An Efficient Genetic Algorithm for the Uncapacitated Single Allocation Hub Location Problem

Mohammad Naeem and Beatrice Ombuki-Berman

Abstract— **Hub location problem is a NP-hard problem that frequently arises in the design of transportation and distribution systems, postal delivery networks, and airline passenger flow. We propose a simple but effective genetic algorithm (GA) for the uncapacitated single allocation hub location problem (USAHLP). Our main contribution is two new simple chromosome encoding schemes based on indirect representation and two crossover operators. We performed an empirical study to evaluate the effectiveness of the proposed GA using well-known benchmark problems from the Civil Aeronautics Board (CAB) and Australian Post (AP) data sets. The GA found all best-known solutions for the 80 CAB problems and introduced new solutions for the larger problem instances for AP data. The proposed GA can easily be extended to other variants of location problems arising in network design planning in transportation and distributed systems.**

I. INTRODUCTION

Hub Location Problems (HLPs) are classical combinatorial optimization problems that arise in telecommunication and transportation networks where nodes send and receive commodities (i.e., data transmissions, passengers, express packages, mail, etc.) through special facilities or transhipment points called hubs. Hubs consolidate flows from origin nodes and re-route them to destination nodes sometimes via other hubs. The sending and receiving nodes in such networks are called spokes. The networks are called hub-spoke networks. The assumption in hub-spoke networks is that, hubs are fully-connected through low-cost high-volume pathways that allow a discount factor to be applied to the transportation cost of the flow between a given hub pair. Another assumption in these networks is that, all the internodal flow takes place through at least one hub and at most two.

Due to their wide applicability and economic importance in determining efficient distribution strategies to reduce operational costs, variants of HLPs have been studied with increased interest. A comprehensive survey on HLPs and their classification can be found in Kara *et. al.* [8]. In this paper, we focus on hub-and-spoke networks which have wide applicability in many areas including passenger airlines [9, 10, 11], express package delivery firms [12], message delivery networks [13], trucking industry [14], telecommunication systems [21], supply-chain of chain stores such as Wallmart [22], and many other areas.

This work is supported in part by the Natural Sciences and Engineering Council of Canada (NSERC). This support is gratefully acknowledged.

Mohammad Naeem is with the Department of Computer Science, Brock University, St.Catharines, Ontario, Canada (email: mnaeem198@yahoo.com).

Beatrice Ombuki-Berman is with the Department of Computer Science, Brock University, St.Catharines, Ontario, Canada (email:bombuki@brocku.ca@brocku.ca).

A hub-and-spoke network typically involves three simultaneous decisions: deciding the optimal number of hub nodes, their locations and the appropriate allocation of the non-hub nodes to the hubs [6] with an overall objective of minimizing the total flow cost which consists of both fixed and variable costs. Although different variants of hub-and spoke networks exist, a common goal in all variants is to establish the location of hubs, and the allocation of spokes such that the total flow cost is minimized. Generally, the variations arise due to various considerations such as, imposing capacity limits on the amount of volume at a given hub, allowing single or multiple allocation of non-hub nodes to the hubs, and whether the number of selected hubs is known a priori or is left as decision variable.

In this paper, we focus on the uncapacitated single Allocation hub location problem (USAHLP) which places no capacity constraints on hubs, assigns each non-hub node to exactly one hub, and all traffic flows via the hub nodes. If the number of hubs is fixed a priori, e.g., is equal to p, the problem is then known as the uncapacitated single allocation p-hub median problem (USApHMP). Surveys on different applications related to USAHLP, and the classifications of its variants are found in [5], [15], [16] and [17]. Although the USAHLP is more commonly encountered in practice [1], the USApHMP has attracted more attention than the USAHLP, and thus, one aim of our work is to bridge this gap by proposing an application of genetic algorithms approach for USAHLP.

The SAHLP, whether the capacitated or uncapaciated version is NP-hard [8]. The combinatorial explosion is obvious, and obtaining exact optimal solutions for this type of problems is computationally intractable [23]. Thus, we can rarely accomplish optimal networks within reasonable time for large problems, and meta-heuristics which do not guarantee optimal solutions, but seek good approximate solutions within practical time are often relied on. Due to their usefulness and economic importance, both the capacitated and uncapacitated versions of SAHLP have received a good amount of research attention where both exact and heuristic methods have been proposed to tackle them. Some of these methods include a quadratic integer programming formulation [17] and its linearization [8] , Genetic Algorithm (GA) [3], a hybrid heuristic combining GA and Tabu Search [2], and a Simulated Annealing (SA) and Tabu Search (TS) based hybrid solution method [4] for the USAHLP. The recent work [1] by M. R Silva and C.B. Cunha using a two-stage tabu search approach offer a comprehensive empirical analysis for the USAHLP which other researchers (including our work) can compare their work against.

978-1-4244-8126-2/10/$26.00 ©2010 IEEE

Although a few papers exist in the literature proposing GAs for USAHLP, the performance of the published GA work has been outperformed by other meta-heuristics, such as tabu search [1]. Abdinour-Helm [2] proposed a hybrid approach based on GA and Tabu Search to solve the USAHLP. The GA was used to determine the number and location of hubs and the Tabu Search (TS), to assign spokes to hubs. They reported an improvement over their earlier GA-approach that used distance-based assignment of spokes to hubs. However, their stand-alone GA results are not available. Topcuoglo *et al.* [3] developed a GA-based approach to the USAHLP. They found improved solutions to some Civil Aeronautics Board (CAB) problems. They also used Australian Post (AP) data in their experiments that had not been previously used in any study on USAHLP. Another GA-based study on the USAHLP cited by Kara *et al.* [8] is a hybrid approach by Cunha and Silva [18] that employed GA and Simulated Annealing.

The non-GA heuristics applied to the USAHLP include two hybrid approaches by Chen *et al.* [4] and Silva *et. al.* [1]. Chen *et al.* [4] combined SA with Tabu List(TL) to solve USAHLP. This approach involves applying Simulated Annealing to determine an upper-bound for the number of hubs and then using restricted single location exchange procedure to locate the hubs. Non-hub nodes are first allocated to nearest hubs followed by an improvement procedure for allocation that iteratively re-allocates nodes with less flow to other hubs until no improvement is possible.

In this paper, we propose a new simple but effective GA approach for the uncapacitated SAHLP. We employ an indirect solution encoding based on two schemes. Furthermore, we propose two problem-specific crossovers for the USAHLP. The approach adopted in these crossovers treat clusters, i.e., a hub with associated spokes, as units of gene exchange between the mating parents instead of individual nodes as in the existing GAs for the USAHLP. We probabilistically employed three mutation operators in the GA, i.e., the *shift mutation*, the *swap mutation*, and the *change hub mutation*. The shift and swap mutations have been used in previous GA studies on SAHLP. We perform an empirical study to evaluate the effectiveness of our GA by using two sets of well-known benchmark problems derived from real-world applications.

The performance of the GA on both sets of the benchmark problems is encouraging. The performance of the proposed GA is better than that of GA [3] and GATS [2] on CAB problems and comparable with that of MSTS-3 [1] and HubTS [1] and SATLUHLP [4] on AP problems. The remainder of this paper is organized as follows: Section II gives a formal description of the USAHLP. The details of the proposed GA is given in Section III, followed by experimental discussions in Section IV. Finally the concluding remarks and future work is provided in Section V.

II. UNCAPACITATED SINGLE ALLOCATION HUB LOCATION PROBLEM

The uncapacitated single allocation hub location problem is a special type of hub location problem in which no capacity limits are associated with hubs and a spoke can be assigned to only a single hub. Moreover, the number of hubs is a decision variable in SAHLP and a fixed cost for establishing a hub is also included in the overall transportation cost.

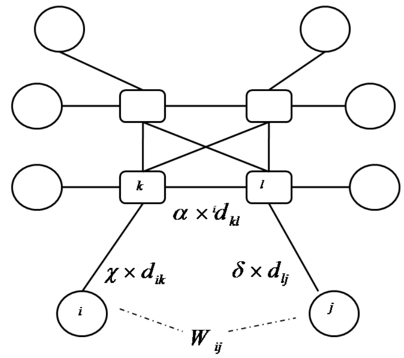

Fig. 1. A single allocation hub-spoke network

The objective in the SAHLP is to minimize the cost of establishing hubs and cost of transportation. This is subject to the constraints that a spoke must be assigned to only a single hub, and flows must be routed only through hubs (at least one and at most two). The transportation cost in the USAHLP has the following three components.

- The *Collection cost*, χ, is the cost incurred on flow from a given spoke to a hub, i.e., cost of spoke-to-hub flow.
- The *Transfer cost*, α, represents the cost of the flow between hubs, i.e., cost of hub-to-hub flow.
- The *Distribution cost*, δ, denotes the cost of the flow from a hub to a spoke, i.e., cost of hub-to-spoke flow.

All the cost types are per unit distance of flow volume between nodes. For example, assume that in Figure 1, W_{ij} volume of a commodity is sent by node i to node j. W_{ij} is first transported from node i to hub k, then from hub k to hub l, and finally from hub l to the destination node j. The net transportation cost C_{ijkl} is,

$$C_{ijkl} = W_{ij} \quad (\chi d_{ik} + \alpha d_{kl} + \delta d_{lj})$$

Where d_{ik} is the distance between node i and hub k, d_{kl} is distance between hubs k and l, and d_{lj} is the distance between hub l and node j. In order to find the transportation cost of the entire network, C_{ijkl} is calculated for all the node pairs in the network. The cost of establishing the required hubs is also included in the total cost. Mathematical formulation for the USAHLP was first given by O'Kelly [6] as a quadratic 0-1 optimization problem with linear constraints. Other mixed integer programming (MIP) formulations were also proposed [5,7]. This work uses the quadratic integer programming formulation by O'Kelley, which is given below.

$$Minimize \quad \sum_{i\in N_N}\sum_{k\in N_N}\sum_{l\in N_N}\sum_{j\in N_N} W_{ij}(\chi d_{ik}+ \alpha d_{kl}+\delta d_{lj})X_{ijkl}+\sum_{k\in N} F_k Z_{kk}$$

Subject to:

$$\sum_{k\in N_N}\sum_{l\in N_N} X_{ijkl}=1, \qquad \forall i,j\in N_N, \qquad (1)$$

$$Z_{ik}\le Z_{kk}, \qquad \forall i,k\in N_N, \qquad (2)$$

$$\sum_{j\in N_N}\sum_{l\in N_N}(W_{ij}X_{ijkl}+W_{ji}X_{jilk}) = (O_i+D_i)Z_{ik} \qquad \forall i,k\in N_N, \qquad (3)$$

$$Z_{ik}\in\{0,1\}, \qquad \forall i,k\in N_N, \qquad (4)$$

$$0\le X_{ijkl}\le 1, \qquad \forall i,j,k,l\in N_N, \qquad (5)$$

Where:

$O_i=\sum_{j\in N_N} W_{ij}$
$D_i=\sum_{j\in N_N} W_{ji}$
N is the number of nodes.
$N_N=\{0,1,2,...,N-1\}$
W_{ij} is the flow between the origin i and destination j.
χ is the *collection* cost (from origin spoke to hub).
α is the *transfer* cost (between hubs).
δ is the *distribution* cost (from hub to destination spoke).
d_{ik} represents the distance between nodes i and hub k.
d_{kl} is the distance between hubs k and l.
d_{lj} is the distance between hub l and node j.
X_{ijkl} is the decision variable that represents the fraction of traffic between origin node i to destination node j through hubs k and l.
F_i is the cost of establishing node i as hub.
Z_{ij} is 1 if node i is assigned to hub j, otherwise it is 0.
Z_{kk} is 1 if node k is also a hub, otherwise it is 0.

Constraint (1) ensures that all the traffic between an origin-destination pair has been routed via the hub sub-network. Constraint (2) prevents non-hub nodes from being allocated to other non-hub nodes while Constraints (3) restricts the commodity flow through each hub. For some hub-spoke networks, e.g., a mail delivery system, the flow may not be symmetric, i.e., $W_{ij}\neq W_{ji}$. Additionally, a node may route flow to itself, i.e., $W_{ii}>0$. In this work, both symmetric and non-symmetric flows are considered.

III. GA METHODOLOGY FOR THE USAHLP

This section presents the details of the chromosome representation, fitness evaluation and genetic operators used. In the GA, each chromosome in the randomly generated population pool is transformed into a hub-spoke network. The chromosomes are then subjected to an evolutionary process until a minimal cost hub-spoke network is evolved or the termination condition is met. The evolutionary process is carried out like in ordinary GA using genetic and selection operators on chromosomes as depicted in Figure 2. Tournament selection with elite retention is used to perform fitness-based selection. Two new solution representation schemes and two new problem-specific crossover operators are proposed in this section for the USAHLP. The GA also probabilistically employs three mutation operators.

```
Read problem instance data
Set GA parameters
Generate an Initial Population
    while(not termination)
        GenerateSolutionsNetwork()
        Evaluate()
        Selection()
        Reproduction()
        Mutation()
    end while
end procedure
```

Fig. 2. An Outline of the genetic Meta-Heuristic for USAHLP

A. Chromosome Encoding and Initial Population Creation

We propose two indirect representation schemes, i.e., *List-based Representation* and *Set-based Representation*, to encode the solution structure of the USAHLP. Note that the words List and Set are used here for differentiation purpose, and do not necessarily represent an actual list or set, as traditionally known. For both approaches, an indirect chromosome representing a hub-spoke network is given an integer string of length N, where N is the total number of nodes in a given problem instance.

1) Set-based Representation: Figure 3 shows how a chromosome is created from a set of initial N nodes, and its corresponding hub-spoke network solution. This hub-spoke network has three clusters, i.e., $C_1=\{\mathbf{12},5,10,2\}$, $C_3=\{\mathbf{1},6,3,4,11\}$, and $C_2=\{\mathbf{8},9,7,0\}$.

The first and bolded numbers in each set is a hub and the remaining nodes are spokes associated with the hub. The creation of an individual is performed in three steps. In the first step, *m, number of hubs* is determined randomly with the GA's initial maximum number of hubs being half the number of nodes in the network and the minimum is 2 (future work will let the GA evolve the number of hubs without starting with a determined upper limit). In the second stage, i.e., *Location Step*, m hubs are randomly chosen from N nodes (N is the total number of nodes in the network). In this way, any node from 1 to N has the chance to become a hub. Lastly, i.e., in the *Allocation Step*, the remaining $N-m$ nodes are allocated to the selected hubs using the distance-based assignment rule, i.e., a given node is assigned to a hub that has the shortest (Euclidean) distance from the hub-node. The above process is applied iteratively to create the entire initial population.

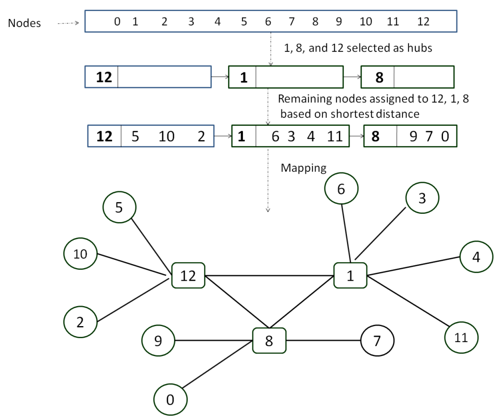

Fig. 3. Set-based chromosome creation with corresponding hub-spoke network

2) List-based Representation: In the list-based encoding, a solution is represented by a list with n hub entries, where n is the number of nodes in the network. The entries are also implicitly indexed by numbers from 0 to $n-1$ that represent spokes. Thus a hub entry in the list is indexed by one of the spokes assigned to the hub. This representation scheme is illustrated in the Figure 4. The network in the Figure 4a has 13 nodes including pre-designated hubs, i.e. 1, 8, and 12. Thus its list representation contains 13 entries as shown in Figure 4b . Every list entry is a hub i.e., either 1, 8, or 12. The hub entries are numbered 0 to 12, such that 0 serves as an index to the first value in the list, 1 to the second value, 2 to the third value, and so on. The hub entry at position 0 of the list is 8, which means spoke 0 is assigned to hub 8. Similarly, hub at position 1 is 1 meaning spoke 1 is assigned to hub 1 and at position 2 is 12 indicating that spoke 2 is allocated to hub 12, etc. In SAHLP formulation, a hub is considered to be assigned to itself. This is indicated by storing values 1, 8, and 12 at positions 1, 8, and 12 of the list.

B. Chromosome Fitness Evaluation

Once each chromosome has been transformed into a feasible hub-spoke network topology the fitness value of each chromosome is determined by using a weighted-sum fitness function. The fitness of an individual $F(x)$ is returned as:

$$F(x) = \sum_{i\in N}\sum_{l\in N}\sum_{k\in N}\sum_{j\in N} W_{ij}(\chi d_{ik} + \alpha d_{kl} + \delta d_{lj})X_{ijkl} + \sum_{k\in N} F_k Z_{kk}$$

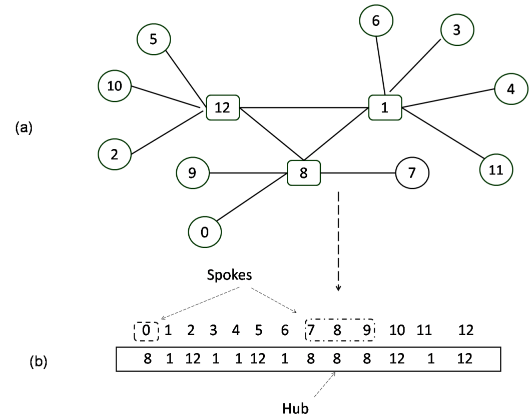

Fig. 4. List-based chromosome

In the above function, the first term represents the cost incurred on the internodal flow and the second term the cost of establishing the selected nodes as hubs. The function as a whole represents the total transportation cost of the network.

C. Reproduction

At every generational stage, the GA selects parents for mating and reproduction. The tournament selection strategy with elite [24] retaining model is used to generate a new population. The tournament selection strategy is a fitness-based selection scheme that works as follows. A set of K individuals are randomly selected from the population. This is known as the tournament set, and the tournament size employed here was 4. An elite model is incorporated to ensure that the best individual is carried on into the next generation.

D. Recombination phase

An important aspect for the successful use of a genetic algorithm is the role that the recombination (usually in the form of a crossover operator) plays. In the Single Allocation Hub Location Problem, combinations of hubs and hub-spoke assignment patterns constitute the building blocks of the solution. Further, the fitness contribution of a cluster in terms of minimizing the objective function depends on the distance and flow between spokes and the hub in the cluster.

Based on the above observations, two problem-specific crossovers were designed for the SAHLP that process clusters instead of individual nodes. A cluster in this context is a hub and its allocated spokes. In these crossovers, one or more clusters are exchanged between the mating parents to produce feasible offspring.

1) Multi-Cluster Exchange Crossover(MCEC): In the Multi-Cluster Exchange Crossover (inspired by the crossover used in vehicle routing by Ombuki-Berman *et al. [20]*), children solutions are produced by swapping one or more randomly selected clusters between the mating parents. The swapping process is followed by a re-adjustment process in which infeasible solutions are corrected. If a hub in a cluster from one parent (i.e., the source parent) is also a hub in a

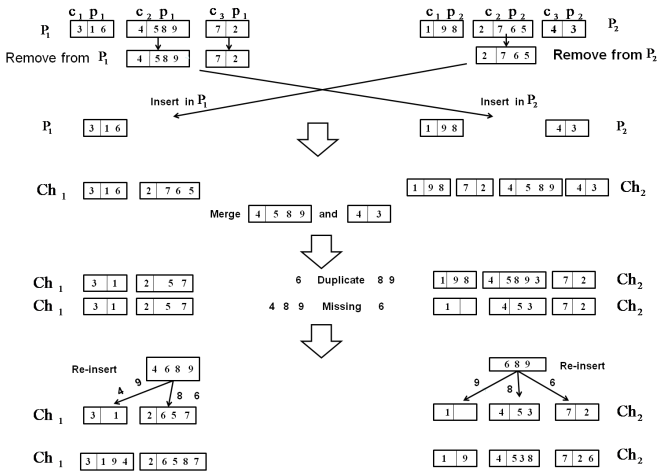

Fig. 5. Multi-Cluster Exchange Crossover (MCEC)

cluster of the other parent (i.e. the destination parent), then both clusters are merged. Duplicate or missing nodes in a child solution resulting from this process are re-assigned based on distance, i.e., a node is assigned to the nearest hub. For illustration, consider two parent solutions P_1 and P_2 with their respective networks in Figure 5 selected for cross-breeding. Multi-Cluster Exchange Crossover (MCEC) is applied to P_1 and P_2 to produce two children solutions Ch_1 and Ch_2 as shown in Figure 5.

2) Double-Cluster Exchange Crossover(DCEC): In Double-Cluster Exchange Crossover (DCEC), two random clusters are iteratively selected from one parent solution and shifted to the other. The same operation is repeated for the other parent. Duplicate nodes in the offspring resulting from the recombination operation are detached from their present hubs and re-assigned to other hubs of the same offspring according to distance i.e., a node is assigned to the closest hub. Likewise, nodes lost by an offspring due to the swap operation are re-inserted in it based on distance.

The process is illustrated In Figure 6. There are two parent solutions, $P_1 = \{\{3,1,2\},\{4,5,8,9\},\{7,6\}\}$ with hubs 3, 4, 7 and $P_2 = \{\{1,9,8\},\{2,7,6,5\},\{3,4\}\}$ with hubs 1,2, and 3 are crossbred using Double-Cluster Exchange Crossover (DCEC).

IV. Mutation

The *Shift Node*[2][3], *Swap Nodes*[2][3], and *Change-Hub* mutations are used probabilistically. In the *Shift Node* mutation, a random node is detached from one cluster and inserted into another random destination cluster. The shift mutation operation can be performed only for clusters with more than one node. In *Swap Node* mutation, two clusters are randomly selected from the given solution and one random node from the first cluster is shifted to the second cluster. Likewise, from the second cluster, a randomly selected node is shifted to the first cluster. In the *Change-Hub* mutation, the hub from a randomly selected cluster of a solution is demoted as spoke whereas a spoke from the same cluster is promoted as hub. The operations are illustrated in the Figure 7.

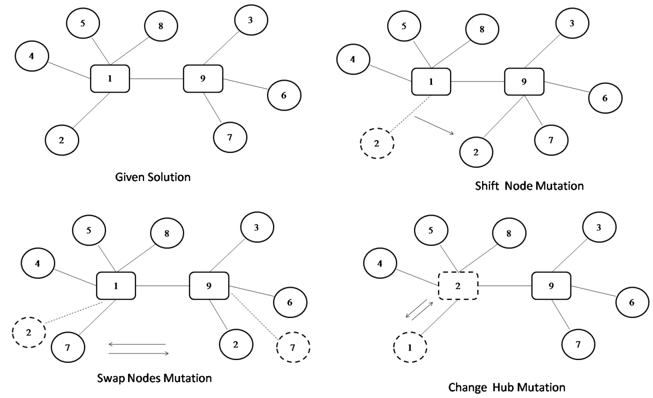

Fig. 6. Mutation Operations

V. Experimental Results and Discussions

The proposed GA which was implemented in Java 1.3, on a Pentium IV 1.5 GHz PC with 512 MB memory on Windows 2007. To evaluate the computational effectiveness of the proposed GA, an empirical study with two versions of the GA, i.e., GA-1 and GA-2 was performed. Each version was based on one of the crossovers introduced in Section 3, that is: GA-1 based on Double-Cluster Exchange Crossover (DCEC) and GA-2 based Multi-Cluster Exchange Crossover (MCEC).

A. Data Sets

Two benchmark data sets for hub location problems, i.e., Civil Aviation Board (CAB) data set [17] and Australian Post (AP) data set [19] were used.

B. CAB Data Set

The CAB benchmark data set by O'Kelley [17] is based on air traffic between 25 cities in USA and has been used extensively as benchmark for the uncapacitated hub location problems. The data set contains test problem instances of 10, 15, 20, and 25 nodes for uncapacitated hub location problems. Unlike the AP data with asymmetric flows between nodes, the internodal flow (W_{ij}) in CAB data set is symmetric i.e., $W_{ij} = W_{ji}$ and is scaled by division with the total network flow, i.e.,

$$\sum_{i=1,j=1}^{i=n,j=n} W_{ij}$$

The unit collection cost χ and unit distribution cost δ in the data set are both fixed at 1.0 [2][3][4]. The transfer cost α i.e., the cost for hub-to-hub flow, is varied between 0.2 and 1.0 to provide discount factors for bulk transportation between hubs [2][3][4].

C. AP Data

The AP data set was introduced by Ernst *et. al.* [19] and is based on a real application to postal delivery system in Australia. AP data is the only data benchmark data set available for capacitated hub location problems. It has also been used by some studies for uncapacited problems [3][4], and we also adopt it here. The set contains problems of up to 200 nodes with each node representing a postal district. The problem sizes are 10, 20, 25, 40, 50, 100, and 200 nodes. The internodal flows in AP data set are asymmetric, i.e., $W_{ij} \neq W_{ji}$. The data set contains hub costs and hub capacities for capacitated hub location problems. The unit collection and distribution costs i.e., χ and δ, in the data set are 3.0 and 2.0 respectively and the discount factor, α, is 0.75.

The AP data set [1] for the SAHLP involves two types of problems. The loose-cost (L) problems have less variation in the hub-cost whereas the tight-cost (T) problems have more variation in the hub-cost. Tight-cost (T) problems tend to be more difficult than the loose-cost (L) problems. These two types of problems have been denoted here by the notation nF where n stands for the number of nodes in the problem and F denotes the cost-type i.e., loose, "L" or tight, "T". For example, notation "10L" denotes the loose-cost (L) problem with 10 nodes.

D. Parameter Setting

Obtaining good parameter settings for a given problem is a crucial factor in the performance of a GA. It is now generally accepted that optimum parameter settings may be problem specific, implying that the GA being designed must first be parameterized in the context of a particular problem [20]. The parameter settings for the GA-1 and GA-2, shown in Table 1 were empirically established. The experiments were based on 30 runs of each of the above GA versions. The execution of the GA was terminated after 1000 generations or when there was no change in fitness for 150 generations. Furthermore, after a mutation decision was made according to the mutation rate in the table above, probabilities of 0.2, 0.6, and 0.2 were used to select one of the mutation types i.e., *shift*, *swap*, or *replace-hub* mutation respectively.

TABLE I

EXPERIMENTAL PARAMETERS

Parameter	GA-1	GA-2
population size	500	500
population	generational	√
chromosome initialization	random	√
generational span	1000	√
probability of crossover	0.55	0.60
probabliity of mutation	0.2, 0.4	0.4, 0.2

The tick marks in the following tables means a best-known solution has been found by the proposed GA, while the bolded values in tables indicated a new best solution has been found. Table II compares the performance of GA-1 and GA-2 in terms of solution quality for both set based and list based representation using the AP data. The results show that both GA-1 and GA-2 are quite efficient for the USAHLP since not only did they find all the best currently known solution, they also found improved solutions for the larger problem instances of 200. In conclusion, we can say based on Table II, that both the proposed crossovers are equally suitable for use with any of the proposed chromosome encoding for the USAHLP, even though the GA-2 seems to perform slightly better than GA-1 for the list representation.

TABLE II

COMPARISON OF GA-1 AND GA-2, AP DATA

Problem	Known best [1]	GA-1 Set	GA-2 Set	GA-1 List	GA-2 List
10L	224250.05	√	√	√	√
20L	234690.95	√	√	√	√
25L	236650.62	√	√	√	√
40L	240986.23	√	√	√	√
50L	237421.98	√	√	√	√
100L	238016.28	√	√	√	√
200L	233803.02	**233802.97**	**233802.97**	233837.42	**233802.97**
10T	263399.94	√	√	√	√
20T	271128.18	√	√	√	√
25T	295667.84	√	√	√	√
40T	293164.83	√	√	√	√
50T	300420.98	√	√	√	√
100T	305097.96	√	√	√	√
200T	272237.78	**272188.11**	**272188.11**	272193.55	**272188.11**

TABLE III

COMPARISON WITH OTHER METHODS, SET REPRESENTATION, AP DATA

Problem	Known best[1]	MSTS-3 [1]	HubTS [1]	GA-1 Set	GA-2 Set
10L	224250.05	√	√	√	√
20L	234690.95	√	√	√	√
25L	236650.62	√	√	√	√
40L	240986.23	√	√	√	√
50L	237421.98	√	√	√	√
100L	238016.28	√	√	√	√
200L	233803.02	√	√	**233802.97**	**233802.97**
10T	263399.94	√	√	√	√
20T	271128.18	√	√	√	√
25T	295667.84	√	√	√	√
40T	293164.83	√	√	√	√
50T	300420.98	√	√	√	√
100T	305097.96	√	√	√	√
200T	272237.78	√	√	**272188.11**	**272188.11**

Figure 8 shows that the proposed GA outperforms other published work by introducing two new improved solutions. The detailed solution quality of the proposed GA as compared to published work is given in Tables III and IV for set based and list based solution representations, respectively.

As can be seen from Tables V to VIII the proposed GA found optimal or the best-known solution for all the problem instances given for CAB data. Based on the performance of the proposed GA for both the available AP and CAB data, we can thus conclude that this GA has been effective in handling the USAHLP. Lastly Figure 8 shows the performance of the proposed GA compared with other published work. Due to space limitation, we do not include the averages of the 30 runs for both the AP and CAB data in this paper.

TABLE IV

COMPARISON WITH OTHER METHODS, LIST REPRESENTATION, AP DATA

Problem	Known best[1]	MSTS-3 [1]	HubTS [1]	GA-1 List	GA-2 List
10L	224250.05	√	√	√	√
20L	234690.95	√	√	√	√
25L	236650.62	√	√	√	√
40L	240986.23	√	√	√	√
50L	237421.98	√	√	√	√
100L	238016.28	√	√	√	√
200L	233803.02	√	√	233837.42	**233802.97**
10T	263399.94	√	√	√	√
20T	271128.18	√	√	√	√
25T	295667.84	√	√	√	√
40T	293164.83	√	√	√	√
50T	300420.98	√	√	√	√
100T	305097.96	√	√	√	√
200T	272237.78	√	√	272193.55	**272188.11**

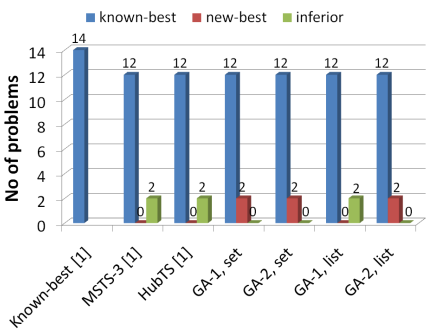

Fig. 7. AP Data: Comparison with other methods in terms of known-best or new-best solutions found

VI. CONCLUSIONS

Hub location forms an important line of research because of its high complexity and intractable nature, as well as its importance in distribution and transportation management. A literature review showed that although the single allocation hub Location problem (SAHLP) is more commonly encountered in practice, the single allocation p-hub median problem (SApHMP) has attracted more attention than the USAHLP. Secondly, the literature review showed that although there exists published work focusing on the development of heuristics for related distribution and transportation problems, little work is reported on the use of GA for the SAHLP. Thus, we aim to further contribute to the development of efficient meta-heuristics for the SAHLP and related problems.

This paper presented the preliminary results of the proposed genetic algorithm for the uncapacitated SAHLP. Our main contribution was to focus on efficient chromosome representation and genetic operators. We proposed two new simple, but efficient solution encoding schemes and two crossover operators. Our proposed genetic algorithm found all best known solutions for the 12 AP problems and improved upon two of these solutions while it found the optimal results for all the 80 publicly available problems for the CAB data. This work suggests that there is much potential to extend the GA to other variants of the SAHLP and related problems. Preliminary work shows that this can be extended to the capacitated SAHLP. Future work involves further evaluating the GA's effectiveness using larger problems, and a thorough statistical analysis of the performance of the proposed GA.

TABLE V

USAHLP: COMPARISON WITH OTHER APPROACHES, $n = 10$, CAB DATA

α	f	Optimal Cost	GA [3] GATS [2] SATLUHLP[4] MSTS-3 [1] HubTS [1]	GA-1 Set	GA-2 Set	GA-1 List	GA-2 List
0.2	100	791.93	√	√	√	√	√
	150	915.99	√	√	√	√	√
	200	1015.99	√	√	√	√	√
	250	1115.99	√	√	√	√	√
0.4	100	867.91	√	√	√	√	√
	150	974.30	√	√	√	√	√
	200	1074.30	√	√	√	√	√
	250	1174.30	√	√	√	√	√
0.6	100	932.62	√	√	√	√	√
	150	1032.62	√	√	√	√	√
	200	1131.05	√	√	√	√	√
	250	1181.05	√	√	√	√	√
0.8	100	990.94	√	√	√	√	√
	150	1081.05	√	√	√	√	√
	200	1131.05	√	√	√	√	√
	250	1181.05	√	√	√	√	√
1.0	100	1031.05	√	√	√	√	√
	150	1181.05	√	√	√	√	√
	200	1131.05	√	√	√	√	√
	250	1181.05	√	√	√	√	√

TABLE VI

USAHLP: COMPARISON WITH OTHER APPROACHES, $n = 15$, CAB DATA

α	f	Optimal Cost	GA [3] GATS [2] SATLUHLP[4] MSTS-3 [1] HubTS [1]	GA-1 Set	GA-2 Set	GA-1 List	GA-2 List
0.2	100	1030.07	√	√	√	√	√
	150	1239.77	√	√	√	√	√
	200	1381.28	√	√	√	√	√
	250	1481.28	√	√	√	√	√
0.4	100	1179.71	√	√	√	√	√
	150	1355.09	√	√	√	√	√
	200	1462.62	√	√	√	√	√
	250	1556.66	√	√	√	√	√
0.6	100	1309.92	√	√	√	√	√
	150	1443.97	√	√	√	√	√
	200	1506.66	√	√	√	√	√
	250	1556.66	√	√	√	√	√
0.8	100	1390.76	√	√	√	√	√
	150	1456.66	√	√	√	√	√
	200	1506.66	√	√	√	√	√
	250	1556.66	√	√	√	√	√
1.0	100	1406.66	√	√	√	√	√
	150	1456.66	√	√	√	√	√
	200	1506.66	√	√	√	√	√
	250	1556.66	√	√	√	√	√

REFERENCES

[1] M. R. Silva and C. B. Cunha, *New simple and efficient heuristics for the uncapacitated single allocation hub location problem*, Computers and Operations Research 36 (2009)3152–3165.

[2] S. Abdinour-Helm, *A hybrid heuristic for the uncapacitated hub location problem*, European Journal of Operational Research 106 (1998)489–499

TABLE VII

USAHLP: COMPARISON WITH OTHER APPROACHES, $n = 20$, CAB DAT

α	f	Optimal Cost	GA [3] GATS [2] SATLUHLP[4] MSTS-3 [1] HubTS [1]	GA-1 Set	GA-2 Set	GA-1 List	GA-2 List
0.2	100	967.74	√	√	√	√	√
	150	1174.53	√	√	√	√	√
	200	1324.53	√	√	√	√	√
	250	1474.53	√	√	√	√	√
0.4	100	1127.09	√	√	√	√	√
	150	1297.76	√	√	√	√	√
	200	1442.56	√	√	√	√	√
	250	1542.56	√	√	√	√	√
0.6	100	1269.15	√	√	√	√	√
	150	1406.04	√	√	√	√	√
	200	1506.04	√	√	√	√	√
	250	1570.91	√	√	√	√	√
0.8	100	1369.52	√	√	√	√	√
	150	1469.52	√	√	√	√	√
	200	1520.91	√	√	√	√	√
	250	1570.91	√	√	√	√	√
1.0	100	1410.07	√	√	√	√	√
	150	1470.91	√	√	√	√	√
	200	1520.91	√	√	√	√	√
	250	1570.91	√	√	√	√	√

TABLE VIII

USAHLP: COMPARISON WITH OTHER APPROACHES, $n = 25$, CAB DATA

α	f	Optimal Cost	GA [3] GATS [2] SATLUHLP[4] MSTS-3 [1] HubTS [1]	GA-1 Set	GA-2 Set	GA-1 List	GA-2 List
0.2	100	1029.63	√	√	√	√	√
	150	1217.34	√	√	√	√	√
	200	1367.34	√	√	√	√	√
	250	1500.90	√	√	√	√	√
0.4	100	1187.51	√	√	√	√	√
	150	1351.69	√	√	√	√	√
	200	1501.62	√	√	√	√	√
	250	1601.62	√	√	√	√	√
0.6	100	1333.56	√	√	√	√	√
	150	1483.56	√	√	√	√	√
	200	1601.20	√	√	√	√	√
	250	1701.20	√	√	√	√	√
0.8	100	1458.83	√	√	√	√	√
	150	1594.08	√	√	√	√	√
	200	1690.57	√	√	√	√	√
	250	1740.57	√	√	√	√	√
1.0	100	1556.63	$\surd^{++}$	1559.19	√	1559.19	√
	150	1640.57	√	√	√	√	√
	200	1690.57	√	√	√	√	√
	250	1740.57	√	√	√	√	√

$^{++}$ GA [3] and GATS [2] values for the $\alpha = 1.0$ and $f = 100$ case are 1559.19 and 1562.16 respectively

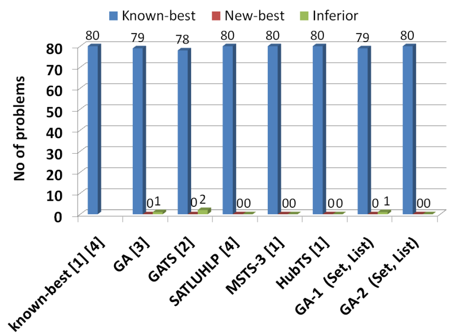

Fig. 8. USAHLP, CAB Data: Comparison with other methods in terms of known-best or new-best solutions found

[3] H. Topcouglu, F. Corut, M. Ermis, G. Yilmaz, *Solving the uncapacitated hub location problem using genetic algorithms*, Computers and Operations Research 32 (2005) 967–984.

[4] Jeng-Fung Chen, *A hybrid heuristic for the uncapacitated single allocation hub location problem*, Omega 35 (2007)211–220

[5] Campbell JF, *Integer programming formulations of discrete hub location problems*, European Journal of Operational Research 72 (1994)387– 405

[6] O'Kelly M., *Hub facility location with fixed costs*, The Journal of the Regional Science Association International 71 (1992)293–306

[7] Ebery J., *Solving large single allocation p-hub problems with two or three hubs*, European Journal of Operational Research 128 (2001)447–58

[8] Amur Sibel, Kara Bahar, *Network Hub Location Problem: State of the Art*, European Journal of Operational Research 190 (2008) 1-21

[9] Aykin T., *Networking policies for hub-and-spoke systems with application to the air transportation system*, Transportation Science 29 (1995) 201-221

[10] Bania, N. P. Bauer, T. Zlatoper, *US air passenger service: a taxonomy of route networks, hub locations, and competition*, Logistics and Transportation Review 34 (1998) 53-74

[11] Skorin-Kapov D., Skorin-Kapov J., *On tabur search for the location of interacting hub facilities*, European Journal of Operational Research 190 (2008) 1-21

[12] Kuby M., R. Gray, *Hub network design problems with stopovers and feeders: case of Federal Express*, Transportation Research 27 (1993) 1-12

[13] Klencewicz J., *Hub location in backbone tributory network design: a review*, Location Science 6 (1998) 307-335

[14] Don T., S. Harit, J. English, G Whisker, *Hub and spoke networks in truckload trucking: configuration, testing, and operational concerns*, Logistics and Transportation 31 (1995) 209-237

[15] O'Kelley ME, Miller H., *The hub network design problem: a review of synthesis*, Journal of Transport Geography 2(1) (1994) 31-40

[16] Cambell J. F., Ernst AT., Krishanmoorthy M., *Hub location problems In Drezner Z, Hamacher HW, editors, Facility location applications and theory Berlin* Springer; 2002.p.373408.

[17] O'Kelley M. E., *A Qudratic integer program for the location of interacting hubs*, European Journal of Operational Research 32(1987) 393-404

[18] Cunha, C.B., Silva, M.R., *A genetic algorithm for the problem of configuring a hub-and-spoke network for a LTL trucking company in Brazil*, European Journal of Operational Research 179(2007)747-758.

[19] T. Ernst Andreas, Krishanmoorthy Mohan, *Solution Algorithms for the Capacitated Single Allocation Hub Location Problem*, Annals of Operations Research 86(1999) 141-159

[20] Ombuki-Berman B. M., Runka A., Hanshar F. T., *Waste Collection Vehicle Routing Problem with Time Windows using Multi-Objective Genetic Algorithms*, From Proceeding (574) Computational Intelligence, 2007

[21] Lee Y, Lim B, Park J. *A hub location problem in designing digital data service networks: Lagrangian relaxation approach*, Location Science 4(1996)185-94

[22] Abdunnour-Helm S., *Using simulated annealing to solve the p-hub median problem*, International Journal of Physical Distribution and Logistics Management 31(2001)203-20.

[23] Garey M.R, and Johnson D.S, *Computers and Intractability, A Guide to The Theory of NP-Completeness*, W. H. Freeman and Company, 1979.

[24] Goldberg D.E, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison Wesley, 1989.