Queen Mary
University of London

MAS 335 **Cryptography**

Notes 5: Stream ciphers, continued **Spring 2008**

One-time pads

A one-time pad is a stream cipher whose key is a random sequence of symbols from the alphabet. This means that, if the size of the alphabet is q and the length of the key is n, then each of the q^n keystrings has probability $1/q^n$. Said another way, each alphabet symbol is equally likely to appear in each position, and the symbols in the various positions are mutually independent. *Shannon's Theorem* asserts that a one-time pad is secure against statistical attack.

Probability: revision

We have a set Ω consisting of all possible *outcomes* of some experiment. *Events* are subsets of Ω; we require that the complement of an event is an event, and that the union of events $A_0, A_1, A_2, \ldots$ is an event. *Probability* is a function P from the set of events to the real numbers satisfying the *Kolmogorov axioms*:

(K1) $P(A) \geq 0$ for any event A.

(K2) $P(\Omega) = 1$.

(K3) If $A_0, A_1, A_2, \ldots$ are *pairwise disjoint* (that is, $A_i \cap A_j = \emptyset$ for $i \neq j$), then

$$P(A_0 \cup A_1 \cup A_2 \cup \cdots) = P(A_0) + P(A_1) + P(A_2) + \cdots.$$

As defined, P is just a function satisfying the axioms; but there are two common ways to think about it:

- $P(A)$ is the limiting frequency of occurrence of A in a long sequence of independent trials (e.g. if I toss a fair coin repeatedly it should come down heads about half the time, so $P(\text{heads}) = 1/2$);

- $P(A)$ is my estimate or calculation of the likelihood that the event A will occur.

Suppose that B is an event with non-zero probability. If I am given the information that B has occurred, this will change my estimate of how likely A is to occur. The *conditional probability* of B given A is

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

If $P(A \mid B) = P(A)$, then knowledge of B does not affect my estimate of the probability of A; we say that A and B are *independent*.

Shannon's Theorem

How does this apply to cryptography?

Suppose that Alice sends an encrypted message to Bob, which is intercepted by Eve.

Before Eve looks at the message, she will have some information about what the plaintext might be. For example, she may expect that it is much more likely to be in English than in Farsi; or she may expect that it will contain confidential information about Alice's bank details. So, if the sample space Ω consists of all q^n strings of length n over an alphabet of size q, then Eve will have some initial estimate $P(p = P_0)$ of the probability that the actual plaintext p is any given string P_0.

Once Eve intercepts the ciphertext Z_0, she uses it to try to decipher the message. If she is successful, and finds that the plaintext is P_1, then she will estimate the probability of P_1 as 1 and the probability of P_0 as 0 for any $P_0 \neq P_1$. These are conditional probabilities; that is,

$$P(p = P_1 \mid z = Z_0) = 1, \quad P(p = P_0 \mid z = Z_0) = 0 \text{ for } P_0 \neq P_1.$$

Even if she does not completely succeed, she is likely to gain some information: for example, she may eliminate some plaintexts completely, and show that others are more likely than she first thought. Thus, she calculates new conditional probabilities $P(p = P_0 \mid z = Z_0)$ which are in general different from the old ones $P(p = P_0)$.

We now specialise to stream ciphers. Recall that, for a stream cipher, the plaintext p, key k, and ciphertext z are all strings of length n over the same alphabet. We encipher one character at a time, by the rule that $z_i = p_i \oplus k_i$, where $x \oplus y$ is the entry in row x and column y of the substitution table. (In order for Bob to be able to decrypt, as we have seen, it is necessary that each column of the substitution table contains each symbol of the alphabet precisely once.)

A *one-time pad* is a stream cipher with the two properties

- the key is random (that is, $P(k = K_0) = 1/q^n$ for any string K_0), and of course is independent of the plaintext;
- the substitution table is a Latin square.

Theorem 15 (Shannon's Theorem) *Suppose that Alice uses a one-time pad. Then Eve's probabilities satisfy*

$$P(p = P_0 \mid z = Z_0) = P(p = P_0);$$

in other words, knowledge of the ciphertext gives no information about the plaintext.

Proof

$$\begin{aligned} P(p = P_0 \mid z = Z_0) &= \frac{P(p = P_0 \text{ and } z = Z_0)}{P(z = Z_0)} \\ &= \frac{P(p = P_0 \text{ and } k = K_0)}{P(z = Z_0)}, \text{ where } P_0 \oplus K_0 = Z_0 \\ &= \frac{P(p = P_0) \cdot P(k = K_0)}{P(z = Z_0)}, \end{aligned}$$

since plaintext and key are independent.

Now $P(k = K_0) = 1/q^n$, since the key is random. So it is enough to show that $P(z = Z_0) = 1/q^n$ to finish the proof.

The event $z = Z_0$ can happen in many ways; any choice of plaintext P' and key K' with $P' \oplus K' = Z_0$ will result in this event. So

$$\begin{aligned} P(z = Z_0) &= \sum_{P' \oplus K' = Z_0} P(p = P' \text{ and } k = K') \\ &= \sum_{P' \oplus K' = Z_0} P(p = P') \cdot P(k = K') \\ &= (1/q^n) \sum_{P' \oplus K' = Z_0} P(p = P') \\ &= 1/q^n. \end{aligned}$$

In the last line we use the fact that, since the substitution table is a Latin square, given any plaintext P' and ciphertext Z_0, there is a unique key K' such that $P' \oplus K' = Z_0$ – that is, in each row of the table, each possible symbol occurs precisely once. (We saw that the corresponding condition on columns is required of any substitution table; together these two conditions define a Latin square). So the sum is over all possible plaintexts P', each occurring once. Now the probabilities $P(p = P')$ add up to 1.

	0	1	2	3	4	5	6	7	8	9
0	4	9	5	3	2	7	0	1	6	8
1	7	5	0	9	3	2	1	8	1	4
2	3	1	7	2	8	0	9	6	9	7
3	0	8	4	7	0	1	3	4	5	2
4	5	3	2	4	9	3	8	2	7	6
5	9	0	1	6	7	5	4	7	2	3
6	2	6	8	0	0	9	7	5	3	1
7	6	2	6	1	4	8	6	0	8	5
8	1	7	9	7	1	4	5	9	0	7
9	8	4	3	5	5	6	2	3	4	0

Table 1: Japanese Army Air Force cipher J6633

In fact an even stronger property holds. If we already know the decryption of part of the ciphertext, then clearly this will alter our estimated probabilities for the rest of the text. However, knowledge of the ciphertext does not give any further information! We have seen that, for a widely used class of stream ciphers (those based on shift registers), this assumption is far from true: knowledge of the ciphertext and a small amount of plaintext enables the cipher to be broken completely.

More about Latin squares

Why do we need a Latin square for the substitution table in a stream cipher?

In the article "Japanese Army Air Force Codes at Bletchley Park and Delhi", by Alan Stripp, in the book *Code Breakers: The Inside Story of Bletchley Park* (edited by F. H. Hinsley and Alan Stripp), the following example is given of a substitution table supposedly used in the Japanese Army Air Force cipher J6633 (Figure 1).

By inspection, it is not a Latin square. It fails in various ways; for example,

(a) symbol 0 occurs twice in column 4 (in rows 3 and 6);

(b) symbol 1 occurs twice in row 1 (in columns 6 and 8).

The consequences of these two flaws are quite different.

Having a repeated element in a column means that the column is not a permutation of the alphabet, and so we cannot use the key to decrypt unambiguously. If the ciphertext letter was 0 and the corresponding key letter was 4, we wouldn't know whether the plaintext letter was 3 or 6.

Having a repeated element in a row does not stop us from decrypting the message. But it destroys the randomness of the key, and gives the cryptanalyst a small amount of leverage: the ciphertext string now carries a small amount of information about the plaintext.

To take this to extremes, suppose that we used a substitution square in which the columns were permutations but all rows were constant, say

	0	1	2	3	4	5	6	7	8	9
0	4	4	4	4	4	4	4	4	4	4
1	7	7	7	7	7	7	7	7	7	7
2	3	3	3	3	3	3	3	3	3	3
3	0	0	0	0	0	0	0	0	0	0
4	5	5	5	5	5	5	5	5	5	5
5	9	9	9	9	9	9	9	9	9	9
6	2	2	2	2	2	2	2	2	2	2
7	6	6	6	6	6	6	6	6	6	6
8	1	1	1	1	1	1	1	1	1	1
9	8	8	8	8	8	8	8	8	8	8

In this case, the plaintext letter 0 is always replaced by the ciphertext letter 4, regardless of the key. In other words, this is a simple substitution cipher, and the key is irrelevant. It can be broken by standard frequency analysis.

Latin squares are very plentiful. Their first practical use was in experimental design in statistics, where they were introduced by R. A. Fisher. (He is commemorated in Caius College, Cambridge, by a stained glass Latin square in a window of the dining hall.) In the early days of the subject, it was recommended that randomization of the experiment should include choosing a random Latin square for the design. The only way this could be done was by tabulating all Latin squares of relatively small order, and choosing one at random from the tables. (The famous tables of Fisher and Yates include such lists.) Subsequently this practice was abandoned. Now, however, a Markov chain method for choosing a random Latin square has been proposed by Jacobson and Matthews.

Another feature of Latin squares is that we can construct them by building up row by row. For $k \leq n$, we define a $k \times n$ *Latin rectangle* to be an array with entries from the set $\{1, \ldots, n\}$ such that each symbol occurs once in each row and at most once in each column. Now any $k \times n$ Latin rectangle with $k < n$ can be "completed" to a Latin square.

Worked example A message in a 3-letter alphabet $\{1,2,3\}$ has been encrypted using a random keystring and the substitution table

	1	2	3
1	2	3	1
2	1	2	2
3	3	1	3

The message has length 3. Before intercepting the ciphertext, your estimates of the probabilities of plaintext strings are

$$P(112) = 0.1, \quad P(231) = 0.2, \quad P(332) = 0.3, \quad P(313) = 0.4,$$

and all other probabilities zero.

You intercept the ciphertext 132. Calculate the conditional probabilities of the plaintext strings given this information.

Does your answer contradict Shannon's Theorem?

Solution We follow the argument in the proof of Shannon's Theorem. First we have to decide which keys would encrypt each possible plaintext as the given ciphertext. We see that $112 \oplus K = 132$ holds for $K = 322$ or 323 (the ambiguity because of the two occurences of 2 in the second row of the table). So $P(z = 132 \mid p = 112) = 2/27$. Similarly, $231 \oplus K = 132$ holds for $K = 111$ or $K = 131$, giving $P(z = 132 \mid p = 231) = 2/27$; and $332 \oplus K = 132$ holds for $K = 212, 232, 213, 233$, so that $P(z = 132 \mid p = 332) = 4/27$. Finally, $313 \oplus K = 132$ is impossible, since 2 does not occur in the third row of the table; so $P(z = 132 \mid p = 313) = 0$.

The Theorem of Total Probability gives

$$P(z = 132) = \frac{2}{27} \cdot \frac{1}{10} + \frac{2}{27} \cdot \frac{2}{10} + \frac{4}{27} \cdot \frac{3}{10} + 0 \cdot \frac{4}{10} = \frac{18}{270}.$$

From Bayes Theorem we find

$$\begin{aligned}
P(p = 112 \mid z = 132) &= \frac{(2/27) \cdot (1/10)}{18/270} = \frac{1}{9}, \\
P(p = 231 \mid z = 132) &= \frac{(2/27) \cdot (2/10)}{18/270} = \frac{2}{9}, \\
P(p = 332 \mid z = 132) &= \frac{(4/27) \cdot (3/10)}{18/270} = \frac{2}{3}, \\
P(p = 313 \mid z = 132) &= 0.
\end{aligned}$$

These are not the same as the prior probabilities, so we have gained some information. However, Shannon's Theorem is not contradicted, since one of its hypotheses asserts that the substitution table is a Latin square, which is not true in this case.

Orthogonal arrays; decryption tables

Suppose that we are using a stream cipher, where the substitution table is a Latin square L over the q-letter alphabet A. Thus, we encrypt the plaintext symbol p with key symbol k as the ciphertext symbol $z = p \oplus k$ occurring in row p and column k of the square. Decryption, however, is more complicated than this simple look-up; to decrypt ciphertext symbol z with key symbol k we must look in column k for the occurrence of symbol z, and then its row p is the plaintext symbol.

It would be much more convenient if we had another square, the *decryption square*, in which we could decrypt just by looking in row z and column k. Such a square does exist; it is called the *adjugate* of L, and is constructed as follows.

First, suppose we are given the alphabet A with q symbols, and also two positive integers k and t with $t \leq k$. An *orthogonal array* with these parameters is a matrix with k rows and q^t columns, with the following property:

> If we select any t of the k rows of the matrix, then for any choice of t symbols $x_1, \ldots, x_t$ from the alphabet A, there is exactly one column which has those entries in the t chosen rows.

From the rest of the definition, it follows that the number of columns must be q^t, since this is the number of ways of choosing t symbols from A.

Proposition 16 *A Latin square of order q is "equivalent" to an orthogonal array with $k = 3$ and $t = 2$ over an alphabet of q symbols, in the sense that each can be constructed from the other.*

Proof Suppose that we are given L. We label the three rows of the array as "rows", "columns" and "entries"; for each position in the square, if its row number is x, its column number y and the entry is z, we add a column $(x,y,z)^\top$ to the array.

This does give an orthogonal array. Let us check that for each pair x,z of symbols, there is exactly one column with x in the first row and z in the third. This means we are looking for a position in row x of the Latin square containing entry z; by definition there is exactly one such position.

The other two pairs of rows are checked similarly.

Example The Latin square

$$L = \begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & 1 & 2 & 3 \\ 2 & 2 & 3 & 1 \\ 3 & 3 & 1 & 2 \end{array}$$

gives us the 3×9 orthogonal array

$$\begin{pmatrix} 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3 \\ 1 & 2 & 3 & 1 & 2 & 3 & 1 & 2 & 3 \\ 1 & 2 & 3 & 2 & 3 & 1 & 3 & 1 & 2 \end{pmatrix}.$$

For example, the fact that the entry in row 2 and column 3 is equal to 1 is recorded in the sixth column of the array, $(2,3,1)^\top$.

Now suppose that L is the substitution table used for encrypting a cipher. Construct the orthogonal array corresponding to L. Then interchange the first and third rows of this array; the result is still an orthogonal array, so we can turn it back into a Latin square. The resulting square is called the *adjugate* of L.

In our example, the new orthogonal array is

$$\begin{pmatrix} 1 & 2 & 3 & 2 & 3 & 1 & 3 & 1 & 2 \\ 1 & 2 & 3 & 1 & 2 & 3 & 1 & 2 & 3 \\ 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3 \end{pmatrix},$$

so the adjugate of our original square is

$$L^\dagger = \begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & 1 & 3 & 2 \\ 2 & 2 & 1 & 3 \\ 3 & 3 & 2 & 1 \end{array}.$$

So, if we use L for encrypting, we use $L^\dagger$ for decrypting. For example, the first square encrypts 2 with key 3 as 1, while the second decrypts 1 with key 3 as 2.

It would be even nicer if the encryption and decryption squares were the same; then the risk of error caused by having two squares and needing to use the right one would be avoided. So we would like a Latin square L which is *self-adjugate*, that is, satisfies $L^\dagger = L$. Not all Latin squares satisfy this (we have seen that it fails for the above example). In particular, the *Vigenère square* or *addition square*, which has alphabet $\mathbb{Z}/(q)$ and (i,j) entry $i+j$ mod q, is not self-adjugate. However, there is a very simple example of a self-adjugate square of any order q, namely the *subtraction square*, in which the alphabet is again $\mathbb{Z}/(q)$ but the (i,j) entry is $j-i$ mod q. This works because, if $k = j-i$ mod q, then $i = j-k$ mod q.

As a final remark, we note that if we take the orthogonal array derived from a Latin square and permute the three rows in any manner, the result is still an orthogonal array, and can be converted back into a Latin square. So there are potentially six Latin

squares associated in this manner with a given one. Of these six, only the adjugate (corresponding to swapping the first and third rows) has an obvious application in cryptography. Note however that the operation of swapping the first and second rows of the array corresponds to taking the *transpose* of the Latin square.

We will see later that orthogonal arrays are used in "secret sharing schemes".

Appendix: Entropy

The concept of entropy originated in nineteenth-century thermodynamics as a measure of the disorder of a complicated physical system. Shannon introduced it into information theory, where it provides a very convenient measure of information. The background probability theory can be found in any book on the subject, or in Peter Cameron's *Notes on Probability* on the Web (they can be found at http://www.maths.qmul.ac.uk/%7Epjc/notes/prob.pdf).

Let X be a random variable on a probability space $\mathcal{S}$ with probability function P. (Recall that this simply means that X is a function on $\mathcal{S}$. In elementary probability theory we assume that the values of X are numbers, but they can be anything at all. Here we only consider finite probability spaces.) The entropy of X is a measure of our uncertainty about the value of X (or, equivalently, the amount of information we would gain if we performed an observation and learned the value of X). This interpretation suggests that the entropy of X should be zero if X is constant (since then measuring X will tell us nothing we don't already know) and maximum if all the values of X have the same probability.

The definition is as follows. The *entropy* of X is given by the formula

$$H(X) = -\sum_{i=1}^{n} P(X = x_i) \log_2 P(X = x_i),$$

where $x_1, \ldots, x_n$ are the possible values of X.

It is easily verified that X has the required properties:

Proposition 17 *(a) $H(X) \geq 0$, with equality if and only if there is a value x such that $P(X = x) = 1$.*

(b) If X takes n values $x_1, \ldots, x_n$, then $H(X) \leq \log_2(n)$, with equality if and only if $P(X = x_i) = 1/n$ for $i = 1, \ldots, n$.

This agrees well with our intuition that entropy measures our initial uncertainty, or the information we gain from performing the experiment. If one of the values has probability 1, then we know it will occur, and so we have no uncertainty. On the other hand, if all values are equally likely, our uncertainty is as large as possible.

Example Suppose that I toss a fair coin n times; the values of the random variable X are the 2^n possible bitstrings produced (where, say, heads $= 1$, tails $= 0$). Then $H(X) = \log_2 2^n = n$. That is, n random bits have entropy n. So the units of entropy are "bits"; observing a random variable X gives us "the same amount of information" as knowledge of $H(X)$ random bits.

If A is an event with non-zero probability, then the *conditional random variable* $X_A = X \mid A$ is defined by the rule that

$$\Pr(X_A = x_i) = \Pr(X = x_i \mid A) = \frac{\Pr(X = x_i \text{ and } A)}{\Pr(A)}.$$

The random variable $X \mid A$ now has entropy $H(X \mid A)$ according to the usual formula.

In particular, let X and Y be random variables. For each value y_j of Y, there is a conditional entropy $H(X \mid (Y = y_j))$. Then we define the conditional entropy of X given Y to be the weighted average (expected value) of $H(X \mid (Y = y_j))$; that is,

$$H(X \mid Y) = \sum_{j=1}^{m} H(X \mid (Y = y_j)) \Pr(Y = y_j),$$

where $y_1, \ldots, y_m$ are the values of Y.

A short calculation shows that

$$H(X \mid Y) = H(X,Y) - H(Y),$$

where $H(X,Y)$ is the entropy of the random variable $Z = (X,Y)$ whose values are pairs (x_i, y_j) of values of X and Y.

We interpret $H(X \mid Y)$ as the remaining uncertainty about X after doing an experiment to measure Y. Indeed, the following holds:

Proposition 18 *For any two random variables X and Y, we have $H(X \mid Y) \leq H(X)$, with equality if and only if X and Y are independent.*

Thus, if X and Y are independent, then knowledge of Y gives no information about X.

Let us apply these ideas to cryptography. If we are in Eve's position, we should regard the plaintext, key, and ciphertext as random variables. We will probably have some assumptions about the relative likelihood of various plaintext messages: a spy is unlikely to be sending a passage of Shakespeare as plaintext (though the plaintext may be hidden in a passage of Shakespeare, or Shakespeare's works may be used in another way in creating a cipher). This knowledge corresponds to a probability distribution on

the plaintexts, from which the entropy $H(P)$ of the plaintext can be calculated. (Here P is the random variable whose values are the actual plaintexts.)

Once Eve intercepts a ciphertext, she can in principle compute some information about the plaintext. This may be complete information (that is, Eve can decrypt the cipher), or perhaps just some change in the probabilities. The conditional entropy $H(P \mid Z)$ is Eve's remaining uncertainty about the plaintext given the ciphertext; it is zero if she can decrypt the message.

In this form, Shannon's theorem states:

> If Alice uses a one-time pad, then $H(P \mid Z) = H(P)$.

In other words, Eve gets no information about the plaintext from knowledge of the ciphertext.

Exercises

1. Let B be an event with non-zero probability. Show that the function P^* on events given by $P^*(A) = P(A \mid B)$ satisfies Kolmogorov's axioms (K1)–(K3).

2. Prove the unproved propositions in the section on entropy.