CS 385 Operating Systems Fall 2013  
Homework Assignment 3  
Searching for Order Among Chaos Using Threads

Due: Friday November 15. Electronic copy due at 2:30, optional paper copy may be delivered to the TA.

**Overall Assignment:**

For this assignment you are to write a program using PTHREADS to analyze a data file looking for patterns. Each thread will analyze a different section of the file in parallel, recording their findings in a common queue of results, which a separate thread will read from and print the latest results to the screen as they arrive. The threads will use a branch-and-bound technique, discontinuing analysis as soon as it can be determined that a result will not be found that is any better than what has already been found (by any thread).

**Command Line Arguments:**

The command line for this program will be as follows:  
 orderSearcher inputDataFile nThreads[…]  
 - inputDataFile is a data file to be read and analyzed. Some potential files will be provided.   
 - nThreads is the number of threads to use to search for patterns in the data, not counting any support  
 threads such as the thread to print out results as they are found. If this value is specified as zero, then   
 the program should do the best job it can of finding patterns in the data without using threads.   
 - you may have additional paramaters that follow the above paramaters, depending on how you choose   
 to handle some of the issues discussed below.

**Input Data:**

The data in the file will be read in as a series of unsigned bytes (0-255), in binary.

**Evaluation Criteria:**

The whole point of the assignment is to use threads to find blocks of data that are as “ordered” as possible, based on a number of different criteria. Note that the idea is to use **all** of the different criteria you can, not to just pick one or two off the list. There are a couple of different ways the criteria can be combined, as discussed below. Here are some potential criteria for determining order:

* The range, defined as the difference between the largest and smallest value in the data set
* Maximum absolute value of change from one value to the next: max\_i|X\_i – X+(i+1)|
* The sum of (absolute value of (change from one value to the next )):∑i|X\_i – X\_(i+1)
* Best regression coefficient fit to a straing line, Nth order polynomial, sine curve, FFT?
* Periodicity? (Cyclic repeating patterns.)

**Branch and Bound:**

The concept of branch and bound is to split the set of all possible solutions into smaller subsets (branching), and then eliminating entire subsets when it becomes clear that no solution within that subset can ever be better than a pervious solution already found in any subset. For this application what that means is that all of the threads will share knowledge of the “best” solution found so far by any thread, and will cease analyzing any new candidates as soon as it is known that they cannot improve upon the best found so far.

For example, suppose we are considering the criteria ∑|X\_i – X\_(i+1)| and we know that the best solution found so far using this criteria has a sum of 100. As soon as the sum we are calculating exceeds 100, then we can stop examining this candidate, at least with respect to this criteria.

For a more challenging example, suppose we are considering the standard deviation sqrt(∑(x\_i – x)^2/N), and the best solution found so far is 10.0. That means that in order for this to be an improvement, the quantity under the root would have to be a smaller than 100, and the sum term would have to be smaller than 100\*N.

**Issues to be resolved:**

* Examine effect of different scope c