Correspondence

The Channel Capacity of a Certain Noisy Timing Channel

Ira S. Moskowitz, *Member, IEEE*, and Allen R. Miller

Abstract—**The effect of noise upon a simple covert timing channel is investigated. Shannon's information theory is used to quantify the resulting information flow across the channel. In particular, how a probabilistic response time to a query by the receiver affects the mutual information and channel capacity is studied. The channel capacity is expressed in terms of the critical probability for the mutual information function which is given in closed form in terms of Wright's hypergeometric function.**

Index Terms—**Channel capacity, covert channel, special functions.**

I. Introduction

We consider an n-user computer system, $n > 2$, where there are two users designated high and low. We assume that certain procedures have been set up so that low may not read high's files and high may not write its files to low. These are the no read up, no write down requirements of the Bell-LaPadula model [1]. However, it may be possible for high to covertly pass information to low over a communication channel that unintentionally, with respect to the system design, exists in the computer system. Such a means of communication is referred to as a covert channel. We are interested in the case where it is possible for high to interfere with the system response time to low's input. We will only be concerned with delays to low's input of a specific query designated by $\check{q}$.

In this correspondence, we do not propose methods of detecting timing channels or of giving specifications which prevent covert channels [2]-[4]. Instead, we continue in the spirit of Millen [5] by giving methods for quantifying the capacity of timing channels. In fact, the first systematic capacity analysis of timing channels can be found in Huskamp's dissertation [6]. The measurement of capacity is necessary for certain levels of "Orange Book" certification [7], which is of great importance to designers of secure systems. We present an idealized situation that we hope will lead to further system-dependent analysis of similar situations.

The communication between high (transmitter) and low (receiver) previously described is a covert timing channel, or more succinctly, a timing channel. We are taking Wray's [8] definition of a timing channel as a "covert channel whose alphabet is constructed from different time values." In [5], Millen discusses a simple timing channel where a reply takes one tick (normalized time unit) if high is not interfering with low, and two ticks if high is interfering. One tick tells the low user in Millen's scheme to interpret the message as the binary number **0** and two ticks as the binary number **1**. (We use bold face characters for the binary numbers to avoid confusion later.) Millen restricted his investigations to noiseless channels. In this correspondence, we obtain Millen's result as a special case.

The noise that we will be studying will not affect the value of the output. The noise will only affect the timing of the output, unlike in [4], where the timing was irrelevant, but the symbols being passed were the important feature. The noise effects in our model are envisioned as being due to time sharing delays of the CPU and I/O caused by many users contending for computing resources. We will refer to this as contention. Of course, it is the contention that causes the noise.

The users have an *a priori* knowledge only of the arrival times of the response to the query $\check{q}$ in the probabilistic sense given in Section II. This probabilistic arrival time is the effect from the noise in our system. A strategy must be developed that exploits this knowledge if high and low are to communicate in an efficient manner.

II. Mathematical Assumptions and Definitions

We shall use a modified exponential distribution to model the uncertainty in arrival times of signals to the low user, thus generalizing the noiseless model of Millen [5]. Suppose that low does its input query $\check{q}$ at time zero. In our noisy system the output will arrive via an exponential distribution starting one tick after $\check{q}$. If high is interfering with low, then the output will arrive via an exponential distribution starting two ticks after $\check{q}$. Again we are assuming that the responses to $\check{q}$ are identical. It is the times at which responses arrive that are different. Thus, we formalize these ideas with the following assumptions.

If high is not interfering with low, then the response time to $\check{q}$, inputted at time zero, is given by the random variable X_1 with probability density function

$$f_1(t) = \begin{cases} \lambda e^{-\lambda(t-1)}, & \text{if } t \geq 1, \\ 0, & \text{otherwise,} \end{cases}$$

and if high is interfering with low, then the response time to $\check{q}$ is given by the random variable X_2 with probability density function

$$f_2(t) = \begin{cases} \lambda e^{-\lambda(t-2)}, & \text{if } t \geq 2, \\ 0, & \text{otherwise.} \end{cases}$$

We model λ as being inversely related to the contention. The parameter λ can be adjusted to demonstrate different scenarios with regard to users contending for resources. Since the expectation of X_1 is $1 + 1/\lambda$, one could estimate λ by using system performance statistics related to mean response time. By letting $\lambda \to \infty$ we obtain the same situation that Millen set up. Later we will show how the channel matrix gives the exact relationship between noise and λ.

The lower λ, the lower the capacity; this is nothing new, noise reduces mutual information. Say, however, that we wish to allow timing channels that have a certain capacity. Thus it may be possible to measure the parameter λ, and if λ is too large, then the computer

Manuscript received April 29, 1991. This work was presented in part at the IEEE Computer Society Symposium on Research in Security and Privacy, Oakland, CA, May 1991.

I. S. Moskowitz is with the Information Technology Division, Code 5543, Naval Research Laboratory, Washington, DC 20375-5000.

A. R. Miller is with the Information Technology Division, Code 5570, Naval Research Laboratory, Washington, DC 20375-5000 and the Department of Mathematics, George Washington University, Washington, DC 20052.

IEEE Log Number 9108023.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JUL 1992		**00-00-1992 to 00-00-1992**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
The Channel Capacity of a Certain Noisy Timing Channel	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Research Laboratory,Information Technology Division,4555 Overlook Avenue, SW,Washington,DC,20375**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**6**	

itself could start up background processes to lower λ so that the capacity falls within an acceptable region. Without a way of quantifying the capacity, this could not be done effectively. This would allow a system to operate at a high level of efficiency and still stay within security guidelines.

Let κ represent the time that the output (response signal) arrives after $\check{q}$ is inputted. Without any restrictions we have that $1 \le \kappa < \infty$, which can lead to a situation where low has an infinite wait for an output to $\check{q}$. Further, let K be the random variable corresponding to κ. The distribution for K is obtained by conditioning on whether high is interfering (denoted by *Int*) or not interfering (denoted by *NoInt*) with low's response to $\check{q}$:

$$P(K \le t) = P(K \le t \mid NoInt)P(NoInt) + P(K \le t \mid Int)P(Int).$$

Notice that the conditional probability $P(K \le t \mid NoInt)$ is just $P(X_1 \le t)$ and $P(K \le t \mid Int) = P(X_2 \le t)$. High will interfere with low depending on whether high wishes to send a **0** or a **1** to low. We assign a probability of p whenever high sends a **0** (*NoInt*); therefore the probability that high will send a **1** (*Int*) is $1 - p$. Thus,

$$P(K \le t) = P(X_1 \le t)p + P(X_2 \le t)(1 - p).$$

Of course the way things stand now, high must have some feedback in order to know whether or not low received the output. Because of the probabilistic nature of the response time to $\check{q}$ we have an unbounded possible response time. Thus, we must make some adjustments in the strategy so that a feasible and realistic communication channel is set up between high and low. We will adopt two different but related strategies. Strategy 1 is the simpler of the two, but Strategy 2 is a more efficient use of the covert communication channel.

Strategy 1: Low will input $\check{q}$ every two ticks. High will interfere or do nothing. If $1 \le \kappa < 2$, then low will interpret the message as a **0**. If two ticks have gone by on low's clock, then low will automatically assume that the message is a **1** and issue an interrupt to its previous query $\check{q}$ before inputting its next query $\check{q}$.

The reason that low must issue an interrupt, if it has not yet received a response to $\check{q}$, is to prevent a response from "leaking" over into the next cycle of query and response. Say for example that low inputs $\check{q}$, two ticks go by and no response is given by the system, and then low again inputs $\check{q}$. How is low to know when it finally does receive a response if it is the response to the first $\check{q}$ or the second $\check{q}$? The issuance of an interrupt after two ticks will prevent this situation. We assume that the interrupt stops the response to $\check{q}$ from reaching the low user and that the interrupt acts instantaneously.

The problem with Strategy 1 is that every cycle takes two ticks and the high and low user are not making the most efficient use of their covert communication channel. The next strategy is a much more efficient use of the resources available.

Strategy 2: Low will input $\check{q}$ as soon as it has received its response from the previous query provided that a response comes in less than two ticks. If, after two ticks, no response has arrived at low, then low will automatically issue an interrupt to its previous query $\check{q}$ and issue its next query $\check{q}$. If $1 \le \kappa < 2$, then low will interpret the message as a **0**. If two ticks have gone by on low's clock, then low will automatically assume that the message is a **1**.

We are assuming that there is no time lag in low deciding, when necessary, to input $\check{q}$, and that the interrupts behave as described for Strategy 1. Ideas similar to the communication protocol in the above strategies are explored more fully in the work of Lee and Davidson [9] where they discuss deadlines in timed synchronous communication.

III. Transmission Errors

There are obvious transmission errors in our strategies which result in noise. The results in this section and the next one hold for both strategies. Let X be the random variable representing the input to the covert communication channel, i.e., the high user, and let Y represent the output random variable corresponding to low. The channel is a discrete memoryless channel.

Let $P(i \mid j)$ be the probability of an i being received by low given that a j was sent by high, where $i, j = \mathbf{0}, \mathbf{1}$. There are no errors if high sends a **1** since $2 \le \kappa$. The low user is watching its clock and as soon as two ticks have gone by, low interprets the message as a **1** which is correct. Thus, we have

$$P(\mathbf{1} \mid \mathbf{1}) = 1 \quad \text{and} \quad P(\mathbf{0} \mid \mathbf{1}) = 0. \tag{1}$$

However, if high wishes to send a **0**, then errors can be introduced. If the output arrives before two ticks have elapsed there is no transmission error. However, if because of contention $2 \le \kappa$, then we do have an error because low will interpret the message as a **1** when it is in fact a **0**. The probability of a **0** being sent and a **0** being received is

$$P(\mathbf{0} \mid \mathbf{0}) = \int_1^2 \lambda e^{-\lambda(t-1)}\,dt = 1 - e^{-\lambda}. \tag{2}$$

Further, the probability of a **0** being sent and a **1** being received is

$$P(\mathbf{1} \mid \mathbf{0}) = \int_2^\infty \lambda e^{-\lambda(t-1)}\,dt = e^{-\lambda}. \tag{3}$$

The channel capacity of the covert timing channel will be calculated in Section IV by using (1), (2), and (3).

IV. Capacity Analysis of Strategy 1

For now we are only trying to calculate the information flow in units of bits per symbol. For Strategy 1, the difference between bits per symbol and bits per tick is trivial, i.e., a factor of 1/2. However, for Strategy 2 there is a substantial difference and we will address this issue in Section V.

As soon as the low user inputs its query $\check{q}$, high inputs either a **0** or a **1**. A **0** corresponds to no interference and a **1** corresponds to interference. The response to $\check{q}$ is always the same for it is the time at which this response arrives that determines the symbol being passed over the channel. If the response arrives between one and two ticks, but not equal to two ticks, Y is set equal to **0**. If the response has not yet arrived at two ticks, or arrives at exactly two ticks, then Y is set equal to **1**. The channel matrix from (1), (2), and (3) is given by

$$\begin{pmatrix} P(\mathbf{0} \mid \mathbf{0}) & P(\mathbf{1} \mid \mathbf{0}) \\ P(\mathbf{0} \mid \mathbf{1}) & P(\mathbf{1} \mid \mathbf{1}) \end{pmatrix} = \begin{pmatrix} 1 - e^{-\lambda} & e^{-\lambda} \\ 0 & 1 \end{pmatrix} \tag{4}$$

and shows how λ influences noise in the communication channel. Let $I(X, Y)$ and C be respectively the mutual information between X and Y and the channel capacity, both of which have units in bits per symbol. The mutual information $I(X, Y)$ expressed as a function of p is given by

$$I(p) = -p \log p + e^{-\lambda} p \log\left(e^{-\lambda} p\right) - \left(1 - p + e^{-\lambda} p\right) \log\left(1 - p + e^{-\lambda} p\right), \tag{5}$$

where logarithms are computed using base 2. We have obtained (5) by calculating the mutual information

$$I(X, Y) = H(X) - H_Y(X)$$

as the difference between the input entropy and the equivocation [10]. The capacity for this channel is the maximum of $I(p)$ with respect to p. Since the mutual information function $I(p)$ is concave down [11, Theorem 5.2.5] with respect to the variable p, it suffices to find the critical point ζ determined by the equation $I'(p) = 0$. Thus, from (5), we have

$$I'(p) = -\log p + e^{-\lambda}\log\left(e^{-\lambda}p\right) + \left(1 - e^{-\lambda}\right)\log\left(1 - p + e^{-\lambda}p\right);$$

so that the critical point is given by

$$\zeta = \frac{1}{1 + e^{\lambda/(e^{\lambda}-1)} - e^{-\lambda}}.$$

Since both $e^{-\lambda} \to 0$ and $\lambda/(e^{\lambda} - 1) \to 0$ as $\lambda \to \infty$ we see that

$$\lim_{\lambda\to\infty} \zeta = 1/2.$$

The capacity is the mutual information function evaluated at ζ, thus

$$C(\lambda) = -\zeta\log\zeta + e^{-\lambda}\zeta\log\left(e^{-\lambda}\zeta\right) - \left(1 - \zeta + e^{-\lambda}\zeta\right)\log\left(1 - \zeta + e^{-\lambda}\zeta\right).$$

The critical point ζ quickly becomes asymptotic to $1/2$. Thus, for $\lambda \gg 0$, $I(p)$ is nearly optimized for an input probability distribution where both **0**'s and **1**'s are sent with equal probabilities of $1/2$. Numerical calculations [12] show that $C(\lambda) - I(1/2)$ is small and quickly approaches zero as $\lambda \to \infty$. This is not surprising in light of a recent result of Majani and Rumsey [13] that for a binary-input discrete memoryless channel, $I(1/2)$ is at least 94.21% of the capacity.

V. Capacity Analysis for Strategy 2

If we do a bit per symbol analysis of both strategies, they are identical. However, if we do a bit per tick analysis they are quite different. This is due to the fact that low will issue its next query $\check{q}$ as soon as it has received a response form its last query, provided that no more than two ticks have elapsed from the issuance of the former query. If $E[T]$ is the average time it takes to send a symbol across the channel, then the mutual information of a discrete memoryless channel, in bits per tick, is defined by

$$I_t \equiv \frac{I(X, Y)}{E[T]}. \tag{6}$$

Here we are using the notational convenience that the subscript t means units are given in bits per tick.

It would seem natural to try to maximize I_t to get the actual channel capacity in units of bits per tick. Verdú [14, Theorem 2] studied the capacity in units of bits per unit cost C_u of a memoryless (stationary) channel. Let $b[X]$ be the cost function associated with the input random variable X. Then Verdú's theorem states that

$$C_u = \sup_X \frac{I(X, Y)}{E[b[X]]}, \tag{7}$$

where the supremum is taken over different probability measures for X with the alphabet of X fixed.

Of course $E[b[X]]$, the expected value of $b[X]$, is given in units of unit cost per symbol. If the cost function is the time it takes to send symbols, then we can replace $E[b[X]]$ by $E[T]$. Combining (6) and (7) we can express the capacity per unit time as the supremum of the mutual information per unit time:

$$C_t = \sup_X \frac{I(X, Y)}{E[T]} = \sup_X I_t. \tag{8}$$

Note that the optimizing process over X involves $I(X, Y)$ and $E[T]$ simultaneously. Obviously one would not want to code the message by just minimizing time because we would lose information by not using enough different symbols.

The actual distribution of the query response random variable T is governed by the distributions X_1, X_2, and Strategy 2. For time values less than one tick or greater than two ticks the probability density function $\hat{f}(t)$ of T is zero since the response can never arrive at those times. For time values greater than or equal to one tick and strictly less than two ticks the behavior of T is governed by $f_1(t)$.

In order to obtain the probability density function $\hat{f}(t)$, we calculate the derivative of the associated cumulative distribution function $F(t) = P(T \le t)$. Hence, $F'(t) = \hat{f}(t)$ and by conditioning we see that

$$P(T \le t) = P(T \le t \mid \mathbf{0})P(\mathbf{0}) + P(T \le t \mid \mathbf{1})P(\mathbf{1}).$$

Now if $1 \le t < 2$,

$$\begin{aligned} P(1 \le T \le t) &= P(1 \le T \le t \mid \mathbf{0})P(\mathbf{0}) + P(1 \le T \le t \mid \mathbf{1})P(\mathbf{1}) \\ &= P(1 \le X_1 \le t)p + P(1 \le X_2 \le t)(1 - p) \\ &= p\int_1^t \lambda e^{-\lambda(\vartheta-1)}\,d\vartheta; \end{aligned}$$

and, since two ticks is the cut-off time,

$$\begin{aligned} P(T = 2) &= p\int_2^{\infty} \lambda e^{-\lambda(\vartheta-1)}\,d\vartheta + (1 - p)\int_2^{\infty} \lambda e^{-\lambda(\vartheta-2)}\,d\vartheta \\ &= e^{-\lambda}p + 1 - p. \end{aligned}$$

Therefore, we have

$$P(T > 2) = 0,$$

$$P(T = 2) = e^{-\lambda}p + 1 - p,$$

$$P(1 < T \le t) = p\int_1^t \lambda e^{-\lambda(\vartheta-1)}\,d\vartheta, \qquad 1 < t < 2,$$

$$P(T \le 1) = 0.$$

Hence, the density function of T is given by

$$\hat{f}(t) = \delta(t - 2)\left[pe^{-\lambda} + 1 - p\right] + \lambda e^{-\lambda(t-1)}p\chi_{[1,2)}(t),$$

where $\delta(\cdot)$ is the Dirac delta function and $\chi_{[1,2)}(\cdot)$ is the characteristic (or indicator) function of the interval $[1, 2)$. To find the expected value of T, since

$$E[T] = \int_{-\infty}^{\infty} t\hat{f}(t)\,dt,$$

we see that

$$E[T] = \int_{-\infty}^{\infty} \left(t\delta(t - 2)\left[pe^{-\lambda} + 1 - p\right]\right)dt + p\int_1^2 t\lambda e^{-\lambda(t-1)}\,dt,$$

and on performing the two integrations we obtain

$$E[T] = 2(pe^{-\lambda} + 1 - p) + p(1 - e^{-\lambda})\left(\frac{1}{\lambda} + \frac{e^{\lambda} - 2}{e^{\lambda} - 1}\right). \quad (9)$$

When λ is infinite there is no contention and hence the channel is noiseless. In this case, the channel matrix (4) becomes

$$\begin{pmatrix} P(\mathbf{0} \mid \mathbf{0}) & P(\mathbf{1} \mid \mathbf{0}) \\ P(\mathbf{0} \mid \mathbf{1}) & P(\mathbf{1} \mid \mathbf{1}) \end{pmatrix}_{\lambda = \infty} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Further, from (9) the expectation is given by

$$E[T]_{\lambda = \infty} = 2 - p.$$

When the contention is maximized, λ is zero. Therefore, the channel matrix (4) becomes

$$\begin{pmatrix} P(\mathbf{0} \mid \mathbf{0}) & P(\mathbf{1} \mid \mathbf{0}) \\ P(\mathbf{0} \mid \mathbf{1}) & P(\mathbf{1} \mid \mathbf{1}) \end{pmatrix}_{\lambda = 0} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

Thus, we see that low cannot infer at all whether high sent a **0** or a **1**. In fact, low will only receive the symbol **1**. Applying L'Hôpital's rule twice shows that the last term of (9) is $3/2$ as $\lambda \to 0$, so that

$$E[T]_{\lambda = 0} = 2.$$

We now express the mutual information in terms of bits per tick. From (5), (6), and (9) we get

$$I_t(p) = \frac{-p \log p + e^{-\lambda} p \log\left(e^{-\lambda} p\right) - \left(1 - p + e^{-\lambda} p\right) \log\left(1 - p + e^{-\lambda} p\right)}{2(pe^{-\lambda} + 1 - p) + p(1 - e^{-\lambda})\left(\frac{1}{\lambda} + \frac{e^{\lambda} - 2}{e^{\lambda} - 1}\right)} \quad (10)$$

and from (8) the channel capacity as a function of λ is given by

$$C_t(\lambda) = \sup_X \frac{-p \log p + e^{-\lambda} p \log\left(e^{-\lambda} p\right) - \left(1 - p + e^{-\lambda} p\right) \log\left(1 - p + e^{-\lambda} p\right)}{2(pe^{-\lambda} + 1 - p) + p(1 - e^{-\lambda})\left(\frac{1}{\lambda} + \frac{e^{\lambda} - 2}{e^{\lambda} - 1}\right)}.$$

When λ is equal to zero or ∞, define $C_t(\lambda)$ by its limiting values. Thus, $C_t(\lambda)$ is a continuous function for $\lambda \in [0, \infty]$ and we may write

$$C_t(\lambda) = \max_{p \in [0,1]} \frac{-p \log p + e^{-\lambda} p \log\left(e^{-\lambda} p\right) - \left(1 - p + e^{-\lambda} p\right) \log\left(1 - p + e^{-\lambda} p\right)}{2(pe^{-\lambda} + 1 - p) + p(1 - e^{-\lambda})\left(\frac{1}{\lambda} + \frac{e^{\lambda} - 2}{e^{\lambda} - 1}\right)}.$$

VI. Exact Result for the Channel Capacity of Strategy 2

Since $I_t(p)$, given by (10), is a nonnegative differentiable function for $p \in (0, 1)$ and its values are zero at the boundary of the interval, it suffices to find a unique critical point, $p_c \in (0, 1)$, for then we know that $I_t(p_c) = C_t(\lambda)$.

Taking the derivative of $I_t(p)$ with respect to p and setting it equal to zero, we arrive at (after some algebraic simplification)

$$x \ln x + (y - u) \ln(1 - yp) - y \ln p = 0, \quad (11)$$

where

$$x \equiv e^{-\lambda}, \qquad y \equiv 1 - x, \qquad 2u \equiv 1 + y/\ln x.$$

Exponentiating both sides of (11) and setting

$$\eta \equiv \frac{y}{y - u}, \qquad r \equiv x^{-x/y} p,$$

we obtain

$$r^{\eta} + (x^x y^y)^{1/y} r - 1 = 0. \quad (12)$$

We recall the Wright function [15] defined by

$${}_1\Psi_1\left[\begin{matrix}(\alpha, A); \\ (\beta, B);\end{matrix} z\right] \equiv \sum_{k=0}^{\infty} \frac{\Gamma(\alpha + Ak)}{\Gamma(\beta + Bk)} \frac{z^k}{k!}$$

and the definition of the Pochhammer symbol

$$(\lambda)_n \equiv \frac{\Gamma(\lambda + n)}{\Gamma(\lambda)},$$

where $\Gamma(z)$ is the Gamma function and n is an integer. For conciseness in what follows we shall write Ψ for ${}_1\Psi_1$.

In [16], Miller showed that Mellin's result [17] concerning the roots of trinomial equations could be extended to include positive non-integer exponents as well. In particular, we have the following.

For $\omega > 1$, the unique positive root of the transcendental equation

$$\xi^{\omega} + \mu\xi - 1 = 0$$

is given by

$$\xi = \frac{1}{\omega} \Psi\left[\begin{matrix}\left(\frac{1}{\omega}, \frac{1}{\omega}\right); \\ \left(\frac{1}{\omega} + 1, \frac{1}{\omega} - 1\right);\end{matrix} -\mu\right]$$

provided that

$$|\mu| < \omega/(\omega - 1)^{1 - 1/\omega}. \quad (13)$$

To apply this result to (12) we must verify that the inequality (13) is satisfied. Considering η as a function of $y \in [0, 1]$, it is easy to show that $4/3 \le \eta \le 2$. Further, since

$$(x^x y^y)^{1/y} \le 1, \qquad x, y \in [0, 1],$$

and

$$\min_{4/3 \le \eta \le 2} (\eta - 1)^{1/\eta - 1} \eta \approx 1.755 > 1,$$

we see that

$$(x^x y^y)^{1/y} < (\eta - 1)^{1/\eta - 1} \eta, \qquad x, y \in [0, 1].$$

Therefore, the previous result may be applied and we arrive at the following.

The channel capacity for $0 < \lambda < \infty$ is

$$C_t(\lambda) = \frac{-p_c \log p_c + e^{-\lambda} p_c \log\left(e^{-\lambda} p_c\right) - \left(1 - p_c + e^{-\lambda} p_c\right) \log\left(1 - p_c + e^{-\lambda} p_c\right)}{2(p_c e^{-\lambda} + 1 - p_c) + p_c(1 - e^{-\lambda})\left(\frac{1}{\lambda} + \frac{e^{\lambda} - 2}{e^{\lambda} - 1}\right)},$$

where the critical probability for the mutual information function $I_t(p)$, $p \in (0, 1)$, is given by

$$p_c = \frac{x^{x/y}}{\eta} \Psi\left[\begin{matrix}\left(\frac{1}{\eta}, \frac{1}{\eta}\right); \\ \left(\frac{1}{\eta} + 1, \frac{1}{\eta} - 1\right);\end{matrix} -yx^{x/y}\right]. \quad (14)$$

Let us consider the two boundary cases. When $\lambda = 0$, there is infinite noise and (10) is identically equal to zero. Therefore, the channel capacity is zero and the critical probability is of no concern. When $\lambda = \infty$, there is no noise which is the situation that Millen studied.

Millen used Shannon's [10] approach that employed finite-difference equations to show that

$$C_t(\infty) = \log\left(\frac{1+\sqrt{5}}{2}\right). \tag{15}$$

We will show this by analyzing the limiting behavior of (14) as $\lambda \to \infty$. In this case, we have

$$p_c = \frac{1}{2}\Psi\left[\begin{matrix}\left(\frac{1}{2},\frac{1}{2}\right);\\ \left(\frac{3}{2},-\frac{1}{2}\right);\end{matrix} -1\right];$$

so that

$$\begin{aligned} 2p_c &= \sum_{k=0}^{\infty} \frac{\Gamma\left(\frac{1}{2}+\frac{k}{2}\right)}{\Gamma\left(\frac{3}{2}-\frac{k}{2}\right)} \frac{(-1)^k}{k!} \\ &= \sum_{k=0}^{\infty} \frac{\Gamma\left(\frac{1}{2}+\frac{2k}{2}\right)}{\left(\frac{3}{2}-\frac{2k}{2}\right)} \frac{(-1)^{2k}}{(2k)!} \\ &\quad + \sum_{k=0}^{\infty} \frac{\Gamma\left(\frac{1}{2}+\frac{2k+1}{2}\right)}{\Gamma\left(\frac{3}{2}-\frac{2k+1}{2}\right)} \frac{(-1)^{2k+1}}{(2k+1)!} \\ &= \sum_{k=0}^{\infty} \frac{\Gamma\left(\frac{1}{2}+k\right)}{\Gamma\left(\frac{3}{2}-k\right)} \frac{1}{(1)_{2k}} + \sum_{k=0}^{\infty} \frac{\Gamma(1+k)}{\Gamma(1-k)} \frac{-1}{(1)_{2k+1}}. \end{aligned}$$

Since $1/\Gamma(1-k)$ vanishes for $k \geq 1$,

$$\left(\frac{3}{2}\right)_{-k} = \frac{(-1)^k}{(-1/2)_k} \quad \text{and} \quad (1)_{2k} = \left(\frac{1}{2}\right)_k k! 2^{2k},$$

we arrive at

$$\frac{2p_c+1}{2} = {}_1F_0\left[-1/2; -; -1/4\right],$$

where ${}_1F_0$ is a generalized hypergeometric function. Since

$${}_1F_0\left[a; -; z\right] = (1-z)^{-a},$$

we have that

$$p_c = \frac{-1+\sqrt{5}}{2};$$

and now after some algebraic manipulation we deduce

$$I_t\left(\frac{-1+\sqrt{5}}{2}\right) = \log\left(\frac{1+\sqrt{5}}{2}\right).$$

Thus, we have obtained Millen's result (15) as a special case. As we did for Strategy 1, numerical calculations also show that for Strategy 2 the channel capacity C_t and the mutual information I_t evaluated at the limiting value of $(-1+\sqrt{5})/2$ are quite close for all values of λ.

VII. Conclusion and Directions for Future Research

We have shown how to incorporate noise into the capacity calculations of certain timing channels. Strategy 1 is rather simplistic, but it is a necessary step to understanding Strategy 2. In addition, Strategy 1 is useful if there is no feedback to high.

In future work, we shall relax the restriction that the responses arrive at 1 or 2 ticks and allow variable response times. We can always normalize the lesser time value to 1 tick so we will investigate the situation where the responses arrive at 1 or β ticks, β being variable. The noise in the channel decreases with increasing β but the time required to send the symbol **1** across the channel increases. Thus, we have an optimization problem for the capacity with respect to β.

At present, in Strategy 2 we allow low to instantaneously interrupt its query. An interesting alternate scenario would permit some delay in the interrupt. Also, the necessity of an interrupt could be mitigated by using a series of distinct queries whose responses are known and inputted in a cyclically repeating order.

Acknowledgment

The authors wish to thank the anonymous referees for their helpful suggestions regarding future areas of research.

References

[1] D. E. Bell and L. J. LaPadula, Secure Computer System: Unified Exposition and Multics Interpretation, MTR-2997, MITRE Corp., Bedford. MA, Mar. 1976.

[2] James W. Gray, III, "Probabilistic interference," in *Proc. IEEE Comput. Soc. Symp. Res. in Security and Privacy*, Oakland, CA, May 1990, pp. 170–179.

[3] J. McLean, "Security models and information flow," in *Proc. IEEE Comput. Soc. Symp. Res. in Security and Privacy*, Oakland, CA, May 1990, pp. 180–187.

[4] I. S. Moskowitz, "Quotient states and probabilistic channels," in *Proc. Comput. Security Foundations Workshop III*, Franconia, NH, June 1990, pp. 74–83.

[5] J. K. Millen, "Finite-state noiseless covert channels," in *Proc. Comput. Security Foundations Workshop II*, Franconia, NH, June 1989, pp. 81–86.

[6] J. C. Huskamp, "Covert communication channels in timesharing systems," Ph.D. thesis, Univ. of California, Berkeley, CA 1978; also tech. rep. UCB-CS-78-02 and Electron. Res. Lab. Memo. No. ERL-M78/37.

[7] Dept. of Defense, Nat. Comput. Security Center, Trusted Comput. Syst. Evaluation Criteria 5200.28-STD, Dec. 1985.

[8] J. C. Wray, "A methodology for the detection of timing channels," *Cipher, Newsletter IEEE Comput. Soc. Technical Committee on Security and Privacy*, Winter 1991.

[9] I. Lee and S. B. Davidson, "A performance analysis of timed synchronous communication primitives," *IEEE Trans. Comput.*, vol. 39, pp. 1117–1131, Sept. 1990.

[10] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press, 1949.

[11] R. E. Blahut, *Principles and Practice of Information Theory*. Reading, MA: Addison-Wesley, 1987.

[12] I. S. Moskowitz, "Variable noise effects upon a simple timing channel," in *Proc. IEEE Comput. Soc. Symp. Res. in Security and Privacy*, Oakland, CA, May 1991, 362–372.

[13] E. E. Majani and H. Rumsey, "Two results on the capacity of binary-input discrete memoryless channels," *IEEE Trans. Inform. Theory*, to appear.

[14] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1019–1030, Sept. 1990.

[15] H. M. Srivastava and H. L. Manocha, *A Treatise on Generating Functions*. New York: Wiley/Halsted, 1984.

[16] A. R. Miller, "Solutions of Fermat's last equation in terms of

Wright's hypergeometric function," *Fibonacci Quarterly*, vol. 29, pp. 52-56, Feb. 1991.
[17] H. Hochstadt, *The Functions of Mathematical Physics*. New York: Wiley-Interscience, 1971.

On the Capacity Region of the Discrete Additive Multiple-Access Arbitrarily Varying Channel

John A. Gubner, *Member, IEEE*

Abstract—**The discrete additive multiple-access arbitrarily varying channel (AVC) with two senders and one receiver is considered. Necessary and sufficient conditions are given for its deterministic-code average-probability-of-error capacity region under a state constraint to have a nonempty interior. In the case that no state constraint is present, the capacity region is characterized exactly. In the case of the noiseless mod-2 adder AVC using state constraint function $l(s) = s$ and subject to a state constraint L less than or equal to 0.13616917, the capacity region is shown to be a 45-degree triangle whose legs have length $1 - h(L)$, where h denotes the binary entropy function.**

Index Terms—**Additive channel, multiple-access, arbitrarily varying channel, state constraint, capacity region.**

I. Introduction

A general multiple-access arbitrarily varying channel (AVC) with two senders and one receiver is a transition probability W from $\mathsf{X} \times \mathsf{Y} \times \mathsf{S}$ into Z, where X, Y, S, and Z are finite sets, each containing at least two elements. We interpret $W(z \mid x, y, s)$ as the conditional probability that the channel output is $z \in \mathsf{Z}$ given that the channel input symbol from sender 1 is $x \in \mathsf{X}$, the channel input symbol from sender 2 is $y \in \mathsf{Y}$, and that the channel *state* is $s \in \mathsf{S}$. When block codes of length n are used, we say the AVC is subject to *state constraint* L if the state-selection mechanism can generate only those state sequences $s = (s_1, \cdots, s_n)$ that satisfy a time-average constraint of the form

$$\frac{1}{n} \sum_{k=1}^{n} l(s_k) \le L, \tag{1}$$

where l is a given nonnegative constraint function defined on S and satisfying $\min_s l(s) = 0$. Note that if $L \ge \max_s l(s)$, then *all* state sequences s satisfy (1); in this case we say that the state constraint is not present, or inactive.

Definition (Additive AVC): Let G be a finite nontrivial commutative group. Suppose that $\mathsf{X} = \mathsf{Y} = \mathsf{Z} = \mathsf{G}$. We say that W is an additive AVC if

$$W(z \mid x, y, s) = V_s(z - x - y),$$

for some transition probability V from S into G.

General multiple-access AVC's subject to a state constraint have been studied in [6]. There, both forward and converse results were proved that enable one to give inner and outer bounds on the capacity region. To obtain meaningful inner bounds, one must exhibit input probability distributions for which certain inequalities are nonvacuous. We show that for the additive AVC such input distributions always exist.

In the absence of state constraints, we exactly characterize the capacity region of the additive AVC.

In the special case of the noiseless mod-2 adder AVC with $l(s) = s$ and state constraint $L \le 0.13616917$, the capacity region is shown to be a 45° triangle whose legs have length $1 - h(L)$, where h denotes the binary entropy function defined in Theorem 3.

Additive AVC's with one sender and one receiver were considered in [4, Section V], but under the assumption that the channel symbols come from a finite subset of $\mathbb{R}^d$ rather than a finite commutative group G. This is in contrast to the results of [4, Section IV] concerning a restricted form of additive AVC called a group adder AVC, which is an additive AVC for which $\mathsf{S} = \mathsf{G}$ and $V_s(t) = \mu(t - s)$ for some probability distribution μ on G. In an earlier paper [3, Section IV] Csiszár and Narayan analyzed the single-user noiseless mod-2 adder AVC.

II. Statement of Results

In order to state our results, we need the following notation. Let $\mathscr{P}(\mathsf{S})$ denote the set of probability distributions on S. For $r \in \mathscr{P}(\mathsf{S})$, let rV denote the distribution on G defined by $(rV)(t) = \sum_s r(s) V_s(t)$. Let $H(rV)$ denote the entropy of rV. Let

$$\mathscr{P}^L(\mathsf{S}) \triangleq \left\{ r \in \mathscr{P}(\mathsf{S}) : \sum_{s \in \mathsf{S}} l(s) r(s) \le L \right\}.$$

Note that if $L \ge \max_s l(s)$, then $\mathscr{P}^L(\mathsf{S}) = \mathscr{P}(\mathsf{S})$. We now state our main results.

Theorem 1: The deterministic-code average-probability-of-error capacity region under state constraint L of an additive multiple-access AVC V has a nonempty interior, if and only if there is no $r \in \mathscr{P}^L(\mathsf{S})$ such that rV is the uniform distribution on G. Furthermore, the capacity region is always contained in the 45° triangle,

$$\left\{ (R_1, R_2) : R_1 \ge 0, R_2 \ge 0, \right.$$
$$\left. \text{and } R_1 + R_2 \le \log |\mathsf{G}| - \max_{r \in \mathscr{P}^L(\mathsf{S})} H(rV) \right\}, \tag{2}$$

where $|\mathsf{G}|$ denotes the cardinality of the set G.

Remark: Since $\mathscr{P}^L(\mathsf{S})$ is compact and since H is continuous,

$$\log |\mathsf{G}| > \max_{r \in \mathscr{P}^L(\mathsf{S})} H(rV), \tag{3}$$

if and only if there is no $r \in \mathscr{P}^L(\mathsf{S})$ such that rV is the uniform distribution on G.

Theorem 2: In the absence of state constraints, the capacity region of the additive multiple-access AVC V is always given by (2), where $\mathscr{P}^L(\mathsf{S})$ is replaced by $\mathscr{P}(\mathsf{S})$.

Proof: Theorem 2 follows from Theorem 1, the preceding Remark, ([7, Theorem 1, p. 214], which says that if the *deterministic-code* average-probability-of-error capacity region has a nonempty interior, then it is equal to the *random-code* average-probability-of-error capacity region), and [6, Section IV], which shows that the random-code average-probability-of-error capacity region of the additive AVC is given by (2). We give an independent proof in Section V. □

Manuscript received November 6, 1990; revised December 20, 1991. This work was supported in part by the Air Force Office of Scientific Research under Grant AFOSR-90-0181. This work was presented in part at the IEEE International Symposium on Information Theory, Budapest, Hungary, June 24-28, 1991.

The author is with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, 1415 Johnson Drive, Madison, WI 53706-1691.

IEEE Log Number 9108024.