**Assignment 4 – Encryption/Decryption**

# Overview

In this assignment, you will be creating five small programs that encrypt and decrypt information using a one-time pad-like system. I believe that you will find the topic quite fascinating: one of your challenges will be to pull yourself away from the stories of real-world espionage and tradecraft that have used the techniques you will be implementing.

These programs serve as a capstone to what you have been learning in this course and will combine the multi-processing code you have been learning with socket-based inter-process communication. Your programs will also be accessible from the command line using standard UNIX features like input/output redirection, and job control. Finally, you will write a short compilation script.

# Specifications

Use the following link as your primary reference on One-Time Pads (OTP):

[**http://en.wikipedia.org/wiki/One-time\_pad**](http://en.wikipedia.org/wiki/One-time_pad)

The following definitions will be important:

**Plaintext** is the term for the information that you wish to encrypt and protect. It is human readable.

**Ciphertext** is the term for the plaintext after it has been encrypted by your programs. Ciphertext is not human-readable, and in fact cannot be cracked, if the OTP system is used correctly.

A **Key** is the random sequence of characters that will be used to convert Plaintext to Ciphertext, and back again. It must not be re-used, or else the encryption is in danger of being broken.

Below is an excerpt from Wikipedia:

“Suppose Alice wishes to send the message "HELLO" to Bob. Assume two pads of paper containing identical random sequences of letters were somehow previously produced and securely issued to both.

Alice chooses the appropriate unused page from the pad. The way to do this is normally arranged for in advance, as for instance 'use the 12th sheet on 1 May', or 'use the next available sheet for the next message'.

The material on the selected sheet is the key for this message. Each letter from the pad will be combined in a predetermined way with one letter of the message. (It is common, but not required, to assign each letter a numerical value, e.g., "A" is 0, "B" is 1, and so on.)

In this example, the technique is to combine the key and the message using modular addition. The numerical values of corresponding message and key letters are *added* together, modulo 26. So, if key material begins with "XMCKL" and the message is "HELLO", then the coding would be done as follows:

H E L L O message

7 (H) 4 (E) 11 (L) 11 (L) 14 (O) message

+ 23 (X) 12 (M) 2 (C) 10 (K) 11 (L) key

= 30 16 13 21 25 message + key

= 4 (E) 16 (Q) 13 (N) 21 (V) 25 (Z) message + key (mod 26)

E Q N V Z → ciphertext

If a number is larger than 26, then the remainder, after *subtraction* of 26, is taken [as the result]. This simply means that if the computations "go past" Z, the sequence starts again at A.

The ciphertext to be sent to Bob is thus "EQNVZ". Bob uses the matching key page and the same process, but in reverse, to obtain the plaintext. Here the key is *subtracted* from the ciphertext, again using modular arithmetic:

E Q N V Z ciphertext

4 (E) 16 (Q) 13 (N) 21 (V) 25 (Z) ciphertext

- 23 (X) 12 (M) 2 (C) 10 (K) 11 (L) key

= -19 4 11 11 14 ciphertext – key

= 7 (H) 4 (E) 11 (L) 11 (L) 14 (O) ciphertext – key (mod 26)

H E L L O → message

Similar to the above, if a number is negative then 26 is *added* to make the number zero or higher.

Thus Bob recovers Alice's plaintext, the message "HELLO". Both Alice and Bob destroy the key sheet immediately after use, thus preventing reuse and an attack against the cipher.”

Your program will encrypt and decrypt plaintext into ciphertext, using a key, in exactly the same fashion as above, except it will be using modulo 27 operations: your 27 characters are the 26 capital letters, and the space character ( ). All 27 characters will be encrypted and decrypted as above.

To do this, you will be creating five small programs in C. Two of these will function as “daemons”, and will be accessed using network sockets. Two will use the daemons to perform work, and the last is a standalone utility.

Here are the specifications of the five programs:

**otp\_enc\_d:** This program will run in the background as a daemon. Its function is to perform the actual encoding, as described above in the Wikipedia quote. This program will listen on a particular port, assigned when it is first run, and receives plaintext and a key via that port when a connection to it is made. It will then write back the ciphertext to the process that it is connected to via the same port. Note that the key passed in must be at least as big as the plaintext.

When **otp\_enc\_d** makes a connection with **otp\_enc (described below),** it must fork off a separate process immediately and be available to receive more connections. Your version must support up to five concurrent socket connections. In the forked off process, the actual encryption will take place, and the ciphertext will be written back.

You may either create a new process every time a connection is made, or set up a pool of five process to handle your encryption tasks. Your system must be able to do five separate encryptions at once, using either method you choose. Use this syntax for **otp\_enc\_d:**

**otp\_enc\_d**

***listening\_port***

***listening\_port*** is the port that **otp\_enc\_d** should listen on. You will always start **otp\_enc\_d** in the background, as follows (the port 57171 is just an example – yours should be able to use any port):

**% otp\_enc\_d 57171 &**

In all error situations, your program must output errors as appropriate (see grading script below for details), but should not crash or otherwise exit, unless the errors happen when the program is starting up. That is, once running, if given bad input **otp\_enc\_d** should recognize the bad input, report an error to the screen, and continue to run.

Your **otp\_enc\_d** should be killable with the -KILL signal, as normal: you may not have to do anything for your program to have this ability.

**otp\_enc:** This program connects to **otp\_enc\_d** and asks it to perform a one-time pad style encryption as detailed above. By itself, **otp\_enc** doesn’t do the encryption. Its syntax is as follows:

**otp\_enc**

***plaintext key port***

In this syntax, ***plaintext*** is the name of a file in the current directory that contains the text you wish to encrypt. Similarly, ***key*** contains the encryption key you wish to use to encrypt the text. Finally, ***port*** is the port that **otp\_enc** should attempt to connect to **otp\_enc\_d** on.

When **otp\_enc** receives the ciphertext, it should output it to *stdout*. Thus, **otp\_enc** can be launched in any of the following methods, and should send its output appropriately:

**% otp\_enc myplaintext mykey 57171**

**% otp\_enc myplaintext mykey 57171 > myciphertext**

**% otp\_enc myplaintext mykey 57171 > myciphertext &**

If **otp\_enc** receives key or plaintext files with bad characters in them, or the key file is shorter than the plaintext, it should exit with an error and set the exit value to 1. If **otp\_enc** cannot find the port given, it should report this error, and the bad port, and set the exit value to 2. Otherwise, on successfully running, **otp\_enc** should set the exit value to 0. **otp\_enc** should NOT be able to connect to **otp\_dec\_d,** **(described below),** even if it tries to connect on the correct port - you'll need to have the programs reject each other.

**otp\_dec\_d:** This program performs exactly like **otp\_enc\_d,** in syntax and usage. In this case, however, **otp\_dec\_d** will decrypt ciphertext it is given, using the passed-in ciphertext and key. Thus, it returns plaintext again to **otp\_dec.**

**otp\_dec:** Similarly with **opt\_enc**, this program will connect to **otp\_dec\_d** and will ask it to decrypt ciphertext using a passed-in ciphertext and key. It will use the same syntax and usage as **otp\_enc**, and must be runnable in the same three ways. **otp\_dec** should NOT be able to connect to **otp\_enc\_d**, even if it tries to connect on the correct port - you'll need to have the programs reject each other.

**keygen:** This program creates a key file of specified length. The characters in the file generated will be any of the 27 allowed characters, generated using the standard UNIX randomization methods. Do not create spaces every five characters, as has been historically done. Note that you specifically do not have to do any fancy random number generation: we’re not looking for cryptographically secure random number generation!

The syntax for keygen is as follows:

**keygen**

***keylength***

Where ***keylength*** is the length of the key file in characters. **keygen** outputs to stdout. Here is an example run, which redirects stdout to a key file of 256 characters called “mykey”:

**% keygen 256 > mykey**

## Files and Scripts

You are provided with 5 plaintext files to use. You are also provided with a [grading script](https://canvas.oregonstate.edu/courses/1516852/files/62576475/download?wrap=1) ("p4gradingscript") that you can run to test your software. If it passes the tests in the script, and has sufficient documentation, it will receive full points (see below).

Finally, you will be required to write a Makefile that compiles all five of your programs and cleans everything with the *make clean* command. The compilation must create all five programs, in the same directory as the makefile, for immediate use by the grading script.

## Example

Here is an example of usage, if you were testing your code from the command line:

% cat plaintext1

THE RED GOOSE FLIES AT MIDNIGHT

% otp\_enc\_d 57171 &

% otp\_dec\_d 57172 &

% keygen 10 > myshortkey

% otp\_enc plaintext1 myshortkey 57171 > ciphertext1

Error: key ‘myshortkey’ is too short

% echo $?

1

% keygen 1024 > mykey

% otp\_enc plaintext1 mykey 57171 > ciphertext1

% cat ciphertext1

GU WIRGEWOMGRIFOENBYIWUG T WOFL

% keygen 1024 > mykey2

% otp\_dec ciphertext1 mykey 57172 > plaintext1\_a

% otp\_dec ciphertext1 mykey2 57172 > plaintext1\_b

% cat plaintext1\_a

THE RED GOOSE FLIES AT MIDNIGHT

% cat plaintext1\_b

WVIOWBTUEIOBC FVTROIROUXA JBWE

% cmp plaintext1 plaintext1\_a

% echo $?

0

% cmp plaintext1 plaintext1\_b

plaintext1 plaintext1\_b differ: byte 1, line 1

% echo $?

1

% otp\_enc plaintext5 mykey 57171

otp\_enc\_d error: input contains bad characters

% otp\_enc plaintext3 mykey 57172

Error: could not contact otp\_enc\_d on port 57172

% echo $?

2

%

## About newlines

You are only supposed to accept the 26 letters of alphabet and the "space" character as valid for encrypting/decrypting. However, all of the plaintext input files end with a newline character. Text files need to end in a newline character for various reasons.

When one of your programs reads in an input file, strip off the newline. Then encrypt and decrypt the text string, again with no newline character. When you send the result to stdout, or save results into a file, tack a newline to the end, or your length will be off in the grading script. Note that the newline character affects the length of files as reported by the wc command! Try it!

## About Reusing Sockets

In the p4gradingscript, you can select which ports to use: I recommend ports in the 50000+ range. However, UNIX doesn't immediately let go of the ports you use after your program finishes! To allow your program to continue to use the same port, read this:

<https://linux.die.net/man/3/setsockopt>

...and then play around with this command:

**setsockopt(sock\_fd, SOL\_SOCKET, SO\_REUSEADDR, &yes, sizeof(int));**

## What to submit

Please submit a single zip file including the following files:

1. All of your program code
2. The Makefile
3. All five plaintext# files, numbered 1 through 5
4. A copy of the grading script named “p4gradingscript”

# Hints

I HIGHY recommend that you develop this program directly on the Cal Poly server. Doing so will prevent you from having problems transferring the program back and forth, which can cause compatibility issues.

If you do see ^M characters all over your files, try this command:

% dos2unix bustedFile

# Grading

The graders will run the Makefile, and will then run the “p4gradingscript”. They will make a reasonable effort to make your code compile, but if it doesn’t compile, you’ll receive a zero on this assignment.

150 points are available in the grading script, while the final 10 points will be based on your style, readability, and commenting. Comment well, often, and verbosely: we want to see that you are telling us WHY you are doing things, in addition to telling us WHAT you are doing.