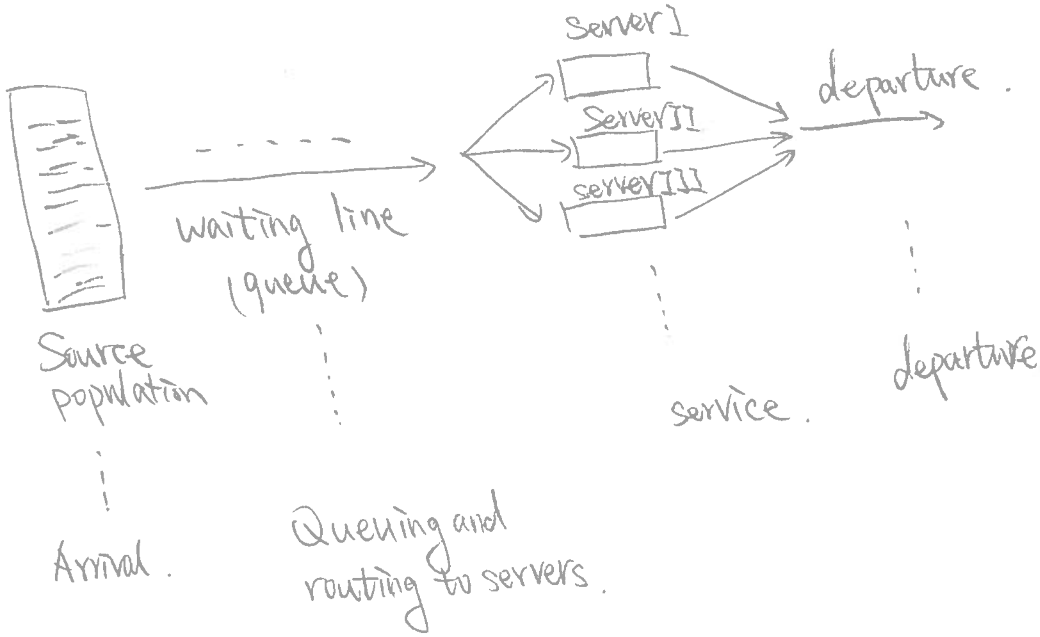
Chapter 7. Queuing System.

A queuing system is a facility consisting of one or several servers designed to perform certain tasks or certain jobs and a queue of jobs waiting to be processed.



① Arrival:

Arrival rate: $\lambda_A = \frac{E_A(t)}{t}$.

The expected time between arrivals: $u_A = \frac{1}{\lambda_A}$.

② Queuing and routing to servers:

$X_w(t)$: # of people waiting at time t.

③ Service:

The average service time $u_s = \frac{1}{\lambda_s} \rightarrow$ service rate.

utilization $r = \frac{\lambda_A}{\lambda_s} = \frac{u_s}{u_A}$, arrival-to-service ratio.

$X_s(t)$ = number of jobs receiving service at time t.

$X_w(t)$ = number of jobs in a queue at time t.

$X(t) = X_s(t) + X_w(t)$. = the total number of jobs in the system at time t

S_k = service time of the k-th job.

W_k = waiting time of the k-th job

$R_k = S_k + W_k$. response time

7.3. Bernoulli single-server queuing process.

- discrete-time.
- service time and interarrival time are independent.
- Number of arrivals counted with a Binomial counting process.

$\Downarrow$

- service takes at least one frame Δ
- at most one arrival per frame Δ

$p_A = \lambda_A \cdot \Delta$ → probability of arrival in any frame.

$p_s = \lambda_s \cdot \Delta$ → Probability of departure in any frame.

Now we compute all transition probabilities:

$$P_{01} = P\{\text{new arrival}\} = p_A$$

$$P_{00} = P\{\text{no arrival}\} = 1 - p_A.$$

And for all $i \geq 1$.

$$P_{i,i-1} = P\{\text{no arrivals} \wedge \text{one departure}\} = (1-p_A)\cdot p_s.$$

$$P_{i,i} = P\{\text{no arrivals} \wedge \text{no departure}\} + P\{\text{one arrival} \wedge \text{one departure}\}$$

$$= (1-p_A)\cdot(1-p_s) + p_A \cdot p_s$$

$$P_{i,i+1} = P\{\text{no departure} \wedge \text{one arrival}\} = p_A \cdot (1-p_s)$$

All the other transition probabilities equal 0, because the number of jobs cannot change by more than one during any single frame.

$$
\begin{array}{c c}
 & \begin{array}{ccc} 0 & 1 & \cdots \end{array} \\
P = \begin{array}{c} 0 \\ 1 \\ 2 \\ \vdots \\ n \end{array} & \left[\begin{array}{cccc} 1-P_A & P_A & & \\ P_S\cdot(1-P_A) & (1-P_A)\cdot(1-P_S)+P_A\cdot P_S, & P_A\cdot(1-P_S). & \\ & & \cdots & \\ & & & \ddots \end{array} \right]
\end{array}
$$

BSSQP is not regular, but it will have a steady state if $\lambda_S > \lambda_A$

In BSSQP, since the server processes at most 1 job at a time,

Therefore:

$$E(X_S) = P\{\text{server is busy}\} = 1 - P\{\text{server is idle}\}$$

Therefore,

$$E(X_w) = E(X) - E(X_S).$$