Sécurité des Systèmes d'Information
Series 2 Correction - Entropy

October 6th, 2021

Exercise 1 : Encryption and Entropy

Compute the following entropies :

1. $H(P) = \frac{1}{4}log_2(4) + \frac{3}{20}log_2(\frac{20}{3}) + \frac{3}{5}log_2(\frac{5}{3}) = \frac{1}{2} + \frac{3}{20}log(4) + \frac{3}{20}log(\frac{5}{3}) + \frac{3}{5}log_2(\frac{5}{3}) = \frac{4}{5} + \frac{3}{4}(log_2(5) - log_2(3)) \simeq 1.352724$

2. $H(K) = 3(\frac{1}{3}log_2(3)) = log_2(3) \simeq 1.584963.$

3. Let's first compute the probability of each ciphertext :

   $p(c_1) = p(m_1, k_1) + p(m_1, k_3) = \frac{1}{12} + \frac{1}{12} = \frac{1}{6}$

   $p(c_2) = p(m_2, k_3) + p(m_3, k_1) = \frac{3}{60} + \frac{6}{30} = \frac{1}{4}$

   $p(c_3) = p(m_2, k_2) + p(m_3, k_3) = \frac{3}{60} + \frac{6}{30} = \frac{1}{4}$

   $p(c_4) = p(m_1, k_2) + p(m_2, k_1) = \frac{1}{12} + \frac{3}{60} = \frac{2}{15}$

   $p(c_5) = p(m_3, k_3) = \frac{6}{30} = \frac{1}{5}$

   Then, the entropy of the ciphertexts is : $H(C) = \frac{1}{6}log_2(6) + \frac{1}{4}log_2(4) + \frac{1}{4}log_2(4) + \frac{2}{15}log_2(\frac{15}{2}) + \frac{1}{5}log_2(5)$

   $$= \frac{1}{6}(1 + log_2(3)) + \frac{1}{2} + \frac{1}{2} + \frac{2}{15}(log_2(5) + log_2(3) - 1) + \frac{1}{5}log_2(5)$$

   $$= \frac{1}{6} + \frac{1}{6}log_2(3) + \frac{1}{2} + \frac{1}{2} + \frac{2}{15}log_2(5) + \frac{2}{15}log_2(3) - \frac{2}{15} + \frac{1}{5}log_2(5)$$

$$= \frac{31}{30} + \frac{3}{10}log_2(3) + \frac{1}{3}log_2(5) \simeq 2.282798$$

4. We first compute the conditionnal probabilities, $p(m_i|c_i)$. Some of these doesn't even need calculations.

   These pairs (plaintext,ciphertext) are not possible in this encryption system :
   $p(m_2|c_1) = p(m_3|c_1) = p(m_1|c_2) = p(m_1|c_3) = p(m_3|c_4) = p(m_1|c_5) = p(m_2|c_5) = 0$.

   And for ciphers c_1 and c_5, there's only one possible plaintext :
   $p(m_1|c_1) = p(m_3|c_5) = 1$.

   Then we only have to compute the few ones left :
   $p(m_2|c_2) = \frac{p(m_2 \cap c_2)}{p(c_2)} = \frac{p(m_2,k_3)}{p(c_2)} = \frac{\frac{3}{60}}{\frac{1}{4}} = \frac{12}{60} = \frac{1}{5}$
   (As m_2, k_3 is the only pair for which m_2 becomes c_2). Similarly :
   $p(m_3|c_2) = \frac{\frac{6}{30}}{\frac{1}{4}} = \frac{4}{5}$
   (Since there's only two possibilities of plaintext for c_2, then we could also have computed it as $1 - p(m_2|c_2)$)
   $p(m_2|c_3) = \frac{\frac{3}{60}}{\frac{1}{4}} = \frac{1}{5}$
   $p(m_3|c_3) = \frac{4}{5}$
   $p(m_1|c_4) = \frac{\frac{1}{12}}{\frac{2}{15}} = \frac{15}{24} = \frac{5}{8}$
   $p(m_2|c_4) = \frac{3}{8}$

   And the conditional entropy can be written as :

   $H(P|C) = \sum_{i=1}^{5} p(c_i)(\sum_{j=1}^{3} p(m_j|c_i)log_2(\frac{1}{p(m_j|c_i)}))$

   Which, excluding all cases with probability 0, gives us :

$$H(P|C) = \frac{1}{6}(1log_2(1)) + \frac{1}{4}(\frac{1}{5}log_2(5) + \frac{4}{5}log_2(\frac{5}{4}))$$

$$+\frac{1}{4}(\frac{1}{5}log_2(5) + \frac{4}{5}log_2(\frac{5}{4})) + \frac{2}{15}(\frac{5}{8}log_2(\frac{8}{5}) + \frac{3}{8}log_2(\frac{8}{3})) + \frac{1}{5}(1log_2(1))$$

$$= 0+\frac{1}{2}(\frac{1}{5}log_2(5)+\frac{4}{5}log_2(5)-\frac{4}{5}log_2(4))+\frac{2}{15}(\frac{5}{8}log_2(8)-\frac{5}{8}log_2(5)+\frac{3}{8}log_2(8)-\frac{3}{8}log_2(3))+0$$

$$= \frac{1}{2}log_2(5) - \frac{4}{5} + \frac{2}{15}log_2(8) - \frac{1}{12}log_2(5) - \frac{1}{20}log_2(3)$$

$$= -\frac{4}{5} - \frac{1}{20}log_2(3) + \frac{5}{12}log_2(5) + \frac{2}{15}log_2(8) \simeq 0.488222$$

Exercise 2 : Random Generators

1. $P_G(00) = (0.5+\delta)(0.5+\delta) = 0.25 + \delta^2 + \delta$
   $P_G(11) = (0.5-\delta)(0.5-\delta) = 0.25 + \delta^2 - \delta$
   $P_G(01) = (0.5+\delta)(0.5-\delta) = 0.25 - \delta^2$
   $P_G(10) = (0.5-\delta)(0.5+\delta) = 0.25 - \delta^2$

2. $P_G(01) = P_G(10)$, so $P_A(0) = P_A(1) = 0.5$

3. $P(GenerateABitForA) = 0.5 - 2\delta^2$. That mean for 2 bits in G, we can expect to generate $0.5 - 2\delta^2$ bits for A. If we generate $2\frac{x}{0.5-2\delta^2}$ bits with G, we can expect x bits for A.

4. This method allows us to transform a non-random generator into a random one, at the cost of time and energy, as we'll need to generate multiple bits with G to create one really random bit with A. The more the bias is strong, the more it will cost time and energy.

5. With each bit generated, we have two cases : we either have the same bit as the previous one, or the opposite. If it's the same, we're back to the same situation (the new pair is not usable, and we have the same last bit as before). If it's different, then we have a usable pair (if the previous bit was 0, then 01, if the previous bit was 1, then 10), and the opposite last bit as previously. Meaning this process will generate the pairs 01, 10, 01, 10, 01, 10, 01, ... (with some 11 or 00 in between that will be ignored). This means we won't generate random bits with A, but a deterministic series of 0,1,0,1,0,1,0,1,...
   That would be an atrocious generator, as every bit would be deterministic except for the very first one.

Exercise 3 : Password Entropy

1. We have a randomly chosen date, with 365 days and 2022 years (from 0000 to 2021), so 365 * 2022 = 738030 valid dates :
   $E(mdp1) = 738030(\frac{1}{738030}log_2(738030)) = log_2(738030) \simeq 19.4933$

2. $E(mdp2) = \sum_{c=1}^{11} E(c) = 11*(\sum_{i=1}^{256} p(i)log(\frac{1}{p(i)})) = 11*256*(\frac{1}{256}log(256)) = 11 * log(256) = 88$

3. $E(mdp3) = \sum_{c=1}^{11} E(c) = 10(\frac{1}{10}log_2(10))+26(\frac{1}{26}log_2(26))+26(\frac{1}{26}log_2(26))+194(\frac{1}{194}log_2(194)) + 7 * log(256) = log_2(10) + 2log_2(26) + log_2(194) + 56 \simeq 76.322720$

4. $E(mdp4) = 5*E(mot) = 5*200000(\frac{1}{200000}log_2(200000)) = 5log_2(200000) \simeq 88.048202$

5. Let P_1 to P_6 be the Pokémon. Then :
   $E(P_1) = \sum_{i=1}^{901} p(i) log(\frac{1}{p(i)}) = 901(\frac{1}{901} log_2(901)) = log_2(901)$.
   The same calculation holds for other Pokémon, except we're having 900 choices for the second one, 899 for the third, and so on down to 896 choices for the sixth Pokémon (since we want six different ones). Then, the entropy of the password is :
   $E(mdp5) = E(P_1)+E(P_2)+E(P_3)+E(P_4)+E(P_5)+E(P_6) = log_2(901)+ log_2(900) + log_2(899) + log_2(898) + log_2(897) + log_2(896) \simeq 58.8682$