

Lab1 test report

1. 实验概要

多线程编程是高性能编程的技术之一，实验 1 将针对数独求解问题比较多线程与单线程的性能差异、同一功能不同代码实现的性能差异以及多线程在不同硬件环境下的性能差异。

1.1 程序输入

程序将在控制台接收用户输入，该输入应为某一目录下的一个数独谜题文件，该文件包含多个数独谜题，每个数独谜题按固定格式存储在该文件中。

1.2 程序输出

实验中把数独的解按与输入相对应的顺序写入到一个文件中。

1.3 Sudoku 算法

实验共提供了 4 中不同的 Sudoku 求解算法：BASIC, DANCE, MINA 和 MINAC。其中，DANCE 算法速度最快，BASIC 算法速度最慢。实验中选用的是最快的 DANCE 算法。

1.4 性能指标

实验以求解完单个输入文件里的所有数独题并把数独的解按顺序写入文件所需要的时间开销作为性能指标。

一般而言，可以用加速比直观地表示并程序与串程序之间的性能差异（加速比：串行执行时间与并行执行时间的比率，是串行与并行执行时间之间一个具体的比较指标）。为了精确地测量性能，时间开销均在数独求解进程/线程绑定 CPU 的某个核的条件下测得，这样保证了该进程/线程不被调度到其他核中运行，但不保证该进程/线程独占某个核。更精确的测量方法可以先把 CPU 的某个核隔离，而后再绑定在某个进程/线程上，这样该 CPU 核心不会运行其他的用户程序。当 CPU 资源充足时（CPU 核心数足够多，当前正在运行的进程/线程足够少），是否把核隔离并没有多大影响，因为操作系统的调度策略不会频繁的对线程/进程进行无谓的调度。

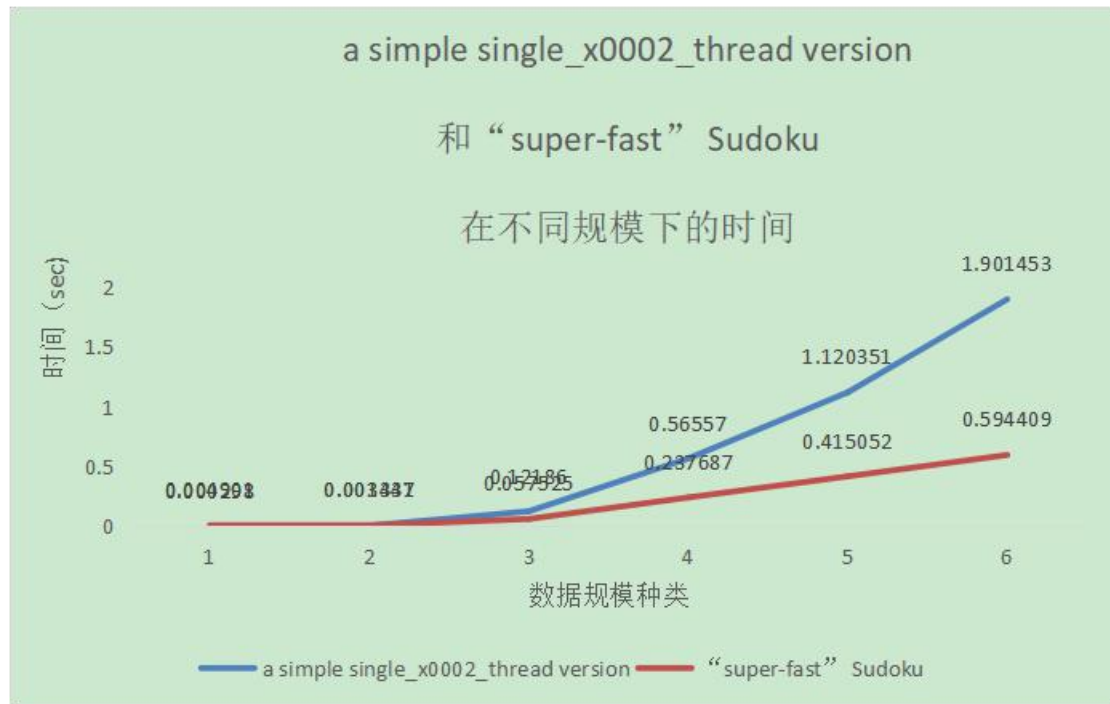
2. 性能测试

程序的性能会受到诸多因素的影响，其中包括软件层面的因素和硬件层面的因素。本节将分析比较多线程程序与单线程程序的性能差异、同一功能不同代码实现的性能差异，以及同一个程序在不同硬件环境下的性能差异。

2.1 单线程和最快数独比较

数据规模分为 6 种，分别为 1,10,1000,5000,10000,15000 分别标记为 1, 2, 3, 4, 5, 6.

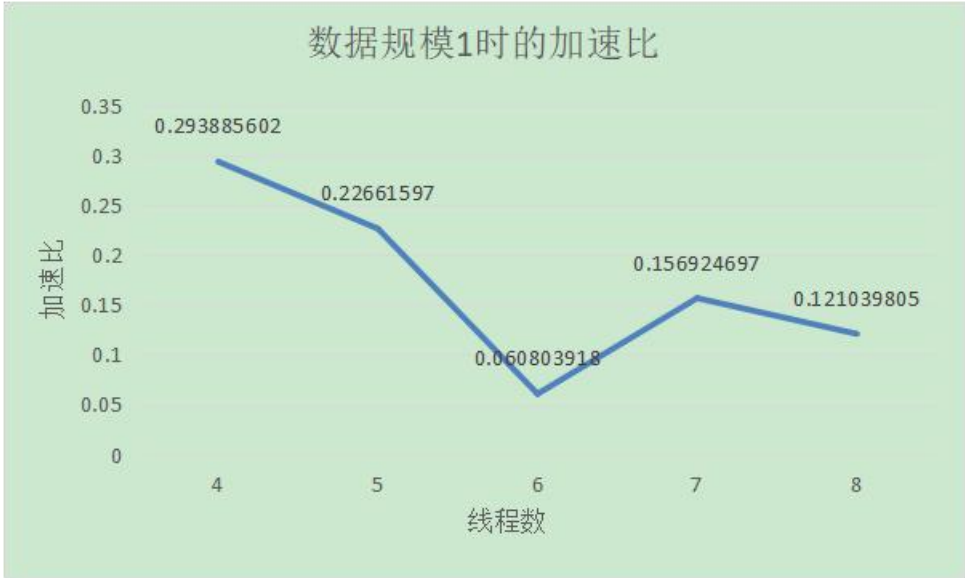
由于多线程的时候，线程也需要时间开销，所以在数据规模较小的时候加速比会比较小，没有可比性，当数据规模为 1000 的时候所需要的时间开销减小，加速比增大。

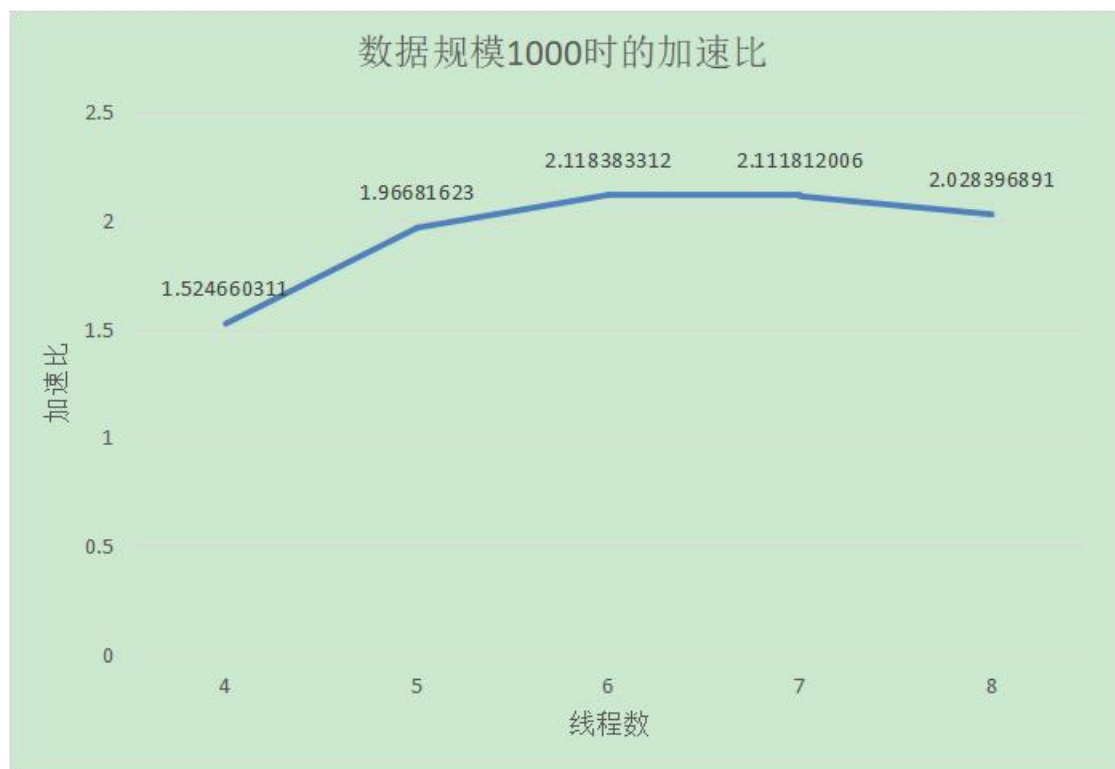


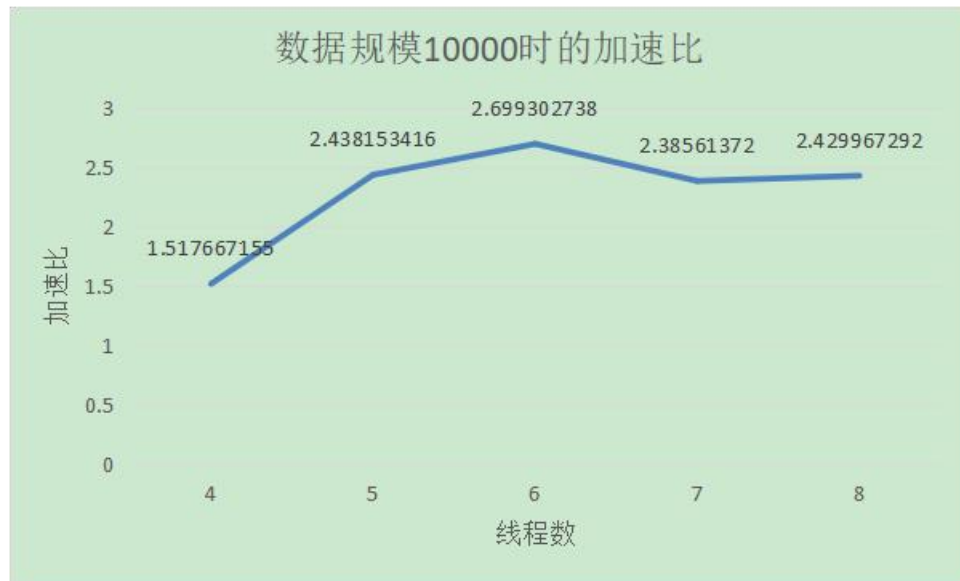
2.2 不同线程比较

输入和输出各占了一个线程，解决问题的线程数从 2 到 6 依次增加。同时由于虚拟机只有 4 个核，由于输入输出很快执行完，所以开的线程为 6 的时候性能最好，当总线程数小于 CPU 总核心数时，随着线程数的增加，所需要的时间开销越小、加速比更高。从 sudoku_solve 线程数为 6 开始，总线程数开始超过 CPU 物理核心数，线程开

始被操作系统调度，调度会有一定的开销，所以性能会有所下降。 同样的，在数据规模较小的时候没有可比性，根据图可以看出。

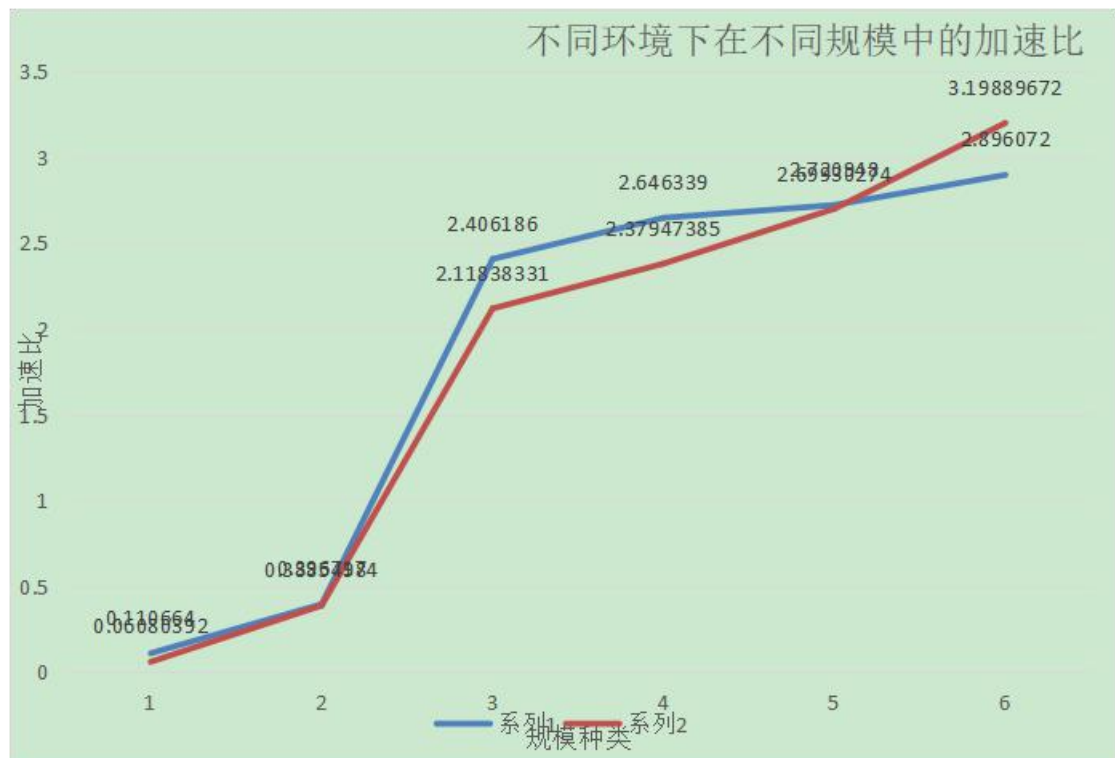






2.3 不同环境下的比较

使用不同版本的虚拟机，选择 2G 和 4G 内存得到的加速比如下：总的来说 4G 和 VMware-workstation-full-15.5.2-15785246.exe 下运行性能更好



三、总结

使用更高版本的虚拟机性能更好，在我们调整线程数的时候，当解决问题的线程等于核心数的时候性能最好，超过这个数性能就会略微下降。同时由于创建线程有时间开销，所以小规模数据不能体现多线程对单线程的时间优化。当然我们从第一个图表可以看出，在不同的输入下，多线程所需的时间要优于单线程。