Proof that $|\mathcal{K}| \geq |\mathcal{M}|$ is necessary for perfect privacy

Let us first remind ourselves of the definition of perfect privacy.

Definition 1. An encryption scheme $\Pi = (\mathcal{E}, \mathcal{D})$ has *perfect privacy*, if for any two messages $M_0, M_1 \in \mathcal{M}$, and any ciphertext $C \in \mathcal{C}$, we have:

$$\Pr[\mathcal{E}_K(M_0) = C] = \Pr[\mathcal{E}_K(M_1) = C],$$

where the probability is over the key $K \in \mathcal{K}$, chosen uniformly at random, and any randomness used internally by $\mathcal{E}$.

Theorem 1. *No symmetric encryption scheme has perfect privacy if $|\mathcal{K}| < |\mathcal{M}|$.*

Proof. Let $\Pi = (\mathcal{E}, \mathcal{D})$ be a symmetric encryption scheme, and assume $|\mathcal{K}| < |\mathcal{M}|$. We want to show that Π cannot have perfect privacy.

By the definition of perfect privacy, it is sufficient to find two messages $M_0, M_1 \in \mathcal{M}$, and a ciphertext $C \in \mathcal{C}$, such that $\Pr[\mathcal{E}_K(M_0) = C] \neq \Pr[\mathcal{E}_K(M_1) = C]$ in order to prove the theorem.

To this end, let $C \in \mathcal{C}$ be an arbitrary ciphertext, and let $\mathcal{M}(C)$ denote the set of all messages which are possible decryptions of C; that is:

$$\mathcal{M}(C) = \{M \mid M = \mathcal{D}_K(C) \text{ for some } K \in \mathcal{K}\}.$$

In other words: $M \in \mathcal{M}(C)$ if there exists some key $K \in \mathcal{K}$, such that $M = \mathcal{D}_K(C)$.

First, let M_0 be an arbitrary message in the set $\mathcal{M}(C)$. Hopefully, it should be clear that $|\mathcal{M}(C)| \leq |\mathcal{K}|$, because for each distinct message $M \in \mathcal{M}(C)$ there corresponds one or more distinct[1] keys in $\mathcal{K}$. However, by the assumption of the theorem, we also have

$$|\mathcal{M}(C)| \leq |\mathcal{K}| < |\mathcal{M}|.$$

This means that there must *exist* some $M_1 \in \mathcal{M}$ which is *not* in $\mathcal{M}(C)$. In other words, $M_1 \neq M_0 \in \mathcal{M}(C)$.

Now let us calculate the probabilities $\Pr[\mathcal{E}_K(M_0) = C]$ and $\Pr[\mathcal{E}_K(M_1) = C]$. For the first one we have $\Pr[\mathcal{E}_K(M_0) = C] = p$ for some probability $p > 0$ (we don't actually care about what this probability is, as long as it is non-zero). However, for the second one we have $\Pr[\mathcal{E}_K(M_1) = C] = 0$, since we explicitly chose M_1 not to be in the set $\mathcal{M}(C)$! Hence, we have

$$\Pr[\mathcal{E}_K(M_0) = C] \neq \Pr[\mathcal{E}_K(M_1) = C],$$

which proves the theorem. □

[1] If two different messages M' and M'' in $\mathcal{M}(C)$ both corresponded to the same key K, then decryption would be ambiguous: should $\mathcal{D}_K(C)$ decrypt to M' or M''?