



Московский государственный университет имени М.В. Ломоносова
Факультет вычислительной математики и кибернетики

Отчет по учебному курсу «Распределенные системы»

Автор: Губанов Федор

группа 420

Москва, 2021 г.

Содержание

Содержание	2
Постановка задачи	3
Реализация программы и оценка времени выполнения	4
Заключение	6

Постановка задачи

Требуется сделать следующее:

- Реализовать программу, моделирующую выполнение протокола голосования для 11 файловых серверов при помощи пересылок MPI типа точка-точка.
- Доработать MPI-программу, реализованную в рамках курса “Суперкомпьютеры и параллельная обработка данных”. Добавить контрольные точки для продолжения работы программы в случае сбоя.

После реализации программы, моделирующую выполнение протокола голосования, необходимо оценить сколько времени потребуется для выполнения одним процессом 3-х операций записи и 10 операций чтения N байтов информации с файлом, расположенным (размноженным) на 11 серверах. А также определить оптимальные значения кворума чтения и кворума записи для $N=300$. Время старта равно 100, время передачи байта равно 1 ($T_s=100, T_b=1$).

Реализация программы и оценка времени выполнения

Идея метода голосования - запрашивать чтение и запись файла у многих серверов.

Для реализации этого метода используется кворум чтения - далее N_r и кворум записи - далее N_w .

1. Для выполнения чтения достаточно обратиться к N_r серверам и использовать данные того сервера, который располагает последней версией файла.
2. Для выполнения записи необходимо, чтобы все N_w сервером выполнили запись и увеличили свои версии файлов на 1. При этом у всех серверов должно быть согласие относительно номера текущей версии.

Для обеспечения корректности данных значения для кворума должны удовлетворять неравенству: $N_r + N_w > N$, где N - количество всех серверов. Поскольку чтение является более частотной операцией, то естественно взять $N_r = 1$.

Данный алгоритм был реализован с помощью функций `MPI_Send` и `MPI_Recv`. А также работа с директориями/файлами с помощью стандартных библиотек C++.

Репозиторий с кодом и файлами сборки/запуска:

https://github.com/Fedor1533/MSU_Dis_Systems

Оценка времени выполнения

Оценим время работы алгоритма. Если время старта равно 100, время передачи байта равно 1 ($T_s=100, T_b=1$), то время выполнения операции передачи N байт рассчитывается следующим образом:

$$time = T_s + N \cdot T_b$$

Тогда для десяти операций чтения потребуется отправить N_r серверам запрос на чтение, после чего каждый из N_r серверов пришлет версию файла, которую он хранит. Предположим каждая из посылок содержит 4 байта. В итоге время выполнения 10 операций чтения:

$time1 = 10 * (N_r * Ts + N_r * 4 * Tb)$ - запрос на чтение

$time2 = 10 * (N_r * Ts + N_r * 4 * Tb)$ - посылка локальной версии с сервера

$$timeR = time1 + time2 = \underline{2000 * N_r + 80 * N_r}$$

Для трех операций записи N байт потребуется отправить N_w серверам запрос на запись, отправить N байт N_w серверам, после чего каждый из N_w серверов отправляет новую версию, подтверждая изменения, при этом версии должны быть согласованы.

В итоге время выполнения 3 операций записи:

$time1 = 3 * (N_w * Ts + N_w * 4 * Tb)$ - запрос на запись

$time2 = 3 * (N_w * Ts + N_w * N * Tb)$ - отправка N байт серверам

$time3 = 3 * (N_w * Ts + N_w * 4 * Tb)$ - посылка локальной версии с сервера подтверждающая запись

$$timeW = time1 + time2 + time3 = \underline{900 * N_w + (3 * N + 24) * N_w}$$

Оптимальное значение

Определим оптимальные значения кворума чтения и кворума записи для $N=300$. Для этого используем тождества, полученные при вычислении времени выполнения операций чтения и записи.

В результате, задача сводится к минимизации данного выражения:

$$\underline{2000 * N_r + 80 * N_r + 900 * N_w + 924 * N_w}, \text{ при этом } \underline{N_w + N_r > 11}$$

Минимум достигается при $N_w = 11$ и $N_r = 1$, то есть согласие на запись должны дать все серверы, а на чтение один любой процесс.

Заключение

Таким образом, была реализована программа, моделирующую выполнение протокола голосования для 11 файловых серверов при помощи пересылок MPI типа точка-точка.

Оценено время выполнения одним процессом 3-х операций записи и 10 операций чтения N байтов информации с файлом, расположенным (размноженным) на 11 серверах. Определено оптимальное значение кворума чтения и кворума записи для $N=300$.