Math 4981 Spring 2021
Cryptography HW 5 Solutions
Due Thursday, February 18

1. Let X be a random variable with possible outcomes $x_1, \ldots, x_n$, and Y a random variable with possible outcomes $y_1, \ldots, y_m$. Let Z be a random variable that corresponds to testing X followed by Y, so the possible outcomes are pairs (x_i, y_j) with $P(x_i, y_j) = P(x_i)P(y_j)$.

   Use the definition of entropy to prove that $H(Z) = H(X) + H(Y)$. This is a special case of property 3 from Shannon's theorem.

   Solution:

$$\begin{aligned}
H(Z) &= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i)P(y_j)\log_2(P(x_i)P(y_j)) \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i)P(y_j)\left(\log_2(P(x_i)) + \log_2(P(y_j))\right) \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i)P(y_j)\log_2(P(x_i)) - \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i)P(y_j)\log_2(P(y_j)) \\
&= -\sum_{i=1}^{n}\left(P(x_i)\log_2(P(x_i))\sum_{j=1}^{m}P(y_j)\right) - \sum_{j=1}^{m}\left(P(y_j)\log_2(P(y_j))\sum_{i=1}^{n}P(x_i)\right) \\
&= -\sum_{i=1}^{n}\left(P(x_i)\log_2(P(x_i))\cdot 1\right) - \sum_{j=1}^{m}\left(P(y_j)\log_2(P(y_j))\cdot 1\right) \\
&= H(X) + H(Y).
\end{aligned}$$

   Definition. The *Key Equivocation* of a cryptosystem is $H(K|C) = H(K) + H(M) - H(C)$. (There's a more complicated formula in terms of random variables, which I'm omitting here). It measures the amount of information about the key revealed by the ciphertext.

   In particular, it tells us how much *more* information we get from the key if we already know the ciphertext. If it is low, knowing the ciphertext tells us a lot about the key. If it's zero, we can determine the key and message purely from the ciphertext.

2. Suppose we have a cryptosystem with two keys $\mathcal{K} = \{k_1, k_2\}$ and three plaintext $\mathcal{M} = \{m_1, m_2, m_3\}$. Suppose the plaintexts have probabilities $P(m_1) = 1/2, P(m_2) = P(m_3) = 1/4$, while the keys are equally likely with $P(k_1) = P(k_2) = 1/2$.

(a) Create an encryption function with three ciphertexts $\mathcal{C} = \{c_1, c_2, c_3\}$, such that $P(c_1) = 1/2$.

(b) Compute $H(K), H(M), H(C)$.

(c) Compute the equivocation $H(K|C)$.

(d) How secure is this cipher?

Solution: There are many possible solutions here. I'll work through one.

(a)

	m_1	m_2	m_3
k_1	c_2	c_1	c_1
k_2	c_1	c_3	c_3

(b)

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

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

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

(c) $H(K|C) = H(K) + H(M) - H(C) = 1$.

(d) This is reasonably secure. If we already have the ciphertext, knowing the key gives us 1 bit of information; but in a vaccuum, it's also true that knowing the key gives us 1 bit of information. So the ciphertext carries no information.

(For other choices of cryptosystem, the equivocation would be different. If the equivocation is higher, the

3. How does key equivocation relate to unicity distance? (Hint: if your message is much longer than the unicity distance, what should happen to the key equivocation?)

Solution: Unicity distance is the message length at which you expect to have a 50% chance of completely decrypting the ciphertext without any additional info; so it's the distance where you expect the key equivocation to be close to zero. With much shorter messages the key equivocation will be high; with much longer messages, the key equivocation will be zero.

Note that the unicity distance isn't deterministic, so we can't say that the equivocation is zero any time the message length is over the unicity distance.

4. Compute the unicity distance for

(a) An autokey cipher with a N-letter keyword.

(b) An affine cipher

(c) A Hill cipher with a block size of 2. (Note: only count matrices that are valid keys! Assume 12/26 of possible matrices are valid keys; this isn't quite right but it's close enough for our purposes.)

(d) A Hill cipher with a block size of 5.

Solution:

(a) If the key word has length N then there are 26^N possible keys, which is $4.7N$ bits of entropy. We divide by 3.2 to get a unicity distance of $1.47N$.

(b) An affine cipher has $12 \cdot 26 = 312$ keys, which is $\log_2(312) \approx 8.3$ bits of entropy. So the unicity distance is about $\frac{8.3}{3.2} \approx 2.6$.

(c) There are 26^4 possible matrices, but only 12/26 of them will be invertible. So there are $26^3 \cdot 12$ possible keys, which is 17.7 bits of entropy. We divide by 3.2 to get 5.5.

(Actually, this is wrong—there are actually somewhat fewer valid keys than this, and the true entropy of the keyspace is about 16.86 bit of entropy, so the unicity distance is about 5.268.).

(d) There are 26^{25} matrices and again only 12/26 will work. So there are $26^{24} \cdot 12$ possible keys, which corresponds to about 116.4 bits of entropy. Dividing by 3.2 gives a unicity distance of 36.4.

(Again, this is wrong—the true entropy of the keys is about 114 bits, and the real unicity distance is about 35.718).

5. From the definition of big-O notation, prove that $x^2 + \sqrt{x} = O(x^2)$.

Solution: When $x \geq 1$, we have $\sqrt{x} \leq x^2$ so $x^2 + \sqrt{x} \leq x^2 + x^2 = 2x^2$. Thus we can take $C = 1$ and $c = 2$.

6. Prove (using the definition or the limit property) that:

(a) $k^{300} = O(2^k)$

(b) $(\log_2(k))^{100} = O(k)$.

Solution:

(a) $\lim_{k\to\infty} \frac{k^{300}}{2^k} = 0$ so $k^{300} = O(2^k)$.

(b) $\lim_{k\to\infty} \frac{\log_2(k)^{100}}{k} = 0$ so $\log_2(k)^{100} = O(k)$.

7. (a) Prove that $2^{\sqrt{k}} = O(2^{\varepsilon k})$ for any $\varepsilon > 0$.

Solution:

Fix some $\epsilon > 0$. Take $c = \frac{1}{\varepsilon^2}$, and if $k > c$ then $\varepsilon k > \varepsilon\sqrt{k}\frac{1}{\varepsilon} = \sqrt{k}$, and so $2^{\varepsilon k} > 2^{\sqrt{k}}$.

(b) Prove that $k^n = O(2^{\sqrt{k}})$ for any n.

Solution: Let $n \in \mathbb{N}$. Then by L'Hospital's Rule:

$$\begin{aligned}
\lim_{k\to\infty} \frac{k^n}{2^{\sqrt{k}}} &= \lim_{k\to\infty} \frac{nk^{n-1}}{2^{\sqrt{k}} \cdot {}^{\ln(2)}/_{2\sqrt{k}}} \\
&= \lim_{k\to\infty} \frac{nk^{n-1/2}}{2^{\sqrt{k}-1}\ln(2)}
\end{aligned}$$

Repeated application of L'Hospital's Rule, $2n$ times, will give

$$\lim_{k\to\infty} \frac{k^n}{2^{\sqrt{k}}} = \lim_{k\to\infty} \frac{n(n-1/2)(n-1)\dots(3/2)(1)}{2^{\sqrt{k}-2n}\ln(2)^{2n}} = 0.$$