BIT 25 (1985), 289–292

# NP-COMPLETENESS OF THE HAMMING SALESMAN PROBLEM

JARMO ERNVALL[1], JYRKI KATAJAINEN[2] and MARTTI PENTTONEN[2]

[1] *Department of Applied Mathematics, University of Jyväskylä, 40100 Jyväskylä, Finland*

[2] *Department of Computer Science, University of Turku, 20500 Turku, Finland*

**Abstract.**

It is shown that the traveling salesman problem, where cities are bit strings with Hamming distances, is NP-complete.

*Keywords:* NP-completeness, traveling salesman problem, Hamming distance.

## 1. Introduction.

The traveling salesman problem is one of the first and best known *NP*-complete problems. Originally it was stated for a general graph with arbitrary distances between cities [4]. Later it has turned out that the problem remains *NP*-complete even when the vertices of the graph are in a plane with Euclidean or rectilinear distances [3, 5]. In this paper we study the problem in the case of Hamming distances.

The Hamming salesman problem has an interest of its own. One could think that with Hamming distances the problem would be easier than with Euclidean or rectilinear distances, because apparently the distances between the cities are perhaps even more closely bound to each others. So one could hope that with Hamming distances a polynomial time algorithm for finding an optimal tour could be found. However, we shall see that the problem remains *NP*-complete also in this case.

We came to the Hamming salesman problem from a data compression problem; some related problems can be found in [1, 2]. When certain data compression techniques are used, it is important that records in a list are ordered in such a way that successive records differ from each other in as few fields as possible. Hence, a fast algorithm for finding an optimal salesman path would be useful. Unfortunately, our result frustrates this effort.

Received October 1984.

## 2. Definitions and result.

In metric spaces all graphs can be considered weighted and complete; for each pair of vertices there is a connecting edge whose weight is determined by the metric. Hence, edges need not be listed.

DEFINITION. An *instance of the Hamming salesman problem* ($HTSP$) is a string $P = v_1\#\ldots\#v_n\#L$, where $v_i \in \{0, 1\}^k$, for some $n$ and $k$, and $L$ is an integer in binary representation. The *size* of $P$ is its length, i.e. $n(k+1)+|L|$, where $|L|$ is the number of bits in $L$. The *distance* between two vertices $v_i = v_{i1} \ldots v_{ik}$, $v_j = v_{j1} \ldots v_{jk}$ $(i, j = 1, \ldots, n)$ is defined by

$$d(v_i, v_j) = \text{card}\,\{t | v_{it} \neq v_{jt}\}.$$

A *solution* of $P$ is a permutation $(i_i, \ldots, i_n)$ of $(1, \ldots, n)$ such that

$$\sum_{j=1}^{n-1} d(v_{i_j}, v_{i_{j+1}}) + d(v_{i_n}, v_{i_1}) \leqslant L.$$

An *instance of an optimizing Hamming salesman problem* ($OHTSP$) is a string $P = v_1 \ldots v_n$ where $v_i$'s are as above. A solution of $P$ is a permutation $(i_i, \ldots, i_n)$ of $(1, \ldots, n)$ such that the sum

$$\sum_{j=1}^{n-1} d(v_{i_j}, v_{i_{j+1}}) + d(v_{i_n}, v_{i_1})$$

is a minimum over all permutations of $(1, \ldots, n)$.

The definition of the distance can be extended to binary strings of infinite length. Note that infinite strings with this distance form a metric space.

The definition of the Hamming salesman problem could be generalized by allowing an arbitrary set $S$ instead of $\{0, 1\}$ without any change in the definition of the distance. The only requirement on $S$ is that the equality of its elements can be tested in polynomial time with respect to their lengths. If, for example, $S$ is the set of integers, each element of $S$ needs a string representation and the comparison of two elements needs several elementary steps. However, it is obvious that this generalized problem is polynomially time equivalent with the original problem.

As in the case of the graph traveling salesman problem [4], it is easy to verify that also in the Hamming case the upper bound and the optimizing problem are equivalent.

LEMMA 1. *HTSP and OHTSP are transformable to each others in polynomial time.*

PROOF. If the *OHTSP* algorithm is known, then *HTSP* can be solved by first computing the optimal tour and then comparing its length with the length of the *HTSP*.

On the other hand, *HTSP* can be used for finding the minimal tour of the *OHTSP* as follows. Clearly $M = nk$ is an upper bound for the length of the optimal tour. Now the shortest tour can be found by solving *HTSP*'s for $M$, $M-1$, $M-2$ and so forth until such an instance is met for which no solution is found. The total running time will be polynomial, because at most $M$ rounds are needed and $M$ is smaller than the size of the instance of the problem.

LEMMA 2. *HTSP and OHTSP are in NP.*

PROOF. By lemma 1, it suffices to prove that *HTSP* is in *NP*. This is very easy, and only a nondeterministic guess and testing is needed.

We show the *NP*-hardness of the Hamming salesman problem by reducing the rectilinear traveling salesman problem to it. We shall now briefly define the mentioned problem.

DEFINITION. An *instance of the rectilinear traveling salesman problem* (*RTSP*) is a string $P = x_1\#y_1\#\ldots\#x_n\#y_n\#L$, where $x_i$'s and $y_i$'s are nonnegative integers in binary representation. As above, the *size* of $P$ is its length. The *distance* by rectilinear metric is defined as follows:

$$d(x_i\#y_i, x_j\#y_j) = |x_i - x_j| + |y_i - y_j|.$$

Again, the *solution* is a tour whose length is not greater than $L$.

We shall need the following result:

LEMMA 3 [3, 5]. *RTSP is NP-complete. This holds true even for the problems where the numerical values of each $x_i$ and $y_i$ is bounded by cn, where c is a small constant.*

REMARK. A rectilinear modification of Theorem 2 in [5] itself is not sufficient for us. We need the boundedness of the coordinates, which is an immediate corollary of the proof of the theorem.

We can now state our result:

THEOREM. *The Hamming salesman problem and the optimizing Hamming salesman problem are NP-complete.*

PROOF. By lemmas 1 and 2, it suffices to prove that $HTSP$ is $NP$-hard. We shall transform any instance $P = x_1\#y_1\#\ldots x_n\#y_n\#L$ of the $RTSP$ to an instance of the $HTSP$ in the following way. Let $X = \max\{x_1,\ldots,x_n\}$ and $Y = \max\{y_1,\ldots,y_n\}$. Each city $x_i\#y_i$ $(i = 1,\ldots,n)$ is now transformed to $1^{x_i}0^{X-x_i}1^{y_i}0^{Y-y_i}$. Obviously, this transformation preserves the distances of the cities. The size of the problem grows but not too much. According to lemma 3, the size of $P$ is $O(n \log n)$, while the size of the new problem is $O(n^2)$. Hence, a polynomial time algorithm for $HTSP$ would yield a polynomial time algorithm for $RTSP$. Consequently, $HTSP$ is $NP$-hard.

## 3. Conclusions.

It was proved that the traveling salesman problem with Hamming metric is $NP$-complete. It can be easily seen that also the respective path problems are $NP$-complete. Our proof was strongly based on the proof of Papadimitriou [5] for a traveling salesman in $N \times N$ with rectilinear metric. It would be interesting to find general conditions for the metric guaranteeing the $NP$-completeness of the problem.

## REFERENCES

1. J. Ernvall, *On the construction of spanning paths by Gray-code in compression of files*, RAIRO Informatique, to appear.
2. J. Ernvall and O. Nevalainen, *Compact storage schemes for formatted files by spanning trees*, BIT 19 (1979), 464–475.
3. M. R. Garey, R. L. Graham and D. S. Johnson, *Some NP-complete problems*, Proc. 8th ACM Symposium on Theory of Computing (1976), 10–29.
4. M. R. Garey and D. S. Johnson, *Computers and Intractability*, Freeman, 1979.
5. C. H. Papadimitriou, *The Euclidean traveling salesman problem is NP-complete*, Theoretical Computer Science 4 (1977), 237–244.