COL819: Programming Assignment 1: Implement Pastry

February 18, 2021

1 Logistics

- Release date: 16 Feb 2021
- Due date: To be decided
- Maximum marks: 100
- Individual assignment
- Submissions on Moodle: The code and the report has to submitted.
- Languages: Java, C, C++, and Python3.X
- Report must be in LaTeX; use vector graphics for images. Do not save your images in the jpeg, png, or bmp formats. Only save your images in the pdf or svg formats. Draw your diagrams using inkscape and plot the graphs using the matplotlib library.

2 Pastry [Total 60 Marks]

In this part of the assignment, you need to implement/simulate a Pastry network.

The implementation has to support some key aspects of the algorithm such as the addition of nodes, deletion of nodes, and look up queries for data. Along with this, there have to be mechanisms for th addition of key-value pairs into the DHT.

You need not create a multithreaded/ multiprocess implementation. You can simulate the behavior of a multithreaded system using a single thread.

2.1 Pastry routing table [30 Marks]

Along with these mandatory aspects, your implementation should also support printing of the routing table of a given node, which can be identified by using the node id. The node id is a unique number that is assigned to each node in the network. Along with the routing table of a node, a summary of the complete network has to be printed:

Total number of nodes : XX.
Total number of data elements : XX.
Total search queries : XX.
Total node add queries : XX.
Total node delete queries : XX.
Total data add queries : XX.

Routing table. ID: XX

|—————————-|
|S.No.|Target|Successor|
|—————————-|
1	XX	Node X
2	XX	Node X
3	XX	Node X
4	XX	Node X
5	XX	Node X
.
.
.
|—————————-|

Print the value of the *target* in the hex format. Print multiple entries in a row (if there are).

2.2 Evaluation [30 Marks]

You need to evaluate the performance of your network for different numbers of nodes in the network. There will be three configurations: 100, 500, and 1000 nodes. For each configuration you need to perform 1 million random search queries, after populating the network with 10,000 data points. You need to report the following for each of the configurations:

- Average number of hops for a search query.
- A histogram showing the distribution of the hops required.
- Operation: Delete half of the nodes from the network, randomly.
- Average number of hops for search queries in this reduced set.
- A histogram showing the distribution of the hops required in this reduced set.

3 Report [Total 40 Marks]

There will be a single report, which must be in the PDF format.

3.1 Experiments [30 Marks]

The report should contain:

- Experimental setting (number of nodes, search queries etc.)
- Graphs with clear labels, x-axis, y-axis and caption for the graph.
- Few lines about the behavior observed.
- Justification of the trends observed.
- The routing table of a single node of your choice for Pastry.

3.2 How to run the code [10 Marks]

The report should also contain details on how to run the code, along with few screenshots.

4 General guidance

- Please stick to basic packages during implementation.
- If you are not sure if a particular package is allowed, ask on Piazza.
- Grading will be done based on the report. There might be a demo, if required. So please ensure that the submitted code executed on your machine correctly.
- We will run MOSS on the submissions. Anyone found with a copied code, either from Internet or from another student, will be dealt with as per the class policy.