Clustering Nodes and Discretizing Movement to Increase the Effectiveness of HEFA for a CVRP

Ubassy Abdillah¹, Suyanto Suyanto² School of Computing, Telkom University Jl. Telekomunikasi No. 01, Terusan Buah Batu Bandung, West Java, Indonesia 40257

Abstract—A Capacitated Vehicle Routing Problem (CVRP) is an important problem in transportation and industry. It is challenging to be solved using some optimization algorithms. Unfortunately, it is not easy to achieve a global optimum solution. Hence, many researchers use a combination of two or more optimization algorithms, which based on swarm intelligence methods, to overcome the drawbacks of the single algorithm. In this research, a CVRP optimization model, which contains two main processes of clustering and optimization, based on a discrete hybrid evolutionary firefly algorithm (DHEFA), is proposed. Some evaluations on three CVRP cases show that DHEFA produces an averaged effectiveness of 91.74%, which is much more effective than the original FA that gives mean effectiveness of 87.95%. This result shows that clustering nodes into several clusters effectively reduces the problem space, and the DHEFA quickly searches the optimum solution in those partial spaces.

Keywords—Swarm intelligence; capacitated vehicle routing problem; firefly algorithm; differential evolution; hybrid evolutionary firefly algorithm

I. INTRODUCTION

A Vehicle Routing Problem (VRP) model plays an important role in various industrial sectors, ranging from productionbased industries to logistical issues. Various methods are used to solve this problem, which are grouped into two categories: deterministic optimization algorithms [1], [2], [3], [4], [5], and probabilistic optimization algorithms [6], [7], [8], [9].

An optimization algorithm determines such a great solution in solving VRP that finding the global optimum solution takes a long time. Not only the deterministic optimization algorithms but also the probabilistic ones have some specific problems. The deterministic algorithms guarantee to give a global optimum solution, but their processes take a long time. In contrast, the probabilistic algorithms are commonly fast, but they do not always produce a global optimum solution. In practice, probabilistic algorithms are preferable in terms of fast processing time.

Therefore, many probabilistic algorithms are developed to tackle optimization problems, such as genetic algorithm (GA) [10], [11], [12], [13], [14], particle swarm optimization (PSO) [15], [16], bee colony optimization [17], [18], cuckoo search [19], and Firefly Algorithm (FA) [20], [21], [22], [23]. There are also many new their hybrid versions or variants, such as parallel genetic algorithm [24], fuzzy optimization [25], ant colony optimization and variable neighbourhood search (ACO-VNS) [26], hybrid firefly algorithm (HEFA) [27], and Hybrid Evolutionary Firefly Algorithm (HEFA) [28].

In practice, HEFA has been proven to produce high performances for many optimization problems [28]. Hence, in this paper, a discrete version of HEFA, which is called as DHEFA, is exploited to develop a CVRP optimization model. A new idea of HEFA-based clustering is also proposed to make the model more effective in searching the minimum-cost route.

Next, the fundamental theory of CVRP and HEFA will be clearly described in Section II. The proposed models of HEFA-based clustering and DHEFA-based optimization are then explained more detail in Section III. After that, Section IV discusses the simulation results. Section V eventually provides conclusion and the further plan.

II. FUNDAMENTAL THEORY

VRP is a combinatorial optimization problem, which is an extension of a Traveling Salesman Problem (TSP) [3]. It has a basic form called Capacitated VRP (CVRP) [7]. Unlike VRP, the CVRP has an additional problem when searching for optimum vehicle order schedules. Each node visited has a load that should be accommodated, and each vehicle has a maximum capacity that cannot be violated. This not only makes the optimum solution depend on the results of vehicle scheduling but also considers the burden that each vehicle can accommodate. The total distance on the scheduling is formulated as

$$x_{\text{tot}} = \sum_{i=1}^{k} \sum_{j=1}^{c_i} x_j,$$
(1)

where x_{tot} is the total distance, k is the number of vehicles, c_i is number of nodes contained in the *i*th vehicle, and x_j is the route traversed by the *j*th vehicle.

A. Firefly Algorithm

FA is inspired by the movements of fireflies looking for a partner, which based on two things: the attraction between fireflies and the intensity of light. The light intensity is basically the value of a function. Unfortunately, the light intensity is not the same in every place. Therefore, in [20] the formula of light intensity in one firefly against the others can be formulated as

$$I = I_0 e^{\gamma r},\tag{2}$$

where I_0 is the value of fitness, γ is the value of light absorption, and r is the distance between the chasing individual and the individual being pursued in a scalar value. Just like the light intensity, the attraction is dynamic since the distance determines its change. The further the distance between the fireflies, the smaller the interest. Hence, the attractiveness function is formulated as

$$\beta = \beta_0 e^{-\gamma r^2},\tag{3}$$

where β_0 is the initial attractiveness value between two individuals and it is generally set to 1. In the original version for continuous-problem optimization, the distance is calculated using an Euclidean distance as

$$r(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2},$$
(4)

where both x and y are n-dimensional vectors.

Meanwhile, the firefly movement is calculated using the formula

$$x_i = x_i + \beta(x_j - x_i) + \alpha \epsilon_i, \tag{5}$$

where β is the attractiveness value and α is a random value from 0 to 1.

B. Differential Evolution (DE)

DE is one of the Evolutionary Algorithms (EAs) [28], where the key processes of this algorithm are mutation, crossover, and selection. The mutation process in DE uses velocity vectors from two random vectors. This velocity vector then becomes the driving force for new vectors, which are not the two previous vectors. The DE mutation formula is represented as

$$v_{i(t+1)} = x_{i(t)} + F(x_{k(t)} - x_{j(t)}), \tag{6}$$

where $v_{i(t+1)}$ is a vector of mutations, $x_{i(t)}$ is an old vector, and $F(x_{k(t)} - x_{j(t)})$ is a random vector difference from other individuals.

Meanwhile, the crossover scheme in HEFA is simply represented as [28]

$$u_{i(t+1)} = v_{i(t)}$$
 if (rand > c_r), (7)

where $u_{i(t+1)}$ is the vector of the crossover result and $v_{i(t)}$ is the result of the exchange of elements between the vector $x_{1(t)}$ and the vector $v_{i(t+1)}$ with c_r is a cross-over rate or a constant value when the element must be crossed-over.

The selection process is then formulated as

$$x_{i(t+1)} = \begin{cases} x_i, (f(x_i) > f(u_{i(t+1)})) \\ \\ u_{i(t+1)}, (f(x_i) \le f(u_{i(t+1)})) \end{cases}$$
(8)

where this process only selects between the old vector x_i and the result of the crossover vector $u_{i(t+1)}$ based on its fitness value.



C. Hybrid Evolutionary Firefly Algorithm

When building a good program of collective intelligence, the balance between both exploitation and exploration plays an important role. High exploitation makes the program converges too quickly, which is known as a premature convergence, and consequently, the program fails to find the best solution (global optimum). In contrast, too high exploration affects the program does not converge to a global optimum. The program tends to behave like a random search.

In FA, the process of balancing exploitation and exploration is more focused on regulating the values of γ and α . The α is responsible to the exploration process in the FA, which is usually a little value. The small α keeps the FA from behaving like a random search. But, at the same time, the exploration area became smaller, as illustrated in Fig. 1. A small radius α limits the movement of FA exploration. Each firefly drawn by a dark blue circle cannot explore areas outside its population. In cases where the solution space is greater than the radius of the distribution of fireflies, some areas within the solution space cannot be traced.

Nevertheless, this exploration problem can be solved using a DE. Fig. 2 shows that the DE behavior that moves based on other vectors makes DE has a significant exploration radius. With a broad reach, DE can explore even outside the population area. This feature makes it one of the reasons why DE can complement the FA.

HEFA is a combination algorithm between FA and DE. This algorithm is introduced by Afnizanfaizal Abdullah in [28]. The process of moving the algorithm is quite simple. HEFA only divides the firefly population into two parts based on their fitness values. Half of the population with high fitness values exploits the FA while the rests with poor fitness scores explore using the DE scheme. The experimental results in [28] prove that HEFA is excellent at solving complex problems and nonlinear biological models.



Fig. 3. Proposed DHEFA-based CVRP optimization model

III. PROPOSED MODEL

The proposed DHEFA-based CVRP optimization model is illustrated in Fig. 3. It receives a dataset of nodes. First, the dataset is clustered using a HEFA. The produced optimum centroids are then exploited to initialize a population of fireflies in a DHEFA, where an individual of firefly represents a candidate solution of a route. Finally, the DHEFA searches a minimum-cost route as the best solution.

The most challenging step in this optimization problem is determining the division of the number of nodes against the available vehicles. This division can be done in a purely random way, selecting nodes in sequence until reaching maximum vehicle capacity, and so on. However, clustering the nodes into n cluster, which is the same as the number of available

vehicles, is the best solution since clustering can reach the minimum total distance traveled by each vehicle.

A. Dataset of Nodes

The dataset used in this research is the Augerat et al. Set B. It has three instances: B-n50-k8 that contains 50 nodes with eight vehicles, B-n66-k9 that consists of 66 nodes with nine vehicles, and B-n78-k10 that contains 78 nodes with ten vehicles. All instances do not provide a cluster of nodes to the vehicle, which is important since it affects the total distance traveled by a vehicle. Therefore, a clustering procedure is needed to develop the optimization model.

B. HEFA-based Clustering

A HEFA-based clustering is exploited here since it has been proven to give a high performance. It is expected to produce as high possible as density cluster for each vehicle since the denser the cluster, the lower the total distance for the vehicle. Firefly at the beginning of an iteration contains a random vector with a size of two times the total vehicles. A pair of two vector elements in a firefly represents the centroid coordinates in the form (x, y).

All coordinates of centroids are then used to produce a fitness value obtained from the objective function. Half of the firefly population will move to pursue the best fitness value from its perspective while the rest move as if randomly in search of better fitness value. Once all fireflies move, they renew their respective fitness values. It is repeated until the stop condition is reached, and the HEFA produces the best firefly with the highest fitness value. An example of HEFA-based clustering a set of sixty nodes into three clusters is illustrated in Fig. 4.



Fig. 4. Example of HEFA-based clustering for three clusters

The coordinates of all centroids produced by the best firefly are then used to determine the cluster of nodes. Each node in a cluster is visited by a particular vehicle.

The objective function is simply designed here using a sum of square Euclidean distance (SSE). This function calculates the total distance of all nodes to their respective centroids, which is formulated as

$$SSE = \sum_{j=1}^{k} \sum_{x_i \in c_j} \|c_j - x_i\|_2^2,$$
(9)

where k is the number of cluster, x_i is the *i*th node, and c_j is the *j*th cluster.

Once the optimum clusters are generated, check if there is a vehicle carrying a load that exceeds the maximum capacity. Any node in an over-loaded vehicle is then redistributed to the nearest under-loaded vehicle, as illustrated in Fig. 5. The vehicle capacity *Cap* in cluster c_2 , which exceeds the maximum capacity *MaxCap*, looks for the closest vehicle to redistribute one or more nodes. A node, which is the closest to the cluster c_1 , is selected to move to the cluster c_1 .



Fig. 5. Example of redistribution nodes in the clusters



Finally, a minimum-cost route is searched using DHEFA. Since the problem of determining the route is a discrete problem (sequence of nodes should be visited), the HEFA has to be redesigned into a discrete model. In [29], a discrete firefly algorithm (DFA) is proposed with a high performance. In this paper, the discrete model of FA is designed by following the concept of DFA.

At the beginning of the iteration, a firefly in DHEFA consists of a random vector with a total size of nodes and elements in the range of one to a total non-repeating node. This vector is divided into the total number of vehicles where the nodes contained in each vehicle are following the clusters resulted from the previous HEFA-based clustering and the redistribution procedure. An example of firefly representation is illustrated in Fig. 6.

HEFA uses a distance that is determined by the difference between two firefly vectors while DHEFA calculates the distance as the number of different elements between two fireflies (also known as the Hamming distance [30]), as illustrated in Fig. 7. Another difference is the movement of fireflies. This movement does not use the sum of the *i*th firefly vector with the distance to the followed firefly as in Equation 5, but instead uses an insertion function. This function takes a random node and swaps it with another random node [29], as illustrated in Fig. 8. In this CVRP case, the insertion is limitedly performed just for two nodes in the same cluster since the vectors in fireflies are divided by the number of vehicles. It cannot exchange two elements in two different vehicles. Therefore, when choosing a random element in k_i , the second random element must be in k_i . This exchange is carried out as much as the Hamming distance $\times \gamma$.



Fig. 6. Individual representation of a firefly

$x_1 =$	1	2	3	4
<i>x</i> ₂ =	1	3	2	4

Fig. 7. Hamming distance between two fireflies



Fig. 8. Insertion function for a firefly movement

Just like HEFA, the movement of a firefly in DHEFA also depends on its fitness value that is calculated using Equation (5). Half of the firefly population chases the best fireflies from its perspective while the rest move randomly, expecting to get better fitness values. All fireflies then update their fitness values to be compared in the next iteration. When the stopping condition is reached, the best fireflies are chosen as the minimum-cost solution, as illustrated in Fig. 9.

IV. RESULTS AND DISCUSSION

In this research, the proposed DHEFA-based model is evaluated and compared with the original FA-based model using three cases of CVRP. The experiments are run five times to give a more accurate statistical result. In each case, an effectiveness metric is used here to measure how close the obtained optimum-cost route to the real global optimum-cost route from the dataset. In this evaluation, both FA and DHEFA have the same conditions of parameters: $\gamma = 0.95$, $\alpha = 0.2$, and $c_r = 0.5$. The results are listed in Table I.



Fig. 9. Minimum-cost solution given by the best firefly

TABLE I. EFFECTIVENESS (%) PRODUCED BY BOTH FA AND DHEFA FOR
FIVE RUNS PER CASE OF CVRP

Corr of CUDD		DUEEA with shorts since	
Case of CVRP	FA without clustering	DHEFA with clustering	
B-n50-k8	94.49	96.00	
	90.71	93.71	
	80.71	94.43	
	91.12	90.87	
	94.11	96.24	
Average	90.23	94.25	
B-n66-k9	92.51	96.55	
	96.52	94.70	
	90.52	90.58	
	89.81	95.39	
	91.99	92.00	
Average	92.27	93.84	
B-n78-k10	83.27	89.45	
	82.09	85.20	
	85.00	84.53	
	80.07	90.31	
	76.36	86.13	
Average	81.36	87.13	

In all cases, DHEFA produces higher effectiveness than the original FA. In the CVRP case of B-n50-k8, with 50 nodes and eight vehicles, DHEFA produces an averaged effectiveness up to 94.25% while the original FA just gives 90.23%. In the CVRP case of B-n66-k9, which contains 66 nodes and nine vehicles, DHEFA also reaches a higher averaged effectiveness of 93.84%, but the FA just obtains 92.27%.

Meanwhile, in the CVRP case of B-n78-k10, with 78 nodes and ten vehicles available, DHEFA gets much higher averaged effectiveness of 87.13% while the original FA yields 81.36% only. Thus, for the three cases, DHEFA reaches much higher averaged effectiveness of 91.74% than the original FA that just obtains 87.95%.

This effectiveness of DHEFA is highly supported by the procedure of clustering nodes. Dividing nodes into some clusters is capable of reducing the problem space in some areas so that the optimization can be partially applied. This concludes that the research objective stated in Section I has been reached.

V. CONCLUSION

The proposed model of DHEFA-based CVRP optimization is capable of reaching the averaged effectiveness of 91.74%. This result is better than the original FA that gives mean effectiveness of 87.95%. This fact shows that the proposed clustering significantly increases the effectiveness of DHEFA. It can be simply explained that clustering nodes into some clusters is capable of reducing the problem space in some areas so that the optimization can be partially applied. In the future, an advanced procedure of redistribution can be introduced to ensure all vehicles have fair loads as well as do not violate the maximum capacity.

REFERENCES

- P. Sitek, J. Wikarek, and K. Rutczyńska-Wdowiak, "Capacitated vehicle routing problem with pick-up, alternative delivery and time windows (CVRPPADTW): A hybrid approach," *Advances in Intelligent Systems* and Computing, vol. 1004, pp. 33–40, 2020.
- [2] C. Archetti, M. G. Speranza, M. Boccia, A. Sforza, and C. Sterle, "A branch-and-cut algorithm for the inventory routing problem with pickups and deliveries," *European Journal of Operational Research*, vol. 282, no. 3, pp. 886–895, 2020.
- [3] G. Marques, R. Sadykov, J.-C. Deschamps, and R. Dupas, "An improved branch-cut-and-price algorithm for the two-echelon capacitated vehicle routing problem," *Computers and Operations Research*, vol. 114, 2020.
- [4] C. Tilk and A. Goel, "Bidirectional labeling for solving vehicle routing and truck driver scheduling problems," *European Journal of Operational Research*, vol. 283, no. 1, pp. 108–124, 2020.
- [5] N. Smiti, M. M. Dhiaf, B. Jarboui, and S. Hanafi, "Skewed general variable neighborhood search for the cumulative capacitated vehicle routing problem," *International Transactions in Operational Research*, vol. 27, no. 1, pp. 651–664, 2020.
- [6] I. Sbai, S. Krichen, and O. Limam, "Two meta-heuristics for solving the capacitated vehicle routing problem: the case of the Tunisian Post Office," *Operational Research*, 2020.
- [7] J. Lysgaard, A. D. López-Sánchez, and A. G. Hernández-Díaz, "A matheuristic for the MinMax capacitated open vehicle routing problem," *International Transactions in Operational Research*, vol. 27, no. 1, pp. 394–417, 2020.
- [8] J.-Q. Li, Y.-Q. Han, P.-Y. Duan, Y.-Y. Han, B. Niu, C.-D. Li, Z.-X. Zheng, and Y.-P. Liu, "Meta-heuristic algorithm for solving vehicle routing problems with time windows and synchronized visit constraints in prefabricated systems," *Journal of Cleaner Production*, vol. 250, 2020.
- [9] F. Oudouar, M. Lazaar, and Z. E. Miloud, "A novel approach based on heuristics and a neural network to solve a capacitated location routing problem," *Simulation Modelling Practice and Theory*, vol. 100, 2020.
- [10] C. O. Alenoghena, J. O. Emagbetere, and A. M. Aibinu, "Artificial intelligence based technique for BTS placement," in *IOP Conference Series: Materials Science and Engineering*, vol. 53, no. 1, Kuala Lumpur, 2013.
- Suyanto, An informed genetic algorithm for university course and student timetabling problems, 2010, vol. 6114 LNAI, no. PART
 [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-13232-2_28
- [12] N. R. Emillia, Suyanto, and W. Maharani, "Isolated Word Recognition Using Ergodic Hidden Markov Models and Genetic Algorithm," *TELKOMNIKA*, vol. 10, no. 1, pp. 129–136, 2012. [Online]. Available: http://dx.doi.org/10.12928/telkomnika.v10i1.769
- [13] R. M. Karim and S. Suyanto, "Optimizing Parameters of Automatic Speech Segmentation into Syllable Units," *International Journal on Intelligent Systems and Applications*, vol. 11, no. 5, pp. 9–17, 2019. [Online]. Available: http://www.mecs-press.org/ijisa/ijisa-v11n5/IJISA-V11-N5-2.pdf

- [14] S. Suyanto and A. E. Putra, "Automatic Segmentation of Indonesian Speech into Syllables using Fuzzy Smoothed Energy Contour with Local Normalization, Splitting, and Assimilation," *Journal of ICT Research and Applications*, vol. 8, no. 2, pp. 97–112, 2014. [Online]. Available: http://journals.itb.ac.id/index.php/jictra/article/download/443/528
- [15] G. Shaari, N. Tekbiyik-Ersoy, and M. Dagbasi, "The state of art in particle swarm optimization based unit commitment: A review," *Processes*, vol. 7, no. 10, 2019.
- [16] E. Annan Boah, O. Kwami Senyo Kondo, A. Aidoo Borsah, and E. T. Brantson, "Critical evaluation of infill well placement and optimization of well spacing using the particle swarm algorithm," *Journal of Petroleum Exploration and Production Technology*, vol. 9, no. 4, pp. 3113–3133, 2019.
- [17] A. Pertiwi and Suyanto, Globally evolved dynamic bee colony optimization, 2011, vol. 6881 LNAI, no. PART 1. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-23851-2_6
- [18] V. Clarissa and S. Suyanto, "New Reward-Based Movement to Improve Globally-Evolved BCO in Nurse Rostering Problem," in 2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), dec 2019, pp. 114–117. [Online]. Available: https://ieeexplore.ieee.org/document/9034669
- [19] G. K. Jati, H. M. Manurung, and Suyanto, "Discrete cuckoo search for traveling salesman problem," in 2012 7th International Conference on Computing and Convergence Technology (ICCCT), 2012, pp. 993–997. [Online]. Available: https://doi.org/10.1007/s00521-013-1402-2
- [20] X.-S. Yang, "Firefly algorithm, stochastic test functions and design optimization," *International Journal of Bio-Inspired Computation*, vol. 2, no. 2, pp. 78–84, 2010. [Online]. Available: https://www.inderscienceonline.com/doi/pdf/10.1504/IJBIC.2010.032124
- [21] G. K. Jati, R. Manurung, and Suyanto, Discrete Firefly Algorithm for Traveling Salesman Problem: A New Movement Scheme. Elsevier Inc., 2013. [Online]. Available: http://dx.doi.org/10.1016/B978-0-12-405163-8.00013-2
- [22] X.-S. Yang, Z. Cui, R. Xiao, A. H. Gandomi, and M. Karamanoglu, *Swarm Intelligence and Bio-Inspired Computation*. Elsevier Inc., 2013. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-

s2.0-85013816320&doi=10.1016%2FC2012-0-02754-8&partnerID=40&md5=8f6271973c61d2eddabc47ad1048ed46

- [23] F. Ghaisani and S. Suyanto, "Discrete Firefly Algorithm for an Examination Timetabling," in 2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), dec 2019, pp. 1–4. [Online]. Available: https://ieeexplore.ieee.org/document/9034668
- [24] M. Abbasi, M. Rafiee, M. R. Khosravi, A. Jolfaei, V. G. Menon, and J. M. Koushyar, "An efficient parallel genetic algorithm solution for vehicle routing problem in cloud implementation of the intelligent transportation systems," *Journal of Cloud Computing*, vol. 9, no. 1, 2020.
- [25] S. Z. Aliahmadi, F. Barzinpour, and M. S. Pishvaee, "A fuzzy optimization approach to the capacitated node-routing problem for municipal solid waste collection with multiple tours: A case study," *Waste Management and Research*, vol. 38, no. 3, pp. 279–290, 2020.
- [26] Y. Wu, F. Pan, S. Li, Z. Chen, and M. Dong, "Peer-induced fairness capacitated vehicle routing scheduling using a hybrid optimization ACO–VNS algorithm," *Soft Computing*, vol. 24, no. 3, pp. 2201–2213, 2020.
- [27] A. M. Altabeeb, A. M. Mohsen, and A. Ghallab, "An improved hybrid firefly algorithm for capacitated vehicle routing problem," *Applied Soft Computing Journal*, vol. 84, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494619305095
- [28] A. Abdullah, S. Deris, M. S. Mohamad, and S. Z. M. Hashim, "A new hybrid firefly algorithm for complex and nonlinear problem," *Advances in Intelligent and Soft Computing*, vol. 151 AISC, pp. 673–680, 2012. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-28765-7_81
- [29] G. K. Jati and Suyanto, "Evolutionary Discrete Firefly Algorithm for Travelling Salesman Problem," in *ICAIS*, no. 1, 2011, pp. 393–403. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-23857-4_38
- [30] Suyanto and A. Harjoko, "Nearest neighbour-based Indonesian G2P conversion," *Telkomnika (Telecommunication, Computing, Electronics,* and Control), vol. 12, no. 2, pp. 389–396, 2014. [Online]. Available: http://dx.doi.org/10.12928/telkomnika.v12i2.57