Analysis of Airplane Boarding Times

Eitan Bachmat * Daniel Berend † Luba Sapir ‡ Steven Skiena §
Natan Stolyarov ¶

Abstract

We model and analyze the process of passengers boarding an airplane. We show how the model yields closed-form estimates for the expected boarding time in many cases of interest. The computations reveal a clear link between the efficiency of various airline boarding policies and interior airplane design parameters, such as distance between rows. Comparison of our results with previous work, based on discrete event simulations, shows a high degree of agreement. Our work thus provides an explanation and theoretical foundation for these previous results, while allowing greater flexibility in terms of exploring many parameter settings.

1 Introduction

The process of airplane boarding is experienced daily by millions of passengers worldwide. Airlines have adopted a variety of boarding strategies in the hope of reducing the gate turnaround time for airplanes. Significant reductions in gate delays would improve on the quality of life for long-suffering air travelers, and yield significant economic benefits from more efficient use of aircraft and airport infrastructure. (See Van Landeghem and Beuselinck (2002), Marelli *et al.* (1998) and Van den Briel *et al.* (2005).)

The most pervasive strategy currently employed links boarding time to seat assignment. In particular, airlines tend to board passengers from the back of the airplane first. Such "back-to-front" policies are implemented by announcements of the form "Passengers from rows 30 and above are now welcome to board the plane".

It is not clear *a priori* how to analyze such strategies or to determine which policies are most effective at minimizing the expected boarding time. Airplane boarding has been previously studied through discrete event simulations by Van Landeghem and Beuselinck

*Department of Computer Science, Ben-Gurion University, Email: ebachmat@cs.bgu.ac.il, Research partially supported by an IBM faculty award

†Departments of Mathematics and Computer Science, Ben-Gurion University, Email: berend@cs.bgu.ac.il

‡Department of Industrial Engineering and Management, Ben-Gurion University, Email: lsapir@bgumail.bgu.ac.il

§Department of Computer Science, SUNY at Stony Brook, Email: skiena@cs.sunysb.edu

¶Department of Computer Science, Ben-Gurion University, Email: stolyarov@cs.bgu.ac.il

(2002), Marelli *et al.* (1998) and Ferrari and Nagel (2005). In addition, Van den Briel *et al.* (2003) formulated a non-linear integer programming problem, related to airplane boarding time, to which they applied various heuristics in order to find efficient boarding policies. The policies were then tested using a discrete event simulation.

Somewhat surprisingly, these studies have found that back-to-front policies are not necessarily effective, and might even be detrimental when compared with a boarding process with no boarding policy at all. Van Landeghem and Beuselinck (2002) argue that back-to-front policies are ineffective because they cause local congestion in the airplane, but no explanation is given for the mechanism by which congestion effects boarding time.

These studies also show that outside/in boarding policies, in which window seat passengers board first, followed by middle and then aisle seat passengers, can improve boarding time. In such policies, and others suggested by the studies, passengers are divided into multiple classes or groups which are boarded in sequence.

The study by Marelli *et al.* (1998) of Boeing Corp. emphasizes the effect of airplane interior design on boarding time, again using discrete event simulation methods.

The results and observations of Marelli *et al.* (1998), Van Landeghem and Beuselinck (2002), Van den Briel *et al.* (2003) and Ferrari and Nagel (2005) are of considerable value and interest. However, they do not address the need for a unified, analytic approach which can lead to a deeper understanding of the boarding process. As an example of the limitations of previous methods, we note that the simulations in all studies were carried out with particular airplanes in mind. It is therefore important to know how the success of airline boarding policies is related, if at all, to airplane design parameters, such as distance between rows, if the results are to be extended to other airplanes. The model of Van den Briel *et al.* (2003) does not take such design parameters into account.

In addition, any new scenarios need to be tested from scratch, and there is very little insight to be gained as to the underlying mechanisms of success for various strategies.

The purpose of this paper is to provide a natural framework for modeling analytically the airplane boarding process. Using this framework, we can offer some answers to the above problems.

In Section 2 we introduce a discrete random process which models airplane boarding. The input parameters of the process include distance between rows in the airplane, number of passengers per row, passenger aisle clearing time, airline boarding policy and a model for passenger reaction to the policy.

Section 3 provides a very brief introduction to Lorentzian geometry, which is used for modeling the asymptotics of the airplane boarding process.

In Section 4 we show that the asymptotic behavior of the boarding process is captured by a 2-dimensional space-time structure, also known as a Lorentzian metric, on a domain in the unit square. The Lorentzian metric depends on the input parameters of the boarding process. It is worthwhile to note that Lorentzian geometry was invented and first used to mathematically describe relativity theory, and it seems that airplane boarding is the first process outside physics to be modeled by it.

Quantities which are important from an operational point of view are captured by the geometry of the corresponding space-time. As an example, the total boarding time is given by the length (proper time) of the longest curve in the model.

Since the Lorentzian metric involves all the parameters of the airplane boarding problem, we can study the interactions between them and their mutual effect on boarding time, rather than considering them separately as in previous studies.

In Section 5 we consider the effect of interior airplane design parameters, such as leg room and number of passengers per row, on boarding time. The same issue was studied experimentally by Marelli *et al.* (1998). The parameters effect boarding time through a quantity which we denote by k. This parameter is proportional to the number of seats per row, and inversely proportional to the distance between successive rows (leg room). We obtain a formula for the effect of k on boarding time in the absence of an airline policy. This calculation also forms the basis for subsequent calculations of airplane boarding time in the presence of various airline boarding policies.

In Section 6 we introduce various airline boarding strategies, which we analyze. These include back-to-front boarding policies which are commonly practiced by airlines, as well as multiple class policies.

In Section 7 we provide detailed computations of the asymptotic boarding times of the airline strategies which were introduced in Section 5 in accordance with the boarding model of Section 2.

In Section 8 we apply the calculations of Section 6 to many specific boarding policies which have been considered in the previous studies. A comparison of our calculations with the results of discrete event simulations from Van Landeghem and Beuselinck (2002) and Van den Briel *et al.* (2003) shows a rather remarkable degree of overall agreement on the relative merit of different policies. We believe that this serves to reinforce the conclusions of all studies. The results show that it is difficult to substantially improve upon random boarding (with no policy) without exerting a large degree of control over the boarding process or reverting to policies which incorporate outside/in seat boarding. It seems unlikely that passengers will tolerate strict control boarding policies, even if these are easy to implement. Among policies which exert mild control over passengers, the best candidates for improving upon random boarding are policies combining outside/in boarding with a touch of back-to-front boarding.

Section 8 provides a brief summary and suggestions for future work.

2 The airplane boarding process

2.1 A combinatorial description of the boarding process

In this section we define a discrete process which models the airplane boarding process. We assume that passengers are assigned seats in the airplane in advance of the boarding process. Boarding is from the front of the plane, and the front row is row 1.

The input data which determines an instance of the boarding process is composed of the following items.

- A sequence of passengers $x_1, \ldots, x_n$, where x_i denotes the i'th passenger in the boarding queue. Each passenger x_i has a seat in an assigned row, denoted by $r_i = r(x_i)$.
- A width value $w_i = w(x_i)$, which measures the aisle space which a passenger occupies, baggage included.
- A delay value $d_i = d(x_i)$, which measures the amount of time it takes from the moment passenger x_i has reached his designated row until he clears the aisle. This time includes getting organized, placing carry-on luggage and, possibly, passing by previously seated passengers from the same row on the way to the designated seat. The last operation usually requires the seated passengers to get up and sit back after the newly arrived passenger has taken his/her seat.
- A length parameter l_j, representing the distance between row j and row $j+1$.

Given the above input parameters, the boarding process is described as follows:

Passengers proceed along the aisle as far as they can, either reaching their assigned row or lining up behind other passengers, who are either getting organized during the delay period or are themselves lined up behind other aisle blocking passengers. The location of row j along the aisle is $L_j = \sum_{m=1}^{j-1} l_m$. When passenger x_i reaches his/her row, $j = r_i$, he/she occupies aisle space from L_j to $L_j + w_i$. Passengers arriving at time t_i to their assigned row sit down and clear the aisle at time $t_i + d_i$. Once they clear the aisle, passengers who are behind them continue marching along the aisle as far as they can, and the process repeats.

We note the following feature of this process. If passenger x_h has reached his/her row r_h, and passengers $x_{i_1}, \ldots, x_{i_k}$ are lined up behind x_h, waiting for him/her to clear the aisle so they can further advance towards their seats, then these passengers occupy all the aisle stretch between L_{r_h} and $L_{r_h} - \sum_{j=1}^{k} w_{i_j}$. Thus a passenger x_m with $m > i_k$ cannot reach his/her row before x_h clears the aisle if $L_{r_m} + w_m > L_{r_i} - \sum_{j=1}^{k} w_{i_j}$. It follows that passengers can on occasion block passengers, who are assigned seats in rows behind them, by creating a line of aisle blocking passengers behind them. A concrete and detailed example of the boarding process is provided in Appendix A of the online material.

To better understand the boarding process, we define a natural partial order on passengers. A passenger A *blocks* another passenger B, and we denote $A \prec B$, if B cannot reach his/her row before A clears the aisle, but can do so right after A clears the aisle. The relation as defined is not transitive. We take its transitive closure so as to make it a partial order.

We can translate this partial order relation into an acyclic directed graph in the standard way. The nodes of the graph are passengers. If $x_i \prec x_j$, we place a directed edge from i to j and assign this edge a weight d_i. The boarding time t_i of a passenger is given by the maximal weight path terminating at i. One may compute t_i via the PERT procedure.

When all delay times are equal and normalized to be 1, the boarding process coincides with a well-known "peeling" process, which can be traced to the work of G. Cantor on ordinal arithmetic. The process peels the partially ordered set by successively eliminating (in rounds) the minimal elements in the partial order. This process provides simultaneously a minimal decomposition of the poset into independent sets and the longest chain in the poset. The passengers who are seated at time $t_i = m$ are peeled in the m'th round. In the boarding process, each passenger, who is not seated in the first round, can be assigned a pointer, which points to the last passenger who blocked his/her way to the assigned row. Following the trail of pointers starting from a passenger x_i, we identify a longest chain in the partial order ending at passenger i. In particular, the number of rounds needed is the size of the longest chain in the partial order.

Example 2.1 *When $w_i = 0$ for all i, the partial order condition becomes $x_i \prec x_j$ for $i < j$ if and only if $r_j \geq r_i$. Thus, chains correspond to weakly increasing subsequences in the sequence (r_i). The boarding process is then identified with a well studied process known as patience sorting, a card game process which optimally computes the longest increasing subsequence in a permutation. (See Aldous and Diaconis (1999) for further details on patience sorting.)*

2.2 The random process setting

In this paper we want to explore the asymptotic behavior of the boarding process given some distributions on the input parameters. To this end, we consider the following distributions and functions, which provide a random process setting to the boarding process.

- We represent passengers by pairs (q, r) in the unit square $[0, 1]^2$. As will be explained more fully later on, the passengers are sorted into rows by their r-coordinate. The q-coordinate represents the time at which the passenger decides to join the queue and hence determines his/her queue location.
- D – A delay distribution. The delay values d_i will be sampled from the distribution D.
- h – The number of passengers per row.
- l – Distance between rows (leg room).

For simplicity of presentation we will assume that D, h, l are all constant throughout the paper.

- W – A width distribution. The values w_i will be sampled from W.
- F – An airline boarding policy. We represent an airline boarding policy by a function F. $F(r)$ indicates the first time at which passengers from row r are allowed to join the boarding queue. If passengers follow the boarding policy, their (q, r)-coordinates satisfy $q \geq F(r)$.
- Ω – A passenger's reaction model. We need to make an assumption as to the nature of the reaction of passengers to the airline policy. The reaction determines the effect of the boarding policy on the boarding process. For instance, in the extreme case in which passengers do not pay any attention to the airline's announcements, the boarding policy is irrelevant. In the other extreme, if passengers join the queue immediately after being allowed, the airline can fully control the queuing order. In this paper we will use the following parameterized reaction model:

  - The attentive reaction model with parameter T. In this model, passengers join the queue at uniformly distributed times, within T time units of being allowed to board.

2.3 Sampling the random process

Given the parameters above, we sample an instance of the boarding problem as follows. We first combine the passenger reaction model Ω with the airline boarding policy F to produce a joint distribution $p(q, r)dqdr$ on passengers' row number and queueing time.

Specifically, given the airline boarding policy $F(r)$, and assuming the attentive reaction model with parameter T, the queue joining time q of a passenger with row indicator coordinate r satisfies $F(r) \leq q \leq F(r) + T$ and is uniformly distributed within this range. The corresponding row/queue joint distribution is $p(q, r) = 1/T$ in the range and 0 outside.

The distribution $p(q, r)dqdr$ is sampled n times independently to produce passenger coordinates $x_i = (q_i, r_i)$, $i = 1, \ldots, n$, where we assume that passengers are indexed in increasing order of the q-coordinate, i.e., $q_1 < q_2 < \ldots < q_n$. To determine the rows $r(x_i)$, the passengers are sorted by the value of r_i in increasing order. The first h passengers are assigned to seats in row 1, the next h to seats in row 2 and so on. We set $l_i = l$ and $d_i = D$. The width w_i of each passenger is sampled independently from W.

We are interested in studying the vector (t_i), where t_i is the time passenger x_i takes his seat. In particular, we are interested in the asymptotics of total boarding time $\max_i\ t_i$.

3 A brief introduction to Lorentzian geometry

We briefly consider some basic material on Lorentzian geometry, relevant to our application. Since Lorentzian geometry is intimately tied to relativity theory, and in fact was developed to model it, we also give a brief account of the physical interpretation of various geometrical notions in the theory. For a much more comprehensive treatment, we refer the reader to Penrose (1972).

A Lorentzian metric on a set $D \subset \mathbf{R}^n$ is given by a differentiable function from D to the set of real invertible symmetric $n \times n$ matrices with exactly one positive eigenvalue. Let $g = g(x_1, \ldots, x_n)$ be the matrix attached to $(x_1, \ldots, x_n) \in D$. To the metric g we associate the quadratic form $ds^2 = \sum_i \sum_j g_{i,j} dx_i dx_j$, which in analogy with the Riemannian case may be considered as the square of a "distance" function.

As an example, consider $\mathbf{R}^n$, where the first coordinate is denoted by t and the others by $x_1, \ldots, x_{n-1}$. We attach to all points in $\mathbf{R}^n$ the fixed diagonal matrix $g = \mathrm{diag}(1, -1, -1, \ldots, -1)$. We have $ds^2 = dt^2 - dx_1^2 - \ldots - dx_{n-1}^2$. The space $\mathbf{R}^n$, equipped with this constant Lorentzian metric, is known as the *Minkowski space*. It was introduced by Minkowski to provide a mathematical (geometrical) model for special relativity theory, and in particular represents the notion of space-time as opposed to space and time separately.

A vector v is *time-like* if $vgv^t \geq 0$, *light-like* if $vgv^t = 0$ and *space-like* if $vgv^t < 0$.

Let C be a differentiable curve given by $x(u) = (x_1(u), \ldots, x_n(u))$, $u \in [a, b]$. The curve C is a *time-like curve* if all its tangent vectors are time-like. The definition of light-like and space-like curves is analogous. The *length* $l(C)$ of a time-like curve C is defined by:

$$l(C) = \int_a^b \sqrt{x'(u) g(x(u)) x'(u)^t} du.$$

The set of time-like vectors has two cone-shaped components. When there are no closed time-like curves, we can define notions of past and future, which determine causality via a consistent choice at each point of one of the components, that is, a continuous vector field of non-vanishing time-like vectors. The vectors in the chosen component are *future pointing*. We can then define a past-future partial order on the points of the space-time by letting $B >_g A$ if there is a curve starting at A and ending at B with future pointing (time-like) tangents. Intuitively, this means that B is in the future of A.

In relativity theory, the trajectory of a particle (small body) from space-time point A to space-time point B is described by a curve with future pointing time-like tangents. Such a curve is known as a *world line*. If the particle is massless, it will travel along a light-like curve. Space-like curves correspond to travel beyond the speed of light and are thus excluded.

The length of the curve is the time which passes during the movement, as measured by a clock attached to the particle. This quantity is known in physics as *proper time*.

Special relativity, in which gravity is not considered, is modeled by the Minkowski space.

When no force is exerted on the particle (free particle), it moves from a point A to a point B along a trajectory which maximizes locally the length of the curve (proper time). In the Minkowski space, such curves are straight lines. This is a restatement of Galileo's principle that a free particle travels along a straight line at constant speed.

General relativity is a theory which incorporates gravity. In the presence of gravity, space-time is no longer described by Minkowski space. It is rather described by a Lorentzian metric which changes from point to point. In the presence of gravity alone (free fall), the trajectory of a particle again maximizes proper time (curve length) between endpoints of each small segment of the curve as in the case of a free particle. Curves describing the motion of free falling particles are known as *geodesics*. As an example, when we throw a stone, the parabola traced through time by the stone forms a geodesic in a curved Lorentzian geometry describing space-time.

4 Modeling the asymptotic behavior of airplane boarding with space-time

We associate a Lorentzian metric on the unit square to the airplane boarding process with parameters D, h, l, W, F, Ω.

First we produce from F and Ω a distribution $p(q,r)dqdr$ as explained in Section 2.3. Let M denote the support of $p(q,r)$, namely the closure of the set $\{(q,r) : p(q,r) > 0\}$. Put $\alpha(q,r) = \int_r^1 p(q,z)dz$ and $k = hE(W)/l$, where $E(W)$ denotes the expectation of W. We define a Lorentzian metric on M by:

$$g_{p,p} = 4D^2 p(q,r)k\alpha, \qquad g_{p,q} = g_{q,p} = 2D^2 p(q,r), \qquad g_{q,q} = 0.$$

Equivalently, the length (proper time) element ds is given by:

$$ds^2 = 4D^2 p(q,r)(dqdr + \alpha k dq^2). \tag{1}$$

The following facts serve as motivation for the introduction of the metric.

A) The volume form associated with the metric g, given by $|det(g)|^{1/2} dqdr = 2D^2 p(q,r)dqdr$, is proportional to the passenger density distribution $p(q,r)dqdr$.

B) Asymptotically, the causal relation among space-time points with respect to the metric and the blocking relation in the airplane boarding process coincide locally.

Given a point $x \in M$, let the *(proper time) level* $\tau(x)$ of x be the supremum of the lengths of time-like curves in M ending at x. The *diameter* $d(M)$ of M is the supremum of all $\tau(x)$, $x \in M$. We use the phrase *with high probability*, or *w.h.p.*, to refer to an event that occurs with probability approaching 1 as the number of passengers n tends to infinity.

The motivation for associating the Lorentzian metric with the airplane boarding problem is given by the following statement:

Consider the boarding process with parameters D, h, l, W, F, Ω, and let (M, g) be the space-time geometry associated with the process as determined above. Let $\varepsilon > 0$. Then, w.h.p., for any passenger $x_i = (q_i, r_i)$ we have

$$(\tau(x_i) - \varepsilon)\sqrt{n} \leq t_i \leq (\tau(x_i) + \varepsilon)\sqrt{n}.$$

In addition, any maximal size chain of blocking passengers, ending with passenger x_i, is contained in an ε-neighborhood of a curve of length $\tau(x_i)$ ending at x_i. In particular, the total boarding time is w.h.p. asymptotically equal to $d(M)\sqrt{n}$.

We provide a heuristic argument which explains the construction of the associated metric and provides evidence for the statement. For simplicity, we assume that $D = 1$. Passengers in the boarding process are represented by points $(q, r) \in M$. Thus, in addition to the blocking partial order, they are also related by the causal (past-future) partial order induced from M. We would like to show that w.h.p. these two partial orders asymptotically coincide. Let $X = (q, r)$ and $X' = (q + dq, r + dr)$, $dq > 0$, represent passengers with nearby coordinates. X blocks X' if $dr > 0$. However, as noted earlier, X may block X' even when X' sits behind him/her, namely $dr < 0$. Consider the time when passenger X arrives at his/her designated row. All passengers with row number indicators beyond $r + dr$, which are behind passenger X in the queue but in front of passenger X', will occupy aisle space between them. The number of such passengers is roughly $n\alpha dq$. Each such passenger occupies w units of aisle space, where w is sampled from W, and hence on the average they occupy roughly $E(W)$ units of aisle space each, and altogether an aisle stretch of roughly $(E(W)/l)\alpha dqn$ rows. The row difference between X and X' is roughly $-(1/h)drn$. Thus, passenger X is blocking passenger X', via the passengers between them, roughly when $dr \leq -\alpha k dq$. Upon inspection, we see that this is precisely the condition for X' to be in the future of X. We see that the metric corresponding to a boarding scenario is designed in such a way that the relation of blocking between passengers and the relation of causality between space-time points asymptotically coincide. Consequently, we will replace in the argument the blocking relation with the causal relation in M. Our assumption that $D = 1$ means that boarding time is given by the length of the longest chain with respect to the blocking relation. Replacing the blocking relation by the asymptotically equivalent causal relation implies that the boarding time should roughly equal w.h.p. the length of the longest chain among the n space-time points with respect to the causal partial order We further note that locally the length of the longest causal chain between two points $X = (q, r)$ and $X' = (q + dq, r + dr)$ is roughly $ds\sqrt{n}$. To establish this, we note that locally the metric is nearly constant and points are nearly uniformly distributed. After the volume preserving coordinate change $q' = q$ and $r' = r + k\alpha q$, we may assume that the metric has the form $ds'^2 = 4p(q,r)dq'dr'$. For this metric, the causal structure coincides with the notion of increasing subsequences. By a theorem of Vershik and Kerov (1977), the longest chain among m uniformly distributed points in a rectangle with sides parallel to the axis is roughly $2\sqrt{m}$. In our case $m = p(q,r)dq(dr + k\alpha dq)\sqrt{n}$, as required. The local result can be integrated along a curve C to show that the longest chain clustered in a small neighborhood of the curve C is roughly $l(C)\sqrt{n}$. The statement is obtained by maximizing over all curves.

Remark: When $k = 0$, the statement is a theorem and in fact a restatement in terms of Lorentzian geometry of a result on increasing subsequences of Deuschel and Zeitouni (1995). Our argument in favor of the statement is based on their method.

5 The effect of airplane interior design on boarding time in the absence of a boarding policy

As a first application of our modeling approach, we estimate the effect of distance between rows (leg room) on boarding time. This problem was first considered via simulations by Marelli *et al.* (1998). It will also serve as an example of a typical computation in the analysis of airplane boarding using Lorentzian geometry.

We choose the normalization $D = 1/2$, which eliminates the constant scaling prefactor $4D^2$ from the metric. Furthermore, whenever analyzing the boarding problem with associated model M, we will consider the normalized boarding time $d(M)$ rather than the total expected boarding time for n passengers, given by $d(M)\sqrt{n}$.

We assume in this section that there is no boarding policy, i.e., passengers join the line at times which are independent of their seat assignments. An equivalent statement is that we assume the policy to be given by the function $F(r) = 0$. According to this assumption, $p(q,r)dqdr$ is a uniform distribution, regardless of the attentiveness parameter T.

Having fixed all other boarding parameters, we note that leg room l is inversely proportional to $k = hE(W)/l$, a parameter through which it affects the model M. We shall therefore analyze the dependence of boarding time on k. We will denote the associated model by M_k. Following the modeling statement, we need to compute $d(M_k)$. Since p is uniform we have $\alpha = 1 - r$, the squared length element thus being $ds^2 = dqdr + k(1-r)dq^2$. Note that this definition of the length element ds can be extended to the entire (q,r)-plane. Passengers can only block passengers joining the queue later. Correspondingly, time-like curves can always be parameterized by the q coordinate, leading to the form $r(q)$, $0 \leq q \leq 1$. The length of $r(q)$, which is obtained by integrating the length element ds along the curve, may be written more explicitly for M_k as

$$L(r) = 2\int_0^1 \sqrt{r' + k(1-r)}dq. \tag{2}$$

Any point in M_k can be reached via a time-like curve from $(0,0)$, and the point $(1,1)$ can be reached via a time-like curve from any other point. We conclude that the longest time-like curve starts at $(0,0)$ and ends at $(1,1)$. This observation provides the boundary conditions for the variational problem of maximizing $L(r)$. Since the functional $L(r)$ does not depend explicitly on q, the Euler-Lagrange equation degenerates to the Beltrami equation

$$r'\frac{dL}{dr'} - L = const. \tag{3}$$

or, explicitly for M_k,

$$\frac{r'}{2\sqrt{r' + k(1-r)}} - \sqrt{r' + k(1-r)} = const. \tag{4}$$

The general solution of the equation is

$$r = c_1 e^{2kq} + c_2 e^{kq} + 1. \tag{5}$$

Given such a solution in the range $\alpha \leq q \leq \beta$, the value of the length of r is

$$L(r) = (e^{k\beta} - e^{k\alpha})\sqrt{\frac{c_1}{k}}. \tag{6}$$

After placing the boundary conditions $r(0) = 0, r(1) = 1$, we obtain the solution

$$r = \frac{e^{2kq}}{e^k - 1} - \frac{e^k}{e^k - 1} e^{kq} + 1. \tag{7}$$

The solution remains within the unit square for $0 \leq q \leq 1$ when $k \leq \ln 2$. Applying the functional to the solution, we obtain

$$d(M_k) = L(r) = \sqrt{\frac{e^k - 1}{k}} \tag{8}$$

for $k \leq \ln 2$.

When $k > \ln 2$, the solution is not contained in the unit square anymore. By comparing curve lengths in M_k and the unit square equipped with the constant metric $ds^2 = (dqdr + kdq^2)$, we observe that the longest curve in M_k between the point $(0,0)$ and $(q_0, 0)$, where $0 < q_0 \leq 1$, is given by the q-axis segment between them. In addition, basic differential geometric considerations tell us that the maximal length curve has to be differentiable. We conclude that the longest curve will consist of a segment of the form $[0, q_0]$ on the q-axis, followed by a geodesic between the points $(q_0, 0)$ and $(1,1)$, whose derivative, when r is considered as a function of q, vanishes at q_0. We solve the Beltrami equation with boundary conditions $r(q_0) = 0$ and $r(1) = 1$, requiring in addition that $r'(q_0) = 0$, which leads to $q_0 = \frac{k - \ln 2}{k}$. Plugging the resulting curve into the functional leads to

$$d(M_k) = \sqrt{k} + \frac{1 - \ln 2}{\sqrt{k}}, \qquad k > \ln 2. \tag{9}$$

The maximal length curves are plotted for various values of k in Figure 1.

Depending on boarding policies and reaction models, the space-time models associated with airplane boarding may be non-flat (curved). However, the policies which are considered in the present paper, which are the most naturally occurring ones, do not lead to such complications.

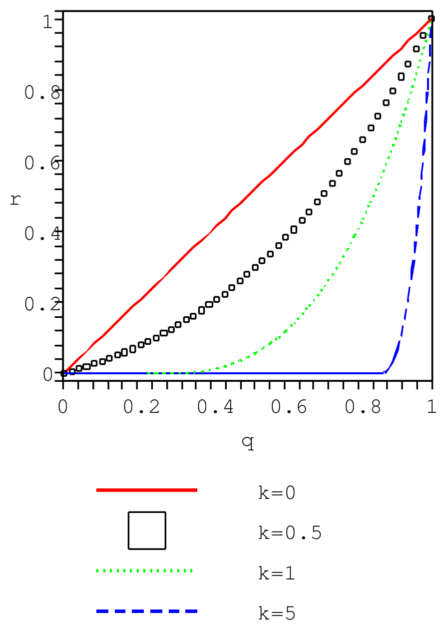

Figure 1: Maximal length curves for several values of k.

6 Airplane boarding policies

In this section we introduce the airline boarding policies which will be analyzed in this paper.

6.1 Announcement policies

Definition: A policy is an *announcement policy* if the corresponding function F is a step function.

An alternative description is the following. Let $0 = q_1 < q_1 < \ldots < q_m < 1$ and $0 = r_1 < r_1 < \ldots < r_m < r_{m+1} = 1$ be decompositions of the q and r unit intervals. Let σ be a permutation on $1, \ldots, m$. Let F be given by $F(r) = q_i$ for $r_{\sigma(i)} \leq r \leq r_{\sigma(i)+1}$. The number m is the *number of groups* in the policy. The i'th group of the policy consists of passengers with row indicators r, satisfying $r_{\sigma(i)} \leq r \leq r_{\sigma(i)+1}$. If σ is the permutation $m, m-1, ..., 1$, the announcement policy is a *back-to-front policy*. Consider a boarding announcement policy F. Once the attentiveness parameter T satisfies

$$T \leq \min_i (q_{i+1} - q_i), \tag{10}$$

the boarding queue order consists of a random permutation of the passengers in the first group, followed by a random permutation of passengers from the second group, and so on. We conclude that, if $T \leq \min_i (q_{i+1} - q_i)$, then the boarding time is independent of T and the q_i's. When considering announcement policies, we shall always assume that this condition holds and hence that the announcement policy is specified only by the r_i's and σ.

Assuming (10), back-to-front announcement policies are implemented by announcements calling, at each time q_i, for passengers from row r_{m-i+1} and above to board the airplane. This is the basic type of boarding policy employed by airlines.

An announcement policy is *uniform* if all blocks of passengers are of equal size, i.e., if $r_i = \frac{i-1}{m}$ for some m. We denote by F_m the uniform back-to-front announcement policy with m groups.

More generally, we denote by $F_{m,\sigma}$ the uniform announcement policy with m groups and the permutation σ. As an example, consider $F_{3,\sigma}$ with $\sigma = (3, 1, 2)$. This policy boards passengers from the back third of the airplane, followed by passengers from the front third, and finally passengers from the middle third. The ordering in each group is random.

6.2 Multiclass policies

Another airline boarding strategy which will be considered is the division of passengers into different classes, so that not all passengers from a contiguous set of rows are in the same class. Passengers from different classes are then allowed to board at different times. An announcement policy can be employed in each class separately. Such strategies have been examined in Van Landeghem and Beuselinck (2002), Van den Briel *et al.* (2003) and Ferrari and Nagel (2005). We first provide some examples of classes.

- Row classes: Let c be an integer. Consider the division of passengers into classes according to the value of r mod c. As an example, consider the case $c = 3$. We board first passengers from rows $3, 6, 9, \ldots$, followed by passengers from rows $1, 4, 7, \ldots$, and finally passengers from rows $2, 5, 8, \ldots$. Such a division effectively triples the distance between rows, while retaining the number of passengers per row and the width of passengers.

- Random classes: Passengers are divided randomly into c classes. For simplicity of analysis, we shall assume that each class contains n/c passengers. This division achieves asymptotically the same goal as the row classification, but in a less structured (rigid) manner.

- Half-row classes: These classes were introduced in Van Landeghem and Beuselinck (2002). There are two classes, consisting of passengers on the right side of the aisle and those to the left of the aisle, respectively. Note that in this 2-class system, unlike the seat type classification system, presented below, the delay distribution remains uneffected.

- Seat type classes: Passengers may be classified according to seat type, instead of row type. We can group all window seat passengers in one class, middle seat passengers in a second class and aisle seat passengers in a third class. Seat type classification has a feature which makes it more advantageous than the other types of classification which were listed. The delay distribution D is greatly reduced by allowing window passengers to board first, followed by middle seats, since passengers do not need to unseat other passengers who are already seated.

In terms of notation, consider a policy with c classes and m announcements per class. Label the group consisting of passengers of the j'th class and from rows $(i-1)/m \leq r \leq i/m$ with index $i+jm$. Let σ be a permutation of the elements $1, \ldots, cm$, which preserves the order of the classes. We denote by $F_{c,m,\sigma}$ the policy which calls passenger groups in the order given by σ.

We note that boarding time does not change if we merely employ classes (other than seat type); it is only in conjunction with an announcement policy within the classes that boarding time can change.

7 Computations

In this section we compute asymptotic boarding times for various policies, introduced in the preceding section.

7.1 Computing boarding time for the F_2 policy

We present a general formula for $d(2,k)$, the normalized boarding time for the F_2 policy with parameter k, when $k \geq 1$. Consider the space-time $M_{2,k}$ which corresponds to the policy F_2 and parameter k. It is easier to study the F_2 policy if we inflate the unit square to the square $[0,2] \times [0,2]$ via the map $(q,r) \longrightarrow (2q,2r)$. We will consider the inflated square with the induced metric. Some care must be taken since the induced metric is not given by the basic modeling formula (1) which assumed normalized coordinates. We will still call the model $M_{2,k}$. Let U be a maximal length geodesic in $M_{2,k}$. U is composed of U_1, which is a maximal curve in the range $1 \leq r \leq 2$ and $0 \leq q \leq 1$, which we denote by R_1, and U_2 in the range $0 \leq r \leq 1$ and $1 \leq q \leq 2$, which we denote by R_2. In the range R_1 the point $u = (0,1)$ is the unique minimal point for the causal relation. We conclude that u is the initial point of U_1. In the range R_2, the points of the form $(1,\delta)$ with $0 \leq \delta \leq 1$ are minimal, and therefore one of these points is the initial point of U_2. In order for U to be a causal curve, we also require that the endpoint of U_1 is in the past (blocks) of the initial point of U_2. If $k > 1$, then there are points of R_1 which are in the past of the initial point of R_2, and therefore both U_1 and U_2 must be non-empty. Assume that (q_1,r_1) is the endpoint of U_1 and (q_2,r_2) is a point in R_1, connected to (q_1,r_1) by a future pointing light-like curve ψ. By the maximality of U, U_1 is the maximal length time-like curve from $u=(0,1)$ to (q_1,r_1). The concatenation U_1 followed by ψ provides a time-like curve between u and (q_2,r_2) whose length is equal to the length of U_1 and such that the endpoint (q_2,r_2) is in the past for the same set of points as (q_1,r_1). In addition, the concatenation is not differentiable at (q_1,r_1) and hence is not a maximal curve between u and (q_2,r_2) in contradiction to the maximality of U. We conclude that the endpoint of U_1 must lie on the boundary of R_1, either on the bottom edge, $r = 1$, or on the right edge, $q = 1$. Points in R_1 with $q = 1$ are not in the past of any point in R_2 and therefore cannot serve as endpoints for U_1 if $k > 1$. Consequently, the endpoint of U_1 must lie on the bottom edge of R_1. Let $(q_1, 1)$ be a point on the bottom edge. As explained in Section 5 the maximal curve between u and $(q_1, 1)$ is the curve $(q, 1)$, $0 \le q \le q_1$. The length of the curve is $q_1\sqrt{k/2}$. We optimize the choice of q_1. Consider a point $(q_1, 1)$ on the bottom edge of R_1. The point $(q_1, 1) \in R_1$ blocks points of the form $(1, \delta(q_1)) \in R_2$ when δ satisfies

$$1 - \delta(q_1) \le (1 - q_1)k, \tag{11}$$

and equality must hold for a maximal curve. The maximal curve in R_2 which begins at the point $(1, \delta)$ must end at $(2, 1)$. Depending on δ, it may either be a geodesic curve lying in the interior of R_2 or contain a segment on the lower edge of R_2. In either case, it must be differentiable, and hence the critical value of δ which separates the two cases is the value for which the geodesic between $(1, \delta)$ and $(2, 1)$ is tangent to the bottom edge of R_1. The tangent geodesic is given by the equation

$$(1 - \sqrt{\delta})^2 (e^{k(q-1)} - \frac{1}{1 - \sqrt{\delta}})^2, \tag{12}$$

and passes through the point $(2, 1)$ when $\delta = (1 - 2e^{-k})^2$. When $\delta \ge (1 - 2e^{-k})^2$, the maximal curve is the geodesic $\frac{1-\delta}{e^k-1} e^{2k(q-1)} - \frac{e^k(1-\delta)}{e^k-1} e^{k(q-1)} + 1$. According to (5), the length of the maximal curve in R_2, beginning at $(1, \delta)$, is

$$\sqrt{\frac{1}{2k}} \sqrt{(e^k - 1)(1 - \delta)}. \tag{13}$$

When $\delta < (1 - 2e^{-k})^2$, the maximal curve consists of three segments: the geodesic of (12) up to the tangent point $(1 + (\ln(\frac{1}{1-\sqrt{\delta}}))/k, 0)$, the line segment joining $(1 + (\ln(\frac{1}{1-\sqrt{\delta}}))/k, 0)$ and $(2 - \ln 2/k, 0)$ and the geodesic from $(2 - \ln 2/k, 0)$ to $(2, 1)$, which is tangent to the bottom edge at $(2 - \ln 2/k, 0)$ according to the computations in Section 5. The length of the three segments combined is

$$\sqrt{\frac{1}{2k}} (\sqrt{\delta} + \ln(1 - \sqrt{\delta}) + k + 1 - \ln 2). \tag{14}$$

Given that in (11) we have equality for maximal curves, the contribution of U_1 is

$$\sqrt{\frac{1}{2k}} (k - 1 + \delta). \tag{15}$$

Let $U(\delta)$ be the curve from $(0, 1)$ to $(1 - \frac{1-\delta}{k}, 1)$, followed by the maximal curve from $(1, \delta)$ to $(2, 1)$. By the analysis presented above we know that the maximal curve has the form $U(\delta)$ for some $0 \le \delta \le 1$. Let $A(\delta)$ denote the sum of the expressions (14) and (15). Differentiating $A(\delta)$ we find that $A(\delta)$ increases towards $\delta = 1/4$ and that it is maximized when $\delta = 1/4$. The length of $U(\delta)$ coincides with $A(\delta)$ in the range $0 \le \delta \le (1 - 2e^{-k})^2$. For $k \ge 2 \ln 2$,

$\delta = 1/4$ is in that range and $A(1/4)$ is the maximal value, while for $k \leq 2\ln 2$, the maximal value in the range is provided by $A((1-2e^{-k})^2)$. Let $B(\delta)$ denote the sum of the expressions (13) and (15). $B(\delta)$ coincides with $U(\delta)$ whenever $\delta \geq (1-2e^{-k})^2$. Differentiating $B(\delta)$ we find that the maximal value is obtained when $\delta = 1 - \frac{e^k-1}{4}$. This value is in the range where $U(\delta)$ coincides with $B(\delta)$ whenever $1 - \frac{e^k-1}{4} \geq (1-2e^{-k})^2$. Writing $x = e^k$ and multiplying both sides by x^3, which is always positive, we obtain the condition $x^3 - 16x^2 + 16x - 1 \leq 0$. Since $x^3 - 16x^2 + 16x - 1 = (x-1)(x-4)^2$, the condition holds for $1 \leq x \leq 4$, which translates to the condition $k \leq 2\ln 2$.

When $k \geq 2\ln 2$, the maximum of $B(\delta)$ in the range $\delta \geq (1-2e^{-k})^2$ is attained at $\delta = (1-2e^{-k})^2$. Since $A((1-2e^{-k})^2) = B((1-2e^{-k})^2)$, we conclude that for $k \geq 2\ln 2$ the maximal value of $U(\delta)$ is attained at $\delta = 1/4$. Similarly, for $k \leq 2\ln 2$ the maximal value of $U(\delta)$ is attained at $\delta = 1 - \frac{e^k-1}{4}$. Plugging these values into $A(\delta)$ and $B(\delta)$, respectively, we see that the length of the maximal curve for $k \geq 2\ln 2$ is

$$d(2,k) = \sqrt{2k} + \frac{3/4 - 2\ln 2}{\sqrt{2k}} \tag{16}$$

and for $1 \leq k \leq 2\ln 2$ is

$$d(2,k) = \sqrt{\frac{1}{2k}}(k + \frac{e^k-1}{4}). \tag{17}$$

7.2 Computations for F_m

We consider the case of $m > 2$. As in the case of F_2, it is more convenient to consider the expanded square $[0,m] \times [0,m]$, with the induced metric. The analogues of the squares R_1 and R_2 are the squares $R_i = [i-1,i] \times [m-i, m-i+1]$, $i = 1,...,m$, which lie along the anti-diagonal. The maximal length curve U is composed as before of segments U_i contained in R_i, with the property that the endpoint of U_i is in the past of the initial point of U_{i+1} for $i = 1,...,m-1$. The arguments presented for the case of F_2 also show that $(0, m-1)$ is the initial point of U (and U_1) and that the initial point of U_i is on the left edge of R_i for all i. In addition, for $i = 1,...,m-1$, the endpoint of U_i lies on the bottom edge of R_i. Thus, the initial point of U_i has the form $(i, m-i+\delta_i)$ and the endpoints have the form $(i+\beta_i, m-i)$, for some δ_i and β_i. The maximality of U implies the relation

$$\beta_i = 1 - \frac{1-\delta_{i+1}}{k}. \tag{18}$$

Consider U_1 and U_2. The initial point and the endpoint of U_1 lie on the bottom edge of R_1, and hence by maximality all of U_1 lies on the bottom edge. Let $(1, m-2+\delta)$ be the initial point of U_2 and $(1+\beta, m-2)$ the endpoint. Given the endpoint of U_2, we can optimize the location of the initial point of U_2. Depending on δ, U_2 may either be a geodesic in the interior of R_2 or be composed of a geodesic segment which is tangent to the bottom edge followed by a segment along the bottom edge. In the first case the length of the geodesic segment is

$$\frac{\sqrt{e^{kb}-1}}{\sqrt{k}} \cdot \frac{\sqrt{(1-\delta)e^{kb}-1}}{\sqrt{e^{kb}-1}}. \tag{19}$$

We may add the contribution of U_1 and differentiate. The sum of contributions is maximized when

$$\delta = -(\frac{e^{kb}}{4} + \frac{4}{e^{kb}}) + \frac{5}{4}\,. \tag{20}$$

Writing $x = \frac{e^{kb}}{4}$, we see that $\delta = -(x+1/x)+5/4$, and since $x+1/x \geq 2$ for all $x > 0$ we see that the optimal δ is negative. Hence the U_2 component of the maximal curve is composed of a geodesic tangent to the bottom of R_2, followed (possibly) by a segment along the bottom. This case was already analyzed in the $d(2,k)$ case. The optimal value for δ is $1/4$ if it is in the range

$$\ln(\frac{1}{1-\sqrt{\delta}})/k \leq b. \tag{21}$$

Let $k \geq 3/4 + \ln 2$. Consider the curve U with $\delta_i = 1/4$ for $i = 1,..,m-1$. By (18), this corresponds to $\beta_i = 1 - \frac{3}{4k}$. We note that (21) is satisfied for $b = \beta_i$ for $k \geq 3/4 + \ln 2$. We claim that U is maximal. Assume to the contrary that U' with parameters β'_i is maximal. Let j be the first index for which $\beta'_j \neq 1 - \frac{3}{4k}$. We claim that $\beta'_j < 1 - \frac{3}{4k}$, as otherwise, by the calculations for U_1 and U_2, the path U'_{i+1} from $(j+1, m-(j+1)+\delta'_{j+1})$ to $(j+1+\beta'_{j+1}, m-(j+1))$ is not optimal, since $\delta'_{j+1} \geq 1/4$. For the same reason we must have $\beta'_i < 1 - \frac{3}{4k}$ for all $i > j$. However, by the computation for $d(2,k)$ we know that for $k \geq 2\ln 2$ the contribution of the union of U_{m-1} and U_m is not optimal. Since $3/4 > \ln 2$, we are done.

The length of the curve U is

$$\sqrt{mk} - \frac{m-2}{\sqrt{mk}}(\ln 2 + 1/4) - \frac{2\ln 2 - 3/4}{\sqrt{mk}}. \tag{22}$$

7.3 Computing $F_{m,\sigma}$

We compute $F_{m,\sigma}$ under some assumptions on σ and k. Using the representation of passengers in the expanded square $[0,m]\times[0,m]$, elements in the i'th group will be represented by points in the square $R_{i,\sigma}$, consisting of points of the form $i \leq q \leq i+1$, and $\sigma(i) \leq r \leq \sigma(i)+1$. Let U be a maximal curve, composed of curves U_i in $R_{i,\sigma}$. The sequence $\sigma(0), \sigma(1), ..., \sigma(m-1)$ decomposes into blocks B_j, $1 \leq j \leq r(\sigma)$, of size b_j of decreasing sequences, that is, $\sigma(0) > ... > \sigma(b_1) < \sigma(b_1+1) > \ldots > \sigma(b_1+b_2) < \sigma(b_1+b_2+1) \ldots$. Let $c_j = \sum_{i=1}^{j} b_j$ be the cumulative sequence of the sequence b_j. Define the *excess descent* of block B_j by $e_j = \sigma(c_j) - \sigma(c_{j-1}) - b_j + 1$ if $b_j > 1$ and $e_j = 0$ if $b_j = 1$. We let $e = \sum_j e_j$. Let $\lambda(\sigma)$ be the maximum over all decreasing pairs of consecutive elements, $\sigma(l) < \sigma(l+1)$, of $\sigma(l+1) - \sigma(l)$. If

$$k \geq 3/4 + \ln 2 + \lambda(\sigma) - 1, \tag{23}$$

then

$$d(m,\sigma,k) = \sum_{j=1}^{r(\sigma)} d(b_j,k)\frac{\sqrt{b_j}}{\sqrt{m}} - e\frac{1}{\sqrt{km}}. \tag{24}$$

To see this, consider a maximal causal curve U in the union of $R_{i,\sigma}$, composed of curves U_i in the respective squares. As in the previous computations, it is easy to verify that the endpoint of U_i is on the bottom edge of $R_{i,\sigma}$, say at $(i+\beta_{i,\sigma}, \sigma(i))$, and the initial point of V_i is on the left border of $R_{i,\sigma}$, say at $(i, \sigma(i)+\delta_{i,\sigma})$. The assumption that U is causal means that the endpoint of U_i blocks the initial point of U_{i+1}. If $\sigma(i) < \sigma(i+1)$, then any point of $R_{i,\sigma}$ blocks any point of $R_{i+1,\sigma}$, and therefore we can make independent computations in the blocks B_j as long as the maximal curves in B_j contain non-empty curves U_i for all i. The condition for blocking when $\sigma(i) > \sigma(i+1)$ is

$$\sigma(i) - \sigma(i+1) - \delta_{i+1} \leq k(1-\beta_i). \tag{25}$$

The process of identifying a maximal curve in B_j is identical to the process in the case of $d(b_j, k)$, subject to the replacement of (18) by (25) and a change in density from $\frac{1}{b_j}$ to $\frac{1}{m}$. As in the case of $d(b_j, k)$, the curve with $\delta_i = 1/4 \; c_{j-1}+2 \leq i \leq c_j$ is optimal when it can be constructed as a causal curve, which by (25) holds when $k \geq \sigma(i) - \sigma(i+1) - 1 + 3/4 + \ln 2$ for all i. The change in density yields a comparison with $d(b_j, k)\frac{\sqrt{b_j}}{\sqrt{m}}$. A comparison of (18) and (25) shows that the portion of the curve U_i on the bottom edge of $R_{i,\sigma}$ is shorter (in the standard Euclidean sense) by $\frac{\sigma(i+1)-\sigma(i)-1}{k}$ than the corresponding curve in the computation of $d(b_j, k)$. Summing over $i \in B_j$ and taking the density $\frac{1}{m}$ into account, we obtain the difference term $-e\frac{1}{\sqrt{km}}$, thus establishing (24).

7.4 Computing $F_{m,c,\sigma}$

The analysis of row class strategies is carried out using our previous techniques. The longest geodesic is composed of longest geodesics in each class, subject to the condition that the endpoint of the geodesic in class i blocks the initial point of the geodesic in class $i+1$. It is simpler to consider multiclass policies in coordinates $0 \leq r \leq m$ and $0 \leq q \leq cm$, where the passengers in the i'th class are modeled in the rectangle S_i, consisting of the points with $(i-1)m \leq q \leq im$.

Let $d(c, m, k)$ denote the normalized boarding time model of a multiclass policy with c (row) classes and the F_m announcement policy within each class. It is easy to verify that, if $m \geq 2$, then the condition, whereby the initial point of the longest geodesic in S_{i+1} is blocked by the endpoint of the longest geodesic in S_i, always holds. In each rectangle S_i there are n/c points and the effective value of k is k/c. Since we have c rectangles, we obtain the following inductive relation of boarding time for $m \geq 2$:

$$d(c, m, k)\sqrt{n} = cd(m, k/c)\sqrt{n/c} = \sqrt{c}d(m, k/c)\sqrt{n}. \tag{26}$$

8 Comparison with previous work

In this section we compare our results and computations with simulation results reported in previous works on airplane boarding, namely, Van Landeghem and Beuselinck (2002) and Van den Briel *et al.* (2003). Another work, Marelli *et al.* (1998), does not provide enough information for the purposes of comparison, although its main conclusions seem to match those of the other works. The recent work of Ferrari and Nagel (2005) is similar to that of Van Landeghem and Beuselinck (2002) both in methods and results, but uses a faster and less detailed simulation. The comparison with the work of Van Landeghem and Beuselinck presented below applies to Ferrari and Nagel (2005) as well. Van Landeghem and Beuselinck (2002) have carried out computer based simulations of 47 different boarding policies on an Airbus A320 with 23 rows and 132 seats, each row (apart from first class) carrying 6 passengers, three on either side of the single aisle. The design and input of the simulations were based on observations of passenger boarding and interviews with personnel at Brussels airport. Each experiment was performed 5 times, and the average and standard deviation are recorded on page 302 of loc. sit. and plotted on page 303. A detailed description of the boarding procedures which were simulated is given on pages 299–300. We shall refer to the simulations as the V-B simulations.

The simulations take into account some observed delays which we have not modeled. Passengers in the V-B simulations travel along the aisle at finite speeds according to some distribution. In our combinatorial boarding model we assume that passengers travel at infinite speeds when unobstructed. The V-B simulations attach to each passenger between 1 and 3 carry-on items, to be stored in the overhead bin compartments. The simulation keeps track of the available bin space. Towards the end of the boarding process, passengers sometimes need to search a large number of bins before they find space for their luggage, causing further delay in aisle clearing time. Also, as noted earlier, late arriving passengers have to unseat other passengers in order to get to their assigned seat. Consequently, the delay becomes larger towards the end of the V-B simulation. The simulations also take into account the situation in which an occasional passenger sits in the wrong place. When the passenger whose assigned seat is taken arrives, the passenger who occupies the wrong seat gets up and moves to the correct place.

In the V-B simulations, the 132 passengers are divided into groups called in succession to join the queue, while the order within the group is random. In 18 of the experiments, the number of groups is very large and the boarding process becomes nearly deterministic, corresponding to various combinatorial instances of the boarding process, as discussed in Section 2.1. The fastest boarding methods according to the simulations belong to these tightly controlled boarding methods. The best method is the one calling window passengers from one side of the aisle first in descending order, followed by window passengers from the other side in descending order, and similarly for middle and then aisle passengers. The problem of families or other small parties of passengers traveling together being separated by such policies can be solved by allowing such parties to board together according to the minimal group number among participants. Van Landeghem and Beuselinck found this policy to be about 2.5 times faster than random boarding. This policy is also the fastest according to our combinatorial model, allowing passengers to board in 6 rounds, far better than the random policy. In fact, the relative ranking among these 18 policies nearly coincides with the total boarding time according to our corresponding combinatorial models. It is unlikely, though, that an airline can exert such detailed control on the order in which passengers board, and we do not know of any example in which such strict methods are practiced. Therefore, we will not analyze these policies in great detail.

The remaining 29 experiments (policies) concern classes defined by blocks of rows or blocks of half-rows, consisting of all passengers on one side of the aisle. We can compare the results of these experiments with the calculations we have performed. We omitted one policy which involves placing first-class passengers first and, in addition, 3 policies which divided the passengers into 20 groups. As noted before, for such a large number of groups the probabilistic analysis tends to be irrelevant. This leaves us with 25 policies. We need some estimate of the parameter k. Given that there are 6 passengers per row, and assuming an estimate of 0.66 for the average width of a passenger, we obtain an estimate of $k = 4$. We performed a complete set of computations assuming this value for the 25 policies considered by Van Landeghem and Beuselinck. We also computed boarding times for the policies with the estimate $k = 3.5$, and observed that the results remain qualitatively the same. Almost all computations were performed using (16), (17), (22), (24) and (26) in a straightforward manner. In a few cases, some additional arguments were needed. An explanation of the basis of these other calculations is given in appendix D of the additional online material.

In terms of our terminology, Van Landeghem and Beuselinck simulated F_1, F_2, F_3, F_4, F_6 and F_{10}. In addition, they simulated $F_{2,2}$, $F_{2,3}$, $F_{2,4}$, $F_{2,6}$ and $F_{2,10}$, using passengers on the right side of the aisle as the first class and passengers on the left side as the second class. The remaining 15 out of 25 simulations were of type $F_{m,\sigma}$ and $F_{c,m,\sigma}$ for various permutations. A list of the permutations used in the V-B simulations is given in Appendix C of the online material.

We consider F_1 to be the basic policy which will serve as a yardstick for measuring all other policies, and normalize its boarding time to be 1 (24.7 minutes according to loc. sit.). We compute for each policy the ratio of the boarding time of the policy to that of F_1. The computations are done with the parameter setting $k = 4$. The results are presented in Figure 3.

The numerical values of the results are tabulated in Appendix B of the online material.

We make the following observations and conclusions.

A) The computational results using the Lorentzian geometry approach and the results of the V-B simulations agree to a large extent. The correlation coefficient between the $k = 4$

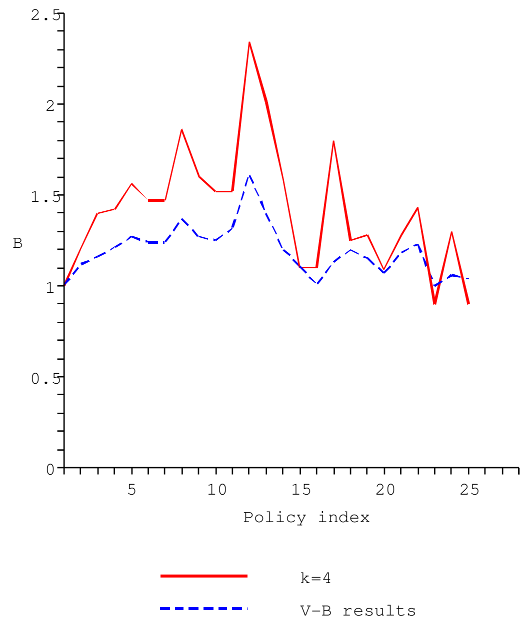

Figure 2: Boarding time (B) according to the model and the V-B simulations.

computations and the V-B results is 0.965. There is also strong agreement with the V-B simulations on the ordering of policies according to boarding time. This result – we believe – reinforces both approaches.

B) The main disagreement between results comes in cases where the number of classes is large. In experiment 12 with policy F_{10}, we are in a situation with $k > 2\ln 2$ and $m = 10$. Each class in this case contains merely 13 passengers. By (22), the boarding time of F_m tends to infinity with m, resulting in an overestimate. Underestimates occur when there are many small groups which do not block each other as in experiments 23 and 25.

C) According to (22) and (26), for fixed values of m and $k \geq c(3/4 + \ln 2$, $d_{c,m}$ is a decreasing function of c. In agreement with this statement, we observe that indeed the boarding times of F_2, F_3, F_4, F_6 are all greater than the corresponding boarding times of $F_{2,2}, F_{2,3}, F_{2,4}, F_{2,6}$ according to the V-B simulations.

D) In some cases the V-B simulations distinguish between pairs of policies which our models cannot distinguish. In the pairs 15 and 16, 19 and 21, and 23 and 25, the groups of passengers consist of passengers from the same rows but from different sides of the aisle. Our models of airplane boarding cannot distinguish between seat assignments differing only by a symmetry with respect to the aisle, and hence the expected boarding time is the same for the pairs. In all three pairs, the V-B results show a difference. This difference is small for 19 and 21, as well as for 23 and 25, but larger for 15 and 16 by about 10 percent. The most natural explanation for the difference is simple statistical fluctuations of the average over the

5 trials of each experiment. The average of the experiments turns out to be in very good agreement with the analytical computations. Another possible explanation is the fact that the V-B simulations keep track of which overhead bins are available. If the simulations also assume that passengers who sit in a given side also try to place their luggage on that same side, then the symmetry is broken and the experiments can be distinguished, even though the differences should still be small.

E) Given observation (C), it is natural to suggest the use of policies with $c = 3$, in particular $F_{3,2}$, which divides passengers into 6 classes. By observation (A), we expect our predictions for this policy to be fairly good since m and the number of groups cm are small and $k/c > 1$. The expected boarding time is 0.95 when we assume $k = 3.5$, 0.99 when we assume $k = 4$ and 1.04 when we assume $k = 4.5$. It seems unlikely that passengers will tolerate a division into more than 6 groups, and within that range $F_{3,2}$ has the best predicted performance. We note that the expected gain over the uniform policy F_1 is rather negligible and is probably not worth the extra complication. We conclude, in agreement with the simulation results of V-B, that it is difficult to improve upon the uniform boarding policy, using policies which do not change the delay distribution D, without excessively burdening the passengers.

F) We are left with the option of using seat type classes which decrease the value of D. Since D operates on the metric as a scaling parameter, seat type classification should uniformly improve any given policy. The policy $F_{3,1}$ with seat type classification, along with some variants which resemble $F_{3,2}$, have been suggested by Van den Briel *et al.* (2003, 2005). They consider a different measure for assessing airplane boarding policies. Essentially, they measure the total number of blocking incidents in the airplane rather than the longest sequence of blockings as in this paper. While the two measures are different, they seem to be positively correlated. Their measure leads Van den Briel *et al.* to consider $F_{3,1}$ with seat type classification, a policy which they compare with F_2, F_3, F_4 and F_5 using simulations based, as in the V-B case, on observations of real instances of airplane boarding. As with the V-B simulations and our calculations, they find that boarding time for an F_m policy increases with m. Van den Briel *et al.* (2005) report on the success of a change in policy at America West airlines. They converted from a (non-uniform) back-to-front policy to a policy which closely resembles $F_{3,2}$ with seat type classification. A 20 percent reduction in boarding time is reported. This empirical observation is also qualitatively in line with our analysis.

9 Summary and future work

We have introduced a multi-parameter discrete random process, which captures the essential features of the airplane boarding process. We have shown that the asymptotic behavior of the random process can be captured by the geometry of a 2-dimensional space-time, which depends on the parameters of the process. In particular, the diameter of the space-time provides the expected boarding time for the process. The theory allows us to compute closed-form estimates for the boarding time in many cases of interest. Comparisons with various simulations, used to study airplane boarding, show that they make similar suggestions in terms of preferred boarding policies. Most of the differences pertain to the question of how much one policy is better than another. This issue should probably be settled by real field experiments rather than simulations. For an airplane with 6 passengers per row, the main candidates for a good boarding policy have been reduced to F_1, the single class policy, $F_{3,1}$ with seat type classification for 3 group policies and $F_{3,2}$ with seat type classification for 6 group policies. It is possible, using the analytic model, to reevaluate the effect of policy changes when the distance between rows or the number of passengers per row/per aisle change. Upon adding passengers per row or squeezing the distance between rows, the boarding process becomes slower, and the F_1 policy more attractive.

The most immediate challenge in terms of the analysis of airplane boarding seems to be the analysis of policies in which seats are unassigned. It is suggested in Van Landeghem and Beuselinck that such policies should behave in a similar fashion to F_1, but with somewhat worse results. Based on sporadic personal evidence, we believe that such policies may in fact prove to be more efficient. The basic difficulty is in establishing a reasonable passenger reaction model. This will require extensive field work and interviews with customers. One insight which may be gained from the present work is that the effective value of the parameter k controls to a large extent boarding time. We believe that passengers which are strangers will tend to space themselves along the airplane, thus effectively lowering the value of k. Other trends, such as a preference for front seats, have detrimental effects. However, until a more detailed study of passenger behavior is performed, we cannot provide a useful model for unassigned seating policies.

Acknowledgments

We wish to thank Percy Deift and Ofer Zeitouni for several helpful conversations and much encouragement. We would also like to thank Aharon Davidson, Matt Visser, Menkes Van den and Jinho Baik for illuminating exchanges and discussions.

References

Aldous, D., P. Diaconis. 1999. Longest increasing subsequences: From patience sorting to the Baik-Deift-Johansson theorem. *Bull. AMS*, 36(4), 413–432.

Deuschel, J.D., O. Zeitouni. 1995. Limiting curves for iid records. *Ann. of Prob.*, 23, 852–878.

Ferrari, P., K. Nagel. 2005. Robustness of efficient passenger boarding in airplanes, *Transportation Research Board Annual Meeting*, paper number 05-0405. Washington D.C., Also http://www.trb.org/.

Marelli, S., G. Mattocks, R. Merry. 1998. The role of computer simulation in reducing airplane turn time. *Boeing Aero Magazine*, Issue 1.

Penrose R. 1972. Techniques of differential topology in relativity. Regional Conference Series in Applied Mathematics 7, published by SIAM.

Van den Briel, M., J. Villalobos, G. Hogg. 2003. The aircraft boarding problem. *Proc. of the 12'th Industrial Eng. Res. Conf., IERC*, CD ROM, article number 2153.

Van den Briel, M., J. Villalobos, G. Hogg, T. Lindemann, A.V. Mule. 2005. America West develops efficient boarding strategies, *Interfaces*, To appear.

Van Landeghem, H., A. Beuselinck. 2002. Reducing passenger boarding time in airplanes: A simulation approach. *European J. of Operations Research*, 142, 294–308.

Vershik, A.M., S.V. Kerov. 1977. Asymptotics of Plancherel measure of the symmetric group and the limiting form of Young tables. *Soviet Math. Dokl.*, 18, 527–531.