YALE UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

CPSC 467a: Cryptography and Computer Security *Notes 4 (rev. 1)*
Professor M. J. Fischer *September 15, 2008*

Lecture Notes 4

20 Hill cipher

The *Hill cipher* is another example of a cipher that encrypts groups of letters at once, thereby tending to mask letter frequencies. It is based on linear algebra. The key is, say, a non-singular 3×3 matrix K. The message m is divided into vectors m_i of 3 letters each. Encryption is just the matrix-vector product Kv. Decryption is the same using K^{-1}.

Unfortunately, the Hill cipher succumbs to a known plaintext attack. Given three linearly independent vectors m_1, m_2, and m_3 and the corresponding ciphertexts $c_i = Km_i$, $i = 1, 2, 3$, it is straightforward to solve for K.

21 Polyalphabetic ciphers

Another way to strengthen monoalphabetic ciphers is to use different substitutions for different letter positions. For example, one might choose 10 different alphabet permutations $k_1, \ldots, k_{10}$, use k_1 for the first letter of m, k_2 for the second letter, and so forth, repeating this sequence after every 10 letters. While this is much harder to break than monoalphabetic ciphers, it turns out that letter frequency analysis can still be used. Every 10th letter is encrypted using the same permutation, so the submessage consisting of just those letters still exhibits normal English language letter frequencies.

22 Transposition techniques

All of the methods discussed so far are based on letter substitution. Another technique is to rearrange the letters of the plaintext. For example, one might write a plaintext message into a matrix by rows and then read it out by columns. While transposition does not seem like a very powerful technique by itself, when used in combination with substitution techniques it can be quite effective. Most practical symmetric cryptosystems are built by composing together many stages of substitutions and transpositions.

23 Composition of Ciphers

Ciphers can be composed to create new ciphers. For example, suppose (E', D') and (E'', D'') are two ciphers. The composition is the cipher (E, D) with keys of the form $k = (k', k'')$, where

$$E_{(k',k'')}(m) = E''_{k''}(E'_{k'}(m))$$

and similarly,

$$D_{(k',k'')}(c) = D'_{k'}(D''_{k''}(m)).$$

We can express this using *functional composition*. Let f and g be functions such that the range of g is contained in the domain of f. Then the composition $h = f \circ g$ is the function such that $h(x) = f(g(x))$. Using this notation, we can write $E_{(k',k'')} = E''_{k''} \circ E'_{k'}$ and $D_{(k',k'')}(c) = D'_{k'} \circ D''_{k''}(m)$.

Composition may lead to a stronger cipher or it may not, and it can be difficult to analyze. Indeed, practical symmetric cryptosystems such as DES and AES tend to be built in this modular way as a composition of simpler systems. Each component offers little security by itself, but when composed, the layers obscure the message to the point that it is difficult for an adversary to recover. The trick is to find ciphers that do successfully hide useful information from a would-be attacker when used in concert.

24 Double Encryption

A special case of composition is when a cryptosystem is composed with itself. We call this *double encryption*. Each message is encrypted twice using two different keys k' and k'', so $E^2_{(k'',k')} = E_{k''} \circ E_{k'}$ and $D^2_{(k'',k')} = D_{k'} \circ D_{k''}$. We say (E, D) is the *underlying* or *base* cryptosystem and (E^2, D^2) is the *doubled* cryptosystem. The doubled cryptosystem has the same message and ciphertext space as the underlying cryptosystem, and the keyspace is $\mathcal{K}^2 = \mathcal{K} \times \mathcal{K}$. The size of the keyspace is squared, resulting in an apparent doubling of the effective key length. Although this makes a brute force attack much more costly, it does not always increase the security of a cryptosystem as much as one might naively think, for other attacks may become possible.

24.1 Example: Double Caesar

Consider Double Caesar, the Caesar cipher composed with itself. One might hope that double Caesar is more resistant to a brute force attack since now there are $26^2 = 676$ possible key pairs (k'', k'). Unfortunately, there are still only 26 possible distinct encryption functions and therefore only 26 possible decryptions of each ciphertext. This is because the Double Caesar encryption function $E^2_{(k'',k')}$ is the same as the single Caesar encryption function E_k with key $k = (k' + k'') \bmod 26$. Even though the key space was enlarged, the number of distinct encryption functions was not, and any attack on the Caesar cipher will work equally well on the Double Caesar cipher. (This is because to the attacker, there is no difference between the two systems. Eve neither knows nor cares how Alice actually computed the ciphertext; all that matters to Eve is probabilistic relationships between plaintexts and ciphertexts.)