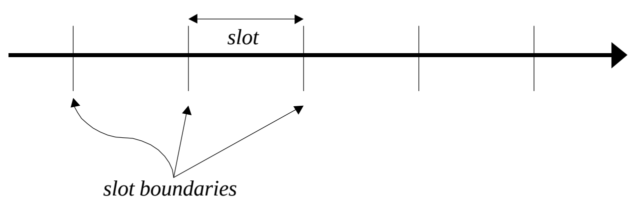
Analysis

of

Discrete Time Queues

(Section 4.6)

Time axis divided into slots



- Arrivals can only occur at slot boundaries
- Service to a job can only start at a slot boundary
- Service duration is always an integral multiple of the slot duration

• Assume that the number of jobs that arrive in successive slots are independent, identically distributed (*i.i.d.*) random variables

Geo Arrival Process: *Geometric (or Bernoulli) Arrival Process*

Only one job can arrive in a slot with probability l and that no jobs arrive in a slot with probability $1-l$, $(0<l<1)$

The inter-arrival time I is geometrically distributed with with mean $1/l$ and with probability-

$P\{I = k \text{ slots}\} = l(1-l)^{-k}$ for $k=1,2,.......$

$$\Lambda(z) = 1 - l(1-z)$$ Generating Function of the number of arrivals in a slot

 3

$Geo^{[X]}$ Arrival Process *Batch Geometric Arrival Process*

L =number of arrivals in a slot (iid random variables)

$l(k) = P\{L = k \text{ arrivals in a slot}\}$ $k=0,1,2,\ldots\ldots\ldots$

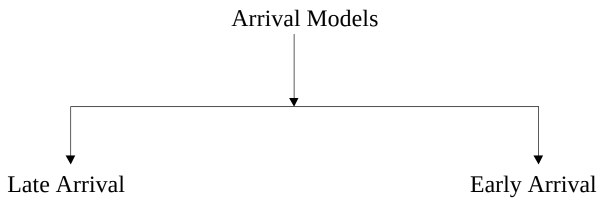
$$\Lambda(z) = \sum_{k=0}^{\infty} l(k) z^k$$ Generating Function of the number of arrivals in a slot

The following definitions of means and *factorial moments* would be useful later

$$l = E\{\Lambda\} = \Lambda^{(1)}(1)$$
$$l^{(2)} = E\{\Lambda(\Lambda-1)\} = \Lambda^{(2)}(1)$$
$$l^{(i)} = E\{\Lambda(\Lambda-1)......(\Lambda-i+1)\} = \Lambda^{(i)}(1) \qquad i = 3,4,........$$

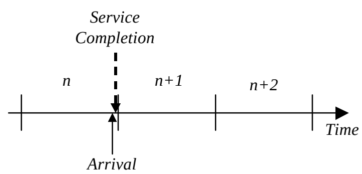
 4

Choice of Arrival Models



- Queue size measured immediately after the slot boundaries
- Waiting time in queue will differ for the late and early models

Late Arrival Model

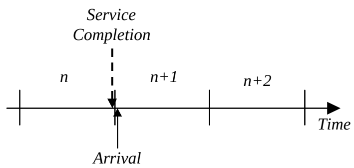


Arrivals come late in the slot, i.e. just before the slot boundary and before the service completions due to occur at the end of that slot

- Number left behind in the queue as seen by a departure will include the arrivals as shown above
- Time spent waiting in queue = Number of slots spent waiting for service not including the slot in which the job arrives

For example, waiting time is zero if a job arrives at the end of the n^{th} slot and starts service from the beginning of the $(n+1)^{th}$ slot, i.e. the next slot after its arrival.

Early Arrival Model



Arrivals come early in the slot, i.e. just after the slot boundary.

Note that service completions occur just before the slot boundary as in the *late arrival* model

• Number left behind in the queue as seen by a departure will not include the arrivals as shown above

• Time spent waiting in queue = Number of slots spent waiting for service including the slot in which the job arrives

For example, waiting time is zero if a job arrives at the beginning of the $(n+1)^{th}$ slot and starts service from that slot itself.

Important points to note -

• Queue size at service completion in the early arrival model will always be lower than in the corresponding late arrival model.

Lower by the number of jobs which actually arrive at that slot boundary

• The waiting time in queue will be the same regardless of the actual model (early/late) being used

Service Times

- Multiples of slot durations - starts and ends at the slot boundaries
- General IID Distribution for the number of slots (say *random variable X)* required to complete service to a job.
- Probability Distribution of Service Times (in terms of the number of slots required for service

Probability Distribution $b(k) = P\{X = k\}$ for $k=1,2,\ldots\ldots$

Generating Function $B(z) = \sum_{k=1}^{\infty} b(k) z^k$

Moments

$$\begin{cases} b = E\{X\} = B^{(1)}(1) \\ b^{(2)} = E\{X^2\} = B^{(2)}(1) + B^{(1)}(1) \\ b^{(i)} = E\{X^i\} \qquad i = 3, 4, \ldots\ldots \end{cases}$$

The Geo/G/1 Queue *(discrete time version of the M/G/1 queue)*

Late Arrival Model

Separate analysis required for the late arrival and the early arrival model

Early Arrival Model

Stability requires that the traffic offered $r = lb < 1$

The Geo/G/1 Queue (Late Arrival Model)

n_i = number of jobs in the queue immediately after the service completion of the i^{th} job

a_i = number of jobs arriving during the service time of the i^{th} job

The random variables a_i i=1,2,...... are independent and identically distributed random variables with the generating function A(z) and mean r.

Homogenous, Discrete-Time Markov Chain

$$\begin{aligned} n_{i+1} &= a_{i+1} & n_i &= 0 \\ &= n_i + a_{i+1} - 1 & n_i &\geq 1 \end{aligned} \qquad (4.91)$$

At equilibrium, the steady-state distribution p_k of this Markov Chain may be found as -

$$p_k = \lim_{i\to\infty} P\{n_i = k\} \qquad k = 0,1,2,.......$$

and
$$P(z) = \sum_{k=0}^{\infty} p_k z^k$$

Show that, using (4.91) under equilibrium conditions gives -

$$P(z) = \frac{(1-r)(1-z)A(z)}{A(z)-z} \qquad \text{with} \quad \begin{cases} p_0 = 1 - r \\ r = lb \end{cases}$$

Show that $A(z) = B(1 - l + lz)$

leading to $$P(z) = \frac{(1-r)(1-z)B(1-l+lz)}{B(1-l+lz)-z} \qquad (4.97)$$

Derived for the instants of service completions of jobs

Also holds for

- Arrival instants
- Just after arbitrary slot boundary
- Arbitrary time instant on the continuous time axis

BASTA/GASTA: Bernoulli/Geometric Arrivals See Time Averages holds as the discrete time version of PASTA

Using *P(z)* from (4.97), show that

Mean number in system $$N = \frac{l^2 b^{(2)} - l r}{2(1-r)} + r \qquad (4.98)$$

The mean time *W* (in number of slots) spent by a job in the system may be found from (4.98) using Little's Result $N = lW$

We can then use $W_q = W - b$ and $N_q = lW_q$ to get the performance measures W_q and N_q

The distribution $G_W(z)$ (i.e. the generating function) of the number of slots for which a job stays in the system may also be obtained for a FCFS Geo/G/1 queue.

Using an approach similar to that followed for the M/G/1 queue, show that for the FCFS Geo/G/1 queue, we will have -

$$P(z) = G_W(1 - l + lz)$$

which leads to $$G_W(z) = \frac{(1-r)(1-z)B(z)}{(1-z) - l(1-B(z))} \qquad (4.102)$$

and $$G_{Wq}(z) = \frac{(1-r)(1-z)}{(1-z) - l(1-B(z))} \qquad (4.104)$$

following usual independence argument between service time and time spent waiting in queue

 15

The Geo/G/1 Queue (Early Arrival Model)

n_i = number of jobs in the queue after the service completion of the i^{th} job and before the next possible arrival point

a_i = number of jobs arriving during the service time of the i^{th} job

$\tilde{a}_i$ = number of jobs arriving in the service time of the the i^{th} job *minus* one slot

Homogenous, Discrete-Time Markov Chain

$$n_{i+1} = \begin{cases} \tilde{a}_{i+1} & n_i = 0 \\ n_i + a_{i+1} - 1 & n_i \geq 1 \end{cases} \qquad (4.106)$$

 16

Generating Function *A(z)* of *a* obtained as before

$$A(z) = B(1 - l + lz)$$

Show that

$$\tilde{A}(z) = \frac{B(1 - l + lz)}{1 - l + lz}$$

From (4.106), show that

$$P(z) = \frac{(1 - r)[A(z) - z\tilde{A}(z)]}{(1 - l)[A(z) - z]}$$

and hence

$$P(z) = \frac{(1 - r)(1 - z)B(1 - l + lz)}{(1 - l + lz)[B(1 - l + lz) - z]}$$

Show that, in this case

$$P(z) = \frac{1}{(1 - l + lz)} G_W(1 - l + lz)$$

which leads to

$$G_W(z) = \frac{(1 - r)(1 - z)B(z)}{(1 - z) - l(1 - B(z))}$$

Note that even though *P(z)* is different for the late/early arrival models, $G_W(z)$ is the same as claimed earlier

The Geo$^{[X]}$/G/1 Queue

Summary of results given here. Please see Sec. 4.6.2 for details

L(z) Generating Function of the number of jobs arriving in a slot

B(z) Generating Function of the number of slot required for service by a job

The Geo$^{[X]}$/G/1 Queue (Late Arrival Model)

Generating Fn. of the number in the system just after a service completion

$$P(z) = \frac{(1 - r)[1 - \Lambda(z)]B(\Lambda(z))}{l\,[B(\Lambda(z)) - z]}$$

Generating Fn. of the number in the system just after an arbitrary slot boundary

$$P^{*}(z) = \frac{(1 - r)(1 - z)B(\Lambda(z))}{B(\Lambda(z)) - z}$$

Generating Fn. of the number of slots spent by a job waiting in the queue, prior to service

$$G_{Wq}(z) = \frac{(1 - r)(1 - z)[1 - \Lambda(B(z))]}{l\,[\Lambda(B(z)) - z][1 - B(z)]}$$

Mean number in system just after an arbitrary slot boundary (or at any arbitrary time instant)

$$N = \frac{l^2 b^{(2)} - l\, r + l^{(2)} b}{2(1 - r)} + r$$

Mean queueing delay (number of slots) encountered by a job

$$W_q = \frac{l^2 b^{(2)} - l\, r + l^{(2)} b}{2(1 - r)}$$

The Geo$^{[X]}$/G/1 Queue (Early Arrival Model)

Generating Fn. of the number in the system just after a service completion

$$P(z) = \frac{(1 - r)[1 - \Lambda(z)]B(\Lambda(z))}{l\Lambda(z)[B(\Lambda(z)) - z]}$$

Generating Fn. of the number in the system just after an arbitrary slot boundary

$$P^*(z) = \frac{(1 - r)(1 - z)B(\Lambda(z))}{B(\Lambda(z)) - z}$$

Generating Fn. of the number of slots spent by a job waiting in the queue, prior to service

$$G_{Wq}(z) = \frac{(1 - r)(1 - z)[1 - \Lambda(B(z))]}{l[\Lambda(B(z)) - z][1 - B(z)]}$$

- *Note that $P^*(z)$ and $G_{W\,q}(z)$ are the same as for the Late Arrival Model*
- *The queue performance parameters N, N_q, W, W_q will be the same as for the late arrival model*