Since the basis of aggregation is simply the fare class designator, often fare classes with substantially
different fare values are grouped together. For example, under the FC system, on the BOS-JFK flight leg,
the local BOS-JFK full fare ODF and the BOS-JFK-LHR connecting transatlantic full-fare ODF will be
aggregated into the same booking class. Although both OD demands are full fare demands, there could be
a very large difference between their associated fare values. Similar large differences can be found in the
discount booking classes as well. Although ODFs share the same booking class, their revenue
contributions may be substantially different.

Under the FC system, seat availability for multiple-leg itineraries is determined by minimum class
booking limit across the legs to be traversed by the ODF. That is, an M class ODF itinerary is accepted
only if along all flight legs of the itinerary the booking limit associated with the M booking class is greater
than zero.

As we have noted in the previous chapter, the conventional leg-based YM approach has some obvious
shortcomings if the airline's objective is to maximize total revenues over the network. Maximizing
revenues over flight legs individually in no way guarantees that total revenues are maximized [43].
Revenue maximization over a network of connecting flight legs requires a combination of two capabilities
[13]. First, there must exist a capability to give preference in terms of seat availability to high revenue
connecting (long-haul) passenger, regardless of the ODF's yield. Second, in certain situations the system
has to prevent these same high revenue long-haul connecting passengers from taking seats away from two
(or more) high-yield local passengers, each of whom occupy seats only on one leg but whose combined
summed revenues can be higher than that of the connecting passenger.

The above two tasks can be achieved simultaneously if the yield management system is able to control
passenger bookings at the ODF level. A basic mathematical formulation for the deterministic seat
inventory problem as a linear program is shown below. The objective is to maximize the total network
revenues subject to the given capacity constraints and forecasted demand:
\[
\begin{align*}
\text{Max} & \sum_{od} \sum_{c} \text{Fare}_{od,c} S_{od,c} \\
\text{s.t.} & \sum_{od} \sum_{c} S_{od,c} \leq Cap_{l} \quad \forall l \\
& S_{od,c} \leq D_{od,c} \quad \forall (od,c) \text{ pairs}
\end{align*}
\]

where the \( S_{od,c} \) decision variables represent the allocation of seats to an OD itinerary, \( od \), and fare class, \( c \). The objective coefficients are the fare values associated with the ODF, \( (od,c) \). The first set of constraints limits the total assigned seats on each flight leg, \( l \), by the aircraft seat capacity limit on the leg, \( Cap_{l} \). The second set of constraints prevents allocation of more seats to an ODF than the actual demand for that ODF, \( D_{od,c} \). A probabilistic formulation for the above problem can be obtained by using the concepts of probabilistic distributions and expected marginal revenues [43].

There are various theoretical and practical obstacles that limit the effectiveness of the linear programming formulation to the OD YM problem. Among them are the problem of very dynamic probabilistic demand, and the problem of small ODF demand values. (For an extensive discussion of the problems refer to [43], or [13]). Nevertheless, there are various yield management optimization and control structure approaches that try to approximate heuristically the ODF level optimization and control. Most of them heuristically estimate a displacement cost for ODF itineraries or incorporate network shadow prices into the decision process. Often these approaches are referred to as "bid price" approaches. In the following, we will describe one of the recently developed approaches as well.

### 3.2.2 Value Class Control (VC)

Value class yield management introduces a certain degree of OD control into the leg-based yield management control strategy, by providing increased availability to high fare long-haul passengers. In this method value classes are defined replacing the booking classes of the FC system. Value classes represent revenue ranges, and the different ODF itineraries are aggregated and mapped into booking classes based on their total fare value instead of fare class/type. This way the aggregation is based more on the network revenue contribution of the ODF, and enables the control system to manage bookings by the actual revenue value to the network. In this approach the value of the itinerary is defined as the total fare of the ODF itinerary (although other values are also possible).

The more value classes that are defined, the finer grained is the classification and the aggregation, and the greater the level of control the yield management system has to differentiate between the value of various ODF itineraries. In this system, value class revenues are also determined as the weighted average of the ODF fares grouped into the value class. Leg-based nested protection levels and booking limits are calculated for each value class on each leg independently using EMSRb or OBL algorithms. Seat availability for an ODF depends on the corresponding value class availability along the ODF’s path.
In this control structure, the conceptual difference is that value classes no longer represent common fare class products (e.g., unrestricted full fare, 14 day advance purchase excursion fare), but rather an aggregation of demands with similar total fares. Thus, it abandons the traditional “fare type” consistency.

The actual result of this approach is that ODF’s with higher total fare are assigned to higher valued booking class buckets, thus yield management optimization algorithms protect more seats for these higher total revenue itineraries; consequently OD fare classes with lower total revenues that are aggregated to lower value classes will be more likely to be spilled.

Since ODF fares are used as the basis for mapping OD fare class itineraries into the value buckets, when this method is combined with using leg-based yield management algorithms like EMSRb or OBL, the result tends to be “greedy” [43]. Since on given leg the connecting long haul ODF fares are more likely to be higher than the local or short haul ODF fares, the approach gives priority to the long haul connecting passenger flows over the short haul or local itineraries. In many cases this decision is a correct solution with respect to the network optimum, however, always giving priority to the long haul OD demands under certain network and demand conditions may also result in negative revenue impacts. This is the case when, for example, long haul connecting passengers may displace the combination of short haul or local passengers with a higher combined total revenue than that of the connecting passenger. In many cases the VC method generates better overall revenues than the FC method [43], yet under certain special network conditions (mostly when most of the leg demand factors are high and balanced) the FC method can lead to better results because in the VC method displacement costs are not considered.

3.2.3 Displacement Cost Bid Price OD Heuristics (BD)

The displacement cost based real-time leg-based control approach is a heuristic to incorporate the positive characteristics of the VC method and at the same time to correct for its „greediness”. In this method, ODF demands are mapped initially to value buckets the same way as in the VC method, but during the booking process they are evaluated with respect to the ODF itinerary’s approximated network revenue contribution. The reason for real-time evaluation is, as we have shown above, the true revenue value of each OD fare class itinerary on a flight leg depends on the demand/capacity ratio and traffic flows across all flight legs in the network [12]. This method’s challenging task is to determine the „network value” of a requested ODF without network optimization. The BD approach uses the concept of displacement cost to determine the network value of an OD fare class booking request and correct for the greediness of the VC approach.

Under leg-based control the total fare value of a connecting ODF itinerary needs to be „reduced” on a flight leg by the forecast revenue displacement on connecting legs. That is, for any ODF itinerary, the network value is its fare minus the expected revenue displacement on connecting legs. The expected
displaced revenue on the connecting legs is a function of the forecast leg demand relative to its capacity, and the proportion of the local demands comprising the total demand of the leg.

The displaced revenues can be derived from the flight leg's Expected Marginal Revenue (EMR) function given in equation (3.2). The EMR functions are generated by the existing leg-based yield management algorithms, therefore, the required data and the calculations are easily obtainable. On any flight leg, displacement costs can be approximated by a function of the EMR value evaluated at the remaining available capacity \( A \) for the flight leg, \( EMR(A) \). Figure 3.1 shows a typical leg EMR function. The EMR value at the available capacity (\( Cap \)), incorporates aggregated information about the total fare value of the last seat on the leg, containing both local and multiple leg demand total revenues. That is, the EMR value overestimates the downline displacement costs, which need to be corrected. The most important issue to be addressed is the displacement of local passenger on all legs of a connecting itinerary. The revenue displacement of the associated with a local passenger is related to the probability that the marginal seat will be booked by a local passenger, which is relatively easy to estimate.

The heuristic approach considered in this thesis approximates the network revenue value, \( N_{ODF1} \), to leg 1 of an ODF itinerary that traverses both leg 1 and 2 as:

\[
N_{ODF1} = Fare_{ODF} - [DISP \times EMR_2(A_2)] \quad \text{where} \quad (3.12)
\]

\[
DISP = P_{loc1} \times P_{loc2}, \quad 0 \leq DISP \leq 1.0. \quad (3.13)
\]

where \( EMR_2(A_2) \) is the expected marginal revenue of the flight leg 2 evaluated at the available seat capacities \( A_2 \) on that leg, and \( P_{loci} \) is the probability of selling seat \( A_i \) to a local passenger on the leg \( i \).

Similarly, the network revenue value to leg 2 for the same ODF itinerary is given by:

\[
N_{ODF2} = Fare_{ODF} - DISP \times EMR_1(A_1). \quad (3.14)
\]

Finally, the decision rule for ODF booking request is to determine the minimum acceptable value for a particular ODF request, and accept the ODF request if the associated fare value exceeds the minimum acceptable value. Thus, the decision rule can be expressed as, accept the ODF booking if:

\[
Fare_{ODF} \geq \max[EMR_1(A_1) + DISP \times EMR_2(A_2), EMR_2(A_2) + DISP \times EMR_1(A_1)]. \quad (3.15)
\]

This type of acceptance rule has been referred as "bid-price" in the context of network revenue optimization. Because, in this decision rule no network optimization is used, this approach can be rather referred to as "leg-based bid price" model. For additional information and the positive impacts on revenues of this model, refer to [13].
3.3 Fleet Assignment Decision Models

Most of the formulations of the fleet assignment problem published since the mid 1980's, have been concerned with assigning aircraft types to already scheduled flights, given various constraints, with the objective of maximizing profits. Hence, in these formulations, the schedule with its flights, with their departure and arrival times, is given, and the decision is made as to what type of fleet should fly the scheduled flights. Note that decisions about frequency of service and route selection are not part of the aircraft assignment model, although certain models were developed with the capability of suggesting which flights of the schedule could be omitted. These models also use integer formulations for the problem, thus resulting in integer solutions. Given the integer solution, assigning the tail numbers to each flight with respect to the required maintenance schedules becomes a straightforward process in a domestic network.

The new technologies that were developed since deregulation also had positive aspects on the fleet assignment problem. The extensive use of the Computer Reservation Systems (CRS) and Yield Management Systems (YM) enabled the airlines to keep track of demand on a much more detailed level. Airlines could provide the optimization problem with demand information that captures the seasonal, day-of-week, and time-of-day variations.

Daskin and Panayotopoulos [22] in 1989 presented a Lagrangian relaxation approach to assigning aircraft to a hub and spoke route structure. A large integer programming formulation of the aircraft assignment problem to a fixed schedule to maximize profits was formulated, which provides an upper bound to the fleet assignment objective function. In the same article, various heuristics are provided to convert the Lagrangian solution into a primal feasible solution, with the assumption that the Lagrangian relaxation problem is embedded into a branch and bound algorithm to obtain integer solutions.

The presented Lagrangian relaxation algorithm is a big step in the modern fleet assignment algorithms, however its applicability in practice may be limited. Lagrange relaxation is practical if by relaxing constraints the new formulation of the problem becomes easily solvable. However, a practical formulation of the fleet assignment problem would contain many more important operational and user specific

---

2 Earlier developed frequency planning models (then called as fleet assignment or aircraft assignment models, but we will refer to them as frequency planning models) tried to find the optimal set of routes to be served, the frequencies to be offered, and the aircraft types used for each flight leg [23], [33]. These models by definition assumed that airlines can freely adjust their schedule to the model recommendations: That is, they assumed that airlines can freely choose routes and change the flight frequencies. However, in today's airline practices traditional decision variables, (like frequency, which route to serve, etc.), could not be decided strictly by mathematical or economic optimization methods and thus the models are not widely used in practice. A good overview of these models is given in [23]. In this dissertation we refer to the fleet assignment model, as defined in 2.1, to the problem of assigning fleet types to a fixed schedule.
constraints than were considered in the proposed Lagrangian relaxation formulation in [22]; thus relaxing one set of constraints would still leave the rest of the problem too complex and not easily solvable.

In the late 1980’s and early 1990’s significant advances in computer hardware and large scale mathematical optimization algorithms made it possible that large scale integer problems, for the size of a major airline, could be solved efficiently. Abara [1] reports an integer linear programming approach used at American Airlines for the fleet assignment problem. His formulation also considers the assignment of fleets to an already designed flight schedule. The objective function can take various forms, including profit maximization, cost minimization, and optimal utilization of fleet type. The model can handle the cases when all flight are to be served, and when some may be dropped. There are five main groups of constraints:

- **flight coverage** -- specifying that each flight is served only once
- **continuity of equipment** -- assuring the integrity of the network
- **schedule balance** -- by station and by aircraft type, assuring continuity
- **aircraft count** -- assuring that only the available number of aircraft are assigned
- **user specific constraints** -- lower or upper limits on any flight related variable, (e.g., aircraft utilization, airport limitations, limits of slots, etc.)

The objective function maximizes the contributions of the selected flights less the cost of aircraft used, the cost of aircraft shortages, and the cost of stations. Each decision variable in the objective function represents a feasible “turn” of an aircraft type in the schedule. (A turn represents a connection of equipment from an arrival to a possible later departure in the constructed time-space network.) There was no explicit explanation given how the $P_{jk}$ objective coefficients were derived and estimated for the model.

The mathematical formulation of the objective function can be written as:

$$\max \sum_{i=0}^{F} \sum_{j=0}^{E} \sum_{k=1}^{K} P_{jk} X_{ijk} - C_{1} \sum_{i=1}^{F} \sum_{k=1}^{K} X_{0ik} - C_{2} \sum_{k=1}^{K} e_{k} - C_{3} \sum_{s=1}^{S} \sum_{k=1}^{K} O_{sk} + T_{sk} - \sum_{k=1}^{K} C_{S_{k}} \sum_{s=1}^{S} Y_{sk} \quad (3.18)$$

subject to the previously defined constraints, where:

- $X_{ijk}$ = feasible turn (flight leg $i$ turns to flight leg $j$ on aircraft type $k$; if $i=0$ then $j$ is a sequence of origination; if $j=0$ then $i$ is a sequence of termination; where sequence represents a daily routing of aircraft)
- $e_{k}$ = extra aircraft of type $k$ used beyond specified (beyond the number of available aircraft)
- $P_{jk}$ = benefit (or profit) of operating flight $j$ on aircraft type $k$,
- $O_{sk}$ = sequence origin shortage of aircraft $k$ at station $s$,
- $T_{sk}$ = sequence termination shortage of aircraft $k$ at station $s$,
- $Y_{sk}$ = indicator of service/no service of aircraft $k$ at station $s$,
- $C_{1}$ = nominal cost of aircraft used (typical value=1)
- $C_{2}$ = large cost per extra aircraft
- $C_{3}$ = large cost of imbalance (shortage)
- $C_{S_{k}}$ = imposed cost (penalty or reward) for each station served by aircraft type $k$,

$F$, $K$, $S$ are the number of flights, aircraft types, stations, respectively.
As reported by Abara, the formulation size becomes very large for medium size problems. Note, that the selected form of the decision variable $X_{ijk}$ is not the best possible, since one variable is needed for each turn possibility. According to Abara [1], a formulation of 400 flight schedule, 60 stations, and 3 aircraft type would require about 6300 columns and 1800 rows. He notes that the continuous solution often is integer or fractional for a few flights. However, in many cases an integer solution requires many integer LP iterations and results in time consuming calculations.

The Coldstart Model of Delta Air Lines [24],[35], has also emerged on the cutting edge of technologies in computer hardware and mathematical optimization algorithms. The Coldstart model is a large scale mixed integer linear program which assigns different types of fleets to flight legs by minimizing the combination of operating and passenger spill costs [35]. The model solves the daily domestic fleet assignment problem with side constraints defined on a time-expanded network [24]. The challenge is that this kind of formulation often degenerated severely, which leads to poor performance using standard methods, and the large number of integer variables make the computation time consuming.

The model uses a time-space network and a fundamental mathematical structure. The network has a circular (24hour) time line for each aircraft fleet at each city. A node represents a point in time where at least one departure or arrival occurs. A flight leg is represented by "sky arcs" for each feasible aircraft type, (where the beginning of the arc corresponds to the departure time and the end of the arc corresponds to the next ready time of the aircraft at the end station in the time line), and the nodes of events in a city are connected by "ground arcs". Hence, the time lines of different cities are connected by "sky arcs"; however, the time line of the different fleets are not connected.

The integer programming formulation of the basic Coldstart model is given by [24]:

$$\text{min} \sum_{i \in L} \sum_{f \in F} \text{cost}_f X_{fi} \quad (3.19a)$$

s.t. \hspace{1cm} \sum_{f} X_{fi} = 1 \hspace{0.5cm} \forall i \in L \quad (3.19b)

\hspace{1cm} \sum_{f} X_{fot} + Y_{for} - \sum_{f} X_{fodn} - Y_{fot} = 0 \hspace{0.5cm} \forall fot \in N \quad (3.19c)

\hspace{1cm} X_{f} - Y_{f} = 0 \hspace{0.5cm} \forall (i, j) \in H \quad (3.19d)

\hspace{1cm} \sum_{i \in L(f)} X_{fi} + \sum_{o \in L(f)} X_{fod} \leq S(f) \hspace{0.5cm} \forall f \quad (3.19e)

\hspace{1cm} Y_{fot} \geq 0 \hspace{0.5cm} \forall fot \in N \quad (3.19f)

\hspace{1cm} X_{fi} \in \{0, 1\} \hspace{0.5cm} \forall i \in L, f \in F \quad (3.19g)

where $X_{fi}$ or also written as $X_{fodi}$ is the decision variable, which has a value of 1 if fleet $f$ flies the flight leg $i$, from $o$ to $d$ departing at time $t$; and 0 otherwise.

$C, F, L$ denotes the set of cities served, available fleets, and flights legs of the schedule, respectively,

$S(f)$ the number of aircraft in each fleet,

$i$ or $odt$ is a flight leg of $L$ with $o, d \in C$ and $t$ is time,
The objective is to minimize the total costs of the assignment. The cost of assigning fleet $f$ to leg $i$, where the cost is a summation of the operating and spill costs. For a more detailed description of spill costs refer to the next section.

The four main sets of constraints (similar to those of Abara) are the following.

The (3.19b) cover constraint forces each flight leg to be flown by exactly one type of aircraft.

The (3.19c) balance constraint assuring the flow conservation at each node for all the fleet types. Basically it says that at each node of the time-space network, for aircraft type $f$, the number of incoming aircraft at time $t$, ($\sum_d X_{fot}$), less the number of departing aircraft at time $t$, ($\sum_d X_{fot}$), plus the number of aircraft being on the ground at city $o$ before $t$ ($\sum_{o} Y_{fot}$), must equal the number of the remaining aircraft on the ground in city $o$ after time $t$ ($\sum_{o} Y_{fot}$). (Note that in the time space network a number of arriving and departing flights could be compressed into one node.)

The (3.19d) constraints are the hookup constraints. It is required sometimes by marketing considerations that the same type of aircraft should fly a multiple leg route. This constraint provides for this scenario.

The (3.19e) constraints are the fleet size constraints, which prevent the model from assigning more aircraft to the schedule than available.

Delta Air Lines expanded the above basic model with additional operational constraints. Among these are maintenance, pilot training, pilot hour, crew break out, crew rest, and noise constraints [35]. The model can be extended and modified to deal with various other problems with minor changes in the constraint matrix.

It is reported that a typical problem for Delta Airlines has about 4000 constraints and 60000 variables. To solve this problem the OB1 interior point code is used, and for the reduction of the problem size they use model enhancements like: node consolidation, island construction, eliminating missed connections. In the solution technique, cost perturbation, branching on set-partitioning constraints, and prioritizing on the order of branching is used [24]. The enhanced model runs one to three hours on the IBM workstation (10 fleets and 25000 flight segments).
The above described models are used extensively and successfully by airlines in the medium term planning practices. These special solution techniques of the above models made it available that very large scale optimization problems could be solved not only heuristically but optimally in reasonable time.

**Short Term (Dynamic) Fleet Assignment Models**

The models discussed thus far solve the fleet assignment problem in advance for a certain time interval (usually one or more months), and the solution serves as the main assignment for that period. The variability of market demand, however, can require dynamic adjustments to the original fleet assignment decisions. Responding to this need, Berge and Hopperstad [15] developed a new operating concept and methodology called "Demand Driven Dispatch (D^3)". Utilizing the demand forecasts of the Yield Management Systems, which improve closer to departure and may show deviations from the earlier forecasted demand, aircraft are reassigned to legs that better match the expected final demand.

The underlying model assumes that data would be supplied from the yield management system using its demand forecasting algorithms. For the model it is also assumed that an airplane family of two or more models, with different seating capacity but with a common flight crew rating, is also available.

The large scale and combinatorial nature of the formulation and the required frequent daily runs led to the use of heuristics in the solution. One of the suggested heuristics, Sequential Minimum Cost Flow (SMCF), is based on the following observation: a two-type aircraft assignment problem can be reduced to a single commodity minimum cost problem. Using this observation they solve the problem in an increasing sequential order with respect to the aircraft capacities. The algorithm is based on the assumption that a decision to place a particular aircraft type on a leg must consider the alternative aircraft of the next size.

The second heuristic: Delta Profit Method (DELPRO) starts from a feasible assignment and performs multiple swaps of airplanes of different types in cases when the total change in revenues (delta revenue) along the paths are positive. The algorithm uses longest path algorithms to look for the biggest profit improvements along the possible paths of the different aircraft assignment. The cost of arcs in the time-span network are equal to the difference in expected profits using the new candidate aircraft versus the currently assigned one. It is reported that the two heuristics approximate the optimal solution very well with a much faster computational time.

Although in D^3 model incorporated certain yield management information in decision process, like many the other fleet assignment approaches, it ignored the existence of possible network effects.
3.4 Overview of Spill Models

*Spill* is the fraction of demand that cannot be accommodated because of the limited capacity of the aircraft. Spilled passengers may be lost to the airline or may be accommodated on other flights of the same airline. The latter case is called the *recapture* of demand. Recaptured passengers may change the original distribution of demand for a flight. In this research, however, we do not take recapture into consideration.

The main inputs for the spill and spill cost estimation models are the unconstrained demand, the fare values, and the aircraft capacity on the flight leg. Given the input data, the expected number of spilled passengers can be calculated. The expected spill cost is the product of the expected number of spilled passengers and the corresponding average fare of spilled passengers.

*Demand* for a flight leg is represented as a probability distribution. Past analyses have suggested that airline demand can be well represented by a normal density function [10],[27],[38]. Normal demand can be characterized by two statistics: the mean and variance.

The probabilistic distribution of demand describes the uncertainty of demand due to randomness of passenger behavior or due to some cyclic variations. The former is due to the uncertainty of the passenger behavior. It is assumed that in general the standard deviation of demand due to randomness scales linearly with the square root of the mean demand [37][34], that is,

\[
\sigma = c \cdot \sqrt{\mu}, \tag{3.20}
\]

where \(c\) is a scaling constant ranging usually between 1\(<c<2.5\). This implies that high mean demand would have low \(K\) factors; while low mean demand would have higher \(K\) factors, that is, bigger variability around its mean. The intuition behind this is that at higher demand levels random effects average or cancel out themselves.

The random variability of demand is due to the process of large group of people independently deciding whether or not to take a flight at a given time [37]. Assuming that the random variables describing the individual's decisions are independent, the Central Limit Theorem would suggest that as the number of individuals is increasing the total demand approaches a normal distribution. Therefore, the assumption that demands for a flight can be correctly represented by a normal density function is also intuitively reasonable.

Another often used representative of the spread of demand is the coefficient of variation, \(K\), which is defined as the ratio of the standard deviation to the mean:

\[
K = \frac{\sigma}{\mu} \tag{3.21}
\]
A large K factor reflects large variations around mean; a low K factor indicates a narrower spread. The K factor for a single flight is typically 10%.

In practice, however, demand is estimated by aggregating flight data over a longer time period. The estimated demand over that time period may include daily, weekly and monthly cyclic variations as well. The changes in the mean demand during the time period increase the standard deviation, and hence also the K factor of the demand distribution. For these cases the typical K factor is 33% [38].

*Spill* is the demand that cannot be accommodated because of the capacity limitations. Hence, given any demand distribution spill on a flight leg can be calculated by the following formula [34]:

\[
Spill = \int_{i=\text{Cap}} f(i) \ast (i - \text{Cap}) \ast di ,
\]  
(3.22)

where \(\text{Cap}\) is the capacity of the aircraft, and \(f(i)\) is the probability density function for demand request \(i\).

In words spill is the integral sum of the probabilities of demand requests \(i\), weighted by the number of passengers turned away, \((i-\text{Cap})\).

An alternative way of deriving the number of spilled passengers is the following: The probability of spill occurring, i.e., the probability of demand to be more than capacity is given by:

\[
\overline{P}(\text{Cap}) = 1 - P(\text{Cap}) = \int_{i=\text{Cap}} f(i) di ,
\]  
(3.23)

where \(P(i)\) represents the cumulative probability function of \(f(i)\). Therefore the total expected number of passengers spilled can be calculated as the integral of \(\overline{P}(i)\) for all seats beyond capacity; hence:

\[
Spill = \int_{i=\text{Cap}} \overline{P}(i) di .
\]  
(3.24)

*Expected Load*, the expected number of passengers to be accommodated, can be calculated by the following formula:

\[
\text{Exp Load} = \int_{i=0}^{\text{Cap}} i \ast f(i) di + \text{Cap} \ast \int_{i=\text{Cap}} f(i) di .
\]  
(3.25)

or similarly by the cumulative probability function approach:

\[
\text{Exp Load} = \int_{i=0}^{\text{Cap}} \overline{P}(i) .
\]  
(3.26)

It is easy to see from equation (3.22) and (3.25) that the sum of the expected load and the spill is equal to the mean of the demand distribution, \(\mu\):

\[
\mu = \text{Exp Load} + Spill
\]  
(3.27)
The above formulae hold for any type of demand distribution.

If we assume that demand, \( f(i) \), is represented by a normal probability function, i.e.,

\[
f(i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-(i - \mu)^2 / 2\sigma^2\right\},
\]

where \( \mu \) and \( \sigma \) are the mean and standard deviation of probability density function \( f(i) \), then substituting equations 3.28 in the spill (and expected load) formulations would result in the unfortunate property of not having a closed form.

In the late 70’s and early 80’s Swan studied the problem of spill estimation thoroughly [37],[38]. In his works he developed the basic spill model, which relies on the concept of a buffer. The buffer, \( b \), is the space between the capacity and mean demand measured in standard deviations, and is available to absorb the variations of demand.

\[
b = (\text{Cap} - \mu) / \sigma
\]

(3.29)

For tractable algebra Swan suggests that the normal distribution be approximated by the logistic distribution. Besides approximating the normal distribution well, the logistic distribution is analytically convenient [14]. Although the approximation is good, the logistic distribution has fatter tails. The logistic distribution is given by:

\[
f(x) = \frac{me^{-mx}}{(1 + e^{-mx})^2},
\]

(3.30)

where \( m \) is a positive scale parameter. Its cumulative probability function, \( P(x) \), is given as:

\[
P(x) = \frac{1}{1 + e^{-mx}}.
\]

(3.31)

From which, \( \bar{P}(x) \) is given by:

\[
\bar{P}(x) = 1 - P(x) = \frac{1}{1 + e^{-mx}}.
\]

(3.32)

Spill is the integral of \( \bar{P}(x) \) for all potential seats beyond capacity. If \( x = (i - k) / \sigma \), (where \( i \) and \( k \) are demand levels and \( \sigma \) is the standard deviation of demand), and \( b \) refers to the "buffer" \( (b = (\text{Cap} - \mu) / \sigma) \), then spill is given by:

\[
\text{Spill} = \sigma \cdot \int_b^{\infty} \frac{1}{1 + e^{mx}} \, dx = \frac{\sigma}{m} \ln(1 + e^{-mb})
\]

(3.33)
Parameter $m$ is the scaling parameter which is designated to correct the variance of the approximation. Swan suggested that for the purpose of spill estimation the value $m=1.7$ be used. He estimated that at the practically used typical load factor ranges this value makes the curve approximate the cumulative normal curve the best. Therefore, Swan's spill approximation formula, also called as the "S-curve approximation", can be written as:

$$Spill = \left(\frac{\sigma}{1.7}\right) \ln(1 + e^{-1.7b}). \quad (3.34)$$

Because the above approximation gives very close results to the real spill estimates, equation (3.34) is a fast and practical tool for estimating the spill for a flight leg when only fare class is considered. Swan's work thus provided the airline industry with a practical formula for estimating spill for a flight leg when demand can be correctly represented in a form of a single Gaussian density function which represents the aggregated total demand for a flight.

During the 1980's, airlines began offering a wide range of discounted fares and many developed multiple fare class structures to manage the capacity available on each flight. To maximize revenues by optimally allocating seats to fare classes, airlines developed Yield Management (YM) systems, as described earlier. Given the demand forecast and fare value for each fare class offered, YM systems set booking limits for each fare class such that expected revenues are maximized. To respond dynamically to the stochastic nature of the booking process, airlines run YM optimization algorithms periodically many times during the booking process of a flight.

Despite the introduction of YM systems and the widespread use of seat allocation algorithms, not much attention has been paid to the impacts of these practices on spill estimation. One exception is Swan, who revisited the basic spill model in an AGIFORS presentation [39] and in an unpublished paper [40].

At the AGIFORS Yield Management conference in Brussels, Swan addressed the issue of using yield management information in aircraft assignment applications [39]. He indicated that information from Yield Management systems can be used to estimate better the demand characteristics of the different aircraft assignments to the different flight legs. Particularly, he suggested the use of Yield Management information to calibrate spill model parameters, which affect the results of the aircraft assignment decisions. He argued that yield management information can help to estimate:

- effective capacity (the average capacity of flights that were closed for bookings),
- $K$ factor ($\sigma/\mu$) for demand,
- fare mix of spill,
- fare value of spill.

For example the fare mix of spill can be estimated by using information about the forecast of bookings after the fare classes were closed. Swan also points out that there are a lot of differences in the Yield
Management, and in the Scheduling "thinking". The former works with detailed bookings, on a daily
time frame; while the latter thinks in terms of onboard passenger loads, with variation and time frames for
a month. Swan's presentation was very important in the sense that it raised the issue and the possible
benefits of incorporating yield management information into aircraft scheduling.

In [40], Swan recognizes that YM systems may influence the correct spill estimates. He argues that the
no-show phenomenon and the presence of discounted fares affect the actual spill values. Swan also notes
that the revenue spilled varies with the fare class mix of demands spilled and the amount of spill. At low
values of spill, the mix is nearly the same as the mix of demands, while at high values of spill, the spill is
dominated by lower fare passengers, due to the use of yield management practices. Swan suggests
approximations to correct for the problem, but does not consider more than two fare classes or present a
specific algorithm to calculate spill in such cases. To incorporate the effects for yield management, he
introduces the term "effective capacity", which as a load factor can range from 85% to 98%. This
effective capacity is affected also by the fare discounts, which he approximates as a value typical to the US
domestic operations and suggests to use an average revenue value for a flight.

Spilled passengers may be lost to the airline or may be accommodated on other flights of the same airline.
The latter case is called the recapture of demand. Hence, recaptured passengers are not lost to the airline.
Recaptured passengers may change the original, "nominal", demand distribution of a flight. The effect of
recapture may result in a flatter demand distribution which is shifted slightly to the right. Swan shows
that if demand is spilled the chances of re-accommodation on other flights are lower than average, due to
correlation of high demand on alternative flights and due to that spilled passengers are the one who book
last [39].

Spill estimates, beside its traditional areas, can play a significant role in other planning processes as well.
Traditionally spill is used in schedule and fleet planning applications. It is used in fleet assignment
models, that assign fleet types to the fixed schedule. In these cases, input demand represent a flight leg
over a longer (one or more month) period. The distribution of demand for this case, beside the
randomness in passenger behavior, will also include the cyclic variations over the period. Spill however,
is used in the short term dynamic fleet assignment models as well, (see previous section). In these
applications, spill is estimated for a single flight and not for a large group of flights. Thus the relevant
input data does not include cyclic variations of demand. Spill (or load) can be an important parameter for
revenue or cost analysis purposes as well. Additionally, one-flight specific spill or load estimates can be
helpful in daily planning for flight related on-board or ground services, and freight planning.
3.5 Summary

The chapter gave an overview of the currently used fleet assignment and yield management algorithms and approaches. Throughout the dissertation we will often refer to these approaches. The chapter also has shown that traditional fleet assignment models assume that flight legs are independent, thus assume leg-independence among the spill estimates as well. The leg-independence assumption may be correct in networks where demand on all flights legs is local. However, in networks where significant connecting OD demand is present, the network is characterized as leg-dependent. In these networks the leg-independence assumption may raise serious concerns about the correctness of the traditional fleet assignment formulations.

The overview of the spill estimation approaches has shown that although it was recognized that yield management information can be used for better estimating spill, effects of the yield management systems on spill estimates were considered only to a very limited extent. The review also noted that traditionally spill estimates were used in medium term planning applications, as an aggregation of observations over a longer time-period. This aggregation may result in some estimation biases, as we will show throughout this dissertation.

Finally, we can conclude that network effects are ignored in both the fleet assignment models and in the spill models as well. Only the paper of Phillips et. al. [29], addressed earlier in Section 2.2 and will be addressed in more detail in Section 6.1, realizes that network effects can affect the actual passenger flows in the network. The paper however, addresses only the deterministic cases and directly does not address the implications on fleet assignment. In addition the paper does not recognize the effects of yield management optimization on calculating accurate OD passenger flows.
4. Network Effects and Leg-Dependence in Airline Networks

The fleet assignment problem, as described in the literature review, has attracted the interest of researchers for many years. Research and development work, however, has concentrated mostly on operational issues of the problem, that is, on the constraints and restrictions that involve the flow of aircraft in the network. Unfortunately, not enough attention has been devoted to the analysis of the fleet assignment problem from a broader, modeling perspective. In this chapter, will show how network effects, which have been disregarded in the traditional fleet assignment formulations, can influence the load and spill estimates on each leg of the network. We also argue that disregarding network effects in the fleet assignment models can endanger the accuracy and correctness of the results. For better understanding, we will start the analysis with cases when demand is represented in a deterministic form and thereafter, we will address the issue when demand is stochastic.

As it was defined in Chapter 2, network effects (or leg-dependence) refers to the phenomenon that a leg-based demand estimate can be influenced not only by the capacity and demand conditions of the actual leg, but also by the capacity and demand conditions on the other legs of an airline network.

According to our knowledge, as it was mentioned in Section 2.2, only one published paper [29] tries to model the interaction of the OD passenger flows in the networks. The paper [29], however, lacks the discussion of cases when demand is given in a probabilistic form. This chapter moves one step forward and the interaction of OD demands in the stochastic environment is addressed and an approach to calculate spill and spill cost in networks are presented. This chapter also shows the actual implications of leg-dependence effects on fleet assignment decisions in both the deterministic and stochastic case, and shows ignoring leg-dependence can lead to sub-optimal fleet solutions.

Basic Model Assumptions

Before advancing to the detailed discussion, let us state the basic modeling assumptions that we will make during or analysis:

1. OD demands are independent,

2. If connecting\(^1\) (multiple leg) OD demand is constrained (or censored) along its path, we will attribute the entire spill of that OD demand to the flight leg that, among the flight legs of the OD demand’s path, becomes full first (and censors demand first) during the booking process, and

\(^1\) In this chapter we will refer to all multiple leg OD demands as connecting demand.
3. OD demands book uniformly over the booking period and the booking rate is proportional to the actual demand value. (This assumption will be relaxed in the later discussion.)

Assumption 2 helps us to resolve the problem of assigning the spill of a multiple-leg OD demand to the flight legs. If only one of the flight legs spill the connecting OD demand, the allocation of the spill to a flight leg is trivial. However, if more than one flight leg becomes full and spills, then it is not obvious to which flight leg to assign the spill and to what extent. Based on the argument that the tightest constraint on the OD passenger flow is established by the leg that becomes full first, we assign the entire spill to that flight leg. This way the double-counting and overestimating spill for the entire network is also avoided.

In the rest of the dissertation, demand will be considered to be deterministic or stochastic. When deterministic demand is considered, the adjective deterministic refers to the deterministic unconstrained total number of the booking requests that will arrive during the booking period. Deterministic demand does not imply an instantaneous booking process: As stated in Assumption 3, we will presume that bookings arrive over a booking period (of length $T$) with a uniform (deterministic) booking rate. The booking rate of the demand is proportional to the actual demand value, (i.e., if the OD demand value is $D$, then the booking rate is $D/T$.) Note that by assuming uniform proportional booking rates, the OD demand mix of the already arrived booking request on a flight leg at any time $t \in T$, will be proportional to the unconstrained OD demand values of demands traversing the flight leg. Similarly, at any time $t$, the OD mix of the remaining demand requests that will arrive until the end of the booking process will also be proportional to the OD demand values. For example, if a local $(y)$ and a connecting $(x)$ OD demand traverse a flight leg, then the ratio of the already arrived booking request at any time $t$ will be $y/x$. We will often refer to the above implication of the assumed uniform booking rates as "proportionality" assumption.

When demand is considered to be stochastic, the uniform booking rate (and proportionality) assumption will still prevail. In our model stochasticity refers to the randomness in demand and not to the randomness in the booking process. We assume that after the stochastic demand values for the booking process (or period) are obtained, the booking rates of the particular OD demands will be uniform and proportional to the actual demand values. With this interpretation the proportional fill assumption prevails in the case of stochastic demands also.

Note that Assumption 3 involves a representation of booking pattern (booking rate) of the OD demand. The booking pattern, which defines the sequence in which bookings arrive for a flight leg, is an important attribute for determining the load and spill for the flight leg. In later chapters we will consider more general booking patterns, and there we will relax Assumption 3.
4.1 Network Effects (Leg-dependence) in Deterministic Networks

In this section we outline the basic fundamentals of leg-dependence effects (that was discussed also in [29] with respect to estimating OD passenger flows) and further show its implications on spill estimates and on fleet assignment solutions. We will analyze the deterministic case in detail, because we believe that it helps to understand the approach that will be used later in the analysis of network effect under the stochastic case.

4.1.1 Traditional Fleet Assignment Models: Leg-Independent Approach

Fleet assignment decisions are leg based decisions, i.e., fleet types must be assigned to flight legs. This requires a *leg based focus* with respect to the objective coefficients and the decision variables. Figure 4.1 shows the underlying concept of the traditional fleet assignment models. To keep the analysis simple, we have chosen a small network example with one fare class only for each OD and assume that demand is deterministic.

Fleet Types:  
F1 Cap=80 seats  
F2 Cap=100 seats

Figure 4.1: Network example with demand represented as local only (traditional representation)

Unfortunately, traditional fleet assignment models assume and require that airline demand be represented in a form of leg-based “local” demands. Therefore, for fleet assignment purposes unconstrained airline demand is given and forecasted in an aggregated form for each flight leg. Traditional methods use this unconstrained aggregation of demand for spill estimation. Figure 4.1 represented this approach. For each flight leg between a connecting city-pair OD an *unconstrained demand estimate* \(D_{od}\) is given (or forecasted). In this case, for leg 1 the given unconstrained leg demand is 100 passengers (pax) and for leg 2 the unconstrained leg demand is 110 passengers. Assume that there are two fleet types available for both legs: An 80 (fleets type *F1*) and 100 (fleets type *F2*) seat aircraft, (see Figure 4.1). Using the leg-independent approach of the traditional methods, the number of spilled passengers for each fleet type and for each flight leg are the following: For flight leg 1 and fleet type *F1* the spill is 20 passengers, for fleet type *F2* it is 0 passengers; for flight leg 2 and fleet type *F1* the spill is 30 passengers, and for fleet type *F2* it is 10 passengers.
Traditional methods would use these deterministic spill estimates to obtain the spill costs and assign the optimal fleet capacities accordingly. Note again that for each fleet type and flight leg combination a single objective coefficient is estimated and provided for the fleet assignment models. Thus this approach, as we already described in section 2.1, is a leg-independent approach. (That is, the cost coefficients corresponding to a flight leg are independent of the actual fleet type assignments of other legs).

The network characteristics shown in Figure 4.1 are such that demands can be characterized as “local”; therefore, a set capacity limit (fleet type assignment) on a flight leg would have no effect on the demand conditions, and thus, on the spill cost estimates of the other leg. Under these network characteristics, the traditional leg-independent spill estimation and fleet assignment approach is judicious and sound.

What, then, is wrong with the leg-independent approach? It is this: Aggregating and representing all of the demand of an airline network as involving one leg only, does not realistically characterize a typical airline network. As we have noted earlier, a large percentage of airline demand today is connecting demand, i.e., the OD demand’s path includes more than one flight leg. Therefore, Figure 4.2 is more representative of a typical airline network.

Fleet Types: F1 Cap=80 seats  
F2 Cap=100 seats

\[ D_{AC} = 50 \text{ pax} \]
\[ D_{AB} = 50 \text{ pax} \]
\[ D_{BC} = 60 \text{ pax} \]

Unconstrained Leg Demand: 100 pax 110 pax

Figure 4.2: Airline network with OD demands represented as local and connecting

By definition, the total unconstrained leg demand is equal to the sum of local and connecting OD demands on each leg. This is the demand that would be accommodated on each leg if seat capacity limits did not constrain bookings. From this figure, however, it is easy to see that if a capacity constraint associated with a fleet type assignment on a flight leg constrains the flow of demand, then because of the multiple leg characteristics of connecting OD demand path, the capacity assignment can influence the demand-related spill and spill cost estimates of the other flight leg as well. That is, the connecting AC OD demand can be affected by the capacity limits of both legs. Therefore, the connecting OD demand ties the fleeting decisions of the two legs together.

\[ \text{In practice spill costs usually calculated as the product of the spill and the estimated spill fare, which is the estimated average fare of the spilled passengers.}\]
Let us assume, for example, that the assigned capacity limits and the underlying booking patterns of the OD demands are such that flight leg 1 becomes full first during the booking process. Then there is no available capacity to accommodate more local and connecting demand on that leg, and we say that the flight leg, or the corresponding capacity, censors or constrains the connecting AC demand. The censoring results in spill. If the connecting demand is censored on leg 1 then the censored part of that cannot appear on the leg 2 either -- even if there is available capacity on leg 2 to accommodate those passengers. The censoring of the connecting demand on leg 1 changes the demand conditions of leg 2. Even if the capacity on leg 2 were unlimited, the part of connecting OD demand that can appear on leg 2 is less than the unconstrained connecting OD demand. We call this part of the unconstrained OD demand as achievable OD traffic on leg 2. The achievable OD traffic is the maximum OD demand that can be accommodated on a leg assuming that no capacity limit set on the leg, but considering the capacity limits set on other legs of the network (on leg 1 in our case). The achievable traffic for a leg is the sum of the local unconstrained OD demand and the achievable OD traffics. The achievable traffic is what should be considered to evaluate the effects of the capacity limit set on the second leg -- instead of the unconstrained leg demand. Clearly, the load and spill estimates on the second leg will not only be a function of the capacity limits set on the second but also on the first leg.

Spill costs, in this example, can be calculated by multiplying the spill by the weighted average fare weighted by the OD demand means. If we assume that the AB fare is $30, the AC fare is $90, and the BC fare is $70, and considering the proportionality assumption (Assumption 3.), the spill and spill costs associated with assigning each fleet type to each leg are given in Table 4.1. For example, for flight leg 2 and fleet type F1 the spill cost was calculated as:

$$spillcost = spill \times \frac{fare_{AC} \times D_{AC} + fare_{BC} \times D_{BC}}{D_{AC} + D_{BC}} = 10 \times \frac{90 \times 50 + 70 \times 60}{50 + 60} = 790.9$$

Other spill costs estimates were calculated similarly.

<table>
<thead>
<tr>
<th></th>
<th>Leg 1</th>
<th>Leg 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fleet Type</td>
<td>Spill</td>
<td>Spill Cost</td>
</tr>
<tr>
<td>F1 (Cap=80)</td>
<td>20 pax</td>
<td>$1200</td>
</tr>
<tr>
<td>F2 (Cap=100)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4.1: Leg-independent spill and spill cost estimates for each fleet type and flight leg

The above table suggests that leg-independent spill estimates, when used in a network where connecting OD demands are present, may presume "infeasible" passenger flows. Consider the case when F1 is
assigned to flight leg 1 and F2 assigned to flight leg 2. The corresponding spill estimates are 20 and 10 passengers. However, if 20 passengers are spilled on leg 1 then, because of the proportionality assumption, 10 of these passengers are connecting AC passengers. In actual fact, on leg 2 the estimated spill of the connecting AC passengers is \((50/(50+60))\times10=4.54\) passengers, not 10 (although they refer both to the same AC demand). On leg 2 the spill estimate is based on the unconstrained leg demand and does not account for the connecting OD passengers spilled on leg 1. The actual demand conditions on leg 2 are overestimated and the achievable traffic principle is ignored when leg-independence is incorrectly assumed.

The above invalid passenger flow controversy and leg-independent assumption result in overestimating the (achievable) demand conditions and may imply that both the load and spill estimates for certain flight legs of the network are double-counted and thus are overestimated.

Summarizing, we could see that the connecting OD passenger flows tie the flight legs together and introduce interdependence. The demand-related load and spill estimates of each flight strongly depend on the supply-related capacity assignment decisions of the other flight legs. We called this phenomenon leg-dependence. We have shown that leg-independence assumption in the traditional fleet assignment methods can result in "infeasible" passenger flows and incorrect spill estimates. The assumed invalid passenger flows may create double counting or overestimation in the load (network revenue) and in spill (spill cost calculations). The following shows why.

### 4.1.2 Spill Calculation Considering Leg-Dependence

Before proceeding further, recall that we assume that if connecting demand is spilled, we will attribute the entire spill of a connecting OD demand to the flight leg that becomes full first along the path of the demand.

Now let us assume that first, fleet type F2 is assigned to flight leg 1. No passengers, and thus no connecting passengers, would be spilled on leg 1, therefore, if fleet type F1 were assigned to flight leg 2, a total of 30 passengers would be spilled on the second leg. Leg-dependence in this case does not affect the spill calculations. According to the proportional fill assumption, the 30 spilled passengers will be brought up by the original demand mean ratios \((A_{Dc}/D_{Br}=50:60)\). Thus 13.63 AC passengers and 16.37 BC passengers would be spilled. If fleet type F2 were assigned to leg 2, then the total spill of 10 passengers would be broken up similarly: 4.54 AC passengers and 5.46 BC passengers would be spilled.

Now suppose that fleet type F1, with capacity of 80 seats, is assigned to flight leg 1. This would result in 20 passengers spilled on leg 1. Because the AB and AC OD demand means are both 50 passengers, 10 AB

\[\text{3 Since we work with expected values and average numbers over a longer planning period non-integer spill numbers are acceptable.}\]
and 10 AC passengers would be spilled. The spill cost on this leg then would be $1200. Due to the spill
of 10 connecting AC passengers on leg 1 the connecting AC achievable OD traffic on leg 2 is 40
passengers. The achievable traffic on leg 2 equals the sum of the 60 local BC demand and the censored
40 AC achievable OD traffic. Using the achievable traffic estimates instead of the unconstrained leg
demands, if fleet type F2 (100 seats) is assigned to leg 2, then leg 2 would spill no passengers at all.
Consequently, for this assignment the spill cost on leg 2 is $0. This result differs from that of the leg-
independent traditional method, which estimates always a spill of 10 passengers, and a spill cost of
$790.9, when fleet type F2 is assigned to leg 2.

Now assume that while fleet type F1 remains assigned to leg 1, fleet type F1 with its 80 seats is assigned
to leg 2 as well. What would be the spill in this case? Could we again assume that due to the censoring
on leg 1 the achievable traffic conditions are same on leg 2? If we assume uniform booking rates and that
aircraft capacities are filled up proportionally until censoring does take place, then the answer is no.
Because the total booking rate is higher on leg 2 than on leg 1, leg 2 would actually reach its capacity first
and only then will be leg 1 full. The leg with the higher total booking rate to capacity ratio reaches full
capacity first. The booking rate to capacity ratio is actually linearly proportional to the total
unconstrained leg demand per capacity ratio, which is identical to the definition of leg demand factor
(DF), frequently used in airline practices. Thus the DF value indicates which leg will reach capacity first
in the booking process. The demand factors of leg 1 and 2, assuming the assignment of fleet type F1 to
both legs, are 100/80=1.25 and 110/80=1.375, respectively. The higher demand factor of leg 2 indicates
that this leg would be full first, and thus this leg censors the connecting AC demand and not vice versa.

Taking this into account, we have to start our analysis with a reversed sequence, in the case when fleet
type F1 is assigned to both legs. We start with leg 2, which fills up first, and calculate the achievable
traffic conditions for leg 1. The spill cost for fleet type 1 on leg 2 is similar to the leg-independent spill
cost estimate, that is $2372.6. Since leg 2 is the leg that reaches capacity first, assuming proportionality
we can determine that 5/11 of the total 30 spilled passengers, that is 13.63 passengers, will be spilled AC
passengers. The achievable traffic on leg 1 then equals the sum of the AC achievable OD traffic, i.e., 50-
13.63=36.37, and the local AB demand. This equals 86.37 passengers. The result is that fleet type F1 on
leg 1 will spill 86.37-80=6.37 passengers. We know that all of these 6.37 passengers will be local AB
passengers, since we have already assumed that 36.37 AC passengers already have been accommodated.
Had this not been the case, flight leg 1 would have been the first leg to be full which according to the
demand factors is not the case. Therefore, the spill cost on leg 1 is $191.

Table 4.2 summarizes the leg-dependent spill and spill cost estimates for each leg under the different
fleeting scenarios.

---

4 The demand factor is defined as $DF = \frac{\text{unconstrained}_\text{leg}_\text{demand}}{\text{capacity}}$. 

57
<table>
<thead>
<tr>
<th></th>
<th>Fleet Type and Capacity on leg 1</th>
<th>Fleet Type and Capacity on leg 2</th>
<th>Leg 1</th>
<th>Leg 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>F1 (80)</td>
<td>F1 (80)</td>
<td>6.37  pax</td>
<td>$191</td>
</tr>
<tr>
<td>II</td>
<td>F1: (80)</td>
<td>F2: (100)</td>
<td>20</td>
<td>1200</td>
</tr>
<tr>
<td>III</td>
<td>F2 (100)</td>
<td>F1 (80)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IV</td>
<td>F2 (100)</td>
<td>F2 (100)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4.2: Leg-dependent spill and spill costs on each leg at different fleeting combinations

We can observe that the leg-dependent spill and spill cost estimates are equal to or smaller than the leg-independent estimates. The leg-independent approach, because it uses always unconstrained leg demands for the analysis, on certain legs (on legs that do not become full first) overestimates not only the spill and spill costs, but alternatively also the load estimates and expected revenues. (This is true because by not taking into account the censoring of connecting OD demands, the actually achievable traffic on the flight legs are overestimated, which must be the basis of load and spill estimation.)

It is important to note, that the overestimation is not uniform. That is, under different fleeting combinations the level of overestimation is different. This is a very important observation with respect to the fleet assignment solution. Fleet assignment compares the costs of different fleeting combinations. If the estimation error due to the network effects were constant, then only the total costs of a fleeting would change but the relative rank of the fleeting would not. In such a case, even though the leg-independent approach would use incorrect spill costs, the fleeting solution would be the same as the solution of the leg-dependent approach. However, since the overestimation is not uniform, the bias may change the actual rank of the different fleeting costs and may have an important effect on the actual fleet assignment solution.

Note that in the above example it was possible that under different capacity conditions the sequence of the legs becoming full might differ as well. In this case, the achievable traffic and spill estimates on flight legs may differ. Therefore, determining the sequence of the flight legs becoming full is an important issue. Since the sequence is in practice determined by the complex relationship of the demands relative to capacity and the booking patterns of the different OD demands, all of these parameters play an important role in determining the correct spill estimates.

A consequence of the above is that at certain capacity intervals the leg-dependent spill values for the network are not a convex function of the capacity. It is possible under certain capacity combinations, that
reducing the capacity on a flight leg will cause the total spill in the network to decrease. While between other capacity ranges, reducing the capacity result in an increase in the total spill. As we will show in Chapter 8, assuming only leg-based control over the ODF passenger flows and a certain fare consistency among the OD demands, the leg-based fleet assignment formulation has a non-convex hull. This would preclude the use of the traditional linear programming formulations to find a truly optimal solution. (For a detailed discussion refer to section 8.2.)

Note that leg-dependence not only affects the spill costs but also the spill fare\(^5\). The basis of the leg-independent spill fare calculation is always the set of the unconstrained OD demands. The spill fare is the weighted combination of the OD demand fares, weighted by the OD mix of spill. The OD mix of spill, since it is based on unconstrained demand, is proportional to the unconstrained OD demand values. However, this is not always the case for the leg-independent estimates -- it is the case only for the leg that spills first. The leg that does not spill first will spill only part of the achievable traffic that can appear on the flight leg. Therefore, the OD mix of spill changes and so does the spill fare on the flight leg. If the differences among the different OD demand fares are large, the magnitude of the differences in the spill fares and spill costs may be large as well.

Note, however, that the spill fare can be affected bi-directionally by leg-dependence. For a moment, consider a simple three leg example (see Figure 4.3), and for simplicity, assume that only two connecting demands and one local demand are present. Also assume that the two connecting fares differ substantially. According to the traditional leg-independent approach, the spill cost on leg 1 is the weighted average of the OD fares, weighted by the OD demand proportions. The leg-dependent spill fare on the other hand, can be larger or smaller than the leg-independent spill fare, depending on which flight leg, leg 1 or 2, becomes full first at the selected capacity assignments. Therefore, the differences between the leg-dependent and leg-independent spill fares can be bi-directional.

![Figure 4.3: Three-leg network example with differing connecting OD fare values](image)

\(^5\) Spill fare was defined as the average fare of the spilled passengers.
4.1.3 Fleeting Solutions Using Leg-independent and Leg-dependent Estimates

Next, let us determine the optimal fleet assignment solution for our two-leg network example, shown in Figure 4.2, under different operating constraints using the leg-dependent and leg-independent spill estimates. Let us assume that leg 1 of the network is a short-haul and leg 2 is a longer-haul segment. Let us also assume that, as is usual in practice, the operating cost of a larger aircraft is greater than that of the smaller aircraft flying on a certain leg. The assumed operating costs associated with flying the two flight legs are the following:

<table>
<thead>
<tr>
<th></th>
<th>Leg 1</th>
<th>Leg 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fleet-1</td>
<td>$2000</td>
<td>$4000</td>
</tr>
<tr>
<td>Fleet-2</td>
<td>$3000</td>
<td>$5500</td>
</tr>
</tbody>
</table>

Table 4.3: Operating costs of the different fleet types

The optimal fleetting, or fleet assignment solution, to the network is the one for which the total costs, i.e., the total operating and spill costs are at the minimized -- subject to various constraints. For the moment let us assume that both fleet types can be assigned to both flight legs, and determine the optimal fleetting solutions using leg-independent and leg-dependent estimates.

The total cost of a fleetting is equal to the sum of the operating costs and the spill costs corresponding to the actual fleet type assignment on each flight leg of the network. Using the leg-independent spill cost estimates in Table 4.1, the leg-independent total fleetting costs are shown in Table 4.4. For example, the total fleetting cost associated with fleetting I, is calculated as (2000+1200)+(4000+2372.6) = $9572.6.

As the results of the table show, the fleetting that comprises the assignment of fleet type F2 to both leg 1 and 2, has the lowest total costs ($9290.9). Therefore, if no additional constraints are present, i.e. both aircraft types can be assigned to both flight legs, then the fleet assignment solution based on the leg-independent estimates would result in assigning fleet type F2 to both legs.

<table>
<thead>
<tr>
<th>Fleeting</th>
<th>Fleet Type and (Capacity) on leg 1</th>
<th>Fleet Type and (Capacity) on leg 2</th>
<th>Fleetting Cost</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>F1 (80)</td>
<td>F1 (80)</td>
<td>$9572.6</td>
<td>4</td>
</tr>
<tr>
<td>II</td>
<td>F1 (80)</td>
<td>F2 (100)</td>
<td>$9490.9</td>
<td>3</td>
</tr>
<tr>
<td>III</td>
<td>F2 (100)</td>
<td>F1 (80)</td>
<td>$9372.6</td>
<td>2</td>
</tr>
<tr>
<td>IV</td>
<td>F2 (100)</td>
<td>F2 (100)</td>
<td>$9290.9</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4.4: Total cost for fleetting combinations considering leg-independent spill cost estimates
If we take the network effects into consideration, and use the leg-dependent spill cost estimates for fleet assignment, then, the results will be different. Let us again assume that there are no additional constraints, that is, that each aircraft type can be assigned to both legs. The total costs of a fleeting combination again are calculated as the sum of the operating costs and spill costs. However, in this case, leg-dependent spill cost estimates are considered, (see Table 4.2). Table 4.5 presents the total costs associated with the different fleeting combinations. For example, the total fleeting cost associated with fleeting combination I, is calculated as \((191+2000)+(2372.6+4000)=\$8563.6\). The presented results show that some of the leg-dependent total fleeting costs differ from the leg-independent costs. This is particularly the case for fleeting combinations I and II. For these cases the differences are substantial.

The leg-dependent costs are less than the leg-independent costs: $8563.6 versus $9572.6 for fleeting I, and $8700 versus $9490.9 for fleeting II. In the leg-independent cases, ignoring the network effects and double counting of spilled passengers is what generates the overestimation of the spill and total costs. For fleeting III and IV, the total costs are similar in the leg-independent and leg-dependent cases. In these fleeting combinations, fleet type F2 with its larger 100 seat capacity is assigned to flight leg 1, and this capacity assignment absorbs even the total leg demand and thus spill does not occur.

<table>
<thead>
<tr>
<th>Fleeting</th>
<th>Fleet Type and (Capacity) on leg 1</th>
<th>Fleet Type and (Capacity) on leg 2</th>
<th>Fleeting Cost</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>F1 (80)</td>
<td>F1 (80)</td>
<td>$8563.6</td>
<td>1</td>
</tr>
<tr>
<td>II</td>
<td>F1: (80)</td>
<td>F2: (100)</td>
<td>$8700</td>
<td>2</td>
</tr>
<tr>
<td>III</td>
<td>F2 (100)</td>
<td>F1 (80)</td>
<td>$9372.6</td>
<td>4</td>
</tr>
<tr>
<td>IV</td>
<td>F2 (100)</td>
<td>F2 (100)</td>
<td>$9290.9</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 4.5: Total cost for a fleeting combinations considering leg-dependent spill cost estimates

Given the leg-dependent total fleeting costs, the best assignment (the fleeting combination with the lowest total costs) corresponds to fleeting combination I (see Table 4.5). That is, it corresponds to the case where fleet type F1 with its 80 seats is assigned to both flight legs. Remember that according to the leg-independent estimates, fleeting combination IV seemed to be the best assignment. There, it seemed to be worth assigning the larger but more expensive fleet type F2 aircraft to both flight legs in order to reduce the high spill costs. However, when network effects are considered, and spilled passengers are not double counted on both legs, then the best result would be to spill more passengers and save on operating costs.

The table also shows that even fleeting combination II would give better results than IV, when leg-dependence is considered.
For estimating the error of using leg-independent estimates and choosing the incorrect fleeting combination IV over I, we have to compare the corresponding leg-dependent total fleeting costs for the two assignments. The additional costs that we take on with the wrong selection equals to the difference with the total leg-dependent fleeting costs, that is, \$9290.9 - \$8563.6 = \$697.3. This is an 8.1\% increase in total incurred costs. That is, using the leg-dependent spill cost estimates over the leg-independent ones for fleet assignment decision, in this network example case, would result in substantial total savings.

Next, let us consider other cases when some additional constraints are introduced in the fleet assignment decisions. First, let us consider the case when the requirement is that both flight legs must be operated by the same aircraft types. This requirement might be the case for certain connecting or “through flights”, where for marketing or operational reasons the same aircraft must fly two connecting flight legs. Therefore, in our example, fleeting combinations I and IV are the possible choices. Since these combinations were the best choices of the leg-independent and leg-dependent cases when all constraints were relaxed, the differences in the best fleeting and the occurring errors are similar to the above case.

Second, let us consider the case when due to operating constraints the same aircraft type cannot be operated on both flight legs. There, fleeting combinations II and III are of concern. In this case, the leg-independent and leg-dependent result would imply different fleeting selections as well. The leg-independent solution would suggest that fleeting combination III is better than II by (\$9490.9-\$9372.6 =) \$118.3, while the leg-dependent approach shows that it is the other way, and fleeting combination II gives lower total costs than III by (\$9372.6-\$8700 =) \$672.6.

Another possible constraint could be that, due to operating or scheduling reasons, fleet type F1 must be assigned to flight leg 1. This limits our consideration to fleeting combinations I and II only. The optimal fleeting selections would differ in this case as well. As Tables 4.4 and 4.5 show, that the best fleeting combination according to the leg-independent approach is fleeting II, while according to the leg-dependent approach is fleeting I.

If the constraint is that fleet type F2 must be assigned to flight leg 1, (fleeting combinations III and IV considered), then the two approaches would result in the same solution. However, from the tables we can also see that had the constraint been that a fleet type is fixed on flight leg 2, the two different approaches would result in different solutions.

To summarize, we have shown that it is possible that leg-dependence effects have a substantial consequence on the optimal fleeting solutions. When network effects are considered, then the optimal fleeting can be different from the result of the leg-independent approach. We demonstrated even in a very simple two-leg example that these differences can arise under typical and reasonable network demand and cost parameters. We speculate, that in real networks, the probability that a leg-dependent solutions will be
substantially different from the leg-independent one is also high. In these cases the benefit of taking leg-dependence into consideration could be substantial.

4.2 Network Effects and Leg-dependence (Stochastic Demand Case)

So far in our example we assumed that demand is given a deterministic form. In the following we will move forward and address the case when demand is stochastic.

4.2.1 Calculating the Achievable Traffic Density

Before starting our discussion, let us recall our modeling interpretation of stochastic demand:

Stochasticity refers to the randomness in demand and not to the randomness in the booking process. We assume that after the stochastic demand values for the booking period are obtained, the booking rate of the particular OD demands will be uniform and proportional to the actual demand values. Therefore, given the random OD demand values for the booking period, the booking process becomes deterministic.

First let us concentrate on determining the achievable traffic distribution. Assume again that our network is a simple two-leg network (see Figure 4.2). For the moment, assume that only a connecting OD demand, is defined for the network. Let assume that \( f_{\text{conn}}(y) \) represents the probability density function (pdf) of the connecting OD demand. Let us also assume that the capacity on flight leg 2 is \( \text{Cap}_2 \), and no capacity limit is assigned to leg 1. We are interested in the achievable traffic density for leg 1, given the capacity limit on leg 2. For this case it is relatively easy to determine the achievable traffic density \( f_T(w) \).

\[
    f_T(w) = f_{\text{conn}}(y) \quad \text{if} \quad w < \text{Cap}_2
\]

\[
    f_T(w) = \int_{\text{Cap}_2}^{w} f_{\text{conn}}(y) dy \quad \text{if} \quad w = \text{Cap}_2 \tag{4.1}
\]

\[
    f_T(w) = 0 \quad \text{otherwise.}
\]

Figure 4.4 shows the shape of the censored achievable traffic density. The curve below \( \text{Cap}_2 \) is identical to the unconstrained distribution, but at \( \text{Cap}_2 \) it is censored and has its peak. This curve represents the connecting achievable OD traffic on leg 1, given the capacity limit on leg 2. That is, had the capacity limit on leg 1 been infinitely large, this curve would have been observed as the traffic on leg 1.
Figure 4.4: Censored probability distribution -- Capacity limit censoring only one OD demand

Let us now follow with the case where the capacity limit on leg 2 censors not only the connecting OD demand, but simultaneously another independent local OD demand as well. In this case both demands are competing for the available capacity on leg 2. As we will show, the random interaction of the local and connecting demands on leg 2 will change substantially the shape and characteristics of the censored achievable OD traffic curve, compared to the curve shown in Figure 4.4. (It is very important to mention once again that we assume that bookings are uniform over the booking period and proportional to the actual random demand values. This uniform booking rate assumption determines the actual sequence of the different OD demand arrivals, which also determines the part of OD demands that are accepted or spilled.)

As Figure 4.5 shows, let us assume that one connecting and two local OD demands are present, and OD demands are independent. For the moment assume that the capacity assigned to flight leg 2 is $Cap_2$, and no capacity limit is assigned yet to flight leg 1. We are interested in the achievable traffic density for leg 1. In order to obtain that, we have to first determine the probability density function of the connecting achievable OD traffic, $f_C(w)$, i.e., the pdf of the connecting demand that can appear on leg 1 given the capacity limit and demand parameters on leg 2. The $f_C(w)$ function represents the part of the connecting OD traffic that may appear on leg 1 given the capacity limit $Cap_2$ and the independent unconstrained OD demands, $f_{local1}(y)$ and $f_{local2}(x)$, on leg 2.

![Diagram](image)

Figure 4.5: Network example with probabilistic OD demands
Note that for any possible value of \( w \) (achievable OD traffic on leg 1) a large number of different combinations of \( x \) and \( y \) (connecting and local leg 2 random deviates) may exist. Because of the proportionality assumption, we can also write that, if \( x \) and \( y \) are the random variables of the AC connecting and BC local demands, respectively, then:

\[
w = \frac{x}{x + y} \text{Cap}_2
\]  

(4.2)

from where:

\[
y = \frac{x(\text{Cap}_2 - w)}{w}.
\]  

(4.3)

If the local and connecting demands are independent, then the probability density function of the achievable traffic of the connecting OD demand can be written as:

\[
f_T(w) = f_{\text{conn}}(w) \int_{y=0}^{\text{Cap}_2-w} f_{\text{local}2}(y)dy + \int_{x=w}^{\text{Cap}_2} f_{\text{conn}}(x)f_{\text{local}2}(x(\text{Cap}_2 - w)/w)dx \quad \text{for } 0 < w < \text{Cap}_2
\]  

(4.4a)

\[
f_T(w) = \int_{x=\text{Cap}_2}^{\text{Cap}_2} f_{\text{conn}}(x)f_{\text{local}2}(0)dx \quad \text{for } w = \text{Cap}_2
\]  

(4.4b)

\[
f_T(w) = 0 \quad \text{otherwise.}
\]  

(4.4c)

Equation (4.4a) refers to the case when the connecting achievable OD traffic is less than the capacity. This would happen under two circumstances: 1) when the connecting demand is less than the capacity and no spill occurs (i.e., the local demand is smaller than the remaining capacity), and 2) when spill does occur (we will later be interested especially in this term when we address the calculation of leg-dependent spill). The first term of equation (4.4a) represents the former and the second part represents the latter case. Equation (4.4b) refers to the case when the connecting achievable OD traffic is equal to capacity. This may happen only in cases when the connecting demand is equal or greater than the capacity and at the same time the local demand is 0. Note that at any value of the local demand other than zero the connecting achievable OD traffic must be less than the capacity. (Intuitively, if the local demand is also present, then the capacity on leg 2 will be shared by the two demands. Hence, the connecting achievable OD traffic cannot be as big as the capacity).

Unfortunately, the calculation of the above achievable traffic pdf is very cumbersome. We used Monte Carlo simulation to estimate the pdf of achievable traffic and to analyze the shape of the curve under different conditions. The simulation results show that the pdf of the achievable traffic density will differ
from the original normal connecting demand pdf, which is also apparent from the above equations. In general, the more connecting demand is censored, the larger the differences will be. If the capacity is large on the censoring leg, i.e., no significant spill occurs, then the integral over the \( f_{\text{local}}(y) \) curve in the first term of equation (4.4a) will approach 1, and the second term of the same equation will approximate 0. This would imply that if no spill occurs then the achievable traffic curve is identical to the underlying connecting demand density. On the other hand, if significant spill occurs, then the integral over the \( f_{\text{local}}(y) \) curve in the first term of equation (4.4a) will be less than 1 (the tighter the capacity the smaller the integral will be), and the more the second term of the equation will modify the value of \( f_T(w) \). In this case, those values of \( w \) will be modified by the proportional fill the most, for which all the possible combinations of the local and connecting demands (that result in \( w \) connecting achievable traffic) may happen with high probabilities. Note also that as equation (4.4b) shows the peak at the capacity limit may happen only in the absence of the local demand. If the probability associated with zero local traffic is low, then there is a very low chance that we can observe a peak at the capacity limit, even in high spill cases.

In our simulation we assumed that the both the connecting and local demands are distributed normally and are independent. The mean and standard deviation of the connecting demands were set to 60 and 20 passengers, respectively. The mean and standard deviation of the local demand were set to 40 and 13 passengers. We then varied the capacity of flight leg 2.

We could observe that as the capacity on flight leg 2 becomes smaller, that is, as the spill increased, the pdf of the achievable OD traffic moved to the left. Figure 4.6 shows the AC achievable OD traffic densities for leg 1 at capacity 120, 100, and 80 on leg 2, respectively. At high capacity, where relatively little spill occurs, the achievable traffic curve resembles more the underlying connecting normal demand density (mean=60, std=20). We can also observe, though, that as the spill increases, the curves become more negatively skewed -- that is, the “longer tail” is in the negative direction, and probabilities are shifted to the lower demand values. Note also that the mode (the point with the highest probability) of the curves is equal to the point associated with the proportional share of the connecting demand of the capacity proportional to the demand means -- that is, in our case: \( \frac{60}{(60+40)} \times \text{Cap} \). For the three cases of our example, the modes are at 0.6*120=72, 0.6*100=60, and 0.6*80=48. We can observe the high probabilities at each of these curves. Although the modes of the curves can be found at the demand levels given above, because of the skewed shape of the curves the mean of the curves can be found at lower demand values, and as the spill increases the standard deviation of the curves also decrease. The means for the three curves are 58.14, 53.84, and 45.96, respectively. The corresponding standard deviations are 17.47, 13.78, and 9.42, respectively.
Figure 4.6: Achievable OD traffic densities on leg 1 for the case when capacity limits censor local and connecting demands (simulated curves)
Figure 4.7: Achievable OD traffic density for the case when capacity limits censor local and connecting demands -- reduced standard deviation for local demand (capacity on leg 2 =100 and local std=5)

We can make another observation about how the standard deviation of the local demand influences the shape and parameters of the achievable traffic curve. Figure 4.7 shows the case when the standard deviation of the local demand was reduced to 5 passengers, (the capacity on leg 2 remained at 100 seats). We can observe that the distribution is more skewed in this case than it was on Figure 4.6b (where we assumed that the standard deviation of the local demand is 20 passengers). This is because when the standard deviation of the local demand is smaller, there is less randomness in the capacity available for the connecting OD demand. This implies that the connecting demand is censored at certain high values with a higher probability, which will result in a more censored right hand side on the pdf of the achievable traffic curve. In our case, the increased probabilities we can observe at demands above 55 passengers. The mean and standard deviation of the curve are 53.11 and 12.76, respectively. These are both smaller than the parameters of the curve associated with Figure 4.6b. We could observe this phenomenon generally at other capacity values and at other demand parameters as well.

Now we can derive some observations about the shape of the achievable traffic curves:

- The greater the censoring, the more the achievable traffic pdf is shifted to the left relative to the unconstrained demand.

- When spill occurs, the mode of the achievable traffic pdf is at the point which represents the proportional share of the connecting demand of the capacity proportional to the demand means (or achievable traffic modes), that is at

\[
\frac{\mu_{\text{conn}}}{\mu_{\text{conn}} + \mu_{\text{local}}} \times \text{Cap}.
\]
• The shape of the achievable traffic curve has a negatively skewed shape. The smaller the variance of the local demand, the more skewed the achievable traffic curve.

• The greater the censoring effect on leg 2 and the more skewed the achievable traffic curve on leg 1, the smaller the mean and variance of the achievable traffic curve.

• In reasonable demand cases, it is very unlikely that we can observe a distinct spike at the right hand side of the achievable traffic curve.

• The achievable traffic density is a function of the interactions of the OD demands on the censoring leg. Consequently, the shapes of the curves are very different when no local OD demand was assumed (Figure 4.4) and when local and connecting OD demands are censored by the capacity limit simultaneously (Figures 4.6 and 4.7).

After calculating and estimating the connecting achievable OD traffic we can obtain the achievable traffic for leg 1, which is the convolution sum of the local OD demand on leg 1 and the connecting achievable OD traffic density function (see Figure 4.8). Because the two functions are independent, we can write that $f_1^T(i)$, the total achievable traffic for leg 1 is:

$$f_1^T(i) = \int_{x=0}^{\infty} f_{\text{local}1}(x)f_T(i-x)dx,$$  \hspace{1cm} (4.5)

where $f_{\text{local}1}(x)$ represents the local demand on flight leg 1, and $f_T(.)$ represents the connecting achievable traffic censored on leg 2.

To calculate the load or spill estimates for leg 1 we have to proceed carefully. We can use the above achievable traffic function as a basis for the calculation only if the assigned capacity limit on leg 1 is relatively large. In this case, we may correctly assume that leg 2 always becomes full before leg 1 and leg 2 censors the connecting OD demand. For example this, would be the case when the capacity limit on leg 1 is 150 seats and on leg 2 is 100 seats. Figure 4.8b shows for this case the estimated achievable traffic density for leg 1 and the unconstrained leg demand density. The latter is the convoluted sum of the unconstrained connecting OD demand and the unconstrained local OD demand on leg 1.
Figure 4.8: Achievable traffic densities on leg 1 (Ach_Traffic) and the leg-independent (Unconstr. Dem.) leg demand
From Figure 4.8 it becomes apparent that the expected load estimate based on the unconstrained leg demand density is larger than the expected load estimate based on the achievable traffic density.

- Thus, since the achievable traffic density is shifted negatively to relative to the unconstrained demand density, it is apparent that the load and spill estimates, that use the achievable traffic density as a basis, will be less than the leg-independent load and spill estimates.

Note that in this situation, since the assumed capacity on leg 1 is large, no spill occurs on leg 1.

4.2.2 Calculating Spill and Spill Costs
To calculate the load and spill estimates at lower capacities on leg 1, where spill may occur, we have to proceed carefully. We have to take into consideration, that at smaller capacity values, it is possible that with a certain probability leg 1 may become full before leg 2. In this case leg 2 will censor the connecting OD demand and not the other way; to prevent double counting the spill of the connecting demand for these cases should be allocated to leg 1. Which leg becomes full first, however, depend on the actual demand levels, which are random.

Therefore, to calculate spill, the probabilities associated with the achievable traffic density and the unconstrained leg demand density (see Figure 4.8) should be weighted by probabilities of the sequence of the legs to become full. Therefore, a weighted linear combination represents the actual curve that should be used for spill estimation. Although this could be also a method to obtain the leg-dependent spill formula for leg 1, we have chosen a different but similar approach, which is easier to understand. This approach uses the basic arguments we have introduced in defining the achievable OD traffic density calculation, presented in equation (4.4).

Remember that spill for a leg can be calculated by the following formula:

\[
Spill = \int_{Cap} f(i)(i - Cap)di, \quad (4.6)
\]

where \( f(i) \) is the probability density function of the leg demand and \( Cap \) is the capacity limit on the flight leg. Let us concentrate on calculating the spill on leg 1 if the assumed capacity on leg 1 is \( Cap_1 \) and on leg 2 is \( Cap_2 \). Flight leg 1 can potentially spill passengers if the total unconstrained leg demand requests are greater than capacity. That is, if the sum of the connecting OD demand, \( x \), and local OD demand on leg 1, \( l_i \), are greater than the capacity:

\[
i = x + l_i > Cap_1,
\]

where \( i \) random variable represents the total unconstrained leg demand on leg 1. The number of \( i \) unconstrained booking requests on leg 1 may result from a large number of different demand
combinations of the local and connecting OD demands. The probability that there will be \( i \) unconstrained booking request on leg 1, \( f(i) \), assuming independence between the OD demands, is:

\[
f(i) = \int_{x=0}^{i} f_{\text{conn}}(x) f_{\text{loc}}(i-x) \, dx, \quad (4.7)
\]

where \( f_{\text{conn}}() \) and \( f_{\text{loc}}() \) are the connecting and local unconstrained OD demand density functions.

The traditional leg-independent spill estimate, \( \text{Spill}^{\text{L-I}} \), for leg 1 then can be calculated as:

\[
\text{Spill}^{\text{L-I}} = \int_{x=0}^{\min(i, \text{Cap_1})} \int_{i=0}^{\text{Cap_1}} f_{\text{conn}}(x) f_{\text{loc}}(i-x) \, di \, dx. \quad (4.8)
\]

To proceed with determining the leg-dependent spill formula for leg 1, there are two cases to be identified.

In Case 1 we model the conditions when leg 1 becomes full first in the booking process, and in Case 2 when leg 2 becomes full first\(^6\).

**Case 1:** In this case leg 1 becomes full before leg 2. Therefore, leg 1 censors the connecting demand and the spill of the local and the connecting OD demands will be attributed entirely to leg 1. The condition for leg 1 to become full before leg 2, is that the demand to capacity ratio (demand factors) is greater for leg 1 than for leg 2. If \( i \) is the total unconstrained demand on leg \( l \) and \( k \) is the total unconstrained demand on leg 2, then the necessary condition is:

\[
\frac{i}{\text{Cap_1}} \geq \frac{k}{\text{Cap_2}}. \quad (4.9)
\]

Note that \( i \) and \( k \) are not independent. Because of the leg-dependence they are strongly correlated. Therefore, we should rather write that:

\[
\frac{i}{\text{Cap_1}} = \frac{x + l_1}{\text{Cap_1}} \geq \frac{x + l_2}{\text{Cap_2}} = \frac{k}{\text{Cap_2}}, \quad (4.10)
\]

where \( x, l_1, \) and \( l_2 \) are the random variables of the connecting, local leg 1, and local leg 2 unconstrained OD demands, respectively.

There are two different cases that we have to distinguish: The case when \( \text{Cap_2} > \text{Cap_1} \) and the case when \( \text{Cap_2} < \text{Cap_1} \). For the former case, we can write that leg 1 becomes full before leg 2 if:

\[
\frac{(x + l_1) \text{Cap_2}}{\text{Cap_1}} - x = \frac{i \text{Cap_2}}{\text{Cap_1}} - x \geq l_2, \quad \text{where} \quad 0 < x < i. \quad (4.11)
\]

\(^6\) When both get full at the same time, which would be a very rare situation, we can choose either case arbitrarily.
We can determine the conditional probability that flight leg 1 becomes full before flight leg 2, \( P[l \rightarrow 2l_i, x] \), conditional to that the total unconstrained demand on leg 1 is \( i \), and of this \( x \) is connecting OD demand:

\[
P[1 \rightarrow 2l_i, x] = \int_{j=0}^{(i/Cap1)Cap2-x} f_{loc2}(j) dj \quad \text{if } Cap2 > Cap1 \text{ and } 0 < x < i. \tag{4.12}
\]

For the case when \( Cap2 < Cap1 \), note that if \( x \) is larger than \( (i/Cap1)Cap2 \), then at any value of \( l_2 \) leg 2 will spill first. If \( x \) is less than \( (i/Cap1)Cap2 \) then the probability that flight leg 1 becomes full first can be calculated as in (4.12). Consequently, the probability that leg 1 becomes full first when \( Cap2 < Cap1 \) is:

\[
P[1 \rightarrow 2l_i, x] = \int_{j=0}^{(i/Cap1)Cap2-x} f_{loc2}(j) dj \quad \text{if } Cap2 < Cap1 \text{ and } 0 < x < (i/Cap1)Cap2.  
\]

**Case 2:** In this case leg 2 becomes full before leg 1. Similar to case 1, we can determine the conditional probability that flight leg 2 becomes full before flight leg 1, \( P[2 \rightarrow 1l_i, x] \), if the total unconstrained demand on leg 1 is \( i \), and of this \( x \) is the connecting OD demand:

\[
P[2 \rightarrow 1l_i, x] = 1 - P[1 \rightarrow 2l_i, x] = \int_{j=(i/Cap1)Cap2-x}^{\infty} f_{loc2}(j) dj \quad \text{if } Cap2 > Cap1, \quad 0 < x < i. \tag{4.13}
\]

The above is valid for the case when \( Cap2 > Cap1 \) and \( 0 < x < i \), and when \( Cap2 < Cap1 \) and \( 0 < x < (i/Cap1)Cap2 \). In the latter case if \( x > (i/Cap1)Cap2 \) then leg 2 becomes full at any value of \( l_2 \). If leg 2 becomes full first, then according to our modeling assumption (Assumption 2) leg 2 will be accounted for the spill of the entire connecting OD demand. Leg 1 in these cases can only attributed for spilling local demand. The local demand that will be spilled is a function of the remaining capacity on leg 1. The remaining capacity is the original capacity minus the connecting achievable OD traffic. Note that if leg 2 becomes full before leg 1, then the part of connecting demand that is accommodated on leg 2, the connecting achievable OD traffic, \( w \), will be also accommodated on leg 1. (Had this not been the case, leg 1 would have been full first). The number of spilled local OD passengers can be calculated as the number of unconstrained local OD requests on leg 1, \( l_i \), minus the remaining capacity (see Figure 4.9), (or zero if the remaining capacity is greater than the local unconstrained demand):

\[
\max[l_i - (Cap_i - w); 0]. \tag{4.14}
\]
The value of \( w \) is a function of the actual OD demand mix on leg 2. If the actual local OD demand deviate on leg 2, \( j \), and the unconstrained connecting OD demand deviate, \( x \), is given, then because of the uniform booking rate and proportionality assumption, \( w \) can be calculated as:

\[
w = \frac{x}{x + j} \text{Cap}_2.
\]  
(4.15)

Note that the value of \( j \) can run from \((\nu \text{Cap}_2)\text{Cap}_1\) to infinity and therefore, the value of \( w \) is not constant either.

We have assumed that the total unconstrained leg demand on leg 1 is \( i \). If \( x \) is the connecting leg demand, then \((i-x)\) must be the local OD demand on leg 1. If we substitute this and equation (4.15) into equation (4.14) and rearrange terms we get that the spill of local demand is equal to:

\[
\max[i - \text{Cap}_1, -(x - \frac{x}{x + j} \text{Cap}_2), 0]
\]  
(4.16)

Figure 4.9: Calculation of local spill on leg 1 -- (\( l_i \) and \( x \) is the unconstrained local and connecting OD demand, and \( w \) is the connecting achievable OD traffic on leg 1)

Now let us first consider the case when \( \text{Cap}_2 > \text{Cap}_1 \):

Based on the above discussions, we can write the conditional expectation of the number of spilled passengers on leg 1 given that the capacity is \( \text{Cap}_1 \) and that the total unconstrained leg demand requests on leg 1 is \( i \):

\[
E[\text{Spill}_i] = \int_{x=0}^{i} f_{\text{conn}}(x) f_{\text{loc1}}(i-x) \int_{y=0}^{(i/\text{Cap}_1)\text{Cap}_2-x} f_{\text{loc2}}(y) \left[ i - \text{Cap}_1 \right] dy dx + \\
+ \int_{x=0}^{i} f_{\text{conn}}(x) f_{\text{loc1}}(i-x) \int_{j=(i/\text{Cap}_1)\text{Cap}_2-x}^{\infty} f_{\text{loc2}}(j) \max[i - \text{Cap}_1, -(x - \frac{x}{x + j} \text{Cap}_2), 0] dj dx
\]
And after rearranging:

\[
E[\text{Spill}_i] = \int_{x=0}^{i \frac{\text{Cap}_2}{\text{Cap}_1}} f_{\text{con}}(x) f_{\text{loc}_1}(i - x) \left\{ \int_{y=0}^{(i \frac{\text{Cap}_2}{\text{Cap}_1}) \frac{x}{x + j} \text{Cap}_2} f_{\text{loc}_2}(y) * [i - \text{Cap}_1] dy + \right. \\
+ \left. \int_{j=(i \frac{\text{Cap}_2}{\text{Cap}_1}) \frac{x}{x + j} \text{Cap}_2} f_{\text{loc}_2}(j) * \max[i - \text{Cap}_1 - (x - \frac{x}{x + j} \text{Cap}_2), 0] dj \right\} dx 
\] (4.17)

The first and second terms in the above equation represent Case 1 and Case 2, respectively. In each term the multipliers in the square brackets represent the actual number of spilled passengers under the corresponding conditions.

Next let us consider the case when \( \text{Cap}_2 < \text{Cap}_1 \):

The calculation here is similar, but we have to take into consideration that at any value of \( x \) that is greater than \( (i \frac{\text{Cap}_1}{\text{Cap}_2}) \) leg 2 will be full before leg 1, regardless of the local demand level on leg 2. Thus the conditional expectation of the number of spilled passengers on leg 1 given that the capacity is \( \text{Cap}_1 \) and that the total unconstrained leg demand requests on leg 1 is \( i \) \( (\text{Cap}_2 < \text{Cap}_1) \):

\[
E[\text{Spill}_i] = \int_{x=0}^{(i \frac{\text{Cap}_1}{\text{Cap}_2}) \frac{x}{x + j} \text{Cap}_1} f_{\text{con}}(x) f_{\text{loc}_1}(i - x) \int_{y=0}^{(i \frac{\text{Cap}_1}{\text{Cap}_2}) \frac{x}{x + j} \text{Cap}_2} f_{\text{loc}_2}(y) * [i - \text{Cap}_1] dy dx + \\
+ \int_{x=(i \frac{\text{Cap}_1}{\text{Cap}_2}) \frac{x}{x + j} \text{Cap}_2} f_{\text{con}}(x) f_{\text{loc}_1}(i - x) \int_{j=(i \frac{\text{Cap}_1}{\text{Cap}_2}) \frac{x}{x + j} \text{Cap}_2} f_{\text{loc}_2}(j) * \max[i - \text{Cap}_1 - (x - \frac{x}{x + j} \text{Cap}_2), 0] dj dx + \\
+ \int_{x=(i \frac{\text{Cap}_1}{\text{Cap}_2}) \frac{x}{x + j} \text{Cap}_2} f_{\text{con}}(x) f_{\text{loc}_1}(i - x) \int_{j=(i \frac{\text{Cap}_1}{\text{Cap}_2}) \frac{x}{x + j} \text{Cap}_2} f_{\text{loc}_2}(j) * \max[i - \text{Cap}_1 - (x - \frac{x}{x + j} \text{Cap}_2), 0] dj dx
\]

and after rearranging:

\[
E[\text{Spill}_i] = \int_{x=0}^{(i \frac{\text{Cap}_1}{\text{Cap}_2}) \frac{x}{x + j} \text{Cap}_1} f_{\text{con}}(x) f_{\text{loc}_1}(i - x) \left\{ \int_{y=0}^{(i \frac{\text{Cap}_1}{\text{Cap}_2}) \frac{x}{x + j} \text{Cap}_2} f_{\text{loc}_2}(y) * [i - \text{Cap}_1] dy + \right. \\
+ \left. \int_{j=(i \frac{\text{Cap}_1}{\text{Cap}_2}) \frac{x}{x + j} \text{Cap}_2} f_{\text{loc}_2}(j) * \max[i - \text{Cap}_1 - (x - \frac{x}{x + j} \text{Cap}_2), 0] dj \right\} dx + \\
+ \int_{x=(i \frac{\text{Cap}_1}{\text{Cap}_2}) \frac{x}{x + j} \text{Cap}_2} f_{\text{con}}(x) f_{\text{loc}_1}(i - x) \int_{j=(i \frac{\text{Cap}_1}{\text{Cap}_2}) \frac{x}{x + j} \text{Cap}_2} f_{\text{loc}_2}(j) * \max[i - \text{Cap}_1 - (x - \frac{x}{x + j} \text{Cap}_2), 0] dj dx
\]

To calculate the leg-dependent spill for leg 1 we have to integrate over all values of total demand \( i \) above the capacity \( \text{Cap}_1 \). Thus, the leg-dependent spill for leg 1 , \( \text{Spill}_{1-D} \), when \( \text{Cap}_2 > \text{Cap}_1 \) can be calculated as:
\[
\text{Spill}_{l-D}^{1} = \int_{i=\text{Cap}_{1}}^{\text{Cap}} \left\{ \int_{x=0}^{1} f_{\text{conn}}(x) f_{\text{loc1}}(i-x) dx \right\} \left( \int_{y=0}^{\text{Cap}_{2}-x} f_{\text{loc2}}(y) \cdot [i - \text{Cap}_{1}] dy + \int_{y=0}^{\text{Cap}_{2}-x} f_{\text{loc2}}(y) \cdot [i - \text{Cap}_{1}] dy \right) \left. \right|_{j=(i/\text{Cap})\text{Cap}_{2}-x} \right\}
\]

\[
\text{Part I.}
\]

\[
\text{Part II.}
\]

\[
\text{Part III.}
\]

The leg-dependent spill for leg 1, \(\text{Spill}_{l-D}^{1}\), when \(\text{Cap}_{2}<\text{Cap}_{1}\) can be calculated as:

\[
\text{Spill}_{l-D}^{1} = \int_{i=\text{Cap}_{1}}^{\text{Cap}} \left\{ \int_{x=0}^{\text{Cap}_{2}} f_{\text{conn}}(x) f_{\text{loc1}}(i-x) dx \right\} \left( \int_{y=0}^{\text{Cap}_{2}-x} f_{\text{loc2}}(y) \cdot [i - \text{Cap}_{1}] dy + \int_{y=0}^{\text{Cap}_{2}-x} f_{\text{loc2}}(y) \cdot [i - \text{Cap}_{1}] dy \right) \left. \right|_{j=(i/\text{Cap})\text{Cap}_{2}-x} \right\}
\]

\[
\text{Part I}
\]

\[
\text{Part II}
\]

\[
\text{Part IIIa}
\]

\[
\text{Part IIIb}
\]

In the above equations, Part I represents the combination of the local and connecting demands that add up to \(i\) on leg 1; Part II represents the Case 1 (when leg 1 becomes full before leg 2); and, Part III represents Case 2 (when leg 2 becomes full before leg 1).

The number of spilled passengers in Part II, \((i - \text{Cap}_{1})\), is always greater than 0 (since \(i\) runs from \(\text{Cap}_{1}\) to infinity), and always greater than the number of spilled passengers in Part III. This is true because substituting the minimum value of \(j\), (consider the case when \(\text{Cap}_{2}>\text{Cap}_{1}\), \((i/\text{Cap})\text{Cap}_{2}-x\), into

\[
\frac{x}{x+j}\text{Cap}_{2},
\]

and after rearranging terms we get:
\[
\frac{Cap_i}{i^x}.
\]  
(4.19)

Since as we noted already \((i-Cap_1)>0, (Cap_1/i)\) is always less than 1 and thus the above equation is also less than \(x\). If the value of \(j\) is larger than its minimum value by \(dj\), then after substitution into (4.19a) we get:

\[
\frac{1}{(i/Cap_1) + (dj/Cap_2)^x},
\]  
(4.20)

which is also less than \(x\) (all dividends are positive and \(i/Cap_1>1\)). Consequently, at all possible values of \(i\) and \(x\),

\[
i - Cap_1 \geq \max\{i - Cap_1 - \left(x - \frac{x}{x + j}Cap_2\right), 0\}.
\]  
(4.21)

The above observation implies that if at any values of \(i\) and \(x\), the conditional probabilities that leg 2 becomes full before leg 1, \(P[2 \rightarrow 1 \mid i, x]\), are greater than 0, then the leg-dependent spill estimate (equation 4.18) will estimate less spill for a flight leg than the leg-independent (traditional) spill estimate (see equation (4.8)). (The analysis with the case when \(Cap_2=Cap_1\) would lead to similar results).

Consequently, the leg-independent spill estimate of a leg \(l\), in cases when the probability that a leg other than \(l\) can be full before leg \(l\) and censors connecting demand is larger than zero, the traditional leg-independent approach overestimates spill.

It is easy to extend the formula given in equation (4.18) and formulate the leg-dependent spill costs for leg 1. The spill costs are calculated by multiplying the spilled OD demand with its corresponding fare value. Note that for Case 1, where the spill is equal to \((i-Cap_1)\), the OD mix of spill is also proportional to the actual unconstrained OD demand deviates. That is, \((x/i)(i-Cap_1)\) connecting OD demands and \((i-x/i)(i-Cap_1)\) local OD demands will be spilled, given that the total leg demand is \(i\) and of this \(x\) is connecting OD demand request. Therefore, the leg-dependent spill cost, for leg 1, \(SC_{L-D}^I\), when \(Cap_2>Cap_1\), can be calculated as:

---

\(^7\) The two equations are equal only when \(i=Cap_1\) and \(j\) is at its minimum value.
\[ SC_{L-D}^{i} = \int_{i=Cap_1} f_{conn}(x)f_{loc1}(i-x) \]
\[ \left\{ \begin{array}{l}
\text{Part I.} \\
\int_{y=0}^{(i-Cap_1)Cap_2-x} f_{loc2}(y) \left[ (i-Cap_1) \left( \frac{x}{i} fare_{conn} + \frac{i-x}{i} fare_{loc1} \right) \right] dy + \\
\text{Part II.} \\
\int_{j=(i-Cap_1)Cap_2-x} f_{loc2}(j) \max\left\{ (i-Cap_1)-(x-\frac{x}{x+j} Cap_2) fare_{loc1},0 \right\} dj dx \end{array} \right\} \]

where \( fare_{conn} \) and \( fare_{loc1} \) are the OD fare values associated with the connecting and local leg 1 OD demands.

The spill costs when \( Cap_2 < Cap_1 \) is:

\[ SC_{L-D}^{L-D} = \int_{i=Cap_1} f_{conn}(x)f_{loc1}(i-x) \]
\[ \left\{ \begin{array}{l}
\text{Part I} \\
\int_{y=0}^{(i-Cap_1)Cap_2-x} f_{loc2}(y) \left[ (i-Cap_1) \left( \frac{x}{i} fare_{conn} + \frac{i-x}{i} fare_{loc1} \right) \right] dy + \\
\text{Part II} \\
\int_{j=(i-Cap_1)Cap_2-x} f_{loc2}(j) \max\left\{ (i-Cap_1)-(x-\frac{x}{x+j} Cap_2) fare_{loc1},0 \right\} dj dx \end{array} \right\} \]
\[ \left\{ \begin{array}{l}
\text{Part IIIa} \\
\int_{j=0}^{i} f_{conn}(x)f_{loc1}(i-x) f_{loc2}(j) \max\left\{ (i-Cap_1)-(x-\frac{x}{x+j} Cap_2) fare_{loc1},0 \right\} dj dx \end{array} \right\} \]
Note again, that similar to case of spill, the spill cost estimate of the traditional approach for a flight leg overestimates the spill costs compared to the leg-dependent spill cost estimation approach. This difference can be magnified even more when we assume that the fare structure on a flight leg is so that the fare of the local OD demand is substantially less than that of the connecting OD demand. In this case the overestimation in spill is further magnified by the differences of the fare values\(^8\).

The consequence of the overestimated spill of the traditional approach can lead to an non-necessary assignment of a larger capacity aircraft, which could be used more profitably on another flight leg.

We have presented an analytical approach to calculate leg-dependent spill and spill costs. The above formulations were very complex and cumbersome to calculate even for a two-leg one fare class per OD demand case. In Chapter 6 we will present a more practical, Monte Carlo simulation based approach that can calculate leg-dependent spill and spill costs for larger and more realistic networks as well. The additional advantage of that method is that we can relax many of the basic modeling assumptions, and more complex and realistic attributes and environments of the booking process can be modeled (e.g., general booking patterns, effects of yield-management systems, etc.).

Note that estimating the expected load or the expected revenues for a flight leg or the network can be easily obtained by following the above modeling approach. These values can be very useful to estimate more accurately loads and expected revenues that a given fleet (capacity) assignment can obtain under the given demand characteristics.

### 4.3 The Interpretation of Spill in Networks

Spill is traditionally defined as the expected demand that goes unaccommodated because of insufficient capacity. Note that as we have seen, spill is not the main underlying process, but rather a by-product of the interactions of ODF demand requests their arrival over the booking period given the capacity limits in the airline network. Since spill refers to demand, theoretically it should be also defined on the same basis: on an origin-destination level. Fleet assignment decisions, on the other hand, because of the nature of the problem, are leg-based decisions. Therefore, the focus for spill estimates remains leg-based.

We have seen that in networks where all demands are accommodated on single-leg non-stop flights, the origin-destination demand corresponds directly to the flight legs -- in these cases the supply and demand networks directly overlap and are both leg based. That was the case in the assumed representation of the airline network demand in traditional fleet assignment and spill estimation models. In this leg-independent representation, it is assumed that if a flight leg capacity constrains demand, it will affect the

---

\(^{8}\) Although note that as we have shown for the deterministic case, the spill fare estimation bias can be bi-directional, which may affect the spill fare also bi-directionally.
passenger flows, and the spill estimates, of that leg only. In networks where demands are all local, the traditional estimation and interpretation of spill is straightforward and cause no modeling problems.

Not as forthright as above is the situation in networks where OD demand is accommodated on multiple leg paths. In this case, the OD demands do not correspond to the leg-based capacity decisions. The challenge is to find an approach that can deal with the conflicting nature of the leg-based supply decisions and the origin-destination path based spill concept, described in this chapter.

The main use of spill estimates is to help evaluate the effects of a capacity assignment or capacity change on a flight leg. Thus, a leg-based spill estimate should indicate the expected number of passengers that is spilled on the particular flight leg due to the assigned capacity limit on that leg. Only the spill of those passengers, that are being spilled by the capacity constraint of that leg should be attributed to a flight leg. To avoid double-counting, it is also important that the spill of a passenger is assigned to one flight leg only.

Spill on a flight leg must be interpreted as a manifestation of the complex interactions of the origin-destination based passenger flows and the effects of the capacity limits in the network. Keeping this in mind, a spill estimate of a flight leg should be a function of all the relevant capacity limits of the networks -- and not only a function of the capacity of the flight leg. This leg-dependent nature of spill alternatively also implies that, while evaluating the spill for a flight leg, all capacity limits on other legs should be taken into consideration and must be assumed as being fixed. Without this ceteris paribus assumption, leg-based spill estimates may be incorrect and misleading.

Using the achievable traffic concept presented earlier, spill can be decomposed and interpreted correctly at the flight leg levels. The achievable traffic approach incorporates the ceteris paribus assumption and also takes the leg-dependence effects into consideration. It also avoids the overestimation (or double-counting) of spill by using the leg achievable traffic estimate as opposed to the unconstrained leg demand.

Note that a leg-based spill estimate cannot be directly used in linear programming formulations in networks where multiple-leg ODF demands are present, and capacity limits are set on flight-leg (this issue is addressed in Chapter 8). Thus, in these networks we have to interpret any spill estimate as an estimate related to a set of network conditions, including the relevant flight legs of the sub-network it refers to, the demand conditions, and the set capacity limits. Spill in airline networks with multiple leg OD demands has to be always interpreted with this network perspective and the leg-based definition and interpretation of spill has to be abandoned.

4.4 Summary

Above we have shown that leg-independent estimates of spill may assume OD passenger flows that are in fact infeasible given existing capacity constraints. Particularly, leg-independent estimates do not take into
consideration that if a multiple leg OD passenger is spilled on a flight leg, then he/she cannot appear on other flight legs of the OD path. Therefore, due to double counting, leg-independent spill estimates on several flight legs of a network can overestimate spill and thus spill costs. We have also shown that deterministic and stochastic leg-independent estimates overestimate spill costs (and also load and revenue estimates) by different magnitudes -- i.e., the level of overestimation changes across the flight legs and across different fleeting combinations. This non-constant relative overestimation may result in the selection of a different sub-optimal fleeting solutions as compared to the solution where network effects, by using the leg-dependent estimates, are considered. Had the overestimation been uniform over all fleeting combinations (i.e., the difference between the leg-dependent and leg-independent spill costs had been a constant value), the relative ranks of the fleeting combinations would have not been changed. In this case, both the leg-dependent and the leg-independent approach would lead to similar fleet assignment solutions -- consequently, the leg-dependent approach, although using incorrect estimates, would still provide the correct fleet assignment solution. However, since the overestimation is not uniform (as the above simple network example has also shown) if network are considered effects by using leg-dependent estimates, the solution may differ from that of the leg-independent approach.

Using a simple two-leg example we have shown that network effects can substantially influence OD passenger flows, and the traditional leg-independent approach to spill estimation may lead to sub-optimal fleet assignment solutions, which may result in substantial losses of cost savings. Therefore, for globally optimal fleet assignment, leg-dependent spill estimates should direct the fleet assignment procedure.

We also noted that spill in airline networks, where multiple-leg ODF itineraries exist and capacity constraints are applied on flight-leg levels, should be always interpreted with a network visor; i.e., a spill estimate should be interpreted as an estimate that refers to a set of flight legs with a given demand characteristics and an assumed fixed capacity assignments on all of the flight legs in the set. That is, in these network a leg-based spill estimate should be always interpreted with a network perspective (ceteris paribus assumption) or the leg-based interpretation should be abandoned.
5. Attributes Influencing Leg-Dependent Effects

We have shown that in a typical airline network, leg-dependence can significantly influence the actual passenger flows on the flight legs. These influenced passenger flows affect the load and spill estimates, which in turn affect the optimal fleet assignment solutions. In the following, we will try to determine the attributes that cause leg-dependence and influence the significance of its effects. We would also like to identify under which set of network conditions the leg-independent assumption, i.e., a leg-independent formulation, would be incorrect and under which conditions it would give a sufficiently good solution. Hence, in our analysis we will often compare the results of the leg-dependent approach to the load and spill estimation against that of the leg-independent approach. In each of the following sections, we will identify likely factors and important demand and network characteristics that influence the leg-dependence effect and also its extent in the network.

5.1 Basis for leg-dependence: Connecting Passenger Ratio

The basis for leg-dependence is the existence of ODF demands with connecting (multiple-leg) itineraries (we will shortly refer to them as “connecting demand”). If all origin-destination demands were accommodated on non-stop flights, that is, if only local demand existed, then capacity constraints would affect only local demands and would have no effects on passenger flows of other legs. In this case the network would be leg-independent. Then, the leg-independent assumption of the traditional fleet assignment optimization methods would be theoretically correct and the obtained fleet assignment solution would be optimal.

However, flight leg demand often comprises local and connecting demands. If the percentage of the local passengers dominates the OD demand mix, then the leg-dependence in the network may be less consequential and its effects in certain cases may be ignored. Alternatively, if the share of connecting passengers is large, then we expect a more significant role for leg-dependence. The connecting passenger ratio may be an initial indicator for the existence and possible significance of leg-dependence issues in the network.

We define the connecting passenger ratio (CR) as the ratio of the connecting demand over the total leg demand. A small CR value indicates low connecting demands compared to the local demands, and a CR equal to 1 implies that all demand of a leg is connecting. The CR varies significantly over different airline networks and different markets. In a hub-and-spoke airline network, where the percentage of the connecting passengers is high, most of the CRs are relatively high as well -- compared to networks that do not operate under a hub-and-spoke structure. This is especially true for the US hub-and-spoke networks. In a US hub-and-spoke network the CRs typically vary between 0.3 to 0.8. The ratios may be lower in international hub-and-spoke networks, e.g., in European hubs.
The CR value in a network can vary significantly. It is possible that on average the network CR value is relatively low, although on certain legs of the network the CR values are high. In this case the leg-dependence effects in the sub-network, where the high CR is characteristic, may be very significant and should not be ignored. Therefore, not only the average values but the variance of the CR should be evaluated.

Although the CR value for the network or for a part of the network may indicate the significance of leg-dependence effects, by itself it cannot give a correct estimate about the significance of the leg-dependence effects. For that, other aspects and attributes of the network must also be considered.

5.2 Amount of Spill
Spill is prerequisite for leg-dependence to be a concern -- particularly the spill of connecting OD demands. If no connecting demand is spilled, leg-dependence will have no effect on the passenger flows of other flight legs. Further, the amount of spill determines the potential significance of the leg-dependence effect. The more connecting OD demands are spilled, the more extensively passenger flows of other flight legs could be altered, thus the more significant the leg-dependence effect may be. Therefore, more spill may indicate potentially more significant leg-dependence effects.

However, one has to be very careful with deriving any implications about the significance of leg-dependence effect solely from the level of spill. We will show that, even in cases when only very little spill occurs on each leg, due to network connectivity characteristics, leg-dependence effects may concentrate and add up on flight legs; making them significant in the aggregate. On the other hand, again due to network connectivity characteristics, it is also possible that when the spill of connecting OD demands is very high on certain legs, leg-dependence effects can be dispersed throughout the network and therefore become negligible. Hence, although the spill of connecting OD demands is a prerequisite for leg-dependence, the level of spill by itself cannot determine its significance. Although in general there is a positive correlation between spill and leg-dependence consequences, in order to assess the overall network-wide consequences of leg-dependence, additional network connectivity attributes must be considered.

5.3 Booking Patterns of Origin-Destination Demands
Booking patterns of the different OD demands may play an important role in the OD mix of spill. Imagine the situation when a capacity constraint, either in the form of total leg capacity or a booking limit, censors both local and connecting OD demands. These demands, representing different markets, may exhibit very different booking behaviors -- either because of the nature of the demand itself or because of different advance purchase requirements. Assume, for example, that the connecting demand tends to book more in advance and also has an advance purchase requirement, while the local demand has
no purchase requirements and tends to book late. The two different booking patterns (booking curves) are shown in Figure 5.1. This case we may correctly infer that, unless the total expected spill is very large, spill will involve mostly the later booking local passengers. A reversed booking pattern characteristic may result in a reversed OD mix of spill.

![Booking Curves](image)

Figure 5.1: Booking curves (Booking patterns) of a later booking (a) and earlier booking (b) OD demand

By affecting the OD mix of spill, the OD demand booking patterns can indirectly affect on the following:

- They determine the spill fare and the spill cost of the flight leg. In cases when the difference in the OD fares are large, different booking patterns may result in substantially different spill costs.

- By affecting the OD demand that becomes part of spill, they also determine the other flight legs of the network that can be affected by leg-dependence. For example, if the booking patterns of concern are such that the OD mix of spill involves mostly local passengers, then it will preclude leg-dependence. In this case, the spill will not result in changes to the achievable traffic estimates on other legs. (We will address this issue in more detail in the next chapter, where the mutual effects of leg-dependence and yield management systems on each other are analyzed.)

### 5.4 Network Connectivity -- Origin-destination demand distribution

The connectivity and distribution of OD demands in the network can affect the significance of the leg-dependence effects. It is possible that under certain conditions, censoring effects add up and concentrate on a flight leg. In this case, even if flight legs spill only a small proportion of the connecting demands, the overall effects may become very substantial. On the other hand it is also possible that the high spill of connecting demands is dispersed in the network and the effects on each of the influenced flight legs are
negligible. Thus, network connectivity and the direction of the effect propagation can strongly influence the consequences of leg-dependence.

5.4.1 Dispersion of Censoring Effects
Assume that the OD demands of leg \( i \) are fed from or distributed to many different upstream or downstream legs (see Figure 5.2). If leg \( i \) spills demand and involves all OD demands that traverse the leg equally (i.e., the connecting OD demands are balanced and have similar booking patterns), then the censoring effect due to the spill is evenly distributed on to many continuing downstream legs or to many feeding upstream legs. In both cases, even if the spill of connecting demands on leg \( i \) is significant, the censoring effects are scattered over many different OD flow paths. In this way, the censoring effect is allocated to different flight legs and its magnitude is thereby decreased. The greater the number of different OD demands involved in spill, the more leg-dependence effects are dispersed; thus, the less the influence of the spill on leg \( i \) will be on each of the feeding legs. Therefore, even if spill on leg \( i \) is very substantial, special attributes of network connectivity may render its actual leg-dependence effect on each of the relevant flight legs negligible.

![Diagram showing dispersion of effects](image)

Figure 5.2: Example for dispersion of the leg-dependence effects (Flight leg \( i \) spills)

This situation could happen in a number of cases. For example, this is the case when a transoceanic or transcontinental flight is fed by or feeds many connecting flight legs. This situation may also happen in a hub-and-spoke network, in a case when either one or a few of the incoming or departing flight legs encounter significant spill. Similar is the situation when only the flight leg that connects two hubs spills passengers.

To verify the above, we have constructed a simulation test-bed. The sample network structure constituted of a main leg \( i \) that are fed by \( n \) feeding legs. For clarity of analysis we have assumed that only
connecting demands are present in the network. That is, only passengers who are willing to travel from city A, B,...,N to city Z exist; see Figure 5.2a. We also assumed that demands are identically, independently and normally distributed. (The identical distributions ensure the balanced OD shares.) The sequence of the bookings in the simulation was allocated randomly. Therefore, on average, the equal share of OD demands in load and spill must prevail. We set the capacity of legs 1,...,n to infinity, and the capacity on leg i to Cap_i. Thus only leg i censored the OD demands and we could observe the achievable traffic on each of the n feeding legs. The number of iterations in the simulation was 4000.

The significance of the leg-dependence effect can be measured by the deviations between the simulated achievable traffic probability density function and the unconstrained leg demand density. For the achievable traffic density, due to the censoring effects taking place, we expect that it will be shifted to the left relative to the unconstrained demand density. This is because the probabilities associated with higher demand values are reduced because of the censoring effect taking place on leg i.

In the simulation the total unconstrained leg demand on leg i was 300 passengers. In order to experience relatively high spill, we set the capacity limit on this leg to 300. This corresponds to a demand factor of DF=1.0. We analyzed the scenarios assuming that the standard deviation of each OD demand is equal twice the square root of its mean. We analyzed different cases of n, the number of feeding legs involved. The OD demands were identical, and were set so that the total mean of the leg demand on leg i resulted in 300 passengers.

Figure 5.3 shows the expected traffic densities and the unconstrained demand densities for cases when the standard deviation of each OD demand is the square root of the mean and the number of legs are n=2,5, 10, and 15. Each figure shows the simulated achievable traffic on one of the n feeding legs (since the OD demand on each leg was identical, the achievable traffic on each leg was closely similar), and the uncensored connecting OD demand. Note that the uncensored (unconstrained) OD demand represents the demand that would be assumed under the leg-independent approach. The figures show our a priori assumption that as the number of legs, n, increases, the achievable traffic curve approaches more closely the original demand curve. This implies, that as n increases, the importance of the leg-dependence effect diminishes. Note also that when n is small, i.e., 2 or 5, the achievable traffic densities deviate from the unconstrained demand densities by a large extent. This implies that leg-dependence effects are significant and cannot be ignored at low values of n.

We are interested in the achievable traffic densities because, as we have shown in the previous Chapter, they indicate the difference between the leg-dependent and leg-independent spill estimates.
a) $n = 2$ feeding legs

b) $n = 5$ feeding legs
Figure 5.3: Simulated achievable traffic densities of a feeding leg and the unconstrained connecting demand densities

The above lead to the conclusion that if the number of feeding legs is large, (and spill from other legs of the network does not affect the feeding flight legs), then the leg-dependence effect due to the spill on leg \( i \) may be ignored. Note that this holds only if the OD demands are balanced. If this is not the case, for example if one of the OD demands represents proportionally a very high share, then the censoring effects are also dispersed proportionally. In this case, the leg-dependence effects on legs of lower proportional demand will be less while the legs with high proportional demand will experience a more significant leg-
dependence effect. In this case, even if the number of the feeding flight legs is big, leg-dependence could be ignored only on a certain number of the legs, while for the others must be considered.

5.4.2 Concentration of Censoring Effects

Very different is the situation when the censoring takes place on the feeding legs. This situation is shown on Figure 5.4. Here we assume that on each feeding leg a capacity limit spills the OD demand and the capacity limit on leg \( i \) is large enough for no spill to occur. In this case the leg-dependence effects on leg \( i \) add up, due to censoring on the feeding legs; this situation is the opposite of what happened in the previous case. Even if each of the feeding legs spills only a little demand, the cumulative leg-dependence effect on leg \( i \) can be very significant. This case models the situation when, despite the fact that the amount of spill is very small on each of the flight legs, the overall leg-dependence effect may be significant.

![Figure 5.4: Example of concentrating leg-dependent effects (legs 1..n spill and leg \( i \) has large capacity)](image)

This situation may be common in airline networks, for example, in cases when a trunk leg is fed by many flight legs or when a trunk leg distributes demand onto many continuing flights. Usually the connection of international flights to domestic networks constitutes this architecture. Similar cases can be observed in a hub and spoke network as well -- assuming the situation when many flight legs feed or distribute traffic to or from one flight leg, where all feeding legs censor demand.

If a capacity constraint on the feeding leg censors an OD flow, then the achievable OD traffic that can appear on leg \( i \) follows the shape of a censored distribution. (Figure 4.4 has shown a censored distribution curve.) The achievable traffic on leg \( i \) is the sum of the censored probability density functions of each of the feeding legs.
We present our results that we have obtained using Monte Carlo simulation. In the simulations we assumed that $n$ feeding legs with $n$ OD demands connecting to leg $i$. The capacity limit on leg $i$ was set to an infinitely large value, while the capacity limits on each of the feeding legs were set so that they would censor a part of the OD demands. Each OD demand was identically normally distributed and the capacity limits on each leg were set also identically. We were interested in the achievable traffic density on leg $i$. Small deviations from the unconstrained demand density on leg $i$ would suggest little, while large deviation would imply significant leg-dependence effects. We expect that as the number of the feeding legs increases, the leg-dependence effects add up and become more significant. We also expect that leg-dependent effects increase as spill on each leg increases. Our simulation results confirm these expectations.

Figure 5.5 shows the results of our simulations for $n=5$, $n=10$, and $n=15$, respectively. In these simulations each of the OD demands was identical and their means were equal to 10 passengers; the standard deviations were equal to the square root of the means, i.e., 3.16. On each of the feeding legs the capacity limits were set to 11 seats. This represents a demand factor of $DF=0.9$. Spill on each of the feeding legs then is equal to 0.8 passengers, which is less than a passenger. The figures show the simulated achievable traffic density, and the total unconstrained leg demand density for leg $i$ -- the latter curve would be assumed if leg-dependence effects were not considered. The figures show that as the number of legs becomes bigger, the censoring effects add up on leg $i$ and the achievable traffic density shifts more to the left. Note, that although as $n$ increases the achievable traffic is shifted more to the left, its shape approaches better the bell shape of the normal curve. Nevertheless, the differences in spill estimates are larger in the case of large $n$'s.
Figure 5.5: Simulated achievable traffic densities and the leg-independent total unconstrained leg demand densities for different, \( n \), number of feeding OD demands (feeding leg DF=0.9).

Apparently, if the spill of each OD demand increases, then the leg-dependence effect on leg \( i \) will be more significant. This implies that if the capacity limits on the feeding legs are tighter, then the leg-dependence effects become more significant as well. Also if the standard deviations of the OD demands are larger, then spill is larger, hence the leg-dependence effect also becomes more significant.
Summarizing, in the case when capacity limits censor the individual OD demands on the feeding legs, then the censoring effects add up on the leg that is being fed, and hence the leg-dependence effects may become significant even in cases when the spill on each of the feeding legs is small.

5.5 The Direction of Leg-Dependence Effect Propagation

So far we have analyzed cases when either the feeding legs spill passengers or the leg that is fed spills the demand. Although both of these cases may happen often in reality, more general cases, when all or many flight legs censor OD demands, are more common in practice. In these cases, the dispersion and the concentration of the censoring effects may take place simultaneously. Which effect dominates the other is a complex task to determine. If demand is deterministic, then the effects are deterministic, and thus the analysis is straightforward. In case of stochastic demand the analysis is more complex.

The major problem of modeling leg-dependence effects is that, unlike flight delays for example, leg-dependence effects do not have a unique propagation direction\(^1\). In case of leg-dependence, time does not limit the propagation of the leg-dependence effects. That is, a capacity constraint on a particular leg may have effects on flight legs that are scheduled in time either earlier, the same time, or later than the flight of consideration.

In order to determine the derived effects of leg-dependence in a more general situation, we need to understand which flight leg would have the dominant effect on the rest of the flight legs. In order to answer this question we will concentrate on determining the direction of the leg-dependence effect propagation. Recall that leg-dependence effects arise because one of the flight legs censors connecting passenger flows. Apparently the flight leg that actually censors the passenger flow will affect the achievable traffic on the other legs of the network and not vice versa. A flight leg spills connecting passenger flows if its capacity becomes full. As a rule for leg-dependence propagation, we can say that leg-dependence effects always propagate from the leg that becomes full earlier towards the leg that becomes full later in the booking process\(^2\).

The above rule makes it relatively easy to determine the leg-dependence effect propagation in the network, at least for deterministic demand. One needs only to determine the time at which flight legs will become full during the booking process. This sequence also determines the leg-dependence effects propagation. (The deterministic algorithm that calculates consistent OD passenger flows, presented in [29] is based on this similar argument as well.)

---

\(^1\) For example if a flight is delayed, then only flights that depart later are affected because of the late arrival of passengers or equipment. The flights that departed or landed before the flight that experience the delay are not affected. Time gives a unique direction for the effect propagation.

\(^2\) Here we assume that capacity limits are only the aircraft seat capacities assigned to each flight leg. Thus in this model booking limits do not affect the passenger flows.
When demand is stochastic it is more complex to determine which flight leg would become full and thus it is more complex to calculate the expected loads and spill in the network. Due to the random nature of demand, the specific leg that becomes full first is also stochastic. Although the demand factor in the case of stochastic demand could be also a useful indicator for evaluating when a flight leg becomes full, it does not give sufficient accurate information. The PFF (Probability of full flight) evaluated at capacity, i.e., the probability that demand is greater or equal to capacity, is a more appropriate measure for evaluating the leg that becomes full first. This can be calculated as:

\[
PFF \equiv \overline{P}(\text{Cap}) = \int_{\text{Cap}} f(i)di,
\]

(5.2)

where \( f(i) \) represents the total unconstrained leg probability density function for \( i \) requests on a flight leg. Calculating the fill rate for each leg of the network, we obtain the probabilities of each leg to be full first during the booking process. To calculate the flight leg that becomes full first next is a little bit harder to calculate. The PFF values cannot be used, because they are leg-independent measures.

The following procedure could be applicable for the stochastic case as well if applied in a Monte Carlo simulation setting. To calculate the expected sequence of the flight legs to become full, a larger number of demand parameters must be drawn iteratively from the OD demand probability distributions. In each iteration, the actual sequence of the filling can be calculated similarly to the deterministic case. By using a large number of iterations the expected sequence can be approximated by simulation. We will address this procedure in more detail in Chapter 6, where the Monte Carlo simulation based approach is discussed.

### 5.6 Boundaries of Leg-dependence Effect Propagation

In this section we would like to define the boundaries of leg-dependence effect propagation. It is important to know that if the capacity limit of a leg is modified, which legs of the network will be affected by this change. This analysis would give us an insight about the network connectivity with respect to leg-dependence and particularly would help us to determine the flight legs that may become affected and consequently need to be revisited due to leg-dependence effects in the NSFA algorithm that we present in Chapter 8.

Note again that leg-dependence exists only if multiple-leg itinerary OD paths are present in the system. A change in capacity limit on leg \( i \) has the potential to affect at least those legs of the network that share OD demand flows with flight leg \( i \). Yet, the effects on the actual passenger flows may propagate even further in the network. Overlapping multiple-leg OD demand flows may transfer the leg-dependence effects. Let us illustrate this situation on a small network example. Figure 5.6 shows a small three-leg network.
Assume for simplicity that all connecting multiple-leg path OD demands consist of a maximum of two legs. The figure shows all the assumed OD demands with the overlapping multiple-leg path demand flows on leg 2.

Figure 5.6: Network example for propagation analysis

Since connecting demands tie legs together, it is obvious that from the leg-dependence point of view flight leg 1 is connected to leg 2. (That is, if the capacity limit on leg 1 changes, it will affect the OD traffic flows of this leg and thus the achievable traffic on leg 2 as well.) Similarly, leg 2 is connected to leg 1 and leg 3 because of the connecting demands AC and BD. Further, leg 3 is connected to leg 2. The above relationships show the direct dependence of flight legs in the network. An even more interesting question is, whether leg 3 could be affected by a capacity decision on leg 1. In other words, can all legs be dependent on each other in this network?

The answer to this question is yes. It is possible that the a change in the capacity limit on flight leg 1 would not only affect the achievable traffic flows of flight leg 2 but also that of flight leg 3. Figure 5.6 illustrates this situation. As the figure shows, let us assume that the capacity on flight leg 2 is 130 seats and 140 seats on flight leg 3. Given the total leg demands (also shown on the figure), the corresponding demand factors (DF) for leg 2 and leg 3 are 1.23 and 1.142, respectively. Let us assume that the capacity we assign to flight leg 1 is 120 seats. That is, its demand factor is 1.25. We again adopt the proportional fill assumption. Given this assumption, the spill of the network can be calculated with the following approach. Since leg 1 is the leg with the highest demand factor, we know that this leg will be the first to become full during the booking process -- thus this leg is the bottle-neck of the network. Given the proportional fill assumption, the OD mix of load and spill on the first leg will be proportional to the OD demand means. That is, the 30 spilled passengers will be 20 local AB and 10 connecting AC passengers. Since leg 1 was the bottle-neck leg, we know that leg 2 cannot be full before leg 1, therefore the 50-10=40 AC passengers will be accommodated on leg 2 also.
In order to calculate the next leg that becomes full we have to calculate a new demand factor for leg 2 to replace the unconstrained connecting demand with the connecting achievable traffic. The achievable traffic on leg 2 is (40AC+50BC+60BD=) 150 passengers. Thus the new demand factor on flight leg 2 is \( \text{DF2}=1.153 \), which is bigger than the demand factor of leg 3 (\( \text{DF3}=1.142 \)). This implies that leg 2 will be full before leg 3. Therefore, we know that the remaining seat capacities that were not occupied by the AC demand flow, i.e., 130-40=90 seats, will be proportionally filled with local BC and connecting BD demands. Similarly, the OD mix of spill is also proportional. Therefore, 90*(50/(50+60))≈40.9 BC and 49.1 BD passenger will be accommodated. From the combined BC and BD demands, 20 passengers will be spilled -- the mix of this demand is also proportional, that is, 9.1 BC and 10.9 BD passengers will be spilled. Finally, on leg 3 we know that 49.1 BD passengers have to be accommodated. The rest of the capacity, or 140-49.1=90.9 seats, is available for the local CD demand. That is, the spill on that leg is 9.1 CD passengers.

Note that if the BD demand had not been censored by leg 2, then on leg 3 only 87.5 CD bookings would have been accepted or 12.5 CD passengers would have been spilled (because of the proportional fill assumption). Therefore, the OD mix of load and spill has changed by the censoring effect that has taken place on leg 2. However, the amount of the censoring effect on leg 2 was a function of the censoring effects on the AC demand flow, which is a function of the capacity limit set on leg 1. It implies that, the leg-dependence effect could propagate along all legs by the overlapping connecting OD traffic flows.

There are conditions when leg dependent effects do not propagate in the network. For example, leg-dependence effects propagate only over the path of legs that all spill. If a leg does not spill any passengers, then it serves as a buffer; leg-dependence effects do not propagate beyond that leg. Therefore, the leg that does not spill connecting passengers at all, defines a boundary to the effect propagation.

Another case leg-dependence effect is bounded, when along a path a flight leg exist that becomes full earlier than any of the other flight legs in the path. This would have been the case if in our example we had assigned a capacity 110 seats to flight leg 1. The highest demand factor in this case would still be that of leg 1. The leg would spill 13.3 connecting AC passengers, thus the achievable traffic on leg 2 would be (50-13.3)AC+50BC+60BD=) 146.66 passengers. Then the demand factor for leg 2 would be equal to \( \text{DF2}=1.12 \), which is smaller than \( \text{DF3}=1.14 \), the demand factor for leg 3. This implies that leg 3 will be full before leg 2. Remembering the rule that leg dependence effects always propagate from the leg that becomes full earlier towards the leg that becomes full later, we realize that leg 3 would have effect on leg 2 and not vice versa. This would also indicate that the leg-dependence effects emanating from leg 1 will stop at leg 2 and would not propagate beyond it.

We can summarize our findings about the conditions for leg-dependence propagation:
• Leg-dependence effects propagate on a path of overlapping demands only.

• Along the path of leg-dependence propagation each leg needs to spill (achievable traffic) in order for the effect to propagate.

• The relative demand factor on each leg establishes the direction of the propagation of leg-dependence effects. Leg-dependence effects always propagate from the leg with a higher demand factor towards the legs with lower demand factors.

• Because of the above, in airline networks the leg-dependence effect propagation due to a capacity change on a leg may be very much bounded and in such case the boundaries can be easily determined (at least in the deterministic demand case).

In a typical airline network one can find further distinct network boundaries for leg-dependence. For example, they can be found between legs that do not share any or any significant amount of the same OD demand flows. In a typical airline network, flights are scheduled so that the overlapping OD demands are clustered. This is the case in a hub-and-spoke structure. In a "connecting bank," flights are scheduled in a way that various OD demand flows find the connections attractive. In a connecting bank the OD demands overlap and connect the flight legs together. However, it is not typical to find OD demands that connect flights that belong to different connecting banks or complexes. Therefore, connecting OD demands can be characterized as being clustered in a connecting complex. This implies that the leg-dependence effects are bounded in the connecting complexes, i.e., since different connecting banks at the same hub do not share significant connecting OD flows, they are independent from each other. The only ties between two connecting banks are established by the flights connecting two hub cities. Flights within the connecting bank, however, can be strongly dependent on each other. Flight legs of a hub complex, even if they are scheduled to fly at the same time, may be dependent on each other. Figure 5.7 shows this situation. Although flights scheduled to fly at the same time (leg 1 and leg 3) are not connected directly by connecting OD flows, they can be connected indirectly by overlapping OD demand flows. Thus two inbound hub flights can be also dependent. They may be connected through an outbound hub flight (via leg 2). From the prospective of leg-dependence, Figure 5.7 and 5.6 are identical. The same rules and forces apply in both cases. This example also shows the time independent nature of leg-dependence propagation, i.e., that time dimension does not limit the leg-dependence effects to propagate.
There are natural boundaries of leg-dependence because of typical passenger travel behavior. For example, if a flight leaves from city A to city B then the next flight that leaves from city B to city A will very likely not carry the same passengers. Certain flights, because of scheduling or geographical reasons do not share common OD demands either. For example a flight leaving Boston for Washington DC is not expected to carry passengers whose next immediate connection would be the flight from Washington DC to New York. Thus the passenger's behavior and rational route, and path selection defines the actual passenger OD demand paths which at the same time define the connectivity of the network from the leg-dependence point of view.

We will analyze further these issues in the next chapters, and will use and exploit the above observations in Chapter 8, where we propose a heuristic fleet assignment algorithm (NSFA) that takes leg-dependence effects into consideration.

5.6.1 Leg-dependence Effect Propagation in Case of Stochastic Demand

In the above we have assumed that demand is deterministic. The analysis of leg-dependence propagation for stochastic demand is more complicated. Nevertheless, the above analysis gives us good insights and solid ground to approach this problem.

The random nature of demand complicates the analysis because it implies also that the leg of the network that first becomes the bottle-neck leg is also stochastic. Furthermore, the sequence of the rest of legs becoming full is also random. Nevertheless, many of the earlier observations hold for the stochastic case as well. For example, leg-dependence effects can propagate over a path of overlapping demands only. Further, along propagation's path all legs need to spill passengers -- otherwise no censoring effect would be caused that could serve as the basis for leg-dependence propagation. In the deterministic case, we used demand factors as indicators for the sequence of flight legs getting full. Determining the expected
sequence is important in the stochastic case as well, because the sequence always determines the direction of the leg-dependence propagation.

Let us start our analysis with the small network shown on Figure 5.8. The question we want to ask is whether leg 4 can be affected by the capacity limit set on leg 1, or more correctly stated: What is the probability that the capacity limit of leg 1 may affect the achievable traffic on leg 4? Remember that leg-dependence effects propagate along a path only if all legs spill. Therefore, the upper bound on the probability that the conditions on leg 4 are influenced by the capacity limit set on leg 1 is that legs 1, 2, and 3 all spill. If we assume that total leg demands OD demands are independent\(^3\), then the probability that all legs will spill is the product of the individual PFF values evaluated at the legs' capacities, that is:

\[
P = \prod_{l \in \text{path}} P_l(Cap_l),
\]

where \(Cap_l\) is capacity of leg \(l\), and \(P_l\) represents the PFF value based on the total leg demand of leg \(l\).

\[\text{Figure 5.8: Network example for boundaries of effect propagation -- stochastic demand}\]

We can see that the longer the path is, the smaller the probability that a capacity limit will propagate along the path. In addition, recall that the propagation of leg-dependence effects along the path has another requirement: The sequence of the legs becoming full during the booking process has to start at the source leg and must follow the paths. For our example it means that first leg 1, then leg 2 and finally leg 3 must become full. Any other sequence would limit the leg dependence propagation. For example, if leg 3 became full before leg 2, then leg 3 would affect leg 2 and not vice versa -- which would restrict the propagation of the leg-dependence effect emanating from leg 1. Therefore, in addition to the probabilities that all three legs become full, we have to take into consideration the probabilities that leg 1 becomes full before leg 2, and that leg 2 becomes full before leg 3. This will be the probability of a capacity limit set on leg 1 have the potential to affect leg 4.

We claim that the probability that a capacity limit on flight leg 1 can affect the achievable traffic on leg 4 is even smaller. We give some intuitive arguments for that. Above we have assumed that leg demands are independent. Note that even when there is independence among the different OD demands, this is not

---

\(^3\)Total independence is a simplifying assumption, because at least the overlapping demands make each pair of legs correlated. For the moment assume though independence.
the case between any neighboring flight legs. The overlapping OD demands introduce correlation between their probability density functions. This correlation is further extended by the leg-dependent effects, which in this case means that the connecting demand becomes censored. Thus, let us assume that leg 1 becomes full. If this is true, then that leg will censor a part of the connecting AC demand. This would mean that the achievable traffic on flight leg 2 is less then the total leg demand was, conditional that leg 1 become full first. If however, the conditional achievable traffic is less than the unconstrained leg demand, then the new PFF, which represents the probability that the next flight leg that becomes full is leg 2, will be less than the PFF that was calculated from the total unconstrained leg demand. Leg-dependence effect propagates along the path if leg 3 is the one which becomes full next. But part of the demand on leg 3 became censored on leg 2, resulting on smaller fill rates for leg 3 as well. Thus if we take the correlation and leg-dependence effects into consideration for calculating the probability for a leg-dependence effect to propagate along a path, the probabilities are even smaller than the product of the individual fill rates for each leg based on total leg demands. That is, the resulting probabilities must be bounded above by the probabilities calculated in (5.3).

If this is the case, then we can argue that the probability that a leg-dependence effect propagates along a path over more than 3-4 legs is very small because of the multiplication of the small probability values. In realistic airline situations, where we relax the assumption of maximum 2 leg itineraries, we can argue that we can ignore the effect of a capacity change on legs that are more than one leg beyond the legs that are directly effected by a censored OD demand. We refer to a leg being directly affected if it shares an OD demand with a leg that spills the OD demand. In our example, it would mean that we may assume that leg-dependence effect does not have effect beyond leg 3 if a capacity limit was changed on leg 1.

The basis for another intuitive argument for leg-dependence effects not to propagate far in the network is the proportionally diminishing size of the effects. Remember, that we assume that spill affects all OD demands proportionally. Therefore, if for example due to the change of capacity on leg 1 the difference in spill of AC passengers is x, then there will be x more seats becoming available to fill on leg 2. If then leg 2 still spills, then only a proportion of the BD demand will be able to use this x extra capacity. This also implies that the change that will affect leg 3 will be a proportional part of x. Therefore, as we go farther and farther, the size and magnitude of the capacity change on flight leg 1 diminish. Thus, not only the probability but also the expected magnitude of the leg-dependence effect diminishes along the path of legs.

\[\text{Note that here we assumed that a multiple-leg passenger itinerary includes maximum 2 legs. If there more legs can be included in an OD path then the propagation effects may propagate with a higher probability over a longer path.}\]

\[\text{Only in cases when no censoring effects from other legs add up, like the case when censoring of feeding flight legs add up and concentrate.}\]
The above imply the very important consequence that if a new capacity limit is assigned to flight leg 1, then we can limit our analysis of leg-dependence effects on passenger flows to the sub-network of legs 1, 2, and 3. In a more general case it would also mean that if we are decide to change the capacity, that is the fleet assignment, on one or few legs of the network, then the sub-network that we may consider as being affected, is bounded closely and can be easily determined. This would limit the size of the sub-network that needs to be examined to take leg dependence effects into consideration. As we will see later in Chapter 8, these characteristics will give us the basis for successfully applying a heuristic algorithm to solve the fleet assignment problem while acknowledging the leg-dependent nature of the network.

5.7 Summary

In this chapter, we have addressed under which demand and network conditions leg-dependence affects passenger flows and thus spill and spill cost estimates. These effects significantly influence the actual spill costs associated with assigning a certain fleet type to a flight leg.

We have also identified that the bases of leg-dependence are:

- The existence of connecting multiple leg path demands in the network, and
- The spill of the connecting multiple leg path passenger demands.

We have also shown that network connectivity plays an important role in the magnitude of the leg-dependence effects. If spill due to a bounding capacity constraint on a flight leg is dispersed onto many connecting flight legs, the actual leg-dependence effect on each influenced flight leg may diminish significantly. This represents a case when although the spill on a flight leg is high, its effect on the rest of the flight legs may become negligible. On the other hand, when spill occurs on flight legs that feed a connecting flight leg, censoring effects add up. In this case, the leg-dependence effect becomes very significant. This case represents the situation when, despite the fact that the amount of spill is very small on each of the spilling flight legs, the overall leg-dependence effect on the leg being fed is very significant.

Finally, we addressed the issue of the direction and boundaries of leg-dependence propagation. We showed that the essential conditions for leg-dependence propagation are overlapping OD demand paths, and the existence of spill on all flight legs along the propagation’s path. We have also argued that in typical stochastic airline networks, leg-dependence effect propagation due to a capacity limit change on a flight leg may be relatively bounded and the boundaries can be easily determined.
6. Approaches for Calculating Leg-dependent Load and Spill Estimates in Networks

It is very important for an airline to be able to calculate leg-dependent load and spill estimates that take into consideration network effects. Not only are they relevant for the purpose of fleet assignment decisions, they provide more realistic estimates for other planning purposes as well. Since load and spill estimates are the bases for calculating expected revenues or spill costs for the planning period, ignoring network effects may lead to invalid predictions and expectations in operational and financial analyses. Further, load or spill estimates may be bases for other planning decisions and forecasts that require the knowledge of the expected number of passengers on board. They may affect the optimal planning of onboard services, the optimal allocation of ground personal, the estimation of the expected available freight capacity in the belly of the aircraft, as well as everything that is in some way related to the expected number and the mix of OD passengers on a flight leg.

In order to obtain valid load and spill estimates, all aspects of the booking process and the operations that can affect the OD passenger flows should be considered. We argued in Chapter 2 that, besides network effects, yield management decisions also have a direct affect on the passenger flows; therefore, effects of yield management should also be considered.

In this chapter, we will outline two previously developed algorithms for estimating leg-dependent passenger flows along with their limitations. Next, a new simulation based approach to calculate leg-dependent load and spill estimates is outlined. An important feature and advantage of the simulation based approach is that, besides networks effects, it can capture directly the different impacts of the various yield management approaches. The new simulation based approach is used as one of the building blocks of the leg-dependent fleet assignment approach that we will present in Chapter 8.

6.1 Deterministic Demand Case

According to our knowledge, the only published paper that addresses the issue of leg-dependence in the context of calculating OD itinerary flows is attributed to Phillips et. al. [29], who identify the issue of network effects and recognize that they can influence the actual OD passenger flows. The authors present an algorithm to calculate “consistent itinerary flows” in airline networks for cases where demand is given in a deterministic form and the total capacity limits on all flight legs are known. The paper recognizes that leg-dependence may influence passenger flows in ways that can affect other planning decisions as well. It also addresses the case of recapture and implications for revenue management, only in the case
when demand is deterministic. The implications of yield management systems on estimating passenger flows in an airline network were not considered.

The algorithm given in [29] follows the arguments and modeling assumptions presented in Section 4.1. In each iteration the algorithm selects the next flight leg that becomes full, calculates load and spill, and the new censored passenger flows in the network. The next flight leg that becomes full is the one with the highest demand to capacity ratio (i.e., with the highest demand factor). This leg will censor OD demands and therefore, changes the achievable leg traffic on other legs of the network. To determine the next flight leg that becomes full, the new achievable traffic (demand) to capacity ratios are considered. To calculate these new traffic per capacity ratios, the following logic is used: The same amount of censored OD demands must be accepted on the all legs of the OD’s path as were accepted on the leg that censors the flight leg. Had this not been the case, the sequence of legs becoming full would have been different. Therefore, on each flight leg \( l \) of the OD demand’s path, the capacity that can be used by other OD demands, (we refer to it as remaining capacity), is the original capacity of leg \( l \) reduced by the amount of accepted passengers on the censoring leg (we called this part of the demand achievable OD traffic). Note that achievable OD traffic values at any point in the algorithm represent the demands that are censored on flight legs that became full earlier. To calculate the new demand factors in order to determine the next flight leg that becomes full, this newly calculated remaining capacity value is considered. This capacity is assumed to be filled by those OD demands that traverse the flight leg and which are not yet censored by any leg of the OD demand’s path. Thus, the new demand factors of an affected flight leg can be calculated as the sum of all non-censored OD demands, divided by the remaining capacity. Using this logic iteratively, consistent passenger flows, load, and spill can be determined for each flight leg. The algorithm, based on the above logic, can be easily adapted to the case when the different OD demands book differently.

To summarize, Philips et. al. [29] present a very important recognition of leg-dependent passenger flows and gives an algorithm to calculate passenger flows in networks where total capacity limits are given. They also recognize that network effects can affect yield management decisions, but they do not address the possibility that yield management decisions may affect the calculation of passenger flows in the network. Hence, they did not address the issue that nested booking limit capacity constraints influence the passenger flows. Additionally, only the deterministic demand case was addressed, despite the fact that one of the most fundamental characteristics of airline demand is stochasticity. Unfortunately, as we will see in the next section, stochasticity complicates the calculation of leg-dependent passenger flows to a large extent, and the basic ideas of the above algorithms cannot be directly extended to the stochastic case.

Next, we describe two models that try to estimate load and spill for the stochastic demand cases. The first model we present is a model developed by researchers at Boeing. The second is a new simulation-based algorithm that we have developed for load and spill calculation under stochastic demand conditions.
6.2 Stochastic Demand Case

In this section, we will assume the more realistic stochastic representation of OD demands. It will be assumed that OD demand can be represented by a probability density function and we will assume independence among different OD demands. First, we will present the model developed by researchers at the Boeing Co.; second, we will present a Monte Carlo simulation based approach, that overcomes some limitations of the first model and provides additional important modeling considerations.

6.2.1 Boeing’s P-Flow Model

In the early 1980’s, Hopperstad at Boeing Co. realized that early work on passenger flows concentrated on the analysis of single OD markets (market share models) and of single flight legs (spill) only. With his colleagues, he developed the “PFlow” Passenger Flow Allocation model that allocates passengers to individual flights considering the total network (non published internal working paper [25]). This algorithm is the probabilistic ancestor of the deterministic algorithm presented above.

The algorithm allocates passengers to individual flight legs considering the total route network. The algorithm follows iterative steps similar to those of the deterministic algorithm: Select a flight leg that has the highest probability to spill; calculate the load and spill for the flight leg and determine the new leg-dependent demand conditions on the affected flight legs; and, determine the new decision parameters for selecting the flight leg that becomes full next.

Theoretically, the modeled problem is broken into two parts: Determining the spill and passenger flows in the network, and reallocating spill to the network. The latter part reallocates spill to new OD paths. In this dissertation we have assumed that spill is lost and thus cannot be reaccommodated; therefore we will not describe this part of the algorithm.

There are two main cycles in the algorithm, see Figure 6.1. The first cycle (inner cycle) selects a flight leg with the highest probability to spill, and recalculates residual demands on all paths that are affected by the spill. Within each cycle, PFlow allocates the remaining demand to passenger paths, calculates leg demands, and determines the probability of passengers getting a seat for each leg. The other cycle reassigns spilled demands to alternative paths.
At each step of the process, the flight leg with the highest probability to spill is chosen (critical or choke leg). Spill for a flight leg is calculated using the Boeing normal distribution spill model (Swan's spill model given in equation 3.34). The accepted OD demand on this flight leg (achievable traffic) is used to determine loads on all flight legs included in the censored OD paths.

The PFlow algorithm is useful and instructive, but is subject to some limitations which compromise its accuracy. First of all, it does not incorporate any effects of yield management. This can be explained by the fact that when the algorithm was developed, yield management practices were not widely used in the industry. The second limitation of the algorithms is that the sequence of the flight legs that are selected as critical legs (i.e., the leg with the next highest probability to spill) is determined by the expected number of spilled passengers on the previous legs. That is, the leg that is selected as the critical leg in an iteration depends on the load and spill (which are expected values) calculated in the previous iterations. For determining the actual probability for a flight leg to be the critical leg, the probabilities associated with the OD demand combinations (that determine the sequence of legs being full) should be evaluated. Under different OD demand combinations, the sequence of the flight legs that become the critical leg can be different. The relative frequency of the different OD demand combinations are a function of the probability distributions. With different probabilities, different sequences of legs becoming full can happen, which determine the direction and magnitude of the network effects. In determining the sequence of flight legs that become full, the relative frequencies associated with the OD demand combinations needs to be considered (see Section 4.2.2). The possibility of the different sequences are not captured in the PFlow model, because expected load and spill values drive the calculations and the selection of the next critical leg (which determine only one sequence of legs becoming full).

Consequently, the PFlow algorithm may have been useful for the purpose it was developed in the early
1980's, however, it has limitations with respect to calculating leg-dependent load and spill in today's airline networks.

Next, we will present a Monte Carlo simulation based approach that overcomes the above limitations and gives a more relevant spill and load estimate in stochastic networks under the given assumptions. Other advantages of this approach are that it can handle any type of the more realistic stochastic representation of demands and other important aspects that affect the passenger flows of the network (e.g., the effects of nested yield management booking limits) can also be easily modeled.

6.2.2 The SMCS Model -- Calculating Load and Spill by Monte Carlo Simulation

In order to determine load and spill at the flight leg level or across the whole network, the interactions of the different OD demands in the stochastic environment should be modeled. In Section 4.2.2 we described an analytical approach for calculating leg-dependent spill and spill costs, but we could see that even for a two leg case, the calculation is very cumbersome. In larger networks with more OD demands, the analytic approach becomes practically infeasible.

In this section, the load and Spill estimation by Monte Carlo Simulation (SMCS) model is outlined. This approach can handle large and complex probabilistic network representations and obtain more realistic feasible load (revenue) and spill (spill cost) estimates for the flight legs and the entire network. The flexibility of the simulation process, in addition, provides the opportunity to incorporate dynamically other decisions of the airline that affect the actual passenger flows of the network (e.g., yield management decisions). The approach also overcomes many of the limitations of the two previously described approaches.

Remember that in dissertation we assume that a spilled passenger is lost to the airline, that is, a spilled passenger cannot be recaptured and be accommodated on another flight or on another path. Note, that the approach outlined below is not fully useful for a general assessment of flow, but for the purpose of determining leg-dependent flows under the assumption of no recapture is suitable.\(^1\)

A simulation approach is a procedure in which a mathematical model representation of reality is employed to perform various experiments and accumulate statistics about the reactions of the simulated system. The simulation technique that we present here is a Monte Carlo based microscopic simulation. The simulation is run at the level of the individual booking request. In this model, actual OD demands (random deviates) are randomly generated from various probability density functions characterizing the OD market. The actual OD demands are loaded on to the underlying airline network according to their predetermined path, observing the prevailing capacity limits on each leg. The entire process of generating

\(^1\) For an approach that addresses the calculation of flow under a more general environment where passenger behavior, route choice selection, recapture is modeled, the reader is referred to [26].
demand and loading it into the network is repeated a large number of times, providing statistics for approximating the expected load, revenue, spill, and spill cost estimates for each flight leg and across the entire network.

The main process in the simulation is the generation of random deviates from the given assumed demand densities and processing them. Figure 6.2 presents the basic steps of the simulation.

**Step 0:** Simulation Initialization

**Step 1:** Initialization for an iteration

**Step 2:** Generate random deviates from the probability density functions for each OD demand

**Step 3:** Calculate weighted probabilities for each not fully processed OD demand where probabilities are weighted by the actual OD demand deviate values generated in Step 2.

**Step 4:** From all OD deviates select randomly an OD demand, where the probability of selecting an OD demand is weighted by the actual OD demand deviate (this step assures the proportional uniform booking rates).

**Step 5:** Process a booking request of the OD demand selected in Step 4. If the capacity limits on any of the flight legs along the OD demand’s path are not binding, then accept the request and update load on each affected flight leg. Otherwise, attribute all spilled OD demand, (the part of OD demands that are generated in Step 2 but at this point not yet accommodated,) to the flight leg that is full. If more than one flight leg is full at this point, then attribute spill to an arbitrary selected flight leg.

**Step 6:** If all OD demands generated in Step 2 are processed, goto Step 7; otherwise goto Step 3.

**Step 6:** If the number of iterations are achieved goto Step 7; otherwise, goto Step 1.

**Step 7:** Calculate Summary Statistics

Figure 6.2: Presentation of the basic simulation process

Through the random selection of the demand deviates, the random combination of the actual OD demands are generated by the probability densities. In any iteration, the probability that a demand value $d$ will be generated for an OD demand $ij$ is $f_{ij}(d)$, where $f_{ij}(d)$ is the probability density function of the demand (see Figure 6.3). By running the simulation for a large number of iterations and processing the actual bookings according to the prevailing capacity conditions, the different unconstrained OD demand combinations, which result in different sequences of legs becoming full and thus in different load and spill values on each leg and across the network, are generated and considered by the correct relative frequencies.

![Figure 6.3: Probability density function](image)

106
Note that any type of probability density function can be assumed for the demand densities. In our simulation, we assumed that demand is represented by a Gaussian or a Poisson distribution, which have been shown to be most relevant for airline demand [10], [27], [38].

In the simulation process, it is possible to assume more than one fare classes for each OD demand. In this case, however, we have to make some assumption about the sequence of the different ODF bookings. In practice, airline booking patterns can be complex, with interspersed arrivals of passenger requests from different fare classes over the course of the booking process. The different booking patterns in each fare class affect the sequence of requests received by the airline, which has an effect on fare class mix of accepted and of spilled passengers. It is possible to capture the booking patterns by dividing the booking process into many booking periods. Then, incremental probabilistic booking demand within each period for each ODF can be determined.

In our simulation approach, we assume that data to support this multiple-period representation of ODF demands is available (i.e., from YM databases). We assume that within each period for each ODF demand uniform proportional booking patterns prevail. Another modeling assumption of the simulation process is that lower-valued fare classes book before higher-valued fare classes but only within period. If the number of periods is relatively large, then the overall interspersed booking patterns, embedded in the multiple-period demand, will dominate. In other words, the effects of the "lower fare classes book before higher fare classes" assumption within each period will diminish as the number of periods increases.

In the simulation we assumed that ODF demands are independent across the fare classes and across the booking periods. Note that it is possible to implement correlation between OD demands or between the demands between the booking period.

The advantage of the simulation approach is that, in addition to total leg capacity limits, it can incorporate the effects of the booking limits set by yield management systems. As we will show in the next chapter, yield management decisions influence significantly the ODF passenger flows of the network and have substantial effects on the resulting load and spill estimates.

In the yield management system the booking process is also divided into booking periods. Demand is forecasted incrementally for each period and nested booking limits are recalculated dynamically at the beginning of each period (also called the "check point"). Using the same division of the booking process that is used in the yield management system, nested booking limits can be obtained and used as inputs in the simulation.

Given a nested multiple-fare class and multiple-period representation of demand, the above outlined simulation procedure needs modification. Figure 6.4 presents the new modified description of the simulation process. The basic concepts of the simulation remain the same; it is only augmented by two
new cycles. One additional cycle represents the iteration over the different fare classes, \( c \), (lower-valued fare classes book before higher-valued fare classes and highest fare class is \( c=1 \)), and the other cycle represents the iteration over the booking periods, \( p \). Note that the requirement for an ODF request to be accepted is that on all flight legs of the ODFs path all corresponding nested booking limits be greater than zero. Depending on the prevailing booking limits, after a booking request is accommodated, nested booking limits need to be updated accordingly. For example, if nested booking limits are assumed, then after a booking request for fare class \( c \) is accepted on a flight leg, the booking limits of fare class \( c \) and of all fare classes that are higher-valued than \( c \) should be reduced.

**Step 0:** Simulations Initialization

**Step 1:** Initialization for an iteration

**Step 2:** Select first booking period (\( p=1 \))

**Step 3:** (Re)calculate (nested) booking limits for each fare class and flight leg for period \( p \) based on the demand forecast for the time-interval between period \( p \) and departure.

**Step 4:** Generate random deviates from the probability density functions for each ODF of booking period \( p \).

**Step 5:** Select the lowest-valued fare class (\( c \))

**Step 6:** Within the fare class calculate weighted probabilities for each not fully processed ODF demand where probabilities are weighted by the actual ODF demand deviates in the fare class generated in Step 4.

**Step 7:** From all OD deviates select randomly an ODF demand, where the probability of selecting an ODF demand is weighted by the actual ODF demand deviate (this step assures the proportional uniform booking rates within the period).

**Step 8:** Process a booking request of the OD demand selected in Step 7. If the booking limits associated with the ODF along the flight legs of the ODF's path are larger than zero, then accept the request and update load on each affected flight leg and update all affected nested booking limits. Otherwise, attribute all spilled ODF demand of period and the periods that are still to come until departure, (i.e., the part of ODF demand that is generated for this period but not yet accommodated and plus the demands generated for the later periods,) to the flight leg that had a booking limit equal to zero. If more than one flight legs have booking limits equal to zero, then attribute spill to an arbitrary selected flight leg.

**Step 9:** If not all individual ODF demand requests are processed (i.e., accommodated or spilled) for booking class \( c \), in period \( p \), then goto Step 6.

**Step 9:** If not the highest-valued fare class (\( c=1 \)) is processed, select the next higher-valued fare class (\( c=c-1 \)) and goto Step 6.

**Step 10:** If not the last booking period is processed, then select next period (\( p=p+1 \)) and goto Step 3.

**Step 11:** If not the last iteration, then goto Step 1;

**Step 12:** Calculate Summary Statistics

Figure 6.4: Presentation of the multiple-fare class multiple-booking period simulation process
By using the above simulation approach, the leg-dependence effects influenced by the nested booking class capacity limits can be also captured. Note that modeling of network effects in a system where nested fare class limits censor the probabilistic ODF demand flows is a more complex process than the case when only total leg capacity limits impose constraints.

Note that by actually tracking all individual demand requests in the simulation, not only the load and spill for the leg but also the corresponding revenue and spill cost estimates can be estimated more accurately. Since the simulation is done on a microscopic booking request level, revenues and spill cost can be directly calculated by multiplying each accepted or spilled booking request of an ODF by the corresponding fare value. This way the OD mix of load and spill, which is subject to the different ODF booking patterns and booking limits, is correctly determined. This is another very important feature provided by the microscopic simulation process.

The simulation process can be further enhanced by the representation of different yield management approaches. In Chapter 7 (Section 7.6), we will introduce different yield management system approaches (where the differences are in the efforts to heuristically evaluate ODFs according to their network values), and will show that the different yield management approaches influence differently ODF passenger flows and thus load and spill estimates. Among the different approaches, we will address an OD heuristic that dynamically determines ODF booking availability by the actual demand and load conditions in the network. It is possible to embed the actual yield management decision process and algorithm into the simulation and experiment its effects on the leg-dependence issue and on actual passenger flows. In this approach, at the beginning of each simulated booking period, yield management booking limits and ODF booking availability are calculated within the simulation process, where the input demand probabilities (forecasts) and the simulated demand conditions in the network are considered. This procedure is repeated over all booking periods and over all iterations.

The output statistics of the simulation provide information about the following:

For each ODF demand:
◊ Simulated actual mean and variance
◊ Load
◊ Spill

For each flight leg:
◊ Simulated unconstrained leg demand
◊ Load
◊ Leg revenue
◊ Spill
◊ Spill Cost

For the network:
◊ Total load and revenues
Total spill and spill cost
Yield
Load factor, etc.

We implemented the simulation process in C language and used different computer platforms to run the code. Our experience has showed that the process is fast. Naturally, the running-time of the simulation is a function of the sizes of each cycle in the process. Namely, the number of the fare classes, the number of booking periods, and the number or iterations all affect the run-time performance. The more booking periods that are considered, the more accurately the actual booking patterns can be replicated and the less are the effects of the simplifying “lower-valued booking classes book earlier” assumption; however, the longer the simulation runs. Similarly, the more iterations that are run, the more representative are the obtained statistics; however the longer the running-time is. Table 6.1 presents simulation running-times for networks with different demand and network conditions. The number of iterations in each case were 100 and 7 fare classes with 15 booking periods were assumed.

<table>
<thead>
<tr>
<th>Processor</th>
<th>3 leg network</th>
<th>2x2 hub network</th>
<th>4x4 hub network</th>
<th>15x15 hub network</th>
<th>25x25 hub network</th>
<th>30x30 hub network</th>
</tr>
</thead>
<tbody>
<tr>
<td>DX2-66 (PC)</td>
<td>3 sec</td>
<td>5 sec</td>
<td>11 sec</td>
<td>3.4 min</td>
<td>6.9 min</td>
<td>8.6 min</td>
</tr>
<tr>
<td>DEC 5000/25</td>
<td>3 sec</td>
<td>5 sec</td>
<td>10 sec</td>
<td>2.9 min</td>
<td>4.7 min</td>
<td>7.6 min</td>
</tr>
<tr>
<td>IBM RS6000/370</td>
<td>2 sec</td>
<td>4 sec</td>
<td>6 sec</td>
<td>55 sec</td>
<td>1.5 min</td>
<td>2.3 min</td>
</tr>
</tbody>
</table>

Table 6.1: Run time performances for different leg-dependent, multiple fare class, multiple period network examples

6.3 Summary

The deterministic algorithm presented in [29] is important in that it recognized that network effects influence passenger flows. The paper presented an algorithm to calculate passenger flows in networks and recognized that network effects can influence yield management decisions as well. However, the paper did not address the issue that yield management decisions can affect the passenger flows of the network also. The deterministic algorithm and modeling approach gives a good insight into the underlying phenomenon, but it fails to account for more characteristic stochastic conditions.

The microscopic Monte Carlo simulation based approach introduced in this Chapter is a flexible and useful tool to estimate demand related estimates (load, spill, revenue, spill cost) in airline networks where complex stochastic network and leg-dependence effects need to be modeled. The simulation based approach generates and processes demand at the level of booking requests and obtains the relevant
statistics. By considering all capacity limits of the network, the complex random interactions of the different ODF booking requests in the network can be modeled and estimated.

We can summarize the advantages of the simulation approach in the following:

◊ ODF demands can be represented and generated by any type of assumed probability density functions.

◊ It is possible to represent multiple-fare class ODF demands by general (non-standard) booking patterns by dividing the booking process into many booking periods.

◊ By representing different types of capacity limits, the complex stochastic leg-dependence effects imposed by the nested booking class capacity limits can be also modeled.

◊ By processing the simulation on a microscopic level, not only the leg-dependent load and spill estimates, but also the corresponding leg-dependent revenues and spill costs can be correctly estimated.

◊ In the simulation it is relatively easy to model the effects of yield management decisions, which also modify the actual passenger flows and the corresponding demand related estimates. It is possible to embed into the simulation the assumed dynamic yield management optimization algorithm and evaluate the effects of the different yield management optimization algorithms and approaches.

◊ Using advanced software techniques and available advanced computer hardware, the simulation process is a relatively fast approach to obtain the required estimates.

◊ The approach is not directly applicable for a general assessment of flow when recapture of spilled passengers is possible. In this case passenger behavior and path preference decisions need to be modeled additionally [26]. Nevertheless, the simulation approach is useful and gives accurate estimates under the assumption of no recapture.
7. Yield Management Impacts on Airline Spill Estimation

While making advances in the solution of the large-scale fleet assignment optimization problem, airlines have continued to use an aggregate approach for spill estimation developed over a decade ago. This aggregate approach ignores the effects of yield management practices which have been widely implemented by airlines during the past decade. The purpose of this chapter is to demonstrate the extent to which airline yield management practices can affect the correct estimation of both spill and spill costs. We show the limitations of the state-of-the-practice methods and present a new approach for more accurate spill cost estimation. We also show that by aggregating flight leg demand over a longer time period into one demand density, spill and spill cost estimations can be further biased.

In the second part of the chapter we will address the effects of different yield management approaches with respect to their influence on network effects. We will show that different approaches control the network passenger flows in a different way, and therefore the leg load and spill estimates. Consequently, the revenue and spill cost estimates on flight legs may differ under the different approaches.

Numerical examples are presented through the chapter to illustrate the extent to which the outcomes of the different estimation approaches differ and how these differences can be large enough to have an impact on optimal fleet assignment.

7.1 Current Practice in Spill Estimation for Fleet Assignment

Fleet assignments are repeated daily over a longer time interval. Most airlines represent the expected demand for a given flight leg during the planning time interval as a single Gaussian probability function [1], [24],[27], [35], [36]. The demand that represents the whole planning interval, few weeks or months, is expressed in a form of a single, aggregated, average normal probability function. Representing demand this way, the variation in the average daily passenger demand is attributed not only to random variations, but also to differences in day-of-week demand variations, seasonality, and other cyclic effects. The aggregation of demand over this longer time interval into an average demand function we call horizontal aggregation.

Because multiple fare class demands are aggregated into a single density, we refer to this demand representation as the total demand function. Yet, it is a fundamental assumption of yield management systems that different probability density functions, or “demand densities”, exist for each fare class offered on a flight. The use of a single demand density to represent total flight demand thus involves combining multiple fare class demands, which does not account for the different fare class demands nor the corresponding different fare values. We call this aggregation of fare class demands vertical aggregation.
As we shall see in what follows, both vertical and horizontal aggregation causes biases in the estimation of spill and spill costs if the traditional spill approach is employed.

Despite the introduction of YM systems and the widespread use of seat allocation algorithms, not much attention has been paid to the impacts of these practices on spill estimation. One exception is Swan [40], who revisited the basic spill model in an unpublished paper. Among other important issues he recognizes that YM systems can have effect on the spill estimates. He argues that the no-show phenomenon, the overbooking practices by the airlines, and the presence of discounted fares affects the actual spill values. Swan also notes that the revenue spilled varies with the fare class mix of demands spilled and the amount of spill. At low values of spill, the mix is nearly the same as the mix of demands, while at high values of spill, the spill is dominated by lower fare passengers, due to the use of yield management practices. Swan suggests approximations to correct for the problem, but does not consider more than two fare classes or present a specific algorithm to calculate spill in such cases. (For a more detailed overview of the spill estimation literature refer to Section 3.4 in Chapter 3.)

Although we did not find any published information on how airlines estimate the revenues of spilled passengers for different aircraft capacities, informal discussions with airline fleet assignment modelers revealed some alternative approaches. As a general rule, these airlines employ one of two approaches for estimating the average fare of the spilled passengers, which we will refer to as spill-fare:

(A) the average fare weighted by the mean demands of each fare class.

(B) a constant approximated value in the range of the lowest available fares.

The latter approach represents an attempt to adjust for the effects of yield management by using a fare value lower than the flight's average fare as an approximation of the spill-fare.

Thus, many airlines are aware that their yield management systems are affecting spill costs, and are adjusting their estimates of spill costs by using “rules of thumb” with respect to the proper spill-fare. Even though the spill-fare is adjusted on an ad-hoc basis, it is still common practice to use a single demand density for estimating the number of passengers spilled at various capacities. This practice implies that airlines believe that yield management systems affect the fare class mix of spill, but not necessarily the magnitude of spill. In the sections that follow, we present new algorithms for more accurate spill estimation for multiple fare classes under conditions of yield management, and then demonstrate that in fact both the fare class mix and the magnitude of spill can be affected by yield management systems.
7.2 A New Approach for Spill Estimation Under Yield Management

As noted in Chapter 2, yield management algorithms set booking limits to optimize the fare mix of passengers. Booking limits affect the actual load in each fare class, and at the same time affect the number of spilled passengers. Therefore, the actual passenger flows resulting from the ODF demands are also affected. To assess the extent to which they are affected and the parameters that have a major influence we have to recall the basic principles of yield management optimization.

Yield management algorithms protect seats for fare classes \( l \) to \( c \), as long as the combined expected marginal seat revenue for classes \( l \) to \( c \) is greater than \( fare_{c+1} \), the fare of fare class \( c+1 \). Hence the protection level for class \( c \), \( Prot_c \), is the number of seats \( x \) for which the following equation holds:

\[
P(x)_{x_c} \cdot fare_{x_c} \geq fare_{x+1},
\]  

where \( P(x)_{x_c} \) and \( fare_{x_c} \) represent the combined probability of receiving \( x \) booking requests and the combined fare for fare classes \( l \) to \( c \), respectively. The booking limit for class \( c \) is \( BL_c = Cap - Prot_{c-1} \).

The protection level changes as the relative fare values change, assuming constant demand distributions. With the change in protection levels, booking limits change as well, thereby influencing the expected number of spilled passengers. The smaller the discount ratio \( d \) is (where \( d = fare_{c+1}/fare_{c} \)), the greater the effects of the yield management system on spill will be. This is because the smaller the discount ratio, the more seats will be protected for the higher fare classes; which also means that more low fare class demand will be spilled.

It is incorrect to assume that changes in the fare ratios change only the mix of the spilled demand, but not its magnitude. This reasoning would suggest that there would be no difference between spill estimates of the total demand density and the spill estimates when different fare classes and booking limits are considered. The different fare ratios, through their impacts on the optimal booking limits, change not only the mix but also the total number of spilled passengers.

In the following we present a new approach for spill and spill cost estimation that takes into consideration the effects of the nested booking limits being applied by YM systems. In this section we will assume the case when spill and spill costs are evaluated for a single flight leg only.

In this section, we use an approach similar to that used by Curry [20] in the OBL algorithm to derive a formula for calculating the expected spill and spill cost for nested multiple fare class structures common to airline reservations and yield management systems. We assume the following about the booking process:

1. The different fare class demands are independent from each other.
2. There are no cancellations of bookings.

3. Spilled passengers are lost to the airline, meaning we do not account for recapture.

With multiple fare class demand, we have an additional assumption about the booking patterns and sequence of fare class bookings:

4. All of the lower-valued fare classes book before the higher-valued fare classes.

This last assumption is relaxed later in our discussion. All of these assumptions are in fact consistent with those employed by the seat allocation optimization models reviewed previously.

Booking limits for a fare class limit the number of passengers that can be booked in that fare class; hence they determine the actual spill of a fare class. In the nested multiple class case, the available capacity for a fare class is a function of both the booking limits and the seats already sold. Because of Assumption 4, the available capacity for a fare class is always equal to the booking limit for that fare class minus the total seats already sold in all lower-valued fare classes.

Let $Spill_c$ denote the total spill from fare classes 1 to c, where fare class 1 corresponds to the highest-valued fare class. Because of the nested structure of the reservations system, $Spill_c$ depends on the actual fare class demand, and the number of seats sold to the lower-valued fare classes. Consequently, in our recursive formulation $Spill_c$ is a function of $S$, the number of seats already sold. Booking limits are calculated by yield management optimization algorithms and are assumed to be given. $Spill_c[S]$, the total spill for fare classes 1 to c is given by:

$$Spill_c[S] = \int_0^{BL_c-S} f_c(i) \cdot Spill_{c-1}[i+S]di + \int_{BL_c-S}^{\infty} f_c(i) \cdot [i-(BL_c-S) + Spill_{c-1}[BL_c]]di,$$

$$Spill_0 = 0.\tag{7.2}$$

where $f_c(i)$ is the probability density of fare class c requests, $BL_c$ is the booking limit for each class c, and $S$ denotes the seats already sold for the flight.

At the beginning of the booking process $S=0$. The recursion starts with the lowest valued booking class. The stopping condition in the recursion is: $Spill_0 = 0$. Note that $BL_c$ is calculated as: $BL_c=Cap\cdot Prot_{c-1}$, where $Prot_{c-1}$ is the total protection level for fare classes 1 to c-1. Thus, the information about the aircraft capacity is embedded in the booking limits. The first term of the right hand side in (7.2) represents the case when demand for class c is less than the available seats for that fare class, $BL_c - S$, i.e., when no fare class c demand is spilled. In these cases the total spill for fare classes 1 to c is from fare classes above c, which is included in $Spill_{c-1}$. The spill of the higher-valued fare classes (1 to c-1),
however, depends on the number of seats already sold, that is, \(i+S\). The second term of the right hand side of (7.2) is the spill when the requested bookings in fare class \(c\) are greater than the available seats. In the case of \(i\) booking requests the spill from fare class \(c\) is equal to \(i-(BL_c-S)\), which together with the spill from the higher fare classes is multiplied by the probability of having \(i\) fare class \(c\) booking requests. Because of the nested class structure, when booking class \(c\) spills, the number of seats sold must be equal to the booking limit of fare class \(c\). Note that equation (7.2) reduces to equation (2.2) if \(c=1:\)

\[
Spill_c[S] = \int_0^{BL_c-S} f(i) \cdot i \cdot di + \int_{BL_c-S}^\infty f(i)(i-(BL_c-S)+0)di = \int_{BL_c-S}^\infty f(i)(i-(BL_c-S))di \tag{7.3}
\]

We can then derive directly the expected spill cost formula, where the number of spilled passengers from fare class \(c\) is multiplied by the fare value of that class, \(\text{fare}_c\). Hence the spill cost, \(SC_c[S]\), for \(c\) nested fare classes is given by:

\[
SC_c[S] = \int_0^{BL_c-S} f_c(i)SC_{c-1}[i+S]di + \int_{BL_c-S}^\infty f_c(i)\text{fare}_c \cdot (i-(BL_c-S)) + SC_{c-1}[BL_c]di, \tag{7.4}
\]

\(SC_0 = 0.\)

The most important assumptions in the above models are that fare class demands are independent, and each lower-valued fare class books in its entirety before the next higher-valued class. The latter is a simplifying assumption about actual booking patterns. Note that by partially differentiating \(SC_c[S]\) with respect to the booking limits, \(BL_c\), and setting it equal to 0, we obtain the optimality conditions for determining the optimal booking limits for revenue maximization \(^1\). This is the result reported by Curry [20], Robinson [31], Wollmer [44], and Brumelle and McGill [18].

For the purposes of this chapter, we continue to assume that \(f(i)\) is normally distributed, and is given separately for each fare class \(c\). This assumption makes the above formulae cumbersome to calculate. It is possible, however, to use various numerical methods to calculate spill and spill costs. One way would be to use numerical integration methods or use a type of Monte Carlo simulation.

The above two formulae represent the correct spill and spill cost calculations for the nested multiple fare class problem in which bookings are assumed to occur in ascending order of fare class values. In practice, however, airline booking patterns are more complex, with interspersed arrivals of passenger requests from different fare classes over the course of the booking process. The different booking patterns in each fare class affect the sequence of requests received by the airline, which has an effect on the fare class mix of

\[^1\] where \(MSC_c[S] = \frac{\partial SC_c[S]}{\partial S}\)

116
spilled passengers. Therefore, for an even more realistic spill estimate under yield management, the effects of these booking patterns must be incorporated into an even more detailed model, as described below.

In yield management systems, historical fare class booking patterns are captured by dividing the booking process for each flight into many booking periods. Airlines re-optimize booking limits at “reading points” or “checkpoints”, at the beginning of each period, during the booking lifetime of the flight. Thus, incremental booking demand within each period and for each fare class is available from the yield management system.

In our more detailed model, we assume that this multiple-period representation of fare class demand is available for spill estimation. We also assume that lower-valued fare classes book before higher-valued fare classes only within period. If the number of periods is relatively large, then the overall booking patterns, embedded in the multiple-period demand, will dominate. In other words, the effects of the “lower fare classes book before higher fare classes” assumption within each period will diminish as the number of periods increases.

The following model enables us to calculate the expected spill and spill costs for the nested multiple period, multiple fare class problem. Let period 1 denote the closest period to departure and period p the pth period prior to departure. Let $Spill_{p,c}$ represent the cumulative spill for p periods and for c fare classes, again from the “top down”. The formulae which follow are different from equations (7.2) and (7.4) in the extra time-period dimension, p. Note also that in the formulae, $f_{c,p}(i)$, the demand distribution, represents the density of incremental demand for fare class c within period p. Also the booking limit, $BL_{c,p}$, is the prevailing booking limit in period p and in fare class c. We assume that at the beginning of each period, booking limits for fare class c are calculated as the difference between the capacity and the protection level for fare classes 1 to c-1\(^2\). We also assume independence between the time-period demands. Thus the total spill, $Spill_{p,c}[S]$, for the c fare class p multiple-period demand is given by:

\[^2\text{Note, that in yield management optimization algorithms in each period booking limits are calculated by considering for each fare class the total demand to come until departure. That is, in period } p \text{ for each fare class demand is calculated as the sum of the incremental fare class demands between period } p \text{ and period } 0, \text{i.e., the departure.}\]
\[
\begin{align*}
\text{Spill}_{p,c}[S] &= \int_0^{BL_{c,p}-S} f_{c,p}(i)\text{Spill}_{p-1,c+1}[i+S]di + \\
&+ \int_{BL_{c,p}-S}^{\infty} f_{c,p}(i)(i-(BL_{c,p}-S) + \text{Spill}_{p-1,c}(BL_{c,p}))di & \text{if } c > 1 \\
\text{Spill}_{p,c}[S] &= \int_0^{BL_{c,p}-S} f_{c,p}(i)\text{Spill}_{p-1,NFC}[i+S]di + \\
&+ \int_{BL_{c,p}-S}^{\infty} f_{c,p}(i)(i-(BL_{c,p}-S) + \text{Spill}_{p-1,NFC}(BL_{c,p}))di & \text{if } c = 1 \\
\text{Spill}_{0,c} &= 0
\end{align*}
\]

where $NFC$ stands for the number of fare classes. The total spill cost, $\text{SC}_{p,c}[S]$, for the $p$ period $c$ fare class demand is given by:

\[
\begin{align*}
\text{SC}_{p,c}[S] &= \int_0^{BL_{c,p}-S} f_{c,p}(i)\text{SC}_{p-1,c+1}[i+S]di + \\
&+ \int_{BL_{c,p}-S}^{\infty} f_{c,p}(i)(\text{fare}_c \times (i-(BL_{c,p}-S)) + \text{SC}_{p-1,c}(BL_{c,p}))di & \text{if } c > 1 \\
\text{SC}_{p,c}[S] &= \int_0^{BL_{c,p}-S} f_{c,p}(i)\text{SC}_{p-1,NFC}[i+S]di + \\
&+ \int_{BL_{c,p}-S}^{\infty} f_{c,p}(i)(\text{fare}_c(i-(BL_{c,p}-S)) + \text{SC}_{p-1,NFC}(BL_{c,p}))di & \text{if } c = 1 \\
\text{SC}_{0,c} &= 0
\end{align*}
\]

If the above formulae were to be used in calculations involving many fare classes and many periods, the depth of the recursion would become prohibitively large. The depth of the recursive stack increases dramatically because, in addition to being nested by the fare classes, it is also nested by the booking periods. Even with advanced computational techniques, the calculations would still take a prohibitively long time. A practical solution to this problem is to approximate the above equations with a Monte Carlo simulation of the airline booking process.

The models presented in this section for estimating spill and spill costs in the context of multiple nested fare classes represent a new approach for taking account of the impacts of yield management systems on
passenger spill. As will become apparent in the following sections, these models can be used to overcome the aggregation biases inherent in current spill and spill cost estimation practices.

### 7.3 Vertical Aggregation Bias

In this section, we explain how the spill estimates from the total demand function can be biased due to vertical aggregation of fare class demands. Furthermore, we discuss how, as a result of these incorrect spill estimates, we expect the spill cost estimates to be biased as well. We call this vertical aggregation bias. Consider the case in which demand for a flight leg is represented by a single Gaussian probability function for a single departure date \( \text{(total demand function)} \). We also assume that there is no correlation between fare class demands. Therefore, the mean of the total demand function, \( \bar{\mu} \), is given by:

\[
\bar{\mu} = \sum \mu_c
\]  

where \( \mu_c \) is the mean demand for fare class \( c \). The standard deviation for the total distribution, \( \bar{\sigma} \), is given by:

\[
\bar{\sigma} = \sqrt{\sum \sigma_c^2}
\]  

where \( \sigma_c \) is the standard deviation of demand for fare class \( c \).

We expect that the actual number of spilled passengers in most cases will differ from the spill estimates of the total demand function. The smaller the discount ratio \( d \) is (where \( d = \text{fare}_1/\text{fare}_c \)), the greater the effects of the yield management system on spill will be. This is because the smaller the discount ratio, the more seats will be protected for the higher fare classes; which also means that more low fare class demand will be spilled. We expect that as the discount ratio decreases, the differences between the spill estimates from the demand density versus the spill estimates when detailed fare class information and booking limits are incorporated will increase.

The following two-fare class example for a single flight leg case shows the indirect effect of the fare ratios (through the applied booking limits) on spill. The mean demand for fare classes 1 and 2 are 40 and 60 passengers, respectively. We assumed that the standard deviations of fare class demands are equal to the square root of the mean. Table 7.1 shows the total number of passengers spilled as a function of fare ratios. The results were obtained by calculating the optimal booking limits for the given demand and discount ratios \( (\bar{x}/d) \), and then by using equation 7.3 to obtain the total number of spilled passengers (i.e., we assumed that first fare class 2 books and then fare class 1). Note that the underlying fare class demand distributions were kept the same in all scenarios. We have changed only the fare ratios (and calculated the corresponding booking limits) and the capacities. The table shows that as the fare ratios become
smaller (i.e., the fare differences become bigger), for a given capacity and same demand, the total spill increases.

<table>
<thead>
<tr>
<th>Capacity</th>
<th>I2/I1=0.6</th>
<th>I2/I1=0.7</th>
<th>I2/I1=0.8</th>
<th>I2/I1=0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>116</td>
<td>0.18</td>
<td>0.19</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td>111</td>
<td>0.72</td>
<td>0.73</td>
<td>0.66</td>
<td>0.68</td>
</tr>
<tr>
<td>105</td>
<td>2.21</td>
<td>2.13</td>
<td>1.97</td>
<td>1.98</td>
</tr>
<tr>
<td>99</td>
<td>5.2</td>
<td>4.91</td>
<td>4.62</td>
<td>4.57</td>
</tr>
<tr>
<td>95</td>
<td>8.09</td>
<td>7.56</td>
<td>7.28</td>
<td>7.2</td>
</tr>
<tr>
<td>90</td>
<td>12.39</td>
<td>11.76</td>
<td>11.37</td>
<td>11.17</td>
</tr>
<tr>
<td>85</td>
<td>16.16</td>
<td>15.48</td>
<td>15.04</td>
<td>14.73</td>
</tr>
<tr>
<td>83</td>
<td>19.12</td>
<td>18.39</td>
<td>17.9</td>
<td>17.6</td>
</tr>
<tr>
<td>79</td>
<td>23.15</td>
<td>22.39</td>
<td>21.85</td>
<td>21.5</td>
</tr>
<tr>
<td>76</td>
<td>26.19</td>
<td>25.43</td>
<td>24.85</td>
<td>24.51</td>
</tr>
</tbody>
</table>

Table 7.1: A two fare class example showing that the total spill is a function of the fare ratios (i.e., optimal booking limits).

The above argument assumes the simple booking pattern in which lower fare classes book before higher classes. It turns out that the vertical aggregation bias is even greater when more realistic interspersed booking patterns of the different fare classes are considered. We argue that various booking patterns can influence both the mix and the number of passengers spilled. For simplicity, assume a two fare class scenario, and consider two booking patterns. In Case A, in the early periods mostly class 2 books. Spill will occur when the booking limit for class 2 reaches 0. Recall that the booking limit is the available capacity minus the protection level for fare class 1 while the available capacity is a function of the seats already sold. In Case A, the seats already sold will be filled mostly with fare class 2 passengers.

In Case B, both fare classes fill the plane from the beginning of the booking process. So, assuming the same rate of booking for class 2 in Case B, in each period the total seats sold on the plane is bigger than it was in Case A, since class 1 bookings are also being accepted. It also means that available seats will decrease at a higher rate. Therefore, assuming for a moment similar protection levels to those of Case A, the booking limit for fare class 2 will reach 0 at a time earlier than in Case A. But there is another force at work here. Because a larger portion of class 1 demand is booked earlier, the protection levels for remaining class 1 demand in Case B become lower than they were in Case A. These smaller protection levels for class 1 result in higher booking limits for class 2. Hence, in Case B we have two effects that can work against each other. These effects could change both the fare class mix and the total number of passengers spilled. Numerical examples, presented below, support the above intuitive explanations.

### 7.3.1 Evaluation of Vertical Aggregation Bias

It is important to understand whether the vertical aggregation biases of existing spill estimates are typically large enough to warrant using a much more complicated estimation approach. If the differences between the estimates produced by the different models are not significant and if these differences have
little impact on the ultimate fleet assignment decision, then it would be hard to justify the use of the proposed estimation approaches. To evaluate the magnitude of the biases, we present in this section several numerical examples based on actual airline data. Our goal here was not to evaluate all possible combinations of input scenarios, but to illustrate the magnitude of biases in some typical and realistic situations. In our evaluation, we compared three different methods for calculating spill and spill costs, as follows:

**Method 1 (Traditional Approach):** Spill is calculated from a single total normal demand representation. The total normal distribution is constructed from the multiple class one- or multiple-period demand representation, assuming no correlation between the fare classes and periods. (For example, given the multiple fare class one period demand representation and assuming independence, the mean and standard deviations of the total demand is calculated by: \( \mu_{total} = \sum c \mu_c \) and \( \sigma_{total} = \sqrt{\sum c \sigma_c^2} \), where \( \mu_c \) and \( \sigma_c \) are the mean and standard deviation of fare class \( c \). For the multiple period case, the total demand is obtained by first calculating the mean and standard deviation for each period, and then the overall total demand calculated as it is given above.) Using the total demand density, spill is calculated by using the formula given in equation 3.34 (Swan's formula). Spill costs are estimated by multiplying the spill by an estimate of the spill-fare. Two different approaches were compared for estimating spill costs:

**Method 1-A (Weighted Average Spill Cost):** The total spill estimate is multiplied by the weighted average fare weighted by the mean demand of each fare class,

\[
SpillCost = \text{Spill} \times \frac{\sum c \text{fare}_c \times \mu_c}{\sum \mu_c}
\]

**Method 1-B (Constant Low Fare Spill Cost):** The total spill for each capacity is multiplied by an assumed average constant fare of spill. This value is assumed to be approximately equal to a weighted average of the two lowest-valued class fares.

**Method 2 (Single Period Multiple Fare Class Approach):** The multiple fare class demand representation is used with Assumption 4, that is, lower-valued fare classes book before higher-valued fare classes. Spill and spill costs are calculated using equations (7.2) and (7.4).

**Method 3 (Multiple Period Multiple Fare Class Approach):** The multiple period multiple fare class demand representation was used. Equations (7.5) and (7.6) for calculating multiple period spill and spill costs were approximated by Monte Carlo simulation.
Comparison of Method 1 vs. Method 2

First, we compare the spill estimates of Methods 1 and 2. Because Method 1 does not incorporate information about different fare classes or their revenue values, we expect that as the fare differences among the classes increase, because of the greater effects of the booking limits, Method 1 will underestimate the total spill by a bigger rate. The examples shown below indeed justify our a priori expectation: As, \( d \), the discount ratio decreases, (i.e., the fare differences become bigger,) the differences between the spill estimates of Method 1 and 2 increase. We can also observe that, for a given discount ratio, as the demand factor (DF), the ratio of mean demand over capacity increases, the spill differences also increase.

We used a 5 fare class demand example (see Table 7.2), and assumed that the standard deviations of total fare class demands are equal to the square root of their means. For the various discount ratios we have set the fares such that for all adjacent lower valued fares: \( fare_{c+1} = d \times fare_c \).

<table>
<thead>
<tr>
<th>Fare class</th>
<th>Mean</th>
<th>Std</th>
<th>Fare ( d=0.6 )</th>
<th>Fare ( d=0.7 )</th>
<th>Fare ( d=0.8 )</th>
<th>Fare ( d=0.9 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>10</td>
<td>3.16</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
</tr>
<tr>
<td>B</td>
<td>13</td>
<td>3.6</td>
<td>360</td>
<td>420</td>
<td>480</td>
<td>540</td>
</tr>
<tr>
<td>M</td>
<td>24</td>
<td>4.89</td>
<td>216</td>
<td>294</td>
<td>384</td>
<td>486</td>
</tr>
<tr>
<td>H</td>
<td>25</td>
<td>5</td>
<td>129</td>
<td>206</td>
<td>307</td>
<td>437</td>
</tr>
<tr>
<td>Q</td>
<td>40</td>
<td>6.32</td>
<td>77</td>
<td>144</td>
<td>246</td>
<td>394</td>
</tr>
</tbody>
</table>

Table 7.2: Data Example 1 and fares at various \( d \) discount ratios

Figure 7.1 shows the spill estimates for Method 1 and 2 for the above data at various discount ratios, \( d \). Since Method 1 is insensitive to the fare structure and does not incorporate the effects of booking limits, its spill estimates are constant over all discount ratios. The differences between the spill estimates of Method 1 and 2 increase as the discount ratio decreases. We can observe that the differences in spill estimates are greater at higher demand factors. Note, however, that the standard deviations in this data example are low (square root of the mean), and we have observed higher spills and spill differences under higher demand variability assumptions. Nevertheless, the \( d=0.7 \) case, which is representative of some recent airline fare structures, shows that the spill differences become substantial around DF=1.0 and above. (Note that it is reasonable to evaluate spill at very high DF cases as well. Due to the cyclic and seasonal variations in demand, it is possible that although the average DF over the planning time period is relatively low, at certain instances -- e.g., on high demand days -- the actual demand to capacity ratio is high. Since these high demand days contribute significantly to the overall (average) spill estimates, the biases of these high demand occasions need to be also considered.)
As well as affecting the total number of spilled passengers, the booking limits also influence the passenger mix of spill. Figure 7.2 shows the estimated average fare of spilled passengers, the "spill-fare", for the various discount ratios as function of demand factor, as calculated by Method 2. We can observe a similar shape of the spill-fare curves for the different discount ratios: Spill-fares are high at lower demand factors and decrease gradually as DF increases. At low demand factors, the spill-fare is higher because when little spill occurs, almost all fare classes are involved in spill. This results in relatively high spill-fares. On the other hand, when the demand factor is high, primarily the lower valued fare classes are involved in spill, due to the fare class booking limits. Thus, the average fares of the spilled passengers are lower at
high demand factors. Note that there can be large differences in the spill-fares for a given discount ratio \(d\), even over a relatively small ranges of DF (0.85 to 1.0).

The changes in spill-fares are largest at lower demand factors, which correspond to the most relevant cases in practice. The results suggest that the assumption that spill-fare

![Graph showing spill-fares at various discount ratios](image)

**Figure 7.2: Method 2 Estimates of Spill Fares at various discount ratios \(d\)**

is relatively constant over a wide range of demand factors is invalid. Indeed, spill-fares as calculated by Method 2 are more sensitive to capacity changes at low demand factors and rather constant at the higher demand factors. Note, however, that in our example the mean demand of the lowest fare class was relatively large (i.e., 40), thus at the higher demand factors tested, mostly the lowest fare class demand was spilled. In the case of other examples where lowest fare classes do not represent such a large proportion of the total demand, at higher demand factors not only the lowest fare class but the next to lowest fare classes will also be involved in spill. This would result in an eventual increase of the average spill fare as the demand factor increases.

Figure 7.3 shows the differences in total flight spill cost estimates from applying Method 1-A and 1-B relative to Method 2 for the same example, for discount ratios of 0.7 and 0.9. Recall that, in Method 1-A and 1-B, the spill estimates of Method 1 were used and were multiplied by the weighted fare and a constant low spill-fare, respectively. In Method 2 spill costs were calculated by using equation (7.4). The general observation is that because of the lower spill estimates, Method 1-B underestimates the spill costs relative to Method 2. Method 1-A, on the other hand, at low demand factors underestimates and at high
demand factors overestimates spill costs. For both methods, at low demand factors, when the spill differences are not significant, spill cost differences are insignificant also.

Figure 7.3: Differences in Spill Cost Estimates Relative to Method 2

Figures 7.3 shows that the magnitude of biases in spill cost estimates for a single flight can be significant. Note that for $d=0.7$, the differences range from a few hundred to two thousand dollars.

Note however, that the differences in spill costs are directly a function of the fares that are associated with the spilled lower-valued fare classes. Note that in the $d=0.7$ case, the bottom fare is $144$. Had the fares in this market been higher, the spill cost differences would also have been larger.
Comparison of Method 1 vs. Method 3

In this section we compare Methods 1 and 3. Recall that in Method 3, demand is represented in a multiple-period form, more realistically capturing booking patterns for the different fare classes. We assumed that the standard deviations of demands are equal to the square root of the mean, both for each fare class and within each period. We used a 7 fare class multiple-period example based on actual airline demand data, as summarized in Table 7.4.

<table>
<thead>
<tr>
<th>Fare Class</th>
<th>Mean</th>
<th>Fare d=0.6</th>
<th>Fare d=0.7</th>
<th>Fare d=0.8</th>
<th>Fare d=0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>10</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>700</td>
</tr>
<tr>
<td>B</td>
<td>13</td>
<td>420</td>
<td>490</td>
<td>560</td>
<td>630</td>
</tr>
<tr>
<td>M</td>
<td>8</td>
<td>252</td>
<td>343</td>
<td>448</td>
<td>567</td>
</tr>
<tr>
<td>H</td>
<td>24</td>
<td>151</td>
<td>240</td>
<td>358</td>
<td>510</td>
</tr>
<tr>
<td>Q</td>
<td>25</td>
<td>90</td>
<td>168</td>
<td>287</td>
<td>459</td>
</tr>
<tr>
<td>K</td>
<td>40</td>
<td>54</td>
<td>117</td>
<td>229</td>
<td>413</td>
</tr>
<tr>
<td>L</td>
<td>30</td>
<td>32</td>
<td>82</td>
<td>184</td>
<td>372</td>
</tr>
</tbody>
</table>

Table 7.4: Data Example 2 fare class means and fare values at different discount ratios

Figures 7.4 portrays the expected spill estimates and Figure 7.5 shows differences in spill estimates between Methods 3 and 1 at different discount ratios as a function of demand factors. We see that, as expected, Method 3 generates higher estimates of passengers spilled than Method 1. The differences in the estimated total number of spilled passengers, as before, are greater at lower discount ratios, (i.e., at higher fare differences) where the booking limits have greater influence. The differences are substantial: For example, for \( d=0.6 \) they range from 2-7 passengers; even for \( d=0.9 \) the differences in the spill estimates are in the range of 2-3 passengers at the lower demand factors tested. Note that in this comparison, the maximum differences are reached not at the highest demand factor as was the case in the previous comparison but at demand factors between 0.95 and 1.15. For example, in the case of \( d=0.9 \), the maximum difference of 3.16 passengers is reached at demand factor 0.95.

We also evaluated the differences in spill estimates for the same demand data but assuming that the standard deviations of fare class and booking period demands are twice the square roots of the means. The total spill estimates were naturally higher in this case because of the increased variability in demand. However, the magnitudes of the differences stayed around the same value, changing only slightly. This implies that even with the increase of variability and hence spill, the difference in estimates did not increase significantly.
Figure 7.4: Expected spill estimates for Data Example 2

Figure 7.5: Spill differences for Method 3 vs. Method 1

Figure 7.6 shows the resulting differences of the spill cost estimates between Method 3 (M3), Method 1-A (M1-A), and Method 1-B (M1-B), at discount ratios of 0.7 and 0.9. The differences were calculated by subtracting the spill cost estimates of Method 1-A and Method 1-B from the estimates of Method 3. We see that the spill cost estimates of the different methods can differ by a substantial amount. In our example, the differences are in the range of several hundred to more than a thousand dollars. The results
show that Method 3, which we believe provides more correct and realistic estimates, provides higher spill cost estimates compared to Method 1-B (Constant Low Fare Spill Cost method). Compared to Method 1-A (Weighted Average Spill Cost method) at lower demand factors Method 3 provides higher spill cost estimates, at lower demand factors it provides lower estimates. The results also show that the intersection point (where the positive biases turn into negative biases) is a function of the discount ratios.

Figure 7.6: Differences in spill cost estimates Method 3 minus Method 1

Because Method 1 always underestimates spill relative to Method 3, Method 1-B always underestimates spill costs relative to Method 3. Naturally, the differences in cost estimates are large when the differences in spill estimates are also large. We have already observed that, as the discount ratio $d$ increases, the differences in spill estimates between Method 3 and Method 1-B are larger at lower demand factors. We
can observe the biggest differences in the spill cost estimates in range of demand factors between 0.9 and 1.1. This is an important observation, implying that it is at relatively low demand factors that the biases are actually the most significant.

For example, Figure 7.6 shows that, for $d=0.7$, the maximum difference (from Method 1-B) is reached at $DF=1.05$, where the bias is $686$. The spill cost estimate of Method 3 at this point is $2408$, corresponding to a 28% underestimation of spill costs. In the case of $d=0.9$ the maximum spill cost differences occur at the even lower $DF=0.96$ value. The corresponding bias of Method 1-B in this case is 44%. (Note again, had the fares in this market been higher, -- maintaining the same discount ratios --, the dollar value of spill cost differences would also have been larger.)

**Comparison of Method 2 vs. Method 3**

If both Method 2 and Method 3 provide more realistic estimates of spill and spill costs under yield management, is there a significant difference between the two methods? Is the additional complexity of Method 3 relative to Method 2 justified in terms of obtaining even better spill estimates? By comparing the two multiple fare class approaches we can determine whether the lower-value fare class booking before higher-value fare class assumption is too simplifying. For this comparison, we used the 7 fare class demand data shown in Table 7.5. Once again, we assumed that the standard deviations of demand for each fare class in each period are equal to the square root of the means.

The spill estimates and their differences for this case are shown in Table 7.6 and the spill cost estimates and their differences are shown in Table 7.7. Note that the maximum differences in both the spill and spill cost estimates occur in the range of demand factors around 1.15. Note that the differences between Methods 3 and 2 are as high as $1000$ at certain demand factors.

<table>
<thead>
<tr>
<th>Fare class</th>
<th>Mean</th>
<th>Fare</th>
<th>Discount ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>45</td>
<td>$500</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>460</td>
<td>0.92</td>
</tr>
<tr>
<td>3</td>
<td>23</td>
<td>380</td>
<td>0.82</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>300</td>
<td>0.78</td>
</tr>
<tr>
<td>5</td>
<td>42</td>
<td>260</td>
<td>0.86</td>
</tr>
<tr>
<td>6</td>
<td>57</td>
<td>220</td>
<td>0.84</td>
</tr>
<tr>
<td>7</td>
<td>38</td>
<td>180</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Table 7.5: Data Example 3
From the signs of the spill differences we can conclude that, in this example with realistic booking patterns, Method 2 at low demand factors underestimates and at high demand factors overestimates the total spill -- and more importantly -- the spill costs of a flight. Analysis of other examples has shown that the point where underestimation turns into overestimation can occur at different demand factors and is a function of the prevailing fare ratios (and booking limits). It is also important to note that the differences occur in both directions and in different magnitudes. Thus, we cannot infer a general rule about the biases of the different methods, (i.e., we cannot state that, for example, Method 2 always overestimates or underestimates the real cost nor can we assume that the biases at all demand factors are the same). What the results show is that for each demand example and fare structure, the estimation biases depend also on the demand factors.

The comparisons of alternative methods for spill and spill cost estimation presented in this section have illustrated differences between the estimates under current practice (represented by Method 1) and those based on the new models introduced in this paper. In the examples used for these comparisons, we have seen that Method 1 can underestimate the number of passengers spilled on a flight 3 to 5 passengers and, in turn, underestimate the total spill costs of a given aircraft capacity by $1000 to $1500, when compared to the much more detailed approach of Method 3. These under-estimation biases of the current practice occurred in our examples for representative discount ratios of 0.7 to 0.9, and demand factors of 0.95 to 1.0, which are not at all unreasonable for an airline’s heaviest flights.
7.4 Horizontal Aggregation Bias

We defined horizontal aggregation as the practice of aggregating flight demand over a longer time period. If demand is represented as the average demand of a longer time interval, then the variation in the average daily passenger demand is attributed not only to random variations, but also to differences in day-of-week demand variations, seasonality, and other cyclic effects. In order to concentrate on the effects of the horizontal aggregation bias only, in the beginning of this section we assume that only one fare class is offered on the flight. This way the vertical aggregation bias will not influence our analysis. Later, in the next section, we will assume the more realistic case, when several fare classes are assumed.

The reason that aggregation bias exists is that the expected spill of the “average-aggregated” demand function is not equal to the average of the expected spill estimates of individual demands:

\[ \text{Spill}[\bar{N}(\bar{\mu}, \bar{\sigma}^2)] = 1/n \left( \sum_i \text{Spill}[N_i(\mu_i, \sigma_i^2)] \right) \]  

(7.9)

where \(\bar{N}(\bar{\mu}, \bar{\sigma}^2)\) represents the probability density function of the average aggregated demand. The two sides of the equation is equal only if the spill estimation function were a linear function.

7.4.1 Evaluation of Horizontal Aggregation Bias

We assume that in practice, airlines generate the “average” single demand density for a flight leg over a longer time-period by pooling all individual daily observations relevant for the planning period together. Airlines pool together and combine observations of a flight over a longer time period, and use the statistics of the pooled observations as parameters for the average normal demand representing that period. For example, for estimating and forecasting demand for a flight over a longer time-period, all historic observations that are relevant for the time-period are combined. Forecasts are made using the combined observations as a basis.

In the following we will distinguish between two different approaches. The first approach, which we call as the aggregated or combined approach represents the above outlined practice of the airlines. We will refer to the demand density that is associated with this approach as combined demand density. The other approach is a disaggregated approach, where estimation is based on daily disaggregated demand densities, and the average estimate over the longer time-period is obtained as the average of the disaggregated daily estimates. For the disaggregated approach, it is possible to classify daily demands into classes that can be characterized by similar demand patterns. For example one classification can be day of week classification. In this case the demand over a longer time interval can be classified into day-of-week demand classes. (Note that this day-of-week representation is a disaggregated representation relative to the fully aggregated representation used in practice.)
To model the practice and make possible to compare the results of the aggregated approach to a disaggregated approach, let us assume that the actual demand densities for a flight for each individual day of the time-period are forecasted correctly and known. If this is the case, then based on the individual (daily or aggregated day of week) demand densities, the airlines' practice of aggregating individual demand for the purpose of fleet assignment can be modeled by a finite mixture model\(^3\). We will refer to the probability distribution of the finite mixture model as the combined demand density. Based on the finite mixture model, the two statistics (mean and variance) that serve as the parameters of the combined average normal density can be calculated by the following way\(^4\). Assuming that \(\mu_i\) is the mean and \(\sigma_i\) is the standard deviation of an individual \(i\) day (or aggregated \(i\) day of week) demand, the mean \(\tilde{\mu}\) and variance \(\text{var}(\tilde{x})\) of the combined demand distribution (pooled observations) are given by:

\[
\tilde{\mu} = \sum_i c_i \mu_i, \tag{7.10}
\]

\[
\text{var}(\tilde{x}) = \sum_i c_i \text{Var}(x) + \sum_i c_i (1 - c_i) E_i^2(x) - 2 \sum_{j=1}^{n-1} c_i c_j E_i(x) E_j(x)
\]

or

\[
\text{var}(\tilde{x}) = \sum_i c_i \sigma_i^2 + \sum_i c_i (1 - c_i) \mu_i^2 - 2 \sum_{i=1}^{n-1} c_i c_j \mu_i \mu_j \tag{7.11}
\]

where \(c_i\) represents the weight of demand \(i\) in the pooled observation and \(\sum c_i = 1\). Note, that the above formula is valid for any type of distribution. If all days are represented equally in the pooling, then \(c_i=1/n\) for all days \(i\), where \(n\) is the number of distributions (days) in the aggregation. For this case the above formula can be reduced to:

\[
\text{var}(\tilde{x}) = 1/n \sum_i \text{var}(x) + 1/n^2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (E_i(x) - E_j(x))^2 \text{ or}
\]

\[
\text{var}(\tilde{x}) = 1/n \sum_i \sigma_i^2 + 1/n^2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (\mu_i - \mu_j)^2 \tag{7.12}
\]

We assume that daily demands can be described correctly by a normal demand density \([37], [35]\). Note, that if we pool the observations of the individual demand densities together, the combined data does not approximate a normal density. Thus the assumption of normality is only an approximation here\(^5\).

We can see from equation 7.12 that the variance of the combined demand density is always equal to or bigger than the mean of the variances. That the combined variance will always be bigger than the mean

\(^3\)The finite mixture model is a combination of different probability densities where each density is represented by a relative weight that add up to zero.

\(^4\)Though the combined (pooled) distribution will not necessary follow a normal distribution, we follow the practice and assume that it approximates a normal probability density function.

\(^5\)It would approximate the Normal distribution if we added up the observations
of the variances is an interesting result, though we still cannot say anything how the final direction of the bias in the spill and spill cost estimation procedure will be affected.

Examples of flight leg demands for two different days of the week are given in Table 7.8. Using formulas 5.4 and 5.6, the mean and standard deviation of the combined demand density are given also in the table.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 1</td>
<td>120</td>
<td>10.95</td>
</tr>
<tr>
<td>Day 2</td>
<td>180</td>
<td>13.41</td>
</tr>
<tr>
<td>Combined</td>
<td>150</td>
<td>32.4</td>
</tr>
</tbody>
</table>

Table 7.8: Data for two days and the mean and standard deviation of the "combined" density

Remember that the standard deviation of the combined demand density is bigger than the average of the two original standard deviations, because it also includes the variability of the mean demands. Figure 7.7 shows the two independent days' spill estimates (curve spill_1 and spill_2), their mean (curve avg_spii:!!) and the spill estimates of the combined density (curve Combined_Spill). Figure 7.8 shows the differences in the spill estimates (disaggregated-combined).

![Figure 7.7: Spill estimate curves for two individual days (spill_1, spill_2), their average (avg_spill), and the spill estimate curve based of the combined demand density (CombinedSpill)](image-url)
Figure 7.8: Differences in spill estimates (disaggregated-combined) (avg_spill-CombinedSpill)

Note, that in this example the combined method under capacity 131 and above capacity 172 overestimates spill, and between these capacities underestimates spill. Note, that the biggest underestimation (1.9 passengers) occurs at capacity 151 (close to the average of the means), and the biggest overestimation happens (-1.6 passengers) at capacity 113. Note also that around the mean demand (150 passengers) the spill based on the combined demand distribution underestimates spill by a large amount. The maximum difference does not always occur at the average of the daily means. Our experiments showed, that as the standard deviation increases, the maximum positive difference moves to the right, i.e., larger capacity. However, in the case of realistic standard deviations, (when standard deviation is assumed to be \( c \) times the square root of mean, where \( 1 < c < 3 \)), our results showed that in the neighborhood of the average of the means the spill estimates of the combined demand data underestimate spill; and the maximum underestimation occurs at capacities slightly above to the average of the means.

We expect that as the difference between the original mean demands becomes larger, the estimation bias becomes more significant. Table 7.9 justifies this, showing the maximum over-, and underestimation of the combined demand approach. In the analysis the mean demand for day 1 was kept at 120 and we varied the mean demand for day 2. The analysis was conducted for two scenarios: 1) when we assumed that standard deviation of the daily demand distribution is equal to its mean, 2) when we assumed that the standard deviation of the daily demand is equal to twice its mean.
<table>
<thead>
<tr>
<th>mean demand for day 2</th>
<th>std = sqrt(mean)</th>
<th>max under-estimation</th>
<th>max over-estimation</th>
<th>std = 2 * sqrt(mean)</th>
<th>max under-estimation</th>
<th>max over-estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>130 pax</td>
<td>0.03 pax</td>
<td>0.06 pax</td>
<td>0.01 pax</td>
<td>0.04 pax</td>
<td></td>
<td></td>
</tr>
<tr>
<td>140</td>
<td>0.18</td>
<td>0.23</td>
<td>0.07</td>
<td>0.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>0.49</td>
<td>0.5</td>
<td>0.19</td>
<td>0.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>160</td>
<td>0.91</td>
<td>0.83</td>
<td>0.37</td>
<td>0.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>170</td>
<td>1.39</td>
<td>1.21</td>
<td>0.63</td>
<td>0.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>1.9</td>
<td>1.61</td>
<td>0.95</td>
<td>1.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>190</td>
<td>2.42</td>
<td>2.04</td>
<td>1.32</td>
<td>1.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>2.93</td>
<td>2.48</td>
<td>1.74</td>
<td>1.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>210</td>
<td>3.44</td>
<td>2.93</td>
<td>2.18</td>
<td>2.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>220</td>
<td>3.94</td>
<td>3.39</td>
<td>2.64</td>
<td>2.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>230</td>
<td>4.44</td>
<td>3.86</td>
<td>3.12</td>
<td>3.02</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.5. Spill differences varying mean of day 2 (day 1 mean demand = 120 pax)

Table 7.9 shows that the differences are smaller when we assume twice the square root of the means for the standard deviations. Hence, the differences in the spill estimates between the two methods are smaller at relatively higher standard deviations. This implies that the less variable the individual daily demand are, the more significant the aggregation bias may be. We can explain this as follows. If the variability of the individual demands is small, then we expect that the spill estimates for the individual day’s flights will differ from each other more significantly than in the case when the variability is high. Therefore, in the former case the average of these individual day estimates may deviate more from the estimates of the combined density. Hence, the aggregation is more robust in the case when the variability is high in the original data.

We also found that (assuming the same differences between the original daily means, and assuming that standard deviations are scaled with the square root of means) as the absolute value of the means decreases, the differences in the spill estimates increase. In other words, as the percentage difference in means increases, the aggregation bias also increases.

Figures 7.9 and 7.10 show the differences in the estimates for a 5 day example, assuming standard deviations to be equal to once and twice the corresponding mean demand, respectively. (For calculating the standard deviations for the combined demand data equation 7.12 was used).

<table>
<thead>
<tr>
<th>Day 1</th>
<th>mean</th>
<th>std=sqrt(mean)</th>
<th>std=2*sqrt(mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 2</td>
<td>180</td>
<td>13.41</td>
<td>26.83</td>
</tr>
<tr>
<td>Day 3</td>
<td>100</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Day 4</td>
<td>125</td>
<td>11.18</td>
<td>22.36</td>
</tr>
<tr>
<td>Day 5</td>
<td>180</td>
<td>13.41</td>
<td>26.83</td>
</tr>
<tr>
<td>Combined</td>
<td>157</td>
<td>39.91</td>
<td>45.43</td>
</tr>
</tbody>
</table>

Table 7.10: Data example 4: five day data example with the corresponding “combined” density parameters
The data shows that differences are in excess of 2 passengers at certain ranges. Note that in both cases, around the average of the five single mean demands (157 pax), the combined method underestimates spill and the maximum difference occurs slightly above 150 passengers. The maximum underestimation in the first and second case are 2.1 and 1.18 passengers, respectively.

![Graph](image)

**Figure 7.9:** Spill differences for the 5 day time interval data (disaggregated-combined); standard deviation equals the square root of means

![Graph](image)

**Figure 7.10:** Spill differences for the 5 day time interval data (disaggregated-combined); standard deviation equals twice the square root of means.

As a conclusion we can say, that horizontal aggregation biases exist, although the differences in the given example showed relatively small deviations (1-2 passengers). One of the most important observations is that the biases are in both directions. However, in the practically reasonable range of capacities spill is usually underestimated by the approach that uses the combined demand density. Our results have shown that relatively bigger biases can be observed if the differences in the individual mean demands are big. In other words the biases are big if variability due to cyclical day of week or seasonal effects is high. (Since

136
in our analysis of the horizontal aggregation bias we assumed that only one fare class is offered, spill cost estimates are proportionally related to the spill estimates. In the case of large average fare of spill the differences of the spill cost estimates may be significant.)

Next we will address the cases when vertical and horizontal biases simultaneously affect the spill estimation.

### 7.5 Example of Horizontal and Vertical Aggregation Combined

In this section we present a case where vertical and horizontal aggregation is performed simultaneously. A 5 day long time-interval is considered with a 7 fare class demand flight. The fare class demands were varied over the different days of the week. In our example we assumed that the standard deviation of the fare class demand is always equal to the square root of the mean. Table 7.11 shows the daily demands of our example with the corresponding fares.

<table>
<thead>
<tr>
<th>Fare Class</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fare</td>
<td>Mean</td>
<td>Std</td>
<td>Mean</td>
<td>Std</td>
</tr>
<tr>
<td>Y</td>
<td>700</td>
<td>45</td>
<td>6.71</td>
<td>35</td>
<td>5.92</td>
</tr>
<tr>
<td>M</td>
<td>550</td>
<td>12</td>
<td>3.46</td>
<td>9</td>
<td>3.00</td>
</tr>
<tr>
<td>B</td>
<td>380</td>
<td>23</td>
<td>4.80</td>
<td>15</td>
<td>3.87</td>
</tr>
<tr>
<td>K</td>
<td>266</td>
<td>15</td>
<td>3.87</td>
<td>8</td>
<td>2.83</td>
</tr>
<tr>
<td>G</td>
<td>260</td>
<td>42</td>
<td>6.48</td>
<td>29</td>
<td>5.39</td>
</tr>
<tr>
<td>J</td>
<td>186</td>
<td>57</td>
<td>7.55</td>
<td>41</td>
<td>6.40</td>
</tr>
<tr>
<td>L</td>
<td>120</td>
<td>36</td>
<td>6.16</td>
<td>36</td>
<td>6.00</td>
</tr>
<tr>
<td>Daily Total Demand</td>
<td>232</td>
<td>15.23</td>
<td>173</td>
<td>13.15</td>
<td>179</td>
</tr>
</tbody>
</table>

Table 7.11: Data example 5: five day multiple fare class example with the corresponding combined density parameters

In our analysis we were interested how the spill and spill cost estimates are affected if both vertical and horizontal aggregations exist. In our analysis the spill and spill cost estimates were calculated using the following methods:
• **VH1 (Traditional Combined Approach):** The combined aggregated average demand density for the 5 day multiple fare class data was constructed. First from the daily demand density the daily total demand parameters (vertical aggregation) were calculated (this is shown in the third line from the bottom in Table 7.11). Using the five daily total demand parameters the mean and standard deviation of the combined (pooled) demand distribution are calculated using equations 7.10 and 7.12 (see Table 7.11). Total spill was calculated using the spill formula (equation 3.34) with the parameters of the combined distribution. Based on this spill estimate, spill costs are calculated by the following two approaches:

  • **VH1_C1 (Traditional Approach Using Correct Spill Fares):** Spill cost is calculated by multiplying the above spill estimates by the “correct” average spill fare. Here we assume that the correct average spill fare is known.

  • **VH1_C2 (Traditional Approach Using Constant Low Spill Fares):** Spill cost is calculated by multiplying the above spill estimates by a constant average spill fare, (that is, for all capacities the same spill-fare is used.) This spill-fare is actually the correct average spill-fare at demand factor DF=1.1.

• **VH2 (One Period Disaggregated Approach):** First the total spill and spill cost is calculated for each day assuming that lower-valued fare classes book before higher-valued fare classes. Hence, for each daily demand, equation 7.3 and 7.4 were used. The final average total spill and spill cost are calculated as the average of the five individual daily spill and spill cost estimates.

• **VH3 (Multiple Period Disaggregated Approach):** First the total number of spill and spill cost is calculated for each day considering typical booking patterns. Again for each daily demand the spill and spill costs were calculated, where equation 7.5 and 7.6 were approximated by simulation. The average total number of spill and spill cost for the five day interval was calculated as the average of the five daily spill and spill cost estimates.

In this case we a priori expect that in method VH1, which represents the state-of-the-practice, the vertical and horizontal aggregation simultaneously will bias the results. Based on our previous discussions we expect that the combined method in most of the cases underestimates spill and spill costs. However, remember that at certain capacities spill and spill costs may be overestimated. The effects of the horizontal and vertical aggregation bias appear simultaneously; depending on the directions and the magnitude the biases may add up or cancel out each other. The final result will be the combination of these individual effects. The results for this example follow below.

Table 7.12 shows the spill estimates for methods VH1, VH2, VH3, and it also shows the differences of the estimates. In this example demand factor refers to the average mean demand over the five day period, thus to 207 passengers. All the reported results are calculated at each day at capacities corresponding to this demand factor; the actual capacities are also shown in the table.

138
<table>
<thead>
<tr>
<th>Capacity</th>
<th>DF</th>
<th>VH1</th>
<th>VH2</th>
<th>VH3</th>
<th>diff (VH2 - VH1)</th>
<th>diff (VH3 - VH1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>243</td>
<td>0.85</td>
<td>2.297526</td>
<td>2.764</td>
<td>4.644</td>
<td>0.466474</td>
<td>2.346474</td>
</tr>
<tr>
<td>230</td>
<td>0.9</td>
<td>4.524123</td>
<td>7.78</td>
<td>8.634</td>
<td>3.255877</td>
<td>4.109877</td>
</tr>
<tr>
<td>218</td>
<td>0.95</td>
<td>7.947815</td>
<td>13.216</td>
<td>13.698</td>
<td>5.268185</td>
<td>5.750185</td>
</tr>
<tr>
<td>197</td>
<td>1.05</td>
<td>18.12501</td>
<td>23.996</td>
<td>24.656</td>
<td>5.870986</td>
<td>6.530988</td>
</tr>
<tr>
<td>188</td>
<td>1.1</td>
<td>24.25012</td>
<td>29.596</td>
<td>30.382</td>
<td>5.345882</td>
<td>6.131882</td>
</tr>
<tr>
<td>180</td>
<td>1.15</td>
<td>30.59306</td>
<td>37.038</td>
<td>37.07</td>
<td>6.444944</td>
<td>6.476944</td>
</tr>
<tr>
<td>172</td>
<td>1.2</td>
<td>36.90197</td>
<td>43.724</td>
<td>42.974</td>
<td>6.822031</td>
<td>6.072031</td>
</tr>
<tr>
<td>165</td>
<td>1.25</td>
<td>43.02286</td>
<td>50.226</td>
<td>49.296</td>
<td>7.203139</td>
<td>6.273139</td>
</tr>
<tr>
<td>159</td>
<td>1.3</td>
<td>48.87425</td>
<td>56.01</td>
<td>54.956</td>
<td>7.135752</td>
<td>6.081752</td>
</tr>
<tr>
<td>153</td>
<td>1.35</td>
<td>54.42092</td>
<td>61.616</td>
<td>60.64</td>
<td>7.195084</td>
<td>6.219084</td>
</tr>
<tr>
<td>148</td>
<td>1.4</td>
<td>59.65474</td>
<td>67.502</td>
<td>66.462</td>
<td>7.847258</td>
<td>6.807258</td>
</tr>
<tr>
<td>143</td>
<td>1.45</td>
<td>64.58247</td>
<td>72.38</td>
<td>71.362</td>
<td>7.797526</td>
<td>6.779526</td>
</tr>
<tr>
<td>138</td>
<td>1.5</td>
<td>69.21844</td>
<td>76.574</td>
<td>76.082</td>
<td>7.355555</td>
<td>6.863555</td>
</tr>
</tbody>
</table>

Table 7.12: Spill estimates and differences for the vertical and horizontal aggregation example

The above results show that VH1, the traditional combined method, underestimates the average total number of spilled passengers over the five day period by a substantial amount. As we can see, both methods VH2 and VH3 estimate on average about 5-6 more spilled passengers. Note that the differences are substantial even in the low demand factor cases, where differences are in the range of 2-6 passengers. Thus as we can see that the VH1 traditional combined method, which represents the practice of most airlines, does not capture correctly the average effects of booking limits under variable demand conditions, hence its estimates are biased significantly by both the vertical and horizontal aggregations. Methods VH2 and VH3 have very similar average spill estimates over the five day period. There are only significant differences in the low demand factor case, but for the rest of the cases the differences are small.

The spill cost estimates of the different methods are shown in Table 7.13. Recall that VH1_C1 and VH1_C2 are the two spill cost estimating methods of for the combined demand approach. The spill cost estimates of VH1_C1 and VH1_C2 differ because of the different spill fares (remember the latter uses a constant spill-fare at all demand factors). However, because the variability of the spill fares is not big, the differences in the spill cost estimates (between VH1_C1 and VH1_C2) are not big either.

Method VH2 and VH3 however, produce much higher average spill cost estimates than VH1_C1 and VH1_C2. Table 7.14 shows the actual differences in the spill cost estimates. Note that VH1_C1 and VH1_C2 against both method VH2 and VH3 underestimate spill costs substantially. The differences in estimates are very substantial at high demand factors but also large at relatively low demand factors.

Note also that spill costs are a function of the fares as well. In our example the bottom fares resembled a domestic short or medium-haul market. Had we considered a longer-haul market, (for example where all fares had been twice as big, maintaining the same fare ratios), the spill cost differences would be also twice as large.
<table>
<thead>
<tr>
<th>Capacity</th>
<th>DF (relative to mean demand)</th>
<th>VH1_C1</th>
<th>VH1_C2</th>
<th>VH 2</th>
<th>VH 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>243</td>
<td>0.85</td>
<td>368.01</td>
<td>366.53</td>
<td>627.88</td>
<td>850.93</td>
</tr>
<tr>
<td>230</td>
<td>0.9</td>
<td>793.34</td>
<td>721.74</td>
<td>1336.56</td>
<td>1463.97</td>
</tr>
<tr>
<td>218</td>
<td>0.95</td>
<td>1382.58</td>
<td>1267.93</td>
<td>2070.74</td>
<td>2222.43</td>
</tr>
<tr>
<td>207</td>
<td>1</td>
<td>2053.21</td>
<td>2004.68</td>
<td>2869.75</td>
<td>3001.64</td>
</tr>
<tr>
<td>197</td>
<td>1.05</td>
<td>2855.54</td>
<td>2891.51</td>
<td>3663.86</td>
<td>3799.75</td>
</tr>
<tr>
<td>188</td>
<td>1.1</td>
<td>3868.85</td>
<td>3868.65</td>
<td>4731.11</td>
<td>4839.21</td>
</tr>
<tr>
<td>180</td>
<td>1.15</td>
<td>4924.69</td>
<td>4880.55</td>
<td>5983.88</td>
<td>6020.90</td>
</tr>
<tr>
<td>172</td>
<td>1.2</td>
<td>5771.66</td>
<td>5887.02</td>
<td>7000.28</td>
<td>7004.73</td>
</tr>
<tr>
<td>165</td>
<td>1.25</td>
<td>6786.47</td>
<td>6863.49</td>
<td>8075.65</td>
<td>8077.50</td>
</tr>
<tr>
<td>159</td>
<td>1.3</td>
<td>7709.99</td>
<td>7796.97</td>
<td>9029.71</td>
<td>9037.70</td>
</tr>
<tr>
<td>153</td>
<td>1.35</td>
<td>8729.55</td>
<td>8681.84</td>
<td>10029.83</td>
<td>10054.08</td>
</tr>
<tr>
<td>148</td>
<td>1.4</td>
<td>9752.31</td>
<td>9516.80</td>
<td>11057.55</td>
<td>11082.63</td>
</tr>
<tr>
<td>143</td>
<td>1.45</td>
<td>10621.81</td>
<td>10302.93</td>
<td>11920.59</td>
<td>11952.96</td>
</tr>
<tr>
<td>138</td>
<td>1.5</td>
<td>11657.40</td>
<td>11042.51</td>
<td>12849.11</td>
<td>12912.64</td>
</tr>
</tbody>
</table>

Table 7.13: Spill cost estimates for the vertical and horizontal aggregation example

<table>
<thead>
<tr>
<th>Capacity</th>
<th>DF (relative to mean demand)</th>
<th>VH2 - VH1_C1</th>
<th>VH2 - VH1_C2</th>
<th>VH3 - VH1_C1</th>
<th>VH3 - VH1_C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>243</td>
<td>0.85</td>
<td>259.88</td>
<td>261.35</td>
<td>462.93</td>
<td>464.41</td>
</tr>
<tr>
<td>230</td>
<td>0.9</td>
<td>543.21</td>
<td>614.82</td>
<td>670.62</td>
<td>742.23</td>
</tr>
<tr>
<td>218</td>
<td>0.95</td>
<td>688.16</td>
<td>802.81</td>
<td>839.84</td>
<td>954.50</td>
</tr>
<tr>
<td>207</td>
<td>1</td>
<td>816.54</td>
<td>865.07</td>
<td>948.43</td>
<td>996.96</td>
</tr>
<tr>
<td>197</td>
<td>1.05</td>
<td>808.32</td>
<td>772.35</td>
<td>944.21</td>
<td>908.24</td>
</tr>
<tr>
<td>188</td>
<td>1.1</td>
<td>862.27</td>
<td>862.46</td>
<td>970.36</td>
<td>970.56</td>
</tr>
<tr>
<td>180</td>
<td>1.15</td>
<td>1059.19</td>
<td>1103.33</td>
<td>1096.22</td>
<td>1140.35</td>
</tr>
<tr>
<td>172</td>
<td>1.2</td>
<td>1228.62</td>
<td>1113.26</td>
<td>1233.07</td>
<td>1117.71</td>
</tr>
<tr>
<td>165</td>
<td>1.25</td>
<td>1289.18</td>
<td>1212.16</td>
<td>1291.03</td>
<td>1214.01</td>
</tr>
<tr>
<td>159</td>
<td>1.3</td>
<td>1319.71</td>
<td>1232.73</td>
<td>1327.70</td>
<td>1240.72</td>
</tr>
<tr>
<td>153</td>
<td>1.35</td>
<td>1300.28</td>
<td>1347.99</td>
<td>1324.54</td>
<td>1372.24</td>
</tr>
<tr>
<td>148</td>
<td>1.4</td>
<td>1305.24</td>
<td>1540.75</td>
<td>1330.32</td>
<td>1565.83</td>
</tr>
<tr>
<td>143</td>
<td>1.45</td>
<td>1298.78</td>
<td>1617.66</td>
<td>1331.15</td>
<td>1650.03</td>
</tr>
<tr>
<td>138</td>
<td>1.5</td>
<td>1191.72</td>
<td>1806.60</td>
<td>1255.24</td>
<td>1870.13</td>
</tr>
</tbody>
</table>

Table 7.14: Differences in spill cost estimates for the vertical and horizontal aggregation example

Summarizing, our result showed that simultaneous vertical and horizontal aggregation can result in very significant spill estimation biases. Though theoretically it is possible that biases with different direction may cancel out each other, this example also showed that it is most probably not the case. In most of the cases the biases add up, resulting in more significant estimation biases. Our general example showed that method VH1_C2 that models the state-of-the-practice of airlines may underestimate spill and more importantly spill costs by an amount in the range of $1000 or more. The example also showed that the differences in the spill and spill cost estimates are not uniform. At different demand to capacity ratios the differences vary.
7.6 The Influence of Various Yield Management Systems on Spill and Spill Costs in Networks

Earlier we addressed how yield management systems affect the leg-based spill and spill cost calculations for the single flight leg case, when flight legs are treated as being independent from each other. In this section we will analyze, using the Monte Carlo simulation approach (see Chapter 6), how various yield management control structures and aggregation methods can affect actual passenger flows in a leg-dependent network, and address their implications on spill and spill costs.

We use the simulation approach because the complex interactions of the stochastic ODF itineraries in the networks under yield management control can be captured. Further, many yield management system approaches dynamically update booking limits, depending on the real-time demand to capacity conditions in the network. By embedding the yield management decision process in the simulation, the behavior of the yield management system under different network conditions can be directly captured and its effects on the spill and spill estimates can be evaluated.

In the following we will show that different yield management system approaches can influence differently the leg-independent and the leg-dependent estimates of spill. As a result, we further argue that for estimating spill and spill costs correctly, detailed yield management information should be used and the effects of the yield management approach actually used should be explicitly taken into consideration.

7.6.1 The Modeled YM Approaches

In this section we will address the main characteristics of the modeled YM approaches with respect to their effects on OD passenger flows. (Section 3.2 contains detailed descriptions of the modeled approaches).

Conventional Fare Class (Yield) Based YM (FC)

Under the FC yield management structure, OD fare class demands on each flight leg are aggregated into booking classes according to their associated “fare types”, irrespective of the actual ODF fare values. That is, for example, all Y and M ODF inventories that traverse the flight leg will be mapped into the Y or M booking classes, respectively. Consequently, in this “fare type consistent” system, on a given leg the calculated booking limit that is associated, for example, with the M booking class, controls each M ODF itinerary that traverses the flight leg without the ability of distinguishing among the itineraries.

In the FC approach the revenue\(^1\) value associated with the booking class is usually calculated as the weighted average fare of the total (non-prorated) ODF fares, weighted by the ODF means. Due to the fare

---

\(^1\) For the case when we refer to the revenue value associated with a booking class (or a value classes in the VC and BD approaches) as “revenue”, as opposed to the word “fare” which refers to the ODF fare value.
type aggregation, it is often the case that ODFs with similar fare type but substantially differing total fare values are aggregated into one booking class. That is, in a given booking class local or short haul ODF itineraries with low fare values, and connecting long haul ODF itineraries with substantially higher fare values can belong to the same booking class. Consequently, the revenue value associated with a booking class may represent ODFs with high variance of fares. It is also possible in the FC method that so-called "inverted" booking class revenues result. This is the case when the revenue associated with a higher valued booking class actually has a lower value than that of the next lower valued booking class. In general we can characterize the booking class revenue hierarchy of a flight leg under the FC system as being "narrow" -- which refers to the relatively small difference between the lowest and the highest booking class revenue values, as compared to the difference between the lowest and highest ODF fares traversing the flight leg. Due to the method of aggregation, the lowest booking class revenue values are "pulled up" by the long haul high fare value ODFs belonging to the bottom fare class, while the highest booking class revenue values are "pulled down" by the local or short haul low fare value ODFs belonging to the highest fare class.

The major disadvantage of the yield (fare class) based yield management approach is its lack of OD control. Thus, booking limits affect ODFs in the booking class equally irrespective of their actual revenue potential. If the booking patterns of the different OD demands belonging to the same booking class are similar and spill occurs, then we can assume that the expected OD mix of spill (also the OD mix of the load) will be approximately proportional to the ODF demands. (For the moment assume that capacity constraints of the other legs do not censor the connecting passenger flows). If the booking patterns of the different OD demands belonging to the same booking class differ from each other, then the expected OD mix of spill (and load) will not be proportional. In this case, most likely those OD itineraries which tend to book later in the booking process will be spilled.

In the FC approach similar fare types of different OD itineraries are aggregated into the same booking class, therefore, in most of the cases we do not expect substantial differences among the ODF booking patterns aggregated into the same booking class. That is, for example, we may expect that most of the 7 day advance purchase demands of the different OD itineraries will follow close to similar patterns. We expect significant differences only in cases when the different ODF demands sharing the flight leg belong to very different types of markets, represent different length of haul itineraries, or when different restrictions (advance purchase) apply. For example, domestic discounted M fare class travelers may have different booking behaviors from international M discounted fare class passengers. Often the latter tend to plan and book their flights earlier. It is also possible that the advance purchase requirements differ for domestic and international ODFs, even though they are aggregated into the same M booking class. Consequently, in the case when, for example, the domestic market ODFs tend to book later due to behavioral reasons or due to the less strict advance purchase requirements, then we would see that if M
class demand is spilled it will involve domestic OD demands at a higher proportion than the original proportion of OD means.

If the booking patterns of the different ODF itineraries are similar and other legs do not censor the passenger flows, (i.e., the leg is the bottle-neck leg of the network), then the OD mix of spill on a flight leg will be close to proportional to the OD fare class demand means. This situation is similar to the analysis we have introduced in Chapter 4, but here we also consider the nested structure of the booking limits on each leg.

**Value Class Based YM System (VC)**

In the value based YM approach, ODF itineraries are mapped into value classes according to their total fare values. Due to this aggregation method, the demand parameters and associated revenues of the value classes differ from the FC method\(^2\). Since in the VC approach ODF itineraries are mapped into value classes according to their total fares, the resulting value class revenue hierarchy will be more consistent. Consistency in this case means no inverted revenue values and less variable revenue discount ratios on a flight leg. (Revenue ratios or revenue discount ratios refer to the ratios between revenues of two adjacent fare or value classes). Even though value class revenues may be calculated as weighted average revenues of the ODF inventory fares, due to the way ODFs are aggregated, the weighted revenue value must fall into the fare ranges of the value class. If the value class ranges are defined so as not to overlap, then this would force the aggregated value class revenue hierarchy to be also consistent.

Compared with the booking class revenue hierarchies of the FC method, the revenue hierarchy of the VC method is “stretched”. This refers to the revenues associated with the lowest and highest valued classes. In this method, the revenue associated with the highest and lowest value classes are higher and lower respectively than the revenues associated with the corresponding booking classes under the FC approach. The more value classes that are defined, the finer grained the mapping is and the more stretched the revenue hierarchy becomes.

A very important characteristic of the VC approach is that in typical airline networks mostly local or short haul passengers are being spilled. One reason for this is that local and short haul demands have relatively lower total fares compared to the long haul OD demands, and are mapped into the lower valued classes. The second reason, we assert that in a given value bucket into which local, short, and long-haul ODFs are also mapped, the local and shorter haul ODF inventories are more likely to be spilled. Arguably, in a given class -- where mapped ODFs have similar total fare values -- local (and short haul) OD inventories

\(^2\) They differ mostly because the number of the defined value classes are in most applications bigger than the number of booking classes. But even in the case when the number of the fare and value classes on a flight leg is the same, the demand parameters and revenue values associated with the booking and value classes would be different.
will be represented by high fare type (e.g., Y full fare) ODFs with late booking behaviors and less strict or no advance purchase requirements. On the other hand, the connecting ODFs, belonging to the same value bucket, often represent discounted ODF fare types (e.g., 7 or 14 day advance purchase leisure demand) with early booking behaviors and strict advance purchase requirements. The outcome is that a given value bucket spills passengers, then due to the booking patterns, the later-booking local (or short haul) passengers will be more likely to be spilled.

If spill involves more passengers than local passengers are present in the affected value classes, then naturally connecting passengers must be involved in spill as well. This situation may be present in hub complexes with relatively little local traffic. Further, the fare structures also have important roles. If the fare structure is not consistent, then it is possible that local ODF itineraries will not have relatively lower total fares than connecting ODFs. Although this situation is not typical, it can be found in today's airline prices (e.g., when a fare associated with a flight beyond the hub is lower than the local fare to the hub). In these situations it will not be true that local ODFs are involved in spill to a higher degree. Nevertheless, in many airline networks we can expect significant enough local demands to be present (hubs with strong local demands for example), and also we may correctly assume that in most of the cases local total fares are less than connecting total fares.

If spill involves mostly local demands, then it has an important implication for leg-dependence. Recall that leg-dependence effects arise only if spill involves connecting ODFs. Therefore, if spill mostly involves local demands, then leg-dependence effects should be less than in the FC approach. Due to this different aggregation method, and the preferential treatment of the higher total ODF itineraries, it is expected that the VC approach influences the ODF passenger flows differently than the FC approach. Therefore, it is also expected that the spill and spill cost estimates will be also different from those of the FC approach.

**Displacement Cost Based OD Heuristic YM system (BD)**

In this yield management control approach, initially each ODF is mapped into value classes similar to the VC approach. For evaluation of seat availability, however, the network revenue value for an ODF is calculated as the total ODF fare value discounted by the displacement costs associated with the other flight legs along the ODF's path (see Section 3.2.3). If the displacement cost is high on a connecting flight leg, then the network revenue associated with the ODF will be reduced substantially and thus this ODF will receive availability associated with a lower valued class. Consequently, the ODF belonging the newly assigned lower valued control class may be censored by the prevailing booking limits -- giving priority to ODFs with higher valued network contribution. In most of the cases, this amounts to promoting two local ODF inventories instead of the one connecting ODF. When on both flight legs of a two-leg path seats become scarce, the more optimal network solution is to accept the two local ODF passengers instead of the
one connecting passenger. In general, if local demands are promoted over the connecting demands, then spill will involve connecting demands to a higher degree. Under these conditions, it is expected that the spill and spill cost estimates under the BD approach will differ from those of the VC approach.

If seats are not scarce in the network, or in other words when the demand factors of the flight legs are not very high, then the BD approach will work identically to the VC approach. This would be also the case when the network is asymmetric with respect to demand to available capacity -- that is, when for example mostly either the incoming or the outgoing flight legs to a connecting hub bank have high demand factors. In these cases the optimal network solution is to accept the connecting higher total fare ODF passengers over the local passengers most of the time. This is exactly what the VC method does. Since the displacement costs on the flight legs that do not have high demand factors are low, the network revenues of each connecting ODF remain close to ODF’s total fare value, and hence the BD approach does not readjust the ODF’s revenue value. Consequently, the BD solutions in these cases are similar to the VC solutions. In situations when the average demand factors are low and demand factors are asymmetric on paths, we expect that mostly local ODF passengers are spilled and thus leg-dependence effects do not influence the leg-dependent spill cost estimates as much. In these cases we expect that the spill and spill cost estimates of the VC approach and the BD approach will be similar.

The BD approach, heuristically, is intended to reduce the “greediness” of the VC approach. The approach reallocates ODF’s in the value class booking hierarchy according to leg bid-prices, which are directly a function of the available leg capacities. That is, if the available capacities are changed in the network, the control mechanism of the yield management system responds to it by promoting different ODF itineraries. It is possible to characterize yield management approaches as capacity responsive or capacity non-responsive OD control. The former represents approaches that respond to the capacity changes in the network by changing priorities to different ODF itineraries. The latter represents the approaches where the actual capacity changes do not directly affect the priorities set for the different ODF itineraries. The bid price yield management approach is a capacity responsive system, while the FC and VC approaches are capacity non-responsive. The former is a capacity responsive system because ODF inventory priorities change at different leg demand to capacity levels. The latter are capacity non-responsive, because the main decision variables -- the booking class and value class protection levels -- are not functions of the leg capacity limits. Flight leg capacities enter into the control mechanism as a maximum availability limit in the form of a highest-valued class booking limit. But, the protection levels for the nested booking or value classes are not affected by the available capacities in the network.

It is easier to calculate the ODF passenger flows and spill in non-responsive systems, because the applied protection levels and the aggregations of ODFs to the control classes are predetermined and do not change with different fleeting assignments. In this case, only the interaction of the ODF passenger flows in the network has to be modeled. However, in the BD approach, for different fleeting combinations, not only
the passenger interactions but also the control responses of the yield management system on the demand to capacity ratio changes have to be considered. Therefore, we have to be careful interpreting the differences among the leg-dependent and leg-independent results. We have to understand that the actually observed passenger flows and spills are not only a function of the demand and fare parameters but also the actually assigned capacities on all affected flight legs of the network.

In what follows, we will evaluate the magnitude of the differences in the spill estimates of the different considered YM approaches. Both leg-independent and leg-dependent network situations will be examined.

7.6.2 Leg-independent Spill Estimates Under Different YM Systems
First, let us assume that leg-dependence effects do not influence passenger flows, and let us concentrate our analysis on the leg-independent spill estimates\(^3\). This analysis show that if different YM architectures with different fare class aggregation approaches are used, then spill estimates can differ substantially. This analysis complements the discussion of spill estimation for a flight leg of Section 7.3. It shows that even for estimating spill for a single flight leg, also the underlying YM approach should be taken into consideration.

We outlined that the VC (and BD) approaches have more "consistent" and more "stretched" value class revenue hierarchies than the FC approach. (Note that if leg-independence is assumed, then the VC and BD approaches work identically. Therefore, under leg-independence, all observations about the VC approach hold for the BD approach as well). Because the number of value classes is larger than the number of booking classes in the FC system, and because of the fare value based aggregation technique, the top revenue values are higher and the bottom revenue values are lower in the VC approach compared to the FC approach. This enables the VC system to discriminate on a higher degree between the ODF demands and protect more seats for the higher valued ODF passengers. Further, because of the consistent revenue architecture, no inverted revenues are possible in the VC approach. Based on above observations, we can expect that under the VC approach leg-independent spill estimates will be larger than the leg-independent spill estimates of the FC approach.

Our experimental results on typical airline data proves this. Let us consider the three-leg example shown in Figure 7.11, with the ODF demands and fares given in Table 7.15. Note that the fare discount ratios (\(d\)) within each OD market are set to \(d=0.7\).

---

\(^3\) Leg-independent estimates refer to the case when network effects (i.e., the censoring of connecting ODF demands by other flight legs) are omitted, but yield management information is taken into account.
Figure 7.11: Network example 3L with a short and a medium haul leg feeding the long haul leg

<table>
<thead>
<tr>
<th>OD</th>
<th>AH</th>
<th>BH</th>
<th>AC</th>
<th>BC</th>
<th>HC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fare</td>
<td>mean</td>
<td>fare</td>
<td>mean</td>
<td>fare</td>
</tr>
<tr>
<td>CL 1</td>
<td>10.35</td>
<td>200</td>
<td>9.87</td>
<td>400</td>
<td>10.27</td>
</tr>
<tr>
<td>CL 2</td>
<td>5.2</td>
<td>140</td>
<td>5.28</td>
<td>280</td>
<td>2.16</td>
</tr>
<tr>
<td>CL 3</td>
<td>5.49</td>
<td>98</td>
<td>15.46</td>
<td>196</td>
<td>8.67</td>
</tr>
<tr>
<td>CL 4</td>
<td>10.09</td>
<td>68</td>
<td>15.61</td>
<td>137</td>
<td>12.29</td>
</tr>
<tr>
<td>CL 5</td>
<td>13.9</td>
<td>48</td>
<td>25.58</td>
<td>95</td>
<td>25.97</td>
</tr>
</tbody>
</table>

Table 7.15: ODF demand and fare values for the 5 OD and 5 fare classes -- 3L network example

Tables 7.16 and 7.17 show for the FC and VC approach, respectively, the booking class (value class) demand parameters, associated revenue values, and revenue discount ratios for each leg of the 3L network after aggregation. The first thing to notice is, that in the FC approach the revenue discount ratios after the aggregation, $r$, do not follow the pattern of the discount ratios, $d=0.7$, that were set within each market of the network. This shows that even in a case when an airline uses a common, typical discount ratio across all its markets, the actual flight leg revenue ratios after aggregation may deviate substantially. Table 7.16 illustrates the different revenue ratios for the FC approach. On leg 1, between classes 2 and 3, the phenomenon of inverted revenues can be observed. Note that although the revenue ratios differ significantly on leg 1, on the other two legs the variability is not very significant. We tested the revenue ratios after aggregation on many typical airline networks. Our result show that the revenue discount ratios differ significantly under the FC method. The revenue ratios on a single leg typically varied in the range of $r=0.6..0.85$. Note however, that the bottom revenue discount ratios -- which are the most important with respect to spill -- showed substantial variability across the flight legs ($r=0.6 ..0.96$).
Because the number of value classes in the VC approach is larger than the number of booking classes in the FC approach, and because of the fare value based aggregation technique, the top revenue values are higher (or on leg 3 are equal) and the bottom revenue values are lower in the VC approach compared to the FC approach. Table 7.17 also shows this stretched and consistent revenue hierarchy. There are no inverted revenues presented, and although we can observe variability among the discount ratios of a flight leg, they are more stable across all flight legs. They are mostly in the range of 0.7 and 0.85, because the actual revenue discount ratios are determined by the defined value class ranges, which force the revenue ratios to be consistent. Results in large realistic networks show similar characteristics.

Table 7.16: Booking class demand, revenue, and revenue discount ratio parameters after aggregation in FC approach -- 3L network

Table 7.17: Demand, revenue and discount ratio parameters for value classes on each leg in the VC approach -- 3L network

One of the main implications of the stretched revenue architecture, the finer grained ODF allocation, and the prevailing consistent discount ratios, is that the yield management approach protects more seats for the higher revenue ODFs. Consequently, the total number of passengers spilled under the VC method is
expected to be greater than in the FC method, but at the same time more spill results from the lower value classes.

Table 7.18 shows the leg-independent spill, spill fare, and spill cost estimates for the FC and VC approaches in the 3L network at the following capacity assignment: Leg 1= 105 seats, leg 2= 145 seats, and leg 3= 215 seats. (Leg-independent spill estimates were obtained by assuming no censoring effects on connecting ODF flows -- however, detailed yield management information was incorporated). As we expected, spill in the FC approach is lower, but lower spill values are outweighed by the higher spill fares. Consequently, the spill costs associated with the VC method are lower than the spill costs of the FC method.

<table>
<thead>
<tr>
<th></th>
<th>Total Spill</th>
<th>Total Spill</th>
<th>Spill Fare</th>
<th>Spill Fare</th>
<th>Spill Cost</th>
<th>Spill Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>FC</td>
<td>13.5</td>
<td>21.0</td>
<td>$173</td>
<td>$80</td>
<td>$2384</td>
<td>$1717</td>
</tr>
<tr>
<td>VC</td>
<td>15.5</td>
<td>16.4</td>
<td>214</td>
<td>157</td>
<td>4830</td>
<td>4928</td>
</tr>
</tbody>
</table>

Table 7.18: Comparison of leg-independent estimates in the VC and FC system

We can observe similar result in larger networks as well. Figure 7.12 shows the leg-independent spill estimates in a 15x15 leg hub network (we refer to it as the E-H network -- see Appendix for the network descriptions). In this 30 leg network, the average of the differences were 1.5 passengers. Note however, that the magnitude of the differences are not uniform. On certain flight legs they are larger, on others they are smaller. The average difference for the 25x25 (50-leg) E-C network (see Appendix) were smaller and equal 0.65 passengers. However, on many of the flight legs in this network, differences exceeded two or three passengers.

![Figure 7.12: Leg-independent spill estimates in the E-H network](image-url)
The characteristics of higher spill fares and higher resulting spill costs for the FC approach can be observed also in the two larger networks. Figures 7.13 and 7.14 show the leg-independent spill fares and Figures 7.15 and 7.16 show the spill costs for each flight legs of the E-H and E-C networks, respectively. The figures show that both the leg-independent spill fares and spill costs on almost all legs are substantially higher for the FC system than the VC system.

Figure 7.13: Leg-independent spill fares in the E-H network

Figure 7.14: Leg-independent spill fares in the E-C network
What we have to derive from the above is, that actual yield management system architecture, booking (virtual) class hierarchy, and the underlying aggregation methodology determines the actual spill and spill cost of a flight leg. This observation relates back to the previous sections of this chapter, where we showed that yield management optimization has an influence on spill estimation. In this section, we have
shown that in order to estimate correctly spill for flight legs, the YM architecture and control mechanism that are used by the airline should be modeled. This implies that, it is not enough to use “rule of thumb” types of principles in the spill estimation, but detailed YM information should be used for each flight leg individually.

7.6.3 Comparison of Leg-Dependent Estimates

If connecting ODF passengers are spilled, then leg-dependence effects can arise. Therefore, we expect that leg-dependent spill estimates would differ from the leg-independent spill estimates. In this section we will show that depending on the actual YM system approach, leg-dependent spill estimates can differ from the leg-independent estimates and from each other substantially. These results will indicate that for determining leg-dependence effects on the spill estimates, not only the network connectivity attributes but also the YM approach used should also be taken into consideration.

Let us consider the network example given in Figure 7.11 again, and assume that the following capacities are assigned to the flight legs: Leg 1 = 105 seats, leg 2 = 145 seats, and leg 3 = 215 seats. At the given assignment, the PFF (probability of full flight) values for leg 1 and 2 are much higher than for leg 3, and thus these legs are expected to become full first most of the time. Since flight leg 1 and 2 are both feeding legs, censoring effects should add up on flight leg 3. Consequently, we expect that the achievable traffic density will be censored and “shifted to the left” on flight leg 3, and that the leg-dependent spill estimate on leg 3 will be smaller than the leg-independent estimate. Since we expect that most of the time flight leg 1 and 2 get full before leg 3 and because these legs do not share ODF demands, we also expect that the differences between the leg-dependent and leg-independent estimates will be small on these feeding legs.

Table 7.19 compares the spill estimates of the leg-dependent and leg-independent case in the $3L$ network. As we expected, the differences between the leg-independent and leg-dependent spill for leg 1 and 2 are not large, (1 passenger and couple of hundred dollars), however, the spill estimates for leg 3 differ substantially. Due to the network effects, the leg-independent spill estimation method overestimates the spill in this example by 7.29 passengers. In terms of spill costs, the overestimation is $2191. The leg-dependent estimate is almost 50% lower than the leg-independent estimate. Censored connecting demands not only affect the total number of spilled passengers, but also the OD mix of spill. If ODF itineraries with significantly differing ODF fares are affected disproportionally by censoring effects, then this will affect the OD mix of spill and consequently the spill fares as well. If connecting ODFs are not censored, then assuming similar fare class booking patterns, the OD mix of spill will be approximately proportional to the OD demands. Consequently, the spill fare should be the weighted average of the ODF fares weighted by the OD demands. Theoretically, the bias can be bi-directional, depending on the censoring levels affecting the different ODF itineraries (refer to Figure 4.3 in Section 4.1.2 for explanation). The actual spill cost is the product of the spill and the spill fare. We showed that spill can
be overestimated if leg-dependence is not considered. However, the actual spill fare can either be underestimated or overestimated. The bias will depend in each case on the actual network and demand parameters.

<table>
<thead>
<tr>
<th></th>
<th>Spill</th>
<th></th>
<th></th>
<th>Spill Cost</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Leg 1</td>
<td>Leg 2</td>
<td>Leg 3</td>
<td>Leg 1</td>
<td>Leg 2</td>
<td>Leg 3</td>
</tr>
<tr>
<td>Leg-Independent</td>
<td>13.5</td>
<td>20.5</td>
<td>15.5</td>
<td>$2384</td>
<td>$4412</td>
<td>$4830</td>
</tr>
<tr>
<td>Leg-Dependent</td>
<td>12.5</td>
<td>19.5</td>
<td>8.21</td>
<td>$2082</td>
<td>$3979</td>
<td>$2639</td>
</tr>
</tbody>
</table>

Table 7.19: Comparison of leg-independent and leg-dependent spill and spill cost estimates in the FC system -- 3L network

We analyzed the spill estimates on leg 3 further. We kept the capacities on leg 1 and 2 fixed, (105 and 145 seats respectively), and varied the capacity of leg 3 from 210 seats to 245 seats. Figure 7.17 shows leg-independent and leg-dependent spill estimates at various capacities under the FC approach. The figure shows that in this capacity range, between 215 (DF=0.95) and 245 (DF=0.83) seats, the leg-dependent spill estimates are substantially less than the leg-independent estimates. The largest difference (7.64 passengers) is at capacity 210 seats, but at capacity 245 the difference is still substantial (3.1 passengers).

![Figure 7.17: Leg-independent and leg-dependent spill estimates for leg 3 as a function of capacity under the FC approach (Cap1=105 seats, Cap2=145 seats)](image)

---

4 When the capacity on leg 3 is varied, due to network effects the spill on leg 1 and 2 varies as well. This especially can be the case when the capacity on leg 3 is small. Under the reported capacity values, however, the spill on flight legs 1 and 2 is relatively constant. As the capacity changed from 210 to 245 seats on leg 3, the spill estimates on leg 1 and 2 varied between 12.5..13.4 and 18.9..20.3 passengers, respectively.

153
If we compare the actual spill and spill cost values of the 3L network under the VC approach (see Table 7.20) we can observe that the differences between the leg-dependent and leg-independent spill and spill cost estimates are smaller, as compared to the FC method. The differences in the spill and spill cost estimates on leg 3 in the FC approach are 7.29 passengers and $2191, respectively. The corresponding differences in the VC approach are 0.8 and $440, respectively. The smaller differences can be explained by the tendency that in the VC method mostly local passengers are spilled, which in turn reduces the leg-dependence effects in the network.

<table>
<thead>
<tr>
<th>Spill</th>
<th>Spill Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leg 1</td>
<td>$1717</td>
</tr>
<tr>
<td>Leg 2</td>
<td>$3993</td>
</tr>
<tr>
<td>Leg 3</td>
<td>$5035</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Spill</th>
<th>Spill Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leg 1</td>
<td>$1265</td>
</tr>
<tr>
<td>Leg 2</td>
<td>$3544</td>
</tr>
<tr>
<td>Leg 3</td>
<td>$4595</td>
</tr>
</tbody>
</table>

Table 7.20: Comparison of the leg-dependent and leg-independent spill estimates VC approach -- 3L network

Figure 7.18 shows the leg-dependent spill estimates for leg 3 in the 3L network as a function of the capacity, assuming that the capacities on leg 1 and 2 are fixed at 105 and 145, respectively. The figure shows the leg-dependent estimates for the VC and BD approaches. As we expected for this network example, the leg-dependent spill estimates differ less substantially from the leg-independent estimates. This also supports the expectation that leg-dependence effects can be less significant in some of the networks under the VC architecture. Note also that at lower capacities, the leg-dependent spill estimates differ between the VC and BD approaches. This happens because at lower capacities the BD heuristic starts promoting local passenger over the connecting passengers, which changes the spill levels on the flight legs.

Figure 7.18: Leg-independent and leg-dependent spill estimates for leg 3 as a function of capacity under the VC system (Cap1 = 105 seats, Cap2 = 145 seats)
The above results lead to another important observation about the leg-dependent spill estimates in networks. Above we have shown that leg-dependent spill estimates of different YM system approaches differ to varying degrees from the corresponding leg-dependent estimates. Figures 7.19 and 7.20 show the leg-dependent spill and spill cost estimates, respectively, as a function of the employed YM system. The figures show that the estimates differ from each other substantially. The differences shown in Figure 7.19 range up to 7 passengers. Similarly, the differences between the leg-dependent spill cost estimates of the VC and FC approaches range up to $1700 (see Figure 7.20).

Figure 7.19: Comparison of leg-dependent spill estimates as a function of the employed YM system

Figure 7.20: Comparison of leg-dependent spill cost estimates as a function of the employed YM system
In the following we will show that, in larger typical airline networks, leg-dependent spill and spill cost estimates under different YM approaches can differ from the leg-independent estimates and from each other substantially. Figure 7.21 and 7.22 show the spill estimates for each flight leg in the $E-H$ and $E-C$ networks for a given fleeting, respectively. The results show that leg-dependent spill estimates under the different YM approaches differ considerably. For network $E-H$ for example, the differences are often in the range of couple of passengers on a flight leg and on certain flight legs (e.g., leg 16 and 22) are in the range of 5 passengers. In the $E-C$ network, we can observe similar differences among the spill estimates under the different YM approaches, and on certain flight legs (e.g., on legs 6, 17, and 36) the differences exceed 10 passengers. Note that, most of the time, the differences are greater between the leg-dependent spill estimates under the FC method and the two other methods. As expected, the FC method most of the time spills less passengers, but not in all cases. The spill estimates under the VC and BD approaches are often similar, however, one can find flight legs in both networks on which the three different approaches result in significantly different spill estimates (e.g., legs 10, 22, 23, and 29 in the $E-H$ network).

Figure 7.21: Leg-dependent spill estimates in the $E-H$ network
One measure to determine how much leg-dependence affects spill is to determine the differences between the leg-independent and leg-dependent spill estimates under each YM approach. Large differences suggest more substantial effects, while smaller differences suggest less significant effects. Figures 7.23 and 7.24 show these differences for networks E-H and E-C, respectively, where for each YM approach the differences were calculated by subtracting leg-dependent spill estimates from the leg-independent estimates.
Figure 7.24: Difference (Leg-independent - Leg-dependent) between spill estimates in the $E-C$ network.

The figures clearly show that the largest differences occur under the FC YM approach. In both figures, the points associated with the FC approach on most of the flight legs lie above the other points. As we expected for the VC approach, the differences are much smaller in both networks, meaning that under the VC method leg-dependent effects on spill estimates can be less significant. Although the differences between the spill estimates of the BD and VC approaches are similar, we can observe that on most of the flight legs, the absolute values of the differences are larger for the BD approach than those of the VC approach. The reason that the differences are bigger in the BD approach is that this approach on certain flight legs corrects the "greediness" of the VC approach. In such cases the BD approach spills connecting ODF itineraries to a higher degree. The OD mix of spilled passengers in the network also explains the above observation. Note that in the $E-H$ network, the total number of local ODF passengers spilled under the VC and BD approaches is 199.4 and 185.4 passengers, respectively. In the $E-C$ network, where the differences were systematically lower for the VC approach, total local spill was 795.4 and 741 passengers for the VC and BD approaches, respectively. These data also support the hypothesis that the VC approach spills mostly local ODF itineraries, and therefore, leg-dependent effects are less substantial in this approach compared to the others.

Based on the earlier discussions, we expect that the spill fares for the FC approach will be higher than those of the other two methods. Figures 7.25 and 7.26 show that this is indeed the case in networks $E-H$ and $E-C$, respectively. The figures show that for almost all flight legs, the FC spill fare is higher than the other spill fares. In the $E-H$ network the differences are often a few hundred dollars, while in the $E-C$ network they are a little smaller.
Figure 7.25: Leg-dependent spill fares in the E-H network

Figure 7.26: Leₜ-dependent spill fares in the E-C network

Spill costs, the products of spill and spill fares, are shown in Figures 7.27 and 7.28 for network E-H and E-C respectively. Because higher spill fares are not outweighed by lower spill in the FC approach, the resulting spill costs for most of the flight legs are greater than those of the other two approaches. Note however, that there are certain flight legs for which the relationship is reversed (e.g., legs 11, 18, and 19 in the E-H network and legs 30, and 36 in the E-C network). The differences are modest for certain flight legs, but for some they are very substantial. In both networks, on certain legs the differences are in the
range of several thousands of dollars. Note also that although the differences between the spill costs of the VC and BD methods are small, the magnitudes of the differences vary.

Figure 7.27: Leg-dependent spill cost in the E-H network

Figure 7.28: Leg-dependent spill cost in the E-C network

Summarizing, the simulation results have shown that the number of passengers spilled, the spill fare, and the spill costs for a flight leg are different under various YM approaches. We have also shown that under similar demand and flight leg capacity conditions, leg-dependence under different YM approaches also
affects the spill and spill cost estimates differently. It is also very important to note that the differences in the direction and magnitude are not consistent. The differences on certain flight legs are small, on others very large and they can be bi-directional. This is very important with respect to the fleet assignment decision process.

The results also suggest that the observed network effects have to be evaluated from a network-wide perspective and a flight leg based focus may lead to incorrect estimates of spill and spill costs. This is especially the case for the BD approach, in which the actual ODF itineraries are controlled by the demand to capacity ratios of many flight legs. In this approach, a leg-based focus in the fleet assignment decisions may lead to incorrect conclusions.

As a result, we suggest that spill and spill cost must be estimated jointly for a set of flight legs (sub-network) that are tied together by overlapping ODF itineraries. That is, for these set of flight legs, spill and spill cost should be estimated by taking into consideration the demand and capacity conditions of all affected flight legs. Spill and spill cost for a single flight leg should only be estimated in cases when the flight leg does not carry significant ODF demands or the leg-independence assumption is valid.

### 7.7 Estimating Leg-dependent Spill Costs in Networks for a Longer Time Period

When spill is estimated for a given set of capacity assignments over a longer time period, demand can significantly fluctuate within the time-period. It is possible that, despite the fact that the average demand to capacity ratio (DF) over the time-period does not approach capacity on a flight leg (e.g., DF= 0.8), due to the variability in demand, on certain high demand days the demand to capacity ratio is very high (e.g., DF= 1.2) and on low demand days this ratio is relatively low (e.g., DF= 0.6). In the previous section, we estimated the leg-dependence effects for a sample network for demand that represents a single day, i.e., the data that were used represented ODF demands for a single day (disaggregated demand representation). We have shown that leg-dependence effects can be very significant in these single day demand situations. The results have also shown, that the leg-dependence effects vary and are inconsistent across different DFs. Therefore, it is likely that on days when the average DF is high with respect to a given assignment, spill and leg-dependence effects can be very substantial\(^5\); On days when the ratios are low, however, spill and leg-dependence effects can be small. Not only the average DF for the network can vary, but with a non-uniform variance of the ODF demands over the time period, the DFs of the individual flight leg can vary as well. This may change the interactions of the ODF demands in the network on different days.

---

\(^5\) Since instances (days) with high DF can be also present in the time-period, it is reasonable to consider network situations with very high demand factors and evaluate yield management and network effects for these individual situations. This explains why in the previous analysis high demand factor cases where evaluated.
The result is that on different days, leg-dependence effects can involve various flight legs to a different extent, producing very different leg-dependent spill and spill costs for the network.

To estimate leg-dependent spill and spill costs for leg-dependent networks, it is even more important to estimate leg-dependent spill and spill costs on individual daily (or classified day of week) demand bases. Then the average spill over a longer time period should be obtained by calculating the average of the daily spill and spill cost estimates. We believe that to capture realistically the leg-dependent interactions of the ODF passenger flows in the network under the yield management architecture being used, detailed daily ODF demand densities should be considered. Estimating leg-dependent spill based on a horizontally aggregated “average” combined (or pooled) demand representation, cannot capture these subtle but significant network and YM effects and can lead to biases. (We have shown a similar case for a single flight leg case in Section 7.4.) To estimate the effects of leg-dependence correctly, demand should represent the correct stochastic demand to capacity conditions in the network. In demand that represents longer time period, the associated demand densities differ from the individual demand densities. This may affect the stochastic sequences of flight legs and booking (value) classes becoming full, which in turn, determines the actual leg-dependent spill estimates. To prevent this inaccuracy, first the daily spill estimates for the sub-network should be obtained based on daily demand densities, and averaging should follow suit.

To reduce the computational burden, first the entire network should be clustered into sub-networks, where each sub-network contains flight legs that are tied together by leg-dependence. Spill and spill cost should be estimated and interpreted for a fleeting, taking into consideration the network and YM effects in the sub-network (i.e., taking into consideration capacities and seat availability on all flight legs of the sub-network). To reduce additionally the computational burden, similar daily ODF demand data can be classified or clustered. For example, if the different day of week demands tend to be similar, they can be classified and the demand over the longer time period can be represented by day-of-week demand densities. Spill and spill costs then, should be estimated based on the classified demand data, and the average spill for the time period should be calculated as the average of the individual estimates. Note that daily or classified detailed demand data by class is already utilized by yield management systems and is therefore generally available.

Although the above approach requires additional computational work, it is unavoidable in order to capture the average spill cost in leg-dependent network over a longer time-period. Using the simulation approach described in Chapter 6, and utilizing today’s efficient computer hardware, the additional computational burden can be affordable. (In Chapter 8 we will address this issue again).
7.8 *Fleet Assignment Solutions Under Different Spill Estimation Approaches*

In this chapter, we have shown that spill and spill cost estimates, when network effects and the effects of the applied yield management approach is taken into consideration, can differ from the traditionally obtained spill and spill cost estimates. The ultimate question to ask is whether taking these effects into consideration can and would change the optimal fleet assignment solution.

If it were the case that taking into consideration the above mentioned effects results only in a consistent adjustment in the spill cost estimates (i.e., for example all estimates are over- or underestimated by a relatively constant value), then one could argue that (despite the fact that correct cost values are not used) the actual fleet assignment solutions (fleeting) would not be affected. However, as we have shown in this chapter, the effect of leg-dependence and the YM decisions changes the spill cost estimates across the flight legs, but not consistently. Depending on the actual demand to capacity ratios (DF) and other demand and network parameters, the magnitude of these effects on the estimates may differ. At lower demand to capacity ratios (DF), where spill and leg-dependence effects are less, the differences are smaller, at higher ratios the differences can be also larger. Similarly, depending on the magnitude of the censoring effects taking place on neighboring flight legs, network effects influence the actual spill estimates inconsistently. This inconsistent nature of the differences (biases) in the estimates in typical airline network situations can result in different fleeting solutions compared to the fleeting solution of the traditional approach in which yield management and network effects are ignored.

In the following, using actual airline data, we will show some examples in which taking into consideration yield management and leg-dependence effects result in different "optimal" fleeting solutions than in the case when these effects are ignored. Using these examples we did not intend to evaluate completely the effects of leg-dependence and yield management in airline networks. The goal of these examples is only to show that in typical airline demand and network situations, considering yield management and network effects can often result in different (better) fleet assignment solutions which in turn contribute to additional cost savings, or alternatively, increased profits. First, we will consider the cases where we assume that the network is leg-independent. This allows, the effects of yield management approaches on the fleet assignment solution to be isolated. Second, we will consider the cases when network and yield management effects simultaneously affect the spill estimates and will evaluate their effects on the actual fleet assignment solution.

7.8.1 *Evaluation of the Different Spill Cost Estimates on Fleeting Solutions in Leg-independent Networks*

In this section, we assume that leg-dependence effects do not affect the spill estimates. That is, we assume that all ODF demands are accommodated on local (non-stop) flights. Recall that fleet assignment
decisions are made by considering simultaneously the operating costs and spill costs (or expected revenues) associated with assigning particular aircraft types to flight legs. The total cost associated with assigning a fleet type to a flight leg is the sum of the operating and spill costs. Generally we assume that, the larger the aircraft, the higher is its operating costs. Spill cost, on the other hand, increases in the other direction. The smaller the aircraft, the higher the spill and the higher the spill cost. Figure 7.29 shows the operating costs (we assume for simplicity that the aircraft operating cost on a given flight leg increase linearly as a function of capacity), the spill costs, and the total costs (the sum of the operating and spill costs) for a typical flight leg as a function of the aircraft capacity. As the figure shows, depending on the assumed capacity ranges, the total costs on the flight leg may be decreasing or increasing.

Note that the actual spill cost curve is a function of the actual demand conditions on the flight leg, while the operating cost curve is generally independent of the demand. If the mean demand on the flight leg is relatively lower, then the spill cost curve is shifted to the left -- if it is higher, then is shifted to the right. Between a pair of fleet types (pair of capacity values) the marginal costs may be increasing or decreasing as a function of DF. The fact that operating costs are independent of demand also implies that the portion of the total cost curve estimates that can be biased occurs where the ratio of the spill costs relative to the operating costs is relatively high. This characteristic, in addition to our earlier analysis, indicates that the relative estimation biases are expected to be inconsistent across the capacity levels on a flight leg but will be a function of the actual DFs.

![Figure 7.29: Spill costs, operating costs and the total costs for a flight leg](image)

---

6 Although this is not always the case in reality. Because of technological innovations and other operational advancements (e.g., fewer required crew members), it may be the case that operating a newer technology larger aircraft on a flight leg costs less than operating an older technology smaller aircraft.

7 Ignoring the passenger variable costs.
In the following, we will show on a simple two leg example that the not uniform differences in spill cost estimates can lead into sub-optimal fleeting selections when yield management effects on spill are ignored.

Let us consider two (independent) flight legs with slightly different distances (see Figure 7.30) and assume that two different aircraft types are available: Fleet type $F1$ with capacity 197 seats and $F2$ with capacity 218 seats. Also assume that the mean demand on flight leg 1 is 206 passengers and on leg 2 is 185 passengers, and the fare structure and fare class demand characteristics on the two legs are similar.

![Figure 7.30: Example A -- Leg-independent two-leg example network](image)

Figures 7.31 and 7.32 show the estimated spill costs using two different approaches for the two fleet types on leg 1 and leg 2, respectively. Spill Cost 1 refers to the estimate obtained by the Traditional Approach Using Constant Low Spill Fares (VH1_C2), and Spill Cost 2 refers to the cost estimates of the Multiple Period Disaggregated Approach (VH 3) which takes into consideration yield management and lacks the biases due to horizontal aggregation (for the definitions of the method refer to Section 7.5). The figure also shows the operating costs for the fleet types on both flight legs, as well as the total costs, which are the sum of the operating and spill costs.

The major difference in the two figures is that the spill cost curves on Figure 7.31 are associated with a higher DFs relative to Figure 7.32. This is due to the relatively higher mean demand on flight leg 1 compared to that on leg 2. Therefore, at the given fleet capacity levels, leg 1 spills more and has higher spill costs than leg 2. The figures also show that the estimation biases are not uniform at different demand to capacity ratios. The biases are relatively constant on Figure 7.31 between the two fleet capacities (197 and 218) but vary in Figure 7.32. (The actual cost values are shown in Table 7.21.)

---

*Note that the estimation biases represented in this example can be similar in cases when spill costs are estimated for a single flight: departure only and horizontal aggregation does not disturb the estimation. The assumed magnitude of the biases can be representative for those cases as well.*
Table 7.21: Operating Cost, Spill Cost and Total Cost for leg 1 and 2 in Example A

<table>
<thead>
<tr>
<th></th>
<th>LEG 1</th>
<th>LEG 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity</td>
<td>197 seats</td>
<td>218 seats</td>
</tr>
<tr>
<td>Op_Cost</td>
<td>$5588</td>
<td>$6134</td>
</tr>
<tr>
<td>Spill Cost 1</td>
<td>$2891</td>
<td>$1267</td>
</tr>
<tr>
<td>Spill Cost 2</td>
<td>$3799</td>
<td>$2222</td>
</tr>
<tr>
<td>Cost 1</td>
<td>$8479</td>
<td>$7401</td>
</tr>
<tr>
<td>Cost 2</td>
<td>$9387</td>
<td>$8356</td>
</tr>
</tbody>
</table>

Figure 7.31: Spill cost, operating cost, and total cost estimates for leg 1 (Cost 1 referring to method VH1_C2 and Cost 2 referring to method VH3)

Figure 7.32: Spill cost, operating cost, and total cost estimates for leg 2 (Cost 1 referring to method VH1_C2 and Cost 2 referring to method VH3)
To select the “best” fleet assignment the following two fleeting combinations need to be evaluated: In fleeting combination I fleet type $F1$ (197 seats) is assigned to leg 1 and fleet type $F2$ (218 seats) is assigned to leg 2, and in fleeting combination II, fleet type $F1$ is assigned to leg 2 and $F2$ to leg 1. The total costs associated with assigning a fleet type to a flight leg under the two different estimation approaches (Cost 1 and Cost 2) and the total fleeting costs (the sum of the total costs associated with the actual fleeting) are shown in Table 7.22.

<table>
<thead>
<tr>
<th>Fleetling Combination</th>
<th>I</th>
<th>II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Capacity</td>
<td>Cost 1</td>
</tr>
<tr>
<td>leg 1</td>
<td>197 seats</td>
<td>8479</td>
</tr>
<tr>
<td>leg 2</td>
<td>218 seats</td>
<td>7274</td>
</tr>
<tr>
<td>Fleetling Cost (sum)</td>
<td>$15753</td>
<td>$16771</td>
</tr>
</tbody>
</table>

Table 7.22: Total costs of assigning a fleet type and fleeting costs using two different approaches

The above table shows that according to the Cost 1 estimates the total fleeting cost associated with fleeting II is less ($15653) than the cost associated with fleeting I ($15753). Therefore, the fleet assignment solution based on this approach would select fleeting combination II over I. According to the Cost 2 estimates, however, the total fleeting costs associated with fleeting II are larger ($17642) than the costs associated with fleeting I ($16771). Consequently, the best fleeting according to the more appropriate Cost 2 estimates is fleeting I. If we assume that Cost 2 estimates represent the correct costs, then the actual savings that we incur by using Cost 2 estimates instead of the Cost 1 estimates in the fleet assignment decisions is the difference of the fleeting costs of fleeting I and II considering the Cost 2 estimates. That is, the savings in this example is $17642-$16771 = $871. Therefore by considering yield management effects in the spill estimation and fleet assignment the incurred savings are almost 5% for this example.

It is important to mention that the biases in spill estimates do not always result in different fleeting solutions. It is possible that although incorrect spill costs are used, the obtained fleet assignment solution is similar to the fleet assignment solution of the approach using the correct spill estimates. Nevertheless, in many typical network situations, the one like the above Example A, the more correct approach in spill estimation can lead to different fleeting solution which in turn can lower the overall costs, or increase profits.

To demonstrate the effects of the spill cost biases in a larger network we conducted a number of analyses in hub-network. Here we demonstrate our daily fleet assignment results with the 30 flight leg $E-H$ hub-network (see Appendix 1 for description of the network). The flight legs in the network are of different length -- the flying times for flight legs varied between 1 and 3 hours. It was assumed that 6 different fleet types were available with capacities of 105, 120, 138, 142, 152 and 170 seats. Spill costs were
estimated using the traditional Constant Low Fare Spill Cost approach (Method 1B) and Multiple Period Multiple Fare Class Approach (Method 3) (see Section 7.3.1 for description). The former represents the traditional spill estimation approach and the latter represents the approach that takes into consideration yield management effects. (No horizontal aggregation was considered in this case).

Using the obtained total cost estimates of the two methods as objective coefficients, the "best" fleet assignment solution subject to the flow balance, cover, and size constraints was calculated by using a similar multicommodity flow integer programming fleet assignment formulation given in equation (3.19). After obtaining the best fleeting combinations in both methods, we compared the differences in the fleeting and in the associated fleeting costs. The fleeting cost for Method 3 assumed to be equal to the objective function value. The total fleeting cost for Method 1B was obtained by considering the fleet assignment solution of Method 1B but for estimating the cost of the fleeting the "correct" costs of Method 3 were substituted. Table 7.23 shows the calculated fleeting costs (and their differences) for the "best" fleet assignment solutions using the cost estimates of the two methods for different fleet cases. The cases represent different fleets, i.e., different number of aircraft of each fleet type available in the fleet. (The second row represents the available number of aircraft in each fleet type in increasing order with respect to the capacity. For example, the first column represents the case when 2 105 seat, 1 120 seat, 3 138 seat etc., aircraft are available in the fleet.) The table show the differences associated with each fleeting cost in absolute and relative terms. These differences represent the actual cost savings resulting from the better fleet assignment solution. The last row of the table shows the number of flight legs that have different fleet types being assigned in the fleet assignment solutions of Method 1B and 3.

<table>
<thead>
<tr>
<th>Case</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fleet</td>
<td>2 3 3 3 2 2</td>
<td>3 1 1 3 3 4</td>
<td>2 2 2 3 3 3</td>
<td>2 4 1 4 2 2</td>
<td>1 2 4 4 2 2</td>
<td>0 5 0 5 0 5</td>
<td>3 3 4 0 3 2</td>
</tr>
<tr>
<td>Method 1B</td>
<td>$197761</td>
<td>175725</td>
<td>182829</td>
<td>203165</td>
<td>192413</td>
<td>179125</td>
<td>205631</td>
</tr>
<tr>
<td>Method 3</td>
<td>$195957</td>
<td>173491</td>
<td>180830</td>
<td>201007</td>
<td>191631</td>
<td>179125</td>
<td>203621</td>
</tr>
<tr>
<td>Difference</td>
<td>$1804</td>
<td>2234</td>
<td>1999</td>
<td>2158</td>
<td>782</td>
<td>0</td>
<td>2010</td>
</tr>
<tr>
<td>% difference</td>
<td>0.91%</td>
<td>1.27%</td>
<td>1.09%</td>
<td>1.06%</td>
<td>0.41%</td>
<td>0.00%</td>
<td>0.98%</td>
</tr>
<tr>
<td>Numb. of Diff. assignm.</td>
<td>12</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>4</td>
<td>0</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 7.23: Fleeting cost for the best fleeting using spill cost estimates of Method 1B and Method 3 and their differences in the 30 leg E-H network.

The results show that under different cases (i.e., assuming a different number of aircraft of each type in the fleet), the fleeting costs associated with the "best" fleet assignment under the two approaches, as well as the their differences, vary. For example in Case 6, the "best" fleet assignments of the two approaches are identical. Although Method 1B uses the biased spill cost estimates, for this case, the obtained "best" fleet assignment is identical to that of Method 3. In other cases however, one can find differences among the fleeting solutions and the associated differences between the fleeting costs are noticeable. For example in Case 1 the two approaches have assigned different aircraft types to 12 flight legs. In Case 5 this
number is only 4. The differences between the fleeting costs under the two approaches are in excess of a couple of thousand dollars. These are the actual savings (or additional profits) for a single connecting bank on a single day that the airline can realize by taking into consideration yield management information and decisions in the fleet assignment decisions. Assuming that the schedule is repeated for a large number of times (weeks or months), the total savings over a longer time-period can add up and be very considerable. Assuming larger airline networks, where the number of such network examples are larger, i.e., many connecting banks are present in the network, we speculate that the possible benefits using the more correct spill cost estimates is very substantial.

7.8.2 Evaluation of the Different Spill Cost Estimates on the Fleeting Solutions in Leg-dependent Networks

In leg-dependent networks, due to censoring effects, leg-dependent spill cost estimates on certain legs can be substantially less than the leg-independent spill cost estimates. In the previous chapters it was shown that the leg-independent approach overestimates spill costs (and revenues) by different magnitudes. For example, on the flight legs that mostly censor connecting demands the differences between the leg-independent and leg-dependent estimates are minor, however, for the other flight legs the differences can be very substantial. This inconsistent relative overestimation again may result in sub-optimal fleeting solutions as compared to the fleeting solution where network effects, by using leg-dependent estimates, are considered.

One of the major impacts of ignoring leg-dependence in the fleet assignment is that a larger than required fleet type can become assigned to a flight leg while that aircraft could be used more profitably on an alternative leg. In the following two-leg network example we present this situation. Assume that the airline needs to select a fleet type to serve the two-leg network segment shown in Figure 7.33. In the network, local and connecting ODF demands are also defined and are given in Table 7.24.

![Figure 7.33: Example B -- Two-leg leg-dependent network example with local and connecting OD demands](image-url)
We estimated the leg-independent and leg-dependent spill costs using the simulation approach outlined in Chapter 6 assuming that fare class based (FC) yield management approach being used. Leg-independent estimates were obtained by taking into consideration detailed yield management information, but censoring effects on connecting ODF demands were ignored. Table 7.25 shows the leg-dependent and leg-independent spill cost estimates for the two flight legs and the two fleet types.

<table>
<thead>
<tr>
<th>Capacity</th>
<th>182</th>
<th>204</th>
</tr>
</thead>
<tbody>
<tr>
<td>leg-indep. LEG 1</td>
<td>$3460</td>
<td>$1309</td>
</tr>
<tr>
<td>leg-dep. LEG 1</td>
<td>2850</td>
<td>1211</td>
</tr>
<tr>
<td>leg-indep. LEG 2</td>
<td>4664</td>
<td>2603</td>
</tr>
<tr>
<td>leg-dep. LEG 2</td>
<td>3226</td>
<td>2400</td>
</tr>
</tbody>
</table>

Table 7.25: Leg-dependent and leg-independent spill cost estimates for leg 1 and leg 2 for two fleet types

Table 7.26 shows the total spill costs for each fleeting under the leg-dependent and leg-independent approaches. (For example, the total spill cost for the leg-independent case when the 182 aircraft is assigned to both flight legs is the sum of the spill costs on the two legs associated with a fleet type, i.e., $3460+4664=8124.) The table also shows the estimated marginal cost savings in spill (difference) associated with utilizing the larger aircraft on both flight legs.

<table>
<thead>
<tr>
<th>Leg-Dependent</th>
<th>Leg-Independent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity = 182</td>
<td>$6076</td>
</tr>
<tr>
<td>Capacity = 204</td>
<td>$3611</td>
</tr>
<tr>
<td>Difference</td>
<td>$2465</td>
</tr>
</tbody>
</table>

Table 7.26: Spill costs and their differences for different fleet type assignment to both flight legs

The results in the table show that the leg-independent spill cost difference, the estimated marginal cost saving, is overestimated significantly. The leg-independent approach double counts the spill of the connecting ODF demands and its expected marginal savings ($4212) is significantly larger than that of the leg-dependent approach ($2465). It incorrectly assumes that a connecting ODF demand is spilled on both flight legs, but as we already addressed it once an ODF passenger once spilled on a flight leg cannot
appear and be spilled on the other flight leg of its path. This double counting may result in sub-optimal fleet type selection.

First, for example, assume that the additional (marginal) operating cost for assigning the larger 204 seat aircraft instead of the smaller 182 seat aircraft to both flight legs is $3718, (if the two flight legs are relatively long flights then this additional cost is very realistic). According to the leg-dependent approach the additional operational costs outweigh the cost savings in spill ($2465-$3718=$-1253). However, according to the leg-independent approach the opposite conclusion prevails ($4212-$3718 = $494). Consequently, the leg-dependent approach would assign the 182 seat aircraft to the flight legs, while the leg-independent approach would assign the larger 204 seat aircraft. The actual savings in the overall costs (or profits) that the airline can realize by using the leg-dependent cost estimates is $1253.

Next, we will introduce a 4 leg hub example and show that depending on the yield management approach being used different fleeting solutions can be optimal. Assume the network situation when two shorter flight legs feed two long-haul flight legs (see Figure 7.34). In the network it was assumed that substantial connecting demand was present between city-pairs A-C and B-C, while the connecting OD demands between city-pairs A-D and B-D, although existing, were relatively little. Consequently, the leg-dependence effects between legs 1, 2 and leg 3 are much stronger than among legs 1, 2 and 4. Therefore, censoring effects that take place on legs 1 and 2 are expected to affect legs 3 and 4 not uniformly. It was also assumed that for flight legs 1 and 2 a 105 and a 145 seat fleet type were available, and for legs 3 and 4 a 210 and a 235 seat aircraft were available.

![Figure 7.34: Example C -- 4 leg hub network with leg-dependent flight legs](image)

The 4 different fleeting combinations and the associated operating costs are given in Table 7.27.

<table>
<thead>
<tr>
<th>Fleeting</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leg 1</td>
<td>105</td>
<td>145</td>
<td>145</td>
<td>105</td>
</tr>
<tr>
<td>Leg 2</td>
<td>145</td>
<td>105</td>
<td>105</td>
<td>145</td>
</tr>
<tr>
<td>Leg 3</td>
<td>210</td>
<td>210</td>
<td>235</td>
<td>235</td>
</tr>
<tr>
<td>Leg 4</td>
<td>235</td>
<td>235</td>
<td>210</td>
<td>210</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OPERATING COSTS</th>
<th>Leg 1</th>
<th>Leg 2</th>
<th>Leg 3</th>
<th>Leg 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$1598</td>
<td>$2118</td>
<td>$2118</td>
<td>$1598</td>
</tr>
<tr>
<td></td>
<td>$2118</td>
<td>$1598</td>
<td>$1598</td>
<td>$2118</td>
</tr>
<tr>
<td></td>
<td>$2963</td>
<td>$2963</td>
<td>$3288</td>
<td>$3288</td>
</tr>
<tr>
<td></td>
<td>$6576</td>
<td>$6576</td>
<td>$5926</td>
<td>$5926</td>
</tr>
</tbody>
</table>

Table 7.27: The four different fleeting combinations and associated fleeting cost for Example C
Table 7.28 shows the spill costs for each leg under the different fleeting combinations under the leg-dependent and leg-independent approaches assuming that yield based (FC) yield management system is being used. (Note that in this case leg-dependent estimates represent the *Multiple Period Multiple Fare Class Approach* (Method 3), that is detailed yield management information is considered and only the leg-dependence effects are ignored. The leg-dependent estimates are obtained using the simulation approach outlined in Chapter 6.) The table also shows the total fleeting costs and their ranks (the fleeting with the lowest fleeting costs being ranked as first) for the leg-dependent and leg-independent approach. Note that according to the leg-dependent approach fleeting I is being ranks first, followed by fleeting IV, II, and III. On the other hand, in the leg-independent approach the relative ranks of the fleeting are entirely different -- fleeting IV is ranked first, followed by fleeting I, III, and II. These results also imply that had leg-dependent spill cost estimates been used in the fleet assignment decision, fleeting I would have been selected. On the other hand, had leg-independent estimates been used, fleeting IV would have been selected. The benefit of using leg-dependent estimates in this example is the actual differences in fleeting costs between fleeting I and IV, where the fleeting costs are evaluated using leg-dependent costs. In this example the savings (or the expected additional profits) of using leg-dependent estimates is $25163-$24098 = $1065, or 4.2%. Again assuming that the schedule is repeated for a larger number of days this savings on a long run can be substantial. In addition, large networks may include many connecting complexes, in which considering leg-dependence effects can lead to better fleet assignment solutions.

<table>
<thead>
<tr>
<th>LEG-DEPENDENT</th>
<th>Total Cost (Spill C. + Op. Cost)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fleeting:</td>
<td>I</td>
</tr>
<tr>
<td>Leg 1</td>
<td>1777</td>
</tr>
<tr>
<td>Leg 2</td>
<td>3768</td>
</tr>
<tr>
<td>Leg 3</td>
<td>3498</td>
</tr>
<tr>
<td>Leg 4</td>
<td>1800</td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td><strong>$10843</strong></td>
</tr>
<tr>
<td><strong>Rank</strong></td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LEG-INDEPENDENT</th>
<th>Total Cost (Spill C. + Op. Cost)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fleeting:</td>
<td>I</td>
</tr>
<tr>
<td>Leg 1</td>
<td>2301</td>
</tr>
<tr>
<td>Leg 2</td>
<td>4279</td>
</tr>
<tr>
<td>Leg 3</td>
<td>5834</td>
</tr>
<tr>
<td>Leg 4</td>
<td>1972</td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td><strong>$14386</strong></td>
</tr>
<tr>
<td><strong>Rank</strong></td>
<td>2</td>
</tr>
</tbody>
</table>

Table 7.28: Spill cost and total costs (spill cost + operating cost) for different fleeting combinations in Example C assuming FC yield management approach.

Next, we estimated the same costs for the case when it was assumed that a value class based (VC) yield management approach was in use. Table 7.29 show the estimates similar to the FC case. Note that the leg-independent spill cost estimates are lower in this case for the reasons explained in Section 7.6.2. Note
also that in this case the rank of the two approaches are also similar. In addition, if we compare the leg-dependent results of the FC and VC approaches then we can discover that the optimal fleeting under the FC yield management approach is fleeting I and under the VC system it is fleeting IV. This example shows that depending on the actual yield management architecture not only the spill costs (both leg-dependent and leg-independent) but also the optimal fleeting solutions can differ from each other. This is even the case when leg-dependent solutions are considered.

<table>
<thead>
<tr>
<th></th>
<th>Spill Cost</th>
<th>Total Cost (Spill C.+ Op. Cost)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
</tr>
<tr>
<td><strong>LEG-DEPENDENT</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leg 1</td>
<td>1118</td>
<td>17</td>
</tr>
<tr>
<td>Leg 2</td>
<td>3423</td>
<td>8535</td>
</tr>
<tr>
<td>Leg 3</td>
<td>5287</td>
<td>5059</td>
</tr>
<tr>
<td>Leg 4</td>
<td>1806</td>
<td>1678</td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td>$11634</td>
<td>$15289</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LEG-INDEPENDENT</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leg 1</td>
<td>1678</td>
<td>153</td>
</tr>
<tr>
<td>Leg 2</td>
<td>3896</td>
<td>9137</td>
</tr>
<tr>
<td>Leg 3</td>
<td>5608</td>
<td>5608</td>
</tr>
<tr>
<td>Leg 4</td>
<td>1901</td>
<td>1901</td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td>$13083</td>
<td>$16799</td>
</tr>
</tbody>
</table>

Table 7.29: Spill cost and total costs (spill cost + operating cost) for different fleeting combinations in Example C assuming VC yield management approach

The fact that under the VC approach the leg-independent and leg-dependent optimal fleeting were similar, does not mean that this is always the case. During our analysis we have encountered a number of cases, when under the VC approach leg-dependence effects resulted in different optimal fleeting solutions compared with the leg-independent solution.

We would like to emphasize again that when network and yield management effects are considered in the spill estimation the resulting fleet assignment solution will not always differ from the traditional approach’s solution where these effects are ignored. However, as we have demonstrated in the above examples, in typical network situations there is a high potential that by considering network and yield management effects better fleet assignment can be obtained which can result substantial cost savings, or alternatively, in higher profits.

## 7.9 Conclusion -- Recommended Spill and Spill Cost Estimation Approach

In this chapter, we have examined the impacts of airline yield management systems on the process of spill and spill cost estimation for a single flight leg. We have demonstrated that, depending on the estimation
method used, the resulting spill estimates can differ significantly. If the fare differences between fare classes are substantial, (i.e., the discount ratios small), then the prevailing booking limits can influence the total number of passengers spilled significantly. At a discount ratio of 0.7, which is representative of the fare structures of US airlines, our results have shown that the current approach being used by airlines (Method 1) can underestimate spill at high demand factors by as many as 4-5 passengers.

The booking patterns and interspersed sequences of requests for different fare classes, which are captured in yield management databases as a multi-period representation of demand, also influence both the total number and fare class mix of spilled passengers. We demonstrated that the fare class mix of spill is not constant at various capacities, even for a given demand pattern.

We showed in this chapter that the different control architectures of the various yield management system approaches may additionally affect leg-independent spill costs. This is due to the fact in the different ODF aggregation strategies, the booking or value class demand parameters and associated revenue values may differ significantly. The prevailing revenue discount ratios, which have significant influence on the booking limits and spill, may differ from the assumed typical market-wide fare discount ratios and, for the purpose of spill estimation, should be evaluated for each flight leg individually. This conclusion means that ad-hoc adjustments to either the capacity or the spill-fare based on general rules of thumb are unlikely to provide accurate spill and spill cost estimates when yield management systems are at work. Use of these models can result in spill estimates significantly different from those of current practice, which can in turn lead to different decisions about the optimal aircraft capacity to assign to a flight leg. The recursive fare class spill models introduced in this dissertation can be used to obtain more accurate estimates of spill for each specific flight demand scenario when leg-independent assumptions are valid.

We also showed that the different yield management approaches can influence the ODF passenger flows in the network in different ways. This implies that the leg-dependent spill and spill cost estimates are subject to the yield management approach being used. Using typical airline hub network examples, our results showed that the differences between the spill cost estimates of the different yield management approaches were in the excess of many hundreds and in other cases in a couple of thousands of dollars on a single leg. Our results have also shown that the differences among the spill and spill costs are inconsistent and additionally, they can be bi-directional.

In cases when leg-dependence effects may affect spill estimation, spill and spill costs should be estimated for the affected sub-network of flight legs for a given fleeting combination (assignment). The Monte Carlo simulation based approach outlined in Chapter 6 can be used efficiently for these purposes.

Airlines represent demand over a longer time period by a single density function. We called the practice horizontal aggregation. We showed that horizontally aggregating demand results in additional biases in the spill and spill cost estimates. If vertical and horizontal biases simultaneously affect the estimation,
then the individual biases can either magnify or cancel each other out. Nevertheless, we have shown, using realistic demand data, that the combined biases can be very substantial.

To avoid the horizontal bias in both leg-independent and leg-dependent networks, we recommended a disaggregated approach. Spill and spill costs should be estimated individually for each daily (or for other classified time-period, e.g. day of week ) demand data first, and the average spill and spill cost for the longer time period should be calculated as the average of the individual estimates.

In this chapter, the analysis of various typical airline examples has also shown, that in typical airline networks, ignoring yield management and/or network effects in the spill cost estimation can likely lead to sub-optimal fleet assignment solutions. This can prevent the airline from substantial cost savings or alternatively realizing potential profits. Therefore, leg-dependent spill estimates that take into consideration yield management effects should be used in fleet assignment decisions.
8. Fleet Assignment Considering Leg-dependence and Yield Management Effects

In this chapter we show how network effects influence the accuracy of fleet assignment formulations for typical airline networks. We analyze leg-dependent fleet assignment formulations and indicate the conditions under which they are valid and can be formulated as linear programs.

We divide networks into two distinct categories, based on the assumed ability to control the particular ODF demand flows in the network. In the first category, we assume that airlines can control the actual passenger flows on an OD level, using ODF based booking limits. In the second, we assume that demand can be controlled at flight leg level only, using leg-based total capacity limits or leg based booking limits. This represents networks with either no yield management systems, or those with such systems which can set booking limits only on a flight level. (Note, that most yield management systems are of this type.)

According to our knowledge, no published work in fleet assignment has attempted to model either the complex network interactions of OD passenger flows, or the effects of the different yield management systems. We show the yield management control structure has a major impact on fleet assignment models and their solution methodologies. Considering leg-dependence requires a structural change in the modeling approach. This chapter will introduce new models that incorporate network-effects and repercussions of yield management systems.

8.1 Fleet Assignment with OD Flow Control

This section assumes that the airline is able to control the actual OD flow in the network, by setting booking limits for ODF demand. Given this assumption we will present formulations where the fleet assignment and yield management decisions are combined and made simultaneously.

Note that the assumption OD control is very optimistic with respect to the assumed ability of control in today’s yield management systems. Unfortunately, today’s yield management systems do not control passenger flows on OD levels. Although certain OD heuristics exist, as was presented earlier, these systems can not be characterized by OD level of control. There are a number of problems and issues that prevent the use of direct OD control in yield management. Williamson [43] in her dissertation summarizes these issues. Nevertheless, we believe that it is worth exploring possible formulations under the assumption of OD control. Besides giving rise to a new and interesting formulation, it can be useful for certain applications as well. Although the assumption of the OD control may be unrealistic for the purpose of today’s fleet assignment, in special airline networks or special applications it may be useful.

\[1\] We will refer to a possible application later in this section.
We are interested in finding the fleeting (the assignment of aircraft), for which ODF demands can be allocated to the available flight legs so that expected profits (total expected revenues minus the total operating costs) are maximized. The formulations given below can be characterized as combined yield management (revenue maximization) and fleet assignment models. Since booking limits can be set on ODF levels, it is possible to incorporate the yield management decisions (whose goal is to maximize expected revenues given the capacities) into the fleet assignment models. In these models, the decisions for allocating the different ODF itineraries to the available leg capacities, and fleet assignment decisions are made simultaneously. These formulations enable us to maximize profits while considering leg-dependence effects. Note, again that these models have serious modeling limitations. They assume a static booking process. Even in the case when stochastic demand is assumed, the dynamism that represents airline booking processes is not fully represented.

8.1.1 Combined Passenger Flow and Fleet Assignment Models
Let us first assume that demand is deterministic. In this case, the formulation is a (static) deterministic combined yield management/fleet assignment model. Assume that there is only one path associated with all ODFs defined for an OD city-pair. Also assume that only flight coverage, flow balance, and size (number of equipment) constraints are being considered while doing the fleet assignment (for a detailed description of these constraints refer to Section 3.3).

The flights are represented in a time-space network [24],[8]. For each fleet type a copy of the network is made. The nodes of the network represent an airport at a certain time, and the arcs represent flight legs and ground arcs. A typical flight network for one fleet type is shown in Figure 8.1, (where ground arcs are the horizontal arcs). A flight arc emanates from a node associated with a departure city and a departure time. The time associated with the flight arc includes the flight time and the turn-around time at the arrival station, i.e., the time necessary to prepare the aircraft for the next flight. Therefore, the end node of a flight arc is associated with the arrival city at a time when the aircraft is available for the next flight. Note that the last ground arc is a “wrap around” arc, which makes the network cyclic, to allow daily fleet assignments with a repetitive schedule. Note that it is possible to consolidate nodes when a sequence of arrivals follow one another or a sequence of departures [24], [8]. It is assumed that for each fleet type a copy of the “same” network is defined, that is, the time-space network that was shown in Figure 8.1 is defined for each existing fleet type f.
Given the assumed time-space network, and the deterministic ODF demands, the Deterministic combined Yield Management and Fleet Assignment Model (DYMFAM) can be formulated as:

\[
\begin{align*}
\text{Max} & \quad \sum_{od \in OD} \sum_{ce \in CL} Fare_{od,c} \cdot T_{od,c} - \sum_{l \in L} \sum_{f \in F} OpCost_{f,l} \cdot X_{f,l} \\
\text{s.t.} & \quad \sum_{f} X_{f,l} = 1 \quad \forall l \in L \\
& \quad \sum_{l \in P(f)} X_{f,l} + Y_{f,-n} - \sum_{l \in Q(f)} X_{f,l} - Y_{f,+n} = 0 \quad \forall f \in F, n \in N \\
& \quad \sum_{l \in P(f)} X_{f,l} + \sum_{l \in Q(f)} Y_{f,l} \leq S(f) \quad \forall f \in F \\
& \quad \sum_{od \in OD, ce \in CL} a_{od,l} \cdot T_{od,c} + \sum_{f \in F} (-Cap_f) \cdot X_{f,l} \leq 0 \quad \forall l \in L \\
& \quad T_{od,c} \leq D_{od,c} \quad \forall od \in OD, c \in CL \\
& \quad Y_{f,+n} \geq 0 \quad \forall n \in N, f \in F \\
& \quad X_{f,l} \in \{0,1\} \quad \forall f \in F, l \in L
\end{align*}
\]  

Where

- \( D_{od,c} \) is the unconstrained fare class \( c \) ODF demand between city \( o \) and \( d \).
- \( a_{od,l} \) is a binary variable that equals 1 if the itinerary path of \( D_{od,c} \) (for all \( c \)) includes flight leg \( l \); and equals 0 otherwise.
- \( T_{od,c} \) is the actual ODF traffic of fare class \( c \) that is carried between city-pair \( od \) in the network -- we will choose them as a subset of the decision variables in the objective function.
- \( Fare_{od,c} \) is the fare associated with fare class \( c \) on itinerary \( od \).
- \( OpCost_{f,l} \) is the operating cost of flying fleet type \( f \) on flight leg \( l \).
- \( X_{f,l} \) is the binary decision variable, which has a value of 1 if fleet \( f \) flies the flight leg \( l \), from \( i \) to \( j \) departing at time \( t \); and 0 otherwise.
• \(l(n)\) and \(O(n)\) contain the flight legs that are flying \textit{in} to and flying \textit{out} of node \(n\).

• \(Y_{f,n}\) and \(Y_{f,n^*}\) represent the number of fleet type \(f\) aircraft on "ground arcs" at node \(n\) before and after, respectively.

• \(S(f)\) is the number of fleet type \(f\) aircraft available.

• \(P(f)\) set of flight arcs that cross the so-called "count line", (see Figure 8.1). (To assure that there are no more than the available number of fleets are assigned to the network, aircraft should be counted at a selected time. Flow balance constraints will assure that the conditions are maintained for the rest of the network.)

• \(Q(f)\) contains all the ground arcs that cross the "count line", (see Figure 8.1).

• \(C, L, C, F, L, C\) denotes fare classes, cities served, available fleets, and flight legs of the schedule, respectively.

The above formulation is a mixed integer multicommodity flow programming (IP) formulation. The commodities in the network are the ODF passengers and the different fleet types.

Constraints (8.1b) are \textit{cover} constraints forcing each flight leg to be flown at each node for each fleet type. Constraints (8.1c) are \textit{flow balance} constraints. They assure that at each node of the time-space network, for fleet \(f\), the sum of the number of incoming aircraft (first term in 8.1c) and the number aircraft being on the ground (\(Y_{f,n}\)) equals the sum of the number of outgoing aircraft (third term in 8.1c) and the number of aircraft that stay at the station (\(Y_{f,n^*}\). Constraints (8.1d) are \textit{fleetsize} constraints, which limit the model from assigning more aircraft to the schedule than available of a certain fleet type. Constraints (8.1e) are the \textit{connecting} constraints between the fleet assignment part and deterministic yield management part of the model. They assure that the total number of ODF passengers that are allocated do not exceed the flight leg capacities on any of the flight legs. These constraints connect the two models and at the same time assure that leg-dependence effects are taken into consideration. Constraints (8.1f) assure the assigned ODF traffic is not larger than the existing ODF demand, constraints (8.1g) assure that the number of aircraft on ground arcs are non-negative, and constraints (8.1h) assure that the actual fleet assignment is an integer assignment.

A schematic diagram of these constraints is given in Figure 8.2. This figure enables us to view the broad layout of the formulation, so that we may be able to exploit problem structure.
The above figure also helps us to estimate the size of the problem: The formulation has $(OD*CL)+(F*L)+G$ decision variables (where G represents the number of ground arcs) and the number of constraints is $(F*N)+L+F+L$. Both the fleet assignment and the deterministic yield management parts of the problem are large. Combining the two makes it hard to solve using a simple linear programming solver. For example, take the case of a simple hub-spoke-network with one hub with a connecting complex in which 20 flights arrive and 20 flights depart. Assume that this complex is repeated 5 times a day², that reasonable OD demand exists between flights of a connection complex only, and that there are 5 fare classes defined on each OD market. The number of OD pairs in this case is $OD=5*(20*20+2*20)=2200$. The number of flight legs $L=5*(20+20)=200$. Assume that there are $F=7$ different fleet types available and in the most optimistic case, the nodes can be consolidated so that the number of nodes $N=5*21=105$. For this case the number of columns is $(2200*5)+(7*200)=12400$, and

² These are typical numbers for large international airlines or mid-size US domestic networks
the number of rows is \((7 \times 105) + 200 + 7 + 200 = 1142\). (Note that this size was the theoretical minimum and the assumed network was relatively small. In practice it would be almost impossible to consolidate nodes so efficiently and larger networks should be considered.) It is not easy, although possible, to obtain an optimal solution for a linear multicommodity flow problem that has a 12400x1142 constraint matrix. It is even harder to obtain the integer solution. The size could be prohibitively large for many solvers. If we consider larger, more typical networks, the size of the formulations could increase dramatically.

The above formulation can be expanded to the case when demand is represented in a stochastic form. When demand is stochastic, then the expected revenue for an ODF demand for a given capacity can be calculated as the ODF fare times the expected load. Remember, that load can be calculated given the formula in equation (3.25). Assuming a typical stochastic demand density function (e.g., Gaussian), Figure 8.3 represents the expected revenue curve, which is a concave function [20]. It is possible to approximate the function by a piece-wise linear function. (Note that since the expected revenue curve is a concave function, we do not need to define and use binary variables, which is often the case in piece-wise approximations, in the linear programming maximization formulation.)

![Revenue Curve]

Figure 8.3: Concave revenue curve with piece-wise linearization

This linear representation of the revenue function can be easily incorporated to the formulation given in equation (8.1), by redefining the ODF traffic variables:

- Let \(T_{odc, t}\) be the \(t\)th decision variable of the \(od, c\) fare class ODF demand, where \(t \in T\) represent the pieces of linearization (see Figure 8.3), and \(b_t\) represent the upper limits of the \(t\)th interval. (The more variables that are defined, the better the linear approximation of the revenue curve will be.) We can also write that the revenue for the ODF is given as:

\[ R_{odc} = \sum_{t \in T} T_{odc, t} \times b_t \]

\(^3\)Note that there is no need to define an upper limit for the last interval because the number of passengers that can be allocated are also limited by the flight leg capacity limits.
\[ T_{od,ct} = \sum_{t \in T} T_{od,ct} \] where \[ T_{od,ct} \leq b_t - b_{t-1} \]

- Let \( Rev_{od,ct} \) the approximated marginal revenue associated with the linear approximation in the \( t \)th interval.

The variable \( T_{od,c} \) in this stochastic case can be interpreted, similar to the deterministic case, as the (non-nested) booking limit set for the ODF itinerary. Assuming that the above linearization is obtained for each expected revenue function for each ODF, the Stochastic combined Yield Management and Fleet Assignment Model (SYMFAM) is given by:

\[
\text{Max} \sum_{od \in OD} \sum_{c \in CL} \sum_{t \in T} Rev_{od,ct} T_{od,ct} - \sum_{l \in L} \sum_{f \in F} \text{OpCost}_{f,l} \times X_{f,l} 
\]

s.t.

\[
\sum_f X_{f,l} = 1 \quad \forall l \in L \]

\[
\sum_{l \in O(n)} X_{f,l} + Y_{f,\neg n} - \sum_{l \in D(n)} X_{f,l} - Y_{f,\neg n} = 0 \quad \forall f \in F, n \in N
\]

\[
\sum_{l \in P(f)} X_{f,l} + \sum_{l \in Q(f)} Y_{f,l} \leq S(f) \quad \forall f \in F
\]

\[
\sum_{od \in OD} \sum_{c \in CL} \sum_{t \in T} a_{od,ct} T_{od,ct} + \sum_{f \in F} (-Cap_f) \times X_{f,l} \leq 0 \quad \forall l \in L
\]

\[
T_{od,ct} \leq b_t - b_{t-1} \quad \forall od \in OD, c \in CL, t \in T
\]

\[
Y_{f,\neg n} \geq 0 \quad \forall n \in N, f \in F
\]

\[
X_{f,l} \in \{0,1\} \quad \forall f \in F, l \in L
\]

The advantage of SYMFAM is that it incorporates the stochastic nature of demand. Note however, that in this formulation booking limits are not nested. Williamson in [43] has shown that non-nested booking limits are bad from the perspective of yield management. Further, the model inherently assumes a static booking process. That is, the dynamic yield management optimization process that reacts to the interspersed ODF booking demand requests are not captured.

Note that in this formulation, the size of the problem increased substantially by the inclusion of additional columns in the constraint matrix. Although it is possible that the LP solution can be obtained even for this problem, it is expected that the IP solution will be very hard to obtain.
To overcome this problem, we are proposing an alternative column generation based formulation for the DYMFAM and SYMFAM problems. Using column generation, we will be able to decompose the problem into subproblems and also be able to exploit the existing fleet assignment solution approach capability developed by Hane et. al. [43] and Barnhart [7].

8.1.2 Column Generation Based Combined Passenger Flow and Fleet Assignment Models

Sometimes, LP formulations have many variables. In such cases, it is not just inefficient to solve the full problem, it is also infeasible to enumerate all variables. Such problems are solved by implicit enumeration methods like column generation. Column generation is an algorithm that solves the linear program using subsets of the total number of variables (columns). At each iteration, variables are added as needed. Variables are selected using a subproblem that implicitly prices out all possible variables. Variables that "price-out" are added to the linear program (restricted master problem) and the LP is re-solved over the new set of variables. The algorithm stops when no new variables that can potentially improve the original solution can be generated by the subproblem. The trick is to have an efficient procedure that efficiently prices out new variables. (For a detailed description of the column generation solution technique refer to [2], [6], [7], [8], [48].)

In the following column generation formulation, we decompose the DYMFAM and SYMFAM problem, (which were flight leg based formulations), into a master problem and a subproblem and exploit the reduced sizes due to the decomposition. This decomposition can help us to solve the large scale problem, in cases like DYMFAM and SYMFAM where the formulation may be prohibitively large. Further, the column generation approach can provide a heuristic solution that approximates the optimal solution by a sufficiently small "gap", i.e., with an acceptable small difference between the obtained and the optimal solution's objective function value. Using the column generation approach, in solving the subproblem for finding new columns to be added to the master problem, as we will see below, we also can leverage on the existing solution methodologies developed for solving the traditional large scale fleet assignment problem [24], [7].

The following formulation is a Deterministic combined Yield Management and Fleet Assignment Model with Column Generation (DYMFAMCG):

---

4 A variable "prices out" if it has the potential to improve the solution value of the master problem. In particular, in a maximization problem, any variable that has a positive reduced cost "prices out".
\[
\text{Max} \sum_{od \in OD} \sum_{ce \in CL} \text{Fare}_{od,c} \cdot T_{od,c} - \sum_{i \in Fl} \text{FlCost}_i \cdot X_i
\]  
(8.4a)

s.t.
\[
\sum_{od \in OD} \sum_{ce \in CL} a_{od,l} \cdot T_{od,c} + \sum_{i \in Fl} (-\text{Cap}_{i,l}) \cdot X_i \leq 0 \quad \forall l \in L
\]  
(8.4b)

\[
\sum_{i \in Fl} X_i = 1
\]  
(8.4c)

\[
T_{od,c} \leq D_{od,c} \quad \forall od \in OD, c \in CL.
\]  
(8.4d)

\[
X_i \in \{0,1\} \quad \forall i \in Fl
\]  
(8.4e)

Where

- \(\text{FlCost}_i\) is the total cost of the fleeting \(i\). The cost of the fleeting is the sum of the operating costs associated with the actual assignment of aircraft types to each flight leg.

- \(X_i\) is the binary decision variable, which has the value 1 if fleeting \(i\) is selected, and 0 otherwise.

- \(\text{Cap}_{i,l}\) is the capacity of the aircraft that is assigned to flight leg \(l\) in fleeting \(i\).

- \(Fl\) is the set of different fleeting (columns) that are part of the master problem, and \(i \in Fl\) is a fleeting (column).

- The description for the rest of the variables are similar to what was given for the model in equation (8.1).

The schematic diagram of the above constraints is given in Figure 8.4. The diagram helps to understand the structure of the model. As the diagram shows, the column generation formulation has two distinct parts: The first part, similar to the earlier formulations, assigns the ODF passenger flow mix to the network. The columns in this part of the formulation are fixed and all are always part of the "master problem" (or they could be generated if we assumed that there are more than one path available for an OD pair). The variables associated with these columns are the \(T_{od,c}\) variables. The second part of the formulation contains the fleeting columns. A complete fleeting (a particular assignment to the flight legs) is represented by a variable \(X_i\), with its corresponding column. An element associated with a flight leg in the fleeting column contains the negative of the capacity of the fleet type that is assigned to the particular flight leg in the fleeting. That is, if fleet type \(f\) is assigned to flight leg \(l\) in the fleeting, and the capacity of fleet type \(f\) is \(\text{Cap}\), then the entry associated with flight leg \(l\) will contain the value \(-\text{Cap}\). Each fleeting that is represented by a variable (column) in the master problem, is a valid fleeting, i.e., all cover, flow balance, and size constraints are satisfied. Note that this formulation is a fleeting based formulation, opposite to the earlier formulations which were flight leg based formulations.
The objective function of the formulation maximizes total profit by maximizing the total revenues minus the total fleeting costs. Since the variables associated with the fleeting columns are binary variables, the optimal solution will be the fleeting for which it is possible to allocate the ODF demands to the available capacities so that the resulting total revenues minus the cost associated with the fleeting is maximal.

Constraints (8.4b) are special connecting constraints. These constraints connect the passenger allocation part to the fleet assignment part of the model. The constraints assure that no more ODF passengers can be allocated to any flight leg of the network than the capacity of the flight leg. These constraints are the ones that assure that leg-dependent effects are observed in the passenger allocation part of the model.

Constraint (8.4c) is the bounding, or convexity constraint. They assure that the flight legs will be covered by a fleeting. Constraints (8.4d) assure that no more ODF passengers are allocated to the network than the actual unconstrained ODF demand is. Constraints (8.4e) assure that selected fleeting solution will be integer.

Theoretically it can be possible to generate all possible "valid" fleeting columns that satisfy the three basic constraints (cover, flow balance, size) and add to the master problem. However, we presume that this is either an overwhelming or unnecessary task. This method would be useful only in case of small network sizes. Using the column generation solution methodology, we assume that not all possible fleeting columns would be part of the linear problem. (We refer to the problem with the limited number of fleeting columns as the restricted master problem.) To start with, only one or a few fleeting columns...
would be generated and included. The included columns could represent the currently used fleet in the schedule, and the fleet that is obtained using the traditional fleet assignment formulation (which does not consider leg-dependence effects). Additional fleet can be obtained quickly by swapping fleet types around cycles. For finding swapping cycles efficiently refer to [15]. This procedure would enable the generation of additional fleet for the master problem while maintaining a valid fleet.

The optimal LP solution to the master problem can be found using advanced LP solvers. Since not all fleet columns are part of the master problem, it is not guaranteed that the basic columns (i.e., the columns that are in the basis) of the optimal LP solution to the master problem would be the same as the basic columns of the optimal solution to the entire problem. From linear algebra theory, though, it is known that the solution to a problem can be improved if a column with positive reduced costs can be found and added to the basic columns. If no columns have positive reduced costs, then the solution is optimal [2], [7], [8], [48].

Therefore, our task is to find a column (that is not part of the master problem) with positive reduced costs and add to the master problem. If no such column can be found, then we can stop because we know that the solution to the master problem is globally optimal. If one or more fleet columns are found with positive reduced costs, then they should be added to the master problem. After solving the LP of the master problem, a new LP solution is obtained. The procedure of finding columns with positive reduced costs and adding to the master problem should be performed iteratively until no more such columns are found, or until the obtained solution is good enough (i.e., the difference "gap" in the solution value between the obtained and the optimal solution, or lower bound, is sufficiently small).

Remembering that the $\text{FlCost}_l$ were defined as the sum of the operating costs associated with the fleet types assigned to the flight legs, the reduced cost, $\bar{c}_i$, of a fleet column $i$ can be calculated by the following:

$$\bar{c}_i = -\text{FlCost}_i - \sum_{l \in L} \Pi_i * (-\text{Cap}_{i,l}) - \sigma$$  \hspace{1cm} \text{(8.5a)}

or alternatively:

$$\bar{c}_i = \sum_{l \in L} \sum_{f \in F} \xi_{f,l} * (-\text{OpCost}_{f,l}) - \sum_{l \in L} \Pi_i * (-\text{Cap}_{i,l}) - \sigma$$

$$\bar{c}_i = \sum_{l \in L} \left( \sum_{f \in F} \xi_{f,l} * (-\text{OpCost}_{f,l} + \Pi_i * \text{Cap}_{i,l}) \right) - \sigma$$ \hspace{1cm} \text{(8.5b)}

Where

- $\Pi_i$ is the dual variable of the connecting constraint associated with leg $l$ in the optimal solution of the restricted master problem. (See the schematic diagram in Figure 8.4).
• $\sigma$ is the dual variable associated with the bounding (or convexity) constraints. (See Figure 8.4).

• $\xi_{f,l}^i$ is the assignment variable that has the value 1 if fleet type $f$ is assigned to leg $l$ in the fleeting $i$, 0 otherwise.

• $FlCost_I$ and $Cap_I$ are as were given earlier.

Note that $Cap_I$ will equal $Cap_f$ if fleet type $f$ is assigned to leg $l$ in the fleeting of $i$ (i.e., when $\xi_{f,l}^i$ equals 1). The above also imply that, if a column is found for which the sum of the reduced costs on each selected arcs are greater than $\sigma$, i.e. if

$$\sum_{l \in L} \left( \sum_{f \in F} \xi_{f,l}^i \ast (-OpCost_{f,l} + \Pi_I \ast Cap_f) \right) > \sigma$$

(8.6a)

then adding the column to the master problem as a new variable will improve the solution value of the master problem.

Note that finding the fleeting with the largest reduced cost is a multicommodity flow problem. In this case, the cost that is associated with assigning fleet type $f$ to flight leg $l$ in the subproblem's maximum cost multicommodity flow IP formulation is given by:

$$(-OpCost_{f,l} + \Pi_I \ast Cap_f)$$

If the subproblem is formulated as a minimum cost multicommodity flow problem, then the cost associated with assigning fleet type $f$ to flight leg $l$ is:

$$(OpCost_{f,l} - \Pi_I \ast Cap_f)$$

(8.6b)

The subproblem then can be formulated as:

$$\min \sum_{l \in L} \sum_{f \in F} rcost_{f} X_{f,l}$$

(8.7a)

s.t.

$$\sum_{f} X_{f,l} = 1 \quad \forall l \in L$$

(8.7b)

$$\sum_{l \in l(n)} X_{f,l} + Y_{f,-n} - \sum_{l \in G(n)} X_{f,l} - \sum_{l \in G(n)} Y_{f,+n} = 0 \quad \forall f \in F, n \in N$$

(8.7c)

$$\sum_{l \in P(f)} X_{f,l} + \sum_{l \in Q(f)} Y_{f,l} \leq S(f) \quad \forall f \in F$$

(8.7d)

$$Y_{f,+n} \geq 0 \quad \forall n \in N, f \in F$$

(8.7e)

$$X_{f,l} \in \{0,1\} \quad \forall f \in F, l \in L$$

(8.7f)

Where

• $rcost_{f,l}$ the modified (reduced) cost when fleet type $f$ is assigned to leg $l$ is:
\[ rcost_{i,j} = OpCost_{i,j} - \Pi_i \times Cap_i \]

- and the rest of the variables are similar as defined under equations (8.1) and (8.4).

The above formulation is similar to the formulation of the traditional fleet assignment model [24] introduced in Section 3.3, only with differences in the objective cost coefficients. Solving the above subproblem is not trivial and for large problem sizes it is not fast (approximately 40 minutes on a workstation). Nevertheless, the advanced solution technique and methodology developed by Hane et al. [24], can be exploited by further enhancing by the techniques developed by Barnhart [6], [7].

Solving the above fleet assignment model with the modified reduced costs on each flight leg, a valid fleeting -- that observe the cover, flow balance, and size constraints -- with the minimum total reduced cost can be obtained. If the total modified cost of the fleeting is larger than \( \sigma \) (if it is calculated as given in equation (8.6a)), then the column should be included in the master problem. Once the optimal LP solution to the entire problem is obtained, using a Branch and Price technique [7] the optimal IP can be also obtained.

For the Stochastic case, based on the discussion above, the column generation formulation of Stochastic combined Yield Management and Fleet Assignment Model (SYMFAMCG) can be easily formulated:

\[
Max \sum_{od \in OD} \sum_{ce \in CL} \sum_{ie \in T} Rev_{od,ce,i}T_{od,ce,i} - \sum_{ie \in Fl} FlCost_i * X_i \quad (8.8a)
\]

s.t.

\[
\sum_{od \in OD} \sum_{ce \in CL} \sum_{ie \in T} a_{od,i} * T_{od,ce,i} + \sum_{ie \in Fl} (-Cap_{i,i}) * X_i \leq 0 \quad \forall l \in L \quad (8.8b)
\]

\[
\sum_{ie \in Fl} X_i = 1 \quad (8.8c)
\]

\[
T_{od,ce,i} \leq b_t - b_{t-1} \quad \forall od \in OD, c \in CL, t \in T \quad (8.8d)
\]

\[
X_i \in \{0,1\} \quad \forall i \in Fl \quad (8.8e)
\]

(For the notation refer to the earlier formulations.) The master problem and the subproblem can be constructed in a similar way as they were for the DYMFMAMCG formulation. The modified cost for each flight leg of the network in the subproblem can be calculated using equation (8.6b) and the same multicommodity flow algorithm should be used for finding a fleeting with a positive reduced cost.

In the following we try to summarize the advantages of the column generation based DYMFMAMCG and SYMFAMCG formulations over the DYMFM and SYMFAM formulations:

- Using the column generation formulation the problem can be solved when the size of the DYMFM and SYMFAM formulation is prohibitively large with respect to the optimization software and hardware capabilities. Using column generation, only a portion of the columns are used and only the
relevant columns are calculated and used in the solution process. By decomposing the problem into
two subproblems, the complexity of the entire problem is reduced.

- The advantage of the fleeting based column generation formulations is that they provide a tighter (or
  at least as tight or strong) LP relaxation than the flight leg based formulations. The convex
  combination of the integer solution points (of the fleeting columns of the column generation
  approach) are contained within the feasible area of the convex combination of the LP solution points
  (of the flight leg based formulation). (For a detailed discussion why column generation formulations
  result in a tighter convex hull, the reader is referred to [46].) The consequence is that the optimal LP
  objective function values from the column generation LP are much lower than those of the flight leg
  based formulations. This helps in the Branch and Price procedure because the higher bound is lower,
  therefore the bounding is better. Also, experience shows that it is more likely to obtain fast an integer
  solution within branch-and-bound [7], [8].

- Another advantage of the column generation approach is that it is relatively easy to obtain a heuristic
  solution. Note that it is possible to terminate the column generation iteration at the end of each
  iteration and get an IP solution associated with the solution to the master problem LP of the last
  iteration. Therefore, using the column generation based approaches we speculate that in a few
  iterations a reasonably good solution can be obtained. The procedure can be run to optimality, but if
time does not allow, the iterations can be terminated and at the time available best solution can be
  obtained. The flight leg based formulations do not have this advantage.

- The subproblem of the column generation formulations is a multicommodity flow problem on the
  original airline network, where the costs are the leg based modified costs given in equation 8.6b. The
  use of the reduced costs in the subproblem may give a good indication what the true “leg-based”
  allocations should be. Therefore, some insights could be obtained. The multicommodity flow
  problem unfortunately is not a trivial problem and the solution to the problem is not expected to be
  very fast. Note however, that the problem is identical to the traditional fleet assignment formulation,
given in Hane et. al [24] and discussed in the literature review in Section 3.3, only with different
  objective coefficients. Nevertheless, the advanced solution methodologies and techniques can be
  directly used and applied. This way by the suggested decomposition we can leverage the existing fleet
  assignment capability.

- Summarizing, to solve the combined yield management and fleet assignment problem, when control
  on the ODF level is assumed and possible, we suggest the use of the column generation based
  SYMFAMCG and DYMFAAMCG approaches.

To summarize, we have shown that if the airline is able to control through its yield management system
passenger flows on the OD level, then it is possible to formulate the fleet assignment problem as a linear
problem where leg-dependence effects are taken into consideration. In the above formulations the yield
management decisions were incorporated into the combined decision model. We introduced a flight level
based formulation and a fleeting based formulation for both the deterministic and stochastic demand case.
For the fleeting based formulation we proposed the column generation solution methodology, because it
can handle large network sizes and can obtain fast heuristic solutions.

Note that in the above models, yield management and fleet assignment decisions were combined and
incorporated. Although it is a new approach and formulation, it has a number of limitations. The
incorporated yield management decision model is a very much simplified model. First of all it does not
produce nested booking limits. Previous research has shown that in the dynamic and stochastic demand environment nesting of booking limits is necessary [43], [12]. Further, the modeled yield management system is a static system, contrary to the yield management practice, in which decisions are updated dynamically at the response of the interspersed booking request. Therefore, the direct applicability of the above models for fleet assignment is limited. Nevertheless, we believe that it is important to explore this area and formulate the algorithms. Besides some practical applications, that will be addressed below, using this algorithms can be a good way to begin to evaluate fleet assignment models and leg-based objective function coefficients. In the next section we will further address the applicability issue and outline a possible application of the above formulations.

8.1.3 Applicability and Relevance Issues
There are serious obstacles that characterize the application of these models in current airline environment. The first is that airline demand is stochastic. Therefore, the stochastic models are the ones that can reasonably be applicable in practice. Second, today most airlines can set booking limits only on flight leg level. Although different OD based heuristic approaches are under development and some are in use today (we introduced one in the previous chapter) they cannot be characterized as ODF level control approaches. It is not completely realistic to represent them as we did in the SYMFAM and SYMFAMCG models in this chapter. Another obstacle is the ODF data. For yield management decision purposes, booking limit solutions need to be also integer solutions. Unfortunately, it is the nature of today’s US domestic airline demands, that connecting ODF demands for a single path and departure are very small in size. Their mean is often less than one passenger. Obtaining an integer solution by the SYMFAM or SYMFAMCG method, seats for various ODF demands would be rounded up to 0 or 1. Research findings of Williamson [43] and Belobaba [11] have shown that this approach is inferior in terms of maximizing expected revenues to the OD heuristic approaches that aggregate ODF demands to a certain degree and use nested booking limits to create a buffer for variance in the different ODF requests. Further research is required to determine, however, the relative magnitudes between the losses due to the non-nested optimization approach versus the gains in the fleet assignment solution that considers leg-dependence effects.

Nevertheless, we believe that the above models can be useful in various applications where the ODF demand means are not very small. These may occur in cases of a small hub, in international hub operations, or in non-hub connections and multiple leg flights. For example, the connecting complexes in many international airline networks are not as big as in the US networks. These international airlines do not operate in a fully hub-and-spoke pattern. Although connection opportunities are available, the connecting ODF demands are not as dispersed as that of the US airlines, and thus the fewer connecting OD opportunities have larger ODF demand sizes. (This can be also true for US international connecting
banks). In these cases, the models developed above can be used to give some estimates or possible insights about the profit opportunities in the network.

Another possible application occurs when an airline, very close before departure, due to unexpected reasons, has to reassign its fleet in a part of the network such that the expected profits are maximized. The unexpected reasons may involve weather or airports with labor strikes (the latter is especially typical in Europe). In this case, most of the ODF bookings are already accepted and confirmed; therefore, even the deterministic formulation can be applicable. In this special situation, the airline often needs to make flight cancellations. A higher level of decision would be if, instead of only canceling flights, airlines would cancel individual bookings in different ODF fare classes as well. The airline then could use the DYMFAM or SYMFAM model that would assign the already existing demand and the available fleet types to the network, such that the total profits are maximized. If the number of the aircraft are less than the required number to cover all flights, then the models can be adjusted to the new circumstances, and the model would select those flights that can be canceled. We believe that the above model can be often used in many similar situations.

Although we are aware that the models introduced in this section today have limitations, we believe that for certain special applications, and certainly at some time in the future they may become very useful. A possible application could be the use of the above models after modifications in very short term fleet assignment decisions (see Section 3.3). Further research is needed to determine their applicability. In the next sections we will discuss the cases that are in a bigger accordance with current airline practices. We will assume that it is not possible to directly control demand on an OD level. We will address the implications of this assumption, and propose an approach that, while taking into consideration the leg-dependence effects, also incorporates the actual characteristics of the yield management approach used by the airline.

8.2 Fleet Assignment with No OD Flow Control

In this section we will assume that passenger flows can not be controlled on OD level. We will address the issue when the control of OD passenger flows is limited to the flight leg level only. This represent networks where a leg-based capacity limit, in a form of a total seat capacity or a leg-based booking limit, is applied to all OD demands that traverse the flight legs. Consequently, the interaction of the different OD demands with respect to the common leg-based capacity limits has to be also modeled here.

First, we will address the implications of the flight leg based capacity limits on fleet assignment models. Using a simplified example, we will show that leg-dependence effects have a very serious implication for the fleet assignment models. We will show that under typical fare hierarchies and network characteristics, the feasible area does not form a convex hull.
Let $D_{ij}$ be the deterministic OD demand between city-pair $i$ and $j$ with a fixed itinerary, and $fare_{ij}$ is its associated fare. Also assume that we can assign a “rubber aircraft” to any flight leg $l$ with capacity $Cap_l$, where the only constraint for leg $l$ is that $0 < Cap_l < MaxCap_l$. Also assume that the operating cost $OpCost_l$, associated with the rubber aircraft scales linearly with its capacity, thus

$$OpCost_l = c_l \times Cap_l$$ (8.9)

where $c_l$ is the fixed cost parameter associated with flight leg $l$ and represents the characteristics of the flight leg.

Let $T_{ij}$ represent the traffic of origin-destination demand $ij$ in the market, i.e., $T_{ij}$ represents the origin-destination demand that is actually carried with respect to the prevailing capacity limits. Since in this model we do not assume that the airline can control OD passengers flows, we cannot choose the traffic variables to be our decision variables. The $T_{ij}$'s -- the leg based projections of the underlying leg-dependent passenger flows, -- serve only as variables representing the actual OD flows on each leg in the network. In this model the leg capacity variables, $Cap_l$, which represent the fleet type as well, will serve as decision variables. Figure 8.5 shows our two-leg sample network. In this example our decision variables are $Cap_A$ and $Cap_B$.

![Diagram of OD flows and capacity constraints](image)

Figure 8.5: Two-leg example for the simplified capacity assignment model

If the OD flows are not being controlled by the airline, we need a model of how the available capacities on each leg are filled up by the different OD demand flows. We will assume again the proportional fill assumption, which states that each OD demand book by a constant uniform booking rate. The result of this assumption is that until no legs of the OD demand paths become full, the OD mix of the accepted passengers on all legs will be proportional to the original OD demand means $^5$. The bottle-neck leg, the leg which becomes full first, will always accommodate an OD mix proportional to the original means of OD demands traversing the leg.

$^5$ For a more detailed description of the proportional fill assumption refer to the earlier discussion in section 4.1.2
The OD mix of the next leg that becomes full, call it leg \( l \), can be calculated according to the logic presented in section Chapter 4. According to that, we know, that on leg \( l \) exactly the same number of these OD passengers will be accommodated as were on the leg that became full before leg \( l \). Thus, only the remaining capacity on leg \( l \) is filled with OD demands that do not traverse the legs that became full earlier than leg \( l \). These OD demands again fill up the remaining capacity proportionally. We can use the same logic to calculate the OD mix for the leg that becomes full next.

The sequence of the legs becoming full is also a very important issue. It is relatively easy to calculate the leg that becomes full first. Assuming proportional uniform booking rates for deterministic demands, the demand factor, \( DF \), of a flight leg determines also the first leg that becomes full. The leg with the highest demand factor will be thus become full first during the booking process, and thus this leg is the bottleneck leg of the network. The next leg that becomes full depends on ratio of the unconstrained OD demands and the remaining capacity (capacity minus the achievable OD traffic) of the leg, (refer to Sections 4.1 and 6.1 for detailed explanation).

Given the proportional fill assumption and the capacity limits on both legs, we can calculate the leg-independent traffic volumes, (the approach of the traditional models):

On leg 1:

\[
T_{AB} = \min\left(\frac{D_{AB}}{D_{AB} + D_{AC}} \cdot Cap_1, D_{AB}\right) \quad T_{AC} = \min\left(\frac{D_{AC}}{D_{AB} + D_{AC}} \cdot Cap_1, D_{AC}\right);
\]

and on leg 2:

\[
T_{AB} = \min\left(\frac{D_{BC}}{D_{BC} + D_{AC}} \cdot Cap_2, D_{BC}\right) \quad T_{AC} = \min\left(\frac{D_{AC}}{D_{BC} + D_{AC}} \cdot Cap_2, D_{AC}\right).
\]

This would be our model assumption if we did not take into consideration the leg dependent nature of the problem. This model would work well if we did not expect any spill in the network or if \( T_{AC} \) were always equal to \( D_{AC} \) -- meaning that connecting demand was never spilled. However, if this does not hold, the leg-dependence issue must be taken into account. Otherwise, it would be possible that the \( T_{AC} \) on the two legs do not equal -- which is impossible in reality and shows inconsistency. This is the essence of leg-dependence. Therefore, \( T_{AC} \) should be written rather as the minimum of the \( T_{AC} \)'s over all legs that \( D_{AC} \) traverses, (this is our definition of achievable OD traffic introduced earlier):

\[
T_{AC} = \min\left(\frac{D_{AC}}{D_{AB} + D_{AC}} \cdot Cap_1, \frac{D_{AC}}{D_{BC} + D_{AC}} \cdot Cap_2, D_{AC}\right). \tag{8.10}
\]

The above equation indicates that the value of \( T_{AC} \) depends on the leg that becomes full first, i.e., it takes into consideration the capacity levels on both legs: \( Cap_1 \) and \( Cap_2 \).
The leg-based linear program of the fleet assignment model can be formulated similarly to that of Model 1. The objective again is to maximize the total profits. Thus the formulation is given by:

\[
\max \, \text{Tot. Rev} - \text{Tot. Cost} = \max \sum_{od \in OD} \text{fare}_{od} \, T_{od} - \sum_{l \in L} c_{l} \, \text{Cap}_{l} = \\
\max T_{AB} \, \text{fare}_{AB} + T_{BC} \, \text{fare}_{BC} + T_{AC} \, \text{fare}_{AC} - (c_{1} \, \text{Cap}_{l} + c_{2} \, \text{Cap}_{2})
\]

s.t.
\[
T_{od} \leq D_{od} \quad \forall od \in OD, \tag{8.11}
\]
\[
T_{AC} = \min \left( \frac{D_{AC}}{D_{AB} + D_{AC}} \ast \text{Cap}_{1}, \frac{D_{AC}}{D_{BC} + D_{AC}} \ast \text{Cap}_{2} \right)
\]
\[
\text{Cap}_{l} \leq \text{MaxCap}_{l} \quad \forall l \in L
\]

Note, that the traffic variables \( T_{AB}, T_{BC}, \) and \( T_{AC} \) are a function of \( \text{Cap1} \) and \( \text{Cap2} \), and their values will be determined whether leg 1 or leg 2 becomes the bottle-neck leg of the network. We showed earlier that in case of deterministic demand, the leg with the higher demand factor will be the first to get full. In this example there are two different outcomes, and these two cases need to be evaluated: 1) leg 1 is the bottle-neck leg; and 2) leg 2 is the bottle-neck of the network. Let us concentrate for now on the cases only when the AC demand is spilled, i.e., \( T_{AC} < D_{AC} \). (It is reasonable to assume that we never assign a capacity that is bigger than the total achievable traffic on a leg \(^6\)).

On the bottle neck leg, OD demands are filled proportionally. Since the bottle-neck leg becomes full first, we know that on the other leg of the network will exactly as much connecting traffic be accommodated as much were on the bottle leg neck. Therefore, the local traffic on the non-bottle-neck leg can be calculated as the minimum of the unconstrained local demand and of the leg capacity minus the connecting AC achievable OD traffic.

Below for the two possible cases, we will define the achievable OD traffic equations and the objective functions. The objective functions are obtained by substituting the corresponding achievable traffic equations into the objective function of equation (8.11). Case 1 represents the cases when the capacity variables relate to the unconstrained demands so that leg 1 will become full first, i.e., in these cases the leg demand factors \( DF \) are such that \( DF_{2} < DF_{1} \); while case 2 represents the case when leg 2 becomes full first, thus when \( DF_{2} > DF_{1} \).

\(^6\) This way we do not need to consider the case when on a non bottle-neck leg the local demand cannot fill the available capacity left after the connecting demand gets censored
Case 1: Leg 1 is the bottle neck, therefore:

\[
T_{AC} = \frac{D_{AC}}{D_{AB} + D_{AC}} \text{Cap}_1 < \frac{D_{AC}}{D_{BC} + D_{AC}} \text{Cap}_2 \\
\Rightarrow \frac{\text{Cap}_1}{\text{Cap}_2} < \frac{D_{AB} + D_{AC}}{D_{BC} + D_{AC}} \text{ or } \frac{D_{BC} + D_{AC}}{\text{Cap}_2} < \frac{D_{AB} + D_{AC}}{\text{Cap}_1} \equiv DF_2 < DF_1
\]

\[
T_{AB} = \min(\frac{D_{AB}}{D_{AB} + D_{AC}} \text{Cap}_1, D_{AB})
\]

\[
T_{BC} = \min(\text{Cap}_2 - \frac{D_{AC}}{D_{AB} + D_{AC}} \text{Cap}_1, D_{BC})
\]

Objective fn. = \[
\frac{D_{AB}}{D_{AB} + D_{AC}} \text{Cap}_1 \text{fare}_{AB} + (\text{Cap}_2 - \frac{D_{AC}}{D_{AB} + D_{AC}} \text{Cap}_1) \text{fare}_{BC}
\]

\[
+ \frac{D_{AC}}{D_{AB} + D_{AC}} \text{Cap}_1 \text{fare}_{AC} - (c_1 \text{Cap}_1 + c_2 \text{Cap}_2) =
\]

\[
= \text{Cap}_1 \{\frac{1}{D_{AB} + D_{AC}} [D_{AC} (\text{fare}_{AC} - \text{fare}_{BC}) + D_{AB} \text{fare}_{AB}] - c_1 \}
\]

\[
+ \text{Cap}_2 \{\text{fare}_{BC} - c_2 \}
\]

Case 2: Leg 2 is the bottle-neck leg:

\[
T_{AC} = \frac{D_{AC}}{D_{BC} + D_{AC}} \text{Cap}_2 < \frac{D_{AC}}{D_{AB} + D_{AC}} \text{Cap}_1 \\
\Rightarrow \frac{\text{Cap}_1}{\text{Cap}_2} > \frac{D_{AB} + D_{AC}}{D_{BC} + D_{AC}} \text{ or } \frac{D_{BC} + D_{AC}}{\text{Cap}_2} > \frac{D_{AB} + D_{AC}}{\text{Cap}_1} \equiv DF_2 > DF_1
\]

\[
T_{AB} = \min(\text{Cap}_1 - \frac{D_{AC}}{D_{BC} + D_{AC}} \text{Cap}_2, D_{AB})
\]

\[
T_{BC} = \min(\frac{D_{BC}}{D_{BC} + D_{AC}} \text{Cap}_2, D_{BC})
\]

Objective fn. = \[
(\text{Cap}_1 - \frac{D_{AC}}{D_{BC} + D_{AC}} \text{Cap}_2) \text{fare}_{AB} + \frac{D_{BC}}{D_{BC} + D_{AC}} \text{Cap}_2 \text{fare}_{BC} +
\]

\[
+ \frac{D_{AC}}{D_{BC} + D_{AC}} \text{Cap}_2 \text{fare}_{AC} - (c_1 \text{Cap}_1 + c_2 \text{Cap}_2) =
\]

\[
= \text{Cap}_1 \{\text{fare}_{AB} - c_1 \} + \text{Cap}_2 \{\frac{1}{D_{BC} + D_{AC}} [D_{AC} (\text{fare}_{AC} - \text{fare}_{AB}) +
\]

\[
+ D_{BC} \text{fare}_{BC}] - c_2 \}
\]

As we can see, depending on the values of the actual demand factors associated with each leg and the actual capacity assignments, the objective function may take an entirely different form. In other words the
actual formula of the objective function depends on the relative size of the two ratios:

\[
\frac{Cap_1}{Cap_2} \text{ and } \frac{D_{AB} + D_{AC}}{D_{BC} + D_{AC}}.
\]

Note, that \(Cap_1\) and \(Cap_2\) are the decision variables. Thus if a new value is assigned to any of the decision variables, the above ratio needs to be evaluated repeatedly. Unfortunately, it seems that the objective function is not a convex linear function of the decision variables. We can show that the feasible region does not form a convex hull either. To prove this refer to Figure 8.6.

![Figure 8.6: Portraying the different directions of the objection functions in the two regions (Case 1 corresponds to the region above and Case 2 to the region below the “cut-line”)](image)

The two axes correspond to the values of the two decision variables \(Cap_1\) and \(Cap_2\). They are both limited from above by the corresponding maximum capacity, \(MaxCap\), constraints. The line

\[
\frac{Cap_1}{Cap_2} = \frac{D_{AB} + D_{AC}}{D_{BC} + D_{AC}}
\]

what we call as the “cut-line”, is also shown. This line divides the feasible set into two different areas. Below this line the critical ratio is

\[
\frac{Cap_1}{Cap_2} > \frac{D_{AB} + D_{AC}}{D_{BC} + D_{AC}}
\]

or in words the demand factor of leg 2 is bigger, therefore leg 2 is the bottle-neck leg. In this region the objective function takes the form that was given under case 2 in equation (8.13). For the area above the cut-line, the right hand side of the critical ratio is bigger — therefore the bottle neck leg for this set of
capacity combinations is leg 1. This corresponds to case 1 (equation (8.12)) , and the form and direction of the objective function “switches” to another direction. The directions of the two objective functions, for the two regions, are also shown on the picture.

The figure happened to show the two objective functions in a position pointing away from each other. Let us analyze this situation first. If the objective functions point apart, then it is easy to see that the feasible set is non-convex. Figure 8.7 helps to understand why. The figure shows the two areas of the feasible region have been rotated, so that the two objective functions are pointing now in the same direction. We generated this picture by rotating the area below the cut-line to the left around the point \( P \), where the boundaries of the two areas have met, until the two objective functions became parallel and pointed in the same direction. From the Figure 8.7 we can see that the feasible region of the problem is not convex with respect to the parallel objective functions.

![Graph showing the rotation of objective function and region of Case 2 parallel to the direction of the objective function of Case 1.](image)

**Figure 8.7:** Rotation of objective function and region of Case 2 parallel to the direction of the objective function of Case 1.

It is known that if the feasible region of a problem is not convex, then it is likely that linear programming optimization algorithms will find only a local maximum and may fail to encounter the combination of the decision variable values that give the global optimum.

Note, however, that there are certain conditions under which the convexity is not violated. For example this would be the case when the tangent of the cut-line equals to 1, that is, when the ratio
\[
\frac{D_{AB} + D_{AC}}{D_{BC} + D_{AC}} = 1.
\]

This means that the network is symmetrical with respect to the local demands. The situation when all legs have identical local and connecting demands is theoretically plausible, but in practice is rare and unexpected.

Next, let us analyze the situations when the two objective functions, unlike the case shown on Figure 8 7, point towards each other or are parallel. For the latter case, there is no need for rotating the regions, and one can see that for any reasonable objective function direction, the problem remains convex. For the former case, after rotating one of the regions, such that the two objective functions point to the same direction (this requires a rotation in a different direction than earlier), the feasible set also remains convex. The conclusion therefore is that, in cases when the two objective functions are pointing into the same direction or towards each other, the feasible set is convex. Thus non-convexity does not jeopardize the use of linear programming optimization methods.

The most interesting question at this point is to define under what conditions the objective functions point towards each other, point the same direction, or point apart. We will answer this question with the following proposition:

**Proposition:** Consider the fleet assignment problem with conditions defined in the model. Also let \( OF_i \) represent the form of the objective function associated with an ordered set of capacity assignments.

The different objective function vectors can exhibit the following three cases.

- **They are parallel and point in the same direction** if the fares associated with all \( od \) multiple-leg path OD demands are equal to the sum of the local OD fares of all legs along any \( od \) demand path, that is, if

  \[
  fare_{od} = \sum_{i \in P_{od}} fare_i \quad \forall od \in OD
  \]  

  \( (8.14) \)

  where \( P_{od} \) is the set of flight legs in the path of OD demand \( od \), and \( OD \) is the set of all origin-destination pairs that with demand.

- **They point towards each other** if the fares associated with all \( od \) multiple-leg path OD demands are bigger than the sum of the local OD fares of all legs along any \( od \) demand path, that is, if

  \[
  fare_{od} > \sum_{i \in P_{od}} fare_i \quad \forall od \in OD
  \]  

  \( (8.15) \)

  where \( P_{od} \) is the set of flight legs in the path of OD demand \( od \), and \( OD \) is the set of all origin-destination pairs that with demand.
◊ They point apart from each other if the fares associated with all od multiple-leg path OD
demands are smaller than the sum of the local OD fares of all legs along any od demand
path, that is, if
\[
\text{fare}_{od} < \sum_{i \in P_{od}} \text{fare}_i \quad \forall od \in OD
\]  
(8.16)
where $P_{od}$ is the set of flight legs in the path of OD demand $od$, and $OD$ is the set of all
origin-destination pairs that with demand.

Proof: We will prove the above proposition for the two-leg example, (refer to Figure 8.8). The proof can
be expanded for the general size network case. Let us first rearrange the objective functions for the two
cases presented earlier in equations (8.12) and (8.13). Therefore for case 1, when leg 1 is the bottle leg
neck, we can write:
\[
OF_1 = Cap_1 \left\{ \frac{1}{D_{AB} + D_{AC}} \left[ D_{AC} (\text{fare}_A - \text{fare}_{BC}) + D_{AB} \text{fare}_{AB} \right] - c_1 \right\} + Cap_2 \left\{ \text{fare}_{BC} - c_2 \right\}
\]
\[
= Cap_1 \left\{ \frac{D_{AC}}{D_{AB} + D_{AC}} (\text{fare}_A - \text{fare}_{BC} - \text{fare}_{AB}) + \frac{D_{AB} + D_{AC}}{D_{AB} + D_{AC}} \text{fare}_{AB} \right\} - Cap_1 c_1 +
+ Cap_2 (\text{fare}_{BC} - c_2) =
\]
\[
= Cap_1 \left\{ \frac{D_{AC}}{D_{AB} + D_{AC}} (\text{fare}_A - \text{fare}_{BC} - \text{fare}_{AB}) \right\} + Cap_1 (\text{fare}_{AB} - c_1) + Cap_2 (\text{fare}_{BC} - c_2)
\]  
(8.17)
and for case 2, when leg 2 is the bottle neck,
\[
OF_2 =
\]
\[
Cap_1 \left\{ \text{fare}_{AB} - c_1 \right\} + Cap_2 \left\{ \frac{1}{D_{BC} + D_{AC}} \left[ D_{AC} (\text{fare}_A - \text{fare}_{AB}) + D_{BC} \text{fare}_{BC} \right] - c_2 \right\}
\]
\[
= Cap_1 \left\{ \text{fare}_{AB} - c_1 \right\} + Cap_2 \left\{ \frac{D_{AC}}{D_{BC} + D_{AC}} (\text{fare}_A - \text{fare}_{AB} - \text{fare}_{BC}) +
+ \frac{D_{BC} + D_{AC}}{D_{BC} + D_{AC}} \text{fare}_{BC} \right\} - Cap_2 c_2
\]
\[
= Cap_1 \left\{ \frac{D_{AC}}{D_{BC} + D_{AC}} (\text{fare}_A - \text{fare}_{AB} - \text{fare}_{BC}) \right\} + Cap_1 (\text{fare}_{AB} - c_1) +
+ Cap_2 (\text{fare}_{BC} - c_2)
\]  
(8.18)
Figure 8.8: Description of the objective function deviations at the three different fare structures

From the above transformed versions of the two objective functions, we can see that they differ only in the first part of the summation. The second and third part of the summation are identical in both equations.
Let us start with the case when condition (8.14) holds. In our equation it transforms to the case when
\[ fare_{AC} = fare_{AB} + fare_{BC}. \]
If this is true then \( fare_{AC} - fare_{AB} - fare_{BC} = 0 \), and the first part in both objective functions will be equal to 0. Consequently, the direction of the two objective functions is the same, and parallel. We call this direction as the balanced direction. (See Figure 8.8a).

Next assume that the condition (8.15) holds. That is, the fare associated with the connecting demand is bigger than the sum of the two local fares: \( fare_{AC} > fare_{AB} + fare_{BC} \). This would also imply that \( fare_{AC} - fare_{AB} - fare_{BC} > 0 \), i.e., is positive. The positive sign, then, in the equations, would result in positive first parts. This would translate into increasing the total value of the multipliers of the decision variable \( Cap_1 \) in the first equation and increasing the total value of the multiplier of \( Cap_2 \) in the second equation. These increases each, relative to the balanced direction, turn (or rotate) the objective functions of case 1 and case 2 into the direction towards the \( Cap_1 \) and the \( Cap_2 \) axis, respectively. Since the feasible region belonging to case 1 is located opposite to the axis associated with the capacity of leg 1, and for case 2 vice versa, the two rotations relative to the balanced direction would direct the two objective functions pointing towards each other. This situation is shown on Figure 8.8b. As the figure shows, the objective function of case 1 was rotated to the right, and the objective function for case 2 was rotated to the left relative to the balanced direction.

Finally, the case of condition (8.16) is examined. In this case the fare associated with the connecting demand is smaller than the sum of the local fares: \( fare_{AC} < fare_{AB} + fare_{BC} \). This would also imply that \( fare_{AC} - fare_{AB} - fare_{BC} < 0 \), i.e., is negative. The negative sign would make the first part of the equations also negative. This would translate to decreasing the total value of the multipliers of decision variable \( Cap_1 \) in the first equation, and decreasing the total value of multiplier of \( Cap_2 \) in the second equation. The reductions in each equation, relative to the balanced direction, turn (or rotate) the objective functions for case 1 and case 2 away from the direction of the Cap1 and Cap2 axis, respectively. The result of the rotations in each area would result in objective function directions pointing away from each other. The situation is shown on Figure 8.8c. As the figure shows, the objective function of case 1 was rotated to the left from the balanced direction, and the objective function for case 2 was rotated to the right from the balanced direction.

The proof can be easily expanded for the multiple leg case. The number of cases that need to be evaluated depends on the network structure and the demand structures. For example, when n flight legs fly into a hub and m legs fly out, the number of different cases that need to be evaluated in the worst case may be equal to \((n+m)!\), i.e., the permutation of the possible sequences of flight legs becoming full. Each case would represent a possible sequence of legs becoming full in the booking process. For example, if
condition (8.16) prevailed, then the objective functions associated with the feasible region for each case would point apart from each other in the multidimensional space.

Unfortunately, the necessary conditions (8.14) and (8.15) for which the model remains a convex problem, do not make too much sense in practice. Particularly these conditions require that the connecting fares are not cheaper than the sum of local fares along the OD demand's path. If this were the case in reality, no passenger would buy connecting tickets. Passengers would be better off by buying the local tickets along the path of travel. Logically, this is not the case in airline networks. Thus the consequence is that the convexity conditions are not guaranteed. Condition (8.16) characterizes the real relationship of fares and thus this is typical. Yet, as we have shown in our discussions, this condition causes a non-convex set in the feasible region, and consequently when leg-dependence is considered, in typical cases most of the mathematical programming algorithms cannot be directly used to solve the leg-based fleet assignment problem.

### 8.2.1 Nested Algorithm with Embedded Simulation (NSFA Algorithm)

In this section, we introduce a multiple step iterative algorithm that takes into consideration leg-dependence effects and indirectly overcomes the non-convexity issue discussed above. In addition, the algorithm also is able to incorporate the dynamic effects of various yield management system approaches, hence overcoming many of the limitations of the formulations given in Section 8.1. The main part of the algorithm is an integer multicommodity flow problem formulation, which is enhanced by a nested dynamic Monte Carlo simulation process.

In this algorithm, it is possible to capture all the important issues that, as we have shown throughout the dissertation, may influence the actual ODF passenger flows, and thus, may affect the correct spill or load estimates. Particularly, the algorithm can incorporate to some extent the following issues into the decision process:

- It is able to incorporate the complex stochastic leg-dependent effects imposed by the total leg capacity limits or by the nested leg-based booking limits.

- It is able to capture the subtle differences of the different yield management optimization approaches in affecting the ODF passenger flows.

- It can take into consideration the effects of the actual ODF booking patterns.

Although the algorithm works for all type of airline networks, the algorithm works efficiently only if following two characteristics identify the actual network:

1. Not all flight legs of the network are affected by leg-dependence; and
2. Leg-dependence effects do not propagate far in the network.

We believe that the above two characteristics prevail in most airline networks. In Chapter 5 we addressed many network and demand attributes that affect leg-dependence effects. Recall that leg-dependence effects arise only if significant connecting demand is present in the network and connecting demands are being spilled. We believe that in many of today's airline networks one can find a large number of flights that spill very little, if any, demand. Most of the relatively high spilling flights include the morning and evening flights, when most of the business and leisure passengers are willing to travel. The rest of the flight legs, with some exceptions, have low demand levels relative to the available capacities and thus do not spill any substantial amount of passengers. Hence, these flights are not affected by leg-dependence effects either. Further, many international airlines operate a non, or only a quasi, hub-and-spoke network. In these networks, only a few selected flight legs serve as possible connections between OD pairs. Among these flight legs, the connecting traffic is significant and the leg-dependence issue is important. However, the rest of the flight legs that are not part of the mini connection banks, carry mostly local OD traffic -- thus leg-dependence does not affect these legs either.

We believe that in typical airline networks leg-dependence effects do not propagate far in the network. In Section 5.6 we have already argued, that leg-dependence effects can propagate over a path of overlapping connecting OD demands. We claim that the prevailing ODF passenger path itineraries in most of the airline networks do not overlap over a longer path. Assume, for example, two connecting banks at a city. The ODF demands overlap within the banks, but do not overlap between the banks. Therefore, leg-dependence effects due to a large amount of spill in one bank, do not affect the flight legs in the other bank. The same argument applies to other types of connections as well, where the different connection opportunities are not connected by overlapping connecting ODF demands.

The Nested Simulation Fleet Assignment (NSFA) Algorithm

The basis of this algorithm is that it assumes that leg-dependence effects can be localized to many different restricted sub-networks in the entire airline network. That it, it is possible to define sub-networks containing flight legs that are in a leg-dependent relationship with each other. Within the sub-network leg-dependence effects may affect the passenger flows, however, beyond the sub-network's limits, leg-dependence effects do not propagate. Thus, it is assumed that it is possible to define these mutually exclusive sub-networks in the network for which the above characteristic of limited effect propagation holds.

The advantage of defining these sub-networks is that it can be assumed that the non-convexity problem, discussed earlier, concerns only the selected sub-networks, and the rest of the network is not affected. Therefore, if the sub-networks are extracted from the network, then the remainder of the network is a network for which convexity prevails and linear programming algorithms could be used to find the
optimal fleeting. The trick is, how the extracted sub-networks are replaced in the original network, such
that the demand and network conditions within the extracted sub-networks are still captured and in the
same time the overall modified network remains convex with respect to its hull.

Figure 8.9 shows the main steps of the Nested Simulation Fleet Assignment (NSFA) algorithm.

**Step 0:** Define sub-networks (SNi) in the original network (N) for which it holds that the flight legs of each sub-
network are subject to the leg-dependence effects and these effects are limited within the boundaries of the
sub-network.

**Step 1:** Extract the sub-networks (SNi) from the original network.

**Step 2:** For each sub-network evaluate valid load (revenue) and spill (spill cost) estimates for different fleeting
combinations, taking into consideration leg-dependence effects and effects of the yield management
optimization approaches in use.

**Step 3:** Replace the extracted sub-network into the original network by so called “fleeting sink and source arcs and
“fleeting variables”. The “new fleeting arcs” assure that flow balance constraints are satisfied, and the
objective function updated with the “fleeting variables” will incorporate the valid revenue and spill cost
estimation results obtained in Step 2.

**Step 4:** Solve the modified network as a multicommodity flow integer program, leveraging on the solution
methodologies developed for the traditional fleet assignment models.

**Step 5:** Given the fleet assignment solution obtained in Step 4, find additional sub-networks that can be effect ed by
leg-dependence effects. If new sub-networks are found, then goto Step 1; otherwise Stop.

Figure 8.9: The stepwise description of the NSFA algorithm

**Determination of the Sub-networks (Step 1 and Step 5)**

Let us first concentrate on Step 0 and Step 5 of the algorithm. These steps are concerned with the
selection and identification of the sub-networks in the original network. These steps are not trivial, but
we speculate that, based on the analysis we provided in Chapter 5, an algorithm to select the sub-networks
can be developed.

A sub-network is defined over flight legs for which it holds that within the sub-network leg-dependence
effects may influence the passenger flows substantially and the effects do not propagate (at least in a
major way) beyond the limits of the sub-network (sometimes it may be an approximation of reality).
Recall, that in Chapter 5 we addressed many network and demand attributes that may influence the
magnitudes of the leg-dependence effects. First of all, leg-dependence effects may arise among flight legs
that share significant ODF demands, and it is expected that connecting ODF demands are going to be
spilled. Therefore, the sub-network identification algorithms should identify first those flight legs that
expect to spill significant amount of connecting ODF demand. It is possible to predict to a certain extent
which flight leg would spill much demand. For example, flight legs that have very high unconstrained leg
demands relative to the capacities of fleet types that are expected to be assigned to the flight leg, are all very likely to spill. The connecting passenger ratio defined in Section 5.1 can give additional information about the possible effects of leg-dependence. To determine in advance more flight legs that are expected to spill, historic observation of previous assignments can be considered. Note that not all flight legs that expect to spill connecting demand need to be identified in the first iteration. Since the algorithm we present here is an iterative algorithm, in Step 0 it is possible to define only a limited number of sub-networks. Then, at the end of the iteration, in Step 5, additional sub-networks can be defined based on the fleet assignment solution obtained in the previous step. Searching through the assignments for each leg, it is possible to determine the flight legs (that are not yet part of any sub-networks) that spill significant amounts of passengers.

Once the flight legs that are expected to spill are identified, the other flight legs that could be affected by leg-dependence, and thus should be included in the sub-networks, should be identified. Based on the discussion in Section 5.6, all flight legs that share reasonable ODF demands with the leg that is expected to spill are subject to leg-dependence effects. All these flight legs should be included in a sub-network. We expect that these flight legs can be easily identified from the ODF demand characteristics. In situations when it is not the case, assuming that leg-dependence effects in typical airline networks do not propagate far, only flight legs that are "close" to the spilling leg would be considered as part of the sub-network.

Note that network connectivity issues, that were discussed in Section 5.4, should be also considered in the identification of sub-networks. On each of the identified flight legs, the spilled ODF demands need to be assessed and its effects on other flight legs with respect to network connectivity (concentrating or dispersing leg-dependence effects) issue needs to be evaluated. For example, if either in Step 0 or Step 5, a number of flight legs are identified to possibly spill demand and they are connected by overlapping ODF demands (see for example legs 1 and 3 in Figures 5.7 and 5.8), then the network connectivity issues should be evaluated. In general, if in a connecting complex more than one flight leg expects to spill, then a sub-network should be defined for all spilling flight legs and flight legs that share ODF demands. If however, in the analysis it is shown that only one flight leg is identified to spill potentially and the leg-dependence effects can be dispersed on many other legs, then no sub-network needs to be identified.

In Chapter 5, and throughout the dissertation we claimed that the characteristics of airline OD demands is that they do not overlap over a long path, but rather overlapping OD demands are clustered into mutually exclusive sub-networks (i.e., two connecting banks in a hub city). Further, we have shown that the

---

7 For example there are short haul domestic flight legs that most probably are serviced by only narrow-body aircraft due to marketing or operating constraints. Therefore, even if there are wide-body aircraft available in the fleet they are not expected to be assigned to the flight leg.
probability of leg-dependence effects to propagate far in the network, even in the case when long overlapping paths exist, is small. Therefore, in the process of defining a sub-network, only those flight legs are included in the sub-network which directly share a common ODF demand at least with another flight leg in the same sub-network that expected to spill ODF demand.

Although we know that the selection and identification of the sub-networks is not trivial, we believe that using the above arguments and experience it is possible to determine in advance at least some of these sub-networks and develop an algorithm that does it throughout the iterations. Nevertheless, the identification of sub-networks and the analysis of airline networks in this light requires additional research efforts.

**Estimating Spill and Spill Costs (Load and Revenues) for each Sub-network (Step 2)**

In this algorithm we assume that the airline can not control ODF demands on the OD level. Therefore, as we have shown earlier, linear programming approaches are not applicable for obtaining the optimal fleet assignment. Further, the complex effects of the different yield management approaches may affect the ODF passenger flows in the sub-network that also need to be taken into consideration.

In accordance to our results presented in the earlier chapters of the dissertation, we believe that the complex interactions of the leg-dependent passenger flows under the prevailing yield management optimization can be captured well by the use of a Monte Carlo simulation approach, where all imposed capacity limits of the various yield management optimization approaches can be directly modeled. (The simulation approach was presented and discussed in detail in Section 6.2.2)

Corresponding to our views about the interpretation of spill in leg-dependent networks (refer to Section 7.7), we believe that load and spill should be interpreted as a network-wide concept in networks affected by leg-dependence. Consequently, instead of determining load or spill for the flight legs of a sub-network, we determine load and spill for a given fleeting combination. Given the fleeting combinations for the sub-networks, using the Monte Carlo based simulation approach, it is possible to capture the valid load (revenue) and spill (spill cost) estimates. By embedding the yield management optimization process into the simulation (as discussed in Section 6.2.2), its complex effects on the actual passenger flows in the sub-network can be also captured. Depending on the size of the sub-network and the available fleeting combinations either all or a sub-set of the fleeting combinations are evaluated. Only fleeting

---

8 Here we refer to the different effects of the various yield management OD heuristic approaches addressed in Chapter 7.
combinations that satisfy the cover and size constraints for that sub-network are evaluated. For each fleeting combination then the valid expected revenue of spill costs can be determined.

Either all fleeting combinations are evaluated (when the sub-network size and the number of realistically assigned fleet types to the flight legs are small) or we select a number of possible combinations. At least one of the fleeting combinations that we evaluate is determined by the SYMFAM or the SYMFAMCG models introduced earlier. Other combinations may include the fleeting that is used currently in the network, and a fleeting that is the result of the traditional fleet assignment formulations, (which do not take leg dependence and yield management effects into consideration). Additional fleeting can be generated by swapping different fleet types in the sub-network, by assuring that flow balances at each node and for each fleet types in the sub-network are maintained. During the iterations, some of the considered combinations for a sub-network could incorporate information about the available fleet types at nodes that connect the sub-network to the rest of the network being assigned in the current fleeting solution. It is important to include some fleeting that have seemingly “bad” fleeting costs (sum of the operating and spill costs). In the final overall solution, it is possible that these fleeting combinations are the best with respect to the overall network fleet assignment solution. Note also that not all possible fleeting combinations need to be evaluated. Certain combinations can be discarded with high probability. For example, not all fleet types could be considered for each flight leg. For flight legs that have traditionally wide-body aircraft being assigned, the narrow-body fleet types can be ignored. This limits the number of fleeting combinations.

Estimating the load and spill for different fleeting in each sub-network can be a relatively fast process. If network sizes are small, then the simulation is also fast. Our experience shows, that relatively low iterations in the simulation (between 50 to 100) can give accurate results even when complex yield management approaches are embedded and evaluated. Simulation run-time results for different networks on different computer platforms were given in Table 6.1. Since the sub-networks are disjoint networks, the simulated network sizes become small, which benefits the simulation run times. Further, because of the above, it is possible to parallelize the simulation process, empowering additionally the computation process.

The result of the simulation therefore, a valid spill and spill cost (load and revenue) estimate for each selected fleeting in each sub-network. The total cost (or profit) associated with the fleeting combination is the sum of the spill costs and the operating costs of aircraft types associated with the actual fleeting (expected profits equal to the expected revenues minus the operating costs associated with a fleeting).

\footnote{Although it may valid for the sub-network, may not be valid for entire network. Step 4 will assure that the flow balance and size constraints for the entire network are satisfied.}
Extracting and Replacing the Sub-networks (Step 1 and 3)

In our algorithm we accomplish this task by replacing the extracted sub-network in the original network by so called *fleeting arcs*. (Assume that for each fleeting combination a fleeting node and the corresponding fleeting arcs are defined. Although the representation with fleeting nodes would make the replacement procedure easier to understand, because only the fleeting arcs going to be included in the formulation, we will limit our discussion to them.) To illustrate the replacement process refer to Figure 8.10 which shows two sub-networks defined in an extract of a time-space network example. (Note that a time-space network is defined for each fleet type in the network.) Sub-network $A$ contains flight legs $(4,6)$, $(3,6)$, $(6,11)$, and $(7,10)$; while sub-network $B$ contains flight legs $(9,12)$ and $(12,13)$. The sub-networks will be replaced by *fleeting sink* and *source* arcs by the following rule. The fleeting sink and source arcs are defined to assure the flow balances at each node, taking into consideration the fleet type commodity flows associated with a fleeting combination in the deleted sub-network.

![Diagram of time-space network](image)

Figure 8.10: Two sub-networks (A and B) defined in a time-space network

All flight arcs that part of the sub-network will be deleted. For each node, that has a deleted flight arc, so called *fleeting source* and *sink* arcs are being defined and added in the time-space network of each fleet type. (Note that fleeting arcs carry commodity specific "supply and demand" information, therefore need to be defined in all time-space networks of each fleet type). Depending on the orientation of the deleted flight arc, $m$ number of sink or source arcs are added to a node in each fleet type's time-space network, where $m$ refers to the number of different fleeting that are being evaluated in Step 2 for the corresponding sub-network. If the deleted flight arc was emanating from the node under consideration, then there will be $m$ sink arcs defined and added to the node in each fleet type's network. If the deleted flight arc was pointing towards a node, then there will be $m$ source nodes defined and added to the node. If a node has
both an emanating and an incoming flight arc being deleted, then both sink and source arcs need to be added to the network. Figure 8.11 shows the added sink and source arcs for sub-network A after the replacement procedure in the sample network. For example at node 4, flight leg (4,6), which was emanating from the node, was deleted. Therefore, assuming that there are \( m_A \) different fleeting combinations evaluated for sub-network \( A \), for node 4 in the network of each fleet type, \( m_A \) source arcs are defined and added (see Figure 8.11). Note that if a node has all its flight arcs deleted, (for example node 6 in our sample network), then it is possible to delete the node from the network and merge the two ground arcs associated with the node (see Figure 8.11).

![Figure 8.11: The replacement of sub-network A by \( m \) source and sink arcs](image)

Note that the replacement of the sub-networks involves the deletion of flight arcs, and at the same time the definition and addition of new (sink and source) arcs into the original network. The net effect depends on the size and structure of the sub-networks (certain flight legs can be deleted without replacing it by sink or source arcs) and on the number of different fleeting combinations that are going to be represented.

Although other outcomes are also possible, generally the net effects result in increase of the number of arcs in the network. Although many new arcs are being added to the network, in the replacement process no additional nodes were defined. Moreover, as it is the case for node 6 in the sample network, it is possible to eliminate nodes from the network representation\(^{10}\).

\(^{10}\) This characteristic is very important with respect to the run-time of the algorithms. Experience shows that the number of constraints (rows) are very important in influencing the running time of the algorithm.
Solving the Modified Network as an Integer Combined Multicommodity Flow Problem (Step 4)

Once the sub-networks are replaced by the fleeting nodes, it is possible to solve the fleet assignment problem as an combined integer multicommodity flow problem, similar to that used by the traditional fleet assignment formulations. The difference is that in this formulation the parts of the network that are subject to leg-dependence effects, are replaced by fleeting arcs and the corresponding fleeting decision variables. In this network each fleeting combination of a replaced sub-network is represented by a fleeting variable, which is added to the objective function.

In the modified network, we assume that only those flight legs are represented by a flight arc that are not affected by leg-dependence effects. Therefore, the non-convexity issue, discussed earlier, does not affect this part of the formulation. The network segments (sub-networks), that are subject to leg-dependence effects, are removed from the network and are represented by a fleeting decision variable with an associated single fleeting cost objective coefficient. Since the sub-networks that give rise to the non-convexity issue are eliminated and are represented by different fleeting variables with an associated cost coefficient, the resulting formulation results in a convex hull in the feasible region, and a concave objective function when the fleet assignment formulation is a profit maximization formulation. If the approach is a cost minimization fleet assignment formulation, then the resulting formulation results in a convex objective function\textsuperscript{11}. Consequently, linear and integer programming approaches can be directly applied to the modified (replaced) representation of the fleet assignment problem. It also means that we can leverage on the existing advanced fleet assignment solution methodologies (e.g., the solution methodologies developed for the Coldstart model [43]) only with minor modifications to the overall formulation.

The cost objective coefficients in the mathematical program are the following. For each variable, $X_{fl}$, that is associated with assigning a fleet type $f$ to a single flight leg $l$ of the original network, the associated cost, $Cost_{fl}$, is the sum of the operating cost of fleet type $f$ on leg $l$ plus the spill cost estimated by either the recursive formula given in equation (7.4) or (7.6). Using this spill cost estimation procedure, the effects of the yield management systems on a single leg-independent flight leg are taken into account.

For decision variables, $Z_{sm}$, that represent a fleeting combination $m$ of a replaced sub-network $s$, the associated costs are obtained in Step 2. The cost associated with the fleeting combination are the sum of the total operating costs associated with the actual fleeting in the sub-network and the estimated valid spill costs for the fleeting, taking into consideration leg-dependence effects and effects of the yield management systems.

\textsuperscript{11} In case when we assume the cost minimization fleet assignment formulation, the requirement is that the objective function is convex with respect to the decision variable values.
Two other important variables that are associated with a fleeting combination $m$ at a node $n$ in the time-space network of fleet type $f$ are $SR_{f,n,m}$ that represents the source arcs, and $SK_{f,n,m}$ that represents the sink arcs. $SR_{f,n,m}$ and $SK_{f,n,m}$ represent the number of aircraft of each fleet type provided and/or needed (source or sink) at this location (node $n$) if fleeting combination $m$ is selected for a sub-network. These variables are part of the flow balance constraints and assure the flow balance between the selected sub-network and the rest of the network.

The formulation of the modified multic commodity IP formulation with fleeting nodes is the following:

$$\min \sum \sum_{l \in L} \sum_{f \in F} Cost_{f,l} X_{f,l} + \sum \sum_{s \in SN} \sum_{m \in FC_{s}} FCost_{s,m} Z_{s,m}$$  \hspace{1cm} \text{(8.19a)}$$

s.t.

$$\sum_{f \in F} X_{f,l} = 1 \quad \forall l \in L'$$ \hspace{1cm} \text{(8.19b)}

$$\sum_{m \in FC_{s}} Z_{s,m} = 1 \quad \forall s \in SN$$ \hspace{1cm} \text{(8.19c)}

$$\sum_{l \in l(n)} X_{f,l} + Y_{f,*n} - \sum_{l \in O(n)} X_{f,l} - Y_{f,*n} +$$

$$+ \sum_{m \in FC_{s}} (SR_{f,n,m} - SK_{f,n,m}) Z_{s,m} = 0 \quad \forall f \in F, n \in N', s \in SN(n)$$ \hspace{1cm} \text{(8.19d)}

$$\sum_{l \in P(f)} X_{f,l} + \sum_{l \in Q(f)} Y_{f,l} \leq S(f) \quad \forall f \in F$$ \hspace{1cm} \text{(8.19e)}

$$Y_{f,*n} \geq 0 \quad \forall n \in N', f \in F$$ \hspace{1cm} \text{(8.19f)}

$$X_{f,l} \in \{0,1\} \quad \forall f \in F, l \in L'$$ \hspace{1cm} \text{(8.19g)}

$$Z_{s,m} \in \{0,1\} \quad \forall s \in SN, m \in FC_{s}$$ \hspace{1cm} \text{(8.19h)}

Where

- $Cost_{f,l}$ is the operating cost plus the estimated spill costs of flight leg $l$ when fleet type $f$ is assigned to the leg, where spill costs are estimated by either equation the recursive formula given in equation 7.4 or 7.6, taking into consideration the effects of yield management systems on the single leg-independent leg $l$. (For these legs leg-independence is assumed.)

- $FCost_{s,m}$ is the fleeting cost associated with fleeting $m$ of sub-network $s$, where fleeting costs are estimated in Step 2 by a Monte Carlo simulation.

- $s \in SN$ is a sub-network, where the set $SN$ represents all the sub-networks that are defined in the original network at the time. $SN(n)$ represents the sub-network(s) that are incidental to node $n$, i.e., the sub-network for which either source arcs or sink arcs (or both) were assigned at node $n$.

- $m \in FC_{s}$ is a fleeting combination associated with sub-network $s$, where set $FC_{s}$ represents all fleeting combinations of sub-network $s$. 

211
- $SR_{f,a,n}$ represents the source arc and $SK_{f,a,m}$ represents the sink arc variable associated with fleeting $m$ at node $n$ for fleet type $f$. They represent the number of aircraft of each fleet type provided and/or needed (source or sink) at this location if fleeting combination $m$ is selected for a sub-network.

- $L'$ and $N'$ represent the set of flight legs and set of nodes, respectively, that are currently not part of any sub-networks.

- $X_{f,l}$ is a binary decision variable, which has the value 1 if fleet type $f$ is selected on leg $l$, and 0 otherwise.

- $Z_{s,m}$ is a binary decision variable, which has the value 1 if fleeting $m$ is selected for sub-network $s$, and 0 otherwise.

- The rest of the variables were defined earlier.

The first term of the objective function represents the costs associated with the assignment of the different fleet types to flight legs that are part of the original time-space network and not part of any of the defined sub-networks. The second term represents the costs associated with the selected fleeting in each of the sub-networks that are defined in the network at the time.

Constraints (8.19b) and (8.19c) are cover constraints, assuring that each flight leg and each sub-network is covered by a fleet type or a fleeting, respectively. Constraints (8.19d) are the flow balance constraints, where besides the original flight and ground arcs the source and sink arcs are also taken into consideration. That is, for each fleet type at each node the sum of the incoming aircraft (sum of the incoming flight arcs, the number of aircraft on ground at time before the node, and the number of aircraft in the source arcs) should be equal to the number of the outgoing aircraft (sum of the leaving flight arcs, the number of aircraft on ground at time after the node, and the number of aircraft in the sink arcs).

Constraints (8.19e) are the fleet size constraints. (Note that it is assumed that the time-line along which aircraft are counted does not cut through any sub-networks.) Constraints (8.19f) assures that ground arc variables are not negative. Constraints (8.19g) and (8.19h) assure that the fleeting solution is integer.

The schematic diagram of the above formulation for the network is given in Figure 8.12.
The NSFA algorithm can be used iteratively, until a solution is found where no more sub-networks can be determined, or until the maximum allowable run-time is reached. The algorithm (in case of rational sub-network identification) provides the same or better fleeting solution than the traditional fleet assignment solutions, which do not consider leg-dependence effects or the effects of yield management systems.

Note that it is possible to start the algorithm with no initial sub-networks defined in Step 0. In this case, in Step 5, the traditional fleet assignment problem is being solved. In the next iterations then, by selecting sub-networks and evaluating additional (different) fleeting combinations, the algorithm can take into consideration the network effects and effects of the yield management system and improve on the overall fleeting. This way the NSFA algorithm gets at least as good or better fleeting solutions than the solution of the traditional fleet assignment formulations.

We have already noted above that in the replacement procedure, no additional nodes need to be defined. This also means, although new variables (columns) are being added to the original network formulation,
no additional nodes (additional rows in the constraint matrix) were added. In particular, it can often happen that as a net effect, the number of the nodes (rows) in the network representation decreases. The deletion of a node of the network, actually means that as many rows are deleted from the constraint matrix as different fleet types are available\textsuperscript{12}.

The above technique is very important from the perspective of run-time efficiency. The increase in the number of constraints (number of rows in the constraint matrix) is much inferior to the increase in the number of variables (number of columns) \cite{8}.

The algorithm can work well if the basic assumptions (i.e., that not all flight legs of the network are affected by leg-dependence and that leg-dependence effects do not propagate far) hold. If it is possible to find only a limited number of distinguishable sub-networks in the network, then the algorithm may perform very well and provide good results. If this is not true, then too many sub-networks need to be defined in the original network, and the efficiency of the algorithm deteriorates.

As a heuristic approach, it is possible to limit the definition of the sub-networks to a limited number of important connections. Taking only these limited number of sub-networks into consideration may improve the solution. Also, the solution is subject to the considered fleeting combinations in each sub-network. It is possible to create all fleeting combinations, but time limitations may require that the number of evaluated combinations be limited. It is theoretically possible that the optimal solution would use a combination that is not included in the set of evaluated combinations. However, with a good selection procedure, the number of such cases can be kept low. Nevertheless, it can be guaranteed that the optimal solution is found if all possible combinations are considered. If this is not the case, we can view the approach as a heuristic that improves the solution.

The algorithm can be very useful in airline networks where the not all flight legs are part of a strictly defined hub complex. Networks of many international airlines fall into this category. Despite the hub structure of many international airlines, their actual schedule prevents the characterization of their operations as strictly hub-and-spoke. Consequently, a limited number of flight legs share significant ODF demand. The number of connection complexes (or sub-networks) that provide good connectionas and reasonable shared ODF traffic is low.

Nevertheless, this algorithm can be beneficial to airlines with any type of network. As we have noted, the algorithm, (in the case of rational sub-network determinations), gives at least as good a fleeting solution as the traditional fleet assignment model. However, this algorithm, even if only a few sub-networks are defined, has the potential to improve on the selected fleeting. Therefore, the use of this algorithm, as a

\textsuperscript{12} This is because there is a copy for the underlying time-space network for each fleet type in fleet assignment formulations.
heuristic iterative approach, can be favorable even in fully hub-and-spoke networks, where most of the flights should be defined as part of a sub-network.

To summarize, the following characterize the NSFA algorithm:

- It is based on the assumption that not all flight legs of the network are affected by leg-dependence and that leg-dependence effects do not propagate far.
- It takes into consideration the complex leg-dependence effects in a stochastic environment.
- Detailed yield management information and the effects of the different yield management approaches (OD heuristic) are captured by simulation.
- Valid spill costs can be obtained taking into consideration leg-dependence effects and the effects of the yield management decisions.
- By defining sub-networks with small sizes, the network sizes for simulation and the run-time can be kept small.
- It overcomes the non-convexity problem presented in networks where the ability to control passenger flows on OD level is lacking.
- It leverages on the state-of-the-art existing large scale fleet assignment solution methodologies.
- The algorithm has the potential to provide a better fleeting solution than that of the traditional fleet assignment models in any type of network. As a heuristic approach, it can improve on the fleeting in any network, however, its efficiency is the best in networks for which the basic network assumption of the model hold.

8.3 Summary

In this chapter, we have analyzed how network effects may affect fleet assignment formulations. We classified airline networks into two distinct groups according to the assumed ability of the used yield management system to control OD demands. We showed that in cases when the airline is able to control ODF passenger flows by setting booking limits on an ODF level, then it is possible to formulate the combined yield management and fleet assignment problem that takes into consideration the leg-dependence issue as an IP problem. We have given deterministic (DYMFAM) and stochastic (SYMFMAM) formulations. The yield management decisions were modeled by embedding the yield management decisions into the fleet assignment formulation. For both the deterministic and the stochastic case, formulations based on column generation were also presented (DYMFAMCG and SYMFAMCG). Although the application of these models at the current level of airline practice is mostly limited (due to the lack of ability to control passenger flows on OD level), it was argued that in certain applications, or for
applications in the future, the presented models can be useful. Nevertheless, we find it important to
mention that we do not believe that at the current level of practice and airline network characteristics,
these models should be used for fleet assignment. We believe that instead, the NSFA approach with
embedded simulations should be used for the purpose of fleet assignment.

In the chapter we have also shown that in cases when the airline is not able to control the passenger flows
on OD level, then network effects create an important modeling difficulty. We have shown that under
typical fare structures, fleet assignment formulation constitute a non-convex hull in their feasible set,
while under other fare conditions this problem does not arise. Unfortunately, typical and rational airline
fare structures can be characterized with fare structures that imply non-convexity. Consequently, if leg-
dependence is considered, the fleet assignment problem for typical airline structures cannot be solved by
the direct and sole use of linear programming methods.

This chapter also includes a nested embedded simulation fleet assignment model (NSFA), that overcomes
the problem of non-convexity. The algorithm is based on the assumption that not all flight legs of the
network are affected by leg-dependence and that leg-dependence effects do not propagate far. The
algorithm leverages on the existing fleet assignment solution techniques and solution methodologies,
enhanced by a Monte Carlo based simulation. For parts of the networks where leg-dependence is an issue,
simulation is used to estimate accurately the valid spill and spill costs. The advantage of the method is
that it is possible to incorporate the actual yield management approach being used by the airline,
therefore, the complex dynamic effects of the yield management optimization approaches can be also
captured.
9. Conclusions

9.1 Summary of Research Findings and Contributions

The emphasis of this dissertation has been on the fleet assignment decision process, taking into account network (leg-dependence) effects and yield management practices. In light of this, many of the fundamental issues related to incorporating network and yield management effects into the fleet assignment decision process were addressed.

The general contribution of this dissertation is to recognizing the importance of these effects and incorporating them into the fleet assignment decision process. There are several contributions made in this research in terms of both theoretical analysis and practical approaches. We have identified the important demand and network characteristics under which network and YM effects influence spill estimates and have evaluated the significance of these effects in typical stochastic airline situations. Based on various network and demand examples, we have shown that the influence of network and YM effects on spill estimates are large enough to change the fleet assignment solution. We have shown that incorporating these effects into the decision process airlines can achieve higher profits. With this in mind, we analyzed the modeling implications of these effects on fleet assignment optimization, and as a contribution we outlined various analytic estimation models, optimization algorithms and heuristics that incorporate network and YM effects into the spill cost estimation and the fleet assignment decision process. In the following, we present a detailed overview of the research findings and contributions.

In an effort to isolate the effects of leg-dependence and yield management practices, the analysis was based on a simplified representation of the underlying environment, with such factors and practices as overbooking, recapture of spilled passengers on other flights or paths, no-shows, and upgrades not included. Nevertheless, as one of the initial efforts to identify the fundamental issues and to develop approaches to incorporate the important network and YM effects into the overall decision process, this research took an important step toward a deeper understanding of the underlying issues.

The driving force behind incorporating network and YM effects into the spill estimation and fleet assignment decisions is to capture additional profits beyond those being realized from current approaches. The dissertation has shown that by incorporating these effects into the decision process, airlines can capture significantly higher profits when the assignment is applied over a longer time-period.

Throughout the dissertation we addressed the possible implications of leg-dependence effects on the fleet assignment solution. Starting from the deterministic demand case, we expanded the analysis to more typical stochastic demand cases. We have shown that due to the censoring effects in leg-dependent networks, the "achievable traffic" demand densities on certain flight legs are different from the
unconstrained leg demand densities. The achievable traffic densities, when local and connecting demands are present on a flight leg, are left skewed and shifted to the left (in the direction of lower values) relative to the unconstrained leg demand densities. Since traditional leg-independent approaches use the unconstrained demand densities as a basis for load and spill estimates, these approaches overestimate both loads (expected revenues) and spill (spill costs) on flight legs affected by censored OD demands. We have also shown that the overestimation from leg-dependence effects are not uniform across the flight legs or at different demand to capacity ratios.

By analyzing network effects in different demand and network situations, we identified the important demand and network characteristics that influence the leg-dependence effects and extent on spill estimates in a network. We noted that leg-dependence effects arise only if connecting (multiple-leg path) OD demands are present and spilled in the network. Further, we have shown that network connectivity can play a very important role. If spill due to a bounding capacity constraint on a flight leg is dispersed onto many connecting flight legs, the actual leg-dependence effect on each influenced flight leg may diminish significantly. This represents a case when, although the spill on a flight leg is high, its effect on the rest of the flight legs may become negligible. On the other hand, when spill occurs on flight legs that feed a connecting flight leg, censoring effects add up. This case represents the situation when, despite the fact that the amount of spill is very small on each of the spilling flight legs, the overall leg-dependence effect on the fed leg is significant. The issue of the direction and boundaries of leg-dependence propagation were also addressed. We showed that the essential conditions for leg-dependence propagation are overlapping OD demand paths, and that spill occurs on all flight legs along the propagation’s path. We have also argued that in typical stochastic airline networks, leg-dependence effect propagation due to a capacity limit change on a flight leg may be very much bounded and these boundaries can be easily determined.

We analyzed the interrelationships of yield management practices and fleet assignment optimization. We demonstrated that yield management systems, through their booking limits, influence both the total number of passengers spilled and their fare class mix, which in turn affects the total spill costs of given aircraft capacities assigned to the flight legs. We have shown that, depending on the estimation method used, the resulting spill estimates can differ significantly. Under assumed discount ratios, which are representative of the fare structures of US airlines, our results have shown that by not incorporating yield management effects into the analysis, airlines may underestimate spill by as many as 4 to 7 passengers on a single flight leg. The booking patterns and interspersed sequences of requests for different fare classes, which are captured by yield management systems as a multiple-period representation of demand, also influence both the total number and fare class mix of spilled passengers. We demonstrated that the fare class mix of spill is not constant at various capacities, even for a given demand pattern. On the basis of the numerical examples presented, we have concluded that the direction and magnitude of the bias in spill
estimates from the currently used approaches are functions of many parameters, including the mix of fare class demands, the booking patterns of the different classes prior to departure, the relative fare values of the different fare classes, the overall demand factor for each capacity, as well as the yield management approach being used, with its booking class architecture and type of control being considered. This conclusion means that ad hoc adjustments to either the capacity or the spill-fare based on general rules of thumb are unlikely to provide accurate spill or spill cost estimates when yield management systems are at work. The recursive fare class spill models introduced in the dissertation can be used to obtain the most accurate estimates of spill for each specific flight demand scenario when leg-independent assumptions are valid.

Our analysis also showed that different yield management approaches can influence the ODF passenger flows in the network in different ways. This implies that the leg-dependent spill and spill cost estimates are a function of the employed yield management approach. Using typical airline hub networks, our results showed that the differences between the estimates of the different yield management approaches (Conventional, Value Based, and OD Bid Price Heuristic YM approaches) exceeded many hundreds and in some cases thousands of dollars per flight leg. Our results have also shown that the differences among the spill and spill costs are not uniform. The examples revealed that under different YM approaches the optimal fleet assignment solutions may be different.

Spill for fleet assignment purposes is estimated for a longer time-period. Traditionally, airlines represent the average demand for the time-period by a horizontally aggregated combined demand probability density function. Spill estimates are then based on this combined density. To avoid the horizontal bias, we recommend that spill costs be estimated individually (on daily or classified data) and that the average spill cost of the time-period be estimated as the average of the individual estimates. We argue that yield management and leg-dependence effects can be captured correctly by this disaggregated approach.

We have demonstrated through various typical airline examples that it is possible that ignoring yield management and/or network effects in the spill cost estimation can lead to sub-optimal fleet assignment solutions. With this in mind, a combination of a variety of estimation approaches and optimization models were developed. A detailed Monte Carlo simulation based approach was developed, which is a flexible and useful tool to estimate demand based estimates (spill, spill cost, load, and expected revenues) in airline networks where complex stochastic network and leg-dependence effects need to be modeled. This approach overcomes the limitations of the two previously developed approaches. The advantage of the simulation approach is that aside from the complex network effects, it can directly capture the dynamic interactions of the modeled YM approaches.

By synthesizing the research findings, various optimization approaches have been presented. We classified airline networks into two distinct groups according to the assumed ability of the yield
management system to control OD demands. We showed that in cases when the airline is able to control ODF passenger flows by setting booking limits on an ODF level, it is possible to formulate the fleet assignment problem that takes into consideration the leg-dependence issue as an IP problem. We have given deterministic (DYMFM) and stochastic (SYMFM) combined passenger flow and fleet assignment formulations, in which simplified yield management decisions were modeled by embedding the yield management decisions into the fleet assignment formulation. Due to the anticipated large size in the constraint matrix of a possible application, column generation-based formulations were also presented (DYMFMCG and SYMFMCG). These algorithms, while interesting from the perspective of the solution methodology, do not model realistically the booking process and the dynamic YM environment. Therefore, their direct application to the fleet assignment problem is not recommended. Nevertheless, we believe that it may initiate some further research and lead to possible applicability at a later time.

In the dissertation we have also addressed the more realistic case when an airline is not able to control the passenger flows on OD level. We have shown that network effects create important modeling difficulties in this case. We have shown formally that under certain fare structures, fleet assignment formulations constitute a non-convex hull in the feasible set, whereas under other fare conditions this problem does not arise. Unfortunately, typical and rational airline fare structures can be characterized with fare structures that imply non-convexity. Consequently, if leg-dependence is considered, the fleet assignment problem for typical airline structures cannot be solved by the direct and sole use of linear programming methods.

This dissertation presents a new Nested embedded Simulation Fleet Assignment model (NSFA), that overcomes the problem of non-convexity. The algorithm is based on the assumption that not all flight legs of the network are affected by leg-dependence and that leg-dependence effects do not propagate far. The algorithm leverages on the existing fleet assignment solution techniques and solution methodologies, enhanced by a Monte Carlo based simulation. For parts of the networks, where leg-dependence is an issue, simulation is used to assess valid spill and spill costs. The advantage of this method is that it is possible to incorporate the actual yield management approach being used by the airline; therefore, the complex dynamic effects of the yield management optimization approaches, along with network effects, can be also captured.

We can summarize the research findings and contributions of this dissertation in one paragraph: Our analysis has shown that network and yield management effects can significantly influence fleet assignment solutions in typical airline networks. We developed and presented various models, approaches, and algorithms in this dissertation that can help to capture and incorporate these effects into the decision process. We assumed throughout the dissertation that demand estimates are available at the level of the origin-destination path of a passenger. Since the developed algorithms use this information, airlines have an additional incentive to continue in their effort to develop and maintain OD databases and
estimate OD passenger flows in the network. The dissertation did not consider or solve all the important issues relating to the fleet assignment problem; however, as an initial effort to capture network and yield management effects in the estimation process, this dissertation has made an important step toward a better solution of the fleet assignment problem.

9.2 Further Research Directions

While significant progress into better modeling the fleet assignment problem has been made, there is still further research to be done as an extension of this work. As stated initially, a number of important factors such as the practice of overbooking, recapture of spilled passengers on other flights or paths, no-shows, and upgrades were not included. The influence of these issues on the fleet assignment solutions must be also evaluated, and if significant, included in an advanced fleet assignment approach. Much previous and on-going research tries to address these issues in different research objectives [26], [47]. Those research results should complement the results of this dissertation in developing a fleet assignment solution approach that models the underlying passenger decision process and the operational environment in a more realistic and appropriate way.

In this dissertation we have shown that network and YM effects are significant. We have shown that by considering these effects, the best fleet assignment solutions can be different from traditional approaches that ignore these effects. We have evaluated the cost savings, or additional profits, resulting from better assignment in typical but small airline networks. Additional analysis with respect to cost savings, or profit increases, on a full airline network should be conducted. It would be interesting to know in a full airline network, how many of the flight legs would be served by a different fleet type when these effects are considered, and to determine the overall financial implications.

In this dissertation we have outlined a fleet assignment algorithm (NSFA) that takes into consideration the network and YM effects. The algorithm is based on the assumption that leg-dependence effects can be localized to a number of relatively small size sub-networks. Although it is assumed that this assumption holds for many airline networks, additional analysis is needed to evaluate to what extent these assumptions hold in general. The algorithms also assumed that if these assumption hold, then it is relatively straightforward to determine the sub-networks. Algorithms need to be developed that search for these sub-networks and define their boundaries. In general, by evaluating the credibility of the basic assumptions, the robustness of the algorithm needs to be evaluated.

Another area of further research is to extend the research of the very short run dynamic fleet assignment models (the concepts that were developed by Berge and Hopperstad [15] in their D³ model) with the research finding of this dissertation. The process of swapping fleet types shortly before departure, considering the bookings on hand and expected as well as available fleet types, can be directly extended
and complemented by the basic finding and proposed estimation methods and algorithms of this dissertation.

Another area of further research is to consider network effects in the yield management algorithms. Although advanced yield management algorithms incorporate certain network effects into the decision process, network effects as they were defined throughout this dissertation are not considered. It seems to be relatively straightforward and beneficial to account for censored passenger flows in the YM optimization algorithm.

A strongly related area for further research is considering leg-dependence effects in demand forecasts. When estimating demand for a flight, not only the booking limits and capacity limits of the flight leg should be considered in the statistical estimation procedure, but also the relevant and significant booking and capacity limits of other flight legs.

As one can see, network effects generally point to new directions in the research of various airline scheduling and planning problems. Various applications and planning decisions are affected directly or indirectly by the on-board loads and OD mix of passenger flows. They include on-board and ground services, freight belly utilization models. Since the trend in many airline (and other) networks has been toward more effective capacity utilization and higher load factors, network effects become more significant. Consequently, the need to research ways to capture these effects and to incorporate them into the decision process is expected to increase opportunities for researchers to explore and solve new problems.
Appendix

The two networks described below, are based on actual data representing US domestic hub complexes. In the examples, for an ODF demand the mean and standard deviation of the incremental demand for each booking period was given. It was assumed that ODF demands are distributed by Normal distribution and are independent. Data was generated by disaggregating the booking class demand data kept by the existing yield management systems into estimated total ODF demands. Then each ODF was assigned to a representative booking curve, which apportioned the total ODF demands over the mutually exclusive booking periods prior to departure. Therefore, during the data generation what it believed to be a reasonable set of assumptions were used for generating the actual ODF demand data for the connecting complexes.

Network E.H: This hub network represents 30 different flight legs consisting of 15 flights in and 15 flights out. The flight legs symmetrically include 4 short haul, 6 medium, and 5 long haul flight legs in and out of the hub. Demands exist on 255 OD pairs, and for each OD pair 7 fare classes are defined. Demand is given on an incremental basis for 10 booking periods. We assigned fleet types to the network so that demand factors on each leg are high and balanced. The demand factors are high and vary between 0.84 to 1.18. The percentages of local demands on each leg vary between 37% and 50%, and average 42%. This hub network represents a connecting complex with a medium local demand percentage. The ODF fare values in this network are set so that in each OD market, the fare discount ratios across the fare classes are constant and equal to \( d=0.8 \).

Network E.C: This larger hub network represents a 50 leg, (25 in 25 out) US domestic hub complex. The legs consist of a variety of different flight legs. Demands are given for 566 OD pairs, and 6 fare classes are represented on each OD market with 10 booking periods. The demand factors that correspond to the actual demand data and scheduled flight capacities on the given particular day vary between 0.54 and 1.5. The percentages of local demand on each leg vary between 38% and 74%, with an overall average of 61%. This network example represents a hub with relatively strong local demand. The fare discount ratios across the markets vary and in a number of cases one could find connecting ODF itineraries with lower total fare than the corresponding local ODF's fare in the same fare class.
Bibliography:


