

COMP3712 Computer Programming 3
COMP9712 Computer Programming 3 (GE)

Semester 1, 2020

Examination

Computing and Mathematical Sciences

Examination Duration: 2 Days

Pages: 2

Questions: 1

Instructions To Students:

1. This is an open book, open notes, take-home exam. You may consult any of the course materials (and other outside materials if you wish, though they will probably not be as much help) in figuring out the answers to the questions.
2. You may NOT discuss the exam with ANYONE, whether they are taking the topic or not.
3. Take-Home exams are designed to find out about:
 - what YOU know about the topic
 - how YOU APPLY your knowledge of the topic; and
 - not how much you can find and copy from a source.
4. If you find a solution for a test problem in a book or online, cite it in your submission and **do not** copy it as-is, but make changes which demonstrate you understand what you are writing as you would for a normal assignment. Otherwise we may not accept it. Your submission will be processed through Turnitin, just as an assignment would be.
5. You should submit your answers as a PDF document via the FLO assignment box and clearly identify each section of the question in your answer.

If you have any questions, direct them to me via e-mail (trent.lewis@flinders.edu.au). I will not answer questions that are too specific and would give you an unfair advantage in taking the exam, but I will be happy to give clarification on what is meant by a question. If you are at all uncertain about what sort of answer I am looking for, think you have found an error or typo, please ask me about it.

Question 1 (50 Marks)

A virus has broken out across the world. There is currently no effective anti-viral drugs and all treatments for infected patients are only saving the lives of a few. The disease is highly infectious and only requires people to be within 1.5m for infection to occur. Infected people may be asymptomatic but can still spread the disease. The best course of action appears to be physical distancing. When an infection is detected, quick and effective contact tracing must be conducted. This occurs via a tracking app installed covertly on the public's mobile devices. Identified contacts are quarantined to stop the spread. The tracking app is continuously sending data to the authorities and requires no user intervention for the tracking to occur.

You have been tasked with developing the software that will complete the contact tracing based on the retrieved data from the tracking app. Each user of the app has a unique alphanumeric code.

(i) Describe how you would implement such a program, outlining the data structures and algorithms that you would use. Give examples of how the infected person and their contacts (and the contact's contacts and so on) would be represented in your program and how the contacts would be found. Pictures are encouraged. (15 Marks)

Once the contact tracing has been completed, each person needs to be contacted, however, to maintain a level of anonymity, the tracing only has the alphanumeric code for each person. A data store must be constructed that can contain all estimated 5 billion people whose mobile devices have the app installed (the app can be installed on both smart and non-smart phones). The record in the data store for each person contains their mobile phone number, name, and last known address. It is vitally important that the search is rapid and can be executed many times over.

(ii) How would you initialize the data structure? What is the time and memory complexity of doing this? Express your answer in terms of n , where n is the number of people, and/or m the length of the alphanumeric code. Keep in mind that the solution should be both time and memory efficient, but sometimes a trade-off is necessary. Justify any trade-offs considered. (15 Marks)

Once people have been found they must be tested. New evidence from the scientific community has shown that the virus can be detected in the saliva of infected people, even before they are symptomatic. The virus DNA has unique patterns in the sequence of an unknown origin. The saliva also contains DNA from many protein sequences and the profiling process produces many smaller sequences of DNA that must be tested against the virus DNA. An issue that arises from the sequencing process is that the DNA sequences are not perfect - there are sometimes errors in the process, the sequences from the extraction process are smaller than the full DNA sequence, and the virus may have small mutations.

(iii) How would you test each of the unknown DNA sequences from the saliva against the newly discovered virus DNA sequence? (10 marks)

With the world's computational resources at your disposal, you are now required to write a version of your program that makes use of multithreaded programming techniques to improve the running time of all aspects of the above.

(iv) Discuss what parts of your code can be parallelized and the issues associated with doing so. For example, if you have access to a 1000-core supercomputer, will you see a 1000 times speed up? (10 Marks)

You have now saved the world!