Response to Reviewers' Comments

Reviewer: 1

I do appreciate the revised manuscript and the explanations given in the cover letter. For the newly added complexity of exploring mappings, I'd suggest to support the mathematical derivation through references. (I also understand about space and references limitations w/ ESL. If needed, please consider "outsourcing" references to all the prior art you studied, mentioned as >100 papers in the cover letter, to some webpage of yours.)

Response: We would like to thank the honorable reviewer for the valuable feedback. We have listed all the papers that we read on a separate GitHub page - `https://github.com/nivi1501/ESL/`.

Due to space constraints, the manuscript does not detail all intermediate steps of the derivation presented in Section 5b. We present a detailed derivation in this document.

Lemma 1: *An adversary needs to try out* $\leq n^{O(\log \log n)}$ *candidate mappings to find out the correct X - Y mapping for encryption schemes like AES and PRESENT.*

Proof: In certain algorithms, such as AES, it is necessary to XOR plain text bytes with key bytes prior to accessing T-tables or S-boxes. The Hamming distance between two plaintexts is retained by the XOR operation. Section 3.2 of Reference [1] also emphasizes this point.

Assume, we have address mapping as a countermeasure. This means that the i^{th} entry of the T-table or S-box is actually mapped to the j^{th} entry. In other words, the tables are permuted and this permutation is not known. The attacker can use an ingenious strategy. She can provide two plaintexts that are a Hamming distance of 1 apart. She will then get the indices of the entries in the corresponding tables (using a traditional side-channel attack). These are mapped to T-table or S-box entries that are a Hamming distance of 1 apart. She can continue to do this and find the neighbors of every entry that are a Hamming distance of 1 away. This structure is nothing but a hypercube. In this case, we know the structure of the hypercube, but we still do not know which address is mapped to which entry of the T-table/S-box. We just have the Hamming distance information. In short, we need a labeling of the hypercube. A brute force approach is to go through all labelings and see if we can break the cipher.

Now, it is well known that a hypercube with n vertices has $n(\log n)!$ labelings. We apply the Stirling's approximation to expand $(\log n)!$:

$$m! < \sqrt{2\pi m}\left(\frac{m}{e}\right)^m e^{\frac{1}{12m}} \tag{1}$$

Now, $e^{\frac{1}{12m}} < m$ for $m \geq 2$, and $\sqrt{2\pi m} < 2.51m < m^2$ for $m \geq 3$. Hence, for $m \geq 3$.

$$m! < m^3\left(\frac{m}{e}\right)^m < m^{m+3} \tag{2}$$

Let us now replace m with $\log n$. Now $(\log n)^{\log n} = n^{\log\log n}$. This can be easily proven by taking the log of both sides.

We thus have:

$$(\log n)! < (\log n)^{\log n+3} = (\log n)^3 n^{\log\log n} < n^{\log\log n+3} \tag{3}$$

Given that the number of labelings is $n(\log n)!$, we have the following for $n \geq 8$ ($\log n \geq 3$).

$$n(\log n)! < n^{\log\log n+4} = n^{O(\log\log n)} \tag{4}$$

Hence, the number of labelings of the hypercube is upper-bounded by $n^{O(\log\log n)}$.

Now, we add another lemma to further back up our argument. The results are not included in the main manuscript due to space constraints.

Lemma 2: $I(K;Y|T)$ establishes a lower bound on $I(X;Y|T)$.

Proof: Given the plaintext T, key K, we have a Markov chain $\langle key(K), T\rangle \rightarrow X \rightarrow Y$ [2]. This basically means that the MI between X and Y determines the amount of information that can be extracted out of Y about the key. The two assumptions are that this is a Markov process and the noise distribution N is independent of the key, K.

In the generic case, we can use the data processing inequality in information theory and write $I(K;Y|T) \leq I(X;Y|T)$ (derived from Eqn.8 in [2]).

Using Equation 10 in [2]:

$$I((X,T);(Y,T)) = I(X;Y|T) + H(T) \tag{5}$$

$$I((K,T);(Y,T)) = I(K;Y|T) + H(T) \tag{6}$$

Now, using Equation 8 in [2]:

$$I((K,T);(Y,T)) \leq I((X;T),(Y;T)) \tag{7}$$

Using the previous equations, we get:

$$I(K;Y|T) \leq I(X;Y|T) \tag{8}$$

References

[1] R. Spreitzer and T. Plos, "Cache-access pattern attack on disaligned aes t-tables," in *International Workshop on Constructive Side-Channel Analysis and Secure Design*. Springer, 2013, pp. 200–214.

[2] E. de Chérisey *et al.*, "Best information is most successful," *CHES*, 2019.