SM4 Reversibility Proof

Anciety

October 2019

1 Terms And Notes

The terms and notes used in this proof is as follows.
Given a 128-bit plaintext input, divides into word size (X_0, X_1, X_2, X_3), each of X_i is a word (32-bit).
The output as well, described as (Y_0, Y_1, Y_2, Y_3), each 32-bit.
Round key are denoted as rk_i where i is the ranges $i = 0, 1, .., 31$.
F is the round function maps 5 words to a new word. R is a reverse transformation defined as follows:

$$R(a, b, c, d) = (d, c, b, a) \tag{1}$$

2 Structure

The structure of the SM4 encryption consists a unbalanced Feistel network. Each round, round function is applied to the 4 words and the round key which produces a new word, next round will happen on last 3 words of last round with new word appended.
Formally, for round k, we have

$$\begin{aligned}
input_k &= (X_{k-1}, X_k, X_{k+1}, X_{k+2}) \; for \; 0 \le k \le 31 \\
output_k &= (X_k, X_{k+1}, X_{k+2}, F(X_{k-1}, X_k, X_{k+1}, X_{k+2}, rk_k)) \; for \; 0 \le k \le 31 \\
input_{k+1} &= output_k \; for \; 0 \le k \le 30 \\
ciphertext &= R(output_{31})
\end{aligned} \tag{2}$$

This gives the overall structure of SM4 encryption.

3 Proof

The decryption procedure of SM4 is given as use the same procedure as encryption, except reverse the round key sequence.

Formally, this can be described as, for round k, using the same sub-labels as the encryption, we have:

$$\begin{aligned} input_k &= (X_{35-k}, X_{34-k}, X_{33-k}, X_{32-k}) \\ output_k &= (X_{34-k}, X_{33-k}, X_{32-k}, F(X_{35-k}, X_{34-k}, X_{33-k}, X_{32-k}, rk_{31-k})) \\ input_{k+1} &= output_k \ for \ 0 \leq k \leq 30 \\ plaintext &= output_{31} \end{aligned} \tag{3}$$

The order of the input is reversed by the R function accordingly.

Round Function itself is defined as:

$$F(X_0, X_1, X_2, X_3, rk) = X_0 \bigoplus T(X_1 \bigoplus X_2 \bigoplus X_3 \bigoplus rk) \tag{4}$$

To prove the decryption is actually correct, we can prove for a single round. If each single round reverses the original procedure correspondingly, it can be proved then. This is to say, we need to prove that indeed each X_i is the one used exactly in encryption, so to speak. This can be proved as follows by induction. Initial status, we have:

$$(Y_0, Y_1, Y_2, Y_3) = (X_{32}, X_{33}, X_{34}, X_{35}) = R(X_{35}, X_{34}, X_{33}, X_{32}) \tag{5}$$

This proves when $k = 0$, $input_0$ of decryption input is well-defined. Next we prove for each round, when $input_i$ is well defined. This is to prove:

$$F(X_{35-k}, X_{34-k}, X_{33-k}, X_{32-k}, rk_{31-k}) = X_{31-k} \tag{6}$$

Then:

$$\begin{gathered} F(X_{35-k}, X_{34-k}, X_{33-k}, X_{32-k}) \\ = X_{35-k} \bigoplus T(X_{34-k} \bigoplus X_{33-k} \bigoplus X_{32-k})(by\ definition) \\ \\ = X_{31-k} \bigoplus T(X_{32-k} \bigoplus X_{33-k} \bigoplus X_{34-k}) \\ \bigoplus T(X_{34-k} \bigoplus X_{33-k} \bigoplus X_{32-k})(substitute\ with\ encryption\ round) \\ \\ = X_{31-k} \end{gathered}$$

Thus, the decryption procedure is well-defined, which means the decryption can decrypt the ciphertext to plaintext.