Parallel Design Patterns

Assessed Coursework, 2016
PART ONE

The deadline for PART ONE of the assessed coursework is Monday 22nd of February at 12 noon. You will submit this via the PDP learn pages.

About the coursework

The assessment for Parallel Design Patterns is split into two parts. This first submission ("part one") comprises of a short report. I expect that a complete answer could be expressed succinctly in **two pages**. The second submission ("part two") builds on this same problem and will require you to submit both source code, and a further written report.

Overview of the problem

You will solve a problem similar to one often faced by developers working for EPCC. Red squirrels are in decline in the UK and as such there is considerable effort to save them, however a disease called *squirrel parapoxvirus* is especially perilous to red squirrels and can often kill them, wiping out entire habitats. A team of biologists are working to try and understand the spread of this disease through the squirrel population, with the eventual aim that their findings will help to protect the animals. They have written bits of a serial code to model the disease based upon empirical (experimental) observations of squirrels and they now wish to parallelise this in order to simulate real world habitats. However they have no clear idea about how to achieve this parallelisation, therefore they therefore come to you, with their code, for advice. A more detailed description of the model is later in this document.

Your tasks

For this, coursework part one, you should analyse the problem and write a report answering the following key questions:

1. **[weighting: 50%]**
    a. A statement of which parallelisation pattern (from the algorithm structure/strategy design space is most suitable.)
    b. Based upon the pattern chosen, an explanation of *why* this was chosen. This should include a discussion about how the pattern's context and forces relate to the problem
2. **[weighting: 20%]**
    a. The biologists decide to adopt a regular two-dimensional grid to represent the environment, but want to model the spread of disease into neighbouring cells through means other than squirrels. This corresponds to a short range interaction between neighbouring cells. What pattern might be applicable here, and why?
3. **[weighting: 20%]**
    a. Pick a further two patterns from the course, and discuss why *either* they are not suitable for this problem or could be used in conjunction with the pattern you have selected
4. **[weighting: 10%]**
    a. Your choice of pattern in this assessment will be based on the problem as described here. Given the pattern you have chosen, what aspects of the pattern and its solutions might be relevant to the choice of implementation language or hardware platform?

Details of the biologist's model

The model that the biologists have written exhibits the following behaviour:

- Squirrels can move about their environment
- A squirrel's current position in the environment is described by a coordinate pair (x,y)
- A squirrel moves by stepping to a new position
- Squirrels can be born
    - In this model all squirrels give birth to new squirrels
    - New squirrels start in the same position as their parents
    - A squirrel can move as soon as it is born. We assume all new squirrels are adults, capable as any other squirrel (we ignore the early phases in a squirrel's life.)
    - All squirrels are born healthy (but see note about the initial infection at the start of the simulation.)
- Squirrels can catch squirrel parapoxvirus
    - The animals either have the disease or not (so their state of infection can be represented as a boolean.)
    - A squirrel never recovers from the disease, they have it until they die
- Squirrels can die; they are normally long lived so in this model they only die from being infected by the disease.
- The environment in which the squirrels live is modelled as a set of grid cells. In their current model, this is a regular grid, but they would like the parallel framework to be able to deal with more general grids in the future. The biologists have supplied a function which converts an (x,y) position to the grid cell (an integer cell number).
- Time is modelled coarsely, in terms of months
    - Squirrels can take many steps in a single month
    - Squirrels move throughout the entire month, they have no concept of global time or seasons.
    - The squirrel parapoxvirus virus can live without a host in the environment for 2 months.
- The disease is passed through the environment
- Land cells each have a populationInflux value calculated as the total number of squirrels that have stepped into the specific cell at some point during the past 3 months.
    - Squirrels that are born in a cell do not contribute to the populationInflux at the instant that they are born
    - If a squirrel steps and ends up in the same cell, this counts as a squirrel arriving in the cell. (In other words, every hop can be treated in the same way, as a squirrel arriving in a cell.)
    - Each time the month changes the cells should be updated to represent the past 3 months
- Land cells each have an *infectionLevel* which is calculated as the total number of infected squirrels that have been in the specific cell in the past 2 months.
    - All squirrels are born healthy so won't contribute to this level
    - If a squirrel steps and ends up in the same cell, this counts as a squirrel arriving in the cell. (In other words, every hop can be treated in the same way, as a squirrel arriving in a cell.)
    - Each time the month changes the cells should be updated to represent the past 2 months
- After every 50 steps, a squirrel will reproduce with a probability depending only on the average *populationInflux* of all of the cells they have visited in their last 50 steps.
- After every step, the squirrel will catch the disease with a probability depending on the *infectionLevel* of all the cells they have visited in the last 50 steps.
- An infected squirrel will live for a minimum of 50 steps, after this then they will die with a fixed probability of 1/6 for each step after.

Parameters

The following parameters will give you an idea of the scale of simulations that the biologists wish to perform, but these will not influence your choice of pattern.

- Initial number of squirrels: 34
- Number of cells: 16
- Maximum allowed number of squirrels: 200
  - The biologists don't think the habitat can ever home more than this number, if the number exceeds 200 then your simulation should terminate with an error message
- Initial infection level: 4 squirrels at the start of a simulation
- Months to model: 24 although it should be possible to easily modify this