Proofs of Security Theorems from Lecture IV

Semantic security of randomized CTR mode

Theorem 1. *Let $\mathcal{E}$ be a block cipher over $(\mathcal{K}, \mathcal{X})$ that is ε-secure for some negligible ε. Then for any fixed IV, the cipher $\mathcal{E}_{CTR}$ is δ-semantically secure for*

$$\delta = 2\varepsilon + \frac{n^2}{2|\mathcal{X}|},$$

where n is the maximal number of blocks per message (i.e., $\mathcal{M} = \mathcal{X}^{\leq n}$).

Setup Let $\delta > 0$ be arbitrary, We prove the following: for any semantic security adversary $\mathcal{B}$ which achieves advantage δ against $\mathcal{E}_{CTR}$, there exists a PRP-adversary $\mathcal{A}$ which achieves advantage $\frac{\delta}{2} - \frac{n^2}{4|\mathcal{X}|}$ against the block cipher $\mathcal{E}$. Hence, if all PRP-adversaries achieve advantage at most ε against $\mathcal{E}$ (i.e., if $\mathcal{E}$ is an ε-secure block cipher), we know that there cannot be an adversary against $\mathcal{E}_{CTR}$ achieving advantage larger than $\delta = 2\varepsilon + \frac{n^2}{2|\mathcal{X}|}$.

Constructing $\mathcal{A}$ using $\mathcal{B}$ Let $\mathcal{B}$ be a semantic security adversary and let δ be its advantage against $\mathcal{E}_{CTR}$. We construct a PRP-adversary $\mathcal{A}$ as follows:

1. First, $\mathcal{A}$ samples a bit b from $\{0,1\}$ and a key k from $\mathcal{K}$, both uniformly at random. This is to simulate a semantic security attack game for $\mathcal{B}$.
2. Then $\mathcal{A}$ queries $\mathcal{B}$ for its two messages m_0 and m_1 (we know that they must be of the same length). Let

   $$m_0 = z_{0,1} \,||\, z_{0,2} \,||\, \cdots \,||\, z_{0,n}$$
   $$m_1 = z_{1,1} \,||\, z_{1,2} \,||\, \cdots \,||\, z_{1,n},$$

   where the $z_{i,j}$ are the individual message blocks.
3. $\mathcal{A}$ then outputs n as the number of rounds it is going to play the PRP attack game. For each $1 \leq i \leq n$ it sets x_i to $[i]_{\mathcal{X}}$, i.e. to a block containing the binary representation of number i.
4. $\mathcal{A}$ proceeds to interact with its own challenger: in every round $1 \leq i \leq n$ it sends out the block x_i computed in the previous step and receives the corresponding block y_i.

5. After the final round, $\mathcal{A}$ computes the ciphertext $c = z_{b,1} \oplus y_1 \,||\, z_{b,2} \oplus y_2 \,||\, \cdots \,||\, z_{b,n} \oplus y_n$ and gives c to $\mathcal{B}$.

6. Finally, $\mathcal{A}$ observes $\mathcal{B}$'s output. If the output matches b (i.e., if $\mathcal{B}$ wins the game which $\mathcal{A}$ simulated for him), $\mathcal{A}$ outputs *fake*, otherwise it outputs *rand*.

$\mathcal{A}$'s advantage We now want to compute $\mathcal{A}$'s advantage in its PRP attack game. Let us briefly recall how this game proceeds. The challenger randomly samples $m \in \{rand, fake\}$.

- If $m = rand$, the challenger chooses a permutation $f \in \Pi(\mathcal{X})$ uniformly at random.
- If $m = fake$, the challenger chooses a key $k \in \mathcal{K}$ uniformly at random and sets f to $E(k, \cdot)$, where E is the encryption function of the block cipher $\mathcal{E}$.

In both cases, the challenger then interacts with $\mathcal{A}$ in the n rounds as follows: in every round i, upon receiving the block x_i, it computes $y_i = f(x_i)$ and sends y_i back to $\mathcal{A}$.

For brevity, let us denote by B_w the event that $\mathcal{B}$ wins its simulated game (i.e., correctly guesses $\mathcal{A}$'s bit b) and by B_l the event that $\mathcal{B}$ loses. Similarly, let us denote by M_f the event that $m = fake$ and by M_r the event that $m = rand$. The probability that $\mathcal{A}$ wins this attack game can be written, using the law of total probability[1], as

$$\begin{aligned} \mathbb{P}(\mathcal{A}\ wins) &= \mathbb{P}(M_f) \cdot \mathbb{P}(B_w \mid M_f) + \mathbb{P}(M_r) \cdot \mathbb{P}(B_l \mid M_r) \\ &= \frac{1}{2} \cdot \mathbb{P}(B_w \mid M_f) + \frac{1}{2} \cdot \mathbb{P}(B_l \mid M_r). \end{aligned} \tag{1}$$

Let us evaluate (or at least bound) each term in (1) individually.

Evaluating $\mathbb{P}(B_w \mid M_f)$ Observe that if $m = fake$, then the sequence of blocks $y_1 \,||\, y_2 \,||\, \cdots \,||\, y_n$ produced by the interaction of $\mathcal{A}$ and its challenger is really a keystream that would be produced by the counter mode of $\mathcal{E}$ when using key k and IV equal to $[1]_{\mathcal{X}}$. Hence, in this case the message c is indeed an encryption of m_b produced by $\mathcal{E}_{CTR}$ using these parameters. Hence, from $\mathcal{B}$'s point of view, the whole probabilistic experiment is really a semantic security attack game against $\mathcal{E}_{CTR}$, in which he has, by our assumption, advantage δ. Hence, the probability that he will win in this case equals

$$\mathbb{P}(B_w \mid M_f) = \frac{1}{2} + \delta. \tag{2}$$

[1] `https://en.wikipedia.org/wiki/Law_of_total_probability`

Bounding $\mathbb{P}(B_l \mid M_r)$: Cheating game This part is trickier. First, let's have the following thought experiment: imagine that in the whole process, the PRP challenger is *cheating:* if $m = \mathit{rand}$ he will skip any random choice of $f \in \Pi(\mathcal{X})$. Instead, whenever the challenger a block x_i from $\mathcal{A}$, he will simply sample a block y_i uniformly at random from $\mathcal{X}$ and sends it back to $\mathcal{A}$. Let us call this altered game a *cheating game* and denote its accompanying probability measure by $\mathbb{P}_{ch}$.

Observe that in the cheating game, if $m = \mathit{rand}$, then $y_1 \,||\, y_2 \,||\, \cdots \,||\, y_n$ is a keystream produced according to the perfectly secure one-time pad (OTP) cipher. Hence, in this case, from $\mathcal{B}$'s point of view the whole process is the same as the semantic security attack game against OTP. Since OTP is perfectly secure, any adversary, including $\mathcal{B}$, has zero advantage against it. Hence,

$$\mathbb{P}_{ch}(B_w \mid M_r) = \frac{1}{2} \tag{3}$$

and thus

$$\mathbb{P}_{ch}(B_l \mid M_r) = 1 - \mathbb{P}_{ch}(B_w \mid M_r) = \frac{1}{2}. \tag{4}$$

Cheating vs. non-cheating game However, $\mathbb{P}_{ch}(B_l \mid M_r) \neq \mathbb{P}(B_l \mid M_r)$ since, $\mathbb{P}_{ch} \neq \mathbb{P}$. To see this, note that in the cheating game, if $m = \mathit{rand}$, it is possible for the keystream to contain two occurrences of the same block, i.e. $y_i = y_j$ for some $i \neq j$ might happen with a positive probability. In the original game, this cannot happen, since f is always chosen to be a permutation and the blocks x_i supplied by $\mathcal{A}$ are pairwise distinct (they form an increasing sequence of binary-encoded values).

In essence, the cheating game can be viewed as a variant of the original game in which, in the case $m = \mathit{rand}$, the challenger does not randomly sample a permutation but an arbitrary (possibly non-injective) *function* of type $\mathcal{X} \to \mathcal{X}$. Sampling a random function allows, for each input block, to select the output block uniformly at random from all possible blocks (even those that already appeared before), which is exactly what happens in the cheating game. The question of how this influences the probabilities of winning is addressed in what follows.

Bounding $\mathbb{P}(B_l \mid M_r)$ Let C (short for "collision") be the event that there are $1 \leq i < j \leq n$ such that $y_i = y_j$ and let $\overline{C}$ be the event that no such collision happens. By the law of total probability, we have:

$$\mathbb{P}_{ch}(B_w \mid M_r) = \mathbb{P}_{ch}(C \mid M_r) \cdot \mathbb{P}_{ch}(B_w \mid M_r \wedge C) + \mathbb{P}_{ch}(\overline{C} \mid M_r) \cdot \mathbb{P}_{ch}(B_w \mid M_r \wedge \overline{C}). \tag{5}$$

Noting that all the probabilities are non-negative, we can get rid of the first term on the right-hand side of (5) to get

$$\mathbb{P}_{ch}(B_w \mid M_r) \geq \mathbb{P}_{ch}(\overline{C} \mid M_r) \cdot \mathbb{P}_{ch}(B_w \mid M_r \wedge \overline{C}). \tag{6}$$

Now observe that if a collision *does not* happen in the cheating game, then this game evolves in exactly the same way as the original game. The original game is exactly the cheating game constrained by the condition that a collision never happens in the case when $m = rand$. In particular, $\mathbb{P}_{ch}(A \mid M_r \wedge \overline{C}) = \mathbb{P}(A \mid M_r)$ for any event A.[2] This can be plugged into (6) to get

$$\mathbb{P}_{ch}(B_w \mid M_r) \geq \mathbb{P}_{ch}(\overline{C} \mid M_r) \cdot \mathbb{P}(B_w \mid M_r).$$

By rearranging and using (3) we get

$$\mathbb{P}(B_w \mid M_r) \leq \frac{\mathbb{P}_{ch}(B_w \mid M_r)}{\mathbb{P}_{ch}(\overline{C} \mid M_r)} = \frac{1}{2 \cdot \mathbb{P}_{ch}(\overline{C} \mid M_r)}. \tag{7}$$

Hence,

$$\begin{aligned}
\mathbb{P}(B_l \mid M_r) &= 1 - \mathbb{P}(B_w \mid M_r) \\
&\geq 1 - \frac{1}{2 \cdot \mathbb{P}_{ch}(\overline{C} \mid M_r)} = \frac{2 \cdot \mathbb{P}_{ch}(\overline{C} \mid M_r) - 1}{2 \cdot \underbrace{\mathbb{P}_{ch}(\overline{C} \mid M_r)}_{\leq 1}} \\
&\geq \frac{2 \cdot \mathbb{P}_{ch}(\overline{C} \mid M_r) - 1}{2} = \underbrace{\mathbb{P}_{ch}(\overline{C} \mid M_r)}_{=1-\mathbb{P}_{ch}(C \mid M_r)} - \frac{1}{2} \\
&= \frac{1}{2} - \mathbb{P}_{ch}(C \mid M_r).
\end{aligned}$$

Thus, we proved that

$$\mathbb{P}(B_l \mid M_r) \geq \frac{1}{2} - \mathbb{P}_{ch}(C \mid M_r). \tag{8}$$

To conclude our proof, it remains to bound the probability $\mathbb{P}_{ch}(C \mid M_r)$ of a collision happening in the cheating game.

Probability of collision Recall that a cheating game having a collision (in the case when $m = rand$) means that there are two distinct block indices $1 \leq i < j \leq n$ such that $y_i = y_j$. For a given pair of distinct indices i, j we have

$$\mathbb{P}_{ch}(y_i = y_j \mid M_r) = \frac{1}{|\mathcal{X}|}$$

(the y_i can be fixed arbitrarily but then y_j, which is sampled uniformly from $\mathcal{X}$, must be sampled to the same value as y_i). Using union bound [3] we get

[2] We do not prove this formally, but note that the probability of generating a concrete keystream $y_1 \,||\, \ldots \,||\, y_n$ is the same under $\mathbb{P}_{ch}(\cdot \mid M_r \wedge \overline{C})$ and $\mathbb{P}(\cdot \mid M_r)$ – this can be proved directly using the definition of conditional probability. Since $\mathcal{B}$'s decisions depend only on the ciphertext he receives, and any difference in the ciphertexts he receives in the two games must be caused by the difference of the keystream, the equality follows.

[3] `https://en.wikipedia.org/wiki/Boole%27s_inequality`

$$\mathbb{P}_{ch}(C \mid M_r) \leq \sum_{1 \leq i < j \leq n} \mathbb{P}_{ch}(y_i = y_j \mid M_r) = \sum_{1 \leq i < j \leq n} \frac{1}{|\mathcal{X}|} = \frac{1}{|\mathcal{X}|} \cdot \sum_{i=1}^{n-1} i$$
$$= \frac{(n-1)n}{2|\mathcal{X}|} \leq \frac{n^2}{2|\mathcal{X}|} \tag{9}$$

Completing the proof The bound on the collision probability (9) can be plugged in the bound (8) on the probability of $\mathcal{B}$ when $m = rand$ to get

$$\mathbb{P}(B_l \mid M_r) \geq \frac{1}{2} - \frac{n^2}{2|\mathcal{X}|}. \tag{10}$$

The bounds (10) and (2) can then be plugged into (1) to get

$$\mathbb{P}(\mathcal{A} \text{ wins}) \geq \frac{1}{4} + \frac{\delta}{2} + \frac{1}{4} - \frac{n^2}{4|\mathcal{X}|} = \frac{1}{2} + \frac{\delta}{2} - \frac{n^2}{4|\mathcal{X}|}.$$

Hence, $\mathcal{A}$ achieves advantage $\frac{\delta}{2} - \frac{n^2}{4|\mathcal{X}|}$, as desired.

CPA security of randomized CTR mode

Theorem 2. *Let $\mathcal{E}$ be an ε-secure block cipher over $(\mathcal{K}, \mathcal{X})$, with n being the maximal number of blocks per message. Then the cipher $\mathcal{E}^P_{CTR}$ is δ-secure against all q-bounded adversaries, where*

$$\delta = 2\varepsilon + \frac{q^2 n}{|\mathcal{X}|} + \frac{qn^2}{|\mathcal{X}|}$$

.

The line of reasoning behind the proof is very similar to the semantic security case. Hence, we present only a sketch of the proof focusing on the differences between the semantic security and CPA cases. We advise the reader to first get the grasp of the semantic security proof and then try to generalize it to the CPA security case as an exercise.

Setup We need to show that for each $\mathcal{E}^P_{CTR}$-adversary $\mathcal{B}$ achieving advantage δ there is a $\mathcal{E}$-adversary $\mathcal{A}$ achieving advantage at least $\frac{\delta}{2} - \frac{q^2 n}{2|\mathcal{X}|} - \frac{qn^2}{2|\mathcal{X}|}$.

Constructing $\mathcal{A}$ using $\mathcal{B}$ The adversary $\mathcal{A}$ will play the PRP attack game while internally simulating a CPA attack game to its "black box" $\mathcal{B}$. Formally:

1. First, $\mathcal{A}$ samples a bit b from $\{0, 1\}$ and a key k from $\mathcal{K}$, both uniformly at random.

2. Then it queries $\mathcal{B}$ for the number of rounds q. In each round $1 \leq k \leq q$ it proceeds as follows:

   (a) It queries $\mathcal{B}$ for two messages m_0^k and m_1^k. Let

   $$m_0^k = z_{0,1}^k \,||\, z_{0,2}^k \,||\, \cdots \,||\, z_{0,n}^k$$
   $$m_1^k = z_{1,1}^k \,||\, z_{1,2}^k \,||\, \cdots \,||\, z_{1,n}^k,$$

   where the $z_{i,j}^k$ are the individual message blocks.

   (b) $\mathcal{A}$ then randomly samples the initialization vector IV^k. For each $1 \leq i \leq n$ it sets x_i^k to $[IV + i - 1]_{\mathcal{X}}$, i.e. to a block containing the binary representation of number $IV + i - 1$.

   (c) $\mathcal{A}$ proceeds to interact with its own challenger: for $1 \leq i \leq n$ it sends out the block x_i^k computed in the previous step and receives the corresponding block y_i^k.

   (d) Then, $\mathcal{A}$ computes the ciphertext $c^k = z_{b,1}^k \oplus y_1^k \,||\, z_{b,2}^k \oplus y_2^k \,||\, \cdots \,||\, z_{b,n}^k \oplus y_n^k$ and gives c^k to $\mathcal{B}$.

3. After the final round of the simulated CPA game, $\mathcal{A}$ observes $\mathcal{B}$'s output. If the output matches b (i.e., if $\mathcal{B}$ wins the game which $\mathcal{A}$ simulated for him), $\mathcal{A}$ outputs *fake*, otherwise it outputs *rand*.

$\mathcal{A}$'s advantage Equation (1) is derived in exactly the same way as in the semantic security case. In what follows, we will use the same notation for events as in the previous proof.

Evaluating $\mathbb{P}(B_w \mid M_f)$ As before, in case that the challengers bit is *fake*, then $\mathcal{A}$ truthfully simulates the CPA attack game against $\mathcal{E}_{CTR}^P$ for $\mathcal{B}$. Hence, the probability of $\mathcal{B}$ winning is again $\frac{1}{2} + \delta$, as in (2).

Bounding $\mathbb{P}(B_l \mid M_r)$ This is again the trickier part and we will again use a concept of a cheating game. However, in this case the argument is a bit more complex than in the semantic security case.

Recall that the purpose of the cheating game is to create a game in which $\mathcal{B}$ clearly has zero advantage. We then analyze the "distance" between the original game and the cheating game to bound $\mathcal{B}$'s advantage in the original game.

To ensure that $\mathcal{B}$ has zero advantage, it suffices to ensure that the ciphertexts c^k it receives are statistically independent of the messages m_0^k, m_1^k it sends. This would be the case, e.g., if $\mathcal{A}$ generated c^k in a completely random way. We know that $\mathcal{A}$ is not doing that, but if we knew that each "masking" block y_i^k was generated uniformly at random, then the probability distribution over the resulting c_k's would be the same as if c^k's were generated uniformly at random: let's call such a process a *cheating game* and let $\mathbb{P}_{ch}$ be the associated probability measure.

Two types of collisions Let's again discuss how the cheating game differs from the original game, where y_i^k's are computed by passing x_i^k's through a permutation f randomly sampled at the beginning of the game. There are two types of events on which the two games differ: call them *collision of type 1* and *2*, respectively:

- *Collision of type 1* (or C_1 for short) is the event that there exist $1 \leq i \leq j \leq n$ and $1 \leq k \leq q$ such that $x_i^k \neq x_j^k$ but $y_i^k = y_j^k$.
- *Collision of type 2* (or C_2 for short) is the event that there exist $1 \leq i \leq j \leq n$ and $1 \leq k < \ell \leq q$ such that $x_i^k = x_j^\ell$.

A collision (C) is an event that either C_1 or C_2 happens: $C = C_1 \vee C_2$. We claim that $\mathbb{P}(A \mid \overline{C}) = \mathbb{P}_{ch}(A \mid \overline{C})$ for any event A. The idea is that collision of type 1 can happen only in the cheating game, but not in the original game, since f is always injective. Similarly, if collision of type 2 happens, then y_i^k always equals y_j^ℓ in the original game (since f is a function), but not necessarily in the cheating game, where the y^k's are sampled independently of x^k's. However, once we prohibit the two types of collisions from happening, the two games are statistically indistinguishable.

Bounding $\mathcal{B}$'s advantage Using the same computation as in the semantic security case we derive

$$\mathbb{P}(B_l \mid M_r) \geq \frac{1}{2} - \mathbb{P}_{ch}(C \mid M_r). \tag{11}$$

It thus suffices to bound $\mathbb{P}_{ch}(C \mid M_r)$. By union bound we get

$$\mathbb{P}_{ch}(C \mid M_r) \leq \mathbb{P}_{ch}(C_1 \mid M_r) + \mathbb{P}_{ch}(C_2 \mid M_r).$$

First consider $\mathbb{P}_{ch}(C_1 \mid M_r)$. For each round k we can use the same argument as in (9) to get the bound $\frac{n^2}{2|\mathcal{X}|}$ on the probability of two y-blocks being equal in that round. Since in total we perform q rounds, we get

$$\mathbb{P}_{ch}(C_1 \mid M_r) \leq \frac{qn^2}{2|\mathcal{X}|}. \tag{12}$$

Now consider $\mathbb{P}_{ch}(C_2 \mid M_r)$. Let us *fix* some k, ℓ such that $k < \ell$. The probability that $x_i^k = x_j^\ell$ for some $i, j \leq n$ equals the probability that the intervals $[IV^k, IV^k + n - 1]$ and $[IV^\ell, IV^\ell + n - 1]$ do overlap. The overlap happens if and only if $IV^k - n + 1 \leq IV^\ell \leq IV^k + n - 1$. Since the IV's are sampled uniformly from $\{0, 1, \ldots, |\mathcal{X}| - 1\}$, the probability that IV^ℓ falls into the required interval is $\frac{2n-1}{|\mathcal{X}|} \leq \frac{2n}{|\mathcal{X}|}$. This is a probability of type 2 collision happening for a concrete choice of k, ℓ, so we need to sum it over all the possible $\frac{q(q-1)}{2} \leq \frac{q^2}{2}$ choices, yielding the bound

$$\mathbb{P}_{ch}(C_2 \mid M_r) \leq \frac{nq^2}{|\mathcal{X}|}. \tag{13}$$

Finishing the proof We proceed as in the semantic security case, but with the collision bounds derived in the previous paragraph. Thus,

$$\mathbb{P}(\mathcal{A} \text{ wins}) \geq \frac{1}{4} + \frac{\delta}{2} + \frac{1}{4} - \frac{qn^2}{4|\mathcal{X}|} - \frac{nq^2}{2|\mathcal{X}|},$$

and $\mathcal{A}$ achieves advantage at least $\frac{\delta}{2} - \frac{qn^2}{4|\mathcal{X}|} - \frac{nq^2}{2|\mathcal{X}|} \geq \frac{\delta}{2} - \frac{qn^2}{2|\mathcal{X}|} - \frac{nq^2}{2|\mathcal{X}|}$ as required.