Send to IEEE INF THEORY 1994

Traveling Salesman Problem in the Space of Binary Vectors

Lev B. Levitin, Mark G. Karpovsky
Department of Electrical, Computer and Systems Engineering
Boston University,USA

The traveling salesman problem in the space Z_2^n of all n-dimensional binary vectors can be formulated as follows.

Given: A set $C(n, M)$ of M n-dimensional vectors over GF(2) (M points in Z_2^n).

Find: A Hamiltonian circuit for $C(n, M)$ of minimum Hamming length, i.e., a cycle with minimum sum of Hamming distances between connected points (henceforth called minimal cycle).

This problem is encountered in various situations related to the design and testing of computer hardware. In contrast with "unrestricted" traveling salesman problem, the length of the minimal cycle in Z_2^n is always upperbounded by that for the "worst" possible configuration of M points. Hence we come to the following variational problem.

Problem: For any number M of points in Z_2^n find the function $L(n, M)$ defined as follows:

$$L(n, M) = \max_{\{C(n,M)\}} \min_{P(C(n,M))} \sum_{i=1}^{M} d(x_i), x_{i+1(\mathrm{mod} M)}), \qquad (1)$$

where $x_1, \ldots, x_M$ are the points comprising a binary code $C(n, M)$, $d(x, y)$ is the Hamming distance between points x and y, $P(C(n, M))$ is the set of all possible permutations of the codewords, and the maximum is taken over the set of all possible binary codes of size M and dimension n.

The problem is still open, but partial results are presented below.

Lemma 1. For given n, $L(n, M)$ is a monotonically non-decreasing function of M.

Theorem 1. For any M, such that $2^{n-1} \leq M \leq 2^n$,

$$L(n, M) = 2^n. \quad (2)$$

Theorem 2. (Lower bound) For any n and M,

$$L(n, M) \geq d(n, M) \cdot M, \quad (3)$$

where $d(n, M)$ is the distance of the best code of size M, i.e., of the code of length n with the largest minimum distance.

Conjecture: The best codes are the worst configurations for any given n and M.

The conjecture is known to be true for $M \geq 2^{n-1}$ (see Theorem 1) and for some other cases considered below. However, it is not a trivial question, whether (3) is attainable even in the cases when the best codes are the worst configuration. Indeed, it is easily seen that, in general, it is impossible to build a Hamiltonian circuit for a code using only edges of length equal to the minimum distance. However, as we will show in the next theorem, this is always possible for a cyclic code.

Theorem 3. The length $l(C)$ of a minimal cycle for any cyclic code $C = C(n, 2^k, d)$ with distance d is equal:

$$l(C) = d \cdot 2^k. \quad (4)$$

Theorem 4. (Upper bound).

$$L(n, M) \leq \begin{cases} \frac{n}{2} \cdot \frac{M^2}{M-1}, & \text{for even } M; \\ \frac{n}{2} \cdot (M+1), & \text{for odd } M. \end{cases} \quad (5)$$

The upper bound (5) is attainable for small M ($M \leq n+1$), when the worst configurations are equidistant codes, and all Hamiltonian circuits are minimal cycles.

A family of such codes can be obtained by deleting the zero column in a binary (2^k x 2^k) Hadamard matrix and concatenating the obtained code with itself m times. These equidistant codes have the following parameters: $n = m(2^k - 1)$, $M = 2^k$, $d = m \cdot 2^{k-1}$, where $m = 1, 2, \ldots$. Then $L(n, M) = m \cdot 2^{2k-1} = dM = \frac{M^2}{M-1} \cdot \frac{n}{2}$, thereby satisfying lower bounds (3) and upper bound (5). Another family of equidistant (generally nonlinear) codes can be described as follows. Let each bit be covered by a different list of $M/2$ codewords. Then $n = \binom{M-1}{M/2}, d = \binom{M-2}{M/2-1}$. By concatenating such a code with itself we obtain codes with parameters: $n = m\binom{M-2}{M/2}, d = m\binom{M-2}{M/2-1}$. Codes in both families have even numbers of codewords. Codes with odd M that satisfy both (3) and (5) can be obtained from those codes by deleting one codeword.

For large M ($logM = \Theta(n)$) the gap between (3) and (5) becomes too large, which calls for better upper bounds.

Theorem 5. (Improved upper bound). For any $M \geq 4, n \geq 3$,

$$L(n,M) \leq \frac{8}{3}n + 2\sum_{m=5}^{M} d(n,m). \quad (6)$$

Using the Hamming upper bound $d_H(n,M)$ for the minimum distance

$$d_H(n,M) = 2nh^{-1}(1 - \frac{log_2 M}{n}),$$

where h^{-1} is the inverse to $h(x) = -xlog_2 x - (1-x)log_2(1-x)$, performing integration, instead of summation, in (19) and taking into account the convexity of $d_H(n,M)$ for approximation, one can obtain:

$$L(n,M) \leq 2d_H(n,M)(M-1) + \frac{nM}{lnM} - n. \quad (7)$$

For codes with rates smaller than 1 ($R = \frac{log_2 M}{n} < 1$), the ratio of the second term to the first one is of the order $O(n^{-1})$.

For the lower bound (3), one can use the Varshamov-Gilbert lower bound on the minimum distance $d_{VG}(n,M) = \frac{1}{2}d_H(n,M)$. Thus, the lower and upper bounds differ in a factor of 4.