Problem Set 1 Solutions

Problem 2. [30 points] The ciphertext

```
QFL HCVPS PX V ANSWLCEZK NCJVS; PQ XQVCQX QFL BPSZQL RNZ JLQ ZT PS QFL
BNCSPSJ VSW WNLX SNQ XQNT ZSQPK RNZ JLQ QN DKVXX
```

has been created using a punctuation-respecting substitution cipher on the alphabet of English letters. Your task is to decrypt this ciphertext and recover the plaintext. Show the steps you used to arrive at your solution, the final plaintext, and a table providing, for each letter, its decoding. (*Hint:* J decodes to G.)

An example of this type of cryptanalysis was done in class. It is also in Chapter 2 of the notes, and the latter can serve as a model for how to approach the problem and write the solution.

We typically begin with a frequency analysis. In this case, however, I skipped that stage, since something else sprang to my attention. Namely there are three occurences of the word QFL in the ciphertext, one at the very beginning. What three letter word could be so common and occur at the beginning of the sentence? I felt pretty comfortable guessing it to be THE. This tells me that $\pi^{-1}(\mathtt{Q}) = \mathtt{T}$, $\pi^{-1}(\mathtt{F}) = \mathtt{H}$ and $\pi^{-1}(\mathtt{L}) = \mathtt{E}$, where π denotes the permutation (key) that was used for encryption. The ciphertext contains the word QN. Knowing that Q decodes to T, the only choice for N is O. The second row of the table in Fig. 1 shows where we are. (We have taken into account the hint as well.) Now let us write the ciphertext again, this time indicating above different letters what we believe them to represent:

```
THE             O  E     O G  ;  T  T  T  THE     TE  O  GET       THE
QFL HCVPS PX V ANSWLCEZK NCJVS; PQ XQVCQX QFL BPSZQL RNZ JLQ ZT PS QFL

 O    G      OE   OT  TO    T    O  GET TO
BNCSPSJ VSW WNLX SNQ XQNT ZSQPK RNZ JLQ QN DKVXX
```

The ciphertext contains the word PQ. Since Q stands for T, P could be only I or A. The third word of the ciphertext is PX. This could then be one of: IN, IS, AN, AS, AM, so X could be one of: N, S, M. But we have a ciphertext word XQVCQX and we know that Q stands for T. The only one of the consonants N, S, M that fits before and after a T is S. Conclusion: $\pi^{-1}(\mathtt{X}) = \mathtt{S}$. Now the two choices for the third word of the ciphertext are IS and AS, but only the first seems like a grammatical fit. (I imagine HCVPS is a noun, and we follow it by the verb IS.) So I guess $\pi^{-1}(\mathtt{P}) = \mathtt{I}$. This means that $\pi^{-1}(\mathtt{V}) = \mathtt{A}$. (The symbol V occurs by itself in the ciphertext and can thus only represent I or A.) The two letter word PS occurs in the ciphertext. At this point this could only be IN. (We know P represents I, and so S could only be T or N, but only the latter is still available.) The last three letters of the ciphertext word WNLX being OES, the only choice for W is D. The third row of the

τ	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
$\pi^{-1}(\tau)$						H				G		E		O			T									
$\pi^{-1}(\tau)$						H				G		E		O		I	T		N			A	D	S		
$\pi^{-1}(\tau)$	W	M	R	C	F	H		B		G	L	E		O		I	T	Y	N	P		A	D	S		U

Figure 1: Cryptanalysis of Problem 1.

table in Fig. 1 shows our view of π at this point, and the following shows where we are in terms of decryption of the ciphertext:

```
THE   AIN IS A  ONDE     O GAN; IT STA TS THE  IN TE  O  GET    IN THE
QFL HCVPS PX V ANSWLCEZK NCJVS; PQ XQVCQX QFL BPSZQL RNZ JLQ ZT PS QFL

 O NING AND DOES NOT STO   NTI   O  GET TO   ASS
BNCSPSJ VSW WNLX SNQ XQNT ZSQPK RNZ JLQ QN DKVXX
```

At this point it is pretty easy. `XQVCQX` could only be `STARTS`, meaning `C` represents `R`. So the last four letters of `HCVPS` are `RAIN`, and although many choices of the first letter yield English words, namely `B`, `D`, `G`, `T`, the only one still available is `B`, so `H` represents `B`. You can now pretty much read it off. The fourth row of Fig. 1 shows our final view of π, and the following shows the decrypted ciphertext:

```
THE BRAIN IS A WONDERFUL ORGAN; IT STARTS THE MINUTE YOU GET UP IN THE
QFL HCVPS PX V ANSWLCEZK NCJVS; PQ XQVCQX QFL BPSZQL RNZ JLQ ZT PS QFL

MORNING AND DOES NOT STOP UNTIL YOU GET TO CLASS
BNCSPSJ VSW WNLX SNQ XQNT ZSQPK RNZ JLQ QN DKVXX
```

The message, by the way, is a slight corruption of a quote due to Robert Frost.

Problem 3. **[30 points]** Let $m = 6$, and let $\mathbf{Z}_m$ denote the set $\{0, \ldots, m-1\}$. Let $X \bmod m$ denote the remainder obtained when dividing X by m.

1. **[15 points]** Consider the symmetric encryption scheme in which the encryption of message $M \in \mathbf{Z}_m$ under key $K \in \mathbf{Z}_m$ is $M + K \bmod m$. Is this encryption scheme perfectly secure? Why or why not?

The simplest way to get a handle on what is going on here is to make a table whose row K, column M entry is $\mathcal{E}_K(M)$:

$M =$

$K =$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

Now we claim that

$$\Pr[\mathcal{E}_K(M) = C] = \frac{1}{6} \quad (1)$$

for every $M \in \mathbf{Z}_m$ and $C \in \mathbf{Z}_m$, where the probability is over a random choice of K from $\mathbf{Z}_m$. This implies the scheme is perfectly secure. So why is Equation (1) true? For example consider $M = 2$ and $C = 5$. Then the probability that $\mathcal{E}_K(M) = C$ is the number of times C occurs in column M of the table divided by the number of choices for K. But C occurs exactly once in this column and the number of possible rows is 6 so the probability is 1/6. More generally we can compute

$$\begin{aligned}\Pr[\mathcal{E}_K(M) = C] &= \frac{|\{K \in \mathbf{Z}_m\colon\ K + M \bmod m = C\}|}{|\mathbf{Z}_m|} \\ &= \frac{1}{6}\end{aligned}$$

because the size of the set in the numerator is 1.

Pictorially, there is an easy way to see why the scheme is perfectly secure. It is because in every column of the table, every value in $\mathbf{Z}_m$ shows up exactly once.

2. **[15 points]** Consider the symmetric encryption scheme in which the encryption of message $M \in \mathbf{Z}_m$ under key $K \in \mathbf{Z}_m$ is $M + 2K \bmod m$. Is this encryption scheme perfectly secure? Why or why not?

Again, make a table whose row K, column M entry is $\mathcal{E}_K(M)$:

$K =$ / $M =$

	0	1	2	3	4	5
0	0	1	2	3	4	5
1	2	3	4	5	0	1
2	4	5	0	1	2	3
3	0	1	2	3	4	5
4	2	3	4	5	0	1
5	4	5	0	1	2	3

Let $C = 4$ and $M_1 = 0$ and $M_2 = 1$. Then

$$\begin{aligned}\Pr[\mathcal{E}_K(M_1) = C] &= \frac{2}{6} = \frac{1}{3} \\ \Pr[\mathcal{E}_K(M_2) = C] &= \frac{0}{6} = 0\,.\end{aligned}$$

Since we have found M_1, M_2, C such that

$$\Pr[\mathcal{E}_K(M_1) = C] \neq \Pr[\mathcal{E}_K(M_2) = C]$$

the scheme is not perfectly secure.

Why are the probabilities as claimed? To evaluate the probability that $\mathcal{E}_K(M) = C$ just count the number of times C occurs in column M of the table and divide by the number of keys, which is 6.

In both cases, the key is a randomly chosen element of $\mathbf{Z}_m$ and the message space is also $\mathbf{Z}_m$.
