



CIRRELT

Centre interuniversitaire de recherche
sur les réseaux d'entreprise, la logistique et le transport

Interuniversity Research Centre
on Enterprise Networks, Logistics and Transportation

The Stochastic Eulerian Tour Problem

Srimathy Mohan
Michel Gendreau
Jean-Marc Rousseau

October 2007

CIRRELT-2007-45

Bureaux de Montréal :

Université de Montréal
C.P. 6128, succ. Centre-ville
Montréal (Québec)
Canada H3C 3J7
Téléphone: 514 343-7575
Télécopie: 514 343-7121

Bureaux de Québec :

Université Laval
Pavillon Palasis-Prince, local 2642
Québec (Québec)
Canada G1K 7P4
Téléphone: 418 656-2073
Télécopie: 418 656-2624

www.cirrelt.ca

The Stochastic Eulerian Tour Problem

Srimathy Mohan¹, Michel Gendreau^{2,3,*}, Jean-Marc Rousseau²

1. School of Global Management and Leadership, MC 2451, Arizona State University, P.O. Box 37100, Phoenix, AZ 85069-7100, USA
2. Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT), Université de Montréal, P.O. Box 6128, Station Centre-ville, Montréal, Canada H3C 3J7
3. Department of Computer Science and Operations Research, Université de Montréal, P.O. Box 6128, Station Centre-ville, Montréal, Canada H3C 3J7

Abstract. This paper defines the *Stochastic Eulerian Tour Problem* (SETP) and investigates several characteristics of this problem. Given an undirected Eulerian graph $G = (V, E)$, a subset R ($|R| = n$) of the edges in E that require service, and a probability distribution for the number of edges in R that have to be visited in any given instance of the graph, the SETP seeks an a priori Eulerian tour of minimum expected length. We derive a closed form expression for the expected length of a given Eulerian tour when the number of required edges that have to be visited follows a binomial distribution. We also show that the SETP is NP-hard, even though the deterministic counter part is solvable in polynomial time. We derive further properties and a worst case ratio of the deviation of the expected length of a random Eulerian tour from the expected length of the optimal tour. Finally, we present some of the desirable properties in a good a priori tour using illustrative examples.

Keywords. Arc routing, Eulerian tour problem, stochastic demand.

Acknowledgements. Financial support for this work was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) and by the Fonds FCAR (renamed Fonds québécois de la recherche sur la nature et les technologies, FQRNT). This support is gratefully acknowledged.

Results and views expressed in this publication are the sole responsibility of the authors and do not necessarily reflect those of CIRRELT.

Les résultats et opinions contenus dans cette publication ne reflètent pas nécessairement la position du CIRRELT et n'engagent pas sa responsabilité.

* Corresponding author: michel.gendreau@cirrelt.ca

This document is also published as Publication #1308 by the Department of Computer Science and Operations Research of the Université de Montréal.

Dépôt légal – Bibliothèque nationale du Québec,
Bibliothèque nationale du Canada, 2007

© Copyright Mohan, Gendreau, Rousseau and CIRRELT, 2007

1. Introduction

One of the most common problems in routing is the design of routes for people or vehicles delivering service. Such routing problems are of two types -- node routing and arc routing problems, depending on whether the service request is at a node or on an arc/edge. The underlying problem for most arc routing problems is determining a giant tour that starts and ends at a designated depot, and traverses all edges requiring service at least once. This is the deterministic *Eulerian Tour Problem* (ETP).

A connected graph is Eulerian if there exists a closed walk in the graph containing each edge exactly once. If the given graph is not Eulerian, the first step is to add a least cost set of arcs or edges to the graph to make it Eulerian. This is called the least cost augmentation problem. Edmonds and Johnson (1973) show that this problem can be solved in polynomial time for the undirected Chinese Postman Problem (CPP) using an adaptation of Edmond's blossom algorithm. While the least cost augmentation problem for directed graphs is also solvable in polynomial time, it is NP-hard when the underlying graph contains both arcs and edges. In this case, heuristics are used to make the graph Eulerian.

Given an Eulerian graph, an Eulerian tour can be determined in polynomial time. Edmonds and Johnson (1973) have described three different algorithms for the ETP on an undirected graph. These are the end-pairing algorithm, the next-node algorithm, and the maze-search algorithm. The ETP is well solved for directed and mixed graphs also. van Aardenne-Ehrenfest and de Bruijn describe the spanning arborescence algorithm for the ETP on directed graphs. For mixed graphs, one usually assigns directions to the undirected edges to transform the mixed graph into a symmetric graph, and then completely orient the remaining undirected edges so that the indegree equals to the outdegree for all vertices of the graph. The spanning arborescence algorithm can then be used to determine the Eulerian tour for this graph.

In this paper, we assume that we have solved the least cost augmentation problem and are given an undirected Eulerian graph $G = (V, E)$ in which the set $R (R \subseteq E)$ represents the set of edges that require service. It is important to note that there may be more than one Eulerian tour for a given graph. However, all these tours have the same cost and hence there is no optimization involved in the ETP. But

there exist quite a few situations in practice, when not all the edges that require service need to be visited everyday. In such cases, the number of edges that require a visit on any given day is a random variable. For example, consider a postal carrier who has to deliver mail to n different streets. The postal company wishes to minimize the total walking distance for the carrier. When the carrier has to visit all the n streets every day, any Eulerian tour would suffice, since all the Eulerian tours are of equal length. But in reality, based on the realization of demand, the carrier might have to visit only a subset of the streets requiring service on any particular day.

Consider the following alternative in that situation: the postal carrier follows the predetermined tour as long as he has to visit the next street on the tour to provide service. If at any point on the tour, the postal carrier does not have to visit a street, he skips that street, and takes the shortest path to the next street on the tour that requires a visit. With this alternative, the ETP takes on a different dimension. The different possible Eulerian tours of a graph yield themselves better to skipping certain edges of the graph. We use the example in Figure 1 to illustrate this. All edges of the underlying undirected graph have a length of 1 and the edges represented by solid lines require service. Node **O** represents the depot. The dotted lines represent the edges that are only traversed and not serviced. Tours 1 and 2 are two different Eulerian tours for the same graph. The numbers on the edges of the two tours represent the order in which one visits the edges in these tours.

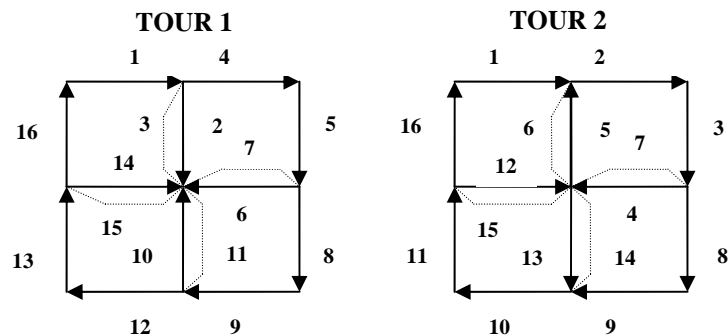


Figure 1. Two different tours for a 3x3 undirected graph

On a particular day, let us assume that edges **A**, **B**, **C**, and **D** require service. This translates to edges 2, 6, 10, and 14 on tour 1 and edges 4, 5, 12, and 13 on tour 2. If we start at the depot, visit the

edges in the same order that they appear in the respective tours and return to the depot, tour 1 results in a total length of 10 (Depot-1-2-7-6-11-10-15-14-15-16-Depot), while tour 2 results in a length of 12 (Depot-1-2-3-4-5-6-15-12-13-10-11-16-Depot). Thus, tour 1 is better for this instance. On the other hand, if edges **A**, **C**, **E** and **F** require service, tour 2 which has a length of 6 (Depot-1-2-3-4-5-1-Depot) is better than tour 1 which has a length of 8 (Depot-1-2-3-4-5-6-3-1-Depot). Hence, under such circumstances, the objective is to determine not just any Eulerian tour, but a particular tour (if more than one tour exists for the given graph) which has the shortest tour length “on an average”. This motivates the investigation of the *Stochastic Eulerian Tour Problem* (SETP) which we define below.

We are given an undirected Eulerian graph $G = (V, E)$, a set $R (R \subseteq E, |R| = n)$ of edges that require service (We shall call them “white” edges following the notation in [6].), and a distance $d(v_i, v_j)$ between every pair of directly connected nodes v_i and v_j . On any instance of the problem, only a subset of the n white edges is present, and hence, requires a visit. The number of present edges follows a specified probability distribution. The objective is to determine an a priori Eulerian tour that visits all the n edges and minimizes the expected length of the tour. On any given instance, one visits and services the present edges in the same order as in the a priori tour, while skipping the ones that are absent.

Our investigation of the SETP has also been motivated by a real-world problem. In the UK postal system, the postal carriers usually deliver mail a second time in the afternoon. During the first mail delivery, the carriers have to visit all the streets almost always, whereas the second mail delivery is typically very light. Only a small subset of the streets requires service during the afternoon delivery. While any Eulerian tour would be sufficient for the first mail delivery, it is definitely advantageous to determine a tour that minimizes the total length in an expected sense for the second mail delivery. It is important to note that even though the ETP is well solved, it is not feasible to determine a new tour for each day, since following a new tour every day would decrease the operating efficiency of the postal carrier considerably. In certain applications, like Canada Post, the mail carrier collects the mail to be delivered at various points along the route from relay boxes. On any given day, the present edges are

known only after the carrier starts his route and thus, it is not possible for the carrier to determine a new route at the start of each day. In such situations, it is certainly efficient to let the mail carrier follow the same route every day, while allowing the flexibility of skipping streets, if necessary.

A considerable amount of research has been done on deterministic arc routing. Eiselt et al. (1995a, 1995b) provide an excellent overview of this area of research. However, stochastic arc routing is a new area of research. Researchers have investigated several stochastic node routing problems over the past decade. In the following section, we present some of the related research on stochastic node routing. Section 3 states the definitions and assumptions for the SETP, and presents the method to obtain the expected length of a given tour efficiently. In this section, we also show that the SETP is NP-hard. We investigate some of the properties and derive bounds for the expected length of a given tour in Section 4. Finally, Section 5 highlights some of the desirable properties in an a priori tour using illustrative examples and Section 6 provides the conclusion and directions for future research.

2. Literature Review

Most of the current literature on arc routing addresses problems in a deterministic context. However, over the past few years, considerable amount of work has been done in understanding the nature of probabilistic node routing problems. We present here a brief summary of the literature on the probabilistic version of the Traveling Salesman Problem (TSP), which is closely related to the SETP. Researchers have also studied the m-TSP and the Vehicle Routing Problem with stochastic customers and stochastic demand. For results about these studies and a recent survey on stochastic vehicle routing, see Gendreat et al. (1995).

Jaillet (1985) introduced the TSP with stochastic customers as the *Probabilistic Traveling Salesman Problem* (PTSP). It is essentially a TSP where each vertex v_i is present with a probability p_i , and hence the number of vertices requiring a visit is a random variable. The recourse action Jaillet uses is to follow the a priori tour and simply skip absent customers. Under the assumption that $p_i = p$ for all

vertices, he derives closed form expressions for computing the expected length of a tour. He also derives bounds and several interesting properties of the problem. Jaillet (1998) and Jaillet and Odoni (1988) summarize most of the results in Jaillet (1985).

Jaillet shows that an optimal TSP tour can be arbitrarily bad for the PTSP. He also shows that an optimal tour for the PTSP may intersect itself in the Euclidean plane. This is in contrast to what we know about optimal TSP tours. These results indicate that algorithms have to be developed specifically with the PTSP in mind. Jaillet has developed a number of heuristics by suitably modifying several well-known TSP heuristics such as the Clarke-Wright algorithm and tour merging algorithms. Rossi and Gavioli (1988) present computational results after testing three of Jaillet's heuristics.

Bertsimas (1988) and Bertsimas and Howell (1993) have developed a few more heuristics based on probabilistic 2-opt edge exchange, vertex moves within a tour, and space filling curves. Laporte and Louveaux (1993) have developed a branch and cut algorithm called Integer L-Shaped method that is applicable to many stochastic programs with recourse. Laporte, Louveaux and Mercure (1989) have applied this method to the stochastic TSP and solved instances with up to 50 vertices optimally.

Recently, Fleury et al. (2005) have studied the Capacitated Arc Routing Problem (CARP) with stochastic demands. They have adapted a hybrid genetic algorithm developed by Lacomme et al. (2001) for the CARP to handle stochastic demands. They have tried several objective functions to study the robustness of the solutions developed by the hybrid genetic algorithm. Computational results indicate that the proposed algorithm produces robust solutions. Fleury et al. (2005) indicate the need for mathematical formulations for stochastic arc routing problems. This paper aims at developing a mathematical representation and analysis of a basic stochastic arc routing problem.

3. Important Results for the SETP

In this section, we first present the basic definitions and assumptions before formally defining the SETP. We then derive a closed form expression for calculating the expected length of a given tour t . We finally show that the SETP is NP-hard even though the ETP is solvable in polynomial time.

3.1. Definitions and Assumptions

$G = (V, E)$ is an undirected Eulerian graph where V is the set of nodes and E is the set of edges. A subset R ($|R| = n$) of the edges in E that requires service denotes the set of white edges. Associated with each edge (v_i, v_j) in E is a non-negative real number $d(v_i, v_j)$, the direct distance from node v_i to node v_j . The graph G has a node designated as the depot where the Eulerian tour starts and ends. In order to facilitate the representation and analysis, we duplicate the depot and represent the duplicated node as v_0 , which now serves as the depot. The duplicated node v_0 is connected to the original depot by two edges of length 0.

Given an Eulerian tour t , we have an ordering of the nodes and edges, and thus, a direction of traversal (and service) for each of the n edges in R . If we traverse edge e_i from node v_k to v_l , we define v_k as the in-node for edge e_i (v_i^{in}) and v_l as the out-node for edge e_i (v_i^{out}). Thus, given the in-node and the out-node for each edge in R , we represent an Eulerian tour t as $t = (v_0, v_1^{in}, e_1, v_1^{out}, v_2^{in}, e_2, v_2^{out}, \dots, v_n^{in}, e_n, v_n^{out}, v_0)$, where the edges e_1, e_2, \dots, e_n are numbered in their order of appearance in tour t . The length of the tour t , $L(t)$ is given by:

$$L(t) = \sum_{i=1}^n l(e_i) + \sum_{i=0}^n d(v_i^{out}, v_{i+1}^{in}) \quad (1)$$

$$\text{with } v_0^{out} = v_{n+1}^{in} = v_0, \text{ and } l(e_i) = \text{length of edge } e_i$$

If nodes v_i and v_j are not directly connected, then $d(v_i, v_j)$ is the shortest distance between v_i and v_j .

Each edge e_i in R is present with probability p_i . Thus, for any given instance, the number of white edges present (i.e., requiring a visit) is a random variable. We assume that if k edges require a visit on a particular day, then every set of k edges out of the n white edges is equally likely. Note that when $p_i = p$ for all i , the number of present edges follows a binomial distribution.

Thus, given $G = (V, E)$, a set of n white edges, a distance matrix D and a probability p_i of white edge e_i being present, the SETP seeks an a priori tour that minimizes the expected length of the tour. Looking at it as a stochastic program with recourse, in the first stage, we construct an a priori Eulerian tour of minimum expected length. Once we know the set of present edges, we can describe the second stage solution as follows -- start at the depot, travel to the in-node of the first present edge via the shortest path, traverse and service the first edge and then take the shortest path from the out-node of the first present edge to the in-node of the second present edge. We continue in a similar manner until we reach the out-node of the last present edge and then take the shortest path back to the depot. Given this recourse action and the precise definition for tour representation, we are ready to present the results for calculating $E[L_t]$, the expected length of a given tour t .

3.2. Expected Length of a Given Tour

The length of any given tour consists of two parts, namely, the total length of the present white edges, and the total distance traveled from the out-node of one present white edge to the in-node of the next present white edge (i.e., the inter-edge traversal distances). The SETP is similar to the PTSP with n white nodes and one depot in certain aspects. The inter-city traversal distances in the PTSP would correspond to the inter-edge traversal distances. The main difference between the two problems is that in the SETP, in addition to the inter-edge traversal distances, we have to consider the length of the white edges also. Thus, many of our results are extensions of Jaillet's (1985, 1988) results for the PTSP.

In order to derive a concise expression for the inter-edge traversal distances, we define the following n quantities.

$$\text{Let } L_t^r = \sum_{j=0}^n d(v_j^{out}, v_{j+r+1}^{in}) \quad \forall r \in \{0, \dots, n-1\} \quad (2)$$

$$\text{where } v_0^{out} = v_{n+1}^{in} = v_0$$

$$\overline{j+r+1} = (j+r) \bmod (n+1) + 1$$

$$d(v_j^{out}, v_{j+r+1}^{in}) = \begin{cases} d(v_j^{out}, v_{j+r+1}^{in}) & \text{if } 0 \leq j \leq n-r \\ d(v_j^{out}, v_0) + d(v_0, v_{j-n+r}^{in}) & \text{if } n-r < j \leq n \end{cases}$$

Note that when $r = 0$, L_t^0 is the total inter-edge traversal distance for the given tour t and $L_t^0 + \sum_{i=1}^n l(e_i)$

is the length of the given Eulerian tour t . For $1 \leq r \leq n-1$, L_t^r is the sum of $(n+1)$ elements. Each element represents the distance from the out-node of edge e_j to the in-node of its $(r+1)^{th}$ successor edge (i.e., edge e_{j+r+1}) with respect to the given tour t . More precisely, we start at the out-node of edge e_j , skip the next r edges on the tour and travel to the in-node of the $(r+1)^{th}$ edge following edge e_j to calculate L_t^r . Note that when $n-r < j \leq n$, to reach the in-node of edge e_{j+r+1} from the out-node of edge e_j , we define $d(v_j^{out}, v_{j+r+1}^{in})$ as reaching the depot from the out-node of e_j and then traveling from the depot to the in-node of e_{j+r+1} . We first obtain the conditional expected length of a given tour t when k of the n white edges are not present.

Lemma 1. *Given a graph G with n white edges, a designated depot v_0 , and a probability of occurrence p for each white edge, the conditional expected length of a tour t , given k of the n white edges are not present in the given tour t is*

$$E[L_t | (n-k) \text{ edges present}] = \begin{cases} 0 & \text{if } k = n \\ (1/n) \left[\sum_{i=1}^n l(e_i) + L_t^{n-1} \right] & \text{if } k = n-1 \\ \left(\frac{1}{\binom{n}{k}} \right) \left[\sum_{i=1}^n \binom{n-1}{k} l(e_i) + \sum_{r=0}^k \binom{n-2-r}{k-r} L_t^r \right] & \text{if } k = 0, 1, \dots, n-2 \end{cases} \quad (3)$$

Proof. (i) $k = n$: Since none of the white edges are present, this case is obvious.

(ii) $k = n - 1$: Only one white edge is present and each one of the n edges is equally likely to be present. We have to travel from the depot to the in-node of this present edge, service the edge, and travel from its out-node back to the depot. Thus, from the definition of L_t^{n-1} , this case follows.

(iii) $k = 0, 1, \dots, n - 2$: The expected length is the sum of the total length of the white edges traversed (the first term) and the total inter-edge traversal distances (the second term in the expression).

When k edges are missing, the resulting total inter-edge traversal distance is a sum of $n - k$ elements. Since some of the elements might be repeated in the various combinations, we regroup them using L_t^r . Consider an element $d(v_j^{out}, v_{j+r+1}^{in})$ of L_t^r for a given $r \in \{0, 1, \dots, n - 2\}$. For this element to be included, the white edges e_j and e_{j+r+1} have to be present and the edges between them have to be absent. Since we have only k edges missing, if $r > k$, $d(v_j^{out}, v_{j+r+1}^{in})$ will never appear and L_t^r will not be used in calculating $E[L_t | (n - k) \text{ edges present}]$. However, if $r \leq k$, we still need to choose the remaining $k - r$ white edges that are missing from a total of $n - 2 - r$ available white edges, and this can be done in $\binom{n - 2 - r}{k - r}$ ways. Since this is valid for all $j \in \{0, 1, \dots, n\}$, it follows that each L_t^r appears $\binom{n - 2 - r}{k - r}$ times while calculating $E[L_t | (n - k) \text{ edges present}]$.

Now, we have to account for the number of times the white edges are actually traversed. When k out of the n edges are missing, we have $\binom{n}{n - k} = \binom{n}{k}$ different combinations of the present edges. Each combination has $(n - k)$ white edges. Since each one of the n white edges appears an equal number of times over all the combinations, the number of times an edge e_i occurs in $E[L_t]$ is

$$\binom{n}{k} (n - k) / n = \binom{n - 1}{k}. \quad \blacksquare$$

Now that we have a closed form expression for the conditional expected length, we can calculate $E[L_t]$ using the appropriate probabilities.

Theorem 1. *Given a graph G , with n white edges, a designated depot v_0 , and a probability of occurrence p for each white edge, the expected length of a given tour t is*

$$E[L_t] = p^2 \left[\sum_{r=0}^{n-2} (1-p)^r L_t^r \right] + p(1-p)^{n-1} L_t^{n-1} + p \left[\sum_{i=1}^n l(e_i) \right] \quad (4)$$

Proof. $E[L_t] = \sum_{k=0}^n E[L_t | n-k \text{ edges present}] \times \text{Prob}\{n-k \text{ edges present}\}$ (5)

$$\text{where, } \text{Prob}\{n-k \text{ edges present}\} = \binom{n}{k} p^{n-k} (1-p)^k$$

$$\begin{aligned} E[L_t] &= \sum_{k=0}^{n-2} \left[\left(\frac{1}{\binom{n}{k}} \right) \left(\sum_{i=1}^n \binom{n-1}{k} l(e_i) + \sum_{r=0}^k \binom{n-2-r}{k-r} L_t^r \right) \right] \left[\binom{n}{k} p^{n-k} (1-p)^k \right] + \\ &\quad (1/n) \left[\sum_{i=1}^n l(e_i) + L_t^{n-1} \right] \left[\binom{n}{n-1} p (1-p)^{n-1} \right] \\ &= \sum_{k=0}^{n-2} \left[\sum_{r=0}^k \binom{n-2-r}{k-r} L_t^r p^{n-k} (1-p)^k \right] + \sum_{k=0}^{n-2} \left[\sum_{i=1}^n \binom{n-1}{k} l(e_i) p^{n-k} (1-p)^k \right] + \\ &\quad \sum_{i=1}^n l(e_i) p (1-p)^{n-1} + L_t^{n-1} p (1-p)^{n-1} \end{aligned} \quad (6)$$

Let us now consider each term of (6).

The first term can be expressed as:

$$\sum_{r=0}^{n-2} L_t^r \left[\sum_{k=r}^{n-2} \binom{n-2-r}{k-r} p^{n-k} (1-p)^k \right] \quad (7)$$

Setting $u = k - r$ and $s = n - 2 - r$, (7) reduces to

$$\sum_{r=0}^{n-2} p^2 (1-p)^r L_t^r \left[\sum_{u=0}^s \binom{s}{u} p^{s-u} (1-p)^u \right] = \sum_{r=0}^{n-2} p^2 (1-p)^r L_t^r \quad (8)$$

The second and third terms can be combined as:

$$\begin{aligned} & \sum_{i=1}^n l(e_i) \left[\sum_{k=0}^{n-1} \binom{n-1}{k} p^{n-k} (1-p)^k \right] \\ &= p \sum_{i=1}^n l(e_i) \left[\sum_{k=0}^{n-1} \binom{n-1}{k} p^{n-1-k} (1-p)^k \right] = p \sum_{i=1}^n l(e_i) \end{aligned} \quad (9)$$

From (6), (8), and (9), we get

$$E[L_t] = p^2 \left[\sum_{r=0}^{n-2} (1-p)^r L_t^r \right] + p(1-p)^{n-1} L_t^{n-1} + p \left[\sum_{i=1}^n l(e_i) \right] \quad \blacksquare$$

Since each L_t^r is the sum of $n+1$ elements, given a tour t , we can calculate $E[L_t]$ in $O(n^2)$ time using (4). Also, note that if the number of white edges present does not follow a binomial distribution, but some other specified probability distribution, we can substitute the appropriate probabilities in (5) to calculate $E[L_t]$. However, the $(n-k)$ present edges must be chosen at random from the set of n white edges to use (5). In certain scenarios when each white edge e_i is present with probability p_i , we can calculate $E[L_t]$ using the following formula.

$$\begin{aligned} E[L_t] &= \sum_{i=1}^n p_i l(e_i) + \sum_{i=1}^n d(v_0, v_i^{in}) p_i \prod_{k=1}^{i-1} (1-p_k) + \sum_{i=1}^n d(v_i^{out}, v_0) p_i \prod_{k=i+1}^n (1-p_k) \\ &+ \sum_{i=1}^n \sum_{j=i+1}^n d(v_i^{out}, v_j^{in}) p_i p_j \prod_{k=i+1}^{j-1} (1-p_k) \end{aligned} \quad (10)$$

We can derive (10) by looking at the probability of the following events:

- each white edge being present (first term)
- link between the depot and the in-node of each white edge being present (second term)
- link between the out-node of each white edge and the depot being present (third term)
- link between the out-node of a white edge e_i and the in-node of each white edge following e_i in tour t being present (fourth term)

Expressions (4) and (10) are similar to the closed form expression for calculating the expected length of a given tour for the PTSP with one black node (the depot) and n white nodes. The main difference is that for the SETP, we have to consider the length of the white edges in addition to the distance traveled between edges. We use this similarity to prove that the SETP is NP-hard.

Theorem 2. *The SETP is NP-hard.*

Proof. The SETP belongs to class NP, since given a tour t , we can calculate $E[L_t]$ in polynomial time ($O(n^2)$) that we can then compare with a bound B . We reduce the PTSP to the SETP to show that it belongs to the class NP-complete.

Given an instance of the PTSP, we construct the graph $G = (V, E)$ as follows:

For each white node v_i in the PTSP, define two nodes v_i^{in} and v_i^{out} .

$$V = \{v_i^{in}, v_i^{out} \mid \forall i = 1, \dots, n\} \cup \{v_0\}$$

$$E = \{(v_i^{in}, v_i^{out}), \forall i = 1, \dots, n\} \cup \{(v_0, v_i^{in}), (v_0, v_i^{out}) \mid \forall i = 1, \dots, n\} \\ \cup \{(v_i^{out}, v_j^{in}), \forall (v_i, v_j) \in E'\}$$

$$R = \{(v_i^{in}, v_i^{out}), \forall i = 1, \dots, n\}$$

$$d(v_i^{in}, v_i^{out}) = 0, \forall i = 1, \dots, n; d(v_i^{out}, v_j^{in}) = d'(v_i, v_j), \forall (v_i, v_j) \in E', i, j = 1, \dots, n$$

$$d(v_0, v_i^{in}) = d(v_0, v_i^{out}) = d'(v_0, v_i), i = 1, \dots, n$$

$$B = B'$$

Let $t = (v_0, v_1^{in}, e_1, v_1^{out}, v_2^{in}, e_2, v_2^{out}, \dots, v_n^{in}, e_n, v_n^{out}, v_0)$ be a feasible Eulerian tour for G with $E[L_t] \leq B$. Since each edge e_i is of length 0, it is clear that from the Eulerian tour t , we can construct a tour $t' = (v_0, v_1, \dots, v_n, v_0)$, which is feasible for G' . Also, note that $L_t^r = L_{t'}^r$ for all $r = 0, \dots, n-1$ by construction of G , and hence $E[L_{t'}] = E[L_t] \leq B = B'$. Similarly, we can show that if the PTSP has a

feasible solution with $E[L_{t'}] \leq B'$, then the SETP has a feasible solution with $E[L_t] \leq B$, and hence the SETP is NP-hard. ■

4. Properties and Bounds for the SETP

In this section, we examine a few properties of the SETP. In order to derive these properties, we express $E[L_t]$ succinctly in a weight-form notation. Let W be the random variable that represents the number of present white edges.

$$E[L_t] = \sum_{r=0}^{n-1} \alpha_r L_t^r + C \sum_{k=0}^{n-1} \beta_k$$

where, $\alpha_r = \sum_{k=r}^{n-2} \left[\frac{\binom{n-2-r}{k-r}}{\binom{n}{k}} \right] \text{Prob}(W = n-k) \quad \forall r \in \{0, \dots, n-2\}$

$$\alpha_{n-1} = \text{Prob}(W = 1)/n, \quad \beta_k = \left(\frac{n-k}{n} \right) \text{Prob}(W = n-k), \text{ and } C = \sum_{i=1}^n l(e_i).$$

We first describe properties of α_r , β_k and L_t^r . We then use these properties to derive an expression for the maximum deviation of the expected length of a given Eulerian tour t from the expected length of the optimal Eulerian tour for the SETP, t^* .

Property 1. *Given a tour t for an Eulerian graph G with n white edges,*

L_t^{n-1} is a constant independent of t ;

L_t^1, \dots, L_t^{n-2} are tour-dependent; and

$L_t^0 + C$ is the length of the tour t .

Proof. From the definition of L_t^r , we see that $L_t^{n-1} = \sum_{i=1}^n [d(v_i^{out}, v_0) + d(v_0, v_i^{in})]$, and hence, it is tour

independent. However, L_t^r , $r = 1, \dots, n-2$ depend on the order in which the edges are visited, and hence

are tour independent. By definition, L_t^0 gives the shortest distances between the out-node of the edge e_i to the in-node of edge e_{i+1} for all white edges. Thus, $L_t^0 + \sum_{i=1}^n l(e_i)$ is the length of the given tour t .

Note that in the PTSP, L_t^0 is the length of the given tour t . ■

Property 2. *The set of edges that define L_t^r along with E , the set of required edges, consists of $(r + 1)$ sub-tours, each starting and ending at the depot v_0 .*

Proof. By the definition used in (2), the term L_t^r contains $(n + 1)$ terms. Of these, exactly r terms are sum of two distances as defined in (2) since these have to pass through the depot before reaching the destination in-node of the next service edge. Also, among the $(n + 1) - r$ terms that have only one distance measure, the depot will be present in exactly two terms. Thus, it follows that the $(n + 1) + r$ terms that make up L_t^r , along with the set of service edges, will contain exactly $(r + 1)$ sub-tours, each starting and ending at the depot. ■

It is important to note that even if the given distance matrix D is symmetric (and hence the matrix of shortest path distances is also symmetric), $d(v_i^{out}, v_j^{in})$ is in general, not equal to $d(v_j^{out}, v_i^{in})$. Hence, several properties of the PTSP do not hold for the SETP. However, if we assume D to satisfy the triangular inequality, then $d(v_i^{out}, v_j^{in}) \leq d(v_i^{out}, v_k^{in}) + d(v_k^{out}, v_j^{in})$, i.e., the inter-edge traversal distances also satisfy the triangular inequality. Under this condition, we can deduce the following.

Property 3. *If D satisfies the triangular inequality, we can show that, for a given tour t of an Eulerian graph G ,*

$$L_t^r \geq L_t^0 \quad \forall r \in \{0, \dots, n-1\}$$

Proof. This follows since,

$$\begin{aligned} L_t^r + C &= \text{length of } (r+1) \text{ sub-tours with } v_0 \text{ as a common node} \\ &\geq \text{length of the given Eulerian tour } t \\ &= L_t^0 + C \end{aligned}$$

From property 2, we know that the set of edges that define L_t^r along with the set of white edges form $(r+1)$ sub-tours with the depot as a common node. Under the triangularity assumption, we can merge these $(r+1)$ sub-tours into a single tour whose length will be less than or equal to $L_t^r + C$. The length of this merged tour is in turn greater than or equal to $L_t^0 + C$, the length of the given Eulerian tour t .

$$\text{Hence, } L_t^r \geq L_t^0 \quad \forall r \in \{0, \dots, n-1\}. \quad \blacksquare$$

Property 4. Given a tour t of an Eulerian graph G with a depot v_0 and n white edges,

$$L_t^r \leq L_t^{r_1} + L_t^{r-r_1-1} \quad \forall r \in \{1, \dots, n-1\}, \text{ and } 0 \leq r_1 \leq r-1.$$

$$\text{Hence, } L_t^r \leq (r+1)L_t^0 \quad \forall r \in \{1, \dots, n-1\}$$

$$\begin{aligned} \text{Proof. } L_t^r &= \sum_{j=0}^n d(v_j^{out}, v_{j+r+1}^{in}) \\ &\leq \sum_{j=0}^n d(v_j^{out}, v_{j+r_1+1}^{in}) + \sum_{j=0}^n d(v_{j+r_1+1}^{out}, v_{j+r+1}^{in}) \end{aligned}$$

by the triangularity assumption

$$\text{But, } \sum_{j=0}^n d(v_{j+r_1+1}^{out}, v_{j+r+1}^{in}) = \sum_{j=0}^n d(v_j^{out}, v_{j+r-r_1-1+1}^{in})$$

$$\text{Hence, } L_t^r \leq L_t^{r_1} + L_t^{r-r_1-1} \quad \forall r \in \{1, \dots, n-1\}, \text{ and } 0 \leq r_1 \leq r-1$$

If we let, $r_1 = r - 1$ in $L_t^r \leq L_t^{r_1} + L_t^{r-r_1-1}$, we get $L_t^r \leq L_t^{r-1} + L_t^0$. Using the same relationship for L_t^{r-1} , we get $L_t^r \leq L_t^{r-2} + 2L_t^0$. Thus,

$$L_t^r \leq (r+1)L_t^0 \quad \forall r \in \{1, \dots, n-1\} \quad \blacksquare$$

Property 5. Given a discrete probability distribution for W ,

$$(i) \quad \sum_{r=0}^{n-1} \alpha_r = E[W]/n \quad \forall n \geq 1$$

$$(ii) \quad \sum_{r=0}^{n-1} (r+1)\alpha_r = 1 - \text{Prob}[W = 0] \quad \forall n \geq 1$$

$$(iii) \quad \sum_{k=0}^{n-1} \beta_k = E[W]/n \quad \forall n \geq 1$$

$$(iv) \quad \sum_{k=0}^{n-1} (n/(n-k))\beta_k = 1 - \text{Prob}[W = 0] \quad \forall n \geq 1$$

$$\begin{aligned} \text{Proof. (i)} \quad \sum_{r=0}^{n-2} \alpha_r &= \sum_{r=0}^{n-2} \left[\sum_{k=r}^{n-2} \left(\binom{n-2-r}{k-r} / \binom{n}{k} \right) \text{Prob}(W = n-k) \right] \\ &= \sum_{k=0}^{n-2} \left(\text{Prob}(W = n-k) / \binom{n}{k} \right) \sum_{r=0}^k \binom{n-2-r}{k-r} \end{aligned} \quad (11)$$

$$\text{But,} \quad \sum_{r=0}^k \binom{n-2-r}{k-r} = \binom{n-1}{k} \quad (12)$$

From (11) and (12), we get

$$\sum_{r=0}^{n-2} \alpha_r = \sum_{k=0}^{n-2} \binom{n-1}{k} / \binom{n}{k} \text{Prob}(W = n-k) \quad (13)$$

By replacing $n-k$ by u in (13), we get

$$\sum_{r=0}^{n-2} \alpha_r = \sum_{u=2}^n \frac{u}{n} \text{Prob}(W = u) = \frac{1}{n} [E[W] - \text{Prob}(W = 1)]$$

Since $\alpha_{n-1} = \frac{\text{Prob}(W = 1)}{n}$, the result follows.

$$\begin{aligned} \text{(ii)} \quad \sum_{r=0}^{n-2} (r+1) \alpha_r &= \sum_{r=0}^{n-2} \left[\sum_{k=r}^{n-2} \left(\frac{\binom{n-2-r}{k-r}}{\binom{n}{k}} \right) (r+1) \text{Prob}(W = n-k) \right] \\ &= \sum_{k=0}^{n-2} \left(\text{Prob}(W = n-k) \frac{\binom{n}{k}}{\binom{n}{k}} \right) \sum_{r=0}^k (r+1) \binom{n-2-r}{k-r} \end{aligned}$$

But,
$$\sum_{r=0}^k (r+1) \binom{n-2-r}{k-r} = \binom{n}{k}$$

Hence,
$$\sum_{r=0}^{n-2} (r+1) \alpha_r = \sum_{k=0}^{n-2} \text{Prob}(W = n-k) = 1 - \text{Prob}(W = 0) - \text{Prob}(W = 1)$$

Since $\alpha_{n-1} = \frac{\text{Prob}(W = 1)}{n}$, the result follows.

$$\begin{aligned} \text{(iii)} \quad \sum_{k=0}^{n-1} \beta_k &= \sum_{k=0}^{n-2} \beta_k + \beta_{n-1} \\ \sum_{k=0}^{n-2} \beta_k &= \sum_{k=0}^{n-2} \left(\frac{n-k}{n} \right) \text{Prob}(W = n-k) \\ &= \sum_{u=2}^n \left(\frac{u}{n} \right) \text{Prob}(W = u) \\ &= \frac{1}{n} [E[W] - \text{Prob}(W = 1)] \end{aligned}$$

Since, $\beta_{n-1} = \text{Prob}[W = 1]/n$, the result follows.

$$\text{(iv)} \quad \sum_{k=0}^{n-1} (n/(n-k)) \beta_k = \sum_{k=0}^{n-1} \text{Prob}[W = n-k] = 1 - \text{Prob}[W = 0] \quad \blacksquare$$

Lemma 2. Given a graph G with n white edges, a depot v_0 , for any given tour t ,

$$(i) \quad E[L_t] \geq (E[W]/n)[L_t^0 + C]$$

$$(ii) \quad E[L_t] \leq (1 - \text{Prob}[W = 0]) [L_t^0 + C]$$

Proof.

$$(i) \quad E[L_t] = \sum_{r=0}^{n-1} \alpha_r L_t^r + C \sum_{k=0}^{n-1} \beta_k$$

$$\geq L_t^0 \sum_{r=0}^{n-1} \alpha_r + C \sum_{k=0}^{n-1} \beta_k \quad \text{[from Prop. 3]}$$

$$= (E[W]/n)[L_t^0 + C] \quad \text{[from Prop. 5]}$$

$$(ii) \quad E[L_t] = \sum_{r=0}^{n-1} \alpha_r L_t^r + C \sum_{k=0}^{n-1} \beta_k$$

$$\leq L_t^0 \sum_{r=0}^{n-1} (r+1) \alpha_r + C \sum_{k=0}^{n-1} (n/(n-k)) \beta_k \quad \text{[from Prop. 4]}$$

$$= (1 - \text{Prob}[W = 0]) [L_t^0 + C] \quad \text{[from Prop. 5]} \quad \blacksquare$$

Next, we derive the worst case ratio for the expected length of a random Eulerian tour when compared to the optimal tour for the SETP.

Theorem 3. *Given a graph G with n white edges, and a designated depot, a distance matrix D that satisfies the triangular inequality, the optimal tour t^* for the SETP and a random Eulerian tour t ,*

$$(E[L_t] - E[L_{t^*}]) / E[L_{t^*}] \leq (1 - E[W]/n - \text{Prob}(W = 0)) / (E[W]/n)$$

Proof. From Lemma 2, $E[L_{t^*}] \geq (L_{t^*}^0 + C)[E[W]/n]$

Since all Eulerian tours are of the same length, $L_t^0 + C = L_{t^*}^0 + C$, and

$$E[L_{t^*}] \geq (L_t^0 + C)[E[W]/n] \quad (14)$$

$$\text{Also, } E[L_t] - E[L_{t^*}] \leq (L_t^0 + C)[1 - E[W]/n - \text{Prob}[W = 0]] \quad (15)$$

Dividing (15) by (14), we get the desired result. ■

Note that when W follows a binomial distribution, with parameter p , $E[W] = np$ and hence, Theorem 3 implies $(E[L_t] - E[L_{t^*}]) / (E[L_{t^*}]) \leq (1 - p - (1 - p)^n) / p$.

5. Illustrative Example

In this section, we present an example to investigate the characteristics of different a priori tours for a given graph. Consider the 3x3 grid network given earlier in Figure 1. All vertical edges are of length 1 and all horizontal edges are of length 2. The dotted lines indicate the edges traversed but not serviced on the Eulerian tour t . All edges represented by solid lines are white and W is binomial with parameter p , ($0 < p < 1$). Note, that the length of any tour is 24. Several tours are possible for this graph. Let us consider tours 3 and 4 given in Figure 2 and compare their expected lengths.

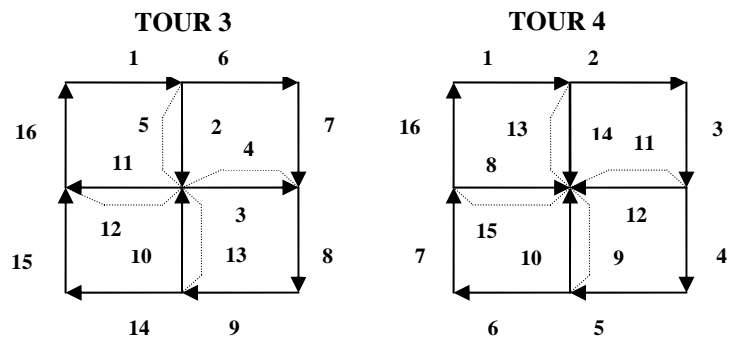


Figure 2. Tours 3 and 4 for a 3x3 undirected graph

The length of both tours is 24. However, the expected length of tour 3, $E[L_{t_3}]$, is less than $E[L_{t_4}]$ for all values of p between 0 and 1. Specifically, for $p = 0.45$, $E[L_{t_3}]$ is 17% lower than $E[L_{t_4}]$. The reason for this is quite obvious from the nature of the tours. In tour 3, edges (2,3,4,5) and edges (10,11,12,13) together form two separate sub-tours. This allows one to skip these sub-tours when the respective edges are not present. On tour 4, we can reach the inner sub-tour only after traversing most

of the outer sub-tour. Thus, inherently tour 4 necessitates traversing more edges than tour 3, on most instances. Thus, we see that the number of sub-tours clearly has an effect on the expected length of the tour.

Another interesting observation is the effect of the orientation of the edges in the inner sub-tour of tour 4. All the edges are serviced from the outside towards the center. If on a particular instance, edges 2, 8 and 14 are present, the length of the tour is 10. However, if we change just the orientation of edge 14, the length reduces to 8. If we change the orientation of edges 10, 12, and 14, $E[L_{t_4}]$ drops from 20.08 to 19.81 for $p = 0.6$. Hence, another factor to take into consideration while developing tours is the orientation of the edges in a sub-tour.

We also tested other tours to understand the impact of the size of the sub-tours. Consider tour 5 with 2 inner sub-tours (Figure 3). One of the sub-tours has 6 edges (edges 2-7) while the other has only 2 edges (edges 10 and 11). $E[L_{t_5}]$ is less than $E[L_{t_4}]$ and greater than $E[L_{t_3}]$ for all values of p . Note that tours 5 and 3 have two sub-tours each, while tour 4 has only one sub-tour. Though tours 3 and 5 have the same number of sub-tours, the sub-tours of tour 5 are not balanced.

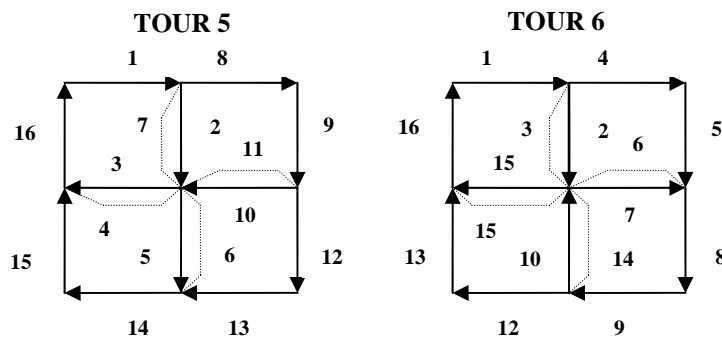


Figure 3. Tours 5 and 6 for a 3x3 undirected graph

Our example also illustrates that having a larger number of balanced sub-tours does not necessarily imply a lower expected length. Consider tour 6 in Figure 3, whose edges are oriented exactly the same way as in tour 3. But tour 6 consists of 4 sub-tours each with 2 edges, while tour 3 consists of 2

sub-tours each with 4 edges. The expected length of tour 3 is marginally less than that of tour 6 for all values of p . This simple example illustrates that the nature (i.e., number and size) and orientation of the sub-tours play an important role in determining the expected length of the tour. Specifically, the better tours in the expected sense, have more balanced sub-tours when compared to the worse tours. Also, the edges of the sub-tours should be oriented to minimize the average inter-edge traversal distances.

6. Conclusion

In this paper, we first defined the SETP as the problem seeking the Eulerian tour of minimum length in the expected sense for an undirected graph, when the number of white edges present follows a specified probability distribution. We then derived a closed form expression for calculating the expected length of a given tour in $O(n^2)$ time. We also showed that the SETP is NP-hard even though the deterministic ETP is solvable in polynomial time. We also derived a worst case ratio of the deviation of the expected length of a random Eulerian tour from the optimal tour. Finally, using an illustrative example, we investigated some of the desirable properties in an a priori tour.

We are currently working on developing and testing solution procedures for the SETP, that take advantage of some of the results presented in this paper. It is important to note that we assume that the least cost augmentation of the underlying graph is already done and define the SETP for an Eulerian graph. One of the main directions for our future research is to study the Stochastic Chinese Postman Problem, i.e., solve the augmentation problem when the edges of the given graph are present according to a specified probability distribution.

References

- Bertsimas, D. J. 1988. *Probabilistic combinatorial optimization problems*. Ph. D. Thesis, Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA.
- Bertsimas, D. J. and L. H. Howell. 1993. Further results on the probabilistic traveling salesman problem. *Eur. J. Oper. Res.* **65** 68-95.

- Edmonds, J. and E. L. Johnson. 1973. Matching, Euler tours and the Chinese postman problem. *Math. Prog.* **5** 88-124.
- Eiselt, H. A., M. Gendreau, and G. Laporte. 1995a. Arc routing problems. Part I: The Chinese postman problem. *Oper. Res.* **43** 231-242.
- Eiselt, H. A., M. Gendreau, and G. Laporte. 1995b. Arc routing problems. Part II: The rural postman problem. *Oper. Res.* **43** 399-414.
- Gendreau, M., G. Laporte, and R. Seguin. 1995. *Stochastic vehicle routing*. Publication CRT-95-02, Centre de Recherche sur les Transports, Université de Montréal, Canada.
- Jaillet, P. 1985. *Probabilistic traveling salesman problem*. Ph. D. Thesis, Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA.
- Jaillet, P. 1988. A priori solution of a traveling salesman problem in which a random subset of customers are visited. *Oper. Res.* **36** 929-936.
- Jaillet, P. and A. R. Odoni. 1988. The probabilistic vehicle routing problem. B.L. Golden and A.A. Assad, Eds. *Vehicle Routing: Methods and Studies*. North-Holland, Amsterdam.
- Lacomme, P., C. Prins, and W. Ramdane-Chérif. 2001. A genetic algorithm for the CARP and its extensions. E.J.W. Boers et al. Eds. *Applications of Evolutionary Computing. Lecture Notes in Computer Science 2037*. Springer, 473-483.
- Laporte, G. and F. V. Louveaux. 1993. The integer L-shaped method for stochastic integer programs with recourse. *Oper. Res. Letters* **13** 133-142.
- Laporte, G., F. V. Louveaux, and H. Mercure. 1989. Models and exact solutions for a class of stochastic location-routing problems. *Eur. J. Oper. Res.* **39** 71-78.
- Rossi, F. A. and I. Gavioli. 1988. Aspects of heuristic methods in the probabilistic traveling salesman problem. G. Andreatta, F. Mason and P. Serafini Eds. *Advanced School on Stochastics in Combinatorial Optimization*. World Scientific, Singapore.