Burt Rosenberg

Problem Set 2

OUT: 13 SEPTEMBER, 1994
DUE: 22 SEPTEMBER, 1994

Goals

To review arrays, files and the handling of characters and text.

Reading Assignment

Read Chapters 3 and 4 from A BOOK ON C.

Programming Assignment

Both the Caesar and Vigenere ciphers are forms of *simple substitution ciphers.* They are examples of a choice of a mapping function which is applied character by character to the *cleartext* to arrive at the *ciphertext.* The inverse map is applied to retrieve the cleartext from the ciphertext.

Label the characters in the cleartext as $s_0 s_2 \ldots s_{r-1}$, where r is the message length. Each s_i becomes σ_i by the substitution map S, $\sigma_i = S(s_i)$, for $i = 0, \ldots, r-1$. The resulting ciphertext is $\sigma_0 \sigma_2 \ldots \sigma_{r-1}$. For the Caesar cipher, the choice of substitutions is limited to one of 26 different cyclical shifts around the alphabet. We can describe this more precisely. If the desired alphabet has N letters, the letters are placed in correspondence with the integers $0, \ldots, N-1$. Because this correspondence is so precise, we can let s_i and σ_i denote simultaneously the character and its corresponding integer value, interpreting according to context. Then the Caesar cipher with integer key K is described by the formula,

$$\sigma_i = S(s_i) = (s_i + K) \bmod N.$$

Strictly speaking, the Caesar cipher is the case $K = 3$. In the case of the Vigenere, the shift value depends on a keyword $\rho_0 \rho_1 \ldots \rho_{k-1}$,

$$\sigma_i = S(s_i) = (s_i + \rho_j) \bmod N, \quad j = i \bmod k.$$

To break a Vigenere cipher, guess k and dissect the message into k pieces, where σ_i is placed in piece j if $j = i \bmod k$. Then break each piece using letter counts, as you would a Caesar cipher. We now discuss a good method for guessing k.

Consider placing an equal number of each letter in an urn. Drawing two letters from the urn, the probability of drawing a pair is 1/26. The first character will be whatever it will be. One in twenty-six times the second character drawn will match it. If, in contrast, you took a book and cut it up into individual characters and threw these characters into an urn, the probability of drawing a pair will be slightly different. It will be 0.06598. Intuitively, since the letter frequencies of English are biased towards certain letters, your hand will be drawn towards a smaller set of candidates for the pair, so the chance of a pair goes up.

One can detect in the ciphertext a rhythm in the number of repeated letters in the cipher text, when pairs are drawn at intervals of l letters apart, varying the l. Listening to this rhythm, one can hear the value of k and begin breaking open this otherwise difficult cipher. This idea is due to William Freedman,

the cryptanalyst responsible for the breaking of the Japanese Purple cipher just before the outbreak of WWII.

Define shifted versions σ^l of the ciphertext σ according to the formula,

$$(\sigma^l)_j = \sigma_i, \quad j = (i + l) \bmod r.$$

(Recall r is the length of σ.) Now form a count of double letters for various values of l,

$$D(l) = \#\{\, t \,|\, (\sigma^l)_t = \sigma_t \,\}.$$

Plot $D(l)$ versus l and look for a regularly spaced pattern of spikes. It is likely that the distance between two adjacent spikes is the key length k.

Write C programs to help you in the task of breaking the following Vigenere cipher. Good Luck.

```
bqmeod ugmqr
ugnaat tsdiu

brizks cvmrni ininkr durkwv hrtdiv a
cfeauh cmvtvn sslzuy ylmfcm ubrsaj k
nqrsem vfnyax izkkpg rboww
wwtfra dswkqm vns
rfquey eewcgk ngomaf zndbfi meouaa dovzmu

iswqav zgcuix himpwr mukybe lvmn
grbehg hqvfwm sflqig qpgnnp fimcka ofazo
cnqarx vzvhrs ajkiuc rnf
xyzwsg ogxfct hrapwr jqvrtt iszcnp hqw
kwyoad qvrbvh ruzjrq niagex rzu

ohtajt wpfhsu seiutu eiepqu
aadflv eqnqed xyivmn nwrfeu
ohtajk pgcuaa wnwwlq caqvjn ifs

flrbvh rnuwcw xeyizi jayefa uh
tpklqr qrzvyo adqvni vcuizk kpgsga dw
zavhra uqrvft ueqru

jgiagn ykugnj eiecsg dvnfsk pgteeq w
```

HINT: Unix is based on the idea that the foundation of all computation is text processing. For this reason, text files are "universal currency", and easy redirection and piping of stdin and stdout is available at the command level. Write a collection of small C programs, each flexible but focused on a single task, and connect them with temporary files or pipes. See also `man sed`.