Entropy in Communication and Chemical Systems

Jean Walrand
Department of Electrical Engineering and
Computer Sciences
University of California
Berkeley, CA 94720
http://www.eecs.berkeley.edu/~wlr

***Abstract*—Entropy plays a central role in communication systems. On the one hand, the objective of communication is to reduce the entropy of some random variable. On the other hand, many useful models of communication networks evolve to a state of maximum entropy given external constraints. Chemical systems also exhibit a similar entropy-maximizing property, as do many systems of interacting particles. This paper reviews this set of fundamental ideas.**

I. Introduction

The goal of communication is to reduce the uncertainty about a random quantity. For instance, one knows the prior likelihood of possible events and learning which events actually occur reduces our uncertainty. Entropy measures the amount of uncertainty.

Users of a communication network place telephone calls or send data at random times. The state of the network can be modeled as a random process. Like many physical systems, some models of networks evolve to a state of maximum entropy given external constraints.

Chemical and interacting particle systems correspond to random interactions between molecules. Models of such interactions also exhibit a maximum entropy property (e.g., [2]).

Thus, in some abstract sense, the second law of thermodynamics applies to a wide class of systems. These systems are characterized by dynamics with specific time-reversibility properties.

This paper reviews the fundamental similarity of those apparently different systems.

II. Information and Entropy

Claude Shannon [7] explains that the *entropy* determines the minimum average number of bits needed to specify the value of a random variable. See [1] for a discussion of the ideas in this section.

Specifically, many sequences $\mathbf{X} = \{X_n, n \geq 1\}$ of random variables have the property that, for large n, the string $X^n = (X_1, \ldots, X_n)$ is almost uniformly distributed in a set of $2^{nH(\mathbf{X})}$ strings that we call *typical*, where $H(\mathbf{X})$ is a quantity called the *entropy rate* of the sequence. In that case, it takes $nH(\mathbf{X})$ bits to enumerate the possible typical values that the string X^n takes and, accordingly, one needs an average of $H(\mathbf{X})$ bits per random variable X_m to specify the value of the string X^n.

The simplest example is when the random variables X_n are independent and identically distributed (i.i.d.). Say that $P(X_n = x_i) = P(x_i) = p_i$ for $i \geq 1$ where $\sum_i p_i = 1$. In that case,

$$H(\mathbf{X}) = H(X_1) := -\sum_i p_i \log p_i.$$

By convention, the logarithms are in base 2. Here, $H(X_1)$ is called the *entropy of the random variable* X_1 with distribution π. Thus, the entropy rate of a sequence of $i.i.d.$ random variables is the entropy of each random variable.

Another easy example is when the random variables X_n are a stationary irreducible Markov chain on $\mathcal{X}$ with invariant distribution π and transition probability matrix $\{P(i,j), i,j \in \mathcal{X}\}$. In that case, one can show that

$$H(\mathbf{X}) = -\sum_{i,j} \pi(i)P(i,j)\log P(i,j).$$

The *Lempel-Ziv* encoding technique exploits the property that long sequences X^n are almost uniformly distributed in a set of $2^{nH(\mathbf{X})}$ typical sequences. The basic idea is to build a dictionary of strings of symbols as they occur in the sequence. As these strings reappear later, one replaces them by a pointer to their location in the dictionary. Imagine that we build a dictionary with strings of n symbols, as they occur in the sequence $X_1, X_2, \ldots$. If n is large, the dictionary will contain only $2^{nH(\mathbf{X})}$ typical sequences. Thus, after a while, one can replace the next n symbols by a pointer with $nH(\mathbf{X})$ bits to indicate the typical sequence that corresponds to these n symbols. Thus, this Lempel-Ziv encoding scheme requires $H(\mathbf{X})$ bits per symbol, which is the minimum number. The key observation is that the dictionary automatically identifies the typical sequences, without requiring any prior knowledge of the statistics of the source.

III. Time Reversal

We briefly review some notions that we use repeatedly in the rest of the paper. See [4] for an elegant presentation of these ideas.

Recall that a continuous-time Markov chain $\{X(t), t \geq 0\}$ is a random process that takes values in some countable set $\mathcal{X}$ and is characterized by the property that, for $x \neq y$,

$$P[X(t+\epsilon) = y \mid X(t) = x; X(s), s \leq t] = q(x,y)\epsilon + o(\epsilon)$$

where $q(x, y) \geq 0$ for $x \neq y$ is given and $o(\epsilon)$ is such that $o(\epsilon)/\epsilon \to 0$ as $\epsilon \downarrow 0$. We call $Q = \{q(x, y), x, y \in \mathcal{X}, x \neq y\}$ the *transition rates* of the Markov chain.

Recall also that a Markov chain is irreducible if it can go from any $x \in \mathcal{X}$ to any other $y \in \mathcal{X}$. An irreducible Markov chain has at most one invariant distribution $\pi = \{\pi(x), x \in \mathcal{X}\}$, which is a probability mass function with the property that if $X(0)$ has that distribution, then so does $X(t)$ for all $t \geq 0$. A distribution π is invariant if and only if it satisfies the following *balance equations*:

$$\sum_{x \neq y} \pi(x) q(x, y) = \pi(y) \sum_{x \neq y} q(y, x), \forall y \in \mathcal{X}. \tag{1}$$

A Markov chain is said to be time-reversible if $\{X(t), t \in [0, T]\}$ has the same distribution as $\{X(T-t), t \in [0, T]\}$ for all $T \geq 0$. This is the case if and only $X(0)$ has a distribution π that satisfies the following *detailed balance equations*:

$$\pi(x) q(x, y) = \pi(y) q(y, x) \text{ for all } x \neq y \text{ in } \mathcal{X}. \tag{2}$$

Note that π must then be invariant since summing the identities (2) over $x \neq y$ shows that π satisfies the balance equations (1).

Assume that π is a distribution and that Q are transition rates such that

$$\pi(x) q(x, y) = \pi(y) q'(y, x), \forall x \neq y \in \mathcal{X} \tag{3}$$

where $q'(x, y) \geq 0$ for all $x \neq y$ and

$$\sum_{y \neq x} q'(x, y) = \sum_{y \neq x} q(x, y), \forall x. \tag{4}$$

Then, summing (3) over $x \neq y$ yields

$$\sum_{x \neq y} \pi(x) q(x, y) = \pi(y) \sum_{x \neq y} q'(y, x) = \pi(y) \sum_{x \neq y} q(y, x)$$

where the last identity comes from (4). Consequently, (3) and (4) imply that π satisfies the balance equations (1) and is invariant. One can show that $q'(x, y)$ are the transition rates of $X(T - t)$, i.e., of the Markov chain reversed in time.

IV. Packet Switching

The Internet transports bits by first grouping them into *packets*. The packets generally have a variable length.

A. Single Buffer

Consider packets that arrive into a buffer where they wait to be transmitted. Following Kleinrock [5], say that the packets arrive as a Poisson process with rate λ and that their lengths are so that the transmitter takes exponentially distributed random times with rate $\mu > \lambda$ to transmit them. In that case, one can model the random number X_t of packets to be transmitted in the buffer at time $t \geq 0$ as a Markov chain with the nonzero transition rates $q(n, n+1) = \lambda$ and $q(n+1, n) = \mu$ for $n \geq 0$. This system is known as an M/M/1 queue.

Lemma 1: The stationary distribution of this Markov chain is $\pi = \{\pi(i), i \geq 0\}$ where

$$\pi(i) = (1 - \rho)\rho^i, i \geq 0 \text{ where } \rho := \frac{\lambda}{\mu}.$$

Proof: Algebra shows that the distribution π satisfies the detailed balance equations (2). ■

Theorem 1: The invariant distribution π maximizes the entropy $H(X)$ subject to $E(X) = \sum_i i p_i = L := \rho/(1 - \rho)$.

Proof: Consider the problem

$$\begin{aligned} &\text{Maximize } -\sum_i p_i \log p_i \\ &\text{subject to } \sum_i i p_i = L \text{ and } \sum_i p_i = 1. \end{aligned}$$

Associating the Lagrange multipliers α to the first constraint and β to the second constraint, the Lagrangian is

$$L(\mathbf{p}, \alpha, \beta) = -\sum_i p_i \log p_i + \alpha(\sum_i i p_i - L) + \beta(\sum_i p_i - 1).$$

Expressing that the partial derivative of $L(\mathbf{p}, \alpha, \beta)$ with respect to p_i is zero, we find

$$-1 - \log p_i + \alpha i + \beta = 0,$$

which shows that $\log p_i = \beta - 1 + \alpha i$, so that $p_i = A b^i$ for some constants A and b. The second constraint ($\sum_i p_i = 1$) implies that $A = 1 - b$. The first constraint then implies that $b = \rho$, so that $p_i = \pi_i$ for $i \geq 0$. ■

This observation shows that buffer queue length is the most "uncertain" given its average value.

B. Multiple Classes

As a minor extension, assume that packets of different types $f \in \{1, \ldots, F\}$ arrive at the buffer as independent Poisson processes with rates $\lambda(f)$. The packet transmission times are independent and exponentially distributed with rate μ. Let $X(t)$ be the list of packet types in the buffer at time t, in their order of arrival. The next result is in [4].

Lemma 2: $X(t)$ is a Markov chain and its invariant distribution π is as follows:

$$\pi(x) = (1 - \rho)\rho^K \Pi_{k=1}^K p(x_k), \forall x = (x_1, \ldots, x_K) \in \mathcal{X}$$

where $\mathcal{X}$ is the set of finite strings with elements in $\{1, \ldots, F\}$, $\rho = \lambda/\mu$ with $\lambda := \sum_f \lambda(f)$ and $p(f) = \lambda(f)\lambda^{-1}$. That is, the number of packets in the buffer is geometrically distributed with parameter ρ (thus with mean $\rho(1 - \rho)^{-1}$), and given the number of packet, their types are i.i.d. and equal to f with probability $p(f)$.

Proof: Let Q be the transition rates of $X(t)$ and Q' be the transition rates of the buffer where the packets flow in the other direction: the packets arrive at the head of the queue and leave from the tail. Then one can verify that π, Q, Q' satisfy (3)-(4). ■

For $x \in \mathcal{X}$, let $|x|_f$ be the number of packets of type f in x. The invariant distribution π is such that

$$E(|X|_f) = a_f := p(f)\rho(1 - \rho)^{-1}.$$

The proof of the next result is similar to that of Theorem 1.

Theorem 2: The distribution π maximizes the entropy $H(X)$ on $\mathcal{X}$ subject to $E(|X|_f) = a_f, f = 1, \ldots, F$.

C. Network

Now consider a simple model of a network. There are J buffers and packets of multiple flows go through the network. The packets of flow f arrive as a Poisson process with rate $\lambda(f)$ and go through a specific sequence of buffers: $1(f), 2(f), ..., e(f)$ where $e(f)$ is the last buffer that these packets go through before they exit the network. Each buffer stores the packets in their order of arrival and serves them one at a time. We assume that the transmission times of packets at buffer i are independent and exponentially distributed with rate μ_i. With these assumptions, the system is a *Jackson Network* [5]. The state of the network is the list $X(t)$ of the flow memberships of all the packets in all the buffers. Following [4], we find the following result.

Lemma 3: The invariant distribution π of $X(t)$ is given by

$$\pi(x) = \Pi_{j=1}^{J}(1-\rho_j)\Pi_k \frac{\lambda(x_{jk})}{\mu_j} \quad (5)$$

where x_{jk} is the flow membership of the packet in position k in buffer j and $\rho_j = \lambda_j/\mu_j$ where λ_j is the sum of the rates $\lambda(f)$ of the flows f that go through buffer j.

Proof: Let Q be the transition rates of $X(t)$ and Q' the transition rates of the same network but where the packets go through the buffers in the opposite direction. One then verifies that (3) and (4) are satisfied. ■

Theorem 3: This distribution maximizes the entropy $H(X_1, \ldots, X_J)$ subject to $E(|X_i|_f) = \lambda(f)/(\mu_i - \lambda_i)$ for $i = 1, \ldots, J$ and $f = 1, \ldots, F$.

Proof: The entropy of a vector of random variables is maximized when they are independent. Also, the entropy of each component is maximized, given the mean values, by the distribution that corresponds to a multiple class single buffer, as we saw in Theorem 2. ■

V. Circuit-Switched Networks

The telephone network transports voice calls as bit streams by setting up *circuits* from the source to the destination. A circuit connects the capacity needed for a voice call on each link along a path from the source to the destination.

A. Single Link

Consider a link that can carry up to K phone calls. There are $N \geq K$ customers attached to the system and they place phone calls independently after an exponential time with rate λ. The durations of the phone calls are independent and exponentially distributed with rate μ.

Let $X(t) = (X_1(t), \ldots, X_N(t))$ where $X_n(t)$ takes the value one if customer n is engaged in a phone call at time $t \geq 0$ and takes the value 0 otherwise.

Lemma 4: The process $\{X(t), t \geq 0\}$ is a Markov chain and its invariant distribution is given by

$$\pi(x) = A(K)\rho^{|X|}$$

where $\mathcal{X} = \{x \in \{0,1\}^N \mid |x| := x_1 + \cdots + x_N \leq K\}$. In this expression, $A(K)$ is the constant such that $\pi(x)$ sums to one.

Proof: Let $e_n \in \{0,1\}^N$ be such that all its components equal to zero except component n that is equal to 1. One then sees that

$$\pi(x)\lambda = \pi(x+e_n)\mu, \forall x \in \mathcal{X} \text{ such that } x + e_n \in \mathcal{X}.$$

This shows that π satisfies the detailed balance equations (2) for the Markov chain. ■

Define $a(K) := E(|X|)$ when X has the distribution π.

Theorem 4: The invariant distribution π maximizes the entropy $H(X)$ of X subject to $E(|X|) = a(K)$.

Proof: The proof is identical to that of Theorem 2. ■

B. Network

Consider a circuit-switched network with J links. Link j can carry up to K_j calls, for $j = 1, \ldots, J$. There are N circuits that correspond to specific paths in the network. Each circuit i places a call after an exponentially distributed random time with rate λ and that call has an exponential duration with rate μ. All these random variables are independent. If a call is placed along a path where one link cannot accept one more call, then the call is blocked (refused).

Let $X(t) = (X_1(t), \ldots, X_N(t))$ where $X_i(t) = 1$ if circuit i is engaged in a phone call at time $t \geq 0$ and $X_i(t) = 0$ otherwise. Let also $\mathcal{X}$ be the set of possible values of $X(t)$. That is, $\mathcal{X} = \{x \in \{0,1\}^N \mid \sum_i x_i A_{ij} \leq K_j, j = 1, \ldots, J\}$ where $A_{ij} = 1$ if circuit i goes through link j and $A_{ij} = 0$ otherwise. The following result is from [4].

Lemma 5: The process $\{X(t), t \geq 0\}$ is a Markov chain on $\mathcal{X}$ and its invariant distribution π has the following form:

$$\pi(x) = A\rho^{|x|}, x \in \mathcal{X} \text{ where } \rho = \frac{\lambda}{\mu}$$

where A is the constant such that $\pi(x)$ sums to one over $\mathcal{X}$.

Proof: Note that π satisfies the detailed balance equations. That is, if $x, x + e_i \in \mathcal{X}$, then $\pi(x)\lambda = \pi(x+e_i)\mu$. ■

This invariant distribution corresponds to some mean values $E(X_i) = a_i = P(X_i = 1)$ for $i = 1, \ldots, N$.

Theorem 5: The invariant distribution π maximizes the entropy $H(X)$ of X subject to $E(X_i) = a_i$ for $i = 1, \ldots, N$.

Proof: The proof is once again identical to that of Theorem 2. ■

VI. Task Processing

Consider the following model of a processing system. There is a set of J processing tasks and a finite set of resources. To perform task j, one needs a subset $\mathcal{R}_j$ of resources and these resources are then monopolized by the task while it is being performed. Finally, it takes an exponential time with rate μ_j to perform task j.

The key observation is that there are subsets $\mathcal{J}_i$ of tasks that can be performed concurrently because they require disjoint sets of resources. An example of such a system is a wireless communication network where the tasks are transmissions by links and the compatible transmissions are limited by interference.

A. Resource Allocation

Consider the following scheme to allocate resources to tasks. Each task j waits an independent random time with rate λ_j before requesting resources. If the subset $\mathcal{R}_j$ of resources that task j needs are available, it gets them and starts performing the task. Otherwise, the task restarts an exponential waiting time with rate λ_j.

Let $X(t) = (X_1(t), \ldots, X_J(t))$ where $X_j(t) = 1$ if task j is being performed at time t and $X_j(t) = 0$ otherwise. The following result is from [6].

Lemma 6: The Markov chain $X(t)$ has the invariant distribution π where

$$\pi(x) = A\Pi_j \rho_j^{x_j}, \text{ for } x \in \mathcal{X}.$$

In this expression, $\rho_j = \lambda_j/\mu_j$, $\mathcal{X}$ is the subset of $\{0,1\}^J$ that corresponds to possible simultaneous executions of tasks and A is the constant such that these probabilities sum to one.

Proof: One verifies that π satisfies the detailed balance equations

$$\pi(x)\lambda_j = \pi(x + e_j)\mu_j \text{ whenever } x, x + e_j \in \mathcal{X}.$$

■

This invariant distribution π corresponds to some values of $r_j = P(X_j = 1)$.

Theorem 6: The invariant distribution π maximizes the entropy $H(X)$ subject to $P(X_j = 1) = r_j$ for $j = 1, \ldots, J$.

Proof: The proof of this result is the same as the previous proofs. ■

B. Achieving Processing Rates

Consider once again the processing system and assume that jobs arrive for task j at an average rate a_j. To keep up with the jobs, one wishes to find the request rates $\{\lambda_j, j = 1, \ldots, J\}$ so that $r_j \geq a_j$ for $j = 1, \ldots, J$. The discussion is borrowed from [3].

This problem admits a solution if the rates $\mathbf{a} = \{a_j, j = 1, \ldots, J\}$ are not too large. Specifically, we say that the rates $\mathbf{a}$ are *strictly feasible* if they can be written as a positive convex combination of vectors x in $\mathcal{X}$, i.e., if

$$\mathbf{a} = \sum_{x \in \mathcal{X}} x p(x)$$

where $p(x) > 0$ for all $x \in \mathcal{X}$ and $\sum_{x \in \mathcal{X}} p(x) = 1$. For two distributions $\mathbf{p}$ and π on $\mathcal{X}$, let

$$d(\mathbf{p}, \pi) = \sum_x p(x) \log \frac{p(x)}{\pi(x)}.$$

To simplify the algebra, let $\lambda_j = e^{\sigma_j}$, so that the invariant distribution π can be written as $\pi(\sigma) = \{\pi(x;\sigma), x \in \mathcal{X}\}$. A direct calculation shows the following result.

Lemma 7: One has

$$\frac{\partial}{\partial \sigma_j} d(\mathbf{p}, \pi(\sigma)) = r_j - a_j, j = 1, \ldots, J.$$

where $r_j = \sum_x x_j \pi(x;\sigma)$ is the processing rate of task j.

The next theorem is a direct consequence of the lemma.

Theorem 7: (a) The distribution $\pi(\sigma)$ that minimizes $d(\mathbf{p}, \pi(\sigma))$ is such that

$$r_j = a_j, j = 1, \ldots, J.$$

(b) A gradient algorithm to minimize $d(\mathbf{p}, \pi(\sigma))$ is as follows:

$$\sigma_j(n+1) = \sigma_j(n) + \alpha(a_j - r_j(n))$$

where α is a step size and $r_j(n)$ is the processing rate of task j that was observed while $\sigma(n)$ was used.

Note that $\sigma_j(n) \approx \alpha q_j(n)$ where $q_j(n)$ is defined as the backlog of jobs that wait for task j. Indeed, that backlog increases at the average rate a_j and decreases with rate $r_j(n)$, so that its increase is approximately $a_j - r_j(n)$. Accordingly, a simple mechanism to achieve a strictly feasible set of rates $\mathbf{a}$ is for task j to request the resources $\mathcal{R}_j$ with rate $\lambda_j(n) = \exp\{\alpha q_j(n)\}$. That is, the *aggressiveness* of the requests for resources by task j should be exponential in the backlog of jobs for that task.

The problem of maximizing the utility of jobs that must go through a sequence of tasks is studied in [3]. It is shown that the aggressiveness of task j should be exponential in the difference between the backlog of task j and that of the next task needed by the job. Moreover, a new job should be admitted in the processing network at a rate that depends on the backlog of the first task that the job requires.

VII. Chemical System

Consider the following abstract model of a hypothetical system of molecules that undergo chemical reactions. There are J types of molecules and let $X(t) = (X_1(t), \ldots, X_J(t))$ where $X_j(t)$ is the number of molecules of type j at time $t \geq 0$. Through chemical reactions, molecules of different types combine and produces molecules of other types. Also, some molecules may decompose into molecules of different types. Such a reaction modifies the vector $X(t)$ from x to $y = x + v$. The vector v indicates how many molecules reacted and how many where produced. For instance, in a reaction $2H_2 + O_2 \rightarrow 2H_2O$, we have $v = (-2, -1, +2)$ if v_1, v_2, v_3 correspond to the number of molecules of hydrogen, oxygen, and water, respectively.

We follow ideas from [4] and [8]. Assume that $X(t)$ is a Markov chain and that the transition rates are such that

$$q(x, x+v) = \Pi_j \phi_j(x_j) a_j^{-v_i^-}, v \in \mathcal{V}, x, x+v \in \{0, 1, \ldots\}^J$$

where $v^- := \max\{0, -v\}$. The meaning of this expression is that the rate of a reaction $x \rightarrow x + v$ is a function of the concentration of reactants and of the reaction itself. Here, $\mathcal{V}$ is a set of possible reactions determined by the laws of chemistry. We assume that if $v \in \mathcal{V}$, then $-v \in \mathcal{V}$. We are not suggesting that this form holds for all chemical reactions; we are only exploring the consequences of this particular law.

Lemma 8: The invariant distribution π of the Markov chain $X(t)$ is such that

$$\pi(x) = A\Pi_j \frac{a_j^{x_j}}{\phi_j(x_j)}$$

where A is a constant so that π sums to one over the set $\mathcal{X}$ of possible values of $X(t)$ starting from some initial condition x_0.

Proof: We check that π satisfies the detailed balance equation (2). Consider $x \in \mathcal{X}$ and $v \in \mathcal{V}$ so that $x + v \in \mathcal{X}$. Note that $\pi(x+v) = \pi(x)\Pi_j[a_j^{v_j}\phi_j(x_j)/\phi_j(x_j+v_j)]$. Also, $q(x+v,x) = \Pi_j\phi_j(x_j+v_j)a_j^{-(-v_j)^-}$. Hence the identity (2) reads

$$\pi(x)\Pi_j\phi_j(x_j)a_j^{-v_j^-} = \pi(x)\Pi_j[a_j^{v_j}\phi_j(x_j)/\phi_j(x_j+v_j)] \times \Pi_j\phi_j(x_j+v_j)a_j^{-(-v_j)^-}$$

and one sees that it is satisfied because $v = (-v)^- - v^-$. ■

We now particularize this result to the case where

$$\phi_j(x_j) = x_j^{b_j}, j = 1, \ldots, J.$$

For instance, for first-order reactions, the speed is proportional to the concentration of reactants and for second-order reactions, it is proportional to the square of those concentrations. With this assumption,

$$\pi(x) = A\Pi_j a_j^{x_j} x_j^{-b_j}.$$

Define $\lambda_j = E(X_j)$ and $\mu_j = E(\log(X_j)), j = 1, \ldots, J$ when X has the stationary distribution π. One then has the following result.

Theorem 8: The invariant distribution π maximizes the entropy $H(X)$ subject to the constraints that $E(X_j) = \lambda_j$ and $E(\log(X_j)) = \mu_j$ for $j = 1, \ldots, J$.

Proof: The Lagrangian of the constrained entropy maximization problem is

$$\begin{aligned} &L(p,\alpha,\beta,\gamma) \\ &= -\sum_x p(x)\log p(x) + \sum_j \alpha_j[\sum_x x_j p(x) - \lambda_j] \\ &+ \sum_j \beta_j[\sum_x \log(x_j)p(x) - \mu_j] + \gamma[\sum_x p(x) - 1]. \end{aligned}$$

The derivative with respect to $p(x)$ being equal to zero implies that

$$-1 - \log p(x) + \sum_j \alpha_j x_j + \sum_j \beta_j \log_j(x_j) + \gamma = 0.$$

Consequently,

$$p(x) = A\Pi_j e^{\alpha_j x_j} x_j^{\beta_j}.$$

The $2J+1$ constants A, α_j, β_j are determined by the $2J+1$ equality constraints and must then be such that $e^{\alpha_j} = a_j$ and $\beta_j = -b_j$, so that $p(x) = \pi(x)$. ■

VIII. Conclusion

The paper explored the role of entropy in communication systems, processing systems, and chemical systems.

As Shannon explained, in communication, entropy measures the number of bits needed to describe an unknown quantity such as an image or a bit file. Thus, in this abstract view, the role of communication is to reduce entropy.

Jackson networks are a simple class of models of packet-switched networks such as the Internet. The stationary distribution of the state of the queues of a Jackson network maximizes the entropy of that state given the average number of packets of each flow in each queue.

We considered a model of circuit-switched networks, such as the telephone network. For that model, the stationary distribution of the state of the customers (engaged or not in a call) also maximizes the entropy subject to the probability that customers are busy.

A model of processing system has a set of tasks. Each task requires a set of resources. The processing times are exponentially distributed. The stationary distribution maximizes the entropy subject to the fraction of time that each task is being performed. Moreover, minimizing the relative entropy between these fraction of times and the rates needed to keep up with arriving jobs provides a simple resource allocation algorithm where tasks request the resources more aggressively as their backlog increases.

Finally, we examined an hypothetical chemical system where the reaction rates have a specific functional dependence on the concentrations of reactants and on the reactions that take place. We showed that this model also exhibits an entropy-maximization property.

Acknowledgment

The author would like to thank Libin Jiang and Jiwoong Lee for their suggestions for improving the text and the U. S. Army Research Office for their support through a MURI grant under BAA 07-036.18.

References

[1] T.M. Cover and J.A. Thomas, Elements of Information Theory, Second Edition, Wiley 20066.

[2] J.W. Gibbs, Elementary principles in statistical mechanics. New York. (1902), C. T. Woodbridge (1981), CT: Ox Bow Press

[3] L. Jiang and J. Walrand A CSMA Distributed Algorithm for Throughput and Utility Maximization in Wireless Networks, submitted July 2008.

[4] F. P. Kelly, Reversibility and Stochastic Networks, Wiley, 1979.

[5] L. Kleinrock, Queueing Systems. Volume 1: Theory, Wiley Interscience, New York, 1975.

[6] X. Wang and K. Kar, Throughput Modelling and Fairness Issues in CSMA/CA Based Ad-Hoc Networks, Proceedings of IEEE Infocom 2005, Miami, March 2005.

[7] C. E. Shannon, "A mathematical theory of communication," Bell System Technical Journal, vol. 27, pp. 379-423 and 623-656, July and October, 1948.

[8] P. Whittle, Systems in stochastic equilibrium, John Wiley & Sons, Inc., 1986, New York, NY, USA.