Jessup Barrueco HW1.2 Question 1

Differential cryptanalysis focuses on the nonlinear operations of the S-DES algorithm, the S-Boxes, where the differences can be analyzed through multiple rounds and operations of the cipher. Differences are computed with the XOR function. XOR is used because it gives information about the two inputs, namely whether or not they were equal. In S-DES, the difference in a ciphertext pair for a specific difference of a plaintext pair is influenced by the key, and certain plaintext differences occurs with a higher probability than other differences.

Consider the S-Box S0. To construct its differential distribution table, see that there are 16^2 or 256 possible input pairs (x, x*). Therefore, a brute force attack is quite possible, as the S-DES algorithm is vulnerable. As the 4-bit quantities x, x* and x'=x(XOR)x* vary over their 16 possible values, the 2-bit quantities y=S(x), y*=S(x*), y'=y(XOR)y*=S(x)(XOR)S(x*) vary over their 4 possible values. The distribution on the differential output y' can be computed for S-box S0 by counting the number of times each value y' occurs as (x,x*) varies over its 256 possible values.

A differential distribution table (DDT) for S0 is given on Slide 39 of the Week2.1.Modes_DiffCrypto lecture slides, uploaded to Piazza. Once the attacker constructs the DDT, the process of a differential cryptoattack involves considering each possible input XOR x', the possible output XORs y', and the frequencies at which they occur. Note that since input pairs always come as duals (a, b), (b, a), all occurrence-counts will be multiples of 2. In the construction of the DDT of S0, the attacker is able to generate a list of all inputs (x, x*) which XOR to the chosen x' value. To begin the cryptoattack, the attacker chooses a spot in the DDT, they choose input and output values x', y', (call chosen values X', Y') and see the frequency at which the input X' yields the output Y'. Then, the process of key determination begins. See that the input XOR value x' stays consistent regardless of the value of the key, a proof for which can be found on page 19 of the Koç paper on Differential Cryptanalysis. From here, the attacker can derive the key with the following relation:

$$S0_I = S0_E \text{ (xor) } S0_K$$
$$S0_K = S0_I \text{ (xor) } S0_E$$

In the construction of the DDT, the attacker is able to generate a list of the possible output values y' given the attacker's chosen input X'. They also know the potential input value pairs (x, x*) which XOR to their chosen input X'. They arbitrarily choose a pair (x, x*) which XOR to their initially chosen X'. Then, they inspect the list of values which all have a difference value of x', which return the initially chosen output Y' from Sbox S0. Call this set of values $\{m_0 \ldots m_{n-1}\}$ Generating a set of possible keys is the process of XORing each value m_i (*i* in *n*) with x, and then again with x* (where (x, x*) XOR to attacker's chosen input X'). The results of that process are the members of the set of possible keys. One "round" is now complete, and the attacker knows that the correct key is present in this set. But how do they know which key? Again, though brute force is quite possible, the formal process would be picking another two values (z, z*) which XOR to the attacker's chosen input value X', and repeating the process with a different chosen output, which will have a different occurrence frequency. At the end of that second round, the attacker would intersect their two sets of possible keys. The correct key is necessarily in that intersection. If, after the second round, the cardinality of the intersected set of possible keys is greater than 1, the attacker repeats this process. This process will repeat until there is only one possible option for a correct key.

Q2. [25pnts] Consider the crypto system below and compute H(K|C)

- P={a,b,c} with $P_P(a)=1/3$ $P_P(b)=1/6$ $P_P(c)=1/2$
- K = (k_1,k_2,k_3) with $P_K(k_1)=1/2$ $P_K(k_2)=1/4$ $P_K(k_3)=1/4$
- C = {1,2,3,4}

$e_{k_1}(a)=1$ $e_{k_1}(b)=2$ $e_{k_1}(c)=2$
$e_{k_2}(a)=2$ $e_{k_2}(b)=3$ $e_{k_2}(c)=1$
$e_{k_3}(a)=3$ $e_{k_3}(b)=4$ $e_{k_3}(c)=4$

$P_C(y) = \sum_{\{k: y \in C(k)\}} P_K(k) \cdot P_P(d_k(y))$

$P_C(1) = \left(\frac{1}{3}\cdot\frac{1}{2}\right) + \left(\frac{1}{2}\cdot\frac{1}{4}\right) = \frac{1}{6}+\frac{1}{8} = \frac{7}{24}$

$P_C(2) = \left(\frac{1}{3}\cdot\frac{1}{4}\right) + \left(\frac{1}{6}\cdot\frac{1}{2}\right) + \left(\frac{1}{2}\cdot\frac{1}{2}\right) = \frac{5}{12}\left(=\frac{10}{24}\right)$

$P_C(3) = \left(\frac{1}{3}\cdot\frac{1}{4}\right) + \left(\frac{1}{6}\cdot\frac{1}{4}\right) = \frac{1}{12}+\frac{1}{24} = \frac{1}{8}\left(=\frac{3}{24}\right)$

$P_C(4) = \left(\frac{1}{6}\cdot\frac{1}{4}\right) + \left(\frac{1}{2}\cdot\frac{1}{4}\right) = \frac{1}{6} \quad \left(=\frac{4}{24}\right)$

if $p_i = \frac{1}{n}$ for $1 \le i \le n$, then:
$H(X) = \log_2 n$

$P_P(a\|1) = \left(\frac{1}{3}\cdot\frac{1}{2}\right)/\frac{7}{24} = \frac{4}{7}$	$P_P(b\|1) = 0$	$P_P(c\|1) = \frac{3}{7}$
$P_P(a\|2) = \frac{1}{5}$	$P_P(b\|2) = \frac{1}{10}$	$P_P(c\|2) = \frac{3}{5}$
$P_P(a\|3) = \frac{2}{3}$	$P_P(b\|3) = \frac{1}{3}$	$P_P(c\|3) = 0$
$P_P(a\|4) = 0$	$P_P(b\|4) = \frac{1}{4}$	$P_P(c\|4) = \frac{3}{4}$

$H(P) = -\left(\frac{1}{3}\log_2\frac{1}{3} + \frac{1}{6}\log_2\frac{1}{6} + \frac{1}{2}\log_2\frac{1}{2}\right) \approx 1.46$

$H(K) = -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{4}\log_2\frac{1}{4} + \frac{1}{4}\log_2\frac{1}{4}\right) = 1.5 = \frac{3}{2}$

$H(C) = -\left(\frac{7}{24}\log_2\frac{7}{24} + \frac{5}{12}\log_2\frac{5}{12} + \frac{1}{8}\log_2\frac{1}{8} + \frac{1}{6}\log_2\frac{1}{6}\right) \approx 1.85$

$H(K|C) = H(K) + H(P) - H(C)$

$\approx 1.5 + 1.46 - 1.85$

≈ 1.11