A Queueing Model for Airport Capacity
and Delay Analysis

H. B. Thiagaraj
Department of Mathematics
CSI College of Engineering, Ketti, India

C. V. Seshiaiah
Head of Department of Science and Humanities
Sri Ramakrishna Engineering College, Coimbatore, India

Copyright © 2014 H. B. Thiagaraj and C. V. Seshiaiah. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Air traffic, worldwide, keeps growing strongly, creating critical capacity situations and traffic congestion. Large delays are suffered by airlines, passengers and airport authorities alike. When a rough estimate is needed, the results of queuing theory can be used to analyse airport runway systems, but when airports are too congested or a more realistic description of the system behaviour is necessary, a simulation approach is a good alternative. This research outlines the limits of the analytical approach and shows how to build a simulation procedure. This procedure is able to measure the performance of an airport runway used only for arrivals, with different traffic mixes and operational variables. The impact of future technological systems is also considered.

Keywords: Air traffic, Delay, Queueing Model, simulation, BIA, Interpolation model

1. Introduction

Air travel delays when encountered it always seems to be at the end of a hard business trip or at the start of a well deserved vacation. We spend them caged in a terminal or plane, usually unable to see the cause of the delay, or worse, able to see that our destination is within
reach, yet unable to reach it. The obvious but naive solution is to add more capacity to the system; more airports, more runways, and more air traffic control ability which is not feasible. We face massive limitations on available land, of which airport need a great deal, of which airport create an excessive amount, and capital, of which airports use a lot. The alternative to large scale expansion of the system is to optimize the facilities we now have. In the air travel system, the bulk of delays are caused by excessive demand on limited facilities, causing queues to form for service, and forcing those who must wait in the queues to incur delay costs. The demand comes from arriving and departing planes, and the service they are demanding is usage of the airport runways, terminals, and other facilities. The capacity of the service facility (i.e. runways and airspace) is often uncertain due to weather conditions, non optimal controller behaviour and equipment failure. A model of air travel delays will incorporate the behaviour and uncertainty of these two components in some framework that allows the experimenter to investigate their interaction and resulting behaviour of the whole system.

There are two branches of modeling theory that can be useful in approaching this problem, Queueing theory and Simulation. Queueing theory permits deeper analysis at less cost with more restrictive assumptions. Simulation permits shallower analysis at greater cost with less restrictive assumptions. In this study we explore the model that these two disciplines provide, that might be useful in understanding airport delays and are simulated. The genesis of this exploration was in the need to analyze the delay characteristics of The Bengaluru International airport (BIA) which is also applicable to the other airports.

2. Background

Airports connect earthbound travellers to the much faster air travel system composed of commercial airliners and private and corporate planes. The services that airports provide and how they do so determine their capacity. Airport capacity in turn determines the capacity of the air travel system. The essential task of the airport is to act as an interface that allows one to pass from the land travel system to the air travel system, and vice versa. The land side of the airport is typical of public transportation facilities that need to move people through a ticketing process and onto different routes. Airport services can be divided into passenger processing and enplaning

The bulk of delays are encountered on the air side of the system, where passengers are delayed as a group in planes or terminals, often for an extended period. The air side system has two ‘modes’ operation, one for when weather and visibility are good, and other for poor weather conditions when instruments are necessary for navigation and landing. The air side services are best understood from the sequence of controllers who direct planes through the stages of arrival and the delays that may be encountered at each stage. Each enroute controller watches a sector of airspace over the Indian, Radar, voice communications, and radio beacons on the ground, called fixes, are used to monitor and direct the aircraft passing through each sector. The airport arrival controller admits planes into the terminal airspace. The terminal airspace extends out in a
Queueing model for airport capacity

20 to 30 mile radius around the airport. The terminal airspace controller directs planes admitted to the airspace to proceed to the airport to land or, in the case of congestion, to delay their landing. The delay can take two forms. A slight delay might be introduced by having the plane reduce its speed or fly a wide arc. More substantial delays are introduced by placing the aircraft in holding stacks. Holding stacks are areas of the terminal airspace where aircraft fly the same oblong flight paths, separated by 1000 feet of altitude, around a ground fix. Up to seven or eight aircraft can be placed in a holding stack and multiple holding stacks can be used for temporarily ‘storing’ aircraft. Planes are taken off the stack from the bottom, and then the planes above them move down in sequence. Holding stacks in the terminal airspace and delay actions of the terminal airspace controllers are generated in response to congestion in succeeding stages of the arrival sequence (closer to the airport), and not by conditions particular to this stage. Some arrival controller actions, and possibly mistakes, can introduce small but significant delays at this stage.

In general, if the same runway is being used for both arrivals and departures, the tower controller only allows takeoffs to occur during gaps in the arrival sequence, thus departing the arrival sequence. Departing planes are often delayed while waiting on the taxiways or at terminals for takeoff clearance. These planes can often make up much of this delay time enroute by burning slightly more fuel. On some occasions the progression of landing will be halted for the planes on the ground to take off, but is rare. Once planes take off they require little controller attention and typically depart the terminal airspace without addition terminal airspace delay.

**Capacity Limitation and Delay Cost**

Delay can be introduced into the system from any overload or poorly performing service component. The stages of service that constitute an airport are arranged in a network. This network can be thought of as a single macroscopic server. We are concerned with the progression of traffic through this server as a whole. The particular arrangement of component servers and their interconnections in the network is often limited in total capacity by only few key components. These are the bottlenecks in the system. Improvements in the capacity of the server as a whole; the primary bottlenecks of the system are the first part that should be investigated.

In typical airport, the primary bottlenecks are always the runway system. The capacity constraint manifests itself through the rate at which the final vector controller brings planes from the terminal airspace to the outer mark of the final descent path. The relatively high demands, and the uncertain service capacity due to weather condition, make it obvious that overload runway systems can explain a majority of the delay encountered at airports. In this paper the runway system will be considered the primary bottlenecks that generate airport delays.
Case Study : BIA

The Bengaluru International Airport is a 4,050 acre international airport that is being built to serve the city of Bengaluru, Karnataka, India. The airport is located in Devanahalli, which is 30 km from the city. The new Bengaluru International Airport at Devanahalli will put Bangalore city on the global destination and offer travellers facilities comparable with the best international airports. The airport will offer quality services and facilities, which will ensure the comfort and ease of travel for all concerned. Construction of the airport began in July 2005, after a decade long postponement. A plan is also being processed for a direct Rail service from Bangalore Cantonment Railway Station to the Basement Rail terminal at the new International Airport. Access on the National Highway is being widened to a six lane expressway, with a 3 feet boundary wall, construction is moving ahead. As of June 2007, a brand new expressway connected the International Airport to the City's Ring Road. The Expressway will began at Hennur on the Outer Ring Road is tolled road. Land Acquisition for the road is completed by December 2007.

Traffic forecast

Air traffic forecasts help to create policy on airport capacity, foretell and give insight to future capacity. It also helps to analyse the economic implications of policy alternatives, and serve as the reference for other government departments.

BIA traffic forecast background

BIA's design and business plans are based on traffic forecasts of SH&E UK, appointed by Government of Karnataka) in the year 2000. Lufthansa Consulting (LHC), appointed by BIA in the year 2002 to revalidate the SH&E traffic forecast. India's economic success in the recent few years has created significant demand for air travel. Further, the Government of India policy to liberalize air traffic has encouraged a number of international airlines to fly into Bengaluru and
new airlines to fly in the domestic sector. With the passenger traffic at the HAL airport increasing from 2.3 million in 2001 to approximately 5 million in 2005, BIAL appointed LHC once more in 2005 to update the traffic forecast and develop planning parameters.

The new traffic forecast
In an extensive report conducted by Lufthansa Consulting, the potential traffic flow from 2005 to 2025 was analyzed. The new forecast shows a significant increase in passenger figures and aircraft movements in the coming years.

In view of the revised traffic forecast, BIAL has redesigned the initial project.

### Total Passengers

<table>
<thead>
<tr>
<th>Scenario</th>
<th>2005</th>
<th>2010</th>
<th>2015</th>
<th>2020</th>
<th>2025</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimistic</td>
<td>4,613,742</td>
<td>10,190,762</td>
<td>13,922,812</td>
<td>18,193,819</td>
<td>23,444,066</td>
</tr>
<tr>
<td>Most Likely</td>
<td>4,470,904</td>
<td>8,504,579</td>
<td>11,369,184</td>
<td>14,536,473</td>
<td>18,441,082</td>
</tr>
<tr>
<td>Conservative</td>
<td>4,328,258</td>
<td>7,144,506</td>
<td>9,777,469</td>
<td>12,284,213</td>
<td>15,377,190</td>
</tr>
</tbody>
</table>

### Total Cargo

<table>
<thead>
<tr>
<th>Scenario</th>
<th>2005</th>
<th>2010</th>
<th>2015</th>
<th>2020</th>
<th>2025</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimistic</td>
<td>124,904</td>
<td>257,263</td>
<td>334,795</td>
<td>426,367</td>
<td>538,844</td>
</tr>
<tr>
<td>Most Likely</td>
<td>122,157</td>
<td>234,017</td>
<td>299,303</td>
<td>375,118</td>
<td>469,179</td>
</tr>
<tr>
<td>Conservative</td>
<td>118,378</td>
<td>198,565</td>
<td>255,033</td>
<td>316,118</td>
<td>391,855</td>
</tr>
</tbody>
</table>

### Total Movements

<table>
<thead>
<tr>
<th>Scenario</th>
<th>2005</th>
<th>2010</th>
<th>2015</th>
<th>2020</th>
<th>2025</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimistic</td>
<td>62,613</td>
<td>125,221</td>
<td>158,100</td>
<td>191,487</td>
<td>228,967</td>
</tr>
<tr>
<td>Most Likely</td>
<td>60,822</td>
<td>106,191</td>
<td>130,464</td>
<td>154,530</td>
<td>182,169</td>
</tr>
<tr>
<td>Conservative</td>
<td>58,553</td>
<td>87,885</td>
<td>112,057</td>
<td>130,467</td>
<td>152,211</td>
</tr>
</tbody>
</table>

### Movement Peak Hours

<table>
<thead>
<tr>
<th>Scenario</th>
<th>2005</th>
<th>2010</th>
<th>2015</th>
<th>2020</th>
<th>2025</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimistic</td>
<td>16</td>
<td>27</td>
<td>30</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>Most Likely</td>
<td>10</td>
<td>16</td>
<td>15</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Conservative</td>
<td>10</td>
<td>19</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>
Below is a diagram of the number of scheduled arrivals over the course of a typical weekdays. The total arrivals can be split into approximately 60% jets, 30% computer aircraft, and 10% general aviation. Since general aviation flights are not scheduled.

![Number of Scheduled Flights Per Hour](image)

The BIA runway system, shows below, is the prime determinant of its capacity to handle aircraft and the primary source of delay. The configuration of runways in use at any particular time is determined primarily by the wind direction and secondarily by noise abatement procedures, traffic demand, and weather conditions in general. The taxiways, which connect terminals and runways, are shown on the diagram as unhighlighted paths. Congestion rarely forms on the taxiways, since crowding there would be alleviated by controllers keeping aircraft at their terminal. Never would a taxiway bottleneck leave a runway out of use.

### 3. Focus of Analysis

Modeling the airport as a landing and takeoff server, and the subsequent analysis shows the best course of action to take in reducing delays. The least expensive sources of increased efficiency are small adjustments to the system as it operates today. For instance, analysis can tell us how much increased landing rates reduce delays. The benefit can then be weighed against the cost of this improvement, in terms of safety, workload, new equipment needed, etc. Analysis can also tell us how much delays will be reduced if the landing or takeoff time for each plane is made less variable. The benefit can also be weighed against the cost of the new equipment and personnel required to reduce the variance of landing times. Another application is in determining the value, in terms of delays, of adding runway capacity by lengthening or adding runways, or by adding equipment so that more aircraft can use the runways in inclement weather.

A more sensitive area of government regulation is in setting user fees for airports. An arriving plane at an airport generates two sources of delay costs. The most obvious is called internal delay, that is, the cost of the passengers time and of aircraft operation while the plane waits to land. If this cost is too high, planes will choose not to come to BIA. Another cost,
called external cost, comes from the added delay. The new arrival adds to other planes that arrive after it and must wait to land. Each planes should be charged both in external and internal cost, the external in the form of landing fees. Small planes, in particular, have low internal cost but create large external costs by delaying large numbers of passengers in big, expensive to operate jets.

**Queueing Models to Characterize Delays in Terminal Areas**

The terminal area is composed of the airport passenger facilities, the runway system and surrounding airspace. The airspace extends up to a 20 to 30 mile radius around the airport and 15 or 20 thousand feet above. Our goal is to model flight operations within this area and the delay incurred by air craft while performing them. Aircraft within the terminal area have one of three goals, to land, to take off or to pass through. They require the services of the airport controllers and use the runway facilities and airspace in order to fulfill the goals. A terminal area system that operates well will fulfill the goals efficiently and without endangering the safety of the aircraft or its passenger.

One can view the terminal area macroscopically as a service facility which provides a complex mix of service which imposes demands on controller, the runway and airspace of the area. As with any service facility, when the demand for the service outstrips the facility’s capacity to provide it, the users of the facility are delayed. If more than one user is delayed, a queue forms. If the disparity in demand and service capacity is large, the queue becomes large and the users are forced to wait for extended period for service. This can incur costs in term of time, money, or opportunity, depending on the type of system and type of delay. A facility with large capacity causes less wait and fewer delay cost.

The terminal area service facilities associated with large urban airports are often overloaded, and the plane wishing to land, take off, or pass through the terminal area are often denied immediate service and forced to wait. In order to improve the performance of the terminal area facilities, that is, increase their capacity and reduce delays, we must understand the behaviour of the system with respect to how delays are generated. Modeling the queueing aspect of the system (the phenomenon of user lining up to wait for service) is the obvious solution. The only feasible solution to improve the terminals area capacity is small scale change in the way the system operates. Modeling and understanding the behaviour of terminal facilities with respect to the generation of delays is essential for making informed decisions about the change and the improvements that are worthwhile to implement.

4. Implementation of Model

**Conceptual Model**

We are primarily interested in analyzing delays in landing aircraft at BIA airport, and especially in analyzing the effects that an improved air traffic control system might have on
delays. A secondary goal is to generate accurate predictions of delays for possible use in the planning and metering aspects of a new air traffic control system. The time frame of interest is intra-day. That is, we are interested in the behaviour of delays over hourly or smaller periods.

Transient analysis is necessary since we are interested in intra-day behaviour and not long term trends. Varying demand and service levels during the day require that the demand and service rates be time dependent. We also wish to evaluate the effect of changes other than the average rate that might affect delays, especially the introduction of an improved air traffic control system. Thus we require flexibility in altering the stochastic process describing service times. We are not considering the effects of multiple servers, nor are we interested in multiple classes of arrivals to the system. The ability to represent sub-components of the terminal server is also not required. A model fulfilling the requirements specified above is diagramed below.

![Diagram](image_url)

The demand consists of aircraft that need to land at the airport. The demand rate can vary during the day. There is no provision for modifying other parameters of the demand process, such as the variance. No distinction is made between aircraft with different characteristics. All arrivals are isomorphic; they are treated the same way by the queue discipline and have exactly the same service characteristics. The queue is a finite capacity FIFO queue. The server has a time varying rate, and the variance of the process governing it can be altered as well.

**Interpolated Model (M/G/1)**

The intuition behind an interpolation model is simple and is based on the relationship between M/M/1, M/D/1, and M/G/1 queues. We might conjecture that by the nature of the different service processes of these queues, the transient result for the M/G/1 queue is bounded above and below by the transient solution to the M/M/1 and M/D/1 queues. In addition, we know that the transient solution to the general system is equivalent to the solution for the M/M/1 models for general systems of order one, therefore one would suspect that the solution moves asymptotically toward the solution for the M/D/1 queue as the order goes to infinity. This leads to a rough method of approximating the M/G/1 queue by interpolation between the results of the M/M/1 and M/D/1 queues.
The M/M/1 queue has a service process that is in some sense as random as possible. If an observer checks the queue at any instant, the probability distribution of the remaining service time is exactly the same. Thus the past events in the queue, such as when the present customer entered service, give no information as to its future behavior. The exponential distribution has a coefficient of variation equal to one. This presents a possible limitation in that distributions with higher coefficients of variation are not well approximated by the exponential. It is possible to postulate distributions with greater variance, but none of them will be independent of the time since the customer has already spent in service.

The M/D/1 queue, on the other hand, has a deterministic service process. The service for all customers always takes the same amount of time. This process has no variance, and its coefficient of variation is zero. The M/G/1 queue has a service process with a service time variance that varies from that of the M/M/1 queue for order one, to deterministic, as in the M/D/1 queue, when the order approaches infinity. These characteristics that allow us to match service time variances by changing the order. The family allows us to specify coefficients of variation between zero and one.

The steady state solution of M/G/1 queueing systems is equivalent to the solution of M/M/1 which approaches asymptotically to the solution of M/D/1 as the other goes to infinity. This is clear by observing the behavior of the steady state waiting solution to the M/G/1 queue.

\[ w_{M/G/1} = \frac{\lambda \sigma^2 + \rho^2}{2(1 - \rho)} \quad \text{where} \quad \rho = \frac{\lambda}{E(s)} \]

By substituting the variance of the M/M/1, M/D/1, and M/G/1 into this formula, we obtain:

General

\[ w_{M/M/1} = \frac{\lambda}{\mu(\mu - \lambda)} \]

Deterministic

\[ w_{M/D/1} = \frac{\lambda}{2\mu(\mu - \lambda)} \]

It is clear that the steady state solution to the M/Er/1 system should always be \(((1 + r)/2r)\) of the steady state solution to the M/M/1 system. The figure below shows different Erlang systems approaching the steady state after a step in demand.
One might also suspect that the transient solution (as opposed to the steady state) to the M/G/1 queue exhibits similar behavior. This is roughly true, but not exactly. In the short run the transient solution of the queue is not simply a fraction of the transient solution of the simpler queues, although it is in the long run (i.e. steady state). The difference lies in the way the different systems approach steady state.

Each system approaches steady state in an exponential manner. The rate at which the part reaches the steady state is determined by the time constant of the system. Unfortunately, queueing systems with more variable service processes have higher time constants and require longer reaching the steady state than those with lower variance. This causes them to react faster to transient changes in the inputs and results in different transient response. The inputs to our systems are always changing, implying that the very short term transient response is more important in determining the transient response than the eventual steady state level.

In trying to approximate the transient response of the M/G/1 queue, we assume that it is always between the transient response to the M/M/1 and M/D/1 queues. This limits our search to a convex interpolation. This assumption is supported by the behaviour of the steady state solutions outlined above. We also assume the transient solution moves monotonically toward the deterministic solution as the order of the Erlang rises, and that it approaches the deterministic solution asymptotically. These assumptions are also supported by the behaviour of the steady state solution. The figure shows the transient response for a number of orders of queues, further supporting this assumption.
One might be able to construct an approximation to the transient solution to the M/G/1 queue by combining the results of the transient solutions to the M/M/1 and M/D/1. First a broad search was made of a variety of functional combinations were tested by producing transient results to simple step input functions using M/D/1, M/M/1, and a variety of orders of M/G/1 queues. Then the solutions to the M/D/1 and M/M/1 queues were used to construct the terms of the functional form in question. The combination of terms was then regressed against the M/G/1 solutions. From this regression the coefficients of the terms are estimated, which completes the functional form, and the residuals could be analyzed to see how well the functional form worked. A number of cycles of solution generation, transformation, regression, and modification of the functional form were performed as the approximation moved closer and closer to the true values.

The best interpolation function found is given below:

\[ w(t)_{M/G/1} = \frac{1}{r} w(t)_{M/M/1} + \frac{r - 1}{r} w(t)_{M/D/1} \]

the Poisson coefficients and half of the deterministic coefficient sums to \((r+1)/2r\), the steady state scaling factor. Thus this interpolation not only works well in approximating the transient solution, but it approximates the steady state solution exactly. By placing a larger weight on the Poisson solution at low orders, it imposes a slower rate of change consistent with the higher time constant. As the order of the general gets larger, it places more weight on the deterministic solution, imposing the faster rate of change inherent in the falling time constant.
In numeric tests with standard input functions the transient solution was corrected to hin 2 or 3%. Below is the transient response to the demand profile shown above of a second order queue. Also shown is the interpolated solution. In heavy black, corresponding to the right hand axis is the vertical distance between the two.

One can see that the interpolated solution drops below the true solution when the input is rising, and rises above when the input is falling. This is a result primarily of the time constant difference. The interpolated model takes longer to ramp up and longer to fall. The differences are not substantial. Overall, the solutions match extremely well, as is demonstrated by the next two figures showing the results for a fifth and tenth order queueing system.
The primary benefit of this approximation is that its order of computation is independent of the order of the system. It requires the calculation of the solution to the deterministic and Poisson queues, and then is able to generate approximate queue solutions of any order with relatively little effort.

**Conclusion**

The transient and steady state delays increase polynomially as the service rate is lowered. The delays generated by systems with low server variance can often be lower than those generated by systems with higher service rates and higher variance. This implies that variance of service times is an important area for improvement in the air traffic system. When the systems are heavily saturated the difference between the delays induced by a Poisson system moves closer to that produced by a deterministic system. Some simulation tests of the accuracy of the Poisson arrivals process were executed for the scheduled arrival profile at BIA. The Poisson was found to be fairly accurate, with some reservations. The simulation distribution of inter-arrival times was found to have greater variance than would be predicted by the Poisson, but this could be attributed to clumping over the day in the schedule. The variance of inter-arrival times was found to be remarkably insensitive to the variance of the lateness distribution used in the simulation, implying that whatever the process generating arrivals is, it is robust to changes in the lateness variance. The distribution of inter-arrival times resembled a low order in general discipline more than exponential, which counts against the Poisson assumption. The variance of the Poisson over a whole day was investigated, and found to be reasonable if one assumes some small percentage of unscheduled aircraft are included in the demand profile. It was concluded that time varying Poisson arrivals is a good assumption given heavy utilization and some amount of unscheduled aircraft.
References


2. Research Plan for Terminal ATC Automation (TATCA)


10. BIA Airport Planning and Design 2012.


Received: March 3, 2014