Dynamic vehicle routing: Status and prospects

Harilaos N. Psaraftis

National Technical University of Athens, Athens, Greece

Although most real-world vehicle routing problems are dynamic, the traditional methodological arsenal for this class of problems has been based on adaptations of static algorithms. Still, some important new methodological approaches have recently emerged. In addition, computer-based technologies such as electronic data interchange (EDI), geographic information systems (GIS), global positioning systems (GPS), and intelligent vehicle-highway systems (IVHS) have significantly enhanced the possibilities for efficient dynamic routing and have opened interesting directions for new research. This paper examines the main issues in this rapidly growing area, and surveys recent results and other advances. The assessment of possible impact of new technologies and the distinction of dynamic problems vis-à-vis their static counterparts are given emphasis.

1. Introduction and background

Efficient distribution of goods entails, among other things, a determination of routes and schedules for the fleet of vehicles so that total distribution costs are minimized, while various requirements (constraints) are met. The constraints concern various facets of the operation, such as vehicle capacities, time windows on pick up and/or delivery, time availability of vehicles, etc.

As the world’s economies become more and more interdependent, efficient distribution of goods (from raw materials to finished products) becomes of paramount importance, not only for the survival of many businesses that depend on such distribution, but ultimately, for the overall competitiveness of the market, both internal and external. As an example, with the implementation of a borderless European market as of January 1, 1993, competitive advantage in many industries becomes critically dependent on their ability to provide efficient and on-time logistical service. Any tangible improvement in the logistical efficiency of the system translates into gains for all consumers, for they ultimately pay the costs of the distribution system, as part of the price of every product they purchase. Similar considerations already apply to the vast distribution markets in North America, but are certain to become more important now that the North American Free Trade Agreement has been ratified.

© J.C. Baltzer AG, Science Publishers
What is likely to distinguish most distribution problems today (a fortiori, tomorrow) from equivalent problems in the past is that the information that is needed to come up with a set of good vehicle routes and schedules is dynamically revealed to the decision maker (in whole or in part). Such real-time availability of information was rare or non-existent in the past. It is becoming more and more common today, and will certainly be pervasive in the years ahead.

Thus, dynamic scenarios have become more common in distribution logistics recently, and are likely to become even more so in the future. In vehicle routing, they span a diverse spectrum of applications, including the delivery of petroleum products or industrial gases, courier services, intermodal services, tramp ship operations, pickup and delivery services, management of container terminals, etc. (more details in section 2).

Revolutionary advances in information and communications technologies (ICT), particularly between vehicle and central dispatching office, have boosted the role of information in such systems, and make it easier for data to be continually (dynamically) updated. The explosive increase in computing speeds at increasingly affordable prices has opened new directions for the implementation of such technologies in distribution logistics. One of the areas that is most ripe to be the beneficiary of such advances is that of vehicle routing (especially dynamic scenarios), for reasons that are summarized below:

(a) Efficient vehicle routing becomes more important as markets tend to become increasingly open.
(b) Real-time distribution scenarios are likely to be the “norm” in the future.
(c) Processing of real-time data is more and more feasible and affordable.
(d) The economic benefits that will be realized if the efficiency of these logistical systems increases are very significant.

From the point of view of an operations research practitioner, a question is whether the state of the art in methodologies in this area is on a par with the rapid evolution (or revolution) in the related technologies and in the pattern of transportation business.

Posing such a question a few years ago might perhaps be irrelevant, for the various technological and other (institutional, economic) developments were at a much less important standing at the time. Nonetheless, in a paper by this author (Psaraftis [48]), it was argued that the methodological base for solution techniques explicitly developed and designed for dynamic vehicle routing problems was scant, and that what could be a dynamic equivalent of the archetypal vehicle routing problem (that is, of the Traveling Salesman Problem) was not even defined, let alone formulated or solved at the time. Indeed, in 1988 most of the traditional approaches for solving a dynamic vehicle routing problem had been straightforward adaptations of static procedures (see Wilson et al. [58], Psaraftis [45], Bell et al. [3],
Psaraftis et al. [47], Brown et al. [10], among others). In these procedures, a static vehicle routing problem (with appropriately defined inputs) was solved, either exactly, or heuristically, each time an input update occurred.

In an attempt to make some suggestions for further research, Psaraftis [48] included a discussion of the possible methodological implications of the differences between static and dynamic vehicle routing, and introduced a generic dynamic vehicle problem, the so-called Dynamic Traveling Salesman Problem (DTSP).

Several years having passed, and given the various institutional and technological developments outlined earlier, it is perhaps time to re-examine this problem area, and ask what has happened in it since then. Thus, without claiming to be encyclopaedic, it is the purpose of this paper to take a more recent look at dynamic vehicle routing, by taking stock of recent developments in this area (both in methodology and in technology) and by discussing some of the relevant issues.

Before we proceed, we state that the scope of our analysis is strictly limited to dynamic vehicle routing (and scheduling) problems, or models. These are a subset of a broader family of models, coming under the general rubric of “dynamic transportation models”. The label “stochastic” is often tagged to this class of models (“dynamic and stochastic transportation models”). In addition to dynamic vehicle routing, this broader class of problems includes dynamic shortest paths (Psaraftis and Tsitsiklis [49]), dynamic traffic assignment (Friesz et al. [23], Boyce et al. [9]), dynamic fleet management (Powell [44], Frantzeskakis and Powell [22]), dynamic air traffic control (Soumis and Odoni [52]), and dynamic facility location (Megiddo [39], Daskin and Hopp [14]).

These other classes of problems will not be examined here. We only note that Operations Research recently produced a special issue on “stochastic and dynamic models in transportation” [19], a volume that contains fifteen papers in this area. This testifies to the recent growth of interest and activity in the general area of dynamic transportation problems.

The rest of this paper is organized as follows: section 2 discusses the contexts in which dynamic vehicle routing problems may be encountered, and why these problems are important. Section 3 presents some of the recent technological advances that make this area more important than it has been thus far. Section 4 presents a taxonomy that characterizes the various attributes of information for this problem class. Section 5 presents some recent methodological developments in this area and discusses them in the context of section 4. Finally, section 6 concludes by recommending directions for further research.

2. Contexts and importance of dynamic vehicle routing

A non-exhaustive inventory of distribution problems that generally involve dynamic vehicle routing is the following (the references that are cited are representative and do not necessarily deal with inherently dynamic formulations of these problems):
(1) **Delivery of petroleum products, industrial gases, or other products** (Brown et al. [10], Dror et al. [16], Fisher et al. [21], Trudeau and Dror [57]). A supplier of oil must arrange a set of routes to replenish the inventories of a set of customers. The amount of oil that each customer needs is not precisely known in advance, and hence there is a possibility that the tank truck that serves that customer does not have enough oil to fully replenish that customer's stock. The possibility of stockouts is costly and should be avoided, and here too the dispatching of the truck fleet should be done efficiently (according to some criterion).

(2) **Courier services** (Mabert et al. [36]). A mini-van goes around a city district collecting parcels and express packages. Requests for service arrive in real time to the courier's central dispatching office and are automatically relayed either by phone or by some other device to the mini-van. The van's on-board computer processes each request, by appropriately inserting it into the sequence of pick ups. Alternatively, the processing of each request is carried out centrally, either by a central computer, or by a human scheduler, who also decides which mini-van should handle each request.

(3) **Intermodal services** (Crainic et al. [12], Min [41]). A liner general-cargo ship arrives at a busy port. Packages are unloaded and assembled into a central warehouse. A fleet of trucks distributes packages to the port hinterland or to other consolidation points, and concurrently picks up other packages that are destined to the warehouse for export. Not all customers are known ahead of time and information on new customers arrives dynamically. The dispatcher has to coordinate the movement of the vehicle fleet in real time.

(4) **Tramp ship operations** (Garmilla [24], Ronen [50], Speidel [53]). A fleet of tramp ships (either chartered or owned) is used by an industrial company to haul bulk commodities (such as oil, coal, metal ores, cement, etc.) from production points (refineries, other factories, loading terminals) to consumption points (storage facilities, unloading silos, other terminals). Demand at these consumption points is dynamic and not precisely known in advance. Also, sailing times are functions of prevailing weather conditions. Schedules and routes for these ships have to be produced on a continual basis to meet the demand.

(5) **Combined pickup and delivery services** (Min [40], Psaraftis et al. [47], Thompson and Psaraftis [56]). Here, a fleet of vehicles is available to service demand requests that arrive by phone or some other communication means (e.g. a "call box" located on the street) to a central dispatching or scheduling center. These requests are not known in advance to the dispatcher, but rather arrive in real time, and must be serviced as soon as possible. The dispatcher must decide in real time which vehicle to send to a specific
request, and how to modify the routing of that vehicle (which may already be committed to pick up and/or drop off some other requests). At the same time, a reasonable level of service should be maintained, and (hopefully) the use of vehicle resources should be “optimized”.

(6) Management of container terminals and loading/unloading of containers (Aslidis [1], ISL [26]). Containers are unloaded from a ship, stored in a container terminal, and loaded onto another ship (or onto a train, or fleet of trucks). The movement of containers is done by special container handling cranes. Information on the movement requirement of these vehicles arrives in real-time. A measure of throughput efficiency has to be optimized.

(7) Miscellaneous other problems
- In industrial manufacturing logistics, one may have the dynamic routing of robots on the production floor (Sabuncuoglu and Hommertzheim [51]).
- In paratransit (dial-a-ride) services, one may have the dynamic routing of small buses, to carry handicapped or other people from specified origins to specified destinations (Psaraftis [45]).
- Share-a-cab services may encounter a similar problem. A taxicab picks up passengers from distinct origins to a common destination (Deng et al. [15]).

One can notice that a common underlying feature of scenarios such as these is that the information that is needed to solve the related vehicle routing problem (that is, the input to the problem) is not known in its entirety ahead of time. Rather, this input (or part thereof) is dynamically revealed as time goes by. This prevents the decision maker from solving the entire problem at once (which he could attempt to do in “batch mode”, perhaps overnight, if all inputs to the problem were known in advance). He is therefore forced to do this in a sequential fashion: solve part of the problem on the basis of the information available now, and when or if this information changes (new input is revealed), try to solve a new problem that incorporates the new information. In doing so, he no longer has the luxury of the batch solution: he has to either produce “good” solutions as quickly as inputs are coming in, or else the solution (and his routing operation) will likely suffer.

Why are these problems increasingly important?

It is well known that the market for the distribution of goods is enormous. Suffice it to mention that in 1990 there were on the order of 13 million trucks operating in countries of the European Union (EU), and the ton-kilometers hauled by the EU road freight system were at least 800 billion. In Germany alone, a country that hauls on the order of 10% of all ton-kilometers in the EC, freight income amounted to about 60 billion DM (or about 35 billion US$), again for the road mode alone. Similar statistics can be found for countries in North America.
The situation in other transport modes is similar. In Europe, with increasing congestion on roads and highways, the importance of shortsea shipping becomes higher, as industrial enterprises depend on this mode especially for large-scale cargoes. Optimized routing of ships is therefore an equally important goal for the overall competitiveness of the market. A similar argument can be made for air transport, although the dependence of industrial enterprises on this particular mode is not of a scale comparable to the previous two.

The conclusion from all this is obvious: Distribution logistics is a big and vital business for the world’s industrial economies. Within systems of such size, even a modest percentage improvement in the overall cost of distribution can be very important for the overall competitiveness of the industries on the market, both internal and external, and might very well be critical for the survival of many businesses that depend on such distribution. In addition, and apart from the industrial enterprises, there are also non-industrial sectors that are relevant too, such as postal services, garbage collection, transportation of the disabled, etc. All of these sectors would benefit from an improved distribution scheme in a dynamic environment.

3. Technological advances

As mentioned earlier, one of the reasons dynamic vehicle routing is more important now than it was a few years ago is due to recent advances in information and communication technologies. Since such advances significantly enhance one’s ability to process information in real time, a whole new spectrum of opportunities that were hitherto nonexistent are revealed. We briefly discuss such advances, with an emphasis on those that have “key words” that have become more and more common these days.

**EDI (Electronic Data Interchange).** In a very general sense, EDI is the electronic transfer from computer to computer of commercial and administrative transactions using an agreed standard to structure the transaction or message data. As it relates to logistics and distribution, EDI aims at improving the speed of communication and control of all aspects of the logistical operation. Carriers, shippers, distributors, forwarders, terminals, warehouses, and all other players in the logistical game increasingly rely on EDI, and this is true for all modes of transport (road, rail, marine, air). In addition to systems that deal with cargo documentation, customs clearance, and cargo tracking, EDI may also include systems for automatic cargo identification (bar codes, programmable tags, etc.) and systems for automated cargo handling, sorting, consolidation, and flow control. The implementation of EDI systems is expanding into almost every aspect of the logistical chain (including manufacturing and retailing) and has enhanced the quality, availability, and value of information on almost any aspect of cargo movement. A recent report of the Commission of European Communities [13] analyzes in depth the impact of EDI...
on transport and makes a series of recommendations for further actions in this growing area. One of the conclusions of this study is that EDI is vital for transport systems that place a high value on speed of operation, and/or systems for which the value of the goods is high. This is typically the case in systems involving vehicle routing in real-time (see also Bollo et al. [8]).

**GPS (Global Positioning Systems).** Advances in satellite communications and cellular technology have made the exchange of data between vehicle and central dispatching office increasingly easy and affordable. In particular, the vehicle's position becomes now much easier to monitor, allowing for a better management of the vehicle fleet. Other vehicle information such as gas consumption, speed, route planning, driver work schedule, etc. can also be exchanged. Telephone, e-mail, fax, or automated data processing by a vehicle on-board computer can be used. The communication interface should be capable of linking all hardware and software systems, performing logical checks within messages, conversion of messages, splitting and merging of messages including access and authorization management. The use of a GPS has to be integrated with the use of GIS, described below.

**GIS (Geographic Information Systems).** In transportation, a GIS is a generalized electronic road map, displayed on a high-resolution graphics screen, and driven by a powerful workstation. This electronic map can be a valuable source of information for the distribution operation, by integrating geographical data (such as street and road information) with other data relevant in the specific application. A modern GIS integrates static and dynamic data. Static information includes maps and transport networks, while dynamic information includes anything that changes in terms of location (e.g. vehicle positions) or other attributes (e.g. appearance of a demand, size of demand, traffic conditions, etc.) The development and implementation of GIS systems has been a rapidly growing activity in recent years in both North America, Europe, and Japan. In Japan, 500,000 in-car autonomous navigation systems had been sold by 1992, and a Commission of European Communities study projected that by the year 2020 the cumulative investment in mobile “advanced telematics” equipment in Europe would be 50 billion ECU, in addition to a further 10 to 15 billion ECU in fixed equipment infrastructure [54]. For a recent survey of GIS, see also Fagan [20].

**IVHS (Intelligent Vehicle–Highway Systems).** Although the term IVHS has been coined in the United States, related activities exist also in Europe (projects PROMETHEUS and DRIVE) and elsewhere (project VICS – Vehicle Intelligent Control System, in Japan). In a general sense, IVHS is the application of ICT technologies in order to better control the flow of vehicles in a host of traffic situations. IVHS attempts to integrate advanced technologies and algorithms for the solution of problems ranging from the road transport of hazardous materials to the
control of traffic congestion, and from the real-time automated guidance of vehicles at high speeds to the reduction of traffic accidents. Chen and Ervin [11] give an overview of activities and policy issues in this rapidly growing area in the United States. In Europe, it is projected that by year 2010, some 90% of the 190 million European vehicles will be fitted with intelligent navigation, communications, and other driver information devices [54].

Many of the above technologies are being used or tested worldwide, although the degree of development is by no means uniform (the full implementation of IVHS, for instance, is still years ahead). In Europe, institutional developments outlined earlier increase the pressure for a better exploitation of such technological advances. In addition to program DRIVE (road transport telematics), other R&D programs such as ESPRIT (information technology), and RACE (telecommunications) are putting special emphasis on these issues, each from its own perspective. All these programs are important contributors to the stated industrial policy of the European Union of improving the competitiveness of its enterprises (and of the EU as a whole). To achieve this goal, all these programs involve (by design) the strong participation of the very industrial enterprises whose competitiveness is on the line, and place a high emphasis on the rapid implementation of the products of the research (for a recent report, see [54]). For the transport sector, as companies will increasingly operate on an international and intermodal basis, one will see more and more use of such technologies for a more efficient operation.

The significance of such developments as far as distribution problems are concerned (and in particular dynamic vehicle routing) is clear: Perhaps for the first time ever, the technologies for implementing an efficient real-time vehicle routing system are reaching (or have already reached) the maturity age: one that permits low-cost, fast processing of information and can integrate the various component technologies together. This means that many technological obstacles that have necessitated the adoption of static (or quasi-static) approaches to dynamic vehicle routing are slowly disappearing. It thus represents a first-class opportunity for the development and application of solution methods and management tools that hitherto were impossible or impractical to implement for this class of problems.

Having established in sections 2 and 3 (a) why the class of dynamic vehicle routing problems has become (and will be) increasingly important, and (b) that now is the time to take advantage of recent technological advances in this area, we are now in a better position to address the question on how the state of the art in dynamic vehicle routing methodology has evolved these last few years. This will be examined in sections 4 and 5.

4. What is really "dynamic"?: Attributes of information

Before we proceed with methodological issues, it is important to be explicit with what we mean by the word "dynamic". Although this may seem obvious or
redundant at first glance, it is important to clarify this issue not only for semantic purposes, but because an appropriate definition of this term may help better delineate this class of problems and related methodologies vis-à-vis the static case.

According to our definition, a vehicle routing problem is "dynamic" (also known as "real-time", or "on-line") if information (input) on the problem is made known to the decision maker or is updated concurrently with the determination of the set of routes. By contrast, if all inputs are received before the determination of the routes and do not change thereafter, the problem is termed "static".

We clarify here that in the above definition the word "problem" refers not to the actual real-world routing problem (most of which are dynamic anyway), but to the abstract problem that is defined by the corresponding analytical model. We make this distinction because it is the formulation of a VRP that is of most importance from a methodological standpoint. However, we should bear in mind that this distinction may have important ramifications, for a problem may be classified as static according to one model formulation and dynamic according to another. An example of this is the vehicle routing problem with stochastic demands, which will be discussed later.

Note also that according to the above definition, even if some of the input to the problem is known in advance and the rest is revealed afterwards, the problem is dynamic, as long as this additional input is explicitly taken into account within the model. Also, some vehicle routing problems that in a literal sense are considered dynamic, are essentially static problems. We clarify these issues with some examples.

EXAMPLE 1: THE TIME-DEPENDENT TSP [37, 38]

This is a traveling salesman problem in which the travel times (costs) between nodes are not constant, but change through time in a known fashion. For instance, due to changing traffic conditions in a certain network it is conceivable that whereas the time to go from node 1 to node 2 is 15 minutes if travel starts before 9:00 a.m., this time is 30 minutes if travel starts between 9:00 and 10:00 a.m., and 45 minutes if travel starts after 10:00 a.m.

It is essential to realize that to the extent that the entire time-wise evolution of the inter-node travel time matrix is known in advance and does not change, one can solve this problem once and for all ahead of time, with no need to reoptimize. Hence, this is a static problem according to our definition, even though the term “time-dependent” might be interpreted as synonymous to “dynamic”.

A similar problem class that is static according to our definition is the one examined by Laporte and Dejax [33] on combined location-routing problems. These are multi-period problems in which primary and secondary facilities may be opened in a period, and closed in another period. All of the problem inputs (and how these evolve in time) are deterministic and known in advance. Hence, this problem too can be solved once and for all (either exactly, or heuristically) with no need to reoptimize.
EXAMPLE 2: THE PROBABILISTIC TSP [27,28]

In the PTSP, travel times are deterministic, but demand at each node occurs with a known probability \( p \) (or does not occur with probability \( 1 - p \)). The PTSP calls for the determination of an "a priori" route \( R \) through all nodes of the network, so that the expected travel time of the actual route \( R' \) that will be traveled on a given day is minimized. The rule here is that before \( R' \) is traveled, the dispatcher is informed which nodes are actually requesting demand and which are not, and that \( R' \) follows the same sequence as \( R \), with nodes that have no demand on a particular day will simply be skipped.

The PTSP is a static problem because for the determination of \( R \) all inputs (travel time matrix and probability \( p \)) are known in advance and do not change. The determination of the optimal \( R \) has to be made before the dispatching of the vehicle. Hence, \( R \) is to be determined once and for all, and does not change. As to \( R' \), the additional information that is needed in order to determine it from \( R \) is which nodes are "on" and which are "off", and this information is made known also before the dispatching of the vehicle on a particular day (of course, a scenario in which the status of each node, and hence the actual route \( R' \), is made known in real-time is also plausible). In any event, note that the PTSP calls only for the determination of the a priori route \( R \), and as such is a static problem in all cases.

EXAMPLE 3: THE VEHICLE ROUTING PROBLEM WITH STOCHASTIC TRAVEL TIMES [34]

This is a VRP in which travel times and on-site service times are stochastic. The problem calls for the determination of a set of planned routes that minimize a certain objective that depends on the formulation. Three formulations are presented: a chance-constrained model, and two distinct recourse models. In all three models, the values realized by the stochastic inputs are revealed after the determination of the planned routes, even though the computation of these planned routes takes into account the stochastic parameters of these random variables (in a different way in each formulation).

As a general rule, if the output of a certain formulation is a set of preplanned routes that are not reoptimized and are computed from inputs that do not evolve in real-time, the problem is static. If, on the other hand, the output is not a set of routes, but rather a policy that prescribes how the routes should evolve as a function of those inputs that evolve in real-time, then the problem is dynamic.

If the above examples essentially refer to static vehicle routing problems, what is an example of a dynamic one?

EXAMPLE 4: THE DYNAMIC TSP [48]

In the simplest version of this problem, demands for service are generated at each node of a network according to a Poisson process of parameter \( \lambda \). The travel
times between nodes are known and deterministic, and the salesman spends a known service time at each node. What is the routing policy that minimizes the average, over all demands, expected time until service of the demand is completed? Alternatively, what is the routing policy that maximizes the average expected number of demands serviced per unit time?

This problem is dynamic because part of the input required to solve it (that is, which nodes actually request service) is revealed to the dispatcher concurrently with the determination of the route. Given this, it is impossible for an optimal route to be produced in advance. At best, what can be produced is a policy, specifying what action should be taken as a function of the state of the system. More about this problem and its variants later.

The above examples show that the way information about a particular routing problem evolves through time and is received by the decision maker is critical for the characterization of the problem as static or dynamic. More important, this plays an important role in determining which methodologies can be used.

The following taxonomy can be useful in characterizing attributes of information that forms the input of a certain vehicle routing problem (some of these concepts are borrowed from Polychronopoulos [43]):

(a) evolution of information: (static/dynamic);
(b) quality of information: (known-deterministic/forecast/probabilistic/unknown);
(c) availability of information: (local/global);
(d) processing of information: (centralized/decentralized).

We now discuss this taxonomy in some detail.

(a) Evolution of information. The differentiation between static and dynamic inputs is relevant here. The former are known for the entire duration of the routing process and are not updated (although they may very well be functions of time, as in the time-dependent TSP). The latter are not known for the entire duration of the routing process (which may be open ended anyway), and will generally be revealed or updated as time goes on.

(b) Quality of information. Some (or all) inputs to the problem can be known with certainty (deterministic), throughout the duration of the routing process. Examples: Number of nodes, number of vehicles, vehicle capacities, inter-node distances, etc. Some other inputs may not be known with certainty, but only as forecasts. These inputs are subject to revision as the routing process evolves. Examples: Demand at a certain node, travel times between certain node pairs, etc. Yet other inputs may be probabilistic, that is, follow prescribed probability distributions or evolve according to known stochastic processes. Examples: Demand location on a Euclidean instance is uniformly
distributed on the unit square, travel time between nodes follows a prescribed distribution, arc costs follow a Markov process, quantity demanded at a node follows a given distribution, etc. Finally, there may be certain inputs on which no information is available at the time of decision. Examples: The time at which the next demand for service is received, the location of that demand, etc.

It is important to realize that the attributes of information quality for a certain input variable are defined for the specific point in time at which the decision has to be made, and may change when time moves along (this is true for dynamic inputs). For instance, a certain variable may be probabilistic now, but becomes deterministic when the realization of its value is revealed. The same is true for forecast and unknown input variables. In addition, the values of probabilities of certain other variables may very well change, as a result of observations during the course of the routing process. As noted in Psarafitis [48], the quality of information in dynamic vehicle routing is usually good for near-term events, and becomes poorer for more distant events.

(c) **Availability of information.** There may be problems in which information is available only on a local basis. For instance, the travel time between two nodes may be random, and its actual value may be revealed (or forecast) only when the vehicle arrives at the starting node. The same may be true when the driver of a tank truck learns of the amount of oil needed to replenish a certain customer's inventory only on location. On the other hand, some inputs may be available globally. For instance, estimated travel times may be received by radio (or by some other device) even for remote parts of the network. Or, customer inventories may be automatically monitored by a device, and this information may be transmitted to the dispatcher on a global and continual basis.

Technologies such as those described in section 3 will generally increase the global availability of information, although the issue of who receives such information and when he receives it is a crucial parameter of system architecture and design. For instance, the central vehicle dispatcher may have at his fingertips all information on a global basis, although he will probably choose to reveal to the driver of a particular vehicle only information that is needed by that vehicle and driver.

(d) **Processing of information.** The two main schemes that exist regarding what is done with the information that is available are the centralized and decentralized ones. According to the former, all information is collected and processed by a central unit (be that a human scheduler, a man–machine computer-assisted dispatching system, or a fully automated system).
On the other hand, it may make sense for some of the information to be processed separately. For instance, if the driver of the truck is given latitude to decide his own routing for a certain set of demand points (either by himself, or by an on-board computer), we have a (partly) decentralized system.

Attribute (d) is very important with respect to the methodology that is used to solve the routing problem. Some of the partitioning or decomposition schemes that have been used for many vehicle routing problems in the past may lend themselves to processing decentralization. It is our opinion that such schemes are more important for dynamic routing problems, given the faster running time requirement of the dynamic scenario vis-à-vis the static one.

5. Methodological advances

To put methodological advances in perspective, it is perhaps useful to summarize the main differences between static and dynamic vehicle routing, as discussed in detail in Psaraftis [48]. These differences have implications on the solution techniques and information processing architectures that can be used:

1. Time dimension is essential.
2. Problem may be open ended.
3. Future information may be imprecise or unknown.
4. Near-term events are more important.
5. Information update mechanisms are essential.
6. Resequencing and reassignment decisions may be warranted.
7. Faster computation times are necessary.
8. Indefinite deferment mechanisms are essential.
9. Objective function may be different.
10. Time constraints may be different.
11. Flexibility to vary vehicle fleet size is lower.
12. Queueing considerations may become important.

The one point we would like to emphasize from the above list is point (9): In dynamic vehicle routing, traditional static objectives may be meaningless. Optimizing only over known inputs may be myopic if some other information (probabilistic) about future inputs is also available. Such information should be explicitly considered by the objective function.

We now examine some recent methodological developments. The focus is papers that have appeared since Psaraftis [48] was published. We first examine the work of Dror et al. [17] on vehicle routing problems with stochastic demands.
This class of problems (labeled as SVRP for stochastic VRP) generally involves the routing of a fleet of vehicles to service a set of demand points under the assumption that the size of the demand is a random variable (in fact, it is the only source of uncertainty in the problem). If customer demand cannot be met by a vehicle, then a route failure occurs, and in such case a remedial (recourse) action must be taken. The recourse action may be for the vehicle to return to the depot to reload and then continue its route following either the same sequence that was initially planned, or some other (reoptimized) sequence. Alternatively, a route may be broken prior to the occurrence of a failure, in order for the vehicle to reload and thus reduce the cost of a possible failure. Clearly, this is an important class of VRPs, and a generalization of the classical VRP.

Anything but a cursory study of this problem class reveals that more than one problem definition (and hence formulation) can be considered. Of those, the case where all demand realizations become known prior to route execution (case of full advance information) is clearly a deterministic (static) VRP. On the other hand, in the case demand is revealed during route execution (case of late information), the problem becomes more complicated and involves some further variants, depending on whether demand information is made available before delivery or during delivery. For these variants, the authors develop several stochastic programming models (chance-constrained programming, and recourse models), and a Markov decision model.

All stochastic programming formulations call for the determination of a preplanned set of routes and do not allow for reoptimization. In the chance-constrained models, constraints that the probability of route failure should be less than a prespecified constant translate into some corresponding deterministic (nonlinear) constraints. In the recourse models, the objective function includes terms that capture (either exactly, or approximately) the expected cost of the specific recourse action that is anticipated in response to a probable break in the routes. Since in both cases all inputs are received before the determination of the set of (preplanned) routes, these formulations are static, according to our earlier definition. The same can be said of the work of Gendreau et al. [25] on the VRP with stochastic customers and demands (this is a generalization of both the PTSP (Jaillet [27, 28]) and the SVRP as formulated above), and of the work of Dror et al. [18] on the SVRP with failures that are “restricted”, either by the data or by constraints.

However, not all SVRP formulations are static. An inherently dynamic SVRP formulation is also included in Dror et al. [17]. The authors formulate the single-vehicle version of the problem as a Markov decision process as follows: Actual demand is revealed upon arrival of the vehicle at the location of the customer. Then the vehicle can either replenish the customer and move to another location, or not replenish and move to the depot or to another customer.

The authors formulate this problem as a sequential decision process under uncertainty, in which the objective is to produce a policy that determines the best action as a function of the state of the system. No computational experience is
reported in the paper. Instead, the authors observe that since the number of possible states can be enormous, the solution would require some form of relaxation or aggregation.

The above is certainly a clue that what looks like a simple dynamic VRP (the single-vehicle VRP with stochastic demands) is probably very difficult to solve exactly. Indeed, the same can be said for a similarly "simple" (or maybe simpler) problem, the Dynamic TSP. As mentioned before, the DTSP can be thought of as a dynamic equivalent of the archetypal vehicle routing problem, the TSP. As such, it can perhaps be considered as the simplest dynamic VRP. Unfortunately, to our knowledge, as of today the general version of this particular problem remains unresolved, and only some properties of its solution in limiting cases have been discussed. For instance, it was conjectured in Psaraftis [48] that if $\lambda$ is extremely low, the DTSP is equivalent to the 1-median problem, for it may make sense for the vehicle to move to the graph's median in anticipation of the next demand. Such a policy would minimize the average expected system time. No other general results for this problem have been obtained, and it seems that this problem, however simple it may look, is inherently complex. This certainly does not bode well for other, more complex dynamic VRPs.

Nonetheless, the DTSP motivated significant research by Bertsimas and van Ryzin [5,6] on a related problem, the so-called Dynamic Traveling Repairman Problem (DTRP), for both the single vehicle and the multiple vehicle cases, with or without vehicle capacity constraints.

The DTRP is the Euclidean-plane equivalent of the DTSP, with an objective function that minimizes the average system time (waiting time plus service time). A general distribution with finite first and second moments is assumed for on-site service time. Using results from geometrical probability, queueing theory, and combinatorial optimization, Bertsimas and van Ryzin obtain some very interesting properties for this problem. These properties provide important insights on the structure of this problem. It is not the scope of this paper to describe this work in detail (much of the relevant research is actually ongoing). However, for motivation purposes we give a brief summary of the main results, starting with the single vehicle uncapacitated case:

1. In "light traffic" ($\lambda \to 0$), the optimal policy is the "stochastic queue median" (SQM) policy, that is, locate the vehicle at the median and serve the customers in an FCFS order, returning to the median after each service. This result formally proves the conjecture of Psaraftis [48] in the Euclidean case.

2. In "heavy traffic", they showed that there exist policies with finite system times $T$ for all $\rho = \lambda s < 1$ (where $s$ is the expected on-site service time), that are independent of the service region size and shape. They then examined several such policies, and showed that all these policies have the same asymptotic behavior, that is,
where $\gamma$ is a constant that depends on the policy, $A$ is the service area, and $v$ is the vehicle speed. Policies that were examined include:

- The partition policy ($\gamma \equiv 1.02$): divide region $A$ (assumed square here) into $n$ subsquares, and serve each region sequentially, with demands in each subsquare served in an FCFS order.
- The TSP policy ($\gamma \equiv 0.72$): as demands arrive, form them into sets of size $n$. When all $n$ demands have arrived, consider it the arrival of a set. Service sets in FCFS order by forming a TSP tour on the set of demands.
- The nearest neighbor policy ($\gamma \equiv 0.64$): after each service completion, serve next the demand that is closest to the vehicle.

In the first two policies, $n$ is optimized according to a method discussed in Bertsimas and van Ryzin [5].

These authors extend these results into the multiple vehicle, capacitated cases (Bertsimas and van Ryzin [6]). Among other things, they show that the stability condition for the queueing system now becomes

$$\rho + 2\lambda r/vq < 1,$$

where $r$ is the expected distance from a uniform location in $A$ to the closest depot, and $q$ is the capacity of the vehicles. The second term in this condition can be interpreted as the radial collection cost, or the average time required to reach a set of $q$ customers from the nearest depot.

For $q$ finite, the authors construct policies that have the same asymptotic behavior, namely,

$$T \sim \gamma^2 \lambda A (1 - 1/q)^2/m^2 v^2 (1 - \rho - 2\lambda r/mq v)^2, \quad \text{as } \rho + 2\lambda r/vq \to 1,$$

where again $\gamma$ is a constant that depends on the policy. The above formula is a generalization of the formula for the single-vehicle, uncapacitated case.

A variety of policies are analyzed, in a similar (but more complicated) fashion as in the single vehicle case. An interesting feature of all this work is that the above type of behavior is not limited only to the above (or other) sub-optimal policies, but characterizes an optimal policy as well. The value of $\gamma$ for the latter is not known precisely, but lower bounds for it can be obtained, depending on the scenario and various assumptions.

Overall, these results can be considered the equivalent of those of Karp [29] for the TSP, and of Jaillet [27, 28] for the PTSP. The routing policies that have been examined are simple, even naive, and they are not outputs of some optimization procedure (as one would perhaps expect). However, the results are remarkable in
that they obtain significant insights into the behavior of a complex dynamic routing problem. These insights can be useful both for strategic planning purposes and in an operational setting. As usual, some caution should be exercised in that most of these results are valid in an asymptotic sense, and it is not clear how slowly they converge (in fact, as Psarasftis [46] points out, asymptotic results for many TSP-related problems typically converge slowly).

What are the information attributes for this problem class? Using the taxonomy of section 4, we can make the following points:

(a) *Evolution of information.* A certain part of the input of this problem is static: The service region and all its properties (such as size, distance matrix, etc.), the number of vehicles, and the vehicle capacity and speed, do not change in time and are not subject to update. The same is true for the parameters of the two random processes that are at play, arrival rate \( \lambda \) and mean service time \( s \). On the other hand, the following inputs are dynamically revealed: time of a demand, location of a demand, location of all vehicles at the time a demand occurs, and service time of a demand.

(b) *Quality of information.* Static inputs are deterministic, whereas dynamic inputs are a priori probabilistic, becoming deterministic when they are realized. The location of all vehicles at the time a demand occurs is conditionally deterministic, that is, is a deterministic function of prior history and of the time a demand occurs.

(c) *Availability of information.* All information is globally available except the realization of on-site service time, which is revealed locally.

(d) *Processing of information.* The authors do not discuss this issue, so a "default" centralized scheme may be assumed. However, it is clear that at least one policy (the partition policy) can be implemented in a decentralized way.

Checking now the list of 12 "methodological" points that appeared in the beginning of this section, we have the following comments about this problem and the approach that was followed:

1. **Time dimension is essential:** Yes, in the DTRP it is.

2. **Problem may be open-ended:** Yes, the DTRP is open-ended.

3. **Future information may be imprecise or unknown:** Yes, future information is imprecise.

4. **Near-term events are more important:** Light-traffic policies have some look-ahead capability, by moving the vehicle to a median in anticipation of future demands. Heavy-traffic policies form routes in response to actual requests, and not in anticipation of future demands.
(5) **Information update mechanisms are essential:** Yes, information update mechanisms are part of the procedure (at the route execution level).

(6) **Resequencing or reassignment decisions may be warranted:** Resequencing is irrelevant here since no tentative routes are formed as a result of the policies. However, in the multiple-vehicle case, once points are partitioned and allocated to vehicles, there is no reassignment decision upon the appearance of a new request. Also, there is no impact of the actual realizations of on-site service times on decisions.

(7) **Faster computation times are necessary:** There is no information on the CPU time of these methods in an actual dynamic situation. Policies are computed off-line anyway and most are easy to compute. On-line CPU time may grow if TSP sub-tours are solved optimally unless the number of points is small.

(8) **Indefinite deferment mechanisms are essential:** There are no explicit mechanisms to prevent indefinite deferment of customer requests. Such may be the outcome for a customer located at the corner of the service region if a nearest neighbor policy is followed (heavy traffic case).

(9) **Objective function may be different:** Yes, the objective function in the DTRP (minimize mean system time) is indeed different from the traditional static TSP-type objective.

(10) **Time constraints may be different:** There are no time constraints in this problem.

(11) **Flexibility to vary vehicle fleet size is lower:** The number of vehicles is fixed.

(12) **Queueing considerations may be important:** Yes, queueing considerations are indeed very important in the DTRP, and are in fact the most important aspect of the overall analysis.

In all fairness, of course, the main thrust of this work was less to develop an on-line procedure that can be used in practice, and more to investigate the properties of this class of problems and the structure of their solutions. The fact that the policies examined are very simple was necessary in order to make the analysis tractable. Nevertheless, we thought it would be insightful to examine this method in the framework of dynamic vehicle routing problems that was presented earlier.

In our literature search, we also identified the work of Bagchi and Nag [2], on an expert systems approach to dynamic vehicle scheduling, and that of Lysgaard [35] on dynamic transportation networks in vehicle routing and scheduling. Last, but not least, Bertsimas and van Ryzin [6] extended their earlier work for the case of general demand and interarrival time distribution, and Bertsimas and Simchi-Levi [4] presented a survey of probabilistic approaches to vehicle routing.

We end this section by mentioning again that there are a host of recent developments in areas that methodologically can be considered "close" to that
of dynamic vehicle routing. Among those, one can mention the recent work of Polychronopoulos [43] and of Psaraftis and Tsitsiklis [49] on dynamic and stochastic shortest distance (or path) problems. Shortest path problems often appear as subproblems in vehicle routing problems, and therefore one may wish to borrow some of the related techniques.

5. Directions for further research

This author believes that the state of the art in inherently dynamic vehicle routing methods has definitely grown these last few years. Perhaps this growth is less spectacular as compared to the revolution that seems to be taking place in some of the related technologies, but it is nevertheless worthy of note. Still, there are important issues that merit continuing investigation. Some of these have been alluded to from the previous analysis.

First, and to the best of our knowledge, the network version of the DTSP defined in section 4 is as of yet unsolved. A challenging task would be to try to use the insights obtained from the DTRP for this problem. Work on “polling systems” (Takagi [55] and Kleinrock and Levy [31]) might also prove methodologically relevant. The development of non “ad hoc” policies for this problem is finally another task.

From a theoretical perspective, another interesting direction (quite distinct from the probabilistic approach) is the analysis of worst-case performance of dynamic vehicle routing algorithms. The pertinent question is: how far from the optimum can a dynamic routing algorithm deviate by only using information currently available and not the entire information (including future information) that could be available if the problem were static? Some research has been performed on dynamic shortest path problems by Papadimitriou and Yannakakis [42], and could prove useful here as well. In addition, Karp [30] has addressed the issue of how much it is worth to know future information (on-line versus off-line algorithms), with a specific focus on the multiple mobile server problem in which the objective is to minimize total travel time. This problem is easier than the multiple vehicle routing problem, but some of the concepts developed for the easier problem could be useful in the vehicle routing version as well.

In closing, we feel it is imperative to take advantage of the opportunities offered by the spectrum of technological advances outlined in section 3 for the development of an efficient dynamic vehicle routing architecture. Decentralization of information processing (by development of parallel system architectures) is a direction that should be given careful consideration.

Acknowledgement

I would like to thank the editors and three anonymous referees for their comments and suggestions.
References


