Genetic Algorithms Applied to Reverse Distribution Networks

Freitas, A. R. R., Silva, V. M. R., Guimarães, F. G., and Campelo, F.

Abstract Reverse Distribution Networks are designed to plan the distribution of products from customers to manufacturers. In this paper, we study the problem with two-levels,with products transported from origination points to collection sites before being sent to a refurbishing site. The optimization of reverse distribution networks can reduce the costs of this reverse chain and help companies become more environmentally efficient. In this paper we describe heuristics for deciding locations, algorithms for defining routes, and problem-specific genetic operators. The results of a comparative analysis of 11 algorithms over 25 problem instances suggest that genetic algorithms hybridized with simplex routing algorithms were significantly better than the other approaches tested.

1 Introduction

When a manufacturer distributes a product, it is usually assumed that the distribution from producer to customers is the only chain that has to be optimized. However,

---

Freitas, A. R. R.
Programa de Pós-Graduação em Engenharia Elétrica - Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, Belo Horizonte, MG, 31270-901 Brazil, e-mail: alandefreitas@gmail.com

Silva, V. M. R.
Instituto de Computação, Universidade Federal Fluminense
Niterói, RJ, 24210-240, Brazil, e-mail: victormrsilva@gmail.com

Guimarães, F. G.
Departamento de Engenharia Elétrica - Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, Belo Horizonte, MG, 31270-901 Brazil, e-mail: fredericoguimaraes@ufmg.br

Campelo, F.
Departamento de Engenharia Elétrica - Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, Belo Horizonte, MG, 31270-901 Brazil, e-mail: fcampelo@ufmg.br

there are many cases in which the products need to be returned to the manufacturer to be replaced, repaired, or recycled (e.g., defective or environmentally hazardous products). Reverse distribution networks are defined by the chain in which the products are returned to the manufacturer. It is an interesting problem as very often the cost of the reverse chain can overtake by many times the price of distributing the product to customers [16, 18]. Besides, environmentally friendly products have become a marketing element for companies [25].

In this work we study a formulation for reverse logistics which is based on two levels [11]. In this area of reverse logistics, the first level represents collection sites, where the products can be temporarily stored until they are sent to refurbishing sites, represented by the second level. This formulation of the problem can be reduced to another NP-Complete problem, and heuristics have been developed to solve it. Besides the known heuristics for the problem, we study the possibilities of using genetic algorithms for the solution of this class of problems.

A model for the two-level reverse distribution network problem takes into account the cost, limits and capacities of each site as well as the costs of transferring the products. A solution for a given problem includes which facilities are open as well as the route for the products. Heuristics for deciding the open facilities, such as greedy heuristics or even genetic algorithms, usually work in parallel with algorithms that calculate the routes for the products.

This work is organized as follows: Section 2 provides a brief description of reverse distribution networks (RDNs), and in Section 3 a mathematical formulation for the two-level RDNs model is presented. Section 4 presents a number of heuristics used for the solution of different aspects of the problem. In Section 5 the proposed GA for the solution of the two-level RDN problem is discussed. Section 6 describes the experimental setup and the statistical analysis of the results, which shows that genetic algorithms performed significantly better than the heuristics on the test set used, particularly when hybridized with the simplex algorithms. Finally, Section 7 presents some final considerations and ideas for continuity.

2 Reverse Distribution Networks

In a distribution system, the reverse chain represents the return of products to the manufacturer, e.g., for repair or recycling. The formulation considered in this work focuses on product recall in which the products are initially located at outlets [11]. The products go from the customer to a collection site or straight to a refurbishing site, forming a two-level problem. Besides the cost of transporting the products, each collection point and each refurbishing site has a cost for being active. There are many variations of this problem, with different products or collection layers [15].

The problem has important implications for retailers who need to consider the reverse chain. Retailers have to be prepared for a number of expectations, risks and impacts that make it important to consider recommendations at the management level [9].

Many different models have been proposed for reverse distribution [7, 2, 24] and different solutions were proposed, such as heuristics [7, 3, 10, 1, 14], linear programming [23], evolutionary computation [12], and Markov chains [9]. For the formulation considered in this work [11], when there are no collection sites, the problem is reduced to the Capacitated Facilities Location Problem, which is NP-Complete [6]. The difficulty of finding an optimal solution for large instances in polynomial time justifies the use of heuristics for the problem.

3 Modeling the Problem

The general model for the problem [11] is described in this section. The objective and constraint functions use the following definitions:

- $I\{i/i\}$, $J\{j/j\}$, $K\{k/k\}$: origination, collection, and refurbishing sites, respectively;
- C_{ijk}: cost of transporting an unit from the origination site i to the refurbishing site k through the collection site j;
- F_j, G_k: cost of opening the collection site j or the refurbishing site k;
- a_i: number of products at the site i;
- B_j, D_k: maximum capacity of the collection site j or of the refurbishing site k;
- P_{min}, P_{max}: minimum and maximum number of collection sites to open;
- Q_{min}, Q_{max}: minimum and maximum number of refurbishing sites to open.

The objective and constraint functions for this problem are given in Equation (1).

$$
\begin{aligned}
\text{minimize } & \sum_i \sum_j \sum_k C_{ijk} a_i X_{ijk} + \sum_j F_j P_j + \sum_k G_k Q_k \\
\text{subject to } & \sum_j \sum_k X_{ijk} = 1 \text{ for all } i \\
& \sum_i \sum_k a_i X_{ijk} \le B_j \text{ for all } j \\
& \sum_i \sum_j a_i X_{ijk} \le D_k \text{ for all } k \\
& X_{ijk} \le P_j \text{ for all } i,j,k \\
& X_{ijk} \le Q_k \text{ for all } i,j,k \\
& P_{min} \le \sum_j P_j \le P_{max} \text{ for all } j \\
& Q_{min} \le \sum_k Q_k \le Q_{max} \text{ for all } j \\
& 0 \le X_{ijk} \le 1 \text{ for all } i,j,k \\
& P_j \in \{0,1\} \text{ for all } j \\
& Q_k \in \{0,1\} \text{ for all } k
\end{aligned}
\tag{1}
$$

Each solution of the problem is defined by:

- $X_{ijk} \rightarrow$ fraction of units at originating site i that will be transported through the sites j and k ($j = 0$ is used to indicate that the products go straight to k)
- $P_j \rightarrow$ if the collection site j is open, $Q_k \rightarrow$ if the refurbishing site j is open

4 Heuristics

The mathematical model for the two-level reverse distribution network problem defines costs, possible locations, capacity of each facility, number of products located in many points, maximum and minimum number of each sort of facility, and cost of transport from each origination site to each refurbishing site through each collection site. Different heuristics are described to decide the open facilities and to calculate the routes for the products.

4.1 Deciding Locations

A simple idea for deciding the open facilities is a greedy heuristic. Based on a greedy principle, we propose the following heuristic to generate solutions:

1. Rank order all collection sites P_j and refurbishing sites Q_k by their capacity/cost;
2. Sort an integer between p_{min} and p_{max} and save it as p_{goal}. Sort an integer between q_{min} and q_{max} and save it as q_{goal};
3. The p_{goal} collection sites with highest capacity/cost value are open. The q_{goal} collection sites with highest capacity/cost value are open.

This simple greedy heuristic can generate and explore solutions with different numbers of open facilities in the two levels of the model. The heuristic can be used by itself or as part of another meta-heuristic.

Another known method used for deciding the open facilities is Heuristics Concentration [20, 11]. That can be done by running an heuristic a number of times in order to identify facilities which are worth further investigation.

Inspired by the heuristics concentration originally proposed for the p-median problem [20], a similar approach has been proposed for the reverse logistics problem [11]:

1. Random Selection: random selection of potential collection and refurbishing sites. For a number of iterations (100 in this work), a subset of size P_{max} of collection sites and Q_{max} refurbishing sites is randomly selected and the routing of products for these sites is solved to by another algorithm. All solutions are saved and the best solution is marked;

2. Heuristics Concentration: the sites most used in the best solutions are added to the best solution found in random selected phase in order to form a new solution and the problem is solved to optimality. This new solution is compared to the best solution from the random selection phase;
3. Heuristics Expansion: add each unused site to the best solution and solve the problem. If a better solution is found, remember this solution and its configuration, but leave the best solution unchanged. Check if any of the other unused sites give a better solution than the new solution found. Repeat this until all the unused sites are checked. Stop when no better solution is found.

4.2 Routing Algorithms

Given the open facilities for each level, we need to define the routes for the products. A greedy heuristic can also be used in this task, to find solutions of good quality in reasonable computing times:

- For each origin site to be examined:
  1. Among the valid routes for the products, find the one with lowest cost;
  2. Send as many products as possible through this route and update its capacity;
  3. If there are still products in the origination site, send them through the next best route. Otherwise, examine the next origin site.

This approach does not guarantee that the optimal route is found, but it has the advantage of being considerably less computationally expensive than exact methods.

Another solution for finding routes from origination to refurbishing sites would be to define a second mathematical model for the problem where the collection and refurbishing sites are considered part of the problem, for which the solution would be only the routes for the products. As we consider that the sites P and Q are no longer decision variables in the routing problem, a simplified formulation arises (2):

$$
\begin{aligned}
\text{minimize } & \sum_i \sum_j \sum_k C_{ijk} a_i X_{ijk} \\
\text{subject to } & \sum_j \sum_k X_{ijk} = 1 \text{ for all } i \\
& \sum_i \sum_k a_i X_{ijk} \leq B_j \text{ for all } j > 0 \\
& \sum_i \sum_j a_i X_{ijk} \leq D_k \text{ for all } k \\
& X_{ijk} \leq P_j \text{ for all } i, j, k \\
& X_{ijk} \leq Q_k \text{ for all } i, j, k \\
& 0 \leq X_{ijk} \leq 1 \text{ for all } i, j, k
\end{aligned}
\tag{2}
$$

This formulation is a linear-programming problem that can be solved by a simplex algorithm to find the optimal values X_{ijk} in polynomial time.

5 Genetic Algorithm

Having all those heuristics as reference, we propose genetic algorithms (GA) that can evolve the solutions generated by the heuristics. In order to efficiently evolve the solutions, problem specific genetic operators are defined and the heuristics are used as approaches to generate new solutions.

In the proposed GAs, each individual is coded as two binary vectors, that represent the variables P, and Q (section 3). The constraints of the problem are all automatically satisfied by the genetic operators developed, which simplifies the solution of the problem by the GAs.

In the proposed approach, the GA searches for good P and Q values, while the heuristics presented in Section 4.2 are used to compute the value of X.

If a special case in which the capacities of the facilities are not enough to keep the products from the origin sites, the fitness of unfeasible solutions are scaled so that they are always worse than the feasible solutions.

The final fitness of the individuals is given by their rank [13]. A stochastic ranking selection [21] is used to decide which parents will generate children.

Besides the fitness functions, the genetic operators are also important to implicitly filter out undesirable solutions [8]. In this work we employ a crossover operator that keeps the open facilities that would probably lead to a good fitness. The number of open facilities of the child is an integer value proportional to the fitness of the parents. In the child's genotype, all facilities that are open in the intersection of both parents are kept open. Other facilities are then taken from the union of the parents facilities until the goal amount of open facilities for the child is met. Note that this operator does not allow the generation of solutions that disrespect the constraints that involve P_{min}, P_{max}, Q_{min}, and Q_{max}.

Mutation in this work works by simply opening or closing a facility at random.

The initial probability of crossover for each individual *ind* was set as $cp_{ind} = 90\%$, while the initial probability of mutation was $mp_{ind} = 5\%$. Those values were adapted [26] at every generation for each individual. In each generation, the whole population is replaced by their children.

The genetic algorithm was used with a population of 40 individuals that were initialized either (i) randomly, (ii) with a greedy heuristic (Section 4.1), or (iii) with Heuristics Concentration (also described in Section 4.1).

6 Experiments

6.1 Instances

The originating, collecting and refurbishment sites were randomly placed on a 100×100 square. Then, based on an existing methodology for creating instances [11], the following values are used:

- $F_j = 0.1([1,10000] + B_j[0,10])$
- $G_k = 0.1([1,25000] + B_j[0,100])$
- $C_{ijk} =$ Euclidian distance from i to j to k
- $a_i = [0,500]$
- $B_j = [0,6000]$
- $D_k = [0,30000]$

The notation $[n_1, n_2]$ represents a uniformly distributed random number between n_1 and n_2, and the parameters $|I|$, $|J|$, P_{max}, $|K|$, Q_{max}, used for generating 5 instance sets, are respectively (i) 30, 14, 4, 12, 2; (ii) 40, 20, 6, 15, 4; (iii) 50, 30, 6, 20, 4; (iv) 70, 30, 6, 20, 4; (v) 100, 40, 8, 30, 6. For each of those 5 parameter configurations, 5 instances were generated. The instances and the best objective function values found in this work are available from the authors[1] and although some parameters could be adjusted in order to represent more realistic problems, the parameters were defined in a way to make comparison with other works possible [11, 4].

6.2 Experimental Design

A comparative experiment was designed to evaluate the performance of GA-based methods relative to other heuristics commonly used in the solution of two-level reverse distribution network problems. This experiment consists in the application of 11 different methods on 25 problem instances, 5 from each instance configuration. The algorithms were used as levels of the experimental factor, and the problems were treated as experimental blocks [17].

We use the following notation for the algorithms compared in this paper: Greedy Heuristic (`Gr`), Concentrations Heuristic (`CH`), Routing Heuristic (`RH`), Simplex Routing (`SR`), Genetic Algorithms (`GA`), Random Solution Generation (`RS`). We compared four non-evolutionary heuristics and seven GA-based heuristics, as described in Table 1. Methods A-D are the heuristics that generate solutions and routes. Methods E-J are Genetic Algorithms with different heuristics for generating the initial solutions (`RS`, `Gr` or `CH`) and calculating the best routes (`RH` or `RH`). Method K is a different configuration of Genetic Algorithm[2] proposed by [4] as efficient for

[1] `http://www.alandefreitas.com/downloads/problem-instances.php`

[2] Binary tournament selection, fusion crossover, swap node mutation, replacement of only half the previous generation, population size $2.|J|$, crossover probability 0.5, mutation probability 0.1

reverse logistics. Twenty replicates of the experiment were performed with a time limit of 3 minutes for Methods E-K.

Table 1 Methods Compared

Method	A	B	C	D
Heuristics	`Gr+RH`	`Gr+SR`	`CH+RH`	`CH+SR`

Method	E	F	G	H	I	J	K
Heuristics	`GA+RS+RH`	`GA+RS+SR`	`GA+Gr+RH`	`GA+Gr+SR`	`GA+CH+RH`	`GA+CH+SR`	`GA2`

For all tests performed, the predefined significance threshold was set as 95%, adjusted using the Bonferroni correction [22]. Also, rank transformation was employed in order to reduce the influence of outliers and heteroscedasticity in the analysis of the data obtained [17].

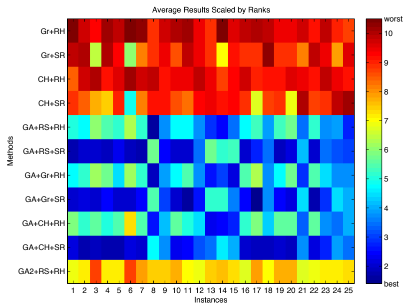
6.3 Analysis of Results

Figure 1(a) shows the difference of mean objective value obtained by the methods in each instance. For better visualization, the values were scaled by average rank. A Friedman test of the data detected highly significant $(p < 10^{-15})$ differences in algorithm performance across the test set used. To evaluate the differences, pairwise testing was performed using FDR-corrected Wilcoxon Signed-Rank tests [5, 19], and bootstrap confidence intervals were derived for the average ranks of each algorithm. The rank effect sizes for each method are presented in Figure 1(b), after removing the problem effects and the overall mean.

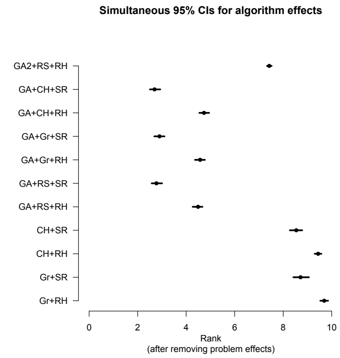
From the results it should be clear that the methods can be divided into five groups, with statistically significant differences between groups but not within them. From best to worst, these groups contain: (i) genetic algorithms based on simplex routing (F, H, J) and (ii) genetic algorithms with the routing heuristic (E, G, I), with both groups presenting above-average performance (mean rank smaller than 0). The other three, below-average groups are composed of: (iii) a less efficient genetic algorithm (K); the heuristics based on simplex routing (B,D); and the heuristics with the routing heuristic (A,C).

7 Conclusion and Future Work

By comparing heuristics for the two-level reverse distribution problem, we conclude that genetic algorithms, specially the ones based on simplex for calculating



(a) Estimated average rank of the methods on each instance



(b) Estimated average ranks and confidence intervals for the methods tested, integrated over the 25 instances.

Fig. 1 Comparison between the methods

the routes, can improve significantly the results when compared to other heuristics proposed for the problem.

Different initial populations, however, do not seem to strongly influence the performance of the genetic algorithms as all of them had comparable results at confidence level $\alpha = 0.05$. On the other hand, all the parameters of the genetic algorithm, such as population size and probability of the operators, could be better explored in order to achieve better results.

Ideas for future works include: (i) comparisons to other algorithms, both nature-inspired and heuristics (ii) tests with Genetic Algorithms for calculating the routes, (iii) search for globally optimal solutions for the proposed instances, (iv) evaluation of other genetic operators, (v) parallel evaluation of solutions, (vi) tests on larger and preferably real-world instances, and (vii) tests with larger time limits.

Acknowledgements This work has been supported by the Brazilian agencies CAPES, CNPq (grants 472446/2010-0 and 305506/2010-2), and FAPEMIG (project APQ-04611-10); and the Marie Curie International Research Staff Exchange Scheme Fellowship within the 7th European Community Framework Programme.

References

1. Alshamrani, A., Mathur, K., Ballou, R.: Reverse logistics: simultaneous design of delivery routes and returns strategies. Computers & Operations Research **34**(2), 595–619 (2007)
2. Barros, A., Dekker, R., Scholten, V.: A two-level network for recycling sand: a case study. European Journal of Operational Research **110**(2), 199–214 (1998)
3. Bautista, J., Pereira, J.: Modeling the problem of locating collection areas for urban waste management. an application to the metropolitan area of barcelona. Omega **34**(6), 617–629

(2006)
4. Costa L. R.and Galvão, R.D.: Mathematical models for reverse logistics: A genetic algorithm for a two-level problem. In: International Conference on Operational Research for Development, pp. 92–104. Fortaleza, CE, Brazil (2007)
5. Crawley, M.J.: The R Book, 1st edn. Wiley (2007)
6. Davis, P., Ray, T.: A branch-bound algorithm for the capacitated facilities location problem. Naval Research Logistics Quarterly **16**(3), 331–343 (1969)
7. Fleischmann, M., Bloemhof-Ruwaard, J., Dekker, R., Van Der Laan, E., Van Nunen, J., Van Wassenhove, L.: Quantitative models for reverse logistics: A review. European journal of operational research **103**(1), 1–17 (1997)
8. Freitas, A., Guimarães, F.: Originality and diversity in the artificial evolution of melodies. In: Proceedings of the 13th annual conference on Genetic and evolutionary computation, pp. 419–426. ACM (2011)
9. Horvath, P., Autry, C., Wilcox, W.: Liquidity implications of reverse logistics for retailers: A markov chain approach. Journal of retailing **81**(3), 191–203 (2005)
10. Hu, T., Sheu, J., Huang, K.: A reverse logistics cost minimization model for the treatment of hazardous wastes. Transportation Research Part E: Logistics and Transportation Review **38**(6), 457–473 (2002)
11. Jayaraman, V., Patterson, R., Rolland, E.: The design of reverse distribution networks: models and solution procedures. European Journal of Operational Research **150**(1), 128–149 (2003)
12. Ko, H., Evans, G.: A genetic algorithm-based heuristic for the dynamic integrated forward/reverse logistics network for 3pls. Computers & Operations Research **34**(2), 346–366 (2007)
13. Kreinovich, V., Quintana, C., Fuentes, O.: Genetic algorithms: What fitness scaling is optimal? Cybernetics and Systems: an International Journal **24**, 9–26 (1993)
14. Lu, Z., Bostel, N.: A facility location model for logistics systems including reverse flows: The case of remanufacturing activities. Computers & Operations Research **34**(2), 299–323 (2007)
15. Melo, M., Nickel, S., Saldanha-da Gama, F.: Facility location and supply chain management - a review. European Journal of Operational Research **196**(2), 401–412 (2009)
16. Min, H.: A bicriterion reverse distribution model for product recall. Omega **17**(5), 483–490 (1989). URL `http://ideas.repec.org/a/eee/jomega/v17y1989i5p483-490.html`
17. Montgomery, D.: Design and Analysis of Experiments, 7th edn. Wiley (2008)
18. R. Chandran, R.L.: Product recall: A challenge for the 1980. International Journal of Physical Distribution and Materials Management **11**(8), 483–490 (1981)
19. R Development Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2011)
20. Rosing, K., ReVelle, C.: Heuristic concentration: Two stage solution construction. European Journal of Operational Research **97**(1), 75–86 (1997)
21. Runarsson, T.P., Yao, X.: Stochastic ranking for constrained evolutionary optimization. IEEE Transactions on Evolutionary Computation **4**, 284–294 (2000)
22. Shaffer, J.P.: Multiple hypothesis testing. Annual Review of Psychology **46**, 561–584 (1995)
23. Sheu, J., Chou, Y., Hu, C.: An integrated logistics operational model for green-supply chain management. Transportation Research Part E: Logistics and Transportation Review **41**(4), 287–313 (2005)
24. Spengler, T., Puchert, H., Penkuhn, T., Rentz, O.: Environmental integrated production and recycling management. European Journal of Operational Research **97**(2), 308–326 (1997)
25. Srivastava, S.: Network design for reverse logistics. Omega **36**(4), 535–548 (2008)
26. Whitacre, J.M.: Adaptation and Self-Organization in Evolutionary Algorithms. Ph.D. thesis, University of New South Wales (2007)