Private Key Encryption - Formal Definitions

Q1: Formally define a symmetric encryption scheme and its correctness property. Write down the full mathematical details.

A1: A symmetric encryption scheme Π is defined as a 3-tuple: $\Pi = (Gen, Enc, Dec)$ and a message space $\mathcal{M}$. There is also a key space $\mathcal{K}$ and ciphertext space $\mathcal{C}$, but these are defined by the $Enc : \mathcal{M} \times \mathcal{K} \to \mathcal{C}$ and $Gen : \{0,1\}^\lambda \to \mathcal{K}$ themselves.

A symmetric ecryption scheme is defined to have the correctness property if $\forall m \in \mathcal{M}$ and $\forall k \in \mathcal{K}$:

$$Dec(k, Enc(k, m)) = m$$

Note here that, the sender is running Enc with key k, and the receiver runs Dec with the same key k. This is why the scheme is known as "symmetric", as the key is present at both sides prior to the actual communication.

Q2: Can the decryption algorithm be randomized? What may happen if it is randomized?

A2: A decryption algorithm can not be randomized. Recall that by definition, a randomized algorithm outputs random values on the same input every time. However, this would break correctness: $Dec(k, Enc(k, m)) = m$, since this would not always output m, and even if it did then we would not call it a randomized algorithm.

Q3: What is deterministic encryption? Give examples.

A3: Deterministic encryption is when the function $Enc(k, m)$ always outputs the same ciphertext c for the same (k, m) pairs. Several examples are:

- *One-Time Pad* uses the function $Enc(k, m) = k \oplus m$ where $\oplus$ is the bitwise XOR operation.
- *Shift Ciphers* use a function in the form of $Enc(k, m) = (k + m) \bmod n$ where n is the alphabet size, and this function applies to the characters of plaintext. For example a *Caesar Cipher* uses $k = 3$ and $n = 26$.

Q4: Why are deterministic encryption schemes insecure for multiple messages?

A4: This question is actually a bit fallacious, I think what is meant is when a key is re-used for multiple messages. Because otherwise, One-Time Pad supports multiple messages (in the assumption of one key per message) and it is a deterministic encryption scheme, and OTP is a perfectly secure scheme. So my answer will be with respect to key re-use on multiple messages.

For a scheme $\Pi = (Gen, Enc, Dec)$, fix a key k and messages m, m'. You have $c = Enc(k, m)$ and $c' = Enc(k, m')$. Suppose that there is a perfectly secret encryption scheme $\Pi' = (Gen, Enc, Dec)$ with message space $\mathcal{M}' = \mathcal{M} \setminus \{m, m'\}$ and key space $\mathcal{K}' = \mathcal{K} \setminus \{k\}$. By definition, $|\mathcal{K}'| \geq |\mathcal{M}'|$. By Shannon's Theorem, if there is one cyphertext per message, then $|\mathcal{M}'| = |\mathcal{C}'| = |\mathcal{K}'|$. Since $|\mathcal{K}| = |\mathcal{K}'| + 1$ and $|\mathcal{M}| = |\mathcal{M}'| + 2$, we clearly see that Π breaks perfect secrecy requirement (as stated in Theorem 2.10 in KL Book).

Q5: Formally define semantic security of a symmetric encryption scheme. Write down the full mathematical details.

A5: Semantic security is an interpretation of security, that is equivalent to indistinguishability. Following definition 3.12 from the KL Book, a symmetric encryption scheme (Enc, Dec) is semantically secure against an eavesdropper (an adversary that is doing a ciphertext-only attack) if for every PPT algorithm $\mathcal{A}$, there exists a PPT algorithm $\mathcal{A}'$ such that for any PPT algorithm $\texttt{Samp}$ and polynomial time computable functions f and h, the following is negligible:

$$|Pr[\mathcal{A}(1^n, Enc(k, m), h(m)) = f(m))] - Pr[\mathcal{A}'(1^n, |m|, h(m)) = f(m)]|$$

where the first probability is taken over uniform $k \in \{0,1\}^n$, m output by $\texttt{Samp}(1^n)$, the randomness of $\mathcal{A}$ and Enc, and the second probability is taken over m output by $\texttt{Samp}(1^n)$ and the randomness of $\mathcal{A}'$.

Here, what is described in layman terms is that in the presence of two adversaries $\mathcal{A}$ and $\mathcal{A}'$, the probability that one "knows" more than other, given that one knows the ciphertext and the other knows the message length, is negligible. It makes perfect sense!

Q6: Formally define semantic security using the indistinguishability definition. Write down the full mathematical details.

A6: Theorem 3.13 from KL Book already states this. Assuming that an encryption scheme has indistinguishable encryptions in the presence of an eavesdropper, then for any PPT algorithm $\mathcal{A}$ there is a PPT algorithm $\mathcal{A}$ such that for any $S \subseteq \{0,1\}^l$ and any function

$f : \{0,1\}^l \to \{0,1\}$, there is a negligible function $\texttt{negl}$ such that:

$$|Pr[\mathcal{A}(1^n, Enc(k,m)) = f(m)] - Pr[\mathcal{A}'(1^n) = f(m)]| \leq negl(n)$$

where the first probability is taken over uniform choice of $k \in \{0,1\}^n$ and $m \in S$, the randomness of $\mathcal{A}$ and Enc, and the second probability is taken over uniform choice of $m \in S$ and the randomness of $\mathcal{A}'$. From the semantic security definition (as defined in definition 3.12 from KL Book) if you remove $h(m)$ which is the apriori knowledge, and also remove $|m|$ from the parameter because the scheme is a fixed-length symmetric encryption scheme, then you get the formula above. Also note that in that definition we required the expression to be negligible, but here we say less than equal to negligible. This makes no difference whatsoever, because anything less than equal to a negligible is also negligible.

Q7: Prove that the symmetric encryption scheme Π would not be indistinguishable in the presence of an eavesdropper if the adversary (I think we meant "challenger" here) can encrypt arbitrary length messages and the adversary is not restricted to output equal length messages in experiment $PrivK^{eav}_{(A,\Pi)}$.

A7: Let us briefly define this specific indistinguishability experiment $\texttt{PrivK}^{\texttt{eav}}_{\mathcal{A},\Pi}(n)$:

(1) Adversary $\mathcal{A}$ is given input 1^n, and outputs a pair of messages m_0, m_1 with **arbitrary lengths**.
(2) A key k is generated and a uniform bit b is chosen. $c \leftarrow Enc(k, m_b)$ is computed and given to $\mathcal{A}$.
(3) $\mathcal{A}$ outputs a bit b'.
(4) If $b = b'$ then the experiment results with 1, 0 otherwise. If the result is 1, $\mathcal{A}$ succeds in distinguishing these messages.

The difference of this experiment to what we have seen in the book is that now the messages can be of different arbitrary lengths. Since this question asks us to delve in to the indistinguishability, I am assuming that correctness is not broken. Also considering that the adversary would easily distinguish the message from the length of ciphertext if they were equal (which adversary would infer from the implementation of Enc, which it has access to), so let us assume that $|c| = max(|m_0|, |m_1|)$. Since the lengths are different, adversary knows that shorter message could normally have another ciphertext of the same length, if it were to be the longest of the two messages, however here it had to be padded. The padding is done by Enc, which the adversary would know exactly how it happens. An example would be

to choose uniform $m_0 = 0$ and $m_1 \in \{0,1\}^{p(n)}$ for some polynomial $p(n)$. What we have is that:

$$Pr[\mathtt{PrivK}^{\mathtt{eav}}_{\mathcal{A},\Pi}(n) = 1] = 1 - Pr[Enc_{padded}(k, m_0) = Enc(k, m_1)]$$

which is non-negligible, thereby showing that $\mathcal{A}$ wins the game with non-negligible probability.