

IMPLEMENTING DISCRETE CUCKOO SEARCH ALGORITHM FOR TSP USING MPI AND BEOWULF CLUSTER

HIGH PERFORMANCE COMPUTING (CSE4014) PROJECT

Ву,

Mahale Shantanu

17BCE1161

Submitted To,

S. Harini

November, 2019

CERTIFICATE

This is to certify that the Project Based Learning entitled "Implementing Discrete Cuckoo Search Algorithm for TSP using MPI and Beowulf Cluster" that is being submitted for CAL in B. Tech High Performance Computing (CSE4014) is a record of bonafide work done under my supervision. The contents of this Project, in full or in parts, have neither been take from any other source nor have been submitted for any other CAL course.

Place: Chennai

Date: 08th November 2019

Signature of Students:

Kumar Shikhar17BCE1198Abhijit Pingle17BCE1254Vinay C Shekhar17BCE1317

Signature of Faculty:

Ayesha

School of Computer Science and Engineering (SCSE)

ABSTRACT

This project is an implementation of a research paper based on High Performance Computing which was published on 8th of June, 2019 at International Journal of Innovative Technology and Exploring Engineering (IJITEE) by

Bhavana V, Varshini Ramesh (students of SCSE, VIT Chennai) and Sivagami M (Professor at VIT Chennai).

The main aim of the project is to implement the proposed idea from the paper.

The main aim of the paper was to find a better-quality solution for Travelling Salesman Problem using Cuckoo Search Algorithm efficiently. Parallelizing to be done to increase the quality of result. To implement parallelism across the nodes, MPI to be used. A Beowulf cluster to be created to achieve parallelism. Combinatorial optimization problems are typically NP-hard, and thus very challenging to solve. In the paper, the random key cuckoo search (RKCS) algorithm for solving the famous Travelling Salesman Problem (TSP) is discussed. A simplified random-key encoding scheme to pass from a continuous space (real numbers) to a combinatorial space is to be used. The Discrete Cuckoo Search optimization code is to be run on a single system with multiple cores and on the Beowulf cluster.

A thorough analysis has been performed to assess the efficacy of this optimization and the results are compared from the results in the paper and the reasons for such deviations is concluded.

INTRODUCTION

Many combinatorial optimization problems are NP-hard, and thus very challenging to solve. In fact, they cannot be solved efficiently by any known algorithm in a practically short time scale when the size of the problem is moderately large. The main difficulty arises with the number of combinations which increases exponentially with the size of the problem. Searching for every possible combination is extremely computationally expansive and unrealistic. An example of these problems is the travelling salesman problem in which a salesperson has to visit a list of cities exactly once, and returning to the departure city, with the aim of minimizing the total travelled distance or the overall cost of the trip.

For these problems, optimization is used to find the 'best' solution. Optimization plays an important role in solving various engineering problems which are Non-deterministic in nature. The goal of the optimization process is to determine either a maximum or a minimum value of the problem being solved, generally known as the objective function. These problems include, but not limited to, systems design, electricity network operation, electricity generation, wireless communications routing and minimization of energy losses during electricity transmission. Proper measures of efficiency of optimization algorithms require assessment of computational time and convergence rate in addition to the accuracy to determine the minimum or maximum values.

Approximate algorithms such as metaheuristics [2] are actually the best choice to solve many combinatorial optimization problems. They are characterized by their 1 simplicity and flexibility while demonstrating remarkable effectiveness. Metaheuristics are usually simple to implement; however, they are often capable to solve complex problems and can thus be adapted to solve problems with diverse objective function properties, either continuous, discrete or mixed, including many real-world optimization problems, from engineering to artificial intelligence.

Brute force often has high computational complexity and is unsuitable for large datasets. Greedy algorithm makes greedy choices at every step. It makes the choice that is best at that moment. This approach often gives only an average solution and not the optimal solution. The greedy algorithms fail especially when global optimal solution is required. Heuristics are used to solve and quickly find solutions by performing optimized search.

So, the question is how to treat combinatorial problems properly without losing the good performance of these metaheuristics. In this paper, we propose the random-key cuckoo search (RKCS) algorithm using the random-key encoding scheme to represent a position, found by the cuckoo search algorithm, in a combinatorial space.

Cuckoo Search Optimization (CSO) is one such meta-heuristic algorithm which replicates brood parasitism found cuckoos. The eggs of the bird represent the potential solutions. Gradually, the algorithm replaces the poor solutions with new and better solutions. The CSO has several applications in various fields such as neural networks, job scheduling, etc. One such problem which can be solved with Cuckoo Search Optimization is Travelling Salesman Problem

A distributed system is a system whose components are located on different networked computers, which communicate and coordinate their actions by passing messages to one another. Cluster computing is a type of distributed system. A computer cluster is a set of connected computers that work together to perform a single task. They can be tightly coupled or loosely coupled.

CUCKOO SEARCH ALGORITHM

Some cuckoo species can have the so-called brood parasitism as an aggressive reproduction strategy. This most studied and discussed feature is that cuckoos lay eggs in a previously observed nest of another species to let host birds hatch and brood their young cuckoo chicks. From the evolutionary point of view, cuckoos aim to increase the probability of survival and reduce the probability of abandoning eggs by the host birds. The behavior of cuckoos is mimicked successfully in the cuckoo search algorithm, in combination with L'evy flights to effectively search better and optimal survival strategy. L'evy flights, named by the French mathematician Paul L'evy, represent a model of random walks characterized by their step lengths which obey a power-law distribution. Several scientific studies have shown that the search for preys by hunters follows typically the same characteristics of L'evy flights. This model is commonly presented by small random steps followed occasionally by large jumps.

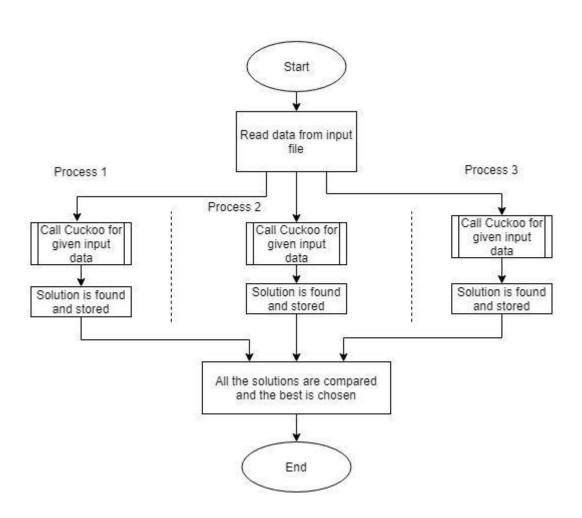
CS is summarized as the following ideal rules:

- 1. Each cuckoo lay one egg at a time and selects a nest randomly.
- 2. The best nest with the highest quality of egg can pass onto the new generations.
- 3. The number of host nests is fixed, and the egg laid by a cuckoo can be discovered by the host bird with a probability pa $\in [0, 1]$.

PROPOSED MODEL

In this approach we have taken as a basis an improved version of CS. This improvement considers a new category of cuckoos that can engage a kind of surveillance on nests likely to be a host. These cuckoos use mechanisms before and after brooding such as the observation of the host nest to decide if the chosen nest is the best choice or not. So, from the current solution, this portion pc of cuckoos searches in the same area a new better solution via L'evy flights.

PSO suffers from partial optimization. The algorithm can get trapped in local optimum, which will lead to non-optimal solution. Probability distribution changes by iteration in PSO, hence the algorithm will be not be efficient. PSO also has a weak local search ability. Hence, Cuckoo Search Optimization is chosen to find the optimum solution in this proposed work. Since the Travelling Salesman Problem is a combinatorial optimization problem and Cuckoo Search Optimization is designed to apply for continuous optimization problem, this research proposes Discrete Cuckoo Search Optimization for solving TSP. Cuckoo search is a non-deterministic algorithm. Therefore, getting one solution may not be the best solution. The algorithm needs to be run multiple times.



Cuckoo Search Optimization executed parallelly

DISCRETE CUCKOO SEARCH OPTIMIZATION (DCSO)

Discrete Cuckoo Search Optimization algorithm uses random walks with Levy flights. Cuckoo Search is an Optimization algorithm developed by Xin-she Yang and Suash Deb in 2009. The algorithm was derived from nature of cuckoo species which lay their eggs in the nests of other birds. If the host bird finds alien eggs, i.e., cuckoo eggs, the host bird will throw these eggs or abandon the nest and create a new nest

In the Cuckoo Search Optimization an egg in a nest represents a solution, and a cuckoo egg represents a new and better solution. The final goal of the algorithm is to throw the bad solutions and replace them with better solutions. Each nest has one egg. Complex cases include multiple eggs in every nest.

Basic rules for the algorithm:

- 1. Cuckoo can lay only one egg in an iteration, and discards its egg in a random nest.
- 2. Only the best solutions will carry on to the next generation.
- 3. Probability of the host bird finding the egg by the cuckoo is higher if the solution is bad. If an egg is detected, it is thrown to a faraway nest.

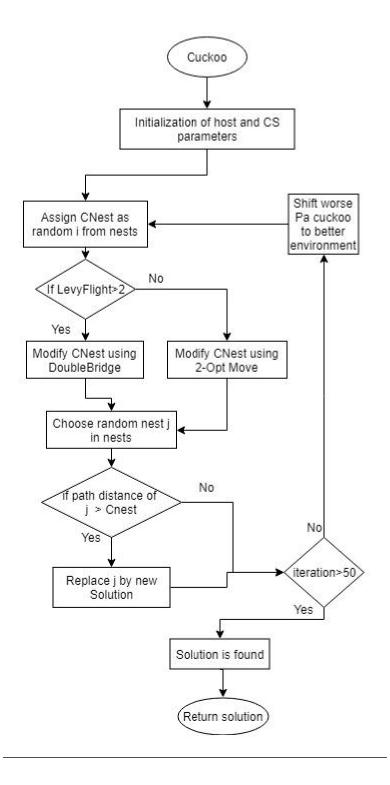
Each node selects a different starting city and computes a traversal path. Randomizing the cities selected helps in a better optimized solution. Each node where this program is run will result in different cuckoo solutions. Finally, all the solutions across the nodes are compared based on their cost and the best solution is selected and presented as the final solution.

ALGORITHM

Algorithm: Discrete Cuckoo Search Optimization

Input: Distance Matrix Output: Optimized path and cost 1: numNest = sizeof(InputMatrix) 2: Pa = 0.2 * numNest3: MaxGen = 504: Generate initial population with randomly chosen city and store in nests. 5: **while** (MaxGen) 6: Assign random index i from nests to Cnest 7: Calculate Levy Flight 8: if Levy Flight>2: Perform Double Bridge for Cnest with random values 9: else: Perform 2-Opt-Move for Cnest with random values end if 10: Get random index j from nests 11: if cost(nest j) > cost(Cnest) Replace j by the new solution; end if 12: A fraction (Pa) of the worse nests are abandoned. 13: Sort the nests based on cost end while 14: nest[0] is the solution

FLOWCHART



Cuckoo function

IMPLEMENTATION DETAILS AND RESULT ANALYSIS

In this proposed work, a new approach is proposed to solve the TSP optimization using multiple nodes and performing the computation in parallelism. By parallelizing we reduce the computational load on each of the individual machines thus improving performance of the algorithm. We have achieved communication between the nodes using Message Passing Interface (MPI, specifically mpi4py) as it is highly suitable for distributed memory models and supports all high-performance computing platforms. Parallelizing is done by splitting the work among different nodes of the cluster.

Beowulf cluster was established for parallelizing. Beowulf cluster is a multi-computer architecture which uses identical computers connected via a network and libraries, which produces a high-performance parallel cluster thus allowing for faster computation. It supports a Non uniform memory access (NUMA) architecture. The cluster follows a client-server architecture. Beowulf clusters often use cheap commodity grade computers, resulting in an inexpensive supercomputer.

First, a sequential algorithm was implemented in python, and it was then converted into a parallelized using MPI libraries designed for python (mpi4py) and then was executed on the Beowulf cluster, with the server node running the driver program. The experimental results are efficacious and are presented to demonstrate the benefits of parallelizing the discrete cuckoo search algorithm for the optimization of the travelling salesman problem.

OUTPUT

Accessing the master node: ssh master

```
mpiuser@abhishek-VirtualBox:~$ ssh master
Welcome to Ubuntu 18.04.3 LTS (GNU/Linux 5.0.0-23-generic x86_64)

* Documentation: https://help.ubuntu.com
   * Management: https://landscape.canonical.com
   * Support: https://ubuntu.com/advantage

* Canonical Livepatch is available for installation.
   - Reduce system reboots and improve kernel security. Activate at: https://ubuntu.com/livepatch

191 packages can be updated.
199 updates are security updates.

Your Hardware Enablement Stack (HWE) is supported until April 2023.
Last login: Wed Nov 6 12:49:08 2019 from 192.168.43.202
mpiuser@shantanu-VirtualBox:~$
```

Serial Implementation: 10 matrices with size = 2^n where n = 1 to 10

mpiuser@shanta	anu-VirtualBox: ~/cloud/project	
File Edit View Search Terminal Help		
CUCKOO SOLUTION from Master: ([3, 2, 0, 7, 1, 6, 14, 10, 8, 5, 15, 4, 12, 9, 13, 11], 491) Hello, World! I am process 0 of 2 on shantanu-VirtualBox.		
UUCKOO SOLUTION from Client - 1: [[4, 7, 8, 15, 2, 6, 0, 14, 13, 9, 10, 3, 12, 1, 5, 11], 472) iello, World! I am process 1 of 2 on abhishek-VirtualBox.		
INAL SOLUTION:(Best of the three solutions): [4, 7, 8, 15, 2, 6, 0, 14, 13, 9, 10, 3, 12, 1, 5, 11], 472) 0.649639630675 seconds		
	7, 8, 27, 13, 21, 4, 16, 18, 25, 26, 19, 10, 0, 2, 31], 1191)	
LUCKOO SOLUTION from Client - 1: [[15, 3, 16, 0, 4, 20, 6, 7, 29, 9, 8, 11, 12, 22, 14, 27, 2, 17, lello, World! I am process 1 of 2 on abhishek-VirtualBox.	19, 18, 5, 21, 13, 23, 1, 25, 26, 24, 10, 28, 30, 31], 1291)	

mpluser@shantanu-VirtualBox: ~/cloud/project

mpluser@shantanu-VirtualBox:~/cloud/project\$

```
1 2 0.0028781890869

2 4 0.0029230117798

3 8 0.0024788379669

4 16 0.0024328231812

5 32 0.0024781227112

6 64 0.0025179386139

7 128 0.0132279396057

8 256 0.0113651752472

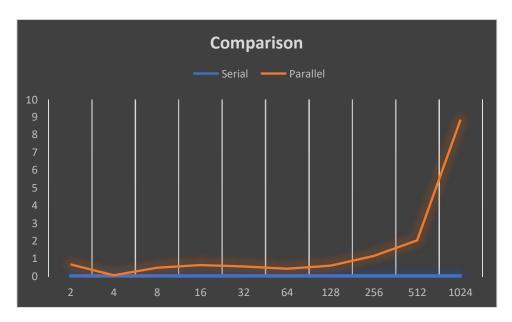
9 512 0.0353300571442

10 1024 0.0672559738159
```

Parallel Implementation: 10 matrices with size = 2^n where n = 1 to 10

--- 0.0672559738159 seconds --mpluser@shantanu-VirtualBox:-/cloud/project\$

04070806503



X-axis: matrix size Y-axis: seconds

CONCLUSION

The graph plotted for the data obtained is abrupt and far from normal as well as misguiding. However, if the matrix size increases beyond a certain point, the serial line will definitely quickly exponentiate, even faster than the parallel line. The above difference between the lines, which is huge, is due to the fact that only two nodes with dual-core processors each were used and the communication overhead between them is too high. If even powerful machines were to be used, we would be able to get outputs for the larger matrix inputs.

Thus, we have successfully implemented the research paper chosen even if we got deviated and illusioned results.