2008-06

Conditional Entropy for Deception Analysis, 13th ICCRTS: C2 for Complex Endeavors

Custy, John; Rowe, Neil C.

Monterey, California. Naval Postgraduate School

International Command and Control Research and Technology Symposium,
Bellevue, WA, June 2008.
http://hdl.handle.net/10945/36946

13th ICCRTS: C2 for Complex Endeavors

Conditional Entropy for Deception Analysis

Topic 4: Cognitive and Social Issues

John Custy (contact)
SPAWAR Systems Center
Code 53527, San Diego, CA 92152
john.cusyt@navy.mil / 619-553-6167

and

Neil C. Rowe
Naval Postgraduate School
Code CS/RP, 1411 Cunningham Road
Monterey, CA 93943
ncrowe@nps.edu / 831-656-2462

Conditional Entropy for Deception Analysis

Abstract

This paper describes how basic concepts from information theory can be used to analyze deception. We show how a general definition of deception can be mapped to a simple communication model known as a Z-channel, and we show that any deception has associated with it a closely related deception we call it's symmetric complement. These ideas allow computation of a specific form of conditional entropy which indicates the average uncertainty, in bits, that a deception imposes on a deception target. This uncertainty provides unique and general insight into a deceptions performance, and also indicates the general counter-deception potential available to a deception target. We then describe two deception-based mechanisms for computer security: the fake honeypot serves to inoculate a computer against intrusions; and the spoofing channel provides a safe and effective means for responding to in-progress computer intrusions. The spoofing channel is of fundamental interest because it is a deception equal to its symmetric complement.

Introduction

One of the most powerful and cost-effective ways of countering an adversary is through deception. Intuition suggests that deception and communication are ‘opposites’ or ‘duals’ in some sense, but any relationship that may exist has never been explored carefully. This is unfortunate because the past few decades have seen sophisticated mathematical tools developed to characterize communication systems ([1], [2], [3]), yet few if any techniques currently exist for the mathematical analysis of deception [4].

In this paper elementary concepts from information theory are used to explore the relationship between communication and deception. Our work is based on the idea that deception is, most generally, an attempt to convince a deception target that some specific incorrect version of reality holds. For a deception to succeed, one bit of valid information about the environment, or ‘reality,’ must be incorrectly inferred by the target. A specific form of conditional entropy quantifies the average uncertainty a target has about the state of their environment, and this uncertainty, together with the deception's probability of success, form a useful measure of a deception's performance. Fundamentally, these results are due to communication and deception models that differ only in how the various participants are mapped onto components of an underlying channel model; the channel model itself is the same for both deception and communication.

It is important to note that our approach gives information only about the average effectiveness of a deception, and not about the outcome of any particular deception encounter. In a similar way, analysis of a communication system will usually not reveal whether a specific symbol transmitted at a specific instant will be received correctly. Rather, in both cases performance is characterized only in terms of averages, and our model is thus most useful for describing the average performance of deceptions that are applied repeatedly.

This paper is structured as follows. We begin by briefly summarizing some basic concepts and terms regarding communication systems. A definition for deception is then given, and we show how this definition allows deception to be mapped to a communication channel model known as a Z-channel. Several common deceptions are used to illustrate this mapping. With this mapping in place, the conditional entropy of an environment with respect to a target's inferences can be computed. This conditional entropy is a measure of the average uncertainty about the environment imposed on a deception target by the deception. We then describe two software tools, the fake honeypot and the spoofing channel, that use deception to support computer security. We close with a summary and brief description of our current work.

Background on Communication Systems

In this section we give an overview of some basic concepts associated with communication systems, with special emphasis on abstract models of binary communication systems like those shown in Figure 1. Models of this sort are commonly used to describe the types and quantities of random errors that binary communication channels introduce as they transfer symbols from one location to another [2].

The two channel models shown in Figure 1 differ in the types of random errors they introduce as they convey a stream of binary symbols from input to output. On the left is a *Z-channel*, which can cause an error in only one of the two symbol types. In general, the input stream $ABAB$ might be

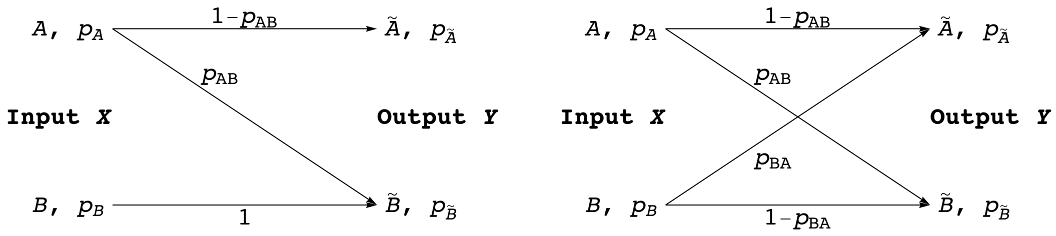


Figure 1: Abstract representations of two different types of discrete binary communication channel. On the left is a Z-channel, and on the right is a Bow-Tie Channel.

delivered as $\tilde{B}\tilde{B}\tilde{B}\tilde{B}$, but could not be delivered as $\tilde{A}\tilde{A}\tilde{A}\tilde{A}$. In other words, reception of an $\tilde{A}$ can only be caused by transmission of an A, but reception of a $\tilde{B}$ can be caused by transmission of either an A or B.

It should be recognized that the specific symbols used to represent inputs and outputs are unimportant, because the phenomena of interest for communication is the statistical *relationship* between the input and output. Two strongly correlated random variables, for example, convey information about each other regardless of how their values are labeled.

On the right of Figure 1 is a *Bow-Tie Channel*, or BTC, which is capable of introducing an error into either of the two symbol types it conveys. In general—that is, as long as the transition probabilities p_{AB} and p_{BA} are both away from zero and one—the input stream $ABAB$ might be delivered as $\tilde{B}\tilde{B}\tilde{B}\tilde{B}$ or as $\tilde{A}\tilde{A}\tilde{A}\tilde{A}$. Stated differently, reception of an $\tilde{A}$ can be caused by transmission of either an A or B, and reception of a $\tilde{B}$ can likewise be caused by transmission of either an A or B.

The special case of the BTC with $p_{AB} = p_{BA} = 0$ is known as the *noiseless channel*, and the special case with $p_{AB} = p_B$ and $p_{BA} = p_A$ is known as the *useless channel*. The Z-channel is simply a special case of the BTC with $p_{BA} = 0$, and the BTC can be viewed as a combination of two 'flipped' Z-channels operating on the same input symbol stream.

The probability of a successful transmission—that is, the probability that a given input symbol will be delivered as a specific output symbol—can easily be computed for either of the channel models shown in Figure 1. For the Z-channel, the probability of a successful transmission is

$$\begin{aligned} p_s &= p_A(1 - p_{AB}) + (1 - p_A) \\ &= 1 - p_A p_{AB}, \end{aligned} \tag{1}$$

and the corresponding expression for the BTC is

$$\begin{aligned} p_s &= p_A(1 - p_{AB}) + (1 - p_A)(1 - p_{BA}) \\ &= 1 - p_A p_{AB} + p_A p_{BA} - p_{BA}. \end{aligned} \tag{2}$$

Surprisingly, however, the probability of successful transmission is not a useful indicator of communication performance. To see this, note that a noiseless BTC has $p_s = 1$, and all transmissions are successful. A channel that causes an error in every transmission, that is, a BTC with $p_{AB} = p_{BA} = 1$, has $p_s = 0$, which indicates that no transmissions are successful. However, we need only invert the symbols at the output of the channel to achieve perfect communication. Even more troubling is that when $p_{AB} = p_{BA} = 1/2$, about half of the transmissions are received without error, yielding $p_s = 1/2$. However, in this case we do not know which transmissions are correct and which are incorrect. A BTC with $p_{AB} = p_{BA} = 1/2$ is of no value for communication despite it's $p_s = 1/2$.

A measure of communication performance can be obtained by, in essence, feeding a long input sequence to a channel and counting the number of output sequences that could result based on the probabilistic constraints imposed by the channel. When this count is low, communication performance is good, and when this count is high, communication performance is bad. To clarify this idea, consider the number of binary sequences, using symbols A and B, consisting of a single B and nine A's: there are $\binom{10}{1} = 10$ such sequences. Similarly, there are $\binom{10}{2} = 45$ sequences consisting of two B's within a sequence of eight A's, and there are $\binom{10}{5} = 252$ sequences consisting of five B's and five A's. As the number of B's increases beyond five, the number of possible sequences decreases.

Counts of this sort can be made for sequences of any finite length, and probabilities can be introduced by specifying the number of symbols of different types as fractions of the entire sequence length. Thus the number of ordered sequences of length N consisting of pN copies of the symbol A's and $(1 - p)N$ copies of the symbol B's is $\binom{N}{pN}$. Stirling's approximation implies that this count increases exponentially with N, and this rate of increase, expressed on a per symbol basis, is denoted *entropy* [3]. The entropy of a binary sequence composed of symbols with probabilities p and $1-p$ is

$$\begin{aligned} H(p) &\equiv \lim_{N\to\infty} \log \binom{N}{pN} / N \\ &= -p\log(p) - (1-p)\log(1-p). \end{aligned} \tag{3}$$

In concrete terms, Equation 3 states that for large N, there are about $2^{N\cdot H(p)}$ binary sequences of length N consisting of pN copies of the symbol A and $(1-p)N$ copies of the symbol B. An important characteristic of this set of $2^{N\cdot H(p)}$ binary sequences is that all members have the same probability: each sequence is composed of the same numbers of underlying symbol types (i.e., A's and B's), with sequences differing only in the arrangement of symbols. There are no distinguished or preferred sequences.

The reader should be aware that the terminology used to denote entropy in different fields is a potential source of confusion. Entropy is often denoted by S in thermodynamics [5], and by H in information theory ([2], [3]). In this paper we will follow the convention used in information theory and denote entropy by H.

To see how communication performance can be described for channels like those shown in Figure 1, imagine a long sequence of symbols received at the output of a BTC. The number of input sequences that could have generated this output sequence can be computed as follows. Each $\tilde{A}$ in the output stream could have been caused by input of an A or B, and more specifically each $\tilde{A}$ observed at the output was generated by input of an A with probability

$$p(A|\tilde{A}) = \frac{p_A(1-p_{AB})}{p_A(1-p_{AB}) + p_B p_{BA}}$$

or by input of a B with probability

$$p(B|\tilde{A}) = \frac{p_B p_{BA}}{p_A(1-p_{AB}) + p_B p_{BA}}.$$

These probabilities specify a value of entropy that we can denote $H(X|\tilde{A})$, with X indicating the input to the channel. In a similar way we can compute $H(X|\tilde{B})$. A weighted combination of these two entropies gives the entropy of the input to the channel, conditioned on the output of the channel. This conditional entropy, denoted $H(X|Y)$, is

$$H(X|Y) = p(\tilde{A})H(X|\tilde{A}) + p(\tilde{B})H(X|\tilde{B}).$$

An interpretation for $H(X|Y)$ is that an output of N symbols, with N large, could have been caused by about $2^{H(X|Y)\cdot N}$ different input sequences. When the conditional entropy $H(X|Y)$ is close to zero, communication performance is good because there are few input sequences that could have generated the given output sequence. Likewise, when $H(X|Y)$ is close to one, communication performance is poor because just about any input sequence could have caused the output. Stated in different terms, when many input sequences could have caused a given output, the uncertainty about the input is large even with the channel output in hand, and conversely, when few input sequences could have caused an output, the uncertainty about the input is small.

Our discussion in this section has focused on the conditional entropy of the input of a channel with respect to its output, but it should be clear that the conditional entropy of a channel's output with respect to its input, $H(Y|X)$, can be computed by similar means, and has an analogous interpretation: $H(Y|X)$ indicates the number of outputs that are consistent with the channel's probabilistic constraints for a given input.

A Deception Model Based on the Z-channel

Our work with deception is based on the following definition. The term *deception target*, or just *target*, refers to anyone who has a deception applied against them, whether or not they 'fall' for it. This definition is based on the discussion in [6], page 2.

Definition 1 Deception *is the presentation of a specific false version of reality by a deceiver to a target for the purpose of changing the target's actions in a specific way that benefits the deceiver.*

That is, deception is the imposition of a *specific* false version of reality onto an adversary: a deceiver does not simply cloak reality in an obscuring fog, but rather replaces reality with a specific and carefully created false version. Deception is thus distinct from the denial of information to an adversary, and distinct also from efforts that direct an adversary in a random, haphazard direction. As stated eloquently in [4], page 71, a successful deception will make an adversary "…quite certain, very decisive, *and wrong*" [emphasis in original].

This definition can be made precise. Any deception attempt will either succeed, or fail. A deception fails whenever the target is able to determine the correct state of the environment. This same outcome results when no deception attempt is made, assuming we ignore ordinary 'honest' mistakes on the part of the target. Alternately, when the evidence presented by the deceiver is strong enough, the target decides in favor of the false environment, and the deception succeeds.

The task carried out by the deception target is identical to that carried out by a binary communication receiver: each uses potentially noisy evidence to make a binary decision. This can be interpreted in terms of the Z-channel on the left of Figure 1. Under any given deception, the *actual*, correct, or valid state of the environment is represented by one symbol, say A. The false, incorrect, or *bogus* version of reality advocated by the deceiver is represented by the symbol B. In the absence of deception, the target is assumed able to correctly infer the state of the environment, but a deception attempt acts as noise in this sensor's observations, and can potentially cause incorrect states to be inferred.

With any given deception we will, to simplify terminology, associate a *partition* of the set of all possible environments. That is, any given deception splits the set of all possible environments into exactly two mutually exclusive and exhaustive subsets. In one of these subsets, the false version of reality specified by the deception actually is false, and the deception can be attempted within environments in this subset. This set of environments is denoted A in Figure 1. In the other set of environments, the 'false version of reality' specified by the deception is not false: it actually holds. In these environments the deception is logically impossible to carry out. These environments are denoted B in Figure 1. We will say that an environment is 'in state A' or 'in state B' when it is a member of set A or B, respectively.

In slightly different words, a deception defines a predicate on the set of possible environments, and a deception target attempts to determine, in the presence of interference from a deceiver, the value taken by this predicate in the actual, or realized, environment. Because A and B partition the set of environments—that is, A and B form mutually exclusive and exhaustive subsets—a given deception can only be attempted when the environment is in set A. Informally, imposing false evidence on a target that the environment is in state B makes no sense when the environment actually *is* in state B.

The Symmetric Complement of a Deception

Before illustrating these ideas with some examples, a subtle asymmetry in our definition of deception must be considered. In paraphrase, Definition 1 states that deception is an attempt by a deceiver to convince a target that the environment is a member of a particular set, which we have labeled B, when in actuality the environment is in another set, which we have labeled A.

This definition is asymmetric in the sense that deception involves an actual and a bogus environment, and these environments 'participate' in a given deception in very different ways. We will define the deception obtained by reversing these roles as the *symmetric complement* of the given deception. The symmetric complement of a deception is formed by complementing the defining predicate of the deception.

Like any other deception, a given symmetric complement can be represented as a Z-channel. However, a deception and its symmetric complement, when deployed together, are represented most naturally and concisely as two Z-channels coupled together into a BTC. We will call an ordinary, non-symmetric deception a *one-sided* deception, and a deception deployed along with its symmetric complement a *two-sided* deception.

Symmetric complements are nothing more than a way to complete, or make symmetric, the Z-channel deception model, and there is no logical reason to expect anything of practical value to result from abstract considerations of this sort. As the next section will show, however, it is not difficult to find examples in which full two-sided deceptions have been successfully deployed.

Examples

We show in this section how one-sided and two-sided deception models are applied to some familiar deceptions. Each example consists of a brief summary of the deception; a description of the participants and their motives; a statement of the actual and false versions of reality; a physical interpretation of the prior and transition probabilities; and a description of the symmetric complement. This information allows each ‘real-world’ deception example to be mapped to the channel models in Figure 1.

Examples 1-3 describe one-sided deceptions, and Examples 4 and 5 involve two-sided deceptions.

Example 1: Sale of a Low Quality Item as High Quality Consider a vendor with a large number of items to sell. Some of the items happen to be of low quality, and the vendor deceptively portrays these low quality items as items of high quality. Here the relevant characteristic of the environment is the quality of the item being sold in any given sales encounter. This can take on the values ‘Item is of high quality,’ denoted B, and ‘Item is of low quality,’ denoted A, with probabilities p_A and $p_B = 1 - p_A$ respectively. The relative frequency at which this deception succeeds—that is, the relative frequency at which low quality items appear as high quality—is given by the transition probability p_{BA}.

- *Sale of a low quality item as a high quality item* is a deception in which the **deceiver** is the seller, and the **deception target** is the buyer. The deceptive seller is motivated by the desire to obtain a high price for the item.
- The **actual version of reality,** A is that the item for sale is of low quality.
- The **bogus version of reality,** B is that the low quality item being sold is of high quality.
- The **prior probability** p_A is the fraction of items that are low quality.
- The **transition probability** p_{AB} is the fraction of low quality items that are successfully sold as high quality.
- The **symmetric complement** of this deception is the deceptive portrayal of high quality items as low quality.

Example 2: Income Tax Evasion Income tax evasion is another deception that can be described in terms of a Z-channel. In this case, a deceptive taxpayer presents to the taxing agency evidence (in the form of false statements, false documents, etc.) for an incorrectly low value of income. The environmental characteristic of interest is the income of the taxpayer, which can take on the value A, 'high income,' or B, 'low income.' The prior probability p_A is the fraction of taxpayers with high incomes, all of whom are assumed to deceptively attempt to appear as low income. The transition probability p_{AB} is the rate at which high income taxpayers successfully portray themselves as low income. Only a high income taxpayer is capable of attempting this deception.

- *Income tax evasion* is a deception in which the **deceiver** is a high-income taxpayer, and the **deception target** is a tax agent. The deceptive taxpayer is motivated by the desire to pay a smaller tax.
- The **actual version of reality** A is that the taxpayer has high income.
- The **false version of reality** B is that the taxpayer has low income.
- The **prior probability** p_A is the fraction of taxpayers who have high income.
- The **transition probability** p_{AB} is the fraction of high income taxpayers who successfully appear as low income.
- The **symmetric complement** of this deception is the deceptive portrayal of a low income taxpayer as high income.

Example 3: Camouflage Camouflage involves an intruder within a monitored spatial region. Presence of the intruder puts the environment in state A, and absence of an intruder puts the environment in state B. This deception can only be carried out by someone who is in the spatial region.

- *Camouflage* is a deception in which the **deceiver** is the intruder, and the **deception target** is a sentry. The deceptive intruder is motivated by the desire to prevent their intrusion from being detected.
- The **actual version of reality** A is that an intruder is present within the surveilled region.
- The **false version of reality** B is that no intruder is present within the surveilled region.
- The **prior probability** p_A is the fraction of times that an intruder is present in the surveilled region.

- The **transition probability** p_{AB} is is the fraction of times that an intruder within the surveilled region is not detected by a sentry.
- The **symmetric complement** of this deception is to cause an intruder to appear within a surveilled region when they are not.

Example 5: Feints An excellent example of a deception with a useful symmetric complement is the feint. In military conflicts, as well as in sports such as fencing and boxing, one party will give a false strike, which requires an opponent to defend or prepare to defend. The symmetric complement of a feint is a 'false feint,' or an actual strike.

- *Feints* are a deception in which the **deceiver** is a potential attacker, and the **deception target** is a potential attack victim. The deceptive attacker is motivated by the desire to make their attack, when it occurs, as effective as possible, and also to cause the opponent to expend effort preparing to defend against an attack that does not occur.
- The **actual version of reality** A is that an attack will not occur.
- The **false version of reality** B is that an attack is about to occur.
- The **prior probability** p_A is rate at which preparations for attack appear.
- The **transition probability** p_{AB} is the fraction of apparent attack preparations on the part of the deceiver that do not result in a real attack.
- The **symmetric complement** is an apparent false attack that becomes a real attack. The **transition probability** p_{BA} is the fraction of apparent attack preparations on the part of the deceiver that *do* result in a real attack.

Example 6: Dummy Aircraft in WWII Another deception with a useful symmetric complement is described in the following exchange regarding the use of 'dummy' aircraft to divert attacks away from real aircraft [7].

> Sometime around mid-1942, Major Oliver Thynne was a novice planner with Colonel Dudley Clarke's 'A' Force, the Cairo-based British deception team. From intelligence, Thynne had just discovered that the Germans had learned to distinguish the dummy British aircraft from the real ones because the flimsy dummies were supported by

> struts under their wings. When Major Thynne reported this to his boss, Brigadier Clarke, the ‘master of deception,’ fired back
>
> “Well, what have you done about it?”
>
> “Done about it, Dudley? What could I do about it?”
>
> “Tell them to put struts under the wings of all the real ones, of course!”

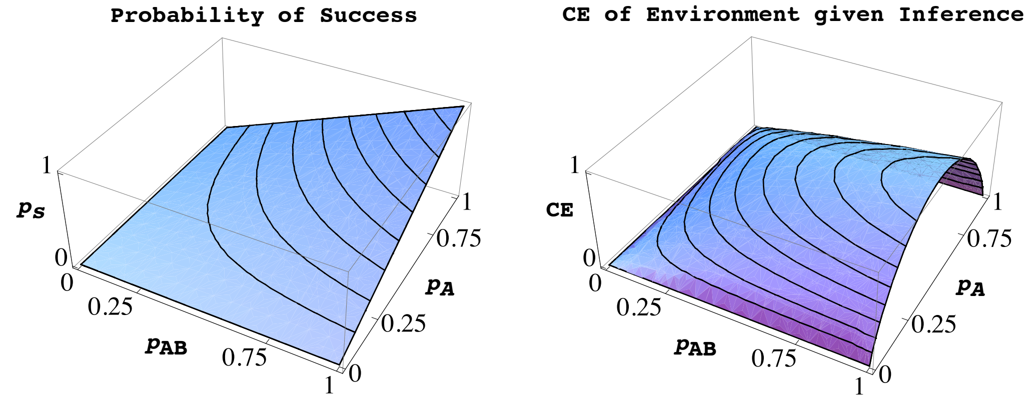
Here worthless dummies stand in place of valuable real aircraft. When the deception target learns to identify the dummies, the symmetric complement of the original deception is deployed.

- The *dummy aircraft* deception is a deception in which the **deceiver** is a group defending aircraft on a runway, and the **deception target** is a group attempting to surveil or attack the aircraft. This deception is motivated by the desire of the defenders to reduce the effectiveness of attacks or surveillance on their aircraft.
- The **actual version of reality** A, is that the dummy aircraft are worthless.
- The **false version of reality** B is that the dummy aircraft are valuable, real aircraft.
- The **prior probability** p_A is the fraction of all aircraft on a given runway that are dummies.
- The **transition probability** p_{AB} is the fraction of dummy aircraft that are believed to be real.
- The **symmetric complement** is the deception in which real aircraft are made to appear false. The transition probability p_{BA} is the fraction of real aircraft that are believed to be dummies.

A variant of this deception involves the use of appropriately painted canvas sheets to make bombed runways appear undamaged, and undamaged runways appear bombed.

Conditional Entropy and Deception

We have seen how the Z-channel and BTC describe one-sided and two-sided deceptions, respectively. These channel models are useful because they allow



Probability of Success

CE of Environment given Inference

Figure 2: Probability of Success (Left) and Conditional Entropy of the Environment given Inference (Right) for One-Sided Deceptions.

us to systematically evaluate the probability of success, or *effectiveness*, of a deception. Of much greater significance, these channel models also allow evaluation of the *uncertainty* about the environment imposed on a deception target by a deceiver. As the following two subsections show, both the effectiveness and the uncertainty associated with a deception are important for understanding the benefits and risks incurred by a deceiver.

One-Sided Deceptions

The surface on the left side of Figure 2 shows probability of success for a one-sided deception, and that on the right shows the conditional entropy of the environment based on the decisions made by the target. To interpret these figures, note that any point in the plane region $0 \leq p_A \leq 1$, $0 \leq p_{AB} \leq 1$ specifies a one-sided model of deception, and that we can informally interpret p_{AB} as the skill of the deceiver, and p_A as an indicator of how often deception opportunities arise.

The left of Figure 2 shows that with increasing opportunities for deception and/or increasing effectiveness of the deceiver, the probability of a successful deception also increases. The right side of Figure 2 shows that as the effectiveness of the deceiver increases, the targets uncertainty about the environment likewise increases, regardless of the value of the prior probability; this increase occurs monotonically as a function of p_{AB}, but at a slower than linear rate.

An ineffective deceiver (i.e., $p_{AB} \approx 0$), or an environment with little uncertainty (p_A near zero or one), causes conditional entropy to be close to zero. In these cases, the target's observations tend to be valid, or the state of the environment can be guessed effectively even without observations. In contrast, when the deceiver is effective (p_{AB} close to one), the target can be burdened with significant uncertainty: observations provide little help in inferring the environment. When a perfectly effective deceiver operates a one-sided deception, the target experiences a constant stream of observations implying that the environment is in state B, with no observations ever implying that the environment is in state A. This stream of observations provides no help in distinguishing state A from B.

The conditional entropy shown in Figure 2 is a concave function of p_{AB}, and a concave function of p_A. That is, if we let x and y be two one-sided deception models, and if we denote conditional entropy by f, then for any $0 \leq \alpha \leq 1$ we have

$$\alpha f(x) + (1-\alpha) f(y) \leq f(\alpha x + (1-\alpha) y) \tag{4}$$

The concave nature of conditional entropy is of some practical significance: it essentially implies that the average of two deceptions imposes less uncertainty on a target than the average deception. To understand this statement, consider for some fixed p_A the two deceptions located at $p_{AB} = 0$ and at $p_{AB} = 1$. The average of the conditional entropies associated with these two extreme deceptions is $1/2$; however, the conditional entropy of the average deception located at $p_{AB} = 1/2$ will be greater than $1/2$. The physical interpretation is that observations associated with $p_{AB} = 0$ are perfect, while those associated with $p_{AB} = 1$ are useless, but because they are associated with distinct deceptions, a deception target can handle them separately. On the other hand, the average deception at $p_{AB} = 1/2$ likewise provides both valid and useless observations, but because these observations are mixed together the deception target suffers greater uncertainty.

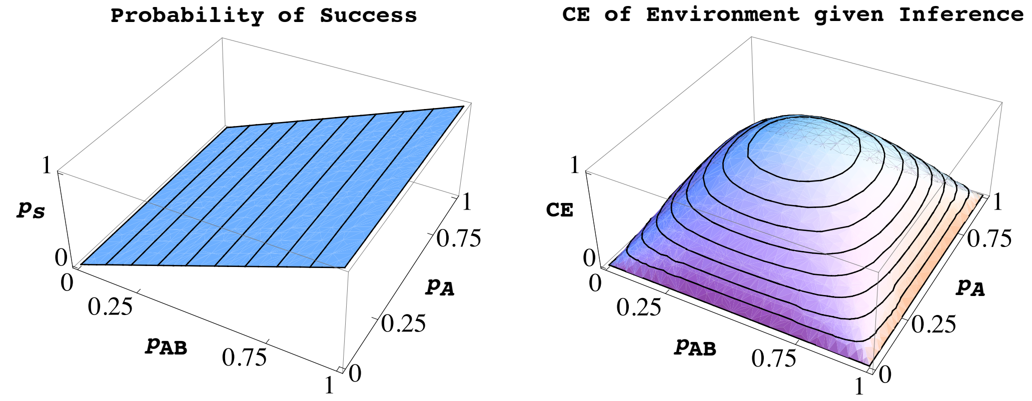


Figure 3: Probability of Success (Left) and Conditional Entropy of the Environment given Inference (Right) for Two-sided Deceptions with $p_{AB} = p_{BA}$.

Two-Sided Deceptions

A general description of a two-sided deception requires, in addition to the prior probability p_A, the two independent parameters p_{AB} and p_{BA}, and so complete information about effectiveness and conditional entropy is awkward to display graphically. We will instead consider the special two-sided deception with $p_{AB} = p_{BA}$, which is representative of the general case.

As shown on the left of Figure 3, since the environment can always support a two-sided deception, the probability of success tends to increase faster than a one-sided deception as a function of p_{AB}. (Note however that the probability of success is independent of the prior probability only in the special case of $p_{AB} = p_{BA}$). The right side of Figure 3 shows that some characteristics of conditional entropy for two-sided deceptions are the same as the one-sided case, for example, the low conditional entropy due to low uncertainty in the environment (i.e., p_A close to zero or one) or an ineffective deceiver ($p_{AB} \approx 0$).

However, certain phenomena of two-sided deceptions are unlike those of one-sided deceptions. In particular, when $p_{AB} \geq 1/2$, conditional entropy *decreases* as p_{AB} increases. That is, the target's uncertainty about the environment decreases as the deceiver becomes more effective, or in other words, the inferences made by the target (or target community) provide valid information about the environment. This is because, in a two-sided deception, the evidence provided by the deceiver can be used to infer the *correct* state of the environment when that evidence is very strong. A target can exploit this situation by deciding in favor of state A when observations indicate that the environment is in state B, and vice versa.

Under these circumstances, the deception is analogous to a binary communication system that delivers as output an 'inverted' form of the input: strictly speaking the system is making an error on every bit, but simply inverting the output stream provides perfect communication. Likewise, a two-sided deception with $p_{AB} = p_{BA} \approx 1$ is very effective, but the low conditional entropy indicates that this deception is vulnerable to a general counter-deception technique, namely, deciding *against* what the evidence supports. Of course, the target must be able to recognize these circumstances. As an example, in the two-sided form of the dummy aircraft example operating with $p_{AB} \approx p_{BA} \approx 1$, attackers can assume that anything that appears to be a real aircraft is a dummy, and anything that appears to be a dummy is a real aircraft.

Another interesting phenomena involves high values of conditional entropy. High values of conditional entropy $H(X|Y)$ indicate high uncertainty about the state of the environment. This occurs in one-sided deceptions when the deceiver is perfectly effective, in which case the deception target observes only evidence that the environment is in state B. In a two-sided deception, as mentioned above, when the deceiver is completely effective, that is, when $p_{AB} \approx 1$ and $p_{BA} \approx 1$, the deception is very effective but there is little uncertainty about the environment—the targets observations are strongly correlated with the state of the environment, whether the target realizes is or not. In a two-sided deception, high values of $H(X|Y)$ occur when the deceiver is only about 'half' effective, that is when $p_{AB} \approx 1/2$ and $p_{BA} \approx 1/2$. Under these circumstances, the target sees a mix of evidence in favor of states A and B, but that evidence is not correlated with the actual states of the environment. These observations are of no more value for inferring the state of the environment than are the flips of a coin.

A two-sided deception operating at $p_{AB} \approx 1/2$ and $p_{BA} \approx 1/2$ can be viewed as the most effective risk-free deception possible for the deceiver. Operating a two-sided deception with $p_{AB} \geq 1/2$ and $p_{BA} \geq 1/2$ can provide a higher probability of success, but also provides the target with some level of (negative) correlation between their inferences and the actual state of the environment. The target may be able to recognize and exploit this correlation.

Deception for Computer Security

The ideas discussed above are being used at the Naval Postgraduate School for development of two deception-based techniques for computer security. The *false honeypot* 'inoculates' a computer against intrusions, and the *spoofing channel* provides a method for responding to successful intrusions once they have been detected.

Honeypots and False Honeypots

One of the most effective ways of gathering information about computer intruders is through the use of *honeypots* ([8], [9]), which are computers placed on a network for the sole purpose of being broken into. Honeypots contain no information of value, and are highly instrumented to capture maximum information about intruders and their activities. Most computer intruders avoid honeypots to protect their intrusion techniques, which often require significant time and expertise to develop. Honeypots are sometimes 'tricked out' to appear as ordinary, non-honeypot computers containing valuable information.

The false honeypot deception can be summarized as follows.

- The **deceiver** is a computer administrator, and the **deception target** is a computer intruder. The deceiver is motivated by the desire to protect the computer under attack.
- The **actual version of reality** A is that the intruder has broken into a specially instrumented computer called a honeypot.
- The **false version of reality** B is that the intruder has broken into an ordinary computer.
- The **prior probability** p_A is the fraction of all computers on a given network that are honeypots.
- The **transition probability** p_{AB} is the fraction of honeypots that are taken to be ordinary.

- ○ The **symmetric complement** is the deception in which an ordinary computer is made to appear as a honeypot.

The fake honeypot deception is simply the symmetric complement of the honeypot deception. That is, a fake honeypot is an ordinary computer that has the appearance, or characteristics, of a honeypot [10]. Such characteristics can consist, for example, of files or daemons that are commonly associated with honeypots, but that would normally be hidden; on a fake honeypot they may be 'carelessly' hidden.

A network populated with honeypots and fake honeypots presents a computer intruder with some interesting challenges. What seems to be a successful intrusion may be nothing more than an intrusion onto a honeypot; this is true even without fake honeypots. However, in addition, a successful intrusion onto an ordinary computer may be dismissed as an intrusion onto a honeypot, potentially protecting valid data from harm.

A mix of honeypots and non-honeypots will reduce and potentially eliminate the value of an intruders observations for identifying the true nature of a computer. Intruders will have to rely on prior probabilities to guide their actions, causing honeypots to become even more valuable sources of information.

As with any two-sided deception, a potential counter-deception opportunity becomes available for high values of p_{AB} and p_{BA}. Under these circumstances, evidence that a compromised computer is a honeypot implies that the computer is in fact ordinary, and evidence that a computer is ordinary implies that it is a honeypot. This problem, if it were to occur, could be remedied by using *only* the symmetric complement of the honeypot deception: that is, by having all computers on a network, even honeypots, appear as honeypots.

Spoofing Channels

In a general sense, the purpose of communication is to convey the result of a random event, experiment, or selection, from one location to another:

> The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. ... The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be

> designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design. ([1], page 31, emphasis in original.)

A communication channel resolves uncertainty at its output about a choice or decision made at its input. This uncertainty, which is inherent in the communication process, provides a unique opportunity for deception. A *spoofing channel* exploits for deception the uncertainty that a conventional communication channel resolves.

In general terms, a spoofing channel works by delivering as output not the message provided to the channel as input, but rather an arbitrary message that has the same statistical structure as the input. The intention is to deliver another of the possible messages that could have been chosen for communication. One spoofing channel application is as an Intrusion Response System (IRS) for use with an Intrusion Detection System (IDS) such as SNORT ([8], [9]). Without an IRS of some sort, an alert from an IDS that an intrusion is in progress leaves the responsible system administrator with few options other than termination of the intruder's connection. This prevents further compromise of data, but also prevents information about the intruder from being gathered. An IRS based on the spoofing channel would prevent further compromise of data by delivering to the intruder nothing more than spoofs of the original data, and in addition would allow the intruder to be observed and information about them gathered as their 'intrusion' continued. Widespread use of IRS's based on spoofing channels would burden all computer intruders with uncertainty about whether hijacked data was valid or invalid.

In discussing the relationship between deception and channel models, we assumed that when no deception attempts are made, a potential deception target is able to correctly infer the state of the environment. This assumption is irrelevant for analysis of the spoofing channel deception because the purpose of ordinary communication is simply to resolve uncertainty about an environment that happens to be non-local in either space or time. If the output of a spoofing channel is statistically consistent with its input, no resources other than the prior probabilities are available at a channels output for deciding whether delivered data is valid or a spoof. For this reason we say that a properly operating spoofing channel is it's own symmetric complement.

Conclusion

In this paper we have not developed a mathematical theory of deception, but we have done something just about as good: we have shown how the existing theory of information provides a mathematical model of deception. As with any mathematical model, it's abstractions allow statements that are very general, and it's mathematical nature allows precision. Thus the material presented in this paper makes possible statements about deception that are both general and precise.

Our focus has been on the uncertainty associated with deception. We have shown that successful one-sided and two-sided deceptions impose very different types of uncertainty onto deception targets. Successful one-sided deceptions impose on a target a constant stream of evidence that the environment is in a given state; this constant stream is of no value for distinguishing which of two states the environment is actually in. A two-sided deception with both deceptions very successful provides a deception target with evidence that is negatively correlated with the actual state of the environment. This situation can provide the deceiver with a high level of success, but this negative correlation also provides the deception target with a general counter-deception method. General counter-deception techniques are unavailable when the target's observations are uncorrelated with the actual state of the environment, but this requires that the effectiveness of the deception be less than maximum.

A number of theoretical and practical topics remain open. We are particularly interested in development of a geometric interpretation of deception based on signal space concepts [3]; and analysis of deception as a game [11] with payoffs quantified by a combination of effectiveness and conditional entropy. Development and evaluation of a spoofing channel for natural language text is also part of our on-going work.

References

[1] Shannon, Claude E., and Warren Weaver. *The Mathematical Theory of Information.* University of Illinois Press, Chicago, Illinois, 1949.

[2] Cover, Thomas M., and Joy A.Thomas. *Elements of Information Theory, 2nd Ed.* John Wiley and Sons, 2006.

[3] Goldman, Stanford. *Information Theory.* Dover Publications, Mineola, NY, 2005.

[4] Whaley, Barton. *Stratagem: Deception and Surprise in War.* Artech House, Boston, Massachusetts, 2007.

[5] Greven, Andreas, Gerhard Keller, and Gerald Warnecke, Eds. *Entropy.* Princeton University Press, Princeton, NJ, 2003.

[6] Godson, Roy, and James J. Wirtz, Eds. *Strategic Denial and Deception: The Twenty-First Century Challenge.* Transaction Publishers, New Brunswick, New Jersey, 2002.

[7] Whaley, Barton. *Conditions Making for Success and Failure of Denial and Deception: Authoritarian and Transition Regimes.* Printed as Ch. 3 of [6].

[8] The Honeynet Project. *Know Your Enemy: Learning About Security Threats, 2nd Ed.* Addison-Wesley, New York, 2004.

[9] `http://www.snort.org`

[10] Rowe, Neil C., Binh T. Duong, and E. John Custy. “Fake Honeypots: A Defensive Tactic for Cyberspace.” Proceedings of the 7th IEEE Workshop on Information Assurance, U.S. Military Academy, West Point, New York, June 2006.

[11] Garg, Nandan, and Daniel Grosu. *Deception in Honeynets: A Game-Theoretic Analysis.* Proceedings of the 2007 IEEE Workshop on Information Assurance, United States Military Academy, West Point, NY, 20-22 June 2007.