

Cumulative Diagrams: An Example

Consider Figure 1 in which the functions $\lambda(t)$ and $\mu(t)$ denote, respectively, the demand rate and the service rate (or “capacity”) over time at the runway system of an airport¹. The demand rate is the expected number of demands per unit of time, i.e., the number of requests for service at the runway system per unit of time. Note that not all these requests will be necessarily satisfied due to capacity limitations. Similarly the service rate is the expected number of service requests that can be satisfied per unit of time when the runway system is continually busy, i.e., it is the maximum throughput capacity of the runway system (details in the next lecture of our course). The units we shall use for both the demand rate and the capacity are “aircraft movements (arrivals or departures) per hour”, unless otherwise stated.

Figure 1 approximates a situation that one sees all the time at busy airports. The time axis shows the busy part of a day at the airport, e.g., the origin may correspond to 07:00 local time. To simplify the analysis we assume the demand rate is constant throughout the busy part of the day, i.e., $\lambda(t) = n$ movements per hour for all t . While the normal capacity h (for “high”) is greater than n , a weather event occurs between $t = a$ and $t = b$ (e.g., this could be a period when fog is present during a particular day) which reduces the capacity to l (for “low”) movements per hour. It will be convenient later on to denote with $T = b - a$ the interval of time during which the capacity is low. We would like to explore quantitatively the implications of this temporary loss of capacity on delay levels at the airport.

¹ We shall use throughout the Greek symbols λ and μ to denote the expected “demand rate” and expected “service rate” (or “capacity”); this is consistent with the standard notation used in queuing theory.

We shall make another important simplification in our model at this point. Namely, we shall assume that demands occur at evenly spaced intervals (e.g., if $n = 60$, a demand occurs exactly every 60 seconds) and, similarly, for any value of the service rate, h or l , the service times are constant (e.g., if $l = 30$ and the runway system is continually busy, an aircraft movement is completed exactly every 2 minutes). In queuing theory, this is called a *deterministic* model, by contrast to *probabilistic* models in which either the times between the occurrence of successive demands or the service times (or, more usually, both) are random variables, i.e., may take on different values, each value associated with a certain probability.

Clearly the situation shown in Figure 1 will result in some delays to aircraft movements during the time period between $t = a$ and $t = b$, at the very least. In fact, it is very easy to plot, as a function of time, the number of aircraft (landing or taking off) which will be in the queue, waiting to use the runway system. This is done in Figure 2. There is no queue until $t = a$ because the runway system's capacity is greater than the demand rate. Beginning at $t = a$ the queue builds up at a rate of $n - l$ aircraft per unit of time, so that, by the time b , the queue will build up to a length of $(n-l) \cdot (b-a) = (n-l) \cdot T$ aircraft. After time b , when the runway system's capacity is "high" again, the queue length will decrease at the rate of $h-n$ aircraft per unit of time, i.e., at the rate at which aircraft receive service minus the rate at which new aircraft join the queue. Thus, it will take an amount of time equal to $\frac{(n-l) \cdot T}{(h-n)}$ for the queue to dissipate and get back to zero aircraft. We therefore have:

$$\text{Total amount of time with a queue present} = T + \frac{(n-l) \cdot T}{(h-n)} = T \cdot \frac{(h-l)}{(h-n)} \quad (1)$$

The area under the triangle in Figure 2 represents the total amount of time that aircraft will spend waiting in the queue, i.e., the total delay time suffered by all delayed aircraft due to the reduction in the capacity from $t = a$ to $t = b$. We have:

$$\text{Total delay time} = \frac{1}{2} \cdot (n-l) \cdot T \cdot \frac{(h-l)}{(h-n)} T = \frac{1}{2} \cdot T^2 \cdot \frac{(n-l)(h-l)}{(h-n)} \quad (2)$$

Note that the total delay time increases with the *square* of the duration of the low capacity interval. From (1) we know that the number of aircraft which demand service during the period when a queue exists is equal to $n \cdot \frac{(h-l)}{(h-n)} \cdot T$, the demand rate multiplied by the duration of the queue. This is the number of aircraft which will suffer some delay. We can then compute another quantity of interest by dividing the total delay time, given by (2), by the number of delayed aircraft to obtain

$$\text{Expected delay per delayed aircraft} = \frac{1}{2} \cdot T \cdot \frac{(n-l)}{n} = \frac{1}{2} \cdot T \cdot \left(1 - \frac{l}{n}\right) \quad (3)$$

Please note the meaning of (3): *given* that an aircraft was delayed, this is the delay that this aircraft suffers “on average”. It is remarkable that this expected delay is a function only of the “low” capacity and of the demand rate and independent of the “high capacity, h .”

An informative way to display the behavior of this queue is through the device of the *cumulative flow diagrams* (or cumulative flow *plots*) shown in Figure 3. In the context of the airport runway example, cumulative flow diagrams typically show (a) the total (“cumulative”) number of requests for service by (arriving or departing) aircraft that have been made between $t = 0$ and the current time t , and (b) the total number of these requests that have been admitted for service up to the current time t . We shall call the former the *cumulative demand diagram*² and the latter the *cumulative admissions to service diagram*. Obviously, the difference between the cumulative demands and the cumulative admissions to service at any time t is the number of aircraft movements

² In fact, the cumulative *demand* diagram simply plots the value of $\int_0^t \lambda(x) \cdot dx$ for all values of t .

queued for admission to the runway system. Figure 3 shows these two cumulative diagrams for our example. Note that the vertical axis, i , of Figure 3 maintains a count of the cumulative number of demands and of admissions to service as they occur. For our example, the demand and admissions-to service cumulative flow diagrams coincide up to $t = a$, because demands are admitted as soon as they show up at the runway system, due to the high capacity h . At $t = a$, however, the number admitted begins lagging behind the demand, as it increases only with a slope of l aircraft per unit of time. This lasts until $t = b$ when the number admitted starts increasing at the rate of h per unit of time, until it eventually “catches up” with the demand curve at time $\frac{(n-l) \cdot T}{(h-n)}$ later. Thereafter the two cumulative flow diagrams coincide again and increase at the rate of n per unit of time, as capacity is higher than the demand rate.

Note that the cumulative flow of admitted-for-service aircraft can never exceed, by definition, the cumulative flow of the demands. At best, it can always be equal to the cumulative flow of demands, in cases where no delays occur due to the capacity being greater than the demand rate for all t . The (positive) vertical distance and the (positive) horizontal distance between the cumulative flow diagrams in Figure 3 both have very real physical interpretations. The *vertical distance* at any time t gives the *number of aircraft in queue* at that time, i.e., it shows the difference between the number of aircraft that have demanded service up to time t and the number of aircraft that have been admitted for service. If we plot that vertical distance as a function of time for Figure 3, we shall obtain again Figure 2. Similarly, the *horizontal distance* gives the *delay suffered by the i -th demand*, i.e., the time that elapses between the instant when the i -th demand requests service at the runway system and the time when that demand is admitted for service assuming that demands are admitted for service in first-come, first-served (FCFS) order³. Note, then, that the total “area” between the two cumulative flow diagrams, is equal to the total delay time lost in units of aircraft time (e.g., aircraft-hours). Thus, this area is given by expression (2) – the reader may wish to confirm this, as an exercise.

³ If a FCFS priority discipline is not in effect, then the horizontal distance is the time that elapses between the i -th demand by an aircraft for service and the i -th admission of a (possibly different) aircraft for service.

The longest delay of all will be suffered by that aircraft which *demand*s service at the instant when the horizontal distance between the cumulative flow diagrams is greatest, assuming a FCFS priority discipline. From Figure 3 it is clear that this is the aircraft which will be *admitted* for service at $t = b$, because the horizontal distance between the two cumulative flow diagrams begins decreasing immediately thereafter. It can be seen that this is the $(n \cdot a + l \cdot T)$ -th aircraft to demand service (and to be admitted for service). This aircraft demands service at the time, t , that satisfies $n \cdot t = n \cdot a + l \cdot T$, i.e., at $t = a + \frac{l \cdot T}{n}$. Since it is already known that this aircraft will be admitted for service at exactly $t = b$, the delay this aircraft will suffer is given by

$$\text{Longest delay suffered by any aircraft} = b - \left[a + \frac{l \cdot T}{n} \right] = T \cdot \left(1 - \frac{l}{n} \right) \quad (4).$$

From Figure 3, taking advantage of the observations in the last paragraph, we can now prepare Figure 4 that shows, for all values of t , the amount of delay that a demand *requesting service* at time t will suffer, under the FCFS assumption.

It is instructive at this point to assign some realistic values to the various parameters of our example and look at the implications. We choose a set of values that may be typical of a major European airport. We let the time units be hours, and set $t = 0$ to correspond to 07:00 local time, $a = 1.0$ and $b = 4.0$, so that the duration, T , of the low capacity period is 3 hours. We also let $n = 60$ aircraft movements per hour, $h = 70$ aircraft movements per hour and $l = 35$ aircraft movements per hour, i.e., the poor weather conditions reduce the capacity to one-half its normal value –not an unusual phenomenon in practice.

From (1) we then find that a queue will be present for a period of 10.5 hours beginning at 08:00 local time, i.e., the after-effects of the poor weather persist for 7.5 hours after the weather event ends at 11:00! With 60 movements scheduled per hour, a

total of 630 movements will suffer some delay. The peak length of the queue is 75 movements (Figure 2 or Figure 3) and occurs at 11:00 local time, but the peak delay time is suffered by the movement/demand scheduled for $t = 2.75$ or for 9:45 local time. This movement will suffer a delay equal to 1.25 hours, or 75 minutes (from (4)) and will actually take place at 11:00, exactly the time of the end of the weather event. The total amount of delay time incurred during the day, from (2), is equal to 393.75 aircraft hours! The economic cost of this delay depends on such factors as the mix of aircraft at this airport, whether the delays will be absorbed while the aircraft are airborne or on the ground, and several other parameters. Just so we can indicate the order of magnitude of that cost, we use here \$1,500 as the direct cost to airlines of one aircraft-hour, a very reasonable amount for a major airport. This gives approximately \$600,000 as the total cost of delay due to the 3-hour weather event, *not* including the cost of delay time to the passengers! Average delay per movement for the 630 movements delayed is 0.625 hours, from (2), or 37.5 minutes, for a cost of approximately \$940 per aircraft (at the assumed \$1,500 per aircraft-hour).

[End of example.] #

Our example above demonstrated, among other things, the preparation and interpretation of cumulative flow diagrams. It is straightforward to extend this approach to the more general case in which (a) both the demand rate and the capacity vary over time and (b) there are several, not just one, instances during the day when the demand is higher than the capacity. Figure 5 shows a typical example of a cumulative flow diagram for this more general case. The only differences from Figure 3 are that (a) the “demand” and the “admitted for service” cumulative flow diagrams will undergo several slope changes and (b) there may be several instances during the day when there is a backlog of demands waiting to be served, i.e., when delays will occur. We also note that the cumulative diagrams need not be piecewise linear functions, as in Figure 5. They can have other functional forms, depending on the functional forms used to describe the demand rate, $\lambda(t)$, and capacity, $\mu(t)$, of the airport during the time interval of interest.

The deterministic model and approach described here have one fundamental deficiency: they capture *only* that part of airport delay which is due to the demand rate exceeding the airport's capacity. *They do not, however, account for the delays which are due to probabilistic fluctuations ("uncertainty")*. For instance, in the numerical example we looked at, the deterministic model states that there would have been *no delay* at the runway system if the airport could operate all day with a high capacity h of 70 movements per hour, while the demand n were equal to 60 movements per hour. But, in practice, one might see some very significant delays under these conditions, if the demands occurred sufficiently randomly in time to create periods of traffic surges and/or if the service times on the runway system exhibited significant variability. Delays of this type are captured by *probabilistic* queuing models and analysis which will be discussed later in the course.