

A Detailed Mathematical Analysis of a Class of Covert Channels Arising in Certain Anonymizing Networks*

Ira S. Moskowitz,¹ Richard E. Newman,²
Daniel P. Crepeau,¹ and Allen R. Miller³

¹ Information Technology Division, Code 5500
Naval Research Laboratory
Washington, DC 20375
USA

moskowitz@itd.nrl.navy.mil
dcrepeau@itd.nrl.navy.mil

&

² CISE Department
University of Florida
Gainesville, FL 32611-6120
USA

nemo@cise.ufl.edu

&

³ Washington, DC
USA

Abstract. There have long been threads of investigation into covert channels, and threads of investigation into anonymity, but these two closely related areas of information hiding have not been directly associated. This paper represents an initial inquiry into the relationship between covert channel capacity and anonymity, and poses more questions than it answers. Even this preliminary work has proven difficult, but in this investigation lies the hope of a deeper understanding of the nature of both areas. MIXes have been used for anonymity, where the concern is shielding the identity of the sender or the receiver of a message, or both. Traffic analysis prevention (TAP) methods are used to conceal larger traffic patterns. Here, we are concerned with how much information a sender to a MIX can leak to an eavesdropping outsider, despite the concealment efforts of MIXes acting as firewalls.

Introduction

Traffic analysis in network communication can be used to open a covert channel from Alice to Eve [12, 13, 23–25]. In this paper we discuss a particular covert channel that exists in an anonymizing network. We present some simplified scenarios as a first step in this analysis.

* Research supported by the Office of Naval Research.

There is always one special transmitting node in a network called *Alice*. Alice and possibly other transmitters have legitimate business transmitting messages to a set of Receivers $\{R_i | i = 1, 2, \dots, M\}$. These transmitters act completely independently of one another, and have no direct knowledge of each other's recent transmission behavior. Alice may have some general knowledge of the long-term traffic levels produced by the other transmitters, e.g., the number of other transmitters and their probabilistic behavior, which can allow Alice to write a code that can improve the covert communication channel's data rate. She cannot, however, perform short-term adaptation to their behavior. Our simplified communication is one-way (the receivers never send to Alice or to the other transmitters). We also assume that there is a clock, and that transmissions only occur in the unit interval of time called a *tick*. Any subset of transmitters can each either send a single message to a single receiver in a tick, or not send a message at all. Each transmitter in a tick can send to a different receiver, and two or more transmitters may send to the same receiver in the same tick. All messages' contents are encrypted end-to-end.

There is also an eavesdropper on the network called *Eve*. Since all transmissions are encrypted, they appear to the eavesdropper Eve as having indistinguishable content. Eve may be either a global passive adversary (GPA), with the ability to see link traffic on every link in the network, or a restricted passive adversary (RPA), with the ability to observe traffic only on certain links.

Alice is not allowed any direct communication with Eve. However, Alice can influence what Eve sees on the network. We present several different scenarios and analyze the subtle ways by which Alice may indirectly communicate with Eve. In particular, we study network scenarios that attempt to achieve a degree of anonymity with respect to the network communication. That is, the networks are designed with various anonymity devices to prevent Eve from learning who is sending a message to whom. Even if a certain degree of anonymity is achieved, it still may be possible for Alice to communicate covertly with Eve. Please keep in mind that anonymous communication networks were *not* designed with this covert channel threat in mind. Rather, it was our study of these anonymity networks that caused us to realize that even in what appears to be a benign form of communication, information may still leak out of the network, contrary to the intent of system design.

The main thrust of this paper is to analyze the situation where there are two enclaves, communication between them is encrypted, and packets are sent only from the first enclave (which contains Alice) to the second (please refer to Figure 1). Eve is able to monitor the communication from the first enclave to the second. Anonymity is "achieved" in that an eavesdropper such as Eve (as RPA) does not know who is sending a message (that is hidden inside of the first enclave) and nor who is receiving the message (this can only be known if one is interior to the second enclave). Eve is only allowed to know how many messages per tick travel from the first enclave to the second. Nonetheless, Alice attempts to communicate covertly with Eve.

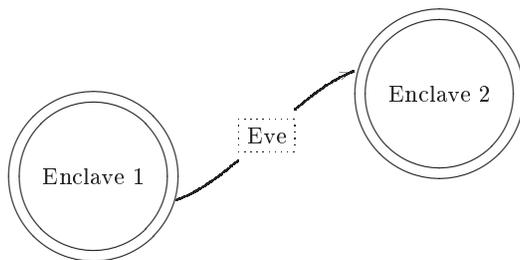


Fig. 1. Restricted Passive Adversary Model.

This paper analyzes the covert communication channel from Alice to Eve. We show that even if anonymity is taken into consideration with respect to system design, covert channels may remain. As a baseline, we first consider situations in which no attempt at anonymity has been made (only encryption of the messages, so that they all appear to be identical to an eavesdropper). Later, we will consider covert channel capacity in networks with the stronger anonymity controls just described. This paper concludes with a summary and some directions for future research.

1 Base Scenario — No anonymity

One transmitter

Alice is the only transmitter, and there are M possible receivers. Eve has knowledge of the network traffic (Eve is a GPA — see Figure 2). The only properties that Eve can discern from a message is its source (trivially Alice) and its destination. Alice can use that fact to send information covertly to Eve. In this simplistic scenario Eve can see if Alice is sending a message, and if Alice is sending a message Eve can determine for which receiver the message is meant. This gives Alice the ability to signal Eve with an alphabet of $M + 1$ symbols: M symbols for the M different receivers, and one symbol (“0”) for the choice of not sending a message.

Since nothing is able to interfere with Alice’s transmission, we have a noiseless discrete memoryless channel (DMC) modeling the covert channel, whose capacity is $\log(M + 1)$ bits per tick.¹

Several transmitters

Now, if there are other transmitters aside from Alice, but their transmissions to any of the M receivers do not affect Alice’s transmissions, then the covert channel from Alice to Eve is as above. This would be the case if the links into a receiver can handle all of the traffic meant for them. Of course, if the link

¹ All logarithms are base 2, and we will also adopt the convenience of no longer stating the units of the capacity. The units will be understood to be bits per tick.

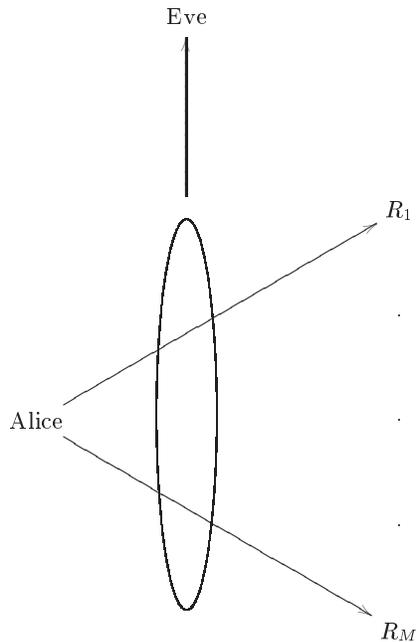


Fig. 2. Global Passive Adversary Model.

capacity into a transmitter *does* affect the number of receivable transmissions then that introduces noise into the channel and the capacity is obviously less than $\log(M + 1)$. This is a course of research worth pursuit.

Anonymity discussion

In the above scenario Alice can obviously leak considerable information to Eve. This is no secret to the anonymity community, *e.g.*, [1–4, 14, 15, 18, 6, 20] (while the preceding list is only a representative sample of papers/URLs on the topic, these papers relate particularly well to what we discuss in this paper). However, in the past the concerns have focused on retaining or regaining anonymity. It is the “anonymity lost” that we exploit for covert communication. If there were “*perfect*” anonymity,² then we would not expect to find a covert channel.

To provide anonymity, transmissions from a transmitter are often first sent to an intermediary, such as a MIX [4] or an onion router [14], before they are forwarded to the receiver. This has the effect of hiding whither the message is going. Thus, these intermediaries serve to anonymize the transmission. Of course, Eve still knows the set of those who receive a message, and she also knows the set of those who sent a message, but she does not know who sent a message

² We intentionally leave the notion of perfect anonymity as fuzzy in this paper. We ask the reader though the somewhat circular question: If we did have perfect anonymity, how could we have covert communication?

to whom. It is interesting that, even when we seem to have “good” statistical anonymity, Alice may still non-trivially be able to communicate covertly with Eve.

The use of a MIX alone does not prevent Alice from covert communication with Eve. In fact there are two possible situations.

1. Alice signals Eve by sending or not sending a message. A MIX alone does nothing to prevent Eve from learning this information (this is not what a MIX is designed to do). We discuss this further at the beginning of the next section. Therefore Alice has a noiseless channel to Eve, with a capacity of one.
2. Alice signals Eve by sending a message to any one of M different receivers. If Alice is the only transmitter, Eve simply sees where messages are going when they leave the MIX (a concern well-known to MIX designers). This allows a covert channel with a capacity of $\log(M + 1)$. If there are other users, their behavior affects what Eve is receiving and the capacity is then less than $\log(M + 1)$.

We will not study the latter situation in this paper, because we do not use pure MIXes. Instead, we use MIXes acting as firewalls.

2 Scenario 2: Indistinguishable Receivers—Two MIX-firewalls

Consider the situation in which every message goes into the anonymizing intermediary referred to as a MIX [4]. The MIX has the effect of hiding the “linking” knowledge of which transmission is sent to which receiver. In other words, Eve knows who is transmitting and who is receiving, but in general, Eve does not know which transmitter is sending to which receiver. This assumes that Eve is a GPA. Of course, if only one transmitter is operating then the MIX hides nothing. In other words the MIX gives statistical anonymity. The amount of anonymity has been measured as the log of the number of transmitters (*anonymity set size*), sometimes in conjunction with probabilistic behavior (e.g., [2–4, 6, 20]).

The main concern of this paper is not with measuring anonymity, rather it is the amount of covert information that may be leaked through less than perfect anonymity. However, we do note the very important observation from our research: *the ability to covertly communicate arises due to a lack of anonymity*. As the number of transmitters goes up and as the transmitters behave in a “uniform (equi-probabilistic) manner,” the anonymity increases and we will show that the covert channel capacity diminishes.

For Scenario 2 we assume that there are transmitters Alice and Clueless $_i$, $i = 1, \dots, N$. The N Clueless $_i$ transmitters behave independently of each other and of Alice, and they all have the same time-invariant probabilistic behavior. Alice and the Clueless $_i$ are hidden from Eve. They submit their messages to a MIX that also functions as a firewall. This first *MIX-firewall* acts as an exit point.

This MIX-firewall sends its encrypted messages to a second MIX-firewall that is an entrance to a second hidden (from Eve) enclave. We further assume that Eve is a GPA *only* between the two MIX-firewalls, *i.e.*, an RPA. That is, Eve only has knowledge of how many messages come out of the first MIX-firewall per tick, and Eve does not know to whom the messages are going. The situation is described by the following diagram (Figure 3).

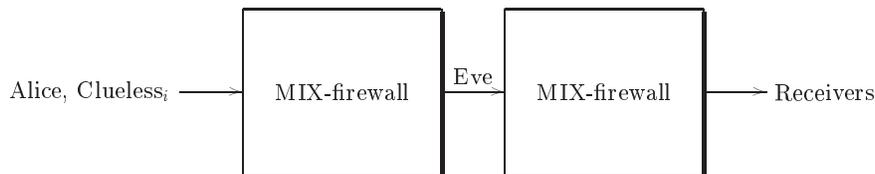


Fig. 3. MIX-firewalls with Restricted Passive Adversary.

This situation is realistic³ if the MIXes are acting as (first) firewall exit and (second) entrance points, or if the MIXes are onion-type routers acting as firewalls. Therefore, as stated above, we assume throughout this scenario that Eve only has knowledge of the number of messages coming out of a MIX acting as a firewall. Transmitters are allowed at most one transmission per tick. Alice attempts to signal Eve by transmitting to one of M possible receivers (which receiver Alice transmits to is immaterial), or by not transmitting at all. However, Clueless _{i} is also transmitting without any regard to what Alice is doing. The transmissions from both Alice and Clueless _{i} go into the first MIX-firewall, which acts as an exit point. Alice does not know what Clueless _{i} is doing (this assumption is made throughout the paper). Eve sees messages coming out of the first MIX-firewall on their way to the second MIX-firewall, but does not know who sent them, or where they are going. All messages go into the second MIX-firewall, which sends them to their receivers. Every tick, Alice and each Clueless _{i} either send or do not send one message each. Therefore, the only knowledge that Eve can get by eavesdropping is the number of messages per tick passing between the two MIX-firewalls. In other words, every tick, Eve observes the number of packets leaving the MIX-firewall and “receives” some number from the set $\{0, 1, \dots, N + 1\}$.

Therefore the only quantity observable by Eve that Alice can affect, per tick, is the number of messages that Eve counts. This covert channel is a discrete memoryless channel *with* noise since the Clueless _{i} ’s randomly affect the output.

³ Consider the case of packets from one LAN/enclave being sent to another LAN/enclave using IPSEC tunneling [8]. In this case, an eavesdropper can only count the number of outgoing messages destined for the receiving enclave. What goes on inside each LAN/enclave is hidden from an eavesdropper. If UDP with no application level ACKs is employed, communication is only one-way [16].

How does Eve regard the transmissions? What is the most information that Alice can send to Eve in this manner? Shannon’s information theory [21] answers these questions for us.

Let us go back to the base scenario; here we stated that the capacity is obviously $\log(M + 1)$. How do we know that some other exploitation of the base scenario will not give us a higher capacity? The reason is that there are at most $M + 1$ symbols in whatever exploitation we use, and if the channel is noiseless we have maximized the capacity (this is related to the maximum entropy as discussed in [11].) For Scenario 2 capacity cannot be explained so easily and is the major study of this paper.

Keep in mind that for Scenario 2 it does not matter if there is one receiver or there are one hundred and one receivers. Eve can only count, and Alice or Clueless_{*i*} can only send one message per tick. Therefore the number of receivers does not matter. It is only important that there is at least one receiver.

We break Scenario 2 down into four cases: 2.0, 2.1, 2.2, and 2.3. Case 2.3 is the general form of Scenario 2 and the first three are simplified special cases.

2.1 Two special cases of Scenario 2: — Alice alone, and with and one additional transmitter

Case 2.0 — Alice

This is the case where $N = 0$. Alice is the only transmitter. Alice sends either 0 (by not sending a message) or 0^c (by sending a message — it does not matter to which receiver Alice sends the message since that is indistinguishable to Eve). Eve receives either $e_0 = 0$ (Alice did nothing) or $e_1 = 1$ (Alice sent a message to a receiver). There is no noise in this channel. The capacity of this covert channel is 1.

We develop the necessary information theory further on in the paper. However, we state the capacity is the maximum, over the probability x for Alice inputting a 0, of the mutual information $I(E, A)$. A is the distribution for Alice described by x , and E is the distribution for Eve. Since there is no noise, I is simply the entropy $H(E)$ describing Eve.

$$I(E, A) = H(E) = -x \log x - (1 - x) \log(1 - x),$$

which is maximized to 1 when $x = .5$.

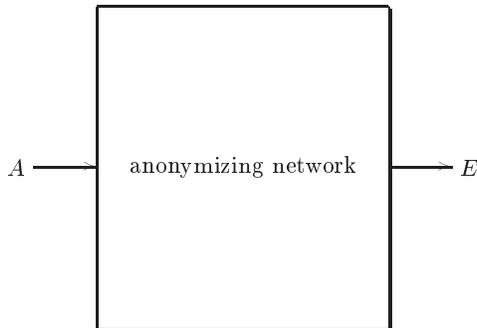
Case 2.1 — Alice and one additional transmitter (Clueless)

In this case $N = 1$. Therefore, Eve receives:

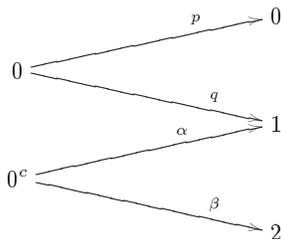
- 0 if neither Alice nor Clueless transmit;
- 1 if Alice does not transmit and Clueless does transmit, or Clueless transmits and Alice does not; or
- 2 if both Alice and Clueless transmit.

In the remainder of subsection 2.1 we develop the information theory to analyze the covert channel for Case 2.1.

Let us model the communications channel as follows: A is the input random variable describing Alice, and E is the output random variable describing Eve. Clueless contributes to the noise, but is not modeled as an input. Alice communicates with Eve via the covert channel. The input symbols for the channel are 0 , which signifies that Alice is not transmitting a message to any receiver, and 0^c , which signifies that Alice is transmitting a message to some receiver.⁴



(a) Channel block diagram



(b) Channel transition diagram

Fig. 4. Channel model for Case 2.1

⁴ At this point we caution the reader not to confuse Alice transmitting a message to a receiver R_i , and Alice communicating to Eve via the covert channel. Eve is *not* the receiver R_i in the sense of Alice or Clueless transmitting a message. Eve receives symbols via the covert channel from Alice. There are two different communication paths that must be kept separate. One is the legitimate network communication that the anonymizing device attempts to keep unknown. The other is the covert communication that Alice has to Eve. A way to stop the covert communication would be for the anonymizing device to pad [11–13, 23, 24] messages so that it would appear to Eve that both Alice and Clueless are transmitting a message. This inefficiency might be tolerated in such an ideal situation as Case 2.1, but such a strategy must be called into question when it comes to real traffic. In Case 2.1 the anonymizing effect is done by a MIX-firewall, which does not *a priori* pad. Of course, before advocating traffic padding one should be fully aware of the threat that the padding is intended to stop. Failure to understand the threat first is inadvisable since padding comes at the pragmatic costs of efficiency and proper network resource utilization.

Figure 4 shows two ways to look at the channel. The top part (a) of the figure is the simple schematic. A is the input, E is the output, and the anonymizing network (the two MIX-firewalls between the transmitters and receivers) adds noise. The bottom part (b) of Figure 4 shows that the inputs symbols are: 0, which represents A not sending a message; and 0^c , corresponding to A actually sending a message to one of the M possible receivers. The output symbols correspond to the three states E might perceive. The output symbol 0 corresponds to no one sending a message; the output symbol 1 corresponds to Alice or Clueless, but not both, sending a message; and the output symbol 2 corresponds to both Alice and Clueless sending a message.

Let us consider the channel matrix.

$$M_{2,1} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 \end{matrix} \\ \begin{matrix} 0 \\ 0^c \end{matrix} & \begin{pmatrix} p & q & 0 \\ 0 & \alpha & \beta \end{pmatrix} \end{matrix}$$

The 2×3 channel matrix $M_{2,1}[i, j]$ represents the conditional probability of Eve receiving the symbol j when Alice sends the symbol i ,

$$M_{2,1}[i, j] = P(E = j \mid A = i).$$

We will show that $p = \alpha$, and thus it trivially follows that $q = \beta$.

The probability $P(\cdot \mid A = i)$ is totally dependent upon what Clueless does (the action of Alice is already fixed at $A = i$, by the fact that it is a conditional probability). Let us consider what happens when Clueless sends a message, and assign a probability $1 - \zeta$ to Clueless sending the message.⁵ Consider $P(E = 0 \mid A = 0)$ and $P(E = 1 \mid A = 0^c)$. The only way for Eve to receive a 0, when Alice *has not* sent a message, is for Clueless not to have sent a message. Therefore, $P(E = 0 \mid A = 0) = \zeta$. The only way for Eve to receive a 1, when Alice *has* sent a message, is for Clueless not to have sent a message. Therefore, we also have $P(E = 1 \mid A = 1) = \zeta$. Thus $p = \zeta = \alpha$, and $q = \beta = 1 - p$. So our channel matrix simplifies to:

$$\begin{matrix} & \begin{matrix} 0 & 1 & 2 \end{matrix} \\ \begin{matrix} 0 \\ 0^c \end{matrix} & \begin{pmatrix} p & q & 0 \\ 0 & p & q \end{pmatrix} \end{matrix}$$

We wish to determine the channel capacity of the above discrete memoryless channel. We let the probability that Alice sends a 0 be $P(A = 0) = x$,

⁵ We will assume from now on that such a distribution can be assigned, and further that the distribution is stationary (it is the same each tick). Without such an assumption we can still study the problem, but if the distribution is non-stationary the analysis becomes much more difficult since the channel is no longer memoryless. We do not feel though that assigning Clueless a stationary distribution is that onerous. The distribution could be assigned via statistical analysis of past behavior (to make this valid one should assume that Clueless is not adapting to Alice's behavior). If one cannot assign a random variable to Clueless then our analysis is erroneous.

and therefore $P(A = 0^c) = 1 - x$. The term x is the only term that can be varied to achieve capacity. Here is where Alice may use knowledge of long-term transmission characteristics of the other transmitters, as well as how many other transmitters there are, to change her (long-term) behavior. As with other studies of covert channels [10] we are not concerned with source coding/decoding issues [21]. Our concern is the limits on how well a transmitter can “optimize” its bit rate to a receiver, given that a channel is noisy. Given a discrete random variable X , taking on the values x_i , $i = 1, \dots, n_X$, the entropy of X is:

$$H(X) = - \sum_{i=1}^{n_X} p(x_i) \log p(x_i) .$$

We use $p(x_i)$ as a shorthand notation for $P(X = x_i)$. Given two such discrete random variables X and Y we define the conditional entropy (equivocation) to be:

$$H(X|Y) = - \sum_{i=1}^{n_Y} p(y_i) \sum_{j=1}^{n_X} p(x_j|y_i) \log p(x_j|y_i) .$$

Given two such random variables we define the mutual information between them to be:

$$I(X, Y) = H(X) - H(X|Y) .$$

Note that $H(X) - H(X|Y) = H(Y) - H(Y|X)$, so we see that $I(X, Y) = I(Y, X)$.

For a DMC whose transmitter random variable is X , and whose receiver random variable is Y , we define the *channel capacity* [21] to be:

$$C = \max_X I(X, Y),$$

where the maximization is over all possible distributions for X (that is, the $p(x_i)$ are all non-negative and sum to one).

In this situation $p(a_0) = P(A = 0) = x$, and $p(a_1) = P(A = 0^c) = 1 - x$. Since varying x is varying all values of the input probabilities, the capacity of the covert channel between Alice and Eve is

$$\max_x \{H(E) - H(E|A)\}.$$

$H(E|A)$ can be trivially determined from the channel matrix. To calculate $H(E)$ we first must determine the distribution for E , which can be determined from the conditional probabilities and the distribution for A . We see that:

$$\begin{aligned} p(e_0) &= P(E = 0) \\ &= P(E = 0|A = 0)P(A = 0) + P(E = 0|A = 0^c)P(A = 0^c) \\ &= px + 0(1 - x) = px, \end{aligned}$$

$$\begin{aligned}
p(e_1) &= P(E = 1) \\
&= P(E = 1|A = 0)P(A = 0) + P(E = 1|A = 0^c)P(A = 0^c) \\
&= qx + p(1 - x),
\end{aligned}$$

and similarly,

$$p(e_2) = P(E = 2) = q(1 - x).$$

Therefore,

$$H(E) = -\{px \log px + [qx + p(1 - x)] \log [qx + p(1 - x)] + q(1 - x) \log q(1 - x)\}.$$

Now, let us calculate the conditional entropy

$$H(E|A) = -\sum_{i=0}^1 p(a_i) \sum_{j=0}^2 p(e_j|a_i) \log p(e_j|a_i).$$

This is:

$$-(P(A = 0)\{p \log p + q \log q + 0 \log 0\} + P(A = 0^c)\{0 \log 0 + p \log p + q \log q\}),$$

which simplifies to

$$H(E, A) = -(x\{p \log p + q \log q\} + (1 - x)\{p \log p + q \log q\}).$$

Thus, $H(E|A) = h(p)$,⁶ so

$$I(E, A) = -(px \log px + [qx + p(1 - x)] \log [qx + p(1 - x)] + q(1 - x) \log q(1 - x)) - h(p),$$

and

$$C = \max_x \left\{ -\left(px \log px + [qx + p(1 - x)] \log [qx + p(1 - x)] + q(1 - x) \log q(1 - x) \right) - h(p) \right\}.$$

One way to find the maximum is to take the first derivative of $I(E, A)$ with respect to x , and set it equal to zero. Since

$$\begin{aligned}
&\frac{d}{dx} \left\{ -\left(px \log px + [qx + p(1 - x)] \log [qx + p(1 - x)] + q(1 - x) \log q(1 - x) \right) - h(p) \right\} \\
&= \frac{-1}{\ln 2} \left\{ p \ln p - (1 - p) \ln(1 - p) + p \ln x - (1 - p) \ln(1 - x) + (1 - 2p) \ln[(1 - 2p)x + p] \right\} \\
&\hspace{15em} (1)
\end{aligned}$$

⁶ The notation $h(p)$ denotes the function $-p \log p - (1 - p) \log(1 - p)$.

(noting that the derivative of $h(p)$ is zero, and $q = 1 - p$), finding the zero of $\frac{d}{dx}I(E, A)$ is equivalent to solving the following equation for x .

$$p \ln p - (1 - p) \ln(1 - p) + p \ln x - (1 - p) \ln(1 - x) + (1 - 2p) \ln[(1 - 2p)x + p] = 0$$

Letting $\beta = e^{(1-p)\ln(1-p) - p \ln p}$, this reduces to solving

$$\beta[(1 - 2p)x + p]^{2p-1} - x^p(1 - x)^{p-1} = 0. \quad (2)$$

When $p = 1/2$, we have that $\beta = 1$, and we are left with $1 - x^{1/2}(1 - x)^{-1/2} = 0$. Thus, when $p = 1/2$, the derivative (1) is maximized when $x = 1/2$. When $p = 0$, $\beta = 1$, and we are left with $x^{-1} - (1 - x)^{-1} = 0$. Hence, when $p = 0$, the derivative (1) also is maximized when $x = 1/2$. When $p = 1$, we have that $\beta = 1$, and we are left with $(1 - x)^1 - x = 0$. Thus, when $p = 1$, the derivative (1) likewise is maximized when $x = 1/2$. All of this might suggest that $x = 1/2$ always maximizes C , but this is not the case (see Figure 5).

Unfortunately, we cannot solve (2) in general. That is, we are unable to derive a closed form expression for the x value that maximizes the derivative (1) as a function of p . Therefore, we numerically solve⁷ for the zero of (1), and use that value to evaluate $I(E, A)$; this gives us the capacity as a function of p . Figure 5 shows plots of both the zero of $\frac{d}{dx}I(E, A)$ as a function of p , and the capacity $C(p)$.⁸ Note that the zero of $\frac{d}{dx}I(E, A)$ is the x value that maximizes $I(E, A)$. That is this choice of x determines the probability distribution of A (as stated earlier $P(A = 0) = x$, and $P(A = 0^c) = 1 - x$) that achieves capacity (maximizes the mutual information).

We see in Figure 5 certain symmetries. The capacity graph is symmetric about $p = .5$, and the graph of the x that achieves capacity is skew-symmetric about $p = .5$ (when $p = .5$ the corresponding x is also $.5$). Consider the two situations where $p = \epsilon$, and where $p = 1 - \epsilon$; in both situations $0 \leq \epsilon \leq .5$. Let x_ϵ be the probability for the input symbol 0 that achieves capacity in the

⁷ At this juncture we could have numerically determined the maximum of $I(E, A)$. We chose instead to use Newton's method to find the zero of the derivative (1). We do this because Newton's method is a fast method, and this way we learn more about the derivative (1). The mutual information function is concave down, see [7] [Thm. 4.4.2]&[5][Thm.2.7.4], as a function of x , and since in this paper the mutual information is never locally constant (see Def. 1 later on in the paper), the maximum (p fixed) is achieved for one and only one x value. Therefore, we can find the capacity as follows. Evaluate, for fixed p , the mutual information as a function of x , letting x go from 0 to 1 in increments of .001. Via the concavity argument this will give the x value that maximizes the mutual information to the nearest .001. This is the capacity. This method and Newton's method gave identical results. Later in the paper we will not differentiate the mutual information due to the complexity of it and since we will not be able to obtain closed form solutions for the x value that maximizes the mutual information. We will instead use this simpler numerical method.

⁸ Holding p fixed we determine the zero of the derivative (1). Using that zero we evaluate $I(E, A)$, using the fixed value of p and the associated zero of (1), to determine the capacity.

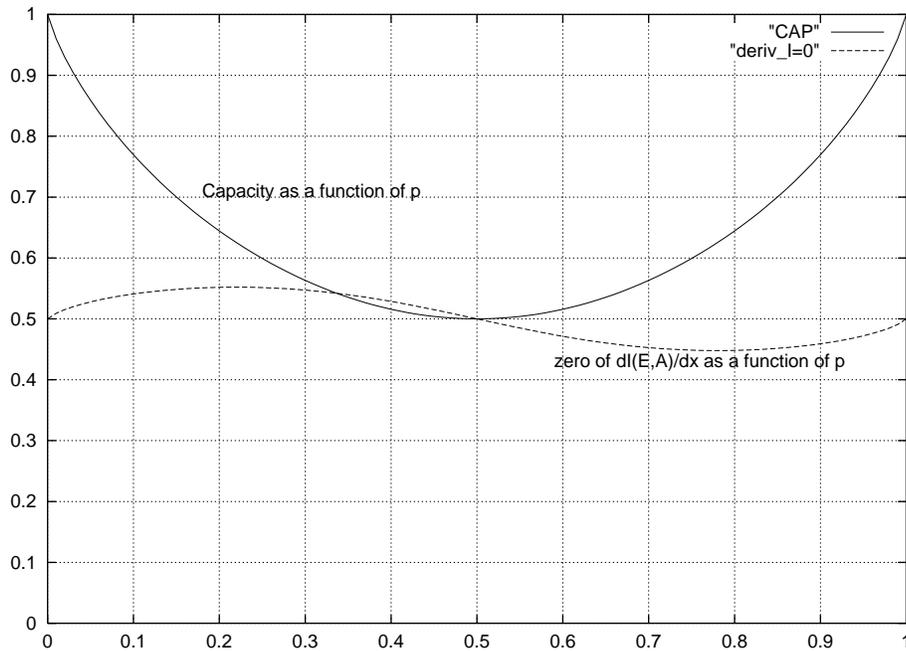


Fig. 5. Plots of covert channel capacity as a function of p , and of the x value that maximizes the mutual information as a function of p .

first situation, and let $x_{1-\epsilon}$ be the probability that achieves capacity for the second situation. For the first situation we have that $1 - x_\epsilon$ is the capacity achieving probability for the output symbol 0^c , and similarly for the second situation $1 - x_{1-\epsilon}$ is the capacity achieving probability for the output symbol 0^c . Physically the two situations are “the same” if we reverse the roles of the outputs symbols 0 and 2. Therefore $x_\epsilon = 1 - x_{1-\epsilon}$. Writing x_ϵ as $x_\epsilon = \frac{1}{2} + \Delta$, we see that $x_{1-\epsilon} = \frac{1}{2} - \Delta$; this is what the lower dotted plot shows in Figure 5 ($\epsilon = 1/2 \Rightarrow \Delta = 0$).

The above discussions bring to light two important observations that also hold when there are N transmitters in addition to Alice.

Observation 1 *In conditions of very little extra traffic, or very high extra traffic, the covert channel from Alice to Eve has higher bit rates.*

Observation 2 *The capacity $C(p)$, as a function of p is strictly bounded below by $C(.5)$, and $C(.5)$ is achieved when the mutual information is evaluated at $x = .5$.*

It is obvious that very little extra traffic corresponds to very little noise. At first glance though, it seems counterintuitive that heavy traffic also corresponds to a small amount of noise. This is because the high traffic is used as a baseline

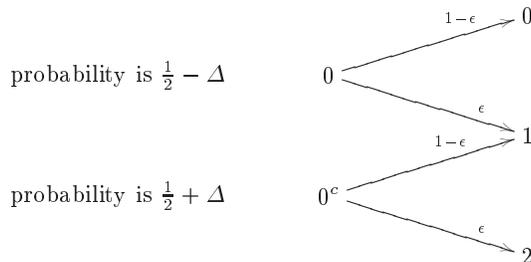


Fig. 6. Channel for Case 2.1 with ϵ interference from Clueless.

against which to signal. This is analogous to transmission of bits over a channel where the bit error rate (BER) P_e is greater than $1/2$. In this case, the capacity of the channel is the same as that of a channel with BER of $1 - P_e$, by first inverting all the bits. It is the in-between situations that negatively affect the signaling ability of Alice. But, even in the noisiest case (i.e., where $p = .5$) Alice can still transmit with a capacity of a half bit per tick.

Note that we can never guaranty error-free transmission, no matter how we group the output symbols. In fact, it is possible that the outputs will always be the symbol 1 (of course the probability of this quickly approaches zero, as the number of transmissions goes up). So this covert channel has a *zero-error capacity* [22] of zero. Capacity is a useful measure of a communication channel if the assumption is that the transmitter can transmit a large number of times. With a large number of transmissions an error-correcting code can be utilized so as to achieve a rate close to capacity. If the transmitter only transmits a small number of transmissions, then using the capacity alone can be misleading.

2.2 Case 2.2—Alice and two additional transmitters ($N = 2$)

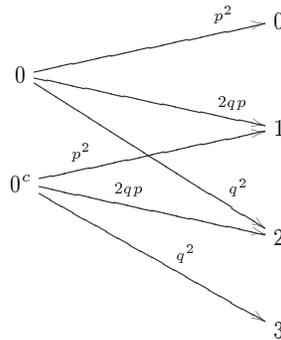
This is similar to Case 2.1, the difference being that we have three possible transmitters, A (random variable as before) for Alice, who is attempting to communicate covertly with E (random variable as before) for Eve, and two other benign “Clueless” transmitters modeled by the random variables C_1 , and C_2 , for Clueless₁ and Clueless₂, respectively. Since the MIX-firewalls only allow Eve to count the number of outgoing messages, our covert channel has four possible output symbols (the inputs are as before 0, for Alice not sending a message, and 0^c , if Alice does send a message). The outputs are:

- 0 — No one sends a message;
- 1 — Alice sends a message, and neither Clueless _{i} send a message; or, Alice does not send a message, and one, and only one, Clueless _{i} sends a message;
- 2 — Alice sends a message and one, and only one, Clueless _{i} sends a message; or, Alice does not send a message and both Clueless _{i} send a message;
- 3 — Alice, Clueless₁, and Clueless₂ all send a message.

As stated earlier we assume that Clueless₁ and Clueless₂ act independently of each other. Therefore, if, as before, p is the probability of a clueless transmitter (Clueless₁ or Clueless₂) not sending a message into the MIX-firewall, and $q = 1 - p$ is the probability of a clueless transmitter sending a message, the conditional probabilities of E given Alice sending 0 are:

- If Alice sends a 0, and Eve receives a 0, then neither Clueless₁ nor Clueless₂ sent a message; the conditional probability is p^2 .
- If Alice sends a 0, and Eve receives a 1, then one, but not both, of Clueless₁ or Clueless₂, sent a message into the MIX; the conditional probability is then $2qp$ from (Clueless₁ yes, Clueless₂ no), or (Clueless₁ no, Clueless₂ yes) - they are disjoint.
- If Alice sends a 0, and Eve receives a 2, then both Clueless₁ and Clueless₂ sent a message into the MIX and the conditional probability is q^2 .
- If Alice sends a 0, Eve never receives a 3, thus the conditional probability is 0.

Similarly we can analyze the case when Alice sends a 0^c . The covert channel diagram and channel matrix are shown in Figure 7.



(a) Channel transition diagram

$$M_{2,2} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 0 \\ 0^c \end{matrix} & \begin{pmatrix} p^2 & 2qp & q^2 & 0 \\ 0 & p^2 & 2qp & q^2 \end{pmatrix} \end{matrix}$$

(b) Channel matrix

Fig. 7. Channel for Case 2.2.

We can easily observe that the zero-error capacity is zero because the output symbols 1 and 2 can both be received if 0 or 0^c is transmitted. Therefore there is always some statistical error in what is received. This is similar to Case 2.1. Now, what about the simpler notion of capacity? We again represent the input random variable as A with distribution $P(A = 0) = p(a_0) = x$, and $P(A = 0^c) =$

$p(a_1) = 1 - x$. The output random variable is E with distribution $P(E = j) = p(e_j)$, $i = 0, 1, 2, 3$, and the mutual information is $I(E, A) = H(E) - H(E|A)$. So,

$$I(E, A) = - \sum_{j=0}^3 p(e_j) \log p(e_j) + \sum_{i=0}^1 p(a_i) \sum_{j=0}^3 p(e_j|a_i) \log p(e_j|a_i) .$$

The $p(e_j|a_i)$ are the i, j terms of the matrix $M_{2,2}$, and $p(a_0) = x$, so all we need are the $p(e_j)$ terms. Since

$$\begin{aligned} p(e_j) &= p(e_j|a_0)p(a_0) + p(e_j|a_1)p(a_1) \\ &= P(E = j|A = 0)P(A = 0) + P(E = j|A = 0^c)P(A = 0^c), \end{aligned}$$

we see that:

$$\begin{aligned} p(e_0) &= p^2 x, \text{ and} \\ p(e_1) &= 2qpx + p^2(1 - x), \\ p(e_2) &= q^2 x + 2qp(1 - x), \\ p(e_3) &= q^2(1 - x) . \end{aligned}$$

So we see that

$$\begin{aligned} H(E) &= - \sum_{j=0}^3 p(e_j) \log p(e_j) \\ &= - \left\{ p^2 x \log p^2 x + (2qpx + p^2(1 - x)) \log (2qpx + p^2(1 - x)) \right. \\ &\quad \left. + (q^2 x + 2qp(1 - x)) \log (q^2 x + 2qp(1 - x)) + q^2(1 - x) \log q^2(1 - x) \right\} . \end{aligned}$$

We also have that

$$-H(E|A) = \sum_{i=0}^1 p(a_i) \sum_{j=0}^3 p(e_j|a_i) \log p(e_j|a_i) = 2[qp - h(p)] .$$

Therefore, we see that the mutual information is

$$I(E, A) = - \left\{ p^2 x \log p^2 x + (2qpx + p^2(1 - x)) \log (2qpx + p^2(1 - x)) \right.$$

$$\begin{aligned}
& + (q^2x + 2qp(1-x)) \log(q^2x + 2qp(1-x)) + q^2(1-x) \log q^2(1-x) \Big\} \\
& + 2[qp - h(p)]
\end{aligned}$$

We will often simply write the mutual information as I instead of $I(A, E) = I(E, A)$. Let us fix the value of p at some boundary values and see what happens to the mutual information.

$$I|_{p=0} = I|_{p=1} = h(x) .$$

Therefore, since capacity is the maximum over x of I , we see that (viewing capacity as a function of p):

$$C(p=0) = C(p=1) = 1$$

Certainly since the input is limited to the two symbols 0 and 0^c , capacity is bounded between zero and one. Let us consider the channel diagrams in these two special cases.

In both of these cases we have a noiseless channel on two symbols. Therefore, the capacity is $\max_x h(x)$ which is simply one. The more interesting cases, when $0 < p < 1$, we solve numerically and plot the results in Figure 9. Of course, the capacity is symmetric about .5 because of the inherent symmetry between p and q .

Figure 10 depicts on one plot the capacity from Case 2.1 (two transmitters — Clueless) and the capacity from Case 2.2 (three transmitters — Clueless₁, Clueless₂).

Except for the boundary values, the capacity is always less for a given p with three transmitters than with two. This is not surprising, the extra clueless transmitter means extra noise. Note that the noisiest case is when $p = .5$; in this case the channel diagram is given in Figure 11. In this case, the capacity is achieved when $P(A = 0) = x = 1/2$, and the capacity is $\sim .3113$ (this can be argued through symmetry; we make it precise below in the general case).

Unfortunately we cannot derive closed form solutions even for these simple cases. Therefore, it seems unlikely that we can derive a closed form for the general case of N clueless transmitters in addition to Alice. Of course, we could still derive the capacity numerically. However, we are able to obtain some bounding results.

2.3 Case 2.3—Alice and N additional transmitters

Case 2.3 is the general form of Scenario 2. Now⁹ we imagine that there are $N + 1$ transmitters, Alice is one of them, and the other N are all independently identical

⁹ One could relax the assumption that all the Clueless_i have identical and independent behavior.

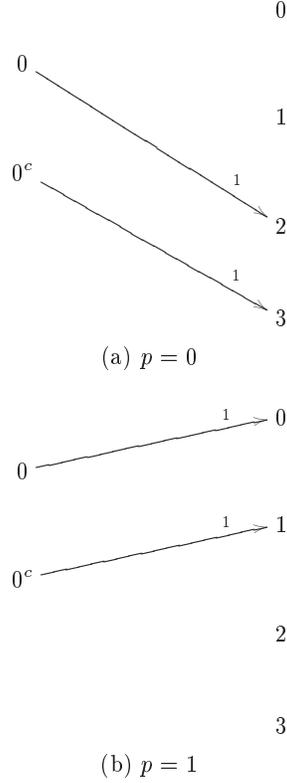


Fig. 8. Special cases for the channel diagram for Case 2.2.

clueless transmitters. That is, there are transmitters $\text{Clueless}_1, \text{Clueless}_2, \dots, \text{Clueless}_N$. Again, Eve can only see how many messages are leaving the first MIX-firewall headed for the second MIX-firewall. Therefore Eve can determine if there are $0, 1, \dots, N + 1$ messages leaving the firewall. That is all Eve can determine. Therefore, there are still the two input symbols $a_0 = 0$ and $a_1 = 0^c$, but we have $N + 2$ output symbols. The probability that Clueless_i does not send a message is still p , and that it does send a message is $q = 1 - p$. Now, calculate the channel matrix.

Alice sends a 0.

- For Eve to receive e_k (that is $E = k$), $0 \leq k \leq N$ we need k of the clueless transmitters to send a message, and $N - k$ not to send a message. Therefore,

$$p(e_k|A = 0) = \binom{N}{k} p^{N-k} q^k, \quad 0 \leq k \leq N.$$

- $p(e_{N+1}|A = 0) = 0$, since the event never happens because Alice is *not* transmitting.

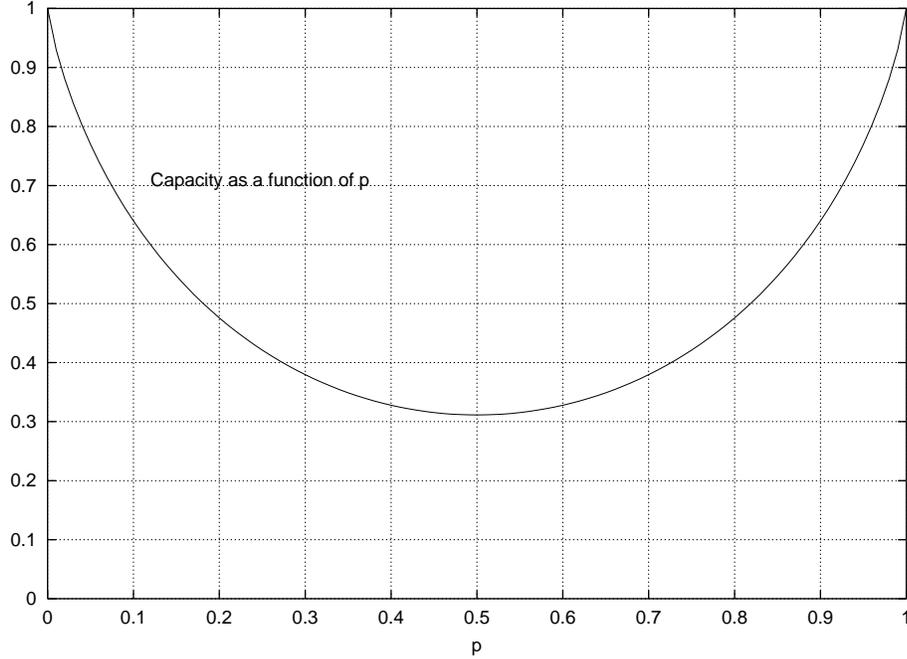


Fig. 9. Capacity as a function of p for three transmitters.

Alice sends a 0^c .

- $p(e_0|A = 0^c) = 0$, since the event never happens, because Alice *is* transmitting so Eve must observe at least one message.
- For Eve to receive e_k (that is $E = k$), $1 \leq k \leq N + 1$ we need $k - 1$ of the clueless transmitters to send a message, and $N - k + 1$ not to send a message. Therefore,

$$p(e_k|A = 0^c) = \binom{N}{k-1} p^{N-k+1} q^{k-1}, \quad 1 \leq k \leq N + 1.$$

Since $p(e_k) = p(e_k|A = 0)P(A = 0) + p(e_k|A = 0^c)P(A = 0^c)$, we have that

$$\begin{aligned} p(e_0) &= xp^N, \\ p(e_k) &= x \binom{N}{k} p^{N-k} q^k + (1-x) \binom{N}{k-1} p^{N-k+1} q^{k-1}, \quad 1 \leq k \leq N, \text{ and} \\ p(e_{N+1}) &= (1-x)q^N. \end{aligned}$$

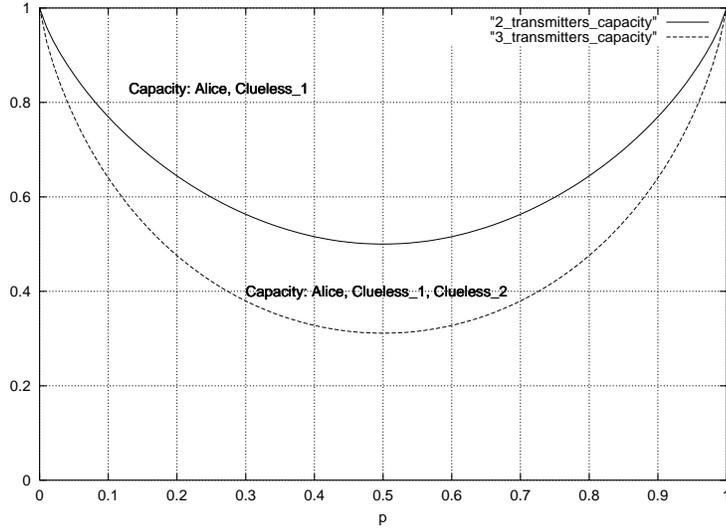


Fig. 10. Capacity as a function of p for both two and three transmitters.

So the entropy of E is

$$H(E) = - \left\{ xp^N \log xp^N + \sum_{k=1}^N \left[x \binom{N}{k} p^{N-k} q^k + (1-x) \binom{N}{k-1} p^{N-k+1} q^{k-1} \right] \log \left[x \binom{N}{k} p^{N-k} q^k + (1-x) \binom{N}{k-1} p^{N-k+1} q^{k-1} \right] + (1-x)q^N \log(1-x)q^N \right\}.$$

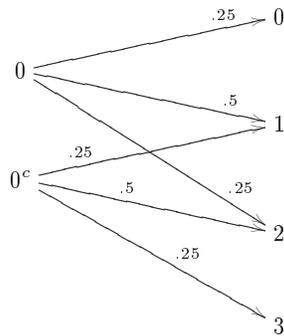
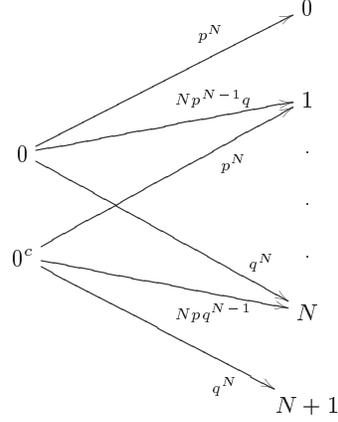


Fig. 11. Channel diagram for noisiest situation for Case 2.2.



$$M_{3,N} = \begin{matrix} & 0 & 1 & 2 & \dots & N & N+1 \\ \begin{matrix} 0 \\ 0^c \end{matrix} & \begin{pmatrix} p^N & Np^{N-1}q & \binom{N}{2}p^{N-2}q^2 & \dots & q^N & 0 \\ 0 & p^N & Np^{N-1}q & \dots & Npq^{N-1} & q^N \end{pmatrix} \end{matrix}$$

(b) Channel matrix

Fig. 12. Channel for Case 2.3, the general case of N clueless users.

The conditional entropy is a little easier to deal with.

$$\begin{aligned}
H(E|A) &= - \left(x \left\{ \sum_{j=0}^N \left[\binom{N}{j} p^{N-j} q^j \right] \log \left[\binom{N}{j} p^{N-j} q^j \right] \right\} \right. \\
&\quad \left. + (1-x) \left\{ \sum_{j=1}^{N+1} \left[\binom{N}{j-1} p^{N-j+1} q^{j-1} \right] \log \left[\binom{N}{j-1} p^{N-j+1} q^{j-1} \right] \right\} \right) \\
H(E|A) &= - \sum_{l=0}^N \left[\binom{N}{l} p^{N-l} q^l \right] \log \left[\binom{N}{l} p^{N-l} q^l \right] \tag{3}
\end{aligned}$$

Observe that $H(E|A)$ is independent of x . Therefore, to maximize the mutual information we only need to maximize $H(E)$.

The **mutual information** is

$$I(E, A) = - \left\{ xp^N \log xp^N \right.$$

$$\begin{aligned}
& + \sum_{k=1}^N \left[x \binom{N}{k} p^{N-k} q^k + (1-x) \binom{N}{k-1} p^{N-k+1} q^{k-1} \right] \\
& \log \left[x \binom{N}{k} p^{N-k} q^k + (1-x) \binom{N}{k-1} p^{N-k+1} q^{k-1} \right] \\
& + (1-x) q^N \log(1-x) q^N \left. \vphantom{\sum_{k=1}^N} \right\} \\
& + \sum_{l=0}^N \left[\binom{N}{l} p^{N-l} q^l \right] \log \left[\binom{N}{l} p^{N-l} q^l \right] \tag{4}
\end{aligned}$$

For Case 2.1 (one Clueless in addition to Alice) and for Case 2.2 (two clueless in addition to Alice) we discussed the symmetry about $p = .5$ informally. Case 2.3 includes Cases 2.1 and 2.2 as special cases, and we prove this symmetry exists for the general case.

Theorem 1 $I(E, A)|_{x,p} = I(E, A)|_{1-x,q}$

PROOF: By inspecting Eq. 4 we see that the last term ($-H(E|A)$) is independent of x , so we can ignore it. The terms $x p^N \log x p^N$ and $(1-x) q^N \log(1-x) q^N$ are interchanged when x and p are interchanged with $1-x$ and q , respectively. This leaves the complicated term in the middle of Eq. 4. We define $A_j(x, p) = \left[x \binom{N}{j} p^{N-j} q^j + (1-x) \binom{N}{j-1} p^{N-j+1} q^{j-1} \right]$, therefore the middle term

is just $\sum_{k=1}^N A_k(x, p) \log A_k(x, p)$. We consider the complementary j and $N-j+1$

indices. Note, $A_{N-j+1}(x, p) = \left[x \binom{N}{N-j+1} p^{j-1} q^{N-j+1} + (1-x) \binom{N}{N-j} p^j q^{N-j} \right]$. (There are always such complementary terms except for when N is odd and j is the ‘‘middle’’ index $\lceil N/2 \rceil$. We will return to this special case.)

Consider $A_j(x, p)$ and the complementary $A_{N-j+1}(x, p)$. Using the identity $\binom{N}{k} = \binom{N}{N-k}$ it trivially follows that $A_j(x, p) = A_{N-j+1}(1-x, q)$. Therefore, since $N - (N-j+1) + 1 = j$ we see that

$$\begin{aligned}
& A_j(x, p) \log A_j(x, p) + A_{N-j+1}(x, p) \log A_{N-j+1}(x, p) \\
& = A_{N-j+1}(1-x, q) \log A_{N-j+1}(1-x, q) + A_j(1-x, q) \log A_j(1-x, q) .
\end{aligned}$$

Now let us look at the special case where N is odd and we are considering $A_{\lceil N/2 \rceil}(x, p)$, which does not have a complementary term, since $\lceil N/2 \rceil = N - \lceil N/2 \rceil + 1$. However, it trivially follows that $N - \lceil N/2 \rceil = \lceil N/2 \rceil - 1$, and hence we also trivially have that $\binom{N}{\lceil N/2 \rceil} = \binom{N}{\lceil N/2 \rceil - 1}$. Therefore, by substitution one sees that $A_{\lceil N/2 \rceil}(x, p) = A_{\lceil N/2 \rceil}(1-x, q)$. \square

We will need the following in the rest of the paper so we will consider $I(E, A)|_{p=.5} = H(E)_{p=.5} - H(E|A)_{p=.5}$ now.

Consider the entropy of E evaluated when $p = \frac{1}{2}$.

$$\begin{aligned}
H(E)|_{p=.5} = & - \left\{ x \left(\frac{1}{2}\right)^N \log x \left(\frac{1}{2}\right)^N \right. \\
& + \sum_{k=1}^N \left[x \binom{N}{k} \left(\frac{1}{2}\right)^N + (1-x) \binom{N}{k-1} \left(\frac{1}{2}\right)^N \right] \\
& \log \left[x \binom{N}{k} \left(\frac{1}{2}\right)^N + (1-x) \binom{N}{k-1} \left(\frac{1}{2}\right)^N \right] \\
& \left. + (1-x) \left(\frac{1}{2}\right)^N \log(1-x) \left(\frac{1}{2}\right)^N \right\} \tag{5}
\end{aligned}$$

Consider the conditional entropy when $p = \frac{1}{2}$.

$$\begin{aligned}
H(E|A)|_{p=.5} = & - \sum_{l=0}^N \left[\binom{N}{l} \left(\frac{1}{2}\right)^{N-l} \left(\frac{1}{2}\right)^l \log \left[\binom{N}{l} \left(\frac{1}{2}\right)^{N-l} \left(\frac{1}{2}\right)^l \right] \right] \\
= & - \sum_{l=0}^N \left[\binom{N}{l} \left(\frac{1}{2}\right)^N \log \left[\binom{N}{l} \left(\frac{1}{2}\right)^N \right] \right] \\
= & - \left(\frac{1}{2}\right)^N \left\{ \sum_{l=0}^N \binom{N}{l} \log \binom{N}{l} - N \sum_{l=0}^N \binom{N}{l} \right\} \\
= & - \left(\frac{1}{2}\right)^N \left\{ \sum_{l=0}^N \binom{N}{l} \log \binom{N}{l} - N 2^N \right\} \\
= & N - \left(\frac{1}{2}\right)^N \sum_{l=0}^N \binom{N}{l} \log \binom{N}{l}
\end{aligned}$$

Note that $H(E|A)|_{p=.5}$ is independent of x . Keep in mind that we may express the mutual information evaluated at (x', p') by the slightly overloaded notation $I(E, A)|_{x=x', p=p'}$. Of course $I(E, A)|_{p=p'}$ is simply a function of x , and $I(E, A)|_{x=x'}$ is a function of p .

Definition 1 *We say that an arbitrary (real valued) function is not locally-constant iff for all x with $f(x)$ defined at x , and for every $\delta > 0$, there exists an x' such that $d(x', x) < \delta$ (i.e., x' in the neighborhood of x) with $f(x') \neq f(x)$.*

That is, for no neighborhood, no matter how small, is the function constant.

Definition 2 *We say that a function $f : [0, 1] \rightarrow \mathfrak{R}$ is symmetric about $x = .5$, iff $f(x) = f(1 - x)$.*

Observation 3 *If $f(x)$ is symmetric about $x = .5$ and it is concave down (convex up) then $f(.5)$ is a maximum (minimum) value. Further, if $f(x)$ is not locally-constant then $.5$ is the only such critical point.*

Theorem 2 $I(E, A)|_{p=.5}$ is symmetric about $x = .5$.

PROOF: By Thm. 1 we see that $I(E, A)|_{x,.5} = I(E, A)|_{1-x,.5}$.
□

Theorem 3 $C(.5) = I(E, A)|_{x=.5, p=.5}$.

PROOF: By Theorem 2, we know that $I(E, A)|_{p=.5}$ is symmetric about $x = .5$, and [7][Thm. 4.4.2]&[5][Thm.2.7.4] show that $I(E, A)|_{p=.5}$ (and in general $I(E, A)$ for fixed p) is concave down. Therefore, from Observation 1, $I(E, A)|_{p=.5}$ obtains its maximum value when $x = .5$. Since capacity, when $p = .5$, is the maximum of $I(E, A)|_{p=.5}$, we are done.
□

Theorem 4 $C(p) \geq I(E, A)|_{x=.5, p=.5}$.

PROOF: By definition $C(p) \geq I(E, A)|_{x=.5}$, since capacity is the maximum of the mutual information. For x fixed $I(E, A)|_x$ is a convex up function of p (see [7][Thm. 4.4.2]&[5][Thm.2.7.4]). By Thm. 1 we see that $I(E, A)|_{x=.5}$ is symmetric about $p = .5$. By Observation 3 we see that $I(E, A)|_{x=.5} \geq I(E, A)|_{x=.5, p=.5}$.
□

This allows us to use the simple single value $I(E, A)|_{x=.5, p=.5}$ as a lower bound for the covert channel capacity.

Corollary 1 $C(p) \geq C(.5)$

PROOF: Apply Theorems 3 and 4 together.
□

Theorem 5 $C(p) = C(1 - p)$ and if x_p is the unique x such that $C(p) = I(E, A)|_{x_p, p}$, then $x_{1-p} = 1 - x_p$.

PROOF: This trivially follows from Thm. 1 and the uniqueness (follows from the concavity properties and the fact that the mutual information is not-locally constant—this follows by inspection of $I(E, A)$) of the critical x value.
□

Let us now use these results to bound capacity from below. We now consider the formula for mutual information when $x = p = .5$. Thus, we study $I(E, A)|_{x=.5, p=.5}$ as N varies. Let us first calculate $H(E|A)|_{x=.5, p=.5}$, since $H(E|A)$ is independent of x :

$$H(E|A)|_{x=.5, p=.5} = H(E|A)|_{p=.5} = N - \left(\frac{1}{2}\right)^N \sum_{l=0}^N \binom{N}{l} \log \binom{N}{l}$$

From Eq. 5 we know $H(E)|_{p=.5}$.

In what follows we use the identity $\binom{N}{k} + \binom{N}{k-1} = \binom{N+1}{k}$, and the fact that

$$\begin{aligned} \sum_{k=1}^N \binom{N+1}{k} &= \sum_{k=0}^{N+1} \binom{N+1}{k} - \binom{N+1}{0} - \binom{N+1}{N+1} \\ &= (1+1)^{N+1} - 1 - 1 = 2^{N+1} - 2. \end{aligned}$$

So

$$\begin{aligned} H(E)|_{x=.5, p=.5} &= - \left\{ \left(\frac{1}{2} \right)^{N+1} \log \left(\frac{1}{2} \right)^{N+1} \right. \\ &\quad + \sum_{k=1}^N \left[\binom{N}{k} \left(\frac{1}{2} \right)^{N+1} + \binom{N}{k-1} \left(\frac{1}{2} \right)^{N+1} \right] \\ &\quad \log \left[\binom{N}{k} \left(\frac{1}{2} \right)^{N+1} + \binom{N}{k-1} \left(\frac{1}{2} \right)^{N+1} \right] \\ &\quad \left. + \left(\frac{1}{2} \right)^{N+1} \log \left(\frac{1}{2} \right)^{N+1} \right\} \end{aligned}$$

$$\begin{aligned} H(E)|_{x=.5, p=.5} &= - \left\{ \left(\frac{1}{2} \right)^N \log \left(\frac{1}{2} \right)^{N+1} \right. \\ &\quad \left. + \sum_{k=1}^N \left(\frac{1}{2} \right)^{N+1} \left[\binom{N}{k} + \binom{N}{k-1} \right] \log \left(\frac{1}{2} \right)^{N+1} \left[\binom{N}{k} + \binom{N}{k-1} \right] \right\} \\ &= - \left\{ \left(\frac{1}{2} \right)^N \log \left(\frac{1}{2} \right)^{N+1} \right. \\ &\quad \left. + \left(\frac{1}{2} \right)^{N+1} \sum_{k=1}^N \binom{N+1}{k} \left[-(N+1) + \log \binom{N+1}{k} \right] \right\} \\ &= - \left\{ \left(\frac{1}{2} \right)^N \log \left(\frac{1}{2} \right)^{N+1} \right. \\ &\quad \left. - \frac{(N+1)}{2^{N+1}} \sum_{k=1}^N \binom{N+1}{k} + \left(\frac{1}{2} \right)^{N+1} \sum_{k=1}^N \binom{N+1}{k} \log \binom{N+1}{k} \right\} \\ &= - \left\{ \left(\frac{1}{2} \right)^N \log \left(\frac{1}{2} \right)^{N+1} \right. \\ &\quad \left. - \frac{(N+1)}{2^{N+1}} (2^{N+1} - 2) + \left(\frac{1}{2} \right)^{N+1} \sum_{k=1}^N \binom{N+1}{k} \log \binom{N+1}{k} \right\} \end{aligned}$$

$$\begin{aligned}
&= - \left\{ - (N+1) \left(\frac{1}{2}\right)^N \right. \\
&\quad \left. - \frac{(N+1)}{2^{N+1}} (2^{N+1} - 2) + \left(\frac{1}{2}\right)^{N+1} \sum_{k=1}^N \binom{N+1}{k} \log \binom{N+1}{k} \right\} \\
&= N+1 - \left(\frac{1}{2}\right)^{N+1} \sum_{k=1}^N \binom{N+1}{k+1} \log \binom{N+1}{k}
\end{aligned}$$

Since $I(E, A)|_{x=.5, p=.5} = H(E)|_{x=.5, p=.5} - H(E|A)|_{x=.5, p=.5}$, we see that:

$$\begin{aligned}
I(E, A)|_{x=.5, p=.5} &= N+1 - \left(\frac{1}{2}\right)^{N+1} \sum_{k=1}^N \binom{N+1}{k} \log \binom{N+1}{k} \\
&\quad - \left\{ N - \left(\frac{1}{2}\right)^N \sum_{l=0}^N \binom{N}{l} \log \binom{N}{l} \right\}
\end{aligned}$$

Therefore,

$$C(.5) = 1 + \left(\frac{1}{2}\right)^N \left\{ \sum_{l=0}^N \binom{N}{l} \log \binom{N}{l} - \frac{1}{2} \sum_{k=1}^N \binom{N+1}{k} \log \binom{N+1}{k} \right\}.$$

Since $\binom{N+1}{0} \log \binom{N+1}{0} = 0$, we can simplify this to:

$$C(.5) = 1 - \left(\frac{1}{2}\right)^N \sum_{k=0}^N \left\{ \frac{1}{2} \binom{N+1}{k} \log \binom{N+1}{k} - \binom{N}{k} \log \binom{N}{k} \right\}. \quad (6)$$

N	$C(.5)$	N	$C(.5)$
1	0.500000	13	0.053593
2	0.311278	14	0.049873
3	0.219361	15	0.046638
4	0.167553	16	0.043799
5	0.135170	17	0.041287
6	0.113278	18	0.039048
7	0.097558	19	0.037039
8	0.085730	20	0.035228
9	0.076502	21	0.033586
10	0.069092	22	0.032090
11	0.063007	23	0.030722
12	0.057917	24	0.029466
		25	0.028309

$C(.5)$ = lower capacity bounds for all p , $N = 1, \dots, 25$

Of course there are further relationships that can be exploited but they do not seem to assist in the analysis, but rather seem to obfuscate the symmetries.

The above table shows the results of numerical calculations of $C(.5)$ to six decimal places.

Note that in the general circumstances of Case 2.3, if $p = 0$ (or similarly $q = 0$), we have a noiseless channel and the capacity is one, which is achieved when $x = .5$. So we see that 1 is a tight upper bound for the capacity. Therefore we have the following result:

For Alice and N ($N > 0$) transmitters: $C(.5) \leq C(p) \leq 1$ and these bounds are tight.

Of course keep in mind the result from Case 2.0:

For Alice and no additional transmitters: Capacity = 1.

Therefore the region between the two plots for the N values represent the region where the capacity falls, depending on the behavior of the other, clueless transmitters (and Alice's knowledge of and long-term adaptation to them). Furthermore, the entire region is spanned by different choices of p (we ignore p for the degenerate case of $N = 0$). See Figure 13.

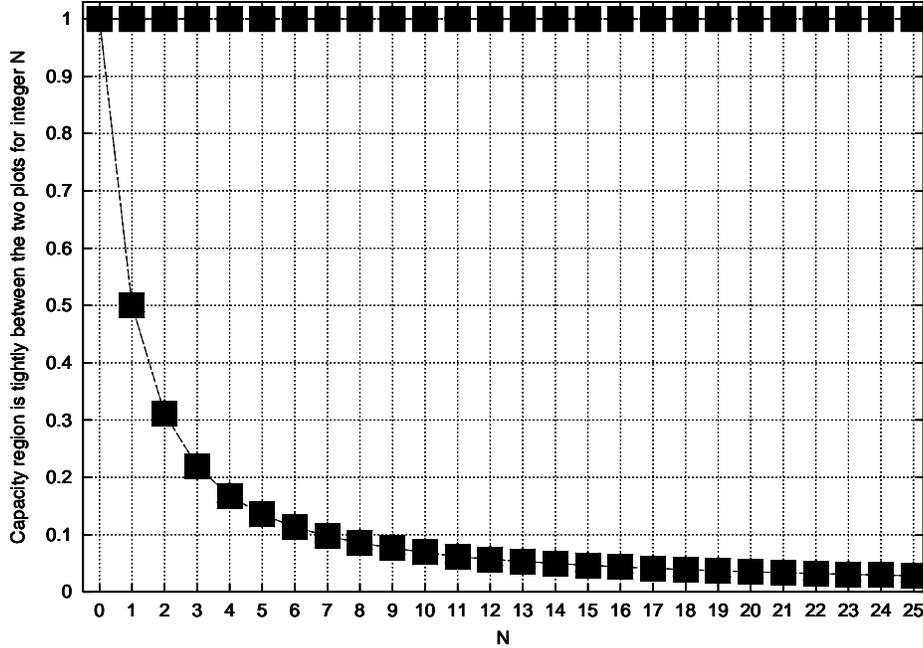


Fig. 13. Capacity region for Case 2, $N = 0$ to 25.

As N grows so does the noise. Therefore, we see that the capacity is non-increasing. We are interested in the lower bound $C(.5)$. We have numerically calculated $C(.5)$ to $N = 7750$ and have shown that $C(.5)$ is monotonically decreasing to zero (for $N=7750$, $C(.5) = .000093$). We can (but do not since it is many pages in length) analytically show $C(.5)$ is monotonic decreasing. That is not surprising since increasing the number of clueless users increases the noise, but it is surprising that it is so difficult to show that $C(.5)$ goes to zero as N goes to infinity. Below we discuss that fact in more detail.

From Eq. (6) we can express $C(.5)$ as

$$C(.5) = 1 - \left(\frac{1}{2}\right)^N S(N),$$

where

$$S(N) \triangleq \sum_{k=0}^N \left\{ \frac{1}{2} \binom{N+1}{k} \log \binom{N+1}{k} - \binom{N}{k} \log \binom{N}{k} \right\}.$$

First we will simplify $S(N)$.

Theorem 6 $S(N) = 2^N \log(N+1) - \sum_{k=0}^N \binom{N}{k} \log(k+1)$

PROOF: Define

$$\alpha(N) \triangleq \sum_{k=0}^N \binom{N}{k} \log \binom{N}{k}.$$

By expanding $\log \binom{N}{k}$ as $\log N! - \log(N-k)! - \log k!$, and using $\sum_{k=0}^N \binom{N}{k} = 2^N$,

$\binom{N}{k} = \binom{N}{N-k}$, and $\sum_{k=0}^N f(k) = \sum_{k=0}^N f(N-k)$, we have that

$$\alpha(N) = 2^N \log N! - 2 \sum_{k=0}^N \binom{N}{k} \log k!.$$

Therefore,

$$\alpha(N+1) = 2^{N+1} \log(N+1)! - 2 \sum_{k=0}^{N+1} \binom{N+1}{k} \log k!.$$

Since $1 \log 1 = 0$, we have that

$$\begin{aligned} S(N) &= \frac{1}{2} \sum_{k=0}^N \binom{N+1}{k} \log \binom{N+1}{k} - \sum_{k=0}^N \binom{N}{k} \log \binom{N}{k} \\ &= \frac{1}{2} \sum_{k=0}^{N+1} \binom{N+1}{k} \log \binom{N+1}{k} - \sum_{k=0}^N \binom{N}{k} \log \binom{N}{k} \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2}\alpha(N+1) - \alpha(N) \\
&= 2^N \log(N+1) + 2 \sum_{k=0}^N \binom{N}{k} \log k! - \sum_{k=0}^{N+1} \binom{N+1}{k} \log k!
\end{aligned}$$

Recalling the identity $\binom{N+1}{k} = \binom{N}{k} + \binom{N}{k-1}$, and using the fact that $\binom{N}{k} = 0$, if $k < 0$ or $k > N$, we further simplify the above to:

$$\begin{aligned}
S(N) &= 2^N \log(N+1) + \sum_{k=0}^{N+1} \binom{N}{k} \log k! - \sum_{k=0}^{N+1} \binom{N}{k-1} \log k! \\
&= 2^N \log(N+1) + \sum_{k=0}^N \binom{N}{k} \log k! - \sum_{k=0}^{N+1} \binom{N}{k-1} \log k!
\end{aligned}$$

after re-indexing the second sum

$$= 2^N \log(N+1) + \sum_{k=0}^N \binom{N}{k} \log k! - \sum_{k=0}^N \binom{N}{k} \log(k+1)!$$

since $\log(k+1)! = \log(k+1) + \log k!$

$$S(N) = 2^N \log(N+1) - \sum_{k=0}^N \binom{N}{k} \log(k+1) \quad \square$$

Keep in mind our goal is to study the behavior of $C(.5)$ as $N \rightarrow \infty$. However, first we need a technical lemma.

Lemma 1 $\sum_{k=1}^N \binom{N}{k} k^p = 2^{N-p} Q_p(N)$, for $p < N$, where $Q_p(N)$ is a monic polynomial in N of degree p .

PROOF: In [17, Formulas 1,2,7,8,9,10 p. 608] or [19, Formula 34 p. 85] it is shown that

$$\sum_{k=1}^N \binom{N}{k} k^p = 2^{N-p} \binom{N}{p} p! + 2 \sum_{i=1}^{p-1} (-1)^i \binom{N}{i} \frac{1}{2^i} \sum_{j=1}^i (-1)^j \binom{i}{j} j^p.$$

The term $2^{N-p} \binom{N}{p} p!$ is simply 2^{N-p} multiplied by $N \cdot (N-1) \cdots (N-p+1)$, which is simply 2^{N-p} times a monic polynomial in N of degree p . The other term, $2 \sum_{i=1}^{p-1} (-1)^i \binom{N}{i} \frac{1}{2^i} \sum_{j=1}^i (-1)^j \binom{i}{j} j^p$, is polynomial in N of degree less than p .
 \square

We are now ready for the major result.

Theorem 7 $\lim_{N \rightarrow \infty} C(.5) = 0$.

PROOF: We will prove the result by showing that $(\frac{1}{2})^N S(N)$ approaches one, since $C(.5) = 1 - (\frac{1}{2})^N S(N)$, this suffices.

Our first step is to use natural logarithms instead of base two logarithms.

$$S(N) = \frac{1}{\ln 2} \left\{ 2^N \ln(N+1) - \sum_{k=0}^N \binom{N}{k} \ln(k+1) \right\}.$$

Consider

$$\begin{aligned} \sum_{k=0}^N \binom{N}{k} \ln(k+1) &= \sum_{k=0}^N \binom{N}{N-k} \ln(N-k+1) \\ &= \sum_{k=0}^N \binom{N}{k} \ln(N-k+1) \\ &= \sum_{k=0}^N \binom{N}{k} \ln(1+N) \left(1 - \frac{k}{1+N}\right) \\ &= \sum_{k=0}^N \binom{N}{k} \ln(1+N) + \sum_{k=0}^N \binom{N}{k} \ln\left(1 - \frac{k}{1+N}\right) \\ &= 2^N \ln(N+1) + \sum_{k=0}^N \binom{N}{k} \ln\left(1 - \frac{k}{1+N}\right). \end{aligned}$$

Now we use the Maclaurin series of $\ln(1-x) = \sum_{n=1}^{\infty} -\left(\frac{x^n}{n}\right)$, which is valid for $|x| < 1$,

$$\sum_{k=0}^N \binom{N}{k} \ln\left(1 - \frac{k}{1+N}\right) = \sum_{k=0}^N \binom{N}{k} \sum_{p=1}^{\infty} \frac{k^p}{p(1+N)^p}$$

(In what follows we do not give an epsilon-delta style proof. Rather we ignore uniform convergence issues and freely pass terms in and out of the sums. This is done in the interest of space and intuition.)

$$= \sum_{p=1}^{\infty} \frac{1}{p} \frac{1}{(1+N)^p} \left[\sum_{k=0}^N \binom{N}{k} k^p \right] = \sum_{p=1}^{\infty} \frac{1}{p} \frac{1}{(1+N)^p} \left[\sum_{k=1}^N \binom{N}{k} k^p \right]$$

(Now we use Lemma 1)

$$= 2^N \sum_{p=1}^{\infty} \frac{(\frac{1}{2})^p}{p} \left[\frac{Q_p(N)}{(N+1)^p} \right]$$

We know can write $(\frac{1}{2})^N S(N)$ as

$$\left(\frac{1}{2}\right)^N S(N) = \frac{1}{\ln 2} \left\{ \sum_{p=1}^{\infty} \frac{\left(\frac{1}{2}\right)^p}{p} \left[\frac{Q_p(N)}{(N+1)^p} \right] \right\}.$$

Since $Q_p(N)$ is a monic polynomial in N of degree p , $\lim_{N \rightarrow \infty} \left[\frac{Q_p(N)}{(N+1)^p} \right] = 1$.

Therefore,

$$\lim_{N \rightarrow \infty} \left(\frac{1}{2}\right)^N S(N) = \frac{1}{\ln 2} \sum_{p=1}^{\infty} \frac{\left(\frac{1}{2}\right)^p}{p} = \frac{-1}{\ln 2} \ln\left(1 - \frac{1}{2}\right) = 1.$$

Since $C(.5) = 1 - \left(\frac{1}{2}\right)^N S(N)$, we are done.

□

2.4 Continuity

For Scenario 2 we wished to say that capacity was a continuous function of p . We thought that we could just use some standard information-theoretic result. Unfortunately, we could not find such a result. We do not think that it would be too hard to argue from the various concavity properties of mutual information that $C(p)$ is a continuous function (of p). However, we decided to present a more general result.

Theorem 8 *Let $F(x, p)$ be a continuous¹⁰ function defined on $[0, 1] \times U$, U an arbitrary subset of the reals, and assume that for each fixed p , $F(x, p)$ achieves a maximum denoted as $\Gamma(p)$. Then $\Gamma(p)$ is a continuous function of p .*

PROOF: If $\Gamma(p)$ is not continuous, then \exists a point of discontinuity p_0 . This means that there is an $\bar{\epsilon} > 0$ such that for any $\delta > 0$, \exists a p_δ such that $|p_\delta - p_0| < \delta$ but $|\Gamma(p_\delta) - \Gamma(p_0)| \geq \bar{\epsilon}$.

There is some x_0 such that $\Gamma(p_0) = F(x_0, p_0) = \max_x F(x, p_0)$ (there may be more than one such “maximizing” x).

Keep in mind though that $F(x, p)$ is a continuous function. This means that for every (t, p_0) , $t \in [0, 1]$, \exists a $\delta_t > 0$ such that¹¹ $d\{(x, p), (t, p_0)\} < \delta_t \Rightarrow |F(x, p) - F(t, p_0)| < \bar{\epsilon}$. The set $\{(x, p) \mid d\{(x, p), (t, p_0)\} < \delta_t\}$ is called a δ_t -neighborhood of (t, p_0) . Every δ_t -neighborhood of (t, p_0) can be replaced with an open square box centered about (t, p_0) with side length δ_t , we call this a δ_t -box neighborhood of (t, p_0) . This δ_t -box neighborhood of (t, p_0) is a proper subset of the δ_t -neighborhood of (t, p_0) .¹² Since $[0, 1] \times p_0$ is a compact set (closed and bounded) and the $\{\delta_t\text{-box neighborhood of } (t, p_0) \mid t \in [0, 1]\}$ is a collection of

¹⁰ In this paper all functions are real valued.

¹¹ d is the standard Euclidean metric in the plane.

¹² Keep in mind that when we form any sort of neighborhood we must intersect it with $\{[0, 1] \times U\}$, therefore our δ_t -balls or δ_t -boxes might not be actual balls or boxes. They might have gaps in them and not extend symmetrically on both sides of p_0 .

open sets that cover $[0, 1] \times p_0$, it can be replaced by a finite subcollection that also acts as a cover. Recall the point x_0 , we require that the δ_{x_0} -box neighborhood of (x_0, p_0) be in this finite subcollection (if it is not, just add it in). Define a set $T \stackrel{\text{def}}{=} \{t^i \mid i \in \{1, \dots, N\}, t^i \in [0, 1], \text{ and } x_0 \text{ is one of the } t^i\}$ such that $\bigcup_{t^i \in T} \{\delta_{t^i}\text{-box neighborhood of } (t^i, p_0)\}$ covers $[0, 1] \times p_0$. For simplicity we refer to the union of these sets as FC . Let $d = \frac{1}{2} \cdot \min\{t^i\}$ (this is where we use finiteness). Note that if $(x', p) \in FC$, then $\exists t^j \in T$ such that $|F(x', p) - F(t^j, p_0)| < \bar{\epsilon}$.

Since $\Gamma(p)$ is not continuous at p_0 we know that there is a p_ζ such that $|p_\zeta - p_0| < d$ but $|\Gamma(p_\zeta) - \Gamma(p_0)| \geq \bar{\epsilon}$. We know that there is some x_ζ such that $\Gamma(p_\zeta) = F(x_\zeta, p_\zeta) = \max_x F(x, p_\zeta)$ (there may be more than one such “maximizing” x). We have two cases to consider:

1. $\Gamma(p_\zeta) > \Gamma(p_0)$: So $\Gamma(p_0) \leq \Gamma(p_\zeta) - \bar{\epsilon}$
 Since d was chosen to be minimal by construction $(x_\zeta, p_\zeta) \in FC$. So for some t^j we have that $|F(x_\zeta, p_\zeta) - F(t^j, p_0)| < \bar{\epsilon}$, which is the same as $|F(t^j, p_0) - \Gamma(p_\zeta)| < \bar{\epsilon}$. So $\Gamma(p_\zeta) - \bar{\epsilon} < F(t^j, p_0)$, therefore $\Gamma(p_0) < F(t^j, p_0)$, which is impossible since $\Gamma(p_0)$ cannot be less than $F(x, p_0)$ for any x .
2. $\Gamma(p_\zeta) < \Gamma(p_0)$: So $\Gamma(p_\zeta) \leq \Gamma(p_0) - \bar{\epsilon}$
 Recall that we constructed FC so that it would contain the δ_{x_0} -box neighborhood of (x_0, p_0) . Therefore, since $|p_\zeta - p_0| < d$, and d was chosen minimal, we have that $(x_0, p_\zeta) \in \delta_{x_0}$ -box neighborhood of (x_0, p_0) . Therefore, $|F(x_0, p_\zeta) - F(x_0, p_0)| < \bar{\epsilon}$, which is the same as $|F(x_0, p_\zeta) - \Gamma(p_0)| < \bar{\epsilon}$. So $\Gamma(p_0) - \bar{\epsilon} < F(x_0, p_\zeta)$, therefore $\Gamma(p_\zeta) < F(x_0, p_\zeta)$, which is impossible since $\Gamma(p_\zeta)$ cannot be less than $F(x, p_\zeta)$ for any x .

Hence we have a contradiction, so $\Gamma(p)$ must be continuous.

□

We note that we used boxes instead of circles because it was easier to construct a distance d so that all points would be guaranteed to be in FC .

It is not important that $x \in [0, 1]$; what is important is that $[0, 1]$ is a compact set. Note that if D is not a compact set there are counter-examples.

Corollary 2 *Let $F(x, p)$ be a continuous function, where $p \in U$ and $x \in D$ where D is a closed and bounded subset of the real line. Assume that for each fixed p , $F(x, p)$ achieves a maximum denoted as $\Gamma(p)$. Then $\Gamma(p)$ is continuous in p .*

This is a technical point that we will not labor upon further. It does not affect the proof. What is important is that they are open sets. We have also used the fact that in a circle of radius r the largest box that can be inscribed (it is also centered about the center of the circle) has side length $\sqrt{2} r$, we use a smaller box. Also when we construct d later we use $\frac{1}{2}$ of a value, that is done since we are only looking at one side of a box.

Since, for Scenario 2, we see by inspection that the mutual information is a continuous function of (x, p) , and $x \in [0, 1]$, we have the following result.

Theorem 9 *For Scenario 2, $C(p)$ is a continuous function.*

We believe that continuity results such as this are important, but they seem to be overlooked in the literature

3 Comments, Generalizations & Future Work

3.1 Comments

We first note that despite the obfuscation provided by MIX-firewalls, and the attendant noise introduced by other transmitters, Alice is still able to transmit information to Eve. At this point, we recall our earlier observations and add to them below.

1. In conditions of very little extra traffic, or very high extra traffic, the covert channel from Alice to Eve has higher bit rates.
2. The capacity $C(p)$, as a function of p is strictly bounded below by $C(.5)$, and $C(.5)$ is achieved when the mutual information is evaluated at $x = .5$ (of course $p = .5$ also in this situation).
3. The capacity $C(p)$, as a function of p is strictly bounded below by a function that decreases monotonically to zero as the number of transmitters increases, but is never zero.
4. The bias in the code used by Alice to achieve the optimum data rate on the channel is not always $x = 0.5$, but it is never far from 0.5, and our preliminary experimental results indicate that the difference in capacity is minor.

The last observation agrees with [9], which presents the general result that in DMCs, capacity obtained by using $x = .5$ is no less than 94.21% of the optimum channel capacity. Even if Alice has no knowledge of the probabilistic behavior of the other transmitters, her data rate will not be too far from optimal if she uses an unbiased code. (Note, however, that the coding rate is very much dependent on knowledge of the number of other transmitters and their behavior.)

3.2 Future Work

Following up the last observation from the preceding subsection, we note that it does not hurt Alice too much if she does not use the optimum bias in her code (i.e., she does not know much about p). However, the choice of code will depend greatly on the channel capacity among other characteristics. It appears that at less noisy conditions (p near 0 or p near 1), the load, $L(N, p) = pN$, in expected packets per tick sent by the other transmitters, dominates in determining the

capacity. That is, for small p or $1-p$, defining $I_N(E, A)$ as the mutual information with N other transmitters,

$$I_N(E, A)|_{kp} \sim I_{kN}(E, A)|_p .$$

For intermediate values of p , (i.e., p near 0.5), the capacity is mostly influenced by N , the number of transmitters. As N increases, experimental results show that the curves of $C(p)$ versus p become increasingly “flat-bottomed,” hence are less sensitive to p for the intermediate values of p . So for Alice, knowing N is crucial unless the loads are rather low, in which case the load is the most important factor.

For Scenario 2 we assume that every Clueless_i was given by the same probability distribution. The probability p measured Clueless_i not sending a message. One can generalize Scenario 2 to allow these probabilities to vary. That is we can assign the probability p_i to Clueless_i not sending a message. Of course this changes the analysis that we have given above. We conjecture that the observations regarding the load and number of transmitters remains true as long as the p_i 's are not too different. The case of varying probabilities will be taken up in future work. However, we feel that our simplistic assumptions serve to show the difficulty of the analysis and to show some general trends. Furthermore, we feel that our assumptions are a good gross model of system behavior.

In future work we will also analyze the situation where we have only an exit point MIX-firewall as shown below.

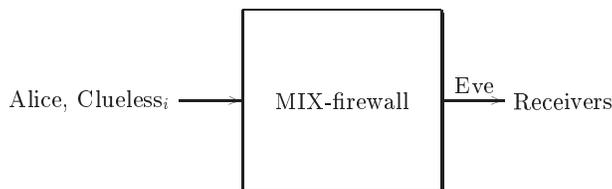


Fig. 14. Scenario with exit point MIX-firewall only.

We have M receivers denoted R_1, \dots, R_M . Eve still does not know directly who sent a message, but Eve does know where messages are going. This increase the capacity of the covert channel. Alice now instead of just sending 0 or 0^c can send: 0 (not transmitting); 1 (message to the first receiver), ... , i (message to the i th receiver, ... , M (message to the M th receiver). The greatest the capacity can be is $\log(M + 1)$. Of course if $M = 1$ the situation reduces to Scenario 2.

As before the Clueless_i are assumed independent and one may allow their distributions to be identical or they may vary.

Related to this is an intermediate question of the nature and capacity of covert channels in a network of MIXes (with Eve as GPA or Eve as an RPA restricted to observing the traffic between MIX-firewalls). Now there are $\text{Clueless}_{i,j}$'s

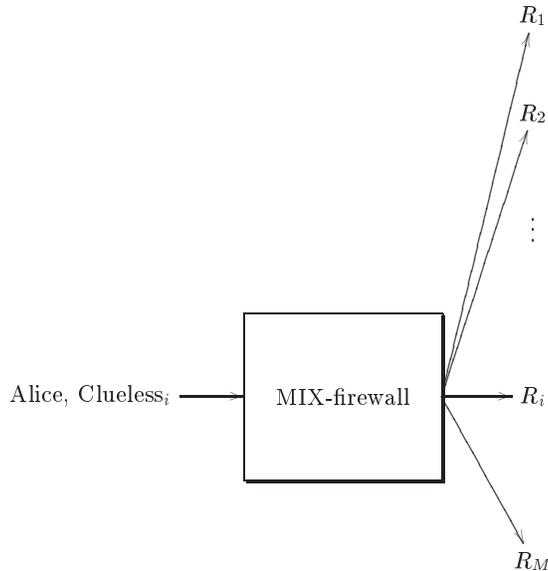


Fig. 15. Exit firewall only

at every ingress MIX and Receiver $_{i,j}$'s at every egress MIX, again with a variety of possible characteristics.

Other areas begging for further investigation include scenarios in which there is limited network capacity (on links or aggregate), and whether or not there is anonymity. We are currently investigating this using the model in which at most B messages can be sent through the network (as output from a sender or as output of a MIX-firewall) in a given tick, and if there are more than B messages awaiting transmission, B of them are chosen at random for delivery. This may relate the work to more sophisticated MIX models, such as pool MIXes, which is also desirable.

A deeper issue raised in this preliminary paper is that of the relationship between anonymity and covert channel capacity (fixing the other factors that affect capacity). It seems evident that as system level anonymity increases in the simple models shown here (i.e., the number of potential senders increases), the minimum capacity decreases to zero. However, as the probability that a clueless sender transmits in a given tick increases, the expected number of actual senders in a given time tick also increases, hence the anonymity increases, but the capacity of the covert channel increases once this probability exceeds 0.5. The relationships are not simple, but their discovery has the potential to increase our understanding of fundamental aspects of network design.

4 Acknowledgements

We are grateful to Paul Syverson for his discussions about anonymity, to LiWu Chang for his assistance with the mathematical results, and also to Gerry Allwein.

References

1. The anonymizer. <http://www.anonymizer.com/>.
2. Oliver Berthold, Hannes Federrath, and Stefan Köpsell. Web MIXes: A system for anonymous and unobservable internet access. In Hannes Federrath, editor, *Designing Privacy Enhancing Technologies: Design Issues in Anonymity and Observability*, pages 115–129. Springer-Verlag, LNCS 2009, July 2000.
3. Oliver Berthold, Andreas Pfitzmann, and Ronny Standke. The disadvantages of free MIX routes and how to overcome them. In Hannes Federrath, editor, *Designing Privacy Enhancing Technologies: Design Issues in Anonymity and Observability*, pages 27–45. Springer-Verlag, LNCS 2009, July 2000.
4. David Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 24(2):84–88, 1981.
5. Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, 1991.
6. Claudia Díaz, Stefaan Seys, Joris Claessens, and Bart Preneel. Towards measuring anonymity. In Paul Syverson and Roger Dingledine, editors, *Privacy Enhancing Technologies (PET 2002)*. Springer-Verlag, LNCS 2482, April 2002.
7. Robert G. Gallager. *Information Theory and Reliable Communication*. Wiley, 1968.
8. S. Kent and R. Atkinson. Security architecture for the Internet Protocol, 1998.
9. E.E. Majani and H. Rumsey. Two results on binary input discrete memoryless channels. In *IEEE International Symposium on Information Theory*, page 104, June 1991.
10. Ira S. Moskowitz and Myong H. Kang. Covert channels — here to stay? In *Proc. COMPASS'94*, pages 235–243, Gaithersburg, MD, June 27- July 1 1994. IEEE Press.
11. Richard E. Newman, Ira S. Moskowitz, Paul Syverson, and Andrei Serjantov. Metrics for traffic analysis prevention. In *PET 2003*, Dresden, March 2003.
12. R. E. Newman-Wolfe and B. R. Venkatraman. High level prevention of traffic analysis. In *Proc. IEEE/ACM Seventh Annual Computer Security Applications Conference*, pages 102–109, San Antonio, TX, Dec 2-6 1991. IEEE CS Press.
13. R. E. Newman-Wolfe and B. R. Venkatraman. Performance analysis of a method for high level prevention of traffic analysis. In *Proc. IEEE/ACM Eighth Annual Computer Security Applications Conference*, pages 123–130, San Antonio, TX, Nov 30-Dec 4 1992. IEEE CS Press.
14. Onion routing home page. <http://www.onion-router.net>.
15. Andreas Pfitzmann and Marit Köhntopp. Anonymity, unobservability and pseudonymity — a proposal for terminology. In Hannes Federrath, editor, *Designing Privacy Enhancing Technologies: Design Issues in Anonymity and Observability*, pages 1–9. Springer-Verlag, LNCS 2009, July 2000.
16. J. Postel. User Datagram Protocol, 1980.

17. A.P. Prudnikov, Yu. A. Brychkov, and O.I. Marichev. *Integrals and Series, Volume 1*. Gordon and Breach, 1986.
18. Michael K. Reiter and Aviel D. Rubin. Crowds: anonymity for web transactions. *ACM Transactions on Information and System Security*, 1(1):66–92, 1998.
19. I.J. Schwatt. *An Introduction to the Operations with Series, 2nd edition*. Chelsea, 1924.
20. Andrei Serjantov and George Danezis. Towards an information theoretic metric for anonymity. In Paul Syverson and Roger Dingledine, editors, *Privacy Enhancing Technologies (PET 2002)*. Springer-Verlag, LNCS 2482, April 2002.
21. Claude E. Shannon. The mathematical theory of communication. *Bell Systems Technical Journal*, 30:50–64, 1948.
22. Claude E. Shannon. The zero error capacity of a noisy channel. *IRE Trans. on Information Theory*, Vol. IT-2:S8–S19, September 1956.
23. B. R. Venkatraman and R. E. Newman-Wolfe. Transmission schedules to prevent traffic analysis. In *Proc. IEEE/ACM Ninth Annual Computer Security Applications Conference*, pages 108–115, Orlando, FL, December 6-10 1993. IEEE CS Press.
24. B. R. Venkatraman and R. E. Newman-Wolfe. Performance analysis of a method for high level prevention of traffic analysis using measurements from a campus network. In *Proc. IEEE/ACM Tenth Annual Computer Security Applications Conference*, pages 288–297, Orlando, FL, December 5-9 1994. IEEE CS Press.
25. B. R. Venkatraman and R. E. Newman-Wolfe. Capacity estimation and auditability of network covert channels. In *Proc. IEEE Symposium on Security and Privacy*, pages 186–198, Oakland, CA, May 8-10 1995. IEEE CS Press.