# Vehicle Routing Problem with Time Windows: A Deterministic Annealing approach

Mayank Baranwal[1,a], Pratik M. Parekh[1,b], Lavanya Marla[2,c], Srinivasa M. Salapaka[1,d], Carolyn L. Beck[2,e]

*Abstract*— The Vehicle Routing Problem with Time-Windows (VRPTW) is an important problem in allocating resources on networks in time and space. We present in this paper a Deterministic Annealing (DA)-based approach to solving the VRPTW with its aspects of routing and scheduling, as well as to model additional constraints of heterogeneous vehicles and shipments. This is the first time, to our knowledge, that a DA approach has been used for problems in the class of the VRPTW. We describe how the DA approach can be adapted to generate an effective heuristic approach to the VRPTW. Our DA approach is also designed to not get trapped in local minima, and demonstrates less sensitivity to initial solutions. The algorithm trades off routing and scheduling in an $n$-dimensional space using a tunable parameter that allows us to generate qualitatively good solutions. These solutions differ in the degree of intersection of the routes, making the case for transfer points where shipments can be exchanged. Simulation results on randomly generated instances show that the constraints are respected and demonstrate near optimal results (when verifiable) in terms of schedules and tour length of individual tours in each solution.

## I. INTRODUCTION

In this paper, our focus is on a class of network-based resource allocation problems that are commonly referred to in the optimization literature as Vehicle Routing Problems with time-Windows (VRPTW). These problems involve both routing and scheduling elements, and are NP-hard in general. VRPTW problems involve the routing and scheduling of multiple vehicles from a depot (or multiple depots) to meet demands at multiple locations, under some time-window considerations. This class of problems is at the core of several problems that involve routing and/or scheduling, and occurs in several application domains including, but not limited to, transportation, logistics and communications. Specific examples include airline scheduling, vehicle routing, service network design, load distribution, production planning, computer scheduling, portfolio selection, and apportionment. These application domains have important economic value, and high importance is attached to achieving economically and computationally efficient solutions to the VRPTW.

VRPTW problems have been some of the most extensively studied problems in the optimization literature. In the past three decades, approaches including exact methods, heuristics, and neighborhood search approaches, or a combination of these have been used. As such, several survey papers on the topic exist. We refer the reader to [1], [2], [3] for the most recent surveys on general vehicle routing problems, and to [1] for some special cases of the VRPTW. Because these problems involve the allocation of multiple heterogeneous and discrete resources, have non-linear costs and constraints, the problems are typically non-linear and non-convex. Heuristic approaches are often subject to achieving local minima rather than global minima, and therefore are dependent on the starting solution that is used.

Specifically, we consider $K$ vehicles (resources) that start at a depot, and need to pick up shipments from various customer locations, denoted $i = 1, ..., N$. Each vehicle $j$ can potentially have capacity $W_j$. Each customer has a time-window $[t_{i,\text{start}}, t_{i,\text{end}}]$ within which the shipment is to be picked up. Associated with each customer/shipment $i$ is a priority weight $p_i$. Additionally, we are also given the travel times between each pair of points in the network including the depot and customer locations. Our goal is to find the set of customers to be visited by each individual vehicle, the sequence and the schedule according to which they should be visited, while minimizing the costs associated with shipments served earlier or later than their time-windows, and the total distance covered by the vehicles. We also consider the case of heterogeneous shipments, where each shipment can also be one of $l = 1, ..., p$ types. Each vehicle $j$, has a holding capacity for the $l$th type of shipment, that we denote as $W_{jl}$.

In this paper we present a solution methodology based on Deterministic Annealing (DA) for the VRPTW. DA is well-suited to combinatorial clustering/resource allocation problems that require obtaining an optimal partition of an underlying domain, and optimally assigning resources to each cell of the partition. DA-based methods have been reported in a vast number of applications such as minimum distortion problems in data compression [4], model aggregation [5], routing problems in multiagent networks [6], locational optimization problems [7], and coverage control problems [8].

Such problems are highly non-convex, computationally complex and suffer from poor local minima that riddle the cost surface [9]. A variety of heuristic approaches have been proposed to address these difficulties, and they range from repeated optimization with different initialization, and heuristics to good initialization, to heuristic rules for cluster splits and merges. In this context, simulated annealing algorithm [10], which capitalizes on the analogy of annealing process in physical chemistry, was shown to be a good iterative method that would achieve a global minimum, but with an annealing rate so slow that this algorithm loses practicality in many applications. In [11], [12], a threshold accepting deterministic variant of the simulated annealing algorithm has been proposed. It should be remarked that in some ex-

[1]Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, 61801 IL, USA

[2]Department of Industrial and Enterprise Systems Engineering, University of Illinois at Urbana-Champaign, 61801 IL, USA

[a]baranwa2@illinois.edu, [b]pparekh2@illinois.edu
[c]lavanyam@illinois.edu, [d]salapaka@illinois.edu
[e]beck3@illinois.edu

isting literature, such deterministic variants are unfortunately also referred to as the deterministic annealing algorithms; however they are significantly different from the maximum-entropy principle based deterministic algorithm [13] used in this article - in terms of the underlying heuristics, implementation, goals, and performance. The DA algorithm used in this paper has the ability to avoid poor local optima while still maintaining a relatively fast convergence rate. However, to the best of our knowledge, DA has not been availed to address the specifications and constraints that arise in the context of simultaneous routing and scheduling problems. Our approach adapts these specifications and constraints in terms of facility location/clustering, thereby enabling the use of DA. Since DA allows for incorporating multiple constraints and different types of specifications, it is well suited to VRPTW (including cases when complex additional constraints are present). For instance, this approach offers the flexibility to tradeoff routing and scheduling in space and time-dimensions to identify clusters in the combined space to allow us to tradeoff between routing and scheduling decisions. We discuss this in detail in the paper.

The main contributions of this paper are to incorporate the DA method to VRPTW, where specific capacity, routing, and scheduling constraints are incorporated simultaneously. Simulation results show that the constraints are respected and demonstrate near optimal results (when verifiable). This method provides a tunable parameter, whereby the relative importance of scheduling vs routing is controlled. Studies with respect to this parameter evidence general features, which can be used for updating scheduling and routing procedures; for instance we show that when schedule is emphasized over routes, there is an increased overlapping of routes and self intersections, which suggests that such scenarios will benefit by having *transfer/exchange locations* where shipments between different vehicle routes can be exchanged.

## II. PROBLEM DESCRIPTION

In order to solve the VRPTW, we first describe a series of related problems, which are in increasing order of complexity.

### A. Uncapacitated Scheduling Problem

This problem involves finding a schedule according to which each vehicle will visit the customer (shipment) locations in order to pick up shipments. In this problem we ignore the locations of customers and assume travel between any two locations is instantaneous. While this is not applicable in practice, this helps illustrate the working of our algorithm. Note that in the context of VRPTW, this problem only considers the vehicle allocation for a given schedule information; it does not have any routing aspect to it. More precisely, the problem reduces to a scenario where shipments with prescribed time-windows are needed to be picked up from a depot; the objective is to ascribe an arrival time for each vehicle so that the maximum number of shipments are serviced.

### B. Capacitated Scheduling Problem

This problem builds upon the uncapacitated scheduling problem by adding capacity constraints on the number of customers (shipments) that can be carried by each vehicle. Vehicles are heterogeneous, that is, can have varying capacities.

### C. Scheduling Problem with Multiple Capacity Constraints

In this version of the capacitated scheduling problem, customers (shipments) picked up have varying types. They thus have different occupancy rates in each vehicle. Each vehicle has a capacity associated with each shipment type it can carry.

### D. Vehicle Routing Problem (VRP)

In the VRP, we find a set of locations that is served by each vehicle in the fleet, as well as the sequence according to which each customer location should be visited. We assume that there are no time-window restrictions on when each customer (shipment) can be visited/served.

### E. Vehicle Routing Problem with Time Windows (VRPTW)

The VRPTW involves solving the VRP with additional time-window restrictions on the time that each customer can be visited. This problem can incorporate heterogeneous shipment sizes as well as heterogeneous capacities.

## III. DETERMINISTIC ANNEALING ALGORITHM: A MAXIMUM ENTROPY PRINCIPLE APPROACH FOR CLUSTERING

At its core, the DA algorithm solves a facility location problem (FLP): For given $N$ customer locations, find $K$ facility locations such that the total *weighted sum of the distance of each customer to its nearest facility is minimized.* In other words, if $s_i$ and $r_j \in \mathbb{R}^n$ denote the locations of $i^{th}$ customer and $j^{th}$ facility, respectively, then the FLP addresses the following optimization problem:

$$\min_{r_j \in \Omega, 1 \leq j \leq K} \sum_{i=1}^{N} p_i \left\{ \min_{r_j, 1 \leq j \leq K} d(s_i, r_j) \right\} \quad (1)$$

where, $d(s_i, r_j) \in \mathbb{R}_+$ denotes the distance between the $i^{th}$ customer location $s_i$ and $j^{th}$ facility location $r_j$, $\Omega \in \mathbb{R}^n$ is a compact domain of interest and $p_i$ is a given positive constant (without loss of generality, we assume $\sum_i p_i = 1$) that denotes the relative weight of the $i^{th}$ customer. Note that in general, $s_i$ and $r_j$ need not necessarily denote physical locations but can belong to any relevant *property space* $\Omega$. Borrowing from the data compression literature [14], we define *distortion* as a measure of the average distance of a customer to its nearest facility, given by $D(s, r) = \sum_{i=1}^{N} p_i \min_{1 \leq j \leq K} d(s_i, r_j)$. Then the equivalent optimization problem is to minimize the distortion function. The solution to a FLP essentially results in a series of clusters, where the facility $j$ is located at the centroid $r_j$ of the $j^{th}$ cluster and each customer is associated only to its nearest facility.

Note that any change in location of a particular customer $i$ affects $d(s_i, r_j)$ only with respect to the *nearest* facility $j$ in

the distortion. This distributed aspect makes most algorithms (such as Lloyd's [15]) overly sensitive to the initial facility location. The DA algorithm [13] overcomes this sensitivity by modifying the distortion $D$ such that every customer $i$ is associated to every facility $j$ through an *association weight* $p(j|i)$:

$$\bar{D}(s,r) = \sum_{i=1}^{N} p_i \sum_{j=1}^{K} p(j|i) d(s_i, r_j). \tag{2}$$

Here we choose $\{p(j|i)\}$ to satisfy $0 \leq p(j|i) \leq 1$ and $\sum_{j=1}^{K} p(j|i) = 1$ without loss of generality. Thus, we replace the *average distance of a customer to its nearest facility* by the *weighted average distance of a customer to all the facilities*. The weights $p(j|i)$ assess the trade-off between decreasing the *local* influence and the deviation of $\bar{D}$ from the original distortion $D$. Since we have no prior knowledge on the association weights, we apply the principle of maximum entropy to estimate them. The Shannon entropy term $H(r|s) = -\sum_{i=1}^{N} p_i \sum_{j=1}^{K} p(j|i) \log(p(j|i))$, widely used in data compression literature [14], measures uncertainties in facility locations with respect to the known customer locations. Thus, maximizing the entropy is equivalent to decreasing the *local* influence.

This trade-off between minimizing the distortion in (2) and maximizing the entropy [4], [13] is achieved by seeking $\{p(j|i)\}$ that minimize the *free energy*, or the Lagrangian, given by $F(r) \triangleq \bar{D}(s,r) - TH(r|s)$, where $T$ is a Lagrange multiplier, referred to as *temperature*. This yields a *Gibbs distribution*

$$p(j|i) = \frac{\exp\{-\beta d(s_i, r_j)\}}{\sum_{k=1}^{K} \exp\{-\beta d(s_i, r_k)\}} \tag{3}$$

with $\beta = 1/T$. By substituting the association weights (3), the free energy in (4) simplifies as

$$F(r) = -\frac{1}{\beta} \sum_{i=1}^{N} p_i \log \sum_{k=1}^{K} \exp\{-\beta d(s_i, r_k)\}. \tag{4}$$

In the DA algorithm, this *free energy* function is then deterministically optimized at successively reduced temperatures over repeated iterations.

The readers are encouraged to refer to [16] for detailed analysis on the complexity of the DA algorithm. For implementation on very large datasets, a scalable modification of the DA is proposed in [17].

## IV. SOLUTION APPROACH: MODIFICATIONS OF THE DA ALGORITHM

DA addresses both routing and scheduling problems as modifications of the clustering problem, wherein by choosing $\Omega$ as a domain of spatial coordinates, the DA solution groups together customers served by the same vehicle, while clustering on the space $\Omega$ of time-windows results in service schedules for each customer. We now describe the DA algorithm in the context of VRP to accommodate various scheduling, routing and capacity constraints.

### A. DA for scheduling

In this scenario, the DA is used to allocate vehicle arrival times $r_j$ to service the shipments with specified service time-windows $[t_{i,\text{start}}, t_{i,\text{end}}]$. Accordingly we choose $\Omega \subset \mathbb{R}$ representing the time domain of interest, the property $s_i$ of the customer $i$ is chosen to be the *mid-point* of the associated time window, i.e.,

$$s_i = \frac{t_{i,\text{start}} + t_{i,\text{end}}}{2}; \tag{5}$$

a convenient choice of the distance metric between the customer $i$ and facility $j$ is the squared-euclidean distance $d(s_i, r_j) = |s_i - r_j|^2$. The distortion $\bar{D}$ captures the penalty of deviating from the *mid-times*. Thus, minimizing the free-energy $F$ in (4) is commensurate with the cost incurred for not serving within the specified time-window being minimized over all shipments.

### B. DA for routing

Some common heuristics to the standard single depot vehicle routing problems include approaches, such as, *cluster first-route second* [18]. We follow the same approach wherein we first cluster the demand nodes based on their geographical distances, and then later design economical routes over each cluster. Executing the DA algorithm directly on the space of geographical coordinates results in a partition of the underlying domain $\Omega$, wherein each cluster $j$ is served by a different vehicle. In the clustering process $\Omega \subset \mathbb{R}^2$ is the spatial domain of interest and we use the parameter $\beta$ to cluster locations. The heuristic behind this approach is that if we use $K = N$ resources and let $\beta \to \infty$ then we expect each customer to be associated with one resource (vehicle).

However, the clustering lacks the sequencing aspect of the routing problem. In order to incorporate the sequencing aspect, we include the minimum tour-length constraint in the original formulation, which essentially amounts to solving a traveling salesman problem (TSP) in each cluster. This is achieved by appropriately modifying the free energy in (4) to obtain the new Lagrangian

$$F'(r) = F(r) + \lambda \left( \sum_{k=1}^{K} d(r_j, r_{j-1}) - L \right), \tag{6}$$

where $L$ is a given tour-length, and $\lambda$ is the Lagrange multiplier related to it. As $L$ is gradually increased from a small value, the tour length between locations $\{r_j\}$ in the same cluster would coincide with the optimal tour length between $\{s_i\}$. Here the spatial coordinates $\underline{x}_i = (x_i, y_i)$ represent the property $s_i$ of the customer $i$. The distance function $d(s_i, r_j)$ between the customer $i$ and facility $j$ is chosen as : $d(s_i, r_j) = \|s_i - r_j\|_2^2$. We adopt the Elastic Net approach [19] for finding the optimal tour-length. The required tour-length is controlled by appropriately varying the Lagrange multiplier $\lambda$ with $\beta$ [13], [20].

### C. DA for scheduling-cum-routing

In this setting, we desire a resource $j$ to be located at a customer $i$ in a pre-ascribed time-window. In this context, the property $s_i$ of the customer $i$ is given by,

$$s_i = \begin{pmatrix} x_i \\ y_i \\ \frac{\lambda}{2}(t_{i,\text{start}} + t_{i,\text{end}}) \end{pmatrix}, \tag{7}$$

where $\underline{x}_i = (x_i, y_i)$ and $[t_{i,\text{start}} \quad t_{i,\text{end}}]$ represent the spatial coordinates and service time-window of the customer $i$, respectively, and the domain of interest is $\Omega \subset \mathbb{R}^2 \times \mathbb{R}_+$. We define a new metric that incorpates both *scheduling* and *routing* through a *velocity* parameter $\lambda$

$$d(s_i, r_j) = \left\| \underline{x}_i - \begin{bmatrix} r_j(1) \\ r_j(2) \end{bmatrix} \right\|_2^2 + \frac{\lambda^2}{4} |(t_{i,\text{start}} + t_{i,\text{end}}) - r_j(3)|^2. \quad (8)$$

Clearly, setting $\lambda = 0$ solves the routing problem, as described in Sec. IV-B, while setting $\lambda$ to a very high value virtually solves the scheduling problem in Sec. IV-A. Thus $\lambda$ captures the trade-off between *distance* based optimization and *time-window* based optimization problems. In practice, $\lambda$ should be set to a value comparable to the average speed of the vehicles.

### D. DA for heterogeneous capacitated resources/vehicles

In this setting, we consider vehicles $\{j\}$ to be of different types. If $\lambda_j$ denotes the total allocated capacity of the resource $j$, i.e. if $\lambda_j$ amount of resources are located at $r_j$, then the modified Gibbs distribution is given by $p(j|i) = \frac{\lambda_j \exp\{-\beta d(s_i, r_j)\}}{\sum_{k=1}^{K} \lambda_k \exp\{-\beta d(s_i, r_k)\}}$. Note that the parameter $\lambda_j$ in the modified distribution specifies the weight of the $j$th resource; the resulting solution $p_j : p_k$ tend to be in the same ratio as $\lambda_j : \lambda_k$, where $p_j$ is marginal distribution of the $j$ resource given by $p_j = \sum_i p_i p(j|i)$ [4], [7], [21].

### E. DA for heterogeneous capacitated vehicles and heterogeneous shipments

In this setting, we have $p$ types of customers, and the total amount of resource $j$ needing to be allocated to all customers of $l^{th}$-type are $\lambda_{jl}, 1 \le l \le p, 1 \le k \le K$. These constraints are incorporated in the Gibbs distribution as $p(j|i) = \frac{\sum_{l=1}^{p} \lambda_{j,l} \exp\{-\beta d(s_i, r_j)\}}{\sum_{k=1}^{K} \sum_{l=1}^{p} \lambda_{kl} \exp\{-\beta d(s_i, r_k)\}}$.

## V. SIMULATIONS AND RESULTS

In this section, we demonstrate the effectiveness of DA for solving scheduling and routing problems of varying degrees of complexity. We generate randomized instances of each problem type and the algorithm is tested for each of the problem instances. However, for the sake of demonstration, we show only one instance of each problem type.

### A. Uncapacitated Scheduling Problem

For the uncapacitated scheduling problem, we consider an instance with 20 shipments needing to be serviced by 3 vehicles. The depot incurs the corresponding cost if the shipment does not get serviced. That is, if a vehicle arrival time does not lie in the time-window for a shipment, the shipment goes to waste and the corresponding cost is incurred by the depot. We randomly sample the time-windows to lie within $[0, 350]$ minutes.

In Fig. 1a, blue lines denote the final clustered information (vehicle arrival times) after direct execution of the DA algorithm (as described in Sec. IV-A). The red bars denote the time-windows for each shipment. Only three shipments indexed $3, 4$ and $11$ do not have any vehicle arrivals within

their time-windows. From the time-intervals seen in Fig. (1a), it is evident that the omission of these shipments is inevitable, as accommodating any of them leads to larger costs as other shipments will not be served.

### B. Capacitated Scheduling Problem

In the capacitated scheduling problem, we provide capacity to each vehicle and the algorithm should attempt to match these capacities. Once again, the time-windows are randomly sampled to lie within $[0, 350]$ minutes. We have 100 shipments needing to be picked up by only 10 vehicles, where the capacities assigned to each vehicle (cluster) are $\{3, 5, 9, 9, 9, 10, 13, 13, 13, 16\}$.

Subjected to these capacities, the algorithm works well and only 30% of the shipments are not serviced. The effectiveness of the DA algorithm for clustering with capacity constraints is observed in the pie chart in Fig. 1c. The pie chart shows the fraction of shipments associated with each vehicle, which is mathematically depicted by $p(r_j) = \sum_i p_i p(j|i)$. The capacities associated with each cluster (vehicle) are in proportion with the capacity values given as constraints.

### C. Scheduling Problem with Multiple Capacity Constraints

For this problem, there are capacities associated with each vehicle and all types of shipments. For the simulation, we choose an instance with 100 shipments in total, 3 types of shipments and only 10 vehicles for pickup. There are 34 shipments of type-1 (red), 36 shipments of type-2 (magenta), and 30 shipments of type-3 (green). The capacities are randomly chosen and subject to these capacities, the algorithm works well.

In Fig. 2a, the blue lines depict the vehicle arrival times, and the red, magenta and green bars indicate the time-windows for the three types of shipments respectively. A total of 71 shipments are picked up. Fig. 2b shows the number of shipments picked up by each vehicle.

Fig. 2c shows the effectiveness of the DA algorithm in respecting the capacity constraints. The plot on the left shows the capacity constraints $(\lambda_{jl})$ for each type of shipment and vehicle. The plot on the right shows the clustered mass information. Clearly, the shapes of the two plots match each other.

### D. Vehicle Routing Problem

For the VRP, we consider a 59 cities network (and a depot city from where the vehicles start) whose geographical coordinates are located in a $[-100, 100] \times [-100, 100]$ unit$^2$ area. The coordinates of the vehicle depot are given by $(0, 0)$. The nodes need to be serviced by a total of six vehicles.

We first cluster the network based on their pairwise square-Euclidean distance $d(\underline{x}_i, \underline{x}_j) = \|\underline{x}_i - \underline{x}_j\|_2^2$, i.e., the DA algorithm described in Sec. IV-B is executed on the set of cities $\{\underline{x}_i\}$ to obtain $K = 6$ partitions of the entire network. This is followed by solving a TSP in each of the individual clusters using the modification described in Sec. IV-B. Fig. 3 shows the effectiveness of the DA algorithm for solving an instance of the VRP.

**Remark**: While the cluster first-route second approach does not necessarily result in an overall global optimal solution, each individual tour obtained by the DA algorithm in
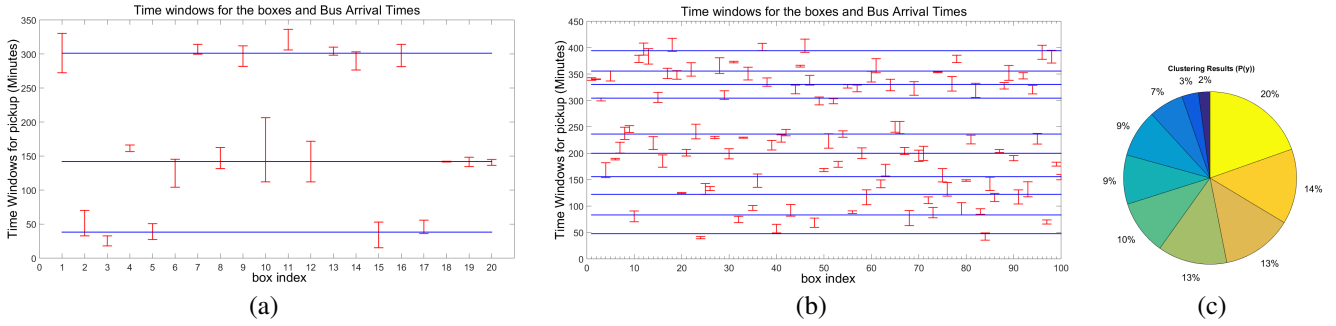
Fig. 1: (a) Time-Windows for 20 shipments (Red Intervals) and the Arrival Times (Cluster locations (blue lines)) for 3 vehicles. (b) Time-Windows for 100 shipments (Red Intervals) and the Arrival Times (Cluster locations (blue lines)) for 10 vehicles. (c) Capacity associated with each cluster for the 100 shipments scenario. Clearly, the capacities obtained by the DA algorithm are in proportion to the capacity values given as constraints.
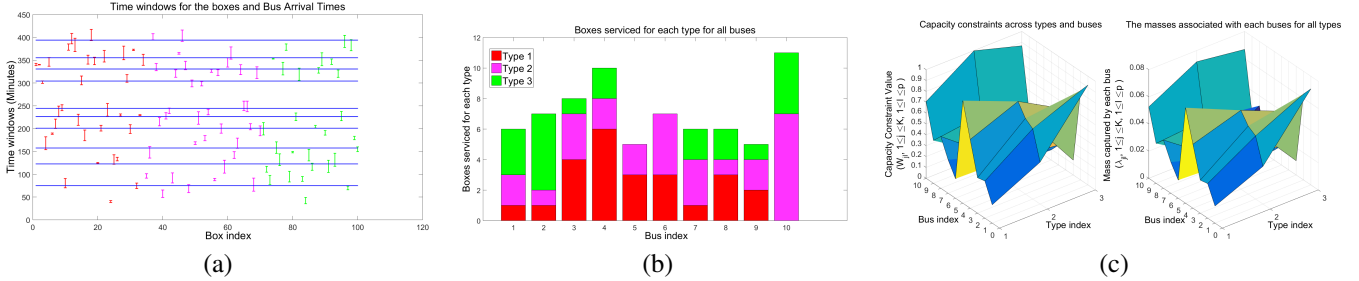


Fig. 2: (a) Time-Windows for 100 shipments (3 types) and the Arrival Times (Cluster locations (blue lines)) for 10 vehicles. The three colors (red, magenta and green) are used to indicate the type of shipments. (b) Number of shipments served by each vehicle for each type. (c) [Right]: Capacity constraint values for each type of shipment and vehicle. [Left]: Capacity associated with each vehicle and each type of shipment. Clearly, the shapes indicate that the capacities obtained by the constrained DA algorithm are in proportion with the capacities provided as constraints.
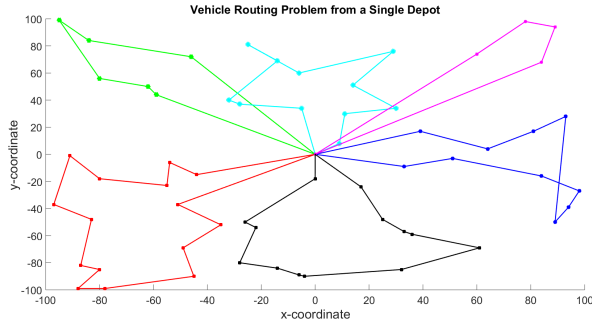


Fig. 3: Solution to the VRP using cluster first-route second approach. The nodes are first clustered based on their pairwise distances, followed by computation of optimal routes over these clusters.

the prescribed is indeed optimal (as verified using exhaustive simulations).

### E. Vehicle Routing Problem with Time Windows

For the VRPTW, we again consider the same network of 59 cities, however, with respective delivery time-windows. The time-windows are chosen to lie in the interval $[0, 350]$ minutes. The maximum speed of each vehicle is chosen to be 10 units/min.

The cities are first clustered using the spatio-temporal metric described in Eq. 8. Introduction of the additional time coordinate in the problem domain results in intersecting partitions in the spatial coordinates (see Fig. 4), i.e., the paths traversed by different vehicles have overlaps. This observation is in consonance with the fact that vehicles may traverse overlapping routes to ensure timely delivery

of packages. Once the cities are clustered in the spatio-temporal domain, the DA algorithm is employed again for solving TSPs over each cluster in this new domain, i.e., the TSP is solved using the new metric. This heuristic ensures timely delivery of shipments so that losses incurred to the depot are minimized. The algorithm achieves this objective by finding self-intersecting (zig-zag) routes. In Fig. 4a, two such routes (green and cyan) are shown with corresponding time-windows. A total of only 11 shipments are not delivered on time.

It should be noted that the parameter $\lambda$ in (8) controls the trade-off between the *distance* based optimization and the *time-window* based optimization. In other words, $\lambda$ can be interpreted as the average velocity of a vehicle. Smaller values of $\lambda$ result in reduced total tour-lengths, but allow for larger number of not-on-schedule deliveries, as shown in Fig. 4b. As $\lambda$ is increased, the paths traversed by each vehicle overlap and appear to have self-intersections as well as intersections with routes of other vehicles. Thus, such a framework also allows for existence of potential *transfer points*, where shipments can be transferred from one vehicle to another. Note that $\lambda = 0$ corresponds to the scenario of VRP (see Fig. 3), where routes do not overlap or self-intersect.

### VI. CONCLUSIONS AND FUTURE WORK

In this paper, scheduling and routing problems of varying degrees of complexity are addressed in the Deterministic Annealing framework. The algorithm surfaces as an efficient heuristic for solving combinatorially complex scheduling and routing problems. Following are some key features and
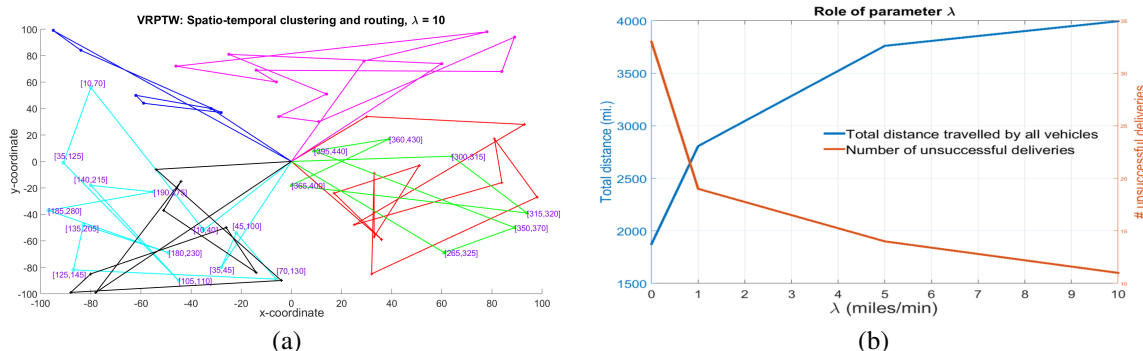
Fig. 4: (a) Computation of optimal routes by the DA algorithm using the proposed spatio-temporal metric. The method ensures timely delivery of packages so that losses incurred to the depot are minimized. Two such routes (green and cyan) are shown with corresponding time-windows, where the corresponding deliveries for most cities are met on time. (b) Effect of parameter $\lambda$ on the total tour-length and the number of unsuccessful deliveries.

observations of the DA algorithm (and its modifications):

(1) The solutions provided by the DA are completely independent of initializations. In fact, the algorithm attempts to solve a completely deterministic optimization problem (maximization of entropy term) in the beginning.

(2) The algorithm has ability to avoid poor local optima. This is illustrated by the quality of solutions for various problem types. For example, in the m-TSP scenario in the VRP in Sec. V-D, the individual tour-lengths are found to be optimal.

(3) When modified appropriately, the algorithm respects various capacity constraints. For example, in Sec. V-C we observe that the capacities obtained by the DA algorithm match the required capacity constraints for each shipment and each vehicle.

(4) A new spatio-temporal metric is proposed which incorporates the concept of time-based distance using a *velocity* parameter $\lambda$. As $\lambda$ is increased, the routes traversed by different vehicles are observed to overlap more, and a relatively larger number of self-intersections are observed within each route to ensure timely delivery of packages. This may facilitate introduction of the *transfer points* where shipments from one vehicle are transferred to another to be delivered by the second vehicle.

The solutions provided by DA are observed to be robust to uncertainties in customer locations or service schedules, which will be demonstrated in the subsequent work. Another future direction is to design distance functions that incorporate the time-windows directly in the problem formulation. We intend to test our algorithm against the existing heuristics on benchmark instances and the results will be reported in the subsequent work. We also aim to include the notion of *transfer points* in our problem formulation as part of our future work.

## ACKNOWLEDGMENT

## REFERENCES

[1] J.-F. Cordeau, G. Laporte, M. W. Savelsbergh, and D. Vigo, "Vehicle Routing," in *Handbook in Operations Research and Management Science*, C. Barnhart and G. Laporte, Eds., no. 14.

[2] J.-F. Cordeau, G. Laporte, J.-Y. Potvin, and M. W. Savelsbergh, "Transportation on Demand," in *Handbook in Operations Research and Management Science*, C. Barnhart and G. Laporte, Eds., no. 14.

[3] B. Golden, S. Raghavan, and E. Wasil, Eds., *The Vehicle Routing problem- Latest advances and new challenges*. Springer, 2008.

[4] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2210–2239, 1998.

[5] Y. Xu, S. M. Salapaka, and C. L. Beck, "Aggregation of graph models and Markov chains by deterministic annealing," *Automatic Control, IEEE Transactions on*, vol. 59, no. 10, pp. 2807–2812, 2014.

[6] N. V. Kale and S. M. Salapaka, "Maximum entropy principle-based algorithm for simultaneous resource location and multihop routing in multiagent networks," *Mobile Computing, IEEE Transactions on*, vol. 11, no. 4, pp. 591–602, 2012.

[7] S. Salapaka, A. Khalak, and M. Dahleh, "Constraints on locational optimization problems," in *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on*, vol. 2. IEEE, 2003, pp. 1741–1746.

[8] Y. Xu, S. M. Salapaka, and C. L. Beck, "Clustering and Coverage Control for Systems With Acceleration-Driven Dynamics," *Automatic Control, IEEE Transactions on*, vol. 59, no. 5, pp. 1342–1347, 2014.

[9] R. M. Gray and E. D. Karnin, "Multiple local optima in vector quantizers (corresp.)," *Information Theory, IEEE Transactions on*, vol. 28, no. 2, pp. 256–261, 1982.

[10] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi *et al.*, "Optimization by simulated annealing," *science*, vol. 220, no. 4598, pp. 671–680, 1983.

[11] K. Braekers, A. Caris, and G. K. Janssens, "A deterministic annealing algorithm for a bi-objective full truckload vehicle routing problem in drayage operations," *Procedia-Social and Behavioral Sciences*, vol. 20, pp. 344–353, 2011.

[12] O. Bräysy, W. Dullaert, G. Hasle, D. Mester, and M. Gendreau, "An effective multirestart deterministic annealing metaheuristic for the fleet size and mix vehicle-routing problem with time windows," *Transportation Science*, vol. 42, no. 3, pp. 371–386, 2008.

[13] K. Rose, "Deterministic annealing, clustering, and optimization," Ph.D. dissertation, California Institute of Technology, 1990.

[14] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[15] S. P. Lloyd, "Least squares quantization in PCM," *Information Theory, IEEE Transactions on*, vol. 28, no. 2, pp. 129–137, 1982.

[16] P. M. Parekh, D. Katselis, C. L. Beck, and S. M. Salapaka, "Deterministic annealing for clustering: Tutorial and computational aspects," in *American Control Conference (ACC), 2015*. IEEE, 2015, pp. 2906–2911.

[17] P. Sharma, S. Salapaka, and C. Beck, "A scalable deterministic annealing algorithm for resource allocation problems," in *American Control Conference, 2006*. IEEE, 2006, pp. 6–pp.

[18] B. E. Gillett and L. R. Miller, "A heuristic algorithm for the vehicle-dispatch problem," *Operations research*, vol. 22, no. 2, pp. 340–349, 1974.

[19] R. Durbin, R. Szeliski, and A. Yuille, "An analysis of the elastic net approach to the traveling salesman problem," *Neural Computation*, vol. 1, no. 3, pp. 348–358, 1989.

[20] B. A. ROEHL, "Maximum-entropy principle approach to the multiple travelling salesman problem and related problems," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 2011.

[21] S. Salapaka and A. Khalak, "Locational optimization problems with constraints on resources," in *Proceedings of the Annual Allerton Conference on Communication Control and Computing*, vol. 41, no. 3. The University; 1998, 2003, pp. 1240–1249.