Exercice - Cours 1

This exercice is derived from the example of the P2P tutorial pages 65 and following.

Motivation: The goal of the exercice is to show an example of performance analysis of an unstructured P2P network.
Question: What is the expected overhead of a search?
Question: How pointer caching helps?

Scenario: Peers cache (store) copies of (or pointer to) content. The caches have limited size (cannot hold everything). The objects have different popularities: different content requested at different rates.

Question: How the cache should be shared among the different contents? Is there an optimal strategy?

Given a system with m objects, n nodes, each node can hold c objects. The total system capacity is hence cn.

q_i is the request rate the i-th object. $q_1 \geq q_2 ... \geq q_m$

p_i is the fraction of the total system capacity to store object i, $\sum_{i=1}^{m} p_i = 1$.

Protocol: To look for an object, a peer successively selects another peer uniformly at random and asks him if it has object i. The peer select peer j with probability $\frac{1}{n}$.
Note that this process has no memory, a peer may ask several time the same peer. We also suppose that a peer may ask himself (this is done to obtain simpler formulas).

Question 1 - Expected length of search, S : A peer looks for object i. What is the expectation of the number of peers (S) he will have to ask? (we suppose here that the probability for a peer to have the data p'_i is proportional to p_i. We have $p'_i = cnp_i$).
Hint to get the formula: One may compute the probability that the number of peers is 1,2, ... k, ...
Hint to compute the formula: Derive the geometric series.

Question 2 - Network bandwidth, B : We consider that to ask one peer uses K units of bandwidth. What is the network bandwidth, B, used to search for all objects? We recall that object i is requested with rate q_i.

The goals for the following of the exercise are:
Goal 1: Find the allocation for the $\{p_i\}$ (as a function of the $\{q_i\}$) that minimizes the bandwidth, B.
Goal 2: Find a distributed method to implement this allocation of the $\{p_i\}$.

Before that we study two allocations often used in practice.
- uniform: $p_1 = p_2 = ... = p_m = \frac{1}{m}$

- proportional: $p_i = aq_i$ where $a = \frac{1}{\sum_{i=1}^{m} q_i}$ is a normalization factor.

Question 3 - Distributed implementation: How would you implement these policies in practice in a distributed way?

Question 4 - Bandwidth usage: What do you expect? Which of the two allocations will use lower bandwidth?

Question 5: What is the network bandwidth usage B for these two allocations?

Question 6: Is it surprising? Discuss. Could you find a better allocation of the $\{p_i\}$?

Square root allocation is optimal.
In this allocation, the p_i are defined as

$$p_i = \frac{\sqrt{q_i}}{\sum_{j=i}^{m} \sqrt{q_j}}$$

Question 7 (difficult): Give a sketch of proof of optimality.
Hint: Consider the search bandwidth B as a function $F(p_1,\dots,p_{m-1})$ and find the minimum of this function.

Question 8 (difficult): How to implement the Square Root allocation in a distributed way?

Question 9 (difficult): Give a sketch of proof that the implementation leads to a square root allocation.

Question 9.1: At time t, what is the average rate r_i at which object i populates?
Hint: Introduce $f_i(t)$ the fraction of locations holding object i at time t.

Question 9.2: When $\frac{f_i(t)}{f_j(t)} < \frac{\sqrt{q_i}}{\sqrt{q_j}}$, what happens to the ratio $\frac{f_i(t)}{f_j(t)}$?
Hint: compute the ratio $\frac{r_i(t)}{r_j(t)}$.

Question 9.3: When $\frac{f_i(t)}{f_j(t)} > \frac{\sqrt{q_i}}{\sqrt{q_j}}$, what happens to the ratio $\frac{f_i(t)}{f_j(t)}$?

Question 9.4: Conclude.