laSalle
UNIVERSITAT RAMON LLULL

Operating Systems 2024-2025

Session 3: Threads 16/10/24

Goals

In this session, you will begin to practice complex techniques typical of operating systems. This session will focus on the creation of threads in order to use several computer cores and thus carry out different tasks simultaneously. Likewise, basic forms of thread synchronization and the passing of variables between them will be practiced. Different thread settings should also be used.

Motivation

More specifically, with this session, the student must exercise:

Threads creation (pthread_create).

Passing parameters to threads.

Threads synchronization (pthread_join).

Attribute configuration (pthread_attr_t).

Previous Documentation

To complete this session, it is recommended to read the following references:

SALVADOR, J. (2009). *Introducció al llenguatge de programació C*.

STEVENS, W. R. & RAGO S.A. (2008). *Advanced Programming in the UNIX Environment*, 2nd edition.

TANENBAUM, A. S. (2009). *Modern Operating Systems*, 3rd Edition.

laSalle
UNIVERSITAT RAMON LLULL

Operating Systems 2024-2025

Session 3: Threads 16/10/24

Concurrent Statistics Calculation

When working with large datasets, such as in data analysis or data science, it is crucial to calculate basic statistics efficiently. In this session, you will develop a program capable of calculating several basic statistics (mean, variance, standard deviation, maximum, and median) using multiple threads.

Since the file containing the data might be large, we will optimize the calculations by dividing the work among different threads. Each thread will handle a specific task, and the results will be combined at the end. This approach will allow the calculations to be performed concurrently, improving efficiency.

Your program will take a filename as an argument, which contains the numerical data to be processed. Example: `./S3 data.txt`. The file data.txt contains numerical data, one number per line.

Task Requirements:

1. Reading the Data:
    - Your program must read all the data from the file and store it in an array (or list).
2. Threaded Calculations:
    - You will create multiple threads, each performing a specific calculation:
        - One thread will calculate the mean.
        - Another thread will calculate the median.
        - A third thread will calculate the maximum value.
        - A fourth thread will compute the variance (the formula will be provided in an annex).
    - Each thread must perform its task independently. The program should display the results for all the calculated statistics once they all finish.
    - In some cases, certain calculations (such as variance) might depend on others (e.g., the mean). If needed, ensure that threads wait for others to finish when there is a dependency.

If the file cannot be found or there is an issue reading the data, an appropriate error message should be displayed. For example: `Error: Unable to open the file 'data.txt'`

The output should summarize the values like this:

```
Mean: 7.36
Median: 4.57
Maximum value: 10.2
Minimum value: 1.0
Variance: 4.5
```

You must wait for all threads to complete their tasks before displaying the results.

laSalle
UNIVERSITAT RAMON LLULL

Operating Systems 2024-2025

Session 3: Threads 16/10/24

Execution Example

```
alumne@matagalls:~/SO/Practiques20242025/ICE/S3$ ./S3.exe
ERROR: Incorrect number of arguments
Please provide the input file name
Usage: ./S3 <data_file>
alumne@matagalls:~/SO/Practiques20242025/ICE/S3$ ./S3.exe data.txt
Mean: 50.271710
Median: 50.505000
Maximum value: 99.960000
Minimum value: 1.170000
Variance: 802.096971
alumne@matagalls:~/SO/Practiques20242025/ICE/S3$
```

APPENDIX:

Variance formula:

$$\sigma^2 = \frac{1}{N} \cdot \sum (xi - mean)^2$$

σ^2 = Variance
N = Total numbers
xi = Each point
$mean$ = The mean computed by the thread

Median:

The median of a set of numbers is the value separating the higher half from the lower half of a data sample, a population, or a probability distribution. For a data set, it may be thought of as the “middle" value

Mean formula:

$$mean = \frac{1}{N} \cdot \sum xi$$

xi = Each point
N = Total numbers

laSalle
UNIVERSITAT RAMON LLULL

Operating Systems 2024-2025

Session 3: Threads 16/10/24

Considerations

- The output of the program must be completely identical to what is shown in the execution examples.
- The use of global variables is prohibited.
- SIGINT must be ignored during the entire execution. Program must not stop if one is received.
- The text file will have the specified format and there will be no mistakes.
- You must ensure that, when the execution ends, all the processes have ended properly.
- When the program ends, all the dynamic memory used must have been freed, and all file descriptors must be closed.
- All inputs and outputs must be done using file descriptors. The use of printf, scanf, FILE*, getchar or similar is NOT allowed.
- The use of the functions “system”, “popen”, or from the same family is NOT allowed.
- You must compile using the –Wall, –Wextra, and –lpthread flags.
- Any deliverable containing warnings or errors will be directly discarded.
- At the start of the .c you must include a comment with your logins, names, and surnames.
- For submitting the session, you must hand over a file “S3.c”, and deliver it through eStudy platform.
- For the use of certain functions, it might be necessary to add:

```
#define _XOPEN_SOURCE 500

#define _POSIX_C_SOURCE 1
```