# Breno Farias da Silva Felipe Archanjo da Cunha Mendes Pamella Lissa Sato Tamura Thaynara Ribeiro Falcão dos Santos

## Laboratório 03: Manipulação de Threads

Relatório técnico de atividade prática solicitado pelo professor Rodrigo Campiolo na disciplina de Sistemas Operacionais do Bacharelado em Ciência da Computação da Universidade Tecnológica Federal do Paraná.

Universidade Tecnológica Federal do Paraná – UTFPR

Departamento Acadêmico de Computação – DACOM

Bacharelado em Ciência da Computação – BCC

Campo Mourão Março / 2022

# Sumário

1	Intro	dução	3
2	Obje	tivos	3
3	Fund	amentação	3
4	Mate	riais	3
5	Proc	edimentos e Resultados	4
	5.1	Parte 1: Análise de Threads	4
	5.2	Parte 2: Programação	7
6	Conc	elusões	4
7	Refer	rências 1	4

## 1 Introdução

O relatório está dividido em duas partes, a primeira reservada para a manipulação de threads além de ser composta por perguntas e respostas. Já a segunda parte, está relacionada à programação da qual engloba o desenvolvimento de programas que criam N threads (5) para diversos objetivos.

## 2 Objetivos

O objetivo do trabalho consiste em, primeiramente, aprimorar os conhecimentos estudados em sala de aula sobre o funcionamento das threads. Além de construir programas que desenvolvam determinadas soluções envolvendo threads e explorar bibliotecas como pthreads.

## 3 Fundamentação

A definição de uma thread segundo Tanebaum é um fluxo de execução que pode ser associado a outras threads para formar maiores processos, ou seja, sao pequenas sequencias de instruções que possuem um fluxo de execução independente podendo ser utilizadas dentro no núcleo dos Sistemas Operacionais (threads de núcleo).

As perguntas e respostas foram fundamentadas pelas aulas e materiais fornecidos pelo próprio Prof. Dr. Rodrigo Campiolo, assim como o desenvolvimento do programa e códigos descritos na parte dois do documento. Além disso, as definições de comandos e conceitos que foram obtidos baseadas em aulas e, parcialmente, por pesquisas sequenciais por cada membro da equipe.

#### 4 Materiais

- Debian 11 LTS (ISO)
- VirtualBox 6.1.32
- Debian GNU/Linux 11 (Bullseye)
- Ryzen 7 3800x
- 8GiB de RAM

#### 5 Procedimentos e Resultados

#### 5.1 Parte 1: Análise de Threads

1. Identifique no seu sistema Linux quantas threads estão em execução. Qual o processo com o maior número de threads?  $\cdot$ 

Para verificar o número de threads em execução, basta executar o htop (Figura 1):

```
brenofarias@breno: ~
                                                                         Q =
                                                 Load average: 1.47 1.25 0.64
                                 0K/7.45G]
                                                                       0:05.75
0:13.16
4504 brenofari
                                                                       0:43.85
                              3024M
                                                                                 /usr/lib/firef
                                                         6.1
4.0
3.4
4085 brenofari
                              34.4G
                                       743N
                                                                       1:39.49 /usr/lib/firefo
                              3201M
5898M
5898M
4736 brenofari
1302 brenofari
                                                                       0:20.68 /usr/lib/firef
0:10.54 /usr/bin/gnome
1303 brenofari
                                        352M
                                                                       0:10.80
1305 brenofari
                           0 5898M
                                       352M
                                                                       0:10.18
1307 brenofari
1309 brenofari
1304 brenofari
                                                                       0:10.51
```

Figura 1 – Demonstração do comando htop

Neste sentido, é possível observar que há 923 threads em execução no sistema em questão. Para verificar qual processo tem o maior número de threads, basta executar o comando *pstree* (Figura 2) e olhar qual processo tem o maior número de filhos. E ainda, a quantidade de filhos pode ser notada ao observar o maior número de processos aninhados/identados.

Figura 2 – Demonstração do comando pstree

Sendo assim, é possível constatar que o processo com maior número de threads em nossa máquina virtual estão relacionados ao systemd, o qual é um processo daemon que gerencia outros processos daemons.

2. Qual o número máximo de threads que o seu sistema suporta? Para verificar o número máximo de threads suportada pelo sistema operacional usado, basta executar o seguinte comando: cat /proc/sys/kernel/threads-max (Figura 3)

Com isso, é possível ver que o limite máximo de threads suportada pela máquina virtual do Debian é de 62494 threads.

brenofarias@breno:~\$ cat /proc/sys/kernel/threads-max
62494
brenofarias@breno:~\$ [

Figura 3 – Demonstração do comando cat /proc/sys/kernel/threads-max

3. Verifique o tempo de execução do programa da questão 3, parte 2, considerando: 1 thread, 2 threads, 4 threads, 8 threads e 16 threads. Descreva o hardware (processador, memória e número de núcleos, tamanho da matriz usada nos testes e o tempo de execução para cada teste).

Considerando uma matriz  $100 \times 200$ , temos os seguintes resultados em uma máquina com um processador Ryzen  $7~3800 \times$  de 8~núcleos e memória RAM de 8GB:

N = 1: 9.542 milissegundos

N = 2: 9.568 milissegundos

N = 4: 10.106 segundos

N = 8: 11.379 milissegundos

N = 16: 12.285 milissegundos

Conclusão: Do exposto, é possível concluir que para valores fixos de linhas (M) e colunas (N), a única variável em questão tornou-se o número de threads usadas. Seguindo o fluxo dos valores para N sendo 1, 2, 4, 8 e 16, torna-se evidente que o menor tempo de execução foi dado para N=1 e o maior tempo de execução foi dado para N=16. Este fato ocorreu devido a problemas com concorrência de dados, paralelismo e sincronização de threads. Há dois tipos de paralelismo, sendo eles, o paralelismo de função e paralelismo de dados.

O paralelismo de função é dado pelo uso de diferentes threads para executar diferentes funções. Neste caso, o paralelismo de função se dá pela divisão de threads diferentes para executar cálculos distintos, definidos pela mediana de cada linha e a média aritmética de cada coluna.

Já o paralelismo de dados ocorre pela execução de mesma função, porém em segmentos diferentes do mesmo conjunto de dados. No caso em questão, esse paralelismo ocorre com a subdivisão das linhas em N subvetores e o mesmo para as colunas. Ou seja, a mesma operação é feita porém em subpartes da matriz. Ao implementar o paralelismo de função, é esperado melhorias significativas de performance, contudo nem sempre o mesmo acontece com a implementação de paralelismo de dados, pois este pode gerar problemas de sincronização os quais nem sempre irão compensar o ganho de desempenho ao adicioná-lo no algoritmo. Se a complexidade computacional da tarefa não é grande e o conjunto de dados também não então, o paralelismo pode simplesmente acarretar em um overhead (sobrecarga) adicional.

Quando criam-se threads, os SO's convencionais precisam criar uma fila para estas threads poderem usar as CPU's disponíveis. Cada thread recebe um tempo para usar a CPU (quantum/time slice) de acordo com a prioridade do processo. Quando o tempo do processo encerra, ele volta para a fila. Assim, a thread executa e para, consecutivamente, até finalizar. Isso é chamado de context switch e, tal processo, tem um custo computacional adicional. Por isso, em certos casos, criar threads demais pode piorar o tempo de processamento.

Sendo assim, a medida que o valor de N aumenta, é notável que o tempo gasto na sincronização das threads não foram suficientes para manter o ganho de performance ao adicionar o paralelismo de dados. Vale relembrar que, sem paralelismo de dados o código não se tornaria mais eficiente ao adicionar mais threads, pois threads diferentes estariam acessando/alterando o mesmo item de dados o que poderia causar problemas como "atualização perdida", pois uma escrita poderia sobrescrever a outra gerando inconsistência nos dados. Além disso, a leitura de dados no disco seria feita uma por vez, o que tiraria o paralelismo do algoritmo.

### 5.2 Parte 2: Programação

1. Faça um programa (Figura 4 e Figura 5) com N threads que localiza um valor em um vetor de inteiros. O espaço de busca no vetor deve ser distribuído para as N threads.

Figura 4 – Demonstração do programa com N threads que localiza um valor em um vetor de inteiros

Figura 5 – Terminal do programa demonstrado acima

2. Faça um programa (Figura 6 e Figura 7) com 2 threads que calcule a frequência de elementos em um vetor que contém somente valores de 0 a 9.



Figura 6 – Demonstração do programa com 2 threads enunciado anteriormente

Figura 7 – Execução do programa demonstrado acima

#### 3. Implemente um programa multithread com pthreads que calcule:

- a) a mediana de cada linha de uma matriz MxN e devolva o resultado em um vetor de tamanho M. (Figura 8 e Figura 9)
- b) a média aritmética de cada coluna de uma matriz MxN e devolva o resultado em um vetor de tamanho N. (Figura 10) O programa deve gerar matrizes MxN com elementos aleatórios para arquivos; usar técnicas de paralelização de funções e de dados; ler matrizes MxN de arquivos no formato em anexo; gravar os resultados em um arquivo texto. (Figura 11 e Figura 12 e Figura 13)

```
Binclude cstdio.h>
Binclude cstdib.h>
Binclude cstdib.h>
Binclude cstring.h>
Binc
```

 ${\bf Figura~8-Demonstração~do~programa~multithread~com~pthreads~c\'alculando~a~mediana~de~uma~matriz}$ 

```
int main(){
   int op;
   pthread_t threads[NUM_THREADS];
   Dados dados[NUM_THREADS];
       printf("Digite:\n[1] - Gerar Matriz\n[2] - Carregar Matriz\n");
       scanf("%d", &op);
   } while(op != 1 && op != 2);
   if (op == 1){
       printf("Digite o número de linhas: ");
       scanf("%d", &rows);
       printf("Digite o número de colunas: ");
       scanf("%d", &cols);
       matriz = create_matrix(rows, cols);
       srand(time(NULL));
       generate_elements(matriz, rows, cols, 99);
       printf("\nMatriz %dx%d criada com sucesso!\n", rows, cols);
   } else{
       matriz = read_matrix_from_file("data_matrix.in", &rows, &cols);
       printf("\nMatriz carregada com sucesso!\n");
   clock_t start = clock();
   printf("\nNumero de Threads: %d\n", NUM_THREADS);
   printf("\nMatriz [%d x %d]\n\n", rows, cols);
```

Figura 9 – Continuação da demonstração do programa iniciado anteriormente

```
printf("----- MEDIA -----\n");
float resultado_medias[cols];
for(int i = 0; i < cols; i++){
    resultado_medias[i] = 0;
medias = resultado_medias;
int quant = cols / NUM_THREADS;
int qnt_aux = cols % NUM_THREADS;
int aux = 0;
for(int i = 0; i < NUM_THREADS; i++){</pre>
   if(i < qnt_aux){</pre>
       dados[i].quantidade = quant + 1;
    } else{
        dados[i].quantidade = quant;
   dados[i].id = i;
    printf("Thread %d: %d colunas\n", i, dados[i].quantidade);
printf("\n");
print_matrix(matriz, rows, cols);
for(int i = 0; i < NUM_THREADS; i++){
   pthread_create(&threads[i], NULL, thread_media, (void*) &dados[i]);
for(int i = 0; i < NUM_THREADS; i++){</pre>
   pthread_join(threads[i], NULL);
```

Figura 10 – Continuação da demonstração do programa iniciado anteriormente

```
printf("\n-----\n");
// Divisão das linhas em "rows" partes
// Divisão das linhas em "row:
quant = rows / NUM_THREADS;
qnt_aux = rows % NUM_THREADS;
aux = 0;
for(int i = 0; i < NUM_THREADS; i++){
   if(i < qnt_aux){
      dados[i].quantidade = quant + 1;
   } else{
      dados[i].quantidade = quant;</pre>
       dados[i].id = i;
 // Ordenando as linhas da matriz principal
for(int i = 0; i < rows; i++){
    selectionSort(matriz[i], cols);</pre>
 int resultado_medianas[rows];
for(int i = 0; i < rows; i++){
    resultado_medianas[i] = 0;</pre>
 }
medianas = resultado_medianas;
 // Criar as threads para a mediana
for(int i = 0; i < NUM_THREADS; i++){
    pthread_create(&threads[i], NULL, thread_mediana, (void*) &dados[i]);</pre>
 // Esperar as threads terminarem
for(int i = 0; i < NUM_THREADS; i++){
    pthread_join(threads[i], NULL);</pre>
 printf("\nMediana: ");
for(int i=0; i<rows;i++){
   printf(" [%d]", resultado_medianas[i]);</pre>
printf("\nMedia: ");
for(int i=0; i<cols;i++){
    printf(" [%.2f]", medias[i]);
}</pre>
 printf("\n");
 // Salvanr os resultados num arquivo
FILE* arq_result = fopen("results.txt", "w");
 fprintf(arq_result, "Mediana: ");
for(int i=0; i<rows;i++){
    fprintf(arq_result, " [%d]", resultado_medianas[i]);</pre>
 fprintf(arq_result, "Media: ");
for(int i=0; i<cols;i++){
    fprintf(arq_result, " [%.2f]", medias[i]);
}</pre>
printf("\nResultados salvos em results.txt com sucesso!\n");
// Fim da contagem de tempo
clock_t end = clock();
double time_spent = (double)(end - start) / CLOCKS_PER_SEC;
```

Figura 11 – Continuação da demonstração da implementação do programa enunciado acima

Figura 12 – Continuação da demonstração da implementação do programa enunciado acima

Figura 13 – Terminal da demonstração do programa multithread com pthreads e suas exigências

**OBS.** Segue anexo, no mesmo diretório desse documento, todos os códigos-fontes referentes aos exercícios da parte 2.

## 6 Conclusões

Após realizados os procedimentos solicitados na descrição da parte 1 e 2 da atividade, finalizou-se a prática do laboratório 3.

## 7 Referências

- $\bullet \ \ https://moodle.utfpr.edu.br/mod/resource/view.php?id=415720$
- $\bullet \ \ https://www.baeldung.com/linux/max-threads-per-process$