Simmons' Subliminal Channel

Héctor Rosario

Héctor Rosario (hrosario@math.uprm.edu) is an Assistant Professor at the University of Puerto Rico, Mayagüez Campus. He received his Ph.D. at Columbia University in May 2003 under the supervision of Profesors Henry O. Pollak and Bruce R. Vogeli. He is interested in the history of mathematics, teacher education, and the education of the gifted.

In 1983 Gustavus J. Simmons gave a talk titled *The Subliminal Channel* during the CRYPTO 83 meeting. This paper would unofficially start the systematic study of modern steganography, the study of methods used to conceal the existence of messages. Somewhat similar to cryptography, it has gained unprecedented attention ever since *USA Today* and *The New York Times* reported that al-Qaeda and other groups are increasingly relying on steganographic techniques to assist them in plotting their attacks ("Terrorist groups hide behind Web encryption," *USA Today*, June 19, 2001, and "Veiled messages of terror lurk into cyberspace," *The New York Times*, October 30, 2001).

Simmons began his landmark talk by describing a situation that has since become the model for communication protocols. In the problem, two accomplices in a crime, say Alice and Bob,[1] are arrested and placed in separate cells. The warden, knowing that the prisoners want to communicate with each other to plan an escape, but likewise hoping to gain some insight into the minds of the prisoners, will allow them to communicate with each other provided he can monitor the contents of their communications. However, the prisoners will only communicate, taking a risk of deception, if they are allowed to authenticate their messages to prevent the warden from sending fraudulent messages. Hence, in order to deceive the warden, Alice and Bob will have to establish a "subliminal channel," that is, a way of sending secret information within the limitations imposed on them. The solution of this problem, known as the prisoners' problem, is the cornerstone of modern steganography.

Authentication without secrecy

Before we realize the subliminal channel, it is important to consider the case when the prisoners simply communicate without trying to exchange secret information, also known as the *authentication without secrecy channel*.

Authentication is simply the process by which Alice introduces prearranged redundant information into a message, which serves as a guarantee to Bob that the message is genuine. Conversely, the absence of such information implies that the communi-

[1] It is historically relevant to note that although many people attribute the "Alice and Bob" naming convention for describing communication exchanges to Simmons, it was Rivest, Shamir, and Adleman who coined the term in 1978. Their landmark paper, *A Mathematical Method for Obtaining Digital Signatures and Public Key Cryptosystems*, introduced the mathematical theory behind the algorithm that even today provides the best security for encrypting messages via public key: the famous RSA algorithm.

cation is not genuine. Needless to say, the warden must allow Alice to encrypt the authentication information, lest the warden just strip off the authentication information and append it to another message. Authentic messages may be required to end in a minimum number of zeroes or in a particular suffix, as is the common practice in military authentication systems. Generally, "the message along with the authenticating information is block encrypted into a cipher using either a single key or a two key cryptoalgorithm" [**2**]. Provided the cryptoalgorithm is adequately secure, the probability $P(A)$ of the warden successfully choosing a cipher that will be accepted by Bob as genuine is related to the information content H_r of the redundant authenticating information. If the information content is, say, 3-bit long, then since for each bit we have two options, the probability of success is 2^{-3}. Hence, if it is H_r-long, $P(A)$ will optimally be 2^{-H_r}.

The warden corroborates that nothing has been concealed in the message by decrypting the cipher. If Alice uses a single-key cryptoalgorithm for encryption, then she must provide the warden with an encryption/decryption key *after* the exchange takes place; if a two-key system is used, then she gives him a decryption key in advance. For single-key cryptographic systems, the warden must be willing to bear the risk of allowing them to communicate prior to his knowing of the contents. However, this should not be a major concern since if he does not receive the decryption key corresponding to the cipher exchange as agreed, he can punish the inmates, say by not allowing further communication between them. However, if the message is long enough, the warden might face an unacceptable level of risk of covert communication. If the warden argues that to avoid bearing this risk he prefers to have the key before hand, Alice will argue that then he would have the unacceptable advantage of being able to forge messages. This problem is solved using a two-key cryptosystem, since this removes the need for even a temporary "trust" by either party since the warden "can have the decryption key in his possession prior to any exchange of messages, and hence have the ability to verify the message content prior to forwarding the cipher" [**2**]. In contrast, since the warden cannot infer the unknown encryption key, Alice is confident that he cannot improve his guessing odds of choosing an acceptable cipher. Actual authentication is frequently much more complex than this simplified description suggests, but as with almost all mathematics, simplified problems often give insight into more complex ones.

The essential points to an authentication without secrecy channel are that

a) Bob authenticates a message through the presence of H_r bits of redundant—yet expected—information in the decrypted cipher, and

b) the warden verifies that nothing has been concealed by decrypting the ciphers and verifying that the resulting message is precisely what he expected based on his foreknowledge of the message.

Although there are operational differences for the warden depending on whether a one- or two-key system is used, namely, that with the former he can check for concealed information prior to the information exchange, and afterwards with the latter. "However this does not alter the way in which he satisfies himself that nothing is concealed—namely, that the cipher decrypts to the expected message" [**2**].

The subliminal channel realized

If Alice wants to communicate H_m bits of information with H_r bits of authentication, then a total of $H_m + H_r$ bits must be sent. Here Simmons conceives the remarkably simple, yet extremely powerful notion of a subliminal channel. Alice might be willing to give up some of her ability to authenticate and use the resulting information capacity, say H_s bits, to communicate secretly. Needless to say, the warden is unaware of this additional risk Alice is taking. For example, suppose the warden tries to deceive Bob sufficiently many times, and that he is told each time whether he succeeded or not, then he would eventually find that his probability of success was $2^{-(H_r - H_s)}$ rather than 2^{-H_r}, as would be the case for the authentication channel agreed to in advance by the warden and Alice.

Let us illustrate these notions with the smallest single key example possible. Let $H_m = H_s = H_r - H_s = 1$, as in Figure 1. Notice that the message consists of a single bit of information, like the outcome of a fair-coin toss, and two bits allocated for authentication, out of which one is subverted for a subliminal channel. Notice also that the original authentication part has a probability of deception of $2^{-H_r} = \frac{1}{4}$, whereas now the probability of deception is $2^{-H_r} = \frac{1}{2}$. This is the risk Alice is willing to take, again, mainly because the warden is not aware of this move and, more importantly, because she has a need to communicate covertly.

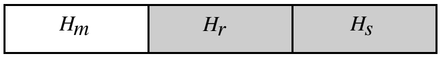

Figure 1. 3-bit signal with a 1-bit subliminal channel

For the first case, authentication without secrecy, our encoding rule relating the outcome of a fair-coin toss to three bit messages will be Heads if the number (encoded message) has even parity and Tails if odd. In other words, when the outcome is Heads, the warden will only allow even parity exchanges, where parity refers to whether the sum of the digits of the binary number representing the message is even or odd, as illustrated in Figure 2.

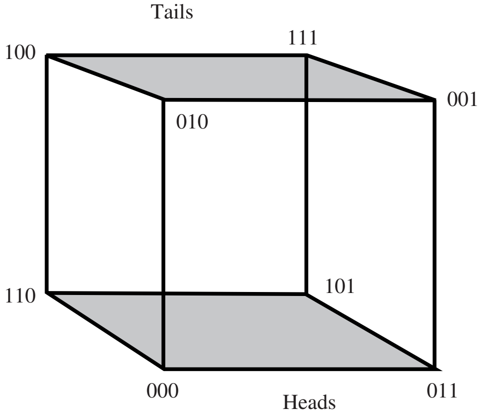

Figure 2.

In reality, what the warden and Alice have agreed to is a key list of the message space by keys such as the one tabulated in Figure 3. However, what the warden does not know is that, before they were apprehended, Alice and Bob randomly selected one of these keys.

There are two ways the warden might attempt to deceive Alice:

a) before the critical coin toss occurs, he could deliver a message of his own choice,
b) after the coin toss has occurred and Alice gives to the warden a message to communicate its outcome, the warden could substitute another message (of the opposite parity).

	Heads	Tails
1	000	100
2	000	010
3	000	001
4	000	111
5	011	100
6	011	010
7	011	001
8	011	111
9	101	100
10	101	010
11	101	001
12	101	111
13	110	100
14	110	010
15	110	001
16	110	111

Figure 3. Keys for a 1-bit channel and a 2-bit authentication without secrecy

Notice that even in the most general scheme, the warden has only these two types of deceit available: impersonation and substitution.

If the sixteen keys are equally likely to occur, then the warden's probability of successful impersonation is just the probability that the message he chooses appears in secret key, in this case, $\frac{4}{16} = \frac{1}{4}$. Alternatively, if he waits for the coin toss to occur, whichever message he is given lies in exactly four keys, corresponding to each of the messages of opposite parity. In this case, his probability of successful substitution is just the probability that he gets the unique message of opposite parity in the secret key, which again is $\frac{1}{4}$. Hence, regardless of whether the warden attempts impersonation or substitution, his probability of success is $\frac{1}{4}$.

Let us now consider a one-bit subliminal channel, where Alice and Bob select a secret key, not from the list of sixteen that the warden believes they are using, but rather from among the eight keys shown in Figure 4.

Each key in this list includes the labels of four corners from Figure 2 and each message is in precisely four of the keys. For example, the message 000 is in keys 1, 2, 3, and 4. The important thing to note about the entries in Figure 4 is that the middle bit in the eight entries in any column is the same as the subliminal message heading the column and that there are columns of even and of odd parity messages headed by both a 0 and by a 1. To communicate a subliminal bit, Alice chooses in the key row corresponding to the secret key she and Bob have chosen the even or odd message (depending on the outcome of the coin toss) that has the desired middle bit.

Given that Bob will now accept any of four different messages as genuine, two of odd parity and two of even parity, the warden's probability of successful impersonation is $\frac{4}{8} = \frac{1}{2}$. Likewise, if the warden waits to see a message sent by Alice, his probability

	Subliminal Message			
	0	1	1	0
1	000	011	111	100
2	000	011	010	001
3	000	110	010	100
4	000	110	111	001
5	101	110	111	100
6	101	110	010	001
7	101	011	010	100
8	101	011	111	001

Figure 4. Keys for a 1-bit overt channel, a 1-bit subliminal channel and a 1-bit authentication without secrecy channel

of successfully substituting a message of opposite parity to misinform Bob of the outcome of the coin toss will also be $\frac{1}{2}$ since the observed message appears in four of the keys, two of which contain each of the messages of opposite parity. Were he to make sufficiently many attempts at deception with this scheme, with an unrealistic "success" feedback on each attempt, he could quickly notice that something strange is happening since the estimate of his probability of success converges to $\frac{1}{2}$ rather than $\frac{1}{4}$. However, in actual applications this is not a problem since the authentication part is chosen to be small enough so that the probability of success is very small, making it practically impossible to estimate it by testing the channel. Furthermore, our simplified model does not provide the essential capability of key changes with successive messages, an essential property in steganography if the warden is to be denied the option of simply substituting an old message from a previous exchange.

As deceptively simple as these ideas might seem, they could have a catastrophic impact on treaty verification and on the national security of a country. In fact, as Burmester et al. [**1**] state, "The second Strategic Arms Limitation Treaty (SALT II) between the former USSR and the USA allowed both nations to authenticate their messages. While it was not possible to hide covert (embedded) data in the messages that were standardized (and could be controlled by the other nation), this was not the case with the authenticators. The capacity of this channel was enough to reveal which silos were loaded with nuclear missiles and which were not. This information would have been crucial to whichever power contemplated a first strike."

References

1. M. Burmester, Y. G. Desmedt, T. Itoh, K. Sakurai, H. Shizuya, M. Yung: A progress report on subliminal-free channels, *Lecture Notes in Computer Science 1174* (1996), 157–168.
2. G. J. Simmons, The prisoners' problem and the subliminal channel, *Advances in Cryptology: Proceedings of Crypto 83* (1983) 51–67.
3. G. J. Simmons, The subliminal channel and digital signatures, *Lecture Notes in Computer Science* (1985) 364–378.