Delay Propagation in Re-Scheduled Queueing Systems

Claus Gwiggner
DRAFT

Abstract

We analyze a queueing system in which customers are scheduled to arrive in certain time slots, and re-scheduled in case of late arrivals. Moreover, the queueing delays are distributed between two waiting rooms. We quantify the propagated delay that is triggered by a late customer. This propagation depends on the ratio of the usage of the waiting rooms. We then identify conditions under which waiting cost functions are convex. This knowledge is useful to derive fuel-efficient aircraft sequencing strategies.

1 Introduction

There are many systems in which customers are scheduled to arrive at certain times and re-scheduled in case of disturbing events, such as delayed arrivals or capacity changes. Railway and aircraft sequencing are just two examples. Today, such schedules are often created by human experts who would like to better understand the impact of the disturbances on the system level: how far do delays propagate? How long does it take until the system is recovered from the disruptions?

For example in aircraft sequencing, it often happens that the schedules are violated due to weather conditions or competition for punctual arrivals [Nolan(1998)]. Aircraft have to be re-scheduled, which creates additional delays. These delays propagate through the airspace and increase the workload of the controllers who are in charge of safe and efficient operations.

While this phenomenon is not new, only few papers analyze it. In the air-traffic literature, [Erzberger(1995)] runs simulations and gives rules of thumb for fuel-efficient scheduling policies in the presence of trajectory uncertainties. [Bayen et al(2005)Bayen, Grieder, Meyer, and Tomlin] analyze the dynamics of speed-control but only based on a deterministic model. Others [Bayen and Tomlin(2004)], [Balakrishnan and Chandran(2010)], [D'Ariano and D'Urgolo(2010)] focus mainly on the computational complexity of the underlying sequencing problem but do not increase the intuitive understanding. This is also the case for the appointment scheduling literature, which begins with the works of [Winsten(1959), Mercer(1960), Mercer(1973)], but loses its traces then. Today, the field is widely occupied with algorithmic questions

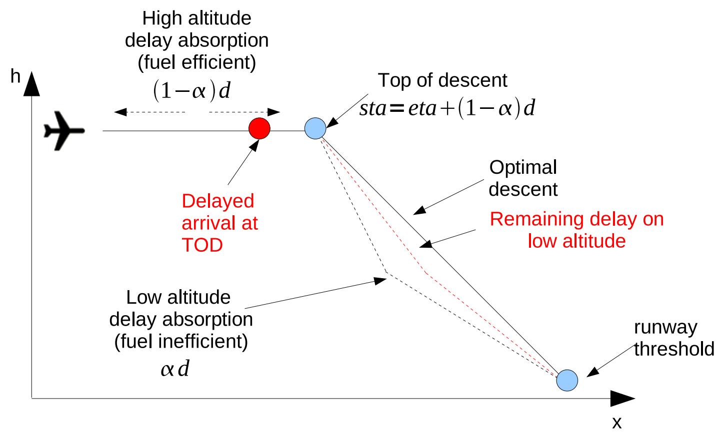

Figure 1: Delay absorption under uncertainty.

[L. V. Green(2008)] (and the references therein). A recent analytical approach is [Guadagni and Ndreca(2010)], but results are not mature yet.

The contribution of this paper is an analysis of delays that propagate through a flow of pre-scheduled customers in a highly congested queueing system. We build an elementary stochastic model and identify the conditions under which the average delays can be minimized. The remainder of the paper is organized as follows: in the next section we introduce the basic model and an analysis of the delay propagation. Then we introduce waiting costs and identify conditions under which delay minimization can be achieved analytically. We conclude with a summary and outlook on the next steps in research.

2 Stochastic Model

A typical instance of the problem occurs during aircraft sequencing. Given the *estimated time of arrival (eta)* of all aircraft, the task is to generate *scheduled times of arrival (sta)*, such that a minimum separation m is guaranteed between successive aircraft. This is a well known scheduling problem, generating *queueing delays* d for every aircraft [Bayen and Tomlin(2004)], [Balakrishnan and Chandran(2010)], [D'Ariano and D'Urgolo(2010)].

Our analysis extends the sequencing problem: when queueing delays are absorbed in high altitudes, fuel burn is minimized for individual flights [Erzberger(1995)]. But due to trajectory prediction errors, there is a risk that lost landing slots propagate back to the remaining aircraft. This increases the total delay, and as a consequence the total fuel burnt. This means that queueing delays have to be distributed between the high altitudes (fuel efficient) and low altitudes (fuel inefficient), even when the objective is to minimize fuel consumption.

2.1 Basic model

As basic model we consider a single arrival trajectory of aircraft i, as depicted in Figure 1. There are two points of interest: the top of descent and the runway threshold (blue circles). One part of the queueing delay d_i is absorbed prior to the top of descent

$$sta_i = eta_i + (1 - \alpha)d_i, \tag{1}$$

and the other part αd_i is included in the sta at the runway threshold. Here, $\alpha \in [0, 1]$ is the percentage of delay to be absorbed on low altitude.

Due to trajectory prediction errors, the *actual time of arrival* at the top of descent (red point) will be

$$ata_i = sta_i + \epsilon_i, \tag{2}$$

where ϵ_i is a random variable describing the error in achieving the scheduled time of arrival. One can guess from the Figure that there is a need to re-schedule the flow when the prediction error ϵ_i is larger than αd_i.

2.2 Delay Propagation

Consider now a flow of aircraft with scheduled times of arrival $sta_1 < sta_2 < \ldots$ as depicted in Figure 2. Since the flow is already scheduled, we know that the spacing between two successive aircraft is $m_i + a_i$, where m_i is the minimum separation and $a_i \geq 0$ is the remaining spacing in the case they were initially far enough behind each other (first line).

A trajectory prediction error ϵ_i will propagate backwards, if it exceeds the amount of delay absorbed on low altitude αd_i (green boxes)

$$\epsilon_i > \alpha d_i.$$

Then, the separation with aircraft $i + 1$ reduces to $m_i + a_i - \epsilon_i$. Thus, aircraft $i+1$ has to be re-scheduled by $\epsilon_i - a_i$, in order to keep the separation m_i (second line in Figure 2). Aircraft $i + 1$ itself absorbs αd_{i+1} on low altitude. Thus, the propagation will continue, if

$$(\epsilon_i - a_i) > \alpha d_{i+1}.$$

Following this mechanism recursively leads to the condition that the error ϵ_i triggers a propagation of the $(k + 1)st$ follower of aircraft i if

$$(\epsilon_i - \sum_{j=0}^{k-1} a_{i+j}) > \alpha d_{i+k}. \tag{3}$$

With a first-come-first-served policy, the a_i's can be easily expressed in terms of the inter-arrival times in the original flow and the queueing delays, leading to $a_i = max(s_i - m_i - d_i, 0)$, where $s_i = eta_i - eta_{i-1}$ (see Appendix A). Thus, the ingredients of (3) are known. More interesting is k, which is the largest number, such that expression (3) is valid. It has the form of a first-visit time or stopping trial (e.g. [Feller(1970), Wolff(1989)]) which depends on the random variable $(\epsilon_i - \alpha d_{i+k})$. Its distribution will be discussed in section 3.2.

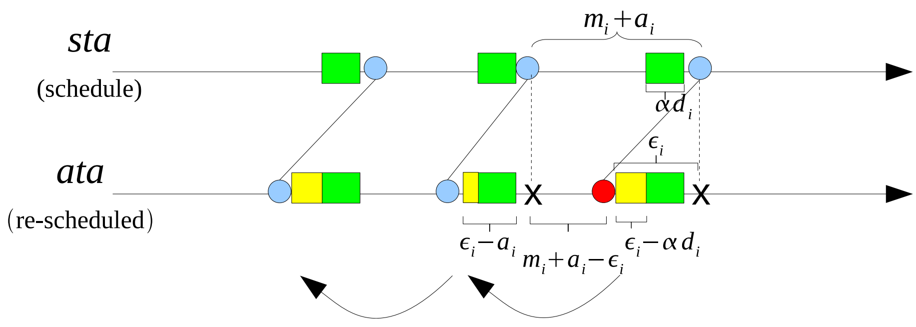

Figure 2: Scheduling process and delay propagation.

Summing all terms up, the amount of propagated delay, triggered by aircraft i is thus

$$D_i = (\epsilon_i - \alpha d_i) + (\epsilon_i - a_i - \alpha d_{i+1}) + \ldots + (\epsilon_i - \sum_{j=0}^{k-1} a_{i+j}) - \alpha d_{i+k} \tag{4}$$

$$= (k+1)\epsilon_i - [(ka_i - \ldots - a_{i+k-1}] - \alpha \sum_{j=0}^{k} d_{i+j} \tag{5}$$

$$= (k+1)\epsilon_i - \underbrace{\sum_{j=0}^{k-1}(k-j)a_{i+j}}_{\approx 0} - \underbrace{\alpha \sum_{j=0}^{k} d_{i+j}}_{\approx (k+1)\alpha d_i}, \tag{6}$$

D is a random variable, because it depends on the queueing delays d, the trajectory prediction error ϵ, the natural spacing a and the number of aircraft affected by the propagation k. This is a complicated expression, but in periods of high traffic density, the pre-scheduled traffic is quite tight. This means that the terms a_i in equation (6) are likely to be very small. A simple idea is thus to approximate the propagated delay of aircraft i by

$$D_i \approx \begin{cases} (k+1)(\epsilon_i - \alpha d_i) & \text{if } \epsilon_i \geq \alpha d_i \\ 0 & \text{else,} \end{cases} \tag{7}$$

In reality, natural spacing a_i between aircraft will absorb the propagated delay and making it smaller and smaller. Moreover, we approximated $\alpha \sum_{j=0}^{k} d_{i+j}$ with $(k+1)\alpha d_i$ which will be justified in the next section. This is why equation (7) is only an approximation.

The trajectory prediction errors ϵ can be assumed to occur independently of the queueing delays d. This is not the case for k, the number of aircraft affected by the propagation: the larger the expression $(\epsilon_i - \alpha d_i)$, the longer the propagation will be. The average propagated delay of aircraft i is then

$$\mathbb{E}(D_i) = \sum_{k=0}^{\infty} \int_{u=0}^{\infty} \int_{v=0}^{u/\alpha} (k+1)(u-\alpha v)P(k,u,v)dvdu \tag{8}$$

$$= \sum_{k=0}^{\infty}(k+1)\left[\int_{u=0}^{\infty} \int_{v=0}^{u/\alpha} (u-\alpha v)P(k \mid u,v)P(u,v)dvdu\right] \tag{9}$$

$$= \sum_{k=0}^{\infty}(k+1)\left[\int_{u=0}^{\infty} \int_{v=0}^{u/\alpha} (u-\alpha v)P(k \mid u,v)f(u)g(v)dvdu\right] \tag{10}$$

where $P(.)$ and $P(k \mid .)$ are the corresponding joint and conditional distributions and f, g the densities of ϵ and d. Note that with 'ordinary' integration limits, the shape of $\mathbb{E}(D)$ would be straightforward.

3 Results

Equation (10) contains three elements: the number of aircraft affected by propagation k, the queueing delay d and the trajectory prediction error ϵ. We obtained results about k (will be published elsewhere) and about the typical shape of D.

3.1 Independence

In order to increase our understanding of equation (10), we first assumed independence between k, the number of aircraft affected by propagation and $\epsilon_i - \alpha d_i$, the size of the initial trajectory prediction error. Then, $P(k|\epsilon, d) = P(k)$ and thus

$$\mathbb{E}(D_i) = \mathbb{E}(k)\mathbb{E}(\epsilon_i - \alpha d_i). \tag{11}$$

In this section, we analyze the shape of model (11) under various distributions. In the next section we come back to the more complicated case of dependent random variables.

Pre-scheduled random arrivals (PSRA/D/1) A regular pre-scheduled and randomly disturbed arrival is given by

$$ata_i = \frac{i}{\lambda} + \epsilon_i \tag{12}$$

where $\lambda \in \mathbb{R}$ is the average arrival rate (min^{-1}).

In periods of over-demand ($\lambda > \mu$), the initial queueing delay of aircraft i is simply $d_i = i(\frac{1}{\mu} - \frac{1}{\lambda})$. Here, μ is the average service rate which is determined by the minimum separations m_i. In a rush of n aircraft, the probability that an aircraft has delay d is thus uniformly distributed $d \sim U[0, n(\frac{1}{\mu} - \frac{1}{\lambda})]$. With

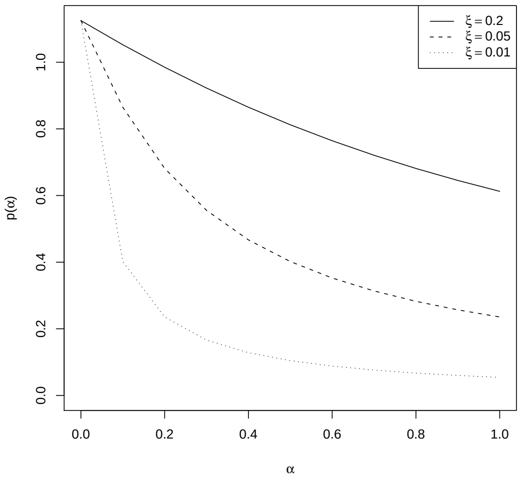

Figure 3: Behavior of $E(\epsilon_i - \alpha d_i)$.

$\epsilon \sim U[-\sigma, \sigma]$ this leads to

$$\mathbb{E}(\epsilon_i - \alpha d_i) = \frac{1}{2\sigma n(\frac{1}{\mu} - \frac{1}{\lambda})} \int_{u=0}^{\sigma} \int_{v=0}^{u/\alpha} (u - \alpha v) dv du \tag{13}$$

$$= \frac{\sigma^2}{12\alpha n(\frac{1}{\mu} - \frac{1}{\lambda})} \propto \frac{1}{\alpha}. \tag{14}$$

This is a sharply decreasing function of α.

General queue in heavy traffic (G/G/1) In congested times, the queuing delay is distributed exponentially with parameter $\xi = \frac{2a}{\sigma^2}$, where a is the difference between average arrival and service rate and σ^2 the sum of the variances of the processes and provided that α/σ is small [Wolff(1989)]. Although this is not the most precise approximation (which we saw by simple simulation examples), it gives us insight into the shape of the propagated delay again. If the trajectory prediction error is also exponentially distributed $\epsilon \sim Exp(\sigma)$, we have

$$\mathbb{E}(\epsilon_i - \alpha d_i) = \frac{\xi}{\sigma(\alpha\sigma + \xi)} \tag{15}$$

$$= \left(\frac{\alpha\sigma^2}{\xi} + \sigma\right)^{-1}. \tag{16}$$

Again, this is a sharply decreasing function of α.

Exponential power distribution Although analytical solutions still exist for a few other distributions, we computed the shape of equation (11), when ϵ

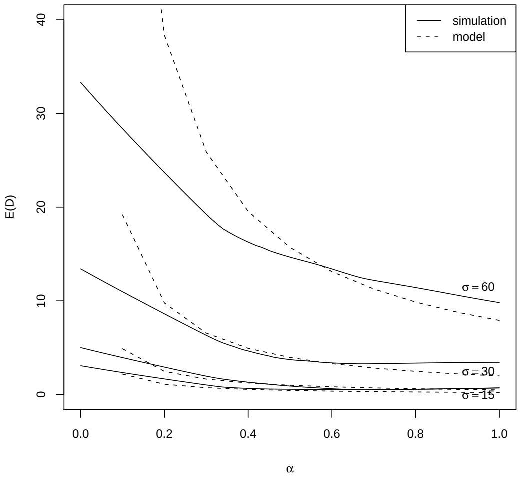

Figure 4: Average propagated delay.

is member of a family of distributions called the the exponential power distribution. Its density is proportional to

$$f(x, \mu, \alpha, \beta) \propto e^{-(\frac{|x-\mu|}{\alpha})^{\beta}}, \tag{17}$$

where μ is a location, α a scale and β a shape parameter. The family includes the Laplace ($\beta = 1$) and Normal distributions ($\beta = 2$) and converges to the uniform distribution for $\beta \to \infty$. For $\beta < 2$ its tails are heavier than a Normal and for $\beta > 2$ lighter.

Anyhow, it turned out that is was the distribution of the queueing delays that had the largest impact on the shape of (11). Figure 3 shows a typical result, where we varied ξ over 0.01, 0.05 and 0.2. The larger ξ grows, the less bended the curve. Note that very small ξ correspond to very high traffic densities. The second parameter with impact on the shape of equation (11) was the variance of the trajectory prediction error. For large variances, the curvature became less involved, as well. For an idea of the interaction of these parameters please look in the Appendix, where we solved two other special cases analytically.

Last but not least, a validation of the model can be seen in Figure 4. The dotted lines are from a Monte-Carlo simulation of the sequencing problem. For small variances of the trajectory prediction errors σ, the model (11) describes the propagated delays well. In case the trajectory prediction errors get large, we obtain less accurate results. But we see that the theoretical model captures the essential nature of the delay propagation. This lets us conjecture that during high congestion, $\mathbb{E}(\epsilon_i - \alpha d_i)$ *always* sharply decreases with α. This property is important for the optimization of fuel-burn.

3.2 Dependent propagation

The independence assumption between k and $\epsilon - \alpha d$ helped to understand the typical behavior of equation (10) but it performed badly in cases of large trajectory errors. With $k \propto \epsilon - \alpha d$, there is just a scale to above. Thus, a non-linear relationship must exist. We have some empirical results on this relationship, that will be reported in another paper.

4 Conditions for Fuel Optimization

In this section we introduce costs for the two waiting rooms and ask the question: what are the conditions, such that a trade-off between delay absorption on high and low altitude exists?

In equilibrium, provided it exists, the average queueing delay is a constant. Minimizing average cost then simplifies to a solution of

$$min_{\alpha} \;\; c(\alpha) = [\alpha + (1-\alpha)\lambda]d(\alpha) \tag{18}$$

$$d(\alpha) = d_o + \mathbb{E}\{D(\alpha)\}, \tag{19}$$

where d_0 is the average queueing delay, $\mathbb{E}(D)$ is the propagated delay, and $\lambda < 1$ is the relative fuel consumption index [kg/minute] in high altitude.

The second derivative of c is

$$c''(\alpha) = 2(1-\lambda)d' + (\alpha + (1-\alpha)\lambda)d''. \tag{20}$$

For c being convex, we need thus

$$\frac{d'}{d''} > -\frac{\alpha}{2} - \frac{\lambda}{2(1-\lambda)}. \tag{21}$$

In the previous section we have seen that the propagated delay decreases sharply with α, the fraction of delay absorbed in low altitude. A one-parameter family with a similar behavior is

$$d_\theta(\alpha) = d_p(1-\alpha)^\theta + d_0, \;\; \theta \geq 1, \tag{22}$$

where $d_p, d_0 \in \mathbb{R}$ are scaling constants [Mesterton-Gibbons(2007)]. For $\theta = 1$ it is a diagonal, and for $\theta \to \infty$, it is parallel to the vertical axis. In this family, the left side of equation (21) becomes

$$\frac{d'}{d''} = \frac{-d_p\theta(1-\alpha)^{\theta-1}}{d_p\theta(\theta-1)(1-\alpha)^{\theta-2}} \tag{23}$$

$$= \frac{1}{\theta-1}\alpha - \frac{1}{\theta-1}, \tag{24}$$

which is linear in α. Thus, the condition (21) for c being convex depends on the intersection of two lines $y_i = a_i\alpha + b_i, \;\; i = 1,2$ with $a_1 = 1/(\theta-1), b_1 = -1/(\theta-1), a_2 = -1/2, b_2 = \lambda/2(1-\lambda)$. Since for $\theta > 1, \;\; -1/2 \neq 1/(\theta-1)$, they have a unique intersection α^*. Thus, c is convex for all $\alpha \geq \alpha^*$, which brings us to a second condition

$$\alpha^* = \frac{b_2 - b_1}{a_1 - a_2} < 0. \tag{25}$$

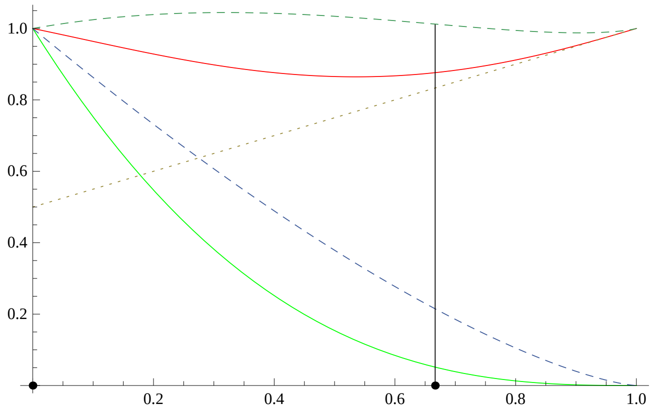

Figure 5: Condition for existence of unique minimum.

The numerator is negative if $b_2 < b_1$, which leads to $\theta > \frac{2}{\lambda} - 1$. For the denominator, the solution is $\theta < -1$, which is infeasible, since we have $\theta \geq 1$.

To summarize, the cost function c is a function of α, the ratio of queueing delay between high and low altitude. The queueing delay is a decreasing function of this ratio. We embed this decreasing function in a one-parameter family and obtain a condition such that c is convex. Figure 5 illustrates the equations. The horizontal axis is α. The vertical axis is dimensionless (all units in the plot are normalized). We visualize two cases. The bold lines represent the case, where the cost function is convex over the whole domain $0 \leq \alpha \leq 1$. Here, the green function is the propagated delay $d - d_0$, and the red one the fuel cost (equation 18). The dashed lines represent the case, where the cost function is not convex. Our second condition for convexity (25) implied that $\theta > \frac{2}{\lambda} - 1$. For the plot we selected $\lambda = 0.5$ and $\theta = 1.4 < \frac{2}{0.5} - 1$. The dot on the horizontal axis represents α^*, where the two lines given in (21) intersect. We can see that the cost function is non-convex before that point, but convex afterwards. Finally, the dotted linear line is the fuel consumption in the case that no trajectory prediction errors occur ($E(D) = 0$). In this case, the most fuel-efficient strategy is to absorb all queueing delays in high altitude ($\alpha = 0$).

We performed the same calculations with other function families (e.g. $d(x) = e^{-\theta x}$ or $d(x) = 1 - x^\theta,\ \theta < 1$), and the results were similar. In particular, the quotient of their first and second derivatives is linear for all of them.

5 Conclusions and Future Work

In this paper we analyzed a system, where customers are initially scheduled to arrive in certain time slots, and re-scheduled in case of missed appointments. The focus of our analysis was the delay propagation due to late arrivals. We formalized the problem as a queueing process, where waiting times are distributed between two waiting rooms. Our first result was that the propagated delay decreases sharply with the fraction of time spent in the first waiting room. This is intuitively clear, since the waiting rooms may be seen as a buffers that protect the server from too long queues. But against intuition was that the rapid de-

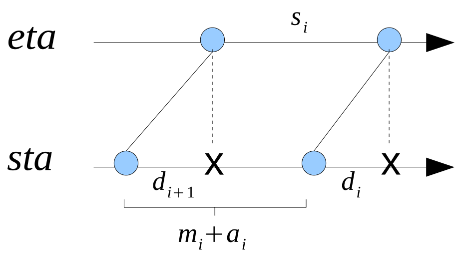

Figure 6: Natural buffers a_i's.

crease of the propagated delay occurs also in a system without any exponential distributions. We then analyzed the conditions under which decreasing functions lead to a convex delay cost. Here again, a non-intuitive property was that linearly decreasing delay propagation functions lead to a concave cost functions, meaning that there is no fuel minimizing trade-off strategy.

The current results are useful to understand the general delay generating mechanisms of our system. For particular case-studies, we need to give a physical meaning to the parameter in the last section. For future work, we think that also more has be known about the number of customers that are affected by delay propagation. Moreover, the problem can be re-formulated as a chain of two servers. Then, new mechanisms to control the propagation, for example with the insertion of variable buffer sizes, may be studied. This knowledge is elementary to understand the conditions for efficient sequencing strategies under uncertainties.

A Natural Buffers

In order to see what the a_i's are, please take a look at Figure 6. Initially, aircraft i and i+1 are separated by a distance s_i. After the scheduling, we know that their separation will be $m_i + a_i$. Both aircraft were delayed by $d_i, d_{i+1} \geq 0$. Thus

$$m_i + a_i = d_{i+1} + s_i - d_i.$$

From the queueing relationship $d_{i+1} = max(d_i + m_i - s_i, 0)$, this leads to

$$a_i = \begin{cases} 0 & if\ d_{i+1} > 0 \\ s_i - m_i - d_i & else \end{cases}$$

Note that it is intuitively clear that $a_i = 0$ in the case that $d_{i+1} > 0$.

B Analytical solution

Let's assume that the system is Markovian with arrival rate λ and service rate μ. Then the equilibrium delay distribution (including processing time) is exponential with parameter $\mu - \lambda$ [Wolff(1989)].

We solved two cases of equation (11) analytically, depending on the distribution of the trajectory prediction errors ϵ:

1. Uniform distribution: $\epsilon \sim U[-\sigma, \sigma]$.
   Then

$$\begin{aligned}\mathbb{E}(\epsilon_i - \alpha d_i) &= \int_{u=0}^{\sigma} \int_{v=0}^{u/\alpha} (u - \alpha v) \frac{1}{2\sigma} (\mu - \lambda) e^{-(\mu-\lambda)v} dv du \\ &= \frac{1}{2\sigma} \int_{u=0}^{\sigma} u + \frac{\alpha}{(\mu - \lambda)} \left(e^{-\frac{(\mu-\lambda)u}{\alpha}} - 1 \right) du \\ &= \frac{1}{2\sigma} \left[\xi^2 (1 - e^{-\frac{\sigma}{\xi}}) - \xi\sigma + \frac{\sigma^2}{2} \right],\end{aligned}$$

   where $\xi = \frac{\alpha}{(\mu-\lambda)}$. The expression is complicated, but has an interesting property: in very high traffic densities ($\rho > 0.9$), it decreases sharply with α, whereas in moderate traffic densities ($\rho < 0.8$), its decline is more or less proportional to α (as can be seen in Figure 3).

2. Gaussian distribution: $\epsilon \sim N(0, \sigma)$

$$\mathbb{E}(\epsilon_i - \alpha d_i) = \int_{u=0}^{\infty} \int_{v=0}^{u/\alpha} (u - \alpha v) \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-u^2}{2\sigma^2}} (\mu - \lambda) e^{-(\mu-\lambda)v} dv du$$

   There is a 'closed-form' solution, but it contains the error function and several other terms (obtained with Mathematica). It is more complicated than the above case, so we do not reproduce it here.

References

[Balakrishnan and Chandran(2010)] Balakrishnan H, Chandran B (2010) Algorithms for scheduling runway operations under constrained position shifting. To appear in Operations Research

[Bayen and Tomlin(2004)] Bayen A, Tomlin C (2004) Optimal arrival traffic spacing via dynamic programming. AIAA Conference on Guidance, Navigation and Control Providence, RI

[Bayen et al(2005)Bayen, Grieder, Meyer, and Tomlin] Bayen A, Grieder P, Meyer G, Tomlin C (2005) Lagrangian delay predictive model for sector-based air traffic flow. Journal of Guidance, Control and Dynamics 28/5

[D'Ariano and D'Urgolo(2010)] D'Ariano A, D'Urgolo P (2010) Optimal sequencing of aircraft take-off and landing at a busy airport. In: Proceedings of the 13th International IEEE Annual Conference on Intelligent Transportation Systems, Madeira Island, Portugal

[Erzberger(1995)] Erzberger H (1995) Design principles and algorithms for automated air traffic management. In: AGARD Lecture Series No. 200 on Knowledge-based Functions in Aerospace Systems

[Feller(1970)] Feller W (1970) An Introduction to Probability Theory and Its Applications, Volume 2. Wiley, 2nd edition

[Guadagni and Ndreca(2010)] Guadagni G, Ndreca S (2010) Queueing systems with pre-scheduled random arrivals. Mathematical Methods of Operations Research Online first

[L. V. Green(2008)] L V Green SS (2008) Reducing Delays for Medical Appointments: A Queueing Approach. Operations Research 56(6):15261538

[Mercer(1960)] Mercer A (1960) A queueing problem in which the arrival times of the customers are scheduled. Journal of the Royal Statistical Society, Series B 22/1:108–113

[Mercer(1973)] Mercer A (1973) Queues with scheduled arrivals: A correction, simplification and extension. Journal of the Royal Statistical Society, Series B 35/1:104–116

[Mesterton-Gibbons(2007)] Mesterton-Gibbons M (2007) A Concrete Approach to Mathematical Modelling. Wiley

[Nolan(1998)] Nolan MS (1998) Fundamentals of Air Traffic Control. Third edition. Brooks/Cole, Wadsworth

[Winsten(1959)] Winsten C (1959) Geometric distributions in the theory of queues. Journal of the Royal Statistical Society Series B (Methodological) XXI/1:1–35

[Wolff(1989)] Wolff RW (1989) Stochastic Modeling and the Theory of Queues. Prentice-Hall, New Jersey