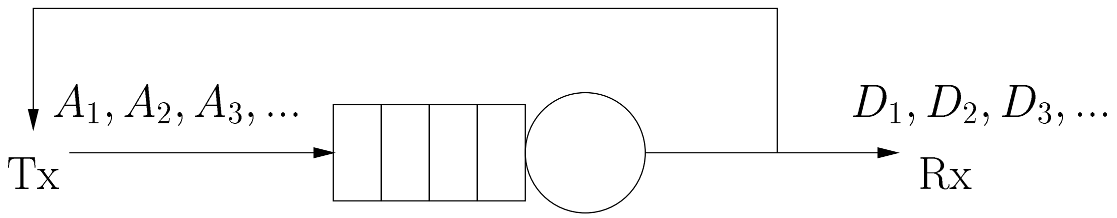
Basic ideas



- A channel with a single-server queue ($\cdot/G/1$, $\cdot/M/1$)

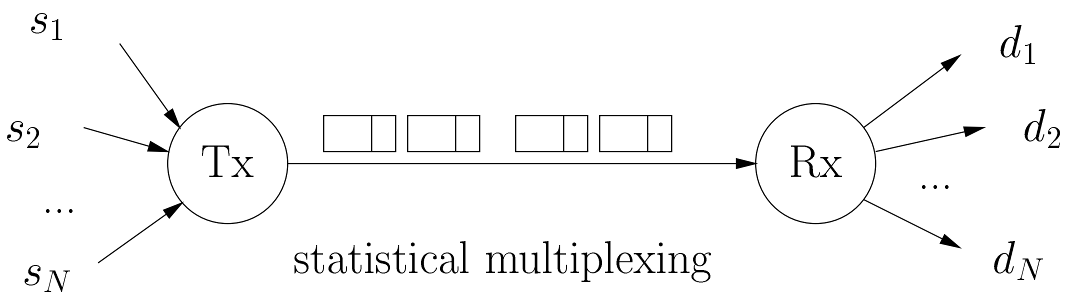
- Capacity of the timing channel

Relevant ideas

- *Protocol information* first analyzed by Gallager (1976)

- The *Poisson channel* first analyzed by Wyner (1988)

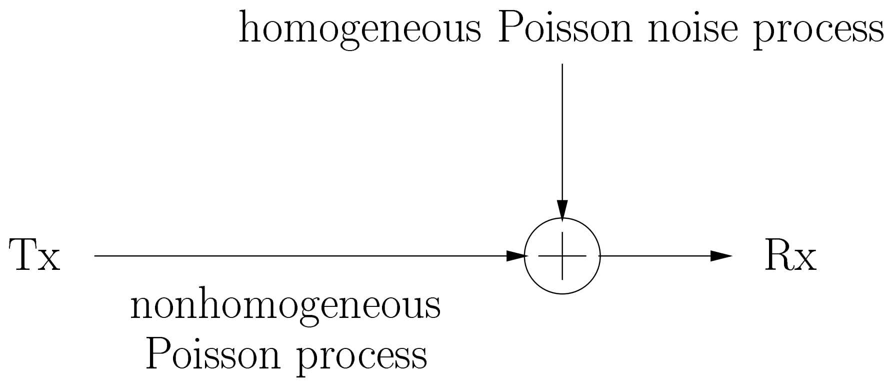
Packet Arrival Times vs. Addresses

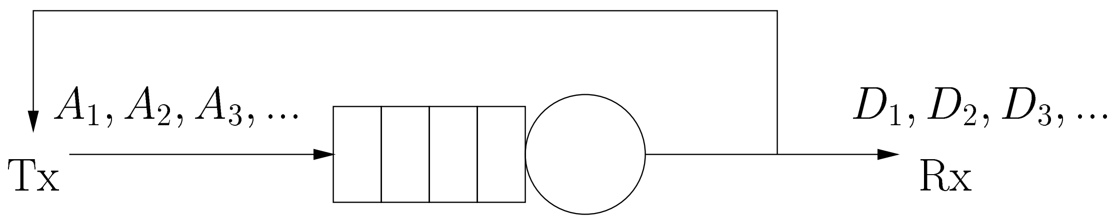


Protocol information $\geq$

I(packet arrival times; packet departure times)

Poisson Channel vs. Channel with a $\cdot/G/1$ Queue





Summary of main results

- For a $\cdot/M/1$ queue,

$$C \;=\; \frac{\mu}{e}\,\text{nats/s.}$$

- Saddle point property of exponential PDF (subject to a fixed mean): $A, A_{\text{expo}}, S, S_{\text{expo}}$,

$$C(A, S_{\text{expo}}) \;\leq\; C(A_{\text{expo}}, S_{\text{expo}}) \;\leq\; C(A_{\text{expo}}, S).$$

- For information bearing packets, $C_I > \mu$ (in bps) is possible.

Outlines of the talk

- Timing channel through a $\cdot/G/1$ queue
- A queue with information bearing packets
- *Telephone signalling channel*

Concentrate on a $\cdot/M/1$ queue.

Timing Channel through a $\cdot/G/1$ Queue

- Except for synchronization, packets contain no information.
- No feedback.
- Channel characterized by the conditional joint PDF

$$P\{\text{packet departure times}|\text{ packet arrival times}\}.$$

- An (n, M, T) code has M codewords.

  A code word has the form $(a_1, ..., a_n)$ with $\Sigma\, a_i \leq T$ on average.

  The transmitter sends packet 0 and then n more packets.

  The decoder observes $(d_1, ..., d_n)$ and makes a decision.

  The rate is $\frac{\log M}{T}$ bits/s.

- A rate R is achievable if there is a sequence of (n, M, T) codes with rates $\geq R$ and $P_e^{\max} \to 0$ as $T \to \infty$.
- The capacity of the queue C is the supremum of all achievable rates.
- A rate R is achievable *at the output (arrival) rate* λ if there is a sequence of $(n, M, n/\lambda)$ codes with rates $\geq R$ and $P_e^{\max} \to 0$ as $n \to \infty$.
- The capacity of the queue $C(\lambda)$ *at the output rate* λ is the supremum of all achievable rates at the output rate λ.

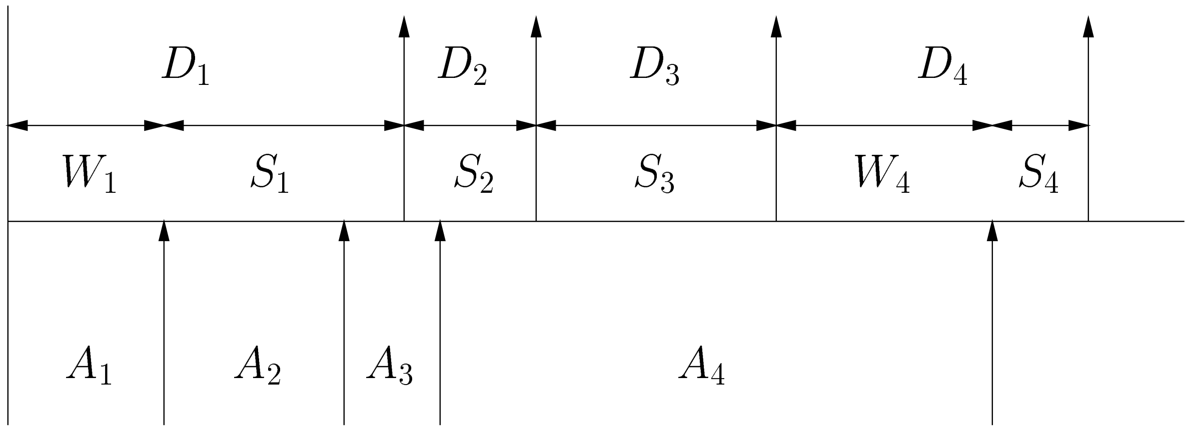
Theorem 1: $C = \sup_{\lambda<\mu} C(\lambda)$.

Theorem 2: *For a $\cdot/G/1$ queue with the mean service time $1/\mu$, let a RV S denote a service time, if $\lambda < \mu$, then*

$$C(\lambda) \ \leq \ \lambda \sup_{W \geq 0, \ E[W] \leq \frac{1}{\lambda} - \frac{1}{\mu}} I(W; W+S),$$

where W is independent of S.

Proof: Let W_i be the waiting time so that $D_i = W_i + S_i$. Let $U, \hat{U} \in \{1, ..., M\}$ be encoded and decoded messages.



By Fano's inequality and data processing inequality,

$$\log M \ \leq\ \frac{1}{1-\epsilon}\left[I(U;\hat{U})+1\right] \ \leq\ \frac{1}{1-\epsilon}\left[I(A_1^n;D_1^n)+1\right] \tag{1}$$

By the chain rule on $I(A_1^n;D_1^n)$,

$$I(A_1^n;D_1^n) \ =\ \sum_{i=1}^{n} I(A_1^n;D_i|D_1^{i-1}). \tag{2}$$

By the chain rule on $I(A_1^n,D_1^{i-1};D_i)$,

$$I(A_1^n,D_1^{i-1};D_i) \ =\ I(D_1^{i-1};D_i)+I(A_1^n;D_i|D_1^{i-1}),\ i>1. \tag{3}$$

Note that

$$\begin{aligned} I(A_1^n,D_1^{i-1};D_i) &= I(W_i;D_i),\ i>1, \\ I(A_1^n;D_1) &= I(W_1;D_1)=0. \end{aligned}$$

Using (2) and (3),

$$\begin{aligned} I(A_1^n;D_1^n) &= \sum_{i=1}^{n} I(W_i;D_i) - \sum_{i=2}^{n} I(D_1^{i-1};D_i) \\ &= \sum_{i=1}^{n} I(W_i;D_i) - D(P_{D_1,\ldots,D_n}||P_{D_1}\ldots P_{D_n}) \\ &\leq \sum_{i=1}^{n} I(W_i;D_i) \ =\ \sum_{i=1}^{n} I(W_i;W_i+S_i). \end{aligned}$$

Define a concave monotonically nondecreasing function

$$c(a) = \sup_{X \text{ ind of } S,\ X \le 0,\ E[X] \le a} I(X; X+S).$$

Note that

$$I(A_1^n; D_1^n) \ \le \ \sum_{i=1}^{n} c(E[W_i]).$$

Since the expected last departure occurs by time n/λ,

$$\frac{1}{n}\sum_{i=1}^{n} E[W_i] \ = \ \frac{1}{n}\sum_{i=1}^{n} E[D_i] - \frac{1}{n}\sum_{i=1}^{n} E[S_i] \ = \ \frac{1}{n}\frac{n}{\lambda} - \frac{1}{n}\frac{n}{\mu} \ \le \ \frac{1}{\lambda} - \frac{1}{\mu}. \tag{4}$$

Using (1), (4), (5), and Jensen's inequality,

$$\begin{aligned}
\log M \ &\le \ \frac{1}{1-\epsilon}\left[\sum_{i=1}^{n} c(E[W_i]) + 1\right] \\
\lambda\frac{\log M}{n} \ &\le \ \frac{\lambda}{1-\epsilon}\left[\frac{1}{n}\sum_{i=1}^{n} c(E[W_i]) + \frac{1}{n}\right] \ \le \ \frac{\lambda}{1-\epsilon}\left[c\left(\frac{1}{n}\sum_{i=1}^{n} E[W_i]\right) + \frac{1}{n}\right] \\
&\le \ \frac{\lambda}{1-\epsilon}\left[c\left(\frac{1}{\lambda} - \frac{1}{\mu}\right) + \frac{1}{n}\right]
\end{aligned}$$

Therefore,

$$R \ \le \ \lambda c\left(\frac{1}{\lambda} - \frac{1}{\mu}\right) \ = \ \lambda \sup_{W \text{ ind of } S,\ W \ge 0,\ E[W] \le \frac{1}{\lambda} - \frac{1}{\mu}} I(W; W+S) \qquad \Box.$$

Theorem 3 (Mutual information saddle point): *Let* $a, b \geq 0$ *be fixed. Define* $\overline{S} \sim$ *Expo*(b)*. Define* $\overline{W}$*, independent of* $\overline{S}$ *with mean* a*, to have the following PD*

$$\begin{aligned} P\{\overline{W} = 0\} &= \frac{b}{a+b}, \\ P\{\overline{W} > w | \ \overline{W} > 0\} &= e^{-w/(a+b)}. \end{aligned}$$

1. $I(\overline{W}; \overline{W} + \overline{S}) = \log\left(1 + \frac{a}{b}\right)$.
2. *For any nonnegative RVs* W *(independent of* $\overline{S}$*) and* S *(independent of* $\overline{W}$*) with mean* a *and* b *respectively,*

$$I(W; W + \overline{S}) \ \leq \ I(\overline{W}; \overline{W} + \overline{S}) \ \leq \ I(\overline{W}; \overline{W} + S).$$

Proof:

1. For $X \sim \text{Expo}(a)$, $h(X) = \ln ea$. $\overline{Y} = \overline{W} + \overline{S} \sim \text{Expo}(a+b)$.

$$\begin{aligned} I(\overline{W};\overline{W}+\overline{S}) &= h(\overline{W}+\overline{S}) - h(\overline{W}+\overline{S}|\overline{W}) \\ &= h(\overline{Y}) - h(\overline{S}) \\ &= \ln e(a+b) - \ln eb \\ &= \ln\left(1+\frac{a}{b}\right). \end{aligned}$$

2. $I(W;W+\overline{S}) \;=\; h(W+\overline{S}) - h(\overline{S}) \;\leq\; h(\overline{W}+\overline{S}) - h(\overline{S})$.

   It remains to show $I(\overline{W};\overline{W}+\overline{S}) \;\leq\; I(\overline{W};\overline{W}+S)$.

$$
\begin{aligned}
&I(\overline{W};\overline{W}+S)\\
&\quad= E\left[\ln \frac{P_{\overline{W},\overline{W}+S}}{P_{\overline{W}+S}P_{\overline{W}}}\right]\\
&\quad= E\left[\ln \frac{P_{\overline{W},\overline{W}+S}}{P_{\overline{W}+S}P_{\overline{W}}}\ \frac{P_{\overline{W}+\overline{S}}}{P_{\overline{W}+\overline{S}}}\ \frac{P_{\overline{W}+S|\overline{W}}}{P_{\overline{W}+S|\overline{W}}}\ \frac{P_{\overline{W}+\overline{S}|\overline{W}}}{P_{\overline{W}+\overline{S}|\overline{W}}}\right]\\
&\quad= D\left(P_{\overline{W}+S|\overline{W}}||P_{\overline{W}+\overline{S}|\overline{W}}\right) - D\left(P_{\overline{W}+S}||P_{\overline{W}+\overline{S}}\right) + E\left[\ln \frac{P_{\overline{W}+\overline{S}|\overline{W}}}{P_{\overline{W}+\overline{S}}}\right]\\
&\quad\geq E\left[\ln \frac{P_{\overline{W}+\overline{S}|\overline{W}}}{P_{\overline{W}+\overline{S}}}\right]\\
&\quad= \ln\left(1+\frac{a}{b}\right) - \frac{1}{b}E[S] + \frac{1}{a+b}E[\overline{W}+S]\\
&\quad= \ln\left(1+\frac{a}{b}\right)\\
&\quad= I(\overline{W};\overline{W}+\overline{S}) \qquad \square
\end{aligned}
$$

For a $\cdot/M/1$ queue, setting $b = 1/\mu$ and $a = 1/\lambda - 1/\mu$,

$$\begin{aligned} C(\lambda) &\leq \lambda \log \frac{\mu}{\lambda}, \quad \lambda < \mu, \\ C = \max_{\lambda<\mu} C(\lambda) &\leq \mu/e \text{ nats/s.} \end{aligned}$$

Theorem 6: *For the $\cdot/M/1$ queue with mean service time $1/\mu$,*

$$\begin{aligned} C(\lambda) &= \lambda \log \frac{\mu}{\lambda}, \ \lambda < \mu, \\ C &= \mu/e \text{ nats/s,} \end{aligned}$$

where C is achieved at $\lambda = \mu/e$.

We use random coding to generate each (n, M, T) code.

- Optimal packet arrivals form a Poisson process with rate λ.

  (1) generate M independent realizations of the process, or

  (2) generate, for each codeword, a Poisson RV N with mean $\mu T/e$, and then distribute N arrivals uniformly over T seconds.

- ML decision rule: the decoder

  (1) computes service times with respect to each codeword,

  (2) select the codeword with nonnegative service times and minimum total service time.

$\cdot/G/1$ Queue with Information Bearing Packets

- C_0: capacity of *information channel* (bits/transmitted packet).
- Errors in information channel are independent from service times.
- At output rate λ, we have two parallel channels

$$C_I = \sup_{\lambda \leq \mu} \left[C(\lambda) + \lambda C_0 \right].$$

Theorem 10: *For an information bearing $\cdot/G/1$ queue with mean service time $1/\mu$,*

$$C_I \text{ (in bits/s)} \geq \begin{cases} \mu e^{C_0}/e, & 0 \leq C_0 \leq 1 \text{ nat/transmission}, \\ \mu C_0, & C_0 \geq 1 \text{ nat/transmission}, \end{cases}$$

with equality iff the service time is exponential.

For a noiseless DMC with the set of packets $\mathcal{A}$,

$$C_I \text{ (in bits/s)} = \begin{cases} \frac{\mu}{e}\log_2 e, & |\mathcal{A}| = 1, \\ 2\frac{\mu}{e}\log_2 e, & |\mathcal{A}| = 2, \\ \mu \log_2 |\mathcal{A}|, & |\mathcal{A}| > 2. \end{cases}$$

For $|\mathcal{A}| = 2$,

$$C_I = 2\frac{\mu}{e}\log_2 e \approx 1.0615\mu \text{ bits/s}.$$

Telephone Signalling Channel

- The caller places the i^{th} call, waits S_i seconds for the first ring, hangs up, waits for W_{i+1} seconds (W_{i+1} depends on message and $S_i, S_{i-1}, ..., S_1$), and places the $(i+1)^{\text{th}}$ call.
- The receiver decodes the message from the observed intervals between calls $D_1, D_2, ...$
- We have an additive-noise memoryless channel with feedback

$$D_i \ = \ W_i + S_i \ \text{ for all } i.$$

- Since feedback cannot increase the capacity of a memoryless channel, the capacity is the same as without feedback.

Theorem 8: *For a telephone signalling channel with mean service time* $1/\mu$*,*

$$\begin{aligned} C_F &= \sup_{\lambda \le \mu} C_F(\lambda) \\ C_F(\lambda) &= \lambda \sup_{W \ge 0,\ E[W] \le \frac{1}{\lambda} - \frac{1}{\mu}} I(W; W+S) \end{aligned}$$

Thus $C_F \ge \mu/e$ *nats/s with equality iff service time is exponential.*

- This channel is equivalent, in capacity, to the $\cdot/G/1$ channel with noiseless and instantaneous feedback.
- Feedback does not increase the capacity of a $\cdot/G/1$ queue.

Summary

- For a $\cdot/M/1$ queue,

$$C \;=\; \frac{\mu}{e}\ \text{nats/s}.$$

- Saddle point property of exponential PDF (subject to a fixed mean): $A, A_{\text{expo}}, S, S_{\text{expo}}$,

$$C(A, S_{\text{expo}}) \;\leq\; C(A_{\text{expo}}, S_{\text{expo}}) \;\leq\; C(A_{\text{expo}}, S).$$

- For information bearing packets, $C_I > \mu$ (in bps) is possible.

- *Feedback does not increase the capacity of a $\cdot/G/1$ queue.*