CS 435: Introduction to Cryptography **2012 Fall**

Solutions 2

Professor Somesh Jha

1. Let $m \in \mathcal{M}$, and consider any distribution over $\mathcal{M}$ that assigns a probability of 0 to m. Then, for any $c \in \mathcal{C}$ for which $\Pr[C = c] > 0$, we have that

$$\begin{aligned}\Pr[M = m \mid C = c] &= \frac{\Pr[M = m \wedge C = c]}{\Pr[C = c]} \\ &\leq \frac{\Pr[M = m]}{\Pr[C = c]} \\ &= 0.\end{aligned}$$

   Since probabilities can't be negative, we then must have that $\Pr[M = m \mid C = c] = \Pr[M = m]$, as desired.

   To show that the scheme is also perfectly secret for any message space $\mathcal{M}' \subseteq \mathcal{M}$, note that any probability distribution over $\mathcal{M}'$ can be extended to one over $\mathcal{M}$, where the probability of any message in $\mathcal{M} \backslash \mathcal{M}'$ is zero. Then, it is trivial to show that since Definition 2.1 holds for the distribution over $\mathcal{M}$, it holds for the distribution over $\mathcal{M}'$ as well.

2. **Solution 1.** There are a couple of ways to interpret how Gen works in the context of the indistinguishability experiment, so it is important that we formulate a precise definition. In this solution, we assume that Gen generates a key by first choosing a length according to some distribution on $\mathbb{N}$ (which must contain finitely many lengths, as per the discussion at the beginning of Chapter 2), and then choosing a string of the chosen length uniformly at random.

   Let ℓ be the maximum length of a string that Gen can generate. Consider the adversary $\mathcal{A}$ that outputs the following two strings of length $2\ell + 1$:

$$\begin{aligned}m_0 &= \underbrace{\mathtt{aaa}\ldots\mathtt{aa}}_{2\ell+1} \\ m_1 &= \underbrace{\mathtt{aaa}\ldots\mathtt{ab}}_{2\ell+1}\end{aligned}$$

   and, upon receiving the ciphertext c, outputs 0 if there is a repeating substring of length $\leq \ell$ and 1 otherwise. If m_0 was the message encoded, then $\mathcal{A}$ will output 0 because there will obviously be a repeating substring. If m_1 was the message encoded, then there will be a substring that repeats at least twice, but the very last character will break the pattern so $\mathcal{A}$ outputs 1. Thus, $\mathcal{A}$ wins the indistinguishability experiment with probability 1, so the Vigenère cipher is not perfectly secret, as desired.

Solution 2. In this solution, we assume that Gen receives some input ℓ, and outputs a string chosen from the uniform distribution on strings of length ℓ. In the context of the indistinguishability experiment, we assume that ℓ is fixed and that the adversary does not know what ℓ is.

For any given adversary we define, it is possible that Gen outputs a key that is longer than the messages it chooses. However, for the scheme to be perfectly secret, it is necessary that $\Pr[\mathsf{PrivK}^{\mathsf{eav}}_{\mathcal{A},\Pi} = 1] = 1/2$ for *every* adversary $\mathcal{A}$, so we only need to show that some adversary exists that can win the game with probability better than $1/2$.

Consider the family of adversaries $\{\mathcal{A}_n\}$, where for each $n \in \mathbb{N}$ the adversary $\mathcal{A}_n$ chooses messages

$$m_0 = \underbrace{\mathtt{aaa\ldots aa}}_{2n+1}$$
$$m_1 = \underbrace{\mathtt{aaa\ldots ab}}_{2n+1}$$

and, upon receiving the ciphertext c, outputs 0 if there is a repeating substring of length n and 1 otherwise. This is very similar to the adversary in Solution 1, and similar logic shows that the adversary $\mathcal{A}_\ell$ wins the indistinguishability experiment with probability 1. Since *there exists* an adversary that can win the game, even though we don't know how to find it beforehand, this shows that the Vigenère cipher is not perfectly secret, as desired.

3. Let p be a distribution over $\mathcal{M}$, and let $m, m' \in \mathcal{M}$ be such that $m \neq m'$ and $p(m) > 0$, $p(m') > 0$. Let $c \in \mathcal{C}$ be such that $\Pr[C = c] > 0$. Then, if we let $c' = c$, we have that
$$\Pr[M = m \wedge M' = m' \mid C = c \wedge C' = c'] = 0$$
since it is impossible for two different plaintexts m and m' to encrypt to the same ciphertext c. Otherwise, for some k we would either have $\mathsf{Dec}_k(\mathsf{Enc}_k(m)) \neq m$ or $\mathsf{Dec}_k(\mathsf{Enc}_k(m')) \neq m'$.

   However, $\Pr[M = m \wedge M' = m'] > 0$ since both m and m' have nonzero probabilities and are sampled independently. Thus,
$$\Pr[M = m \wedge M' = m' \mid C = c \wedge C' = c'] \neq \Pr[M = m \wedge M' = m'],$$
   so no encryption scheme satisfies the definition, as desired.

4. (a) Consider a scheme where the key consists of a completely arbitrary bijection of messages to ciphertexts, chosen uniformly at random from the space of all such bijections. The scheme encrypts messages by finding the corresponding ciphertext in the bijection, and decrypts by reversing the bijection. Intuitively, since the bijection is completely arbitrary, any pair of (nonequal) ciphertexts the adversary sees could have come from any pair of (nonequal) messages, so it should be perfectly secret.

      Formally, for a fixed integer $\ell > 1$, the scheme works as follows:

- The message space $\mathcal{M}$ and ciphertext space $\mathcal{C}$ are each equal to $\{0,1\}^\ell$, the set of all ℓ-length strings over $\{0,1\}$, and the key space $\mathcal{K}$ is the set of all bijections from $\{0,1\}^\ell$ to $\{0,1\}^\ell$.
- Gen works by randomly selecting a key φ from the uniform distribution on $\mathcal{K}$.
- Given a key φ and message $m \in \mathcal{M}$,

$$\mathsf{Enc}_\varphi(m) = \varphi(m).$$

- Given a key φ and ciphertext $c \in \mathcal{C}$,

$$\mathsf{Dec}_\varphi(m) = \varphi^{-1}(m).$$

  Since φ is a bijection, it is trivially invertible, so Dec_φ is well-defined.

Then, given any pair of messages $(m, m') \in \mathcal{M}^2$ and ciphertexts $(c, c') \in \mathcal{C}^2$ such that $m \neq m'$ and $c \neq c'$, we have that

$$\begin{aligned}
&\Pr[M = m \wedge M' = m' \mid C = c \wedge C' = c'] \\
=&\Pr[M = m \mid C = c \wedge C' = c'] \cdot \Pr[M' = m' \mid C = c \wedge C' = c' \wedge M' \neq m] \\
=&\Pr[M = m \mid C = c] \cdot \Pr[M' = m' \mid C' = c' \wedge M' \neq m] \\
=&\frac{\Pr[M = m \wedge C = c]}{\Pr[C = c]} \cdot \frac{\Pr[M' = m' \wedge C' = c' \wedge M' \neq m]}{\Pr[C' = c' \wedge M' \neq m]},
\end{aligned}$$

since M and M' are sampled independently, and each has no dependence on the encryption of the other. Now, let $K(m)$ be the set of all keys for which $\mathsf{Enc}_k(m) = c$ and let $K(m')$ be defined similarly. Note that choosing a random bijection should make m equally likely to map to each ciphertext, so

$$\Pr[K \in K(m)] = \Pr[K' \in K(m')] = \frac{1}{2^\ell}.$$

We can now calculate each of the four probabilities in the above expression.

- $\Pr[M = m \wedge C = c] = \Pr[M = m]/2^\ell$. When the message is m, the ciphertext will be c precisely when the key is in $K(m)$, so

$$\Pr[M = m \wedge C = c] = \Pr[M = m \wedge K \in K(m)].$$

  Since the key and the message are selected independently, we then have that

$$\begin{aligned}
\Pr[M = m \wedge C = c] &= \Pr[M = m] \cdot \Pr[K \in K(m)] \\
&= \frac{\Pr[M = m]}{2^\ell},
\end{aligned}$$

  as desired.
- $\Pr[C = c] = 1/2^\ell$. This is true because given any particular message, a randomly chosen bijection can map the message to any ciphertext with equal probability. Since there are 2^ℓ ciphertexts
- $\Pr[M' = m' \wedge C' = c' \wedge M' \neq m] = \Pr[M' = m']/2^\ell$. The event $M' \neq m$ is redundant, so this can be calculated in the same way as $\Pr[M = m \wedge C = c]$.

- $\Pr[C' = c' \wedge M' \neq m] = \Pr[M \neq M']/2^\ell$. This can also be calculated in much the same way as $\Pr[M = m \wedge C = c]$, with the added observation that $\Pr[M' \neq m] = \Pr[M \neq M']$.

Putting all of these together, we get that

$$\begin{aligned}
&\Pr[M = m \wedge M' = m' \mid C = c \wedge C' = c'] \\
=&\frac{\Pr[M = m]/2^\ell}{1/2^\ell} \cdot \frac{\Pr[M' = m']/2^\ell}{\Pr[M \neq M']/2^\ell} \\
=&\frac{\Pr[M = m] \cdot \Pr[M' = m']}{\Pr[M \neq M']} \\
=&\Pr[M = m \wedge M' = m' \mid M \neq M'],
\end{aligned}$$

so the scheme is perfectly secret, as desired.

(b) An encryption scheme $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ over a message space $\mathcal{M}$ is *perfectly-secret for k messages* if for all distributions over $\mathcal{M}$, all $m_1, m_2, \ldots, m_k \in \mathcal{M}$ with $m_i \neq m_j$ for $1 \leq i < j \leq k$, and all $c_1, c_2, \ldots, c_k \in \mathcal{C}$ with $c_i \neq c_j$ for $1 \leq i < j \leq k$ and $\Pr\left[\bigwedge_{i=1}^k C_i = c_i\right] > 0$:

$$\Pr\left[\bigwedge_{i=1}^k M_i = m_i \mid \bigwedge_{i=1}^k C_i = c_i\right] = \Pr\left[\bigwedge_{i=1}^k M_i = m_i \mid \bigwedge_{i=1}^k \bigwedge_{i \neq j = 1}^k M_i \neq M_j\right],$$

where $m_1, m_2, \ldots, m_k$ are sampled independently from the same distribution over $\mathcal{M}$.

For any k, the scheme given above remains perfectly secret for k messages.