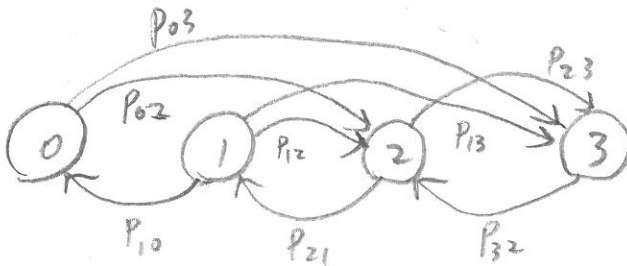


①

with the no-buffering assumption and $m=3$ and $\lambda=1.5$ pkt/sec

 P_1 

$$\begin{cases} P_0(P_{02} + P_{03}) = P_1(P_{10}) \\ P_1(P_{12} + P_{13} + P_{10}) = P_2(P_{21}) \\ P_2(P_{23} + P_{21}) = P_3(P_{32}) + P_1(P_{12}) + P_0(P_{02}) \\ P_3(P_{32}) = P_0(P_{03}) + P_1(P_{13}) + P_2(P_{23}) \\ P_0 + P_1 + P_2 + P_3 = 1 \end{cases}$$

$$\Rightarrow \begin{cases} P_0 + P_1 + P_2 + P_3 = 1 \\ P_3(P_{32}) = P_0(P_{03}) + P_1(P_{13}) + P_2(P_{23}) \\ P_2(P_{21}) = P_0(P_{02} + P_{03}) + P_1(P_{12} + P_{13}) \\ P_1(P_{10}) = P_0(P_{02} + P_{03}) \end{cases}$$

$$P_{10} = Q_a(0,1) Q_r(1,1) = (1 - g_a)^2 g_r$$

$$P_{21} = Q_a(0,2) Q_r(1,2) = (1 - g_a) \cdot 2 \cdot (1 - g_r) g_r$$

$$P_{32} = Q_a(0,3) Q_r(1,3) = 3 \cdot (1 - g_r)^2 g_r$$

$$P_{02} = Q_a(2,0) = 3(1 - g_a) g_a^2$$

$$P_{03} = Q_a(3,0) = g_a^3$$

$$P_{12} = Q_a(1,1)(1 - Q_r(0,1)) = 2(1 - g_a) g_a g_r$$

$$P_{13} = Q_a(2,1) = g_a^2$$

$$P_{23} = Q_a(1,2)(1 - Q_r(0,2)) = g_a(1 - (1 - g_r)^2)$$

1.1 $g_r = 0.2$, $g_a = 1 - e^{-\lambda/3} \approx 0.393$

$$\Rightarrow \begin{cases} P_{10} = 0.073, & P_{02} = 0.281, & P_{13} = 0.154 \\ P_{21} = 0.194, & P_{03} = 0.06, & P_{23} = 0.141 \\ P_{32} = 0.384, & P_{12} = 0.095 \end{cases}$$

$$\Rightarrow P_0 \approx 0.054, P_1 \approx 0.282, P_2 \approx 0.448, P_3 \approx 0.263$$

$$\Rightarrow N = 1.877 \Rightarrow T = 1.251 \text{ seconds}^*$$

1.2 $g_r = 0.8$, $g_a = 0.393$

$$\Rightarrow \begin{cases} P_{10} = 0.294, & P_{02} = 0.281, & P_{13} = 0.154 \\ P_{21} = 0.194, & P_{03} = 0.06, & P_{23} = 0.377 \\ P_{32} = 0.096, & P_{12} = 0.381 \end{cases}$$

$$\Rightarrow P_0 \approx 0.034, P_1 \approx 0.039, P_2 \approx 0.168, P_3 \approx 0.745$$

$$\Rightarrow N = 2.61 \Rightarrow T = 1.74 \text{ seconds}^*$$

2.1

The structure of the introduction section is simple here:

overview \rightarrow motivating example \rightarrow challenges \rightarrow how the paper contributes \rightarrow outline.

Each paragraph serves one purpose above.

The challenge is "with platform constraints and application requirements, it is hard to meet the requirement while being efficient."

P2 (2.2) The contributions are summarized in the fourth paragraph:

- ① a holistic analysis for real-time reliable IoT edge computing;
- ② the ARREC architecture;
- ③ an implementation of ARREC and its empirical performance evaluation.

We may think in this way: an architecture needs to be based on some theoretical analysis that addresses the challenges we are to tackle; the analysis offers guidelines for us to design an architecture. An architecture needs to be implemented in a way that demonstrates the promised performance; the implementation and evaluation offer evidences for us to see the architecture's capability in dealing with the challenges mentioned.

(3.2.1) Here's a reason for the misleading result:

The data in `e2e.delay.misleading` contains some latency results obtained while the broker was still busying accepting new connections. Those results have larger latency. You may verify this by examining the first few lines in `e2e.delay.misleading`. Now, with a longer sampling duration we essentially average out these abnormal results, and therefore the averaged end-to-end latency is shorter.

An important lesson that we may learn here:

A software system often needs some time to "warm up", and thus we should avoid mixing the measurement obtained in the warm-up phase with that obtained in the later phase, to prevent some misinterpretation of the result.