Hiding and Covering in a Compact Metric Space

R. J. McEliece and E. C. Posner
Communications Systems Research Section

This paper studies the relationship between games of search on a compact metric space X and the absolute epsilon entropy I (X) of X. The main result is that $I(X) = -\log v_L^*$, v_L^* *being the lower value of a game on X we call "restricted hide and seek."*

I. Introduction

Let X be a set, S a collection of subsets of X with $\cup S = X$. The two-person zero-sum game "hide and seek" $G(X, S)$ is played as follows. Player 1 (the "hider") chooses a point $x \epsilon X$, and player 2 (the "seeker") chooses $s \epsilon S$. If $x \epsilon s$ player 1 pays player 2 one unit; otherwise no payoff occurs. Let us denote the value of this game, if it exists, by v. (We assume that X has enough structure so that mixed strategies can be defined.)

Now for each integer N let c_N be the smallest integer such that the cartesian power X^N can be covered with c_N sets from S^N, and let

$$c = \lim_{N \to \infty} c_N^{1/N}$$

The main theorem of a previous paper of ours (Ref. 1) was that if X is finite, $v = c^{-1}$. It is the object of this paper to study the relationship between v and c when X is a compact metric space, and S is the set of closed spheres of radius ε.

Our first main result (Theorem 1) is that in this situation, the game G still has a value. For finite X von Neumann's fundamental theorem on finite two-person zero-sum games immediately implies that v exists, and so in Ref. 1 this problem did not arise.

Our second main result is that $c = v^{-1}$ is not true in general, but rather that $c = v^{*-1}$, where v^* is the best expected gain the hider can guarantee himself when he must restrict his sets to a finite subset of X he has chosen in advance. It is always true that $v^* \leqq v$, and for a fixed X, $v^* = v$ except for at most countably many values of ε. In *Section IV*, however, we give an example of a compact metric space for which $v^* < v$. In *Section V* we prove that $c = v^{*-1}$.

These problems arise in information theory. The logarithm of the limit c is the least average number of bits per sample necessary to describe X modulo S; i.e., to identify an s containing x, when block coding is used, and when there is no *a priori* probability distribution on X. We shall show at the end of *Section V* that $-\log v$ represents the maximum, over all Borel *a priori* probability distributions on X, average number of bits per sample necessary to describe X to within an ambiguity of ε, when variable-length coding is used. Thus when $v = c^{-1}$ (the usual state of affairs in spite of our counter-example) there exist probability distributions on X which render variable-length coding useless.

II. General Hide and Seek

If the hider chooses his point x according to a probability distribution λ on (a Borel field containing the points of) X, we say he uses strategy λ. Similarly a strategy μ for the seeker is a probability distribution on (a Borel field containing the points of) S. Let $E = \{(x,s) : x \epsilon s\}$, a subset of the product space $X \times S$. The expected value of the payoff, given that the hider plays strategy λ and the seeker plays μ, is $(\lambda \times \mu)(E) = v(\lambda,\mu)$, $\lambda \times \mu$ being the product measure induced by λ and μ on $X \times S$.

If the hider uses a fixed strategy λ, then from his point of view the worst possible expected payoff is

$$\sup_{\mu} v(\lambda,\mu)$$

Hence he will choose a λ which makes

$$\sup_{\mu} v(\lambda,\mu)$$

as small as possible. Thus we define the *upper value* of $G(X,S)$ as

$$v_U(X,S) = \inf_{\lambda} \sup_{\mu} v(\lambda,\mu). \tag{2.1}$$

Similarly the seeker will choose a μ which makes

$$\inf_{\mu} v(\lambda,\mu)$$

as large as possible, and we define the *lower value* of $G(X,S)$ as

$$v_L(X,S) = \sup_{\mu} \inf_{\lambda} v(\lambda,\mu) \tag{2.2}$$

It is an easy exercise to show that $v_L \leqq v_U$. If it happens that $v_L = v_U$ we denote this common value by $v(X,S)$, and say that the game $G(X,S)$ has a value. If the game has a value, then for every $\eta > 0$, there exist strategies λ and μ such that if the hider plays λ, his expected loss is $\leqq v(X,S) + \eta$ no matter how the seeker plays, and if the seeker plays μ his expected gain is $\geqq v(X,S) - \eta$ no matter how the hider plays. If it happens that there exist strategies λ for the hider which guarantee an expected loss no greater than $v(X,S)$, these strategies are calle l *optimal* strategies. Optimal strategies for the seeker are defined similarly.

There is another form of the definitions of v_U and v_L which will be useful in what follows. By the definition of product measure we can write $v(\lambda,\mu)$ as either of the integrals

$$v(\lambda,\mu) = \int_X \mu(\text{star}(x))\, d\lambda$$
$$= \int_S \lambda(s)\, d\mu \tag{2.3}$$

where $\text{star}(x) = \{s \epsilon S | x \epsilon s\}$. Now if we define the *pure strategy* λ_x for the hider as that strategy which always chooses x; i.e., $\lambda(x) = 1$, $\lambda(x') = 0$ if $x' \neq x$, we see that $\mu(\text{star}(x)) = v(\lambda_x,\mu)$. Similarly if μ_s is a pure strategy for the seeker, $\lambda(s) = v(\lambda,\mu_s)$. Thus from Eq. (2.3) we obtain the estimate

$$v(\lambda,\mu) \leqq \sup_{s \epsilon S} \lambda(s) = \sup_{s \epsilon S} v(\lambda,\mu_s)$$

Hence for a fixed λ,

$$\sup_{\mu} v(\lambda,\mu) = \sup_{s \epsilon S} v(\lambda,\mu_s)$$

and so

$$v_U(X,S) = \inf_{\lambda} \sup_{s \epsilon S} \lambda(s) \tag{2.1'}$$

and similarly

$$v_L(X,S) = \sup_{\mu} \inf_{x \epsilon X} \mu(\text{star}(x)) \tag{2.2'}$$

Let us remark finally that if the set X is finite, it is a consequence of the fundamental theorem of finite two-person, zero-sum games that $G(X,S)$ has a value (Ref. 2, Chap. 7).

III. Hide and Seek in a Compact Metric Space

For the remainder of the paper X will be a compact metric space and S will be the set of closed spheres[1] of radius ε for a fixed $\varepsilon : s_\varepsilon(x) = \{y \epsilon X : d(y,x) \leqq \varepsilon\}$. This game is denoted by $G(X,\varepsilon)$. In this case strategies for the hider and the seeker will both be Borel probability distributions on X, since the seeker need only specify the center of the sphere he wishes to select. In the product space

[1] However, the results in this article also hold when S is the set of closed sets of diameter $\leqq \varepsilon$.

$X \times X$, the set $E = \{(x,y) : d(x,y) \leqq \varepsilon\}$, and for strategies λ and μ, $v(\lambda,\mu) = (\lambda \times \mu)(E)$. Before proceeding we need a result on weak convergence.

Let $B(X)$ be the space of all Borel probability distributions on X, $C(X)$ the space of real-valued continuous functions on X. The topology of weak convergence on $B(X)$ is defined (Ref. 3, Chap. II) as follows: $\mu_n \rightarrow \mu$ in $B(X)$ if for every $f \epsilon C(X)$

$$\int f d_{\mu_n} \rightarrow \int f d_{\mu}$$

B(X) is compact in this topology (Ref. 3, p. 64) and if F is any closed subset of X and $\mu_n \rightarrow \mu$, then

$$\mu(F) \geqq \limsup_{n \rightarrow \infty} \mu_n(F) \qquad (3.1)$$

(Ref. 3, p. 40).

We now consider probability distributions on the product space $X \times X$. The following proof, as are all proofs in this article, is omitted.

Lemma 1. *If $\mu_n \rightarrow \mu$ and $\lambda_n \rightarrow \lambda$ then $\mu_n \times \lambda_n \rightarrow \mu \times \lambda$.*

Lemma 2. *If $\lambda_n \rightarrow \lambda$ and $\mu_n \rightarrow \mu$, then*

$$v(\lambda,\mu) \geqq \limsup_{n \rightarrow \infty} v(\lambda_n,\mu_n)$$

We now have the main theorem of this section.

Theorem 1. *$G(X,\varepsilon)$ has a value $v(\varepsilon)$ which is continuous from above in ε, and the seeker has an optimal strategy. For every $\delta > 0$ the hider has a strategy with finite support which guarantees that he loses no more than $v(\varepsilon) + \delta$. The set of optimal strategies for the seeker is closed in the topology of weak convergence.*

We conclude this section with two examples which show the necessity of certain of the hypotheses in Theorem 1.

Example 1

The hider need not have an optimal strategy. X will be a countable subset of the unit circle, under the geodesic metric. Let $x_n = \exp(\pi i/2^{n+1})$. X will consist of the points $\pm x_n$, $\pm ix_n$ for all n. Then X is closed and so compact. Let $\varepsilon = \pi/2$; for each $x \epsilon X$ we adopt the abbreviation $s(x) = s_{\pi/2}(x)$. Then if the seeker plays ± 1 each with probability ½ his expected gain against any pure hider's strategy will be $\geqq 1/2$ and so $v_L \geqq 1/2$. On the other hand, if the hider uses the strategy λ_N defined by $\lambda_N(x) = 1/2N$ for $x = \pm x_1, \pm x_2, \cdots, \pm x_N$; $\lambda_N(x) = 0$ otherwise, then $\lambda_N(s(x)) = 1/2 + 1/2N$ if $x = \pm ix_n$ for some $n \leqq N$; $\lambda_N(s(x)) = 1/2$ otherwise, and so the hider's expected loss is $\leqq 1/2 + 1/2N$ for any pure seeker's strategy. Thus $v_U \leqq 1/2 + 1/2N$ for any N, and so $G(X,\pi/2)$ has value 1/2. If, however, the hider had an optimal strategy λ, $\lambda(s(x)) \leqq 1/2$ for all $x \epsilon X$, then it would follow from $\lambda(s(x)) + \lambda(s(-x)) = 1 + \lambda(ix) + \lambda(-ix)$ that $\lambda(ix) = \lambda(-ix) = 0$ for all $x \epsilon X$, a contradiction.

Example 2

The set of optimal strategies for the hider, if non-empty, need not be closed. Let X be the closed interval $[0,4]$ under the usual metric, and $\varepsilon = 1$. Then $v(X,\varepsilon) = 1/2$, and if λ_n is the strategy

$$\lambda_n(0) = \lambda_n\left(2 + \frac{1}{n}\right) = \frac{1}{2}$$

then λ_n is optimal for all $n \geqq 1$. However $\lambda_n \rightarrow \lambda$ where $\lambda(0) = \lambda(2) = 1/2$, but λ itself is not optimal, since if the seeker always picks the sphere centered at 1, his gain against λ is always 1.

Example 3

The seeker need not have finitely based nearly optimal strategies such as the hider has; i.e., it is possible that there exists $\delta > 0$ such that if μ is any finitely based strategy (a probability distribution on X which is zero outside a finite subset of X), then $\mu(s_\varepsilon(x)) \leqq v(\varepsilon) - \delta$ for some $x \epsilon X$. This example is best understood in the context of a game we call "restricted hide and seek," introduced in the next section, so we postpone it until then.

IV. Restricted Hide and Seek

In restricted hide and seek, before play begins the seeker is required to choose a finite subset X' of X (unknown to the hider) and then must always choose a sphere of radius ε whose center is in X'. Of course, a referee who knows X' will be needed to keep the seeker honest, since there is no way the hider will be able to tell whether or not the seeker is staying in X'. We denote this game as $G^*(X,\varepsilon)$ and define $v_L^*(\varepsilon), v_U^*(\varepsilon)$ as in *Section II.* Let

$$v(\varepsilon^-) = \lim_{\varepsilon' \uparrow \varepsilon} v(\varepsilon')$$

Lemma 3. $v(\varepsilon^-) \leqq v_L^*(\varepsilon) = v(\varepsilon)$

Lemma 4. *$v_L(\varepsilon) = v(\varepsilon)$ with at most countably many exceptions.*

If X has only two points x,y and $d(x,y) = 1$, then $v(X,1^-) = 1/2$ but $v(X,1) = v_L^*(X,1) = 1$. It is much more difficult to give an example which shows that v_L^* may be strictly less than v_L. We now mention such an example.

Example 4

There exists a compact metric space X such that $v(X,1^-) < v_L^*(X,1) < v(X,1)$.

Let C be a circle of circumference 4, d the geodesic metric on C, and let H_C be the space of closed subsets of C under the Hausdorff metric d':

$$d'(E,F) = \max\left(\max_{e\epsilon E}\min_{f\epsilon F} d(e,f), \max_{f\epsilon F}\min_{e\epsilon E} d(e,f)\right).$$

H_C is a compact metric space under d' (Ref. 4). The set Z of all closed subsets of C of Lebesgue measure 2 is a closed, hence compact, subspace of H_C and is, therefore, separable. Let $\{B_i, i \geqq 1\}$ be a countable dense subset of Z. No finite subset $\{b_k\}$ of C has the property that every B_i contains a b_k. For $\{b_k\}$ can be covered by an open set of arbitrarily small Lebesgue measure and so there exists a set $B \epsilon Z$ and $d_0 > 0$ such that $d(B,b_k) \geqq d_0$ for all k. Thus a B_i such that $d'(B,B_i) < d_0$ cannot contain a b_k.

The space X of this example will have C as a subspace, the metric restricted to C being the geodesic metric. It also contains points $a;a_i\ i \geqq 1$ where

$$d(a,c) = 1 \text{ for } c \epsilon C$$

$$d(a,a_i) = 2^{-i}$$

$$d(a_i,a_j) = |2^{-i} - 2^{-j}|$$

$$d(a_i,c) = 1 + \min(d(B_i,c), 2^{-i-1}) \text{ for all } c \epsilon C$$

In addition, X contains three points c_1', c_2', c_3' which are to be thought of as outside the circle C and equally spaced in angle. The point on C closest to c_i' is labeled c_i. The metric is extended as follows:

$$d(c_i',a) = d(c_i',a_j) = 15/8 \text{ for all } i,j.$$

$$d(c_i',c_i) = 7/8$$

$$d(c_i',c) = \begin{cases} 7/8 + d(c_i,c) & \text{if } d(c_i,c) \leqq 1/8 \\ 1 & \text{if } 1/8 \leqq d(c_i,c) \leqq 1 \\ d'(c_i,c) & \text{if } d(c_i,c) \geqq 1 \end{cases}$$

$$\text{for } c \epsilon C.$$

We assert that (X,d) as defined above is indeed a compact metric space, but omit the tedious verification that d satisfies the triangle inequality. Compactness is best verified by checking sequential compactness, which is equivalent to compactness for a metric space.

It can now be shown that

$$v(X;1^-) = \frac{1}{3}, \qquad v_L^*(X,1) = \frac{2}{5}, \qquad v(X,1) \geqq \frac{1}{2}$$

V. Absolute Epsilon Entropy

Let us use the term "ε-set" to describe a subset of a compact metric space which is contained in some sphere of radius ε. The *epsilon entropy* $H_\varepsilon(X)$ is then defined to be $\log_2 N$, where X can be covered with N ε-sets, but no fewer. $H_\varepsilon(X)$ can be interpreted information theoretically as the minimum average number of bits per sample needed to describe X to within an error of at most ε.

If $(X_i,d_i)\ i = 1, 2, \cdots, n$ are compact metric spaces we shall make the cartesian product $X_1 \times X_2 \times \cdots \times X_n$ into a compact metric space by defining

$$d((x_1, \cdots, x_n), (x_1', \cdot \quad , x_n')) = \max_i d(x_i,x_i')$$

With this definition products of ε-sets are ε-sets and projections of ε-sets onto the coordinate spaces X_i are ε-sets; hence it is a suitable definition for dealing with uniform approximation. If $X_i = X$ for all i we shall write X^n instead of $X_1 \times \cdots \times X_n$.

The *absolute epsilon entropy* $I_\varepsilon(X)$ is defined by

$$I_\varepsilon(X) = \lim_{n\to\infty} \frac{1}{n} H_\varepsilon(X^n)$$

That the limit exists is a consequence of the simple property $H_\varepsilon(X^{n+m}) \leqq H_\varepsilon(X^n) + H_\varepsilon(X^m)$. $I_\varepsilon(X)$ can be interpreted as the minimum average number of bits per sample needed to describe X to within ε when an unlimited number of samples can be stored prior to transmission.

Theorem 2, the main result of this paper, identifies $I_\varepsilon(X)$ in terms of the game "restricted hide and seek."

Theorem 2. $I_\varepsilon(X) = -\log v_L^*(X;\varepsilon)$.

Theorem 2 requires two lemmas.

Lemma 5. $H_\varepsilon(X) = -\log v_L^*(X;\varepsilon)$.

Lemma 6. $v_L^*(X \times Y, \varepsilon) = v_L^*(X,\varepsilon)\, v_L^*(Y,\varepsilon)$.

We conclude the paper with two corollaries to Theorem 2. Let p be a Borel probability measure on X, and let $H_{\varepsilon;p}(X)$ be the infimum, over all partitions

$$X = \bigcup_i A_i,\ A_i \cap A_j = \phi \text{ if } i \neq j,$$

each A_i being a Borel ε-set of X, of the Shannon entropy

$$-\sum_i p(A_i) \log p(A_i)$$

$H_{\varepsilon;p}$ is called the $\varepsilon;p$ entropy of X (Ref. 5). Also define the absolute $\varepsilon;p$ entropy of X by

$$I_{\varepsilon;p}(X) = \lim_{n\to\infty} \frac{1}{n} H_{\varepsilon;p} n (X^n),$$

p^n being the product measure induced on X^n by p. $I_{\varepsilon;p}(X)$ then represents the minimum average number of bits per sample necessary to describe X with an error not exceeding ε, with p as the *a priori* probability distribution on X, when arbitrarily long variable-length codes are used. Combining Theorem 2 with Theorem 2 of Ref. 1, which had

$$-\log v(X,\varepsilon) = \sup_p I_{\varepsilon;p}(X),$$

we conclude

Corollary 1.

$$I_\varepsilon(X) = \sup_p I_{\varepsilon,p}(X)$$

whenever $v_L^*(X,\varepsilon) = v(X,\varepsilon)$; in particular equality holds for all but at most countably many ε.

Hence most of the time "nature" can choose a p on X which is so "bad" that prior knowledge of p could not be used to increase the transmission rate.

Our final result is a simple consequence of Theorem 2 and Lemma 6, and tells us that one cannot save anything by encoding two sources simultaneously.

Corollary 2. $I_\varepsilon(X \times Y) = I_\varepsilon(X) + I_\varepsilon(Y)$.

References

1. McEliece, R. J., and Posner, E. C., "Hide and Seek, Data Storage, and Entropy," *Annals Math. Stat.*, Vol. 42, pp. 1706–1716, 1971.
2. Gale, D., "The Theory of Linear Economic Models," McGraw-Hill Book Co., New York, 1960.
3. Parthasarathy, K. R., *Probability Measures on Metric Spaces,* Academic Press, New York, 1967.
4. Michael, E., "Topologies on Spaces of Subsets," *Trans. Am. Math. Soc.*, Vol. 71, 151–182, 1951.
5. Posner, E. C., and Rodemich, E. R., "Epsilon Entropy and Data Compression," to appear in *Annals Math. Stat.,* Vol. 42, 1971.