

CS 343 Fall 2025 – Assignment 3

Instructor: Peter Buhr

Due Date: Monday, February 24, 2025 at 22:00

Late Date: Wednesday, February 26, 2025 at 22:00

January 6, 2025

This assignment examines synchronization and mutual exclusion, and introduces locks in μ C++. Use it to become familiar with these new facilities, and ensure you use these concepts in your assignment solution. (You may freely use the code from these [example programs](#).) (Tasks may *not* have public members except for constructors and/or destructors.)

1. Given the C++ program in Figure 1, compare stack versus heap allocation in a concurrent program.
 - (a) Compare the versions of the program and different numbers of tasks with respect to performance by doing the following:
 - Run the program after compiling with preprocessor variables STACK, DARRAY, VECTOR1, and VECTOR2. Use compiler flags `-O2 -multi -nodebug`.
 - Time the executions using the time command:

```
$ /usr/bin/time -f "%Uu %Ss %Er %Mkb" ./a.out 2 200000000  
3.21u 0.02s 0:03.32r 4228kb
```

(Output from time differs depending on the shell, so use the system time command.) Compare the *user* (3.21u) and *real* (0:3.32) time among runs, which is the CPU time consumed solely by the execution of user code (versus system) and the total time from the start to the end of the program.
 - Use the second command-line argument (as necessary) to adjust the real time into the range 1 to 100 seconds. (Timing results below 1 second are inaccurate.) Use the same command-line values for all experiments, if possible; otherwise, increase/decrease the arguments as necessary and scale the difference in the answer.
 - Run each of the 4 versions (STACK, DARRAY, VECTOR1, and VECTOR2) with the number of tasks set to 1, 2, and 4.
 - Include all 12 timing results to validate the experiments and the number of calls to malloc.
 - (b) Explain the performance and allocation difference (larger/smaller/by how much) with respect to scaling the number of tasks for each version.
 - (c) Very briefly (2-4 sentences) speculate on the performance scaling among the versions.
2. (a) Quick sort is one of the best sorting algorithms in terms of execution speed on randomly arranged data. It also lends itself easily to concurrent execution by partitioning the data into those greater than a pivot and those less than a pivot so each partition can be sorted independently and concurrently by another thread (divide-and-conquer/embarassingly concurrent).

Write a concurrent quick-sort function with the following interface:

```
template<typename T>  
void quicksort( T values[], unsigned int low, unsigned int high, unsigned int depth );
```

that sorts an array of non-unique values into ascending order. Type T should only need operators `<`, `=`, `>>` and `<<`, and public default and copy constructors. Choose the pivot as follows:

```
pivot = array[low + ( high - low ) / 2];
```

For information-hiding reasons, all sort code must appear within quicksort, except for any required include files.

Implement the concurrent quicksort using:

- i. COBEGIN/COEND statements

```

#include <iostream>
#include <vector>
#include <memory>                                // unique_ptr
using namespace std;

intmax_t tasks = 1, times = 200'000'000, asize = 10;    // default values

_Task Worker {
    void main() {
        for ( int t = 0; t < times; t += 1 ) {
            #if defined( STACK )
                volatile int arr[asize] __attribute__(( unused )); // prevent unused warning
                for ( int i = 0; i < asize; i += 1 ) arr[i] = i;
            #elif defined( DARRAY )
                unique_ptr<volatile int []> arr( new volatile int[asize] );
                for ( int i = 0; i < asize; i += 1 ) arr[i] = i;
            #elif defined( VECTOR1 )
                vector<int> arr( asize );
                for ( int i = 0; i < asize; i += 1 ) arr.at(i) = i;
                asm volatile( "" :: "r"(arr.data()):"memory" ); // prevent eliding code
            #elif defined( VECTOR2 )
                vector<int> arr;
                for ( int i = 0; i < asize; i += 1 ) arr.push_back(i);
                asm volatile( "" :: "r"(arr.data()):"memory" ); // prevent eliding code
            #else
                #error unknown data structure
            #endif
        } // for
    } // Worker::main
}; // Worker

int main( int argc, char * argv[] ) {
    char * nosummary = getenv( "NOSUMMARY" );    // print heap statistics ?

    struct cmd_error {};
    try {                                         // process command-line arguments
        switch ( argc ) {
            case 4:
                asize = convert( argv[3] ); if ( asize <= 0 ) throw cmd_error();
            case 3:
                times = convert( argv[2] ); if ( times <= 0 ) throw cmd_error();
            case 2:
                tasks = convert( argv[1] ); if ( tasks <= 0 ) throw cmd_error();
        } // switch
    } catch( ... ) {
        cout << "Usage: " << argv[0] << " [ tasks (> 0) [ times (> 0) "
            " [ array size (> 0) ] ] " << endl;
        exit( 1 );
    } // try
    uProcessor p[tasks - 1];                    // add CPUs (start with one)
    {
        Worker workers[tasks];                 // add threads
    }
    if ( ! nosummary ) { malloc_stats(); }
} // main

```

Figure 1: Stack versus Dynamic Allocation

- ii. **_Task** type: Do not create tasks by calls to **new**, i.e., no dynamic allocation is necessary for quicksort tasks.
- iii. **_Actor** type: All information about the array must be passed to the actor in an initial message not via the actor's constructor.

A naïve concurrent quicksort partitions the data values as normal, but instead of recursively invoking quicksort on each partition, a new implicit or explicit concurrent quicksort object is created to handle each partition. (For this discussion, assume no other sorting algorithm is used for small partitions.) However, this approach can create a large number of concurrent objects, approximately $2 \times N$, where N is the number of data values. This large number of concurrent objects can slow down the sort.

The number of concurrent objects can be reduced to approximately N by limiting the tree depth of the divide-and-conquer where concurrent sort objects are created (see details below). Basically, the depth argument is decremented on each recursive call and concurrent sort objects are only created while this argument is greater than zero; after which sequential recursive-calls are used to sort each partition. For explicit threading objects, a further reduction is possible by only creating one new concurrent sort task for the left partition and recursively sorting the right partition using the current quicksort task.

The executable program is named `quicksort` and has the following shell interface:

```
quicksort [ unsorted-file | 'd' [ sorted-file | 'd' [ depth (>= 0) ] ] ]
quicksort -t size (>= 0) [ depth (>= 0) ]
```

(Square brackets indicate optional command line parameters, and do not appear on the actual command line.) If no unsorted file is specified, use standard input. If no sorted file is specified, use standard output. If no depth is specified, use 0. **The input-value type is provided externally by the preprocessor variable STYPE (see the Makefile).**

The program has two modes depending on the command option `-t` (i.e., sort or time):

sort mode: read the number of input values, read the input values, sort using 1 processor (which is the default), and output the sorted values. Input and output is specified as follows:

- The unsorted input contains lists of unsorted values. Each list starts with the number of values in that list. For example, the input file:

```
8 25 6 8 -5 99 100 101 7
3 1 -3 5
0
10 9 8 7 6 5 4 3 2 1 0
61 60 59 58 57 56 55 54 53 52 51 50 49 48 47 46 45 44 43 42 41 40 39 38 37
36 35 34 33 32 31 30 29 28 27 26 25 24 23 22 21 20 19 18 17 16 15 14 13 12
11 10 9 8 7 6 5 4 3 2 1 0
```

contains 5 lists with 8, 3, 0, 10, and 61 values in each list. (The line breaks are for readability only; values can be separated by any white-space character and appear across any number of lines.) Since the number of data values can be (very) large, dynamically allocate the array to hold the values, otherwise the array can exceed the stack size of the program main.

Assume the first number of every line in the input file is always present and correctly specifies the number of following values. Assume all following values are correctly formed so no error checking is required on the input data.

- The sorted output is the original unsorted input list followed by the sorted list, as in:

```
25 6 8 -5 99 100 101 7
-5 6 7 8 25 99 100 101
```

```
1 -3 5
-3 1 5
```

blank line from list of length 0 (this line not actually printed)
blank line from list of length 0 (this line not actually printed)

```
9 8 7 6 5 4 3 2 1 0
0 1 2 3 4 5 6 7 8 9
```

```

60 59 58 57 56 55 54 53 52 51 50 49 48 47 46 45 44 43 42 41 40 39
38 37 36 35 34 33 32 31 30 29 28 27 26 25 24 23 22 21 20 19 18 17
16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43
44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

```

End each set of output with a blank line, and start a newline with 2 spaces after printing 22 values from a set of values.

time mode: dimension an integer array to size, initialize the array to values size..1 (descending order), randomize the values using:

```

unsigned int times = sqrt( size );
for ( unsigned int counter = 0; counter < times; counter += 1 ) {
    swap( values[0], values[prng( size ) ] );
} // for

```

sort using $2^{\text{depth}} - 1$ processors, and print no values (used for timing experiments).

Parameter depth is a non-negative number (≥ 0). The default value if unspecified is 0.

Create the additional processors by placing the following declaration in the same block as the quicksort call:

```
uProcessor p[ (1 << depth) - 1 ] __attribute__(( unused )); // 2^depth-1 kernel threads
```

Keep the number of processors small as the [undergraduate machines](#) have a limited number of cores and other students are using the machines.

Print an appropriate error message and terminate the program if unable to open the given files. Check command arguments size and depth for correct form (integer) and range; print an appropriate usage message and terminate the program if a value is invalid.

(b) i. Compare the speedup of the quicksort algorithm with respect to performance by doing the following:

- Bracket the call to the sort in the program main with the following to measure the real time for the sort *within* the program.

```

uTime start = uClock::currTime();
quicksort( ... );
cout << "Sort time " << uClock::currTime() - start << " sec." << endl;

```

- Compile with makefile option `OPT="-O2 -multi -nodebug"`.

- Time the execution using the time command:

```

$ /usr/bin/time -f "%Uu %Ss %E %Mkb" quicksort -t 300000000 0
14.13u 0.59s 0:14.68 1958468kb

```

(Output from time differs depending on the shell, so use the system time command.) Compare the *user* (14.13u) and *real* (0:14.68) time among runs, which is the CPU time consumed solely by the execution of user code (versus system) and the total time from the start to the end of the program.

- Adjust the array size to get the real time in the range 10 to 20 seconds for depth 0. Use the same array size for all experiments.
- For each of CBEGIN, ACTOR, and TASK, run 5 experiments varying the value of depth from 0 1 2 3 4. Include all 15 timing results to validate your experiments.

ii. Explain the performance difference (larger/smaller/by how much) with respect to scaling when using different numbers of processors to achieve parallelism.

iii. Very briefly speculate on the performance difference between the program sort time and the real time from the time command, i.e., why is it the same or different.

3. (a) Implement a generalized FIFO bounded-buffer for a producer/consumer problem with the following interface (you may add only a public destructor and private members):

```
template<typename T> class BoundedBuffer {
public:
    _Exception Poison {};
    BoundedBuffer( const unsigned int size = 10 );
    unsigned long int blocks();
    void poison();
    void insert( T elem );
    T remove() __attribute__(( warn_unused_result ));
};
```

which creates a bounded buffer of size `size`, and supports multiple producers and consumers. Member `blocks` returns the current number of calls to wait in both `insert` and `remove`. Member `poison` marks the buffer as *poisoned* and is called *after* all producers have finished. Once the buffer is poisoned and all values have been consumed, any waiting consumers are unblocked and receive a `Poison` exception; similarly for any arriving consumers. Poisoning is how the consumers know when to terminate. You may *only* use `uCondLock` and `uOwnerLock` to implement the necessary synchronization and mutual exclusion needed by the bounded buffer.

Implement the `BoundedBuffer` in the following ways:

- i. Use busy waiting, when waiting on a full or empty buffer. In this approach, tasks that have been signalled to access empty or full entries may find them taken by new tasks that barged into the buffer. This implementation uses one owner and two condition locks, where the waiting producer and consumer tasks block on the separate condition locks. (If necessary, you may add more locks.) The reason there is barging in this solution is that `uCondLock::wait` re-acquires its argument owner-lock before returning. Now once the owner-lock is released by a task exiting `insert` or `remove`, there is a race to acquire the lock by a new task calling `insert/remove` and by a signalled task. If the calling task wins the race, it barges ahead of any signalled task. So the state of the buffer at the time of the signal is not the same as the time the signalled task re-acquires the argument owner-lock, because the barging task changes the buffer. Hence, the signalled task may have to wait again (looping), and there is no guarantee of eventual progress (long-term starvation).
- ii. Use *no* busy waiting when waiting for buffer entries to become empty or full. In this approach, use *barging avoidance* so a barging task cannot take empty or full buffer entries if another task has been unblocked to access these entries. This implementation uses one owner and two condition locks, where the waiting producer and consumer tasks block on the separate condition locks, but there is (*no looping*). (If necessary, you may add more locks.) Hint, one way to prevent overtaking by bargers is to use a flag variable to indicate when signalling is occurring; entering tasks test the flag to know if they are barging and wait on an appropriate condition-lock. When signalling is finished, an appropriate task is unblocked. (Hint: `uCondLock::signal` returns true if a task is unblocked and false otherwise.)

Before inserting or removing an item to/from the buffer, perform an `assert` that checks if the buffer is not full or not empty, respectively. Both buffer implementations are defined in a single `.h` file separated in the following way:

```
#ifdef BUSY                                // busy waiting implementation
// implementation
#endif // BUSY

#ifdef NOBUSY                              // no busy waiting implementation
// implementation
#endif // NOBUSY
```

The kind of buffer is specified externally by a preprocessor variable of `BUSY` or `NOBUSY`.

Insert the following macros into the `NOBUSY` solution to verify correctness. (These macros must appear in your submitted solution as they are used to test this program for correctness.)

BCHECK_DECL is placed once in the buffer class as a private member, and contains variables and members needed to implement the barging check.

PROD_ENTER is placed immediately *after* mutual exclusion is acquired in `insert`.

INSERT_DONE is placed immediately *after* a value is inserted into the buffer in `insert`.

```

#ifdef NOBUSY // no busy waiting implementation
#include "BargingCheck.h"
template<typename T> class BoundedBuffer {
    ... // regular declarations
    BCHECK_DECL;
public:
    void insert( T elem ) {
        // acquire mutual exclusion
        PROD_ENTER;
        ...
        // buffer insert
        INSERT_DONE;
        ...
        CONS_SIGNAL( cond-lock );
        cond-lock.signal(); // signal consumer task
        ...
        PROD_SIGNAL( cond-lock ); // if necessary
        cond-lock.signal(); // signal producer task
        ...
    }
    T remove() __attribute__(( warn_unused_result )) {
        // acquire mutual exclusion
        CONS_ENTER;
        ...
        // buffer remove
        REMOVE_DONE;
        ...
        PROD_SIGNAL( cond-lock );
        cond-lock.signal(); // signal producer task
        ...
        CONS_SIGNAL( cond-lock ); // if necessary
        cond-lock.signal(); // signal consumer task
        ...
    }
};
#endif // NOBUSY

```

Figure 2: Barging Check Macros

CONS_SIGNAL(cond-lock) is placed immediately *before* signalling a condition lock to unblock a consumer task. If a condition lock has both producers and consumers blocked on it, it is NOT preceded by this macro.

CONS_ENTER is placed immediately *after* mutual exclusion is acquired in remove.

REMOVE_DONE is placed immediately *after* a value is removed from the buffer in remove.

PROD_SIGNAL(cond-lock) is placed immediately *before* signalling a condition lock to unblock a producer task. If a condition lock has both producers and consumers blocked on it, it is NOT preceded by this macro.

Figure 2 shows the macro placement in a buffer implementation, and defining preprocessor variable BARGINGCHECK triggers barging testing (see Makefile). If barging is detected, a message is printed and the program continues, possibly printing more barging messages. Note, when the buffer is poisoned and waiting consumers are signalled, do not include CONS_SIGNAL(cond-lock) before any signalling caused by poisoning. The barging macros are only equipped to detect barging during normal program operation and not during termination due to poisoning. To inspect the program with gdb when barging is detected, set BARGINGCHECK=0 to abort the program.

Test the bounded buffer with a number of producers and consumers. The producer interface is (you may add only a public destructor and private members):

- Time the executions using the time command:


```
$ /usr/bin/time -f "%Uu %Ss %Er %Mkb" ./a.out
3.21u 0.02s 0:03.32r 33640kb
```

 (Output from time differs depending on the shell, so use the system time command.) Compare the *user* time (3.21u) only, which is the CPU time consumed solely by the execution of user code (versus system and real time).
 - Use the program command-line arguments 55 50 20000 20 10 1 and adjust the produce amount (if necessary) to get program execution into the range 1 to 100 seconds. (Timing results below 1 second are inaccurate.) Use the same command-line values for all experiments, if possible; otherwise, increase/decrease the arguments as necessary and scale the difference in the answer.
 - Run both the experiments again after recompiling the programs with compiler optimization turned on (i.e., compiler flag `-O2`).
 - Include 4 timing results to validate the experiments.
- ii. Explain the performance difference (larger/smaller/by how much) between uniprocessor busy and nobusy waiting execution, without and with optimization.
 - iii. Compare the busy and non-busy waiting versions of the program with respect to *multiprocessor* performance by repeating the above experiment with 4 kernel threads.
 - Include 4 timing results to validate the experiments.
 - iv. Explain the performance difference (larger/smaller/by how much) between multiprocessor busy and nobusy waiting execution, without and with optimization.
 - v. Speculate as to the reason for the performance difference between busy and non-busy execution. Use the total number of times a producer/consumer blocks in the bounded buffer to help understand differences.
 - vi. Speculate as to the reason for the performance difference between uniprocessor and multiprocessor execution.

Submission Guidelines

Follow these guidelines carefully. Review the [Assignment Guidelines](#) and [C++ Coding Guidelines](#) *before* starting each assignment. **Each text or test-document file, e.g., `*.{txt,testdoc}` file, must be ASCII text and not exceed 500 lines in length, using the command `fold -w120 *.testdoc | wc -l`. Programs should be divided into separate compilation units, i.e., `*.{h,cc,C,cpp}` files, where applicable. Use the [submit](#) command to electronically copy the following files to the course account.**

1. `q1*.txt` – contains the information required by question 1, p. 1.
2. `q2quicksort.h`, `q2*.{h,cc,C,cpp}` – code for question 2a, p. 1. **Program documentation must be present in your submitted code. No user, system or test documentation is to be submitted for this question. Output for this question is checked via a marking program, so it must match exactly with the given program.**
3. `q2*.txt` – contains the information required by questions 2b, p. 4.
4. `BargingCheck.h` – barging checker (provided)
5. `q3buffer.h`, `q3*.{h,cc,C,cpp}` – code for question 3a, p. 4. **Program documentation must be present in your submitted code. No user, system or test documentation is to be submitted for this question. Output for this question is checked via a marking program, so it must match exactly with the given program.**
6. `q3*.txt` – contains the information required by questions 3b.
7. Modify the following Makefile to compile the programs for question 2a, p. 1 and 3a, p. 4 by inserting the object-file names matching your source-file names.


```

OPT := # -O3
STYPE := int
SIMPL := CBEGIN
BIMPL := NOBUSY
BCHECK := NOBARGINGCHECK

CXX = u++                                # compiler
CXXFLAGS = -g -Wall -Wextra -multi ${OPT} -MMD -DSTYPE="${STYPE}" -D"${SIMPL}" \
           -D"${BIMPL}" -D"${BCHECK}"      # compiler flags
MAKEFILE_NAME = ${firstword ${MAKEFILE_LIST}} # makefile name

OBJECTS1 = # object files forming 1st executable with prefix "q2"
EXEC1 = quicksort                        # 1st executable name

OBJECTS2 = # object files forming 2nd executable with prefix "q3"
EXEC2 = buffer                          # 2nd executable name

OBJECTS = ${OBJECTS1} ${OBJECTS2}       # all object files
DEPENDS = ${OBJECTS:.o=.d}              # substitute ".o" with ".d"
EXECS = ${EXEC1} ${EXEC2}               # all executables

#####

.PHONY : all clean
.ONESHELL :

all : ${EXECS}                          # build all executables

-include MergelImpl

# same implementation concurrency/type as last time ?
ifeq ($(shell if [ "${IMPLSTYPE}" = "${STYPE}" -a "${IMPLSIMPL}" = "${SIMPL}" ]; \
then echo true ; fi },true)
${EXEC1} : ${OBJECTS1}
        ${CXX} ${CXXFLAGS} $^ -o $@
else
        # implementation type has changed => rebuilt
.PHONY : ${EXEC1}
${EXEC1} :
        rm -f MergelImpl
        touch q2quicksort.h
        ${MAKE} ${EXEC1} STYPE="${STYPE}" SIMPL="${SIMPL}"
endif

MergelImpl :
        echo "IMPLSTYPE=${STYPE} \nIMPLSIMPL=${SIMPL}" > MergelImpl
        sleep 1

-include BuflImpl

# same implementation concurrency/type as last time ?
ifeq ($(shell if [ "${BUFIMPL}" = "${BIMPL}" -a "${BCHECKIMPL}" = "${BCHECK}" ]; \
then echo true ; fi },true)
${EXEC2} : ${OBJECTS2}
        ${CXX} ${CXXFLAGS} $^ -o $@
else
        # implementation type has changed => rebuilt
.PHONY : ${EXEC2}
${EXEC2} :
        rm -f BuflImpl
        touch q3buffer.h
        sleep 1
        ${MAKE} ${EXEC2} BIMPL="${BIMPL}" BCHECK="${BCHECK}"
endif

BuflImpl :
        echo "BUFIMPL=${BIMPL} \nBCHECKIMPL=${BCHECK}" > BuflImpl
        sleep 1

```

```
#####

${OBJECTS} : ${MAKEFILE_NAME}          # OPTIONAL : changes to this file => recompile
-include ${DEPENDS}                    # include *.d files containing program dependences

clean :                                # remove files that can be regenerated
    rm -f *.d *.o ${EXECS} MergelImpl BuflImpl
```

This makefile is used as follows:

```
$ make quicksort STYPE=int SIMPL=CBEGIN
$ quicksort ...
$ make quicksort STYPE=double SIMPL=ACTOR
$ quicksort ...
$ make quicksort STYPE=char SIMPL=TASK
$ quicksort ...
$ make buffer BIMPL=BUSY BCHECK=BARGINGCHECK
$ buffer ...
$ make buffer BIMPL=NOBUSY OPT="-O2 "
$ buffer ...
```

Put this Makefile in the directory with the programs, name the source files as specified above, and then type make quicksort or make buffer in the directory to compile the programs. This Makefile must be submitted with the assignment to build the program, so it must be correct. Use the web tool [Request Test Compilation](#) to ensure you have submitted the appropriate files, your makefile is correct, and your code compiles in the testing environment.

Follow these guidelines. Your grade depends on it!