ECE 7680
Lecture Queue – Bits through Queues

Objective: To briefly introduce queues, then describe recent results on the information capacity of queues

Introduction

With all the developments in networking, there have been relatively few "scientific" studies of networking and information theory. The union is, to quote one researcher, as yet "uncosummated." However, some recent work begins to point the way. It deals with passing information through queues. In this lecture we discuss this paper, which is important for a variety of reasons. (1) It serves to illustrate again the concept of capacity. (2) It is of recent research interest; (3) it has application to future technology. (4) It will put a lot of our tools to use and (5) it allows us to introduce (or review) the concepts of queuing, which are important engineering models.

Our primary reference is "Bits through queues" by Venkat Anantharam and Segio Verdu (IEEE Trans. Info. Theory, v. 42, no. 1, pp. 4–18), which won the award for the best paper of 1996.

Queues

A queue: input distrbution; service distribution; number of queues. We denote a queue by $X/Y/1$: is is input distribution, Y is service distribution. We characterize the service time by μ, the service rate, so the average service time is $1/\mu$.

To begin with, we represent information using codes consisting of M codewords, where each codeword constists of n interarrival times. We thus code information by the amount of delay between transmissions (not in the actual information in the packet transmitted).

Fig. 2 of paper:

A_i is the interarrival time, so the kth arrival (at the queue) is at time

$$\sum_{i=1}^{k} A_i$$

D_i = interdeparture time, so the departures occur at times

$$\sum_{i=1}^{k} D_i$$

S_i= service time.

W_i = idling time, the time between the $(i-1)$st departure and the ith arrival (if the ith arrival occurs before the $(i-1)$st departure, then $W_i = 0$. We have

$$D_i = W_i + S_i.$$

Then

$$W_i = \max\{0, \sum_{j=1}^{i} A_j - D_j\}. \tag{1}$$

Discuss how the service time affects the transmission.

Codes

Definition 1 An (n, M, T, ϵ) code for a queue consists of a codebook of M codewords, each of which is a vector of n nonnegative interarrival times $(a_1, \ldots, a_n)$; a decoder which upon observation of all n departures from the queue selects the correct codeword the probability $> 1 - \epsilon$, assuming the queue is initially empty; and the nth departure from the queue occurs on the average no later than T. □

The *rate* of the (n, M, T, ϵ) code is defined as

$$\log M/T$$

Definition 2 The capacity C of the queue is the largest R for which for all $\gamma > 0$ there exists a sequence of (n, M, T, ϵ_T)-codes that satisfy

$$\frac{\log M}{T} > R - \gamma$$

and $\epsilon_T \to 0$. □

The sequence seems to depend upon T.
Let λ be the average output rate.

Definition 3 R is ϵ-achievable at output rate λ if for all $\gamma > 0$ there exists a sequence of $(n, M, n/\lambda, \epsilon)$-codes such that

$$\lambda \frac{\log M}{n} > R - \gamma$$

Rate R is achievable at output rate λ if it is ϵ-achievable at output rate λ for all $0 < \epsilon < 1$. □

Definition 4 The capacity of the queue at rate λ, denoted $C(\lambda)$ is the maximum achievable rate at output rate λ. □

We have the following theorem (which we will not prove): The capacity of a single-server $\cdot/G/1$ queue with service rate μ satisfies

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

We will need the following result. (There are some problems with this derivation, but the authors use it.)

Theorem 1 *(sort of) Fano's inequality leads to*

$$\log |\mathcal{X}| \leq \frac{1}{1 - P_e}[I(X;Y) - \log 2] \tag{2}$$

Proof (Mostly) Starting from

$$P_e \log |\mathcal{X}| \geq H(X|Y) - 1 = H(X) - I(X;Y) - \log 2$$

we find

$$P_e \log |\mathcal{X}| - H(X) \geq -I(X;Y) - \log 2$$

or (here's the problem!)

$$P_e \log |X_c| - \log |\mathcal{X}| \geq -I(X;Y) - \log 2$$

from which the result follows. □

As an important notational simplification, write $D^i = (D_1, \ldots, D_i)$.

The key theorem depends upon the following lemma:

Lemma 1

$$I(A_1, \ldots, A_n; D_1, \ldots, D_n) = \sum_{i=1}^{n} I(W_i; W_i + S_i) - \sum_{i=2}^{n} I(D^{i-1}; D_i). \tag{3}$$

Proof From theorem 2.5.2 from the text, we can write

$$I(D^{i-1}; D_i) = I(A^n, D^{i-1}; D_i) - I(A^n; D_i | D^{i-1})$$

HW: Show that this is true. Summing, we obtain

$$\sum_{i=2}^{n} I(D^{i-1}; D_i) = \sum_{i=2} nI(A^n, D^{i-1}; D_i) - \sum_{i=2}^{n} I(A^n, D_i | D^{i-1})$$

Note that

$$I(A^n, D^{i-1}; D_i) = I(A_1, \ldots, A_n, D_1, \ldots, D_{i-1}; D_i) = I(W_i; D_i),$$

by (1). Thus we obtain

$$\sum_{i=2}^{n} I(D^{i-1}; D_i) = \sum_{i=2}^{n} I(W_i; D_i) - \sum_{i=2}^{n} I(A^n; D_i | D^{i-1}) \tag{4}$$

Also from thm 2.5.2 we have

$$I(A^n; D^n) = I(A^n; D_1) + \sum_{i=2}^{n} I(A^n; D_i | D^{i-1}) \tag{5}$$

Substituting from (4) into (5) we find

$$I(A^n; D^n) = I(A^1; D_1) + \sum_{i=2}^{n} I(W_i; D_i) - \sum_{i=2}^{n} I(D^{i-1}; D_i)$$

Since $I(A^1; D_1) = I(W_1; D_1)$, the result follows. □

The important theorem we will work on is the following:

Theorem 2 *For any $\cdot/G/1$ queue with service time S and $E[S] = 1/\mu$, if $\lambda \leq \mu$ then*

$$C(\lambda) \leq \lambda \sup_{X \geq 0, E[X] \leq 1/\lambda - 1/\mu} I(X; X + S), \tag{6}$$

where X is independent of S.

Proof Let $U \in \{1, 2, \ldots, M\}$ indicate the transmitted message, and let $V \in \{1, 2, \ldots, M\}$ indicate the decoded message. Using the statement of Fano's inequality from the previous theorem, we obtain

$$\log M \leq \frac{1}{1-\epsilon}[I(U; V) + \log 2] \leq \frac{1}{1-\epsilon}[I(A_1, \ldots, A_n; D_1, \ldots, D_n) + \log 2]$$

(the second inequality is essentially the data-processing inequality).

Now define the function

$$c(a) = \sup_{X \geq 0, E[X] \leq a} I(X; X + S)$$

where X is independent of S. By (3) we have

$$I(A^n, D^n) \leq \sum_{i=1}^{n} c(E[W_i]). \tag{7}$$

Now take the expectation $D_i = W_i + S_i$, using the fact that the expected last departure occurs before time n/λ, so $E[D_n] \leq n/\lambda$:

$$\frac{1}{n}\sum_{i=1}^{n} E[W_i] = \frac{1}{n}\sum_{i=1}^{n} E[D_i] - \frac{1}{n}\sum_{i=1}^{n} E[S_i] \leq \frac{1}{\lambda} - \frac{1}{\mu}. \tag{8}$$

So combining (2), (7), and (8) we get

$$\lambda \frac{\log M}{\lambda} \leq \frac{\lambda}{1-\epsilon}\left(\frac{1}{n}\sum_{i=1}^{n} c(E[W_i]) + \frac{\log 2}{n}\right)$$

Now we note that $c(a)$ is (1) concave and (2) monotone. (The first follows, by theorem 2.7.4. The second follows, since as a increases, we form the sup over a larger set of positive numbers.) Then we chain the inequalities:

$$\begin{aligned}
\lambda \frac{\log M}{\lambda} &\leq \frac{\lambda}{1-\epsilon}\left(\frac{1}{n}\sum_{i=1}^{n} c(E[W_i]) + \frac{\log 2}{n}\right) \\
&\leq \frac{\lambda}{1-\epsilon}\left(c\left(\frac{1}{n}\sum_{i=1}^{n} E[W_i]\right) + \frac{\log 2}{n}\right) \qquad \text{(Jensen's inequality)} \\
&\leq \frac{\lambda}{1-\epsilon}\left(c\left(\frac{1}{\lambda} - \frac{1}{\mu}\right) + \frac{\log 2}{n}\right) \qquad (c \text{ monotonic}, E[W_i] \leq ...)
\end{aligned}$$

This is equivalent to the statement of the theorem. □

Two questions: (1) how to compute the maximum mutual information and (2) whether the upper bound is tight (that is, whether it is actually achieved).

Let us now look at maximizing the mutual information in (6). Our result is analogous to the second-moment constrained random variables and Gaussian noise.

Theorem 3 *Let a and b be nonnegative real numbers. Let $\overline{N}$ be exponentially distributede with mean b, so*

$$p_{\overline{N}}(t) = \frac{1}{b}e^{-t/b}, t \geq 0.$$

Let $\overline{X}$ be a nonnegative random variable independent of $\overline{N}$ with the following:

$$P[\overline{X} = 0] = \frac{b}{a+b}$$

$$P(\overline{X} > x | \overline{X} > 0) = e^{-x/(a+b)}$$

Then it can be shown that $E[\overline{X}] = a$. Then:

1. $I(\overline{X}; \overline{X} + \overline{N}) = \log(1 + a/b)$.
2. *For any nonnegative random variable X independent of $\overline{N}$ with mean a,*

$$I(X; X + \overline{N}) \leq I(\overline{X}; \overline{X} + \overline{N}).$$

That is, $\overline{X}$ is an entropy-maximizing r.v.

3. *For any nonnegative random variable N, possibly dependent on $\overline{X}$, with mean b,*

$$I(\overline{X};\overline{X}+\overline{N}) \leq I(\overline{X};\overline{X}+N).$$

4. *For any independent nonnegative random variables X and N with means a and b, respectively,*

$$I(X;X+N) = \log(1+a/b) + D(P_N||P_{\overline{N}}) - D(P_{X+N}||P_{\overline{X}+\overline{N}}).$$

Proof

1. Let $\overline{Y} = \overline{X}+\overline{N}$. Then it can be shown that $\overline{Y}$ is exponential with mean $a+b$. (Use the fact that

$$f_{\overline{X}} = \frac{b}{a+b}\delta(x) + \frac{a}{(a+b)^2}e^{-x/(a+b)}$$

and

$$f_{\overline{Y}} = \frac{1}{b}\frac{1}{s+1/b}.$$

Multiply and take the inverse L.T.)

Now note that $f_{\overline{X}\,\overline{Y}} = f_{\overline{Y}|\overline{X}}f_{\overline{X}} = f_{\overline{N}}(\overline{Y}-\overline{X})f_{\overline{X}}$. Then

$$\begin{aligned} I(\overline{X};\overline{Y}) &= E[\log\frac{f_{\overline{X}\,\overline{Y}}}{f_{\overline{X}}f_{\overline{Y}}} = E[\log\frac{p_{\overline{N}}(\overline{Y}-\overline{X})}{p_{\overline{Y}}(\overline{Y})}] \\ &= E[\log 1/be^{-(\overline{Y}-\overline{X})/b}] - E[\log 1/(a+b)e^{-(y/(a+b))}] \\ &= \log(1+a/b). \end{aligned}$$

2. We have $I(X;X+\overline{N}) = h(X+\overline{N}) - h(X+\overline{N}|X) = H(X+\overline{N}) - h(\overline{N})$, which is maximized over all distributions having mean a by choosing $X = \overline{X}$, since then $X+\overline{N} = \overline{Y}$, which, as discussed above, has the exponential distribution.

3. Choose any nonnegative random variable N with mean b.

$$\begin{aligned} I(\overline{X};\overline{X}+N) &= D(P_{\overline{X}+N,\overline{X}}||P_{\overline{X}+N}P_{\overline{X}}) = EP_{\overline{X}+N|\overline{X}}P_{\overline{X}}\log\frac{P_{\overline{X}+N|X}P_{\overline{x}}}{P_{\overline{X}+N}P_{\overline{X}}} \\ &= EP_{\overline{X}+N|\overline{X}}P_{\overline{X}}\log\frac{P_{\overline{X}+N|X}P_{\overline{X}+\overline{N}|\overline{X}}P_{\overline{X}+\overline{N}}}{P_{\overline{x}+N}P_{\overline{X}}P_{\overline{X}+\overline{N}|\overline{X}}P_{\overline{X}+\overline{N}}} \\ &= D(P_{\overline{X}+N|\overline{X}}||P_{\overline{X}+\overline{N}|\overline{X}}|P_{\overline{X}}) - D(P_{\overline{X}+N}||P_{\overline{X}+\overline{N}}) + E\log\frac{P_{\overline{Y}|\overline{X}}}{P_{\overline{Y}}} \end{aligned}$$

where (for example),

$$D(P_{\overline{X}+N|\overline{X}}||P_{\overline{X}+\overline{N}|\overline{X}}|P_{\overline{X}}) = \sum_x P_{\overline{X}}\sum_N P_{\overline{X}+N|\overline{X}}\log\frac{P_{\overline{X}+N|\overline{X}}}{P_{\overline{X}+\overline{N}|\overline{X}}}$$

is called the conditional divergence. and we use the fact that $\overline{Y} = \overline{X}+\overline{N}$. Now it can be shown that conditioning increases divergence, so the two divergence terms combine to a value ≥ 0. We thus have

$$I(\overline{X};\overline{X}+N) \geq E\log\frac{P_{\overline{Y}|\overline{X}}}{P_{\overline{Y}}} = \log(1+a/b).$$

(after some more work).

4. Let $Y = X + N$. Then expanding we have

$$I(X;Y) = E\log\frac{P_{Y|X}}{P_{\overline{Y}|\overline{X}}} + E\log\frac{P_{\overline{Y}|\overline{X}}}{P_{\overline{Y}}} - E\log\frac{P_Y}{P_{\overline{Y}}} = D(P_N||P_{\overline{N}} + \log(1+a/b) - D(P_{X+N}||P_{\overline{X}+\overline{N}}).$$

□

Theorem 4 *The $\cdot/M/1$ queue with service rate μ satisfies*

$$C(\lambda) \leq \lambda\log\frac{\mu}{\lambda}, \lambda \leq \mu.$$

$$C \leq e^{-1}\mu \textit{ nats/s.}$$

Proof Let $b = 1/\mu$ and $a = 1/\lambda - 1/\mu$. Then by the previous theorem (part 2), $I(X; X+\overline{N}) \leq I(\overline{X}; \overline{X}+\overline{N}) = \log(1+a/b) = \log\mu/\lambda$, and by the first theorem we have

$$C(\lambda) \leq \lambda\log\mu/\lambda.$$

Then to find C, we maximize $C(\lambda)$ over all $\lambda \leq \mu$. The maximum is achieved at $\lambda = e^{-1}\mu$. □

More generally, for a $\cdot/G/1$ queue, we find

$$C(\lambda) \leq \lambda\log\mu/\lambda + \lambda D(P_s||e_\mu), \lambda \leq \mu,$$

where e_μ is an exponential with mean $1/\mu$. This follows from part (4) of two theorems back.

Achievability of the bound

It can also be shown (with a lot of work!) that for an $\cdot/M/1$ queue, the upper bound can be achieved when the input process is Poisson with rate λ. Thus

$$C = e^{-1}\mu \text{ nats/s.}$$

Information-bearing channels

We now generalize to the case where each packet actually carries some kind of information (e.g., bits). I will mostly summarize the results. Let the channel have information-bearing capacity C_0 in nats/sec.

Theorem 5 *The capacity of the information bearing queue is*

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

Observe that if C_0 is sufficiently large, then the λ that maximizes capacity is equal to μ, so no information is carried in the timing.

Combining some theorem, we find:

Theorem 6

$$C_I \geq \begin{cases} \mu e^{-1}\exp(C_0) & \textit{if } 0 \leq C_0 \leq 1 \textit{nat/symbol} \\ \mu C_0 & \textit{otherwise.} \end{cases}$$

For example, if there is noiseless binary transmission, then $C_0 = 1$, and we find

$$C_I = 2\mu e^{-1} \text{ nats/sec } = 1.06\mu \text{ bits/second.}$$

Pretty cool!