4.15 푸아송 Arrival이 아닐 때의 Erlang-B 일반화

“대기열에서 푸아송이 아닌 일반적인 arrival 프로세스에서의 Erlang-B 공식을 이용하여 손실을 근사하여 구한다.”

4.15.1 M/M/S/S 대기열의 트래픽 유형

- Peakedness parameter(z) : 현재 대기열이 가상의 G/G/∞ 대기열이라고 가정하고, 이 상황에서의 트래픽으로 정의한다. 쉽게 말해서, 대기열에 있는 트래픽이 균일하게 오냐 안 오냐 그 정도를 나타낸다.

$$z = \frac{Var[n]}{E[n]}$$

여기서 n은 G/G/∞ 대기열에 있는 요청의 수다.

- $z < 1$: 요청이 발생하는 시간 간격이 일정
  regular arrival이라고 하며, 트래픽이 Smoothed 하다고 표현할 수 있음.

- $z = 1$: Poisson arrival

- $z > 1$: 요청이 발생하는 빈도의 변동성이 큼.
  유명 가수가 SNS에 글을 올리면 순식간에 조회수가 올라가듯이, 특정 짧은 기간 동안 요청이 몰려있는 경우를 나타내며, 트래픽이 peaked 하다고 표현할 수 있음.

돌아와서,

- M/M/S/S 대기열에서의 peakedness parameter를 구하기 위해 대기 중인 요청 수 n의 기댓값과 분산을 구한다.
- 요청 수에 따라서 carried rate가 달라지기 때문에, carried traffic 프로세스의 모델을 유추할 수 있다.

$$E[n] = A_C = \rho(1 - P_B), \qquad Var[n] = V_C = A_C - \rho P_B(S - A_C)$$

- 위의 식을 이용해서 peakedness parameter를 계산한다.

$$z = \frac{V_C}{A_C} = \frac{A_C - \rho P_B(S - A_C)}{A_C} = 1 - \rho P_B\left(\frac{S}{A_C} - 1\right)$$

- 여기서 $z < 1$이므로 carried traffic은 non-Poisson이며 smoothed이다.

- 반대로 거절되는 요청의 arrival process는 $n=S$일 때의 arrival process를 이용해서 구할 수 있다. (A는 기댓값, V는 분산이다.)

$$A_B = \rho P_B, \qquad V_B = A_B\left(1 - A_B + \frac{\rho}{S - \rho + A_B + 1}\right)$$

- 위의 식을 이용해서 peakedness parameter를 계산한다.

$$z = \frac{A_B}{V_B} = 1 - A_B + \frac{\rho}{S - \rho + A_B + 1}$$

- $z > 1$이므로 refused traffic은 non-Poisson이며 peaked이다.

● 부록 : 공식 유도

$$\begin{aligned}
E[n] &= \sum_{i=0}^{S} i P_i \\
&= \sum_{i=0}^{S}\left(i \times \frac{\rho^i}{i!} P_0\right) \\
&= \rho \sum_{i=1}^{S}\left(\frac{\rho^{i-1}}{(i-1)!} P_0\right) \\
&= \rho \sum_{i=1}^{S} P_{i-1} \\
&= \rho(1 - P_B)
\end{aligned}$$

$$\begin{aligned}
E[n^2] &= \sum_{i=0}^{S} i^2 P_i \\
&= \sum_{i=0}^{S}\left(i^2 \times \frac{\rho^i}{i!} P_0\right) \\
&= \rho \sum_{i=1}^{S}\left(i \frac{\rho^{i-1}}{(i-1)!} P_0\right) \\
&= \rho \sum_{i=1}^{S}\left((i-1) \frac{\rho^{i-1}}{(i-1)!} P_0\right) + \rho \sum_{i=1}^{S}\left(\frac{\rho^{i-1}}{(i-1)!} P_0\right) \\
&= \rho^2 \sum_{i=2}^{S}\left(\frac{\rho^{i-2}}{(i-2)!} P_0\right) + \rho \sum_{i=1}^{S} P_{i-1} \\
&= \rho^2 \sum_{i=2}^{S} P_{i-2} + \rho(1 - P_B) \\
&= \rho^2(1 - P_B - P_{S-1}) + \rho(1 - P_B) \\
&= \rho^2\left(1 - P_B - \frac{S}{\rho} P_B\right) + \rho(1 - P_B) \\
&= \rho\{\rho - (S + \rho) P_B\} + \rho(1 - P_B)
\end{aligned}$$

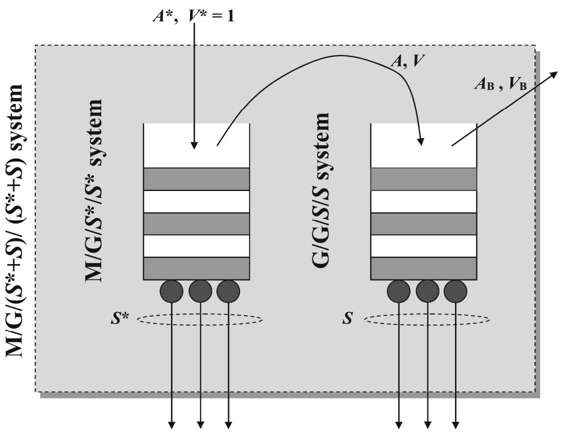
$$\begin{aligned}
Var[n] &= E[n^2] - (E[n])^2 \\
&= \rho\{\rho - (S + \rho) P_B\} + \rho(1 - P_B) - \rho^2(1 - P_B)^2 \\
&= \rho(1 - P_B) + \rho\left\{\rho - (S + \rho) P_B - \rho + \rho P_B^2\right\} \\
&= E[n] + \rho P_B\{-(S + \rho) + \rho P_B\} \\
&= E[n] - \rho P_B\{S + \rho(1 - P_B)\} \\
&= E[n] - \rho P_B\{S + E[n]\}
\end{aligned}$$

4.15.2 Non-Poisson Arrivals를 가진 대기열에서의 Blocking Probability

- 평균 A, 분산 V, peakedness factor $z = V/A$를 가진 트래픽과 G/G/S/S 대기열이 있다고 가정하자. 이때의 blocking probability를 구하고자 한다. Wilkinson 방식 또는 Fredericks 방식 등으로 구할 수 있다.

4.15.2.1 Equivalent Random Theory (ERT) a.k.a. Wilkinson Method

- $V > A$의 특징을 가진 G/G/S/S 트래픽이 있다고 가정하자. 이를 구하기 위해서 우선, 아래의 가상 대기열을 구상해 본다.



- 이는 ERT-Wilkinson 방식을 따른 대기열 시스템이다. $M/G/S^*/S^*$라는 가상의 대기열에서 거절되는 요청들이 G/G/S/S 대기열로 들어간다고 하자.
  트래픽에서 거절되는 요청들은 $z > 1$의 특성을 가졌으므로, 위의 이미지처럼 시스템을 근사할 수 있다.

- 알 수 없는 input intensity와 서버 수를 가진 Poisson Arrival Process를 가진 대기열에서 거절되는 요청의 intensity와 분산이 A와 V로 근사한다.

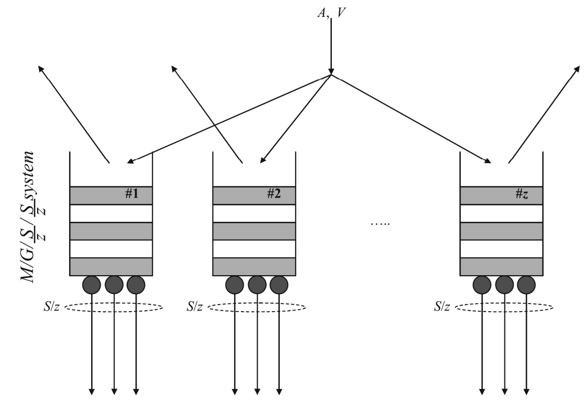
- 4.15.1에서 구한 식들을 이용해서 A와 V를 안다면, S^*와 A^*도 알 수 있다.

$$A^* = V + 3z(z-1),\quad P_B{}^*(S^*, A^*) = \frac{A^{*S^*} e^{-A^*}}{\Gamma(S^*+1, A^*)},\quad P_B = \frac{\lambda_{block}}{\lambda_{avg}} = \frac{A_B}{A} = \frac{A^* P_B{}^*(S+S^*, A^*)}{A}$$

4.15.2.2 Fredericks Method

- 이 방식은 $z > 1$, $z < 1$인 경우 둘 다 적용할 수 있다.

- ERT와 다르게, 가상의 대기열 시스템에서 A^*와 V^*가 없고, 복수의 대기열 시스템이 요청들을 나눠 받는다.

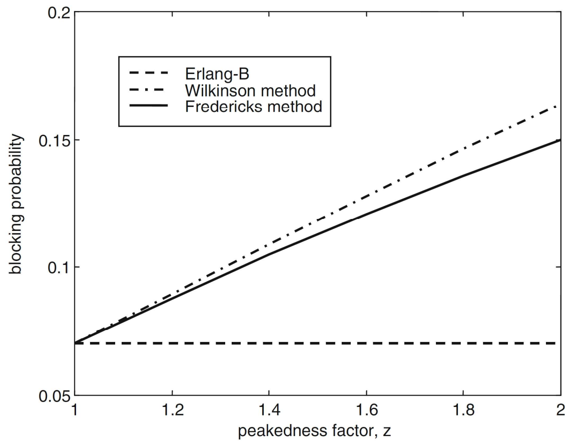


- G/G/S/S 대기열의 blocking probability는 다음과 같이 근사할 수 있다.

$$P_B = P_B\left(\frac{S}{z}, \frac{A}{z}\right)$$

- V와 A의 대소관계에 따라서 근사로 구한 P_B와 푸아송 프로세스가 있는 $P_{B,\,Poisson}$의 대소관계도 달라진다.

1. $V > A \rightarrow P_B(S/z, A/z) > P_{B,\,Poisson}(S, A)$
2. $V \leq A \rightarrow P_B(S/z, A/z) \leq P_{B,\,Poisson}(S, A)$



접근법에 따른 z vs blocking probability