NOTES ON CRYPTOGRAPHY

Based on Katz-Lindell. All random variables considered discrete?

Encryption Scheme: An encryption scheme consists of a tuple $(\mathrm{Gen}, \mathrm{Enc}, \mathrm{Dec})$ where

- Gen is a random variable with values in a set $\mathcal{K}$, called the *keyspace.*
- There is a set $\mathcal{M}$, called the *message space.*
- There is a set $\mathcal{C}$, each element is called a *ciphertext.*
- For each $k \in \mathcal{K}$ and $m \in \mathcal{M}$ we have a random variable $\mathrm{Enc}_k(m)$ taking values in $\mathcal{C}$.
- For each $k \in \mathcal{K}$, we have a map $\mathrm{Dec}_k : \mathcal{C} \to \mathcal{M}$ satisfying
$$\mathrm{Dec}_k(\mathrm{Enc}_k(m)) = m \quad \text{for all } m \in \mathcal{M}$$
.

More formally, $\mathrm{Enc}_k(m) : \Omega \to \mathcal{C}$ on some sample space Ω. The last point means that $\mathrm{Dec}_k(\mathrm{Enc}_k(m)(\omega)) = m$ for all $\omega \in \Omega$. If $\mathrm{Enc}_k(m)$ is a constant map, we simply consider it to be a deterministic function $\mathrm{Enc}_k : \mathcal{M} \to \mathcal{C}$.

Example 0.1 (Caeser cipher)**.** In this example, we identify naturally the lowercase letters $\{a, b, \ldots, z\}$ with $\mathbb{Z}/26\mathbb{Z}$. We let $\mathcal{K} = \mathbb{Z}/26\mathbb{Z}$ and we let $\mathcal{M} = \mathcal{C}$ be the set of all words in the lowercase letters. We now define $\mathrm{Enc}_k(m) \in \mathcal{C}$ to be the word obtained by adding $k \pmod{26}$ to each letter of m. So $\mathrm{Enc}_3(zac) = cdf$. So $\mathrm{Dec}_k = \mathrm{Enc}_{-k}$ is the inverse. The distribution on $\mathcal{K}$ is uniform (i.e., Gen takes values uniformly in $\mathcal{K}$).

Definition 0.2 (Perfect secrecy)**.** An encryption scheme is said to be *perfectly secret* if for all $m, m' \in \mathcal{M}$ and $c \in \mathcal{C}$ we have that
$$Pr(\mathrm{Enc}_k(m) = c) = Pr(\mathrm{Enc}_k(m') = c).$$
More precisely, if $P_{\mathcal{K}}$ is the distribution on K and $P_{\mathcal{C},m,k}$ is the distribution on $\mathcal{C}$ induced by $\mathrm{Enc}_k(m)$ then
$$\sum_{k \in K} P_{\mathcal{K}}(k) P_{\mathcal{C},k,m}(c) = \sum_{k \in K} P_{\mathcal{K}}(k) P_{\mathcal{C},k,m'}(c).$$
That is, Pr refers to the probability distribution where one chooses $k \in \mathcal{K}$ randomly (according to the random variable Gen part of the encryption scheme) and then one runs the random variable $\mathrm{Enc}_k(m)$ to produce $c \in \mathcal{C}$. If Enc_k is deterministic ($\mathrm{Enc}_k(m)$ is constant for all k, m) then obviously Pr is just $P_{\mathcal{K}}$.

Example 0.3 (Caeser cipher is not perfectly secret)**.** Let $m = aa$ and $m' = ab$ and let $c = ff$. Then $Pr(\mathrm{Enc}_k(m) = c) = Pr(k = 0) = \frac{1}{26}$, i.e., the probability that aa gets encoded into ff happens only if $k = 5$, so with probability $\frac{1}{26}$. However $Pr(\mathrm{Enc}_k(m') = c) = 0$ because ab can only be encoded into one of $ab, bc, ce, \ldots, za$.

Example 0.4 (one time pad)**.** Let G be a finite group. We define an encryption scheme where the keyspace and namespace is G. The encryption is $Enc_k(m) = km$. Decryption is given by $Dec_k(m) = k^{-1}m$. The distribution on keyspace is uniform. This scheme is perfectly secret as
$$Pr(Enc_k(m) = c) = Pr(km = c) = Pr(k = cm^{-1}) = \frac{1}{|G|},$$

and this expression is clearly independent of m for each fixed c. In the literature, the one-time pad is actually defined only for the case $G = (\mathbb{Z}/2\mathbb{Z})^\ell$. It is called the one-time pad because one needs to generate a new key for each message, i.e., if we send two differnet messages m and m' then the eavesdroper can compute $Enc_k(m) + Enc_k(m') = k + m + k + m' = m + m'$. Thus the eavesdroper can compute the XOR of two different secret messages, which can be bad. Another drawback is the lack of efficiency in that the keyspace is as large as the message space.

Definition 0.5. Consider an encryption scheme as above. Let $\mathcal{P}_\mathcal{M}$ be any distribution on the message space $\mathcal{M}$. We define the induced distribution Pr on $\mathcal{K} \times \mathcal{M} \times \mathcal{C}$ to be the distribution where

$$Pr(K = k, M = m, C = c) = \mathcal{P}_\mathcal{K}(k)\mathcal{P}_\mathcal{M}(m)P_{\mathcal{C},k,m}(c).$$

In other words we choose $k \in \mathcal{K}$ randomly and then $m \in \mathcal{M}$ independently and then $c \in \mathcal{C}$ is drawn according to the random variable $\mathrm{Enc}_k(m)$.

Proposition 0.6. An encryption scheme is perfectly secret if and only if for any distribution $\mathcal{P}_\mathcal{M}$ the induced distribution Pr satisfies the property that

$$Pr(M = m \mid C = c) = Pr(M = m) \quad \text{for all } m \in M \text{ and } c \in C \text{ with Pr(C=c) ¿0}$$

Example 0.7. Suppose we have a distribution where $\mathcal{P}_M(ab) = 0.4$ and $\mathcal{P}_M(aa) = 0.1$ and $\mathcal{P}_M(bb) = 0.5$. If we are using the Caeser cipher, then $Pr(M = ab` \mid C = dd) = 0 \neq 0.4$ but clearly $P(C = dd) > 0$. Thus the Caeser cipher is not perfectly secret.

Definition 0.8 (Adversary)**.** An adversary to a given encryption scheme $\Pi = (\mathrm{Gen}, \mathrm{Enc}, \mathrm{Dec})$ consists of, for each $m_0, m_1 \in \mathcal{M}$ and $c \in \mathcal{C}$, a random variable $\mathcal{A}(m_0, m_1, c)$ taking values in $\{0, 1\}$. For each fixed m_0, m_1, we define a random variable $\mathrm{PrivK}_{\mathcal{A},\Pi}$ as follows:

- Choose $k \in \mathcal{K}$ randomly according to Gen.
- Choose $b \in \{0, 1\}$ uniformly randomly, then choose $c_b \in \mathcal{C}$ randomly according to $\mathrm{Enc}_k(m_b)$.
- Now choose $b' \in \{0, 1\}$ randomly according to $\mathcal{A}(m_0, m_1, c_b)$.
- Now define $\mathrm{PrivK}_{\mathcal{A},\Pi} = (b == b') \in \{0, 1\}$.

Proposition 0.9. An encryption scheme $\Pi = (\mathrm{Gen}, \mathrm{Enc}, \mathrm{Dec})$ is perfectly secret if and only if for all adversaries $\mathcal{A}$, we have that $Pr(\mathrm{PrivK}_{\mathcal{A},\Pi} = 1) = \frac{1}{2}$.

Intuitively, the adversary constructs any m_0, m_1 that they wish, they then get an assistant to randomly choose $b \in \{0, 1\}$ and pass m_b into the encryption scheme and get a corresponding c_b. The adversary then has to guess, based on this triple m_0, m_1, c_b a $b' \in \{0, 1\}$ and try to get $b' = b$ (they do not know b, only the assistant does). If they can succeed (i.e., $(b == b')$) with probability greater than $\frac{1}{2}$, then the Proposition says that the scheme is not perfect.

Proof of Proposition. Note that

$$Pr(\mathrm{PrivK}_{\mathcal{A},\Pi} = 1) = Pr(\mathcal{A} = 0 \text{ and } b = 0) + Pr(\mathcal{A} = 1 \text{ and } b = 1)$$

We compute the first sum as

$$Pr(\mathcal{A} = 0 \text{ and } b = 0) = \frac{1}{2} \sum_{k \in \mathcal{K}} P_\mathcal{K}(k) \sum_{c \in \mathcal{C}} P_{\mathcal{C},k,m_0}(c) Pr(\mathcal{A}(m_0, m_1, c) = 0)$$

likewise, the second term is

$$Pr(\mathcal{A}=1 \text{ and } b=1)=\frac{1}{2}\sum_{k\in\mathcal{K}}P_{\mathcal{K}}(k)\sum_{c\in\mathcal{C}}P_{\mathcal{C},k,m_1}(c)Pr(\mathcal{A}(m_0,m_1,c)=1)$$

now using the identity $Pr(\mathcal{A}(m_0,m_1,c)=1)=1-Pr(\mathcal{A}(m_0,m_1,c)=0)$ we can add these two terms to get

$$Pr(\text{PrivK}_{\mathcal{A},\Pi}=1)=\frac{1}{2}\sum_{k\in\mathcal{K}}P_{\mathcal{K}}(k)\sum_{c\in\mathcal{C}}P_{\mathcal{C},k,m_1}(c)+\frac{1}{2}\sum_{k\in\mathcal{K}}P_{\mathcal{K}}(k)\sum_{c\in\mathcal{C}}Pr(\mathcal{A}(m_0,m_1,c)=0)(P_{\mathcal{C},k,m_0}(c)-P_{\mathcal{C},k,m_1}(c)).$$

The first sum is

$$\frac{1}{2}\sum_{k\in\mathcal{K}}P_{\mathcal{K}}(k)\sum_{c\in\mathcal{C}}P_{\mathcal{C},k,m_1}(c)=\frac{1}{2}$$

as we are summing over the sample space of a probability measure. Now we have to show that the second sum vanishes for all $\mathcal{A}$ if and only if the scheme is perfectly secret. We rewrite this sum (ignoring the $\frac{1}{2}$ factor) as:

$$\sum_{c\in\mathcal{C}}Pr(\mathcal{A}(m_0,m_1,c)=0)\left(\sum_{k\in\mathcal{K}}P_{\mathcal{K}}(k)P_{\mathcal{C},k,m_0}(c)-\sum_{k\in\mathcal{K}}P_{\mathcal{K}}(k)P_{\mathcal{C},k,m_1}(c)\right)$$

By definition of perfect secrecy, the term inside is 0, i.e., because it is

$$Pr(C=c|M=m_0)-Pr(C=c|M=m_1).$$

Conversely, suppose that this whole expression is 0 for all $\mathcal{A}$. In particular, for each fixed $c_0\in C$, if we choose $\mathcal{A}$ so that $Pr(\mathcal{A}(m_0,m_1,c)=\mathbb{1}_{\{c_0\}}(c)$ then we see that this expression is

$$P(C=c_0|M=m_0)-P(C=c_0|M=m_1)$$

and is equal to 0. □

Proposition 0.10. In a perfectly secret scheme, we have $|\mathcal{K}|\geq|\mathcal{M}|$.

Proof. Suppose for contradiction that $|\mathcal{K}|<|\mathcal{M}|$. Choose $c\in C$ such that $Pr(Enc_k(m)=c)>0$ for some $k,m\in\mathcal{K}\times\mathcal{M}$. Now let $\mathcal{M}(c)=\{Dec_k(c)\mid k\in\mathcal{K}\}$. Then clearly $|\mathcal{M}(c)|\leq|\mathcal{K}|<|\mathcal{M}|$ so we may choose $m'\notin\mathcal{M}(c)$. By perfect secrecy, we have that $Pr(Enc_k(m')=c)=Pr(Enc_k(m)=c)>0$. This means that $Dec_k(c)=m'$ for some $k\in K$. This means $m'\in\mathcal{M}(c)$, a contradiction. □