Project 2

In this project you will synchronize multiple threads. The project must be written in C, and must use the Pthreads library. As you will see, some aspects of the project are not specified; it will be your responsibility to take whatever steps you feel are appropriate to complete the task.

Overview

The purpose of this project is to implement a multi-threaded text file encryptor. Conceptually, the function of the program is simple: to read characters from the input file, encrypt the letters, and write the encrypted characters to the output file. Also, the encryption program counts the number of occurrences of each letter in the input and output files. Any Com S 228 student should be able to implement the single-threaded version of this code (you may want to do this for comparing output). The Com S 352 version will have multiple threads. Specifically, your code will consist of six threads: the main thread, a reader thread, a writer thread, an encryption thread, and two counter threads. Since multiple threads will be accessing the same data structures, you will need to synchronize the threads to avoid race conditions (therein lies the difficulty). Synchronization **must** be done via Pthreads constructs such as semaphores, condition variables, and/or mutexes (i.e., your code must not contain spinlocks). For more information, consult the man pages for the following.

- `pthread_create`
- `pthread_join`
- `pthread_mutex_init`
- `pthread_mutex_lock`
- `pthread_mutex_unlock`
- `pthread_mutex_destroy`
- `pthread_cond_init`
- `pthread_cond_wait`
- `pthread_cond_signal`
- `pthread_cond_destroy`
- `sem_init`
- `sem_wait`
- `sem_post`
- `sem_destroy`

Required Features

5 points: makefile

You get 5 points for simply using a makefile. Name your source files whatever you like. Please name your executable **encrypt**. Be sure that "make" will build your executable.

10 points: Documentation

If you have more than one source file, then you must submit a **Readme** file containing a brief explanation of the functionality of each source file. Each source file must also be well-documented. There should be a paragraph or so at the beginning of each source file describing the code; functions and data structures should also be explained. Basically, another programmer should be able to understand your code from the comments.

10 points: Main thread

The main thread does the following.

1. Obtain the input and output files from the command line. If the number of command line arguments is incorrect, exit after displaying a message about correct usage.
2. Prompt the user for the buffer size *N*.
3. Initialize shared variables. This includes allocating appropriate data structures for the input and output buffers, and appropriate data structures to count the number of occurrences of each letter in the input and output files. You may use any data structure capable of holding exactly *N* characters for the input and output buffers.
4. Create the other threads.
5. Wait for all threads to complete.
6. Display the number of occurrences of each letter in the input and output files.

15 points: reader thread

The reader thread is responsible for reading from the input file (specified by the first argument on the command line) one character at a time, and placing the characters in the input buffer. Each buffer item corresponds to a character. Note that the reader thread may need to block until other threads have consumed data from the input buffer. Specifically, a character in the input buffer cannot be overwritten until the encryptor thread and the input counter thread have processed the character. The reader continues until the end of the input file is reached.

15 points: writer thread

The writer thread is responsible for writing the encrypted characters in the output buffer to the output file (specified by the second argument on the command line). Note that the writer may need to block until an encrypted character is available in the buffer. The writer continues until it has written the last encrypted character. (Hint: the special character EOF will never appear in the middle of an input file.)

15 points: encryption thread

The encryption thread consumes one character at a time from the input buffer, encrypts it, and places it in the output buffer. The encryption algorithm to use is given below. Of course, the encryption thread may need to wait for an item to become available in the input buffer, and for a slot to become available in the output buffer. Note that a character in the output buffer cannot be

overwritten until the writer thread and the output counter thread have processed the character. The encryption thread continues until all characters of the input file have been encrypted.

The encryption algorithm is fairly simple (and is very easy to crack). Only alphabetic characters (i.e., 'A'..'Z' and 'a'..'z') are changed; all other characters are simply copied to the output file. The algorithm either increments, decrements, or does not touch alphabetic characters. The encryption algorithm is as follows.

1. s = 1;
2. Get next character c.
3. if c is not a letter then goto (7).
4. if (s==1) then increase c with wraparound (e.g., 'A' becomes 'B', 'c' becomes 'd', 'Z' becomes 'A', 'z' becomes 'a'), set s=-1, and goto (7).
5. if (s==-1) then decrease c with wraparound (e.g., 'B' becomes 'A', 'd' becomes 'c', 'A' becomes 'Z', 'a' becomes 'z'), set s=0, and goto (7).
6. if (s==0), then do not change c, and set s=1.
7. Encrypted character is c.
8. If c!=EOF then goto (2).

Examples:

```
Original file:    AAAAAAAAAAAAAAAAAA
Encrypted file:   BZABZABZABZABZABZA

Original file:    a aa aaa (aaaa) aaaaa.
Encrypted file:   b za bza (bzab) zabza.

Original file:    This is (probably)
                  easy 2 crack.
Encrypted file:   Ugit hs (qqoczbmx)
                  ebry 2 dqadj.
```

15 points: input counter thread

The input counter thread simply counts occurrences of each letter in the input file by looking at each character in the input buffer. Of course, the input counter thread will need to block if no characters are available in the input buffer.

15 points: output counter thread

The output counter thread simply counts occurrences of each letter in the output file by looking at each character in the output buffer. Of course, the output counter thread will need to block if no characters are available in the output buffer.

Synchronization Requirement

Your program should achieve maximum concurrency. That is, you should allow different threads to operate on different buffer slots concurrently. For example, when the reader thread is placing a character in slot 5 of the input buffer, the encryption thread may process the character in slot 3 of the input buffer and the input counter thread may process the character in slot 2 of the input buffer.

Your program **MUST NOT** do the following: First let the reader thread put *N* characters in the input buffer, and then let the input counter thread and the encryption thread consume all the characters in the input buffer. This does not provide maximum concurrency.

Example

Here are some example input files and their corresponding output files.

> encrypt infile1 outfile1
Enter buffer size: 5
Input file contains
A: 55
Output file contains
A: 18
B: 19
Z: 18

> encrypt infile2 outfile2
Enter buffer size: 1000
Input file contains
A: 14
B: 3
C: 5
D: 3
E: 23
F: 5
G: 1
H: 8
I: 13
K: 1
L: 6
M: 2
N: 10
O: 11
P: 7

R: 10
S: 12
T: 22
U: 6
V: 1
Y: 2
Z: 1
Output file contains
A: 6
B: 8
C: 4
D: 9
E: 12
F: 11
G: 5
H: 6
I: 7
J: 5
K: 1
L: 4
M: 5
N: 9
O: 10
P: 6
Q: 5
R: 4
S: 18
T: 10
U: 12
V: 3
X: 1
Y: 2
Z: 3

Submitting Your Project

You will submit your project on Canvas. Your program must compile and run without errors on pyrite.cs.iastate.edu.

Put all your source files (including the makefile and the README file) in a folder. Then use command `zip -r <your ISU Net-ID> <src_folder>` to create a .zip file. For example, if your Net-ID is ksmith and project2 is the name of the folder that contains all your source files, then you will type `zip -r ksmith project2` to create a file named ksmith.zip and then submit this file.