

On the Addressing Problem for Directed Graphs

F.R.K. Chung¹, R.L. Graham² and P.M. Winkler^{3*}

¹ Bell Communications Research, Morristown, NJ 07960, USA

² AT&T Bell Laboratories, Murray Hill, NJ 07974, USA

³ Emory University, Atlanta, GA 30322, USA

Abstract. The “distance” from vertex u to vertex v in a strongly connected digraph is the number of arcs in a shortest directed path from u to v . The addressing problem, first formulated in the undirected case by Graham and Pollak, entails the assignment of a string of symbols to each vertex in such a way that the distances between vertices are equal to modified Hamming distances between corresponding strings.

A scheme for addressing digraphs is proposed, and the minimum address length is studied both in general and in certain special cases. The problem has some interesting reformulations in terms of matrix factorization and extremal set theory.

1. Introduction

Packet-type communications networks, proposed by J.R. Pierce [5] in the early 70’s and now in operational use, permit computers to send information (in the form of strings of bits, called “packets”) to other computers without waiting for clear lines. In such a system each interchange point, or vertex, has an *address* which can be appended to a packet in order to identify its destination.

Graham and Pollak [2, 3] proposed assigning addresses in such a way that the distance (number of edges in a shortest path) between two vertices can quickly be determined by comparing their addresses. Then, when a packet reaches a vertex u , en route to a destination v , it “knows” the distance $d(u, v)$ to its destination. It then proceeds to the first vertex w that it finds, among those adjacent to u , for which $d(w, v) = d(u, v) - 1$; thus it is guaranteed a minimal-length path to its destination.

In the original formulation all edges of the network are assumed to be capable of carrying packets in either direction, and addresses are strings $a = (a_1, a_2, \dots, a_t)$ of 0’s, 1’s and *’s with distance defined as follows:

$$d(a, b) = |\{k: \{a_k, b_k\} = \{0, 1\}\}|.$$

Graham and Pollak conjectured that address length $t = n - 1$ is sufficient (it is often necessary) for any connected graph on n vertices, and this indeed proved to be the case [7]. Thus the scheme above turned out to be a practical solution of the addressing problem for Pierce’s network.

We now consider the situation where some lines in the network may be capable

* Supported by NSF grant MCS 84-02054.

of carrying packets in only one direction. Then the model for the network becomes a directed graph (digraph) and distance is measured by number of arcs in a shortest directed path. The distance between addresses must be modified to permit asymmetry; we will do this by counting only (0, 1) pairs in one direction, and (1, 0) in the other. The resulting scheme turns out to be about as good as any of its type in terms of address length, but much worse in some cases than the scheme for undirected graphs.

2. Notation and Preliminaries

Let G be a digraph with vertices v_1, v_2, \dots, v_n ; G will always be assumed to be *strongly connected*, so that for any pair of vertices v_i, v_j there is a directed path from v_i to v_j . The number of arcs in a shortest such path will be called the *distance* from v_i to v_j , and denoted by $d(v_i, v_j)$ or simply d_{ij} . Of course d_{ij} satisfies the triangle inequality and is zero only when $i = j$, but is not symmetric in general. We let $D = \{d_{ij}\}$ be the $n \times n$ distance matrix of G .

If $M = \{m_{ij}\}$ is any $n \times n$ matrix then an *addressing* of M of length t is an $n \times t$ matrix A with entries from $\{0, 1, *\}$ such that

$$m_{ij} = |\{k: a_{ik} = 0 \text{ and } a_{jk} = 1, 1 \leq k \leq t\}|$$

for all i, j between 1 and n . An addressing of the digraph G is just an addressing of its distance matrix D .

Theorem 1. *Let $M = \{m_{ij}\}$ be an $n \times n$ matrix with non-negative integer entries and zero diagonal. If M_i is the largest entry in row i then M has an addressing of length $\sum M_i$.*

Proof. We split A into n blocks of M_i columns each, devoting each block to the i^{th} row of M . For each $k, 1 \leq k \leq \sum M_i$, write

$$k = \sum_{i=1}^{r(k)-1} M_i + s(k)$$

where $1 \leq r(k) \leq n$ and $1 \leq s(k) \leq M_{r(k)}$. Then set

$$a_{ik} = \begin{cases} 0 & \text{if } i = r(k); \\ 1 & \text{if } i \neq r(k) \text{ and } s(k) \leq m_{r(k), i}; \\ * & \text{otherwise.} \end{cases}$$

□

Corollary. *Any digraph G on n vertices can be addressed in length $n(n - 1)$.*

Proof. No distance in G can be greater than $n - 1$. The unique extreme case is the directed cycle, which, as we shall see later, does have a more efficient addressing than this one.

Our $\{0, 1, *\}$ addressing scheme can be generalized as follows. Let Q be an $s \times s$ matrix of non-negative integers and define a Q -addressing of an $n \times n$ matrix M to be a matrix $A = \{a_{ik}\}$ such that

$$m_{ij} = \sum_k q_{a_{ik}, a_{jk}}$$

The 0, 1, * scheme then amounts to Q -addressings for

$$Q = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

which is about as good as any in terms of address length:

Theorem 2. For any Q there is a constant c such that for any Q -addressing of a matrix M of length t , there is a $\{0, 1, *\}$ addressing of M of length ct .

Proof. We may assume Q has zero diagonal, since $q_{ii} \neq 0$ implies that any appearance of i in A would foul the zero diagonal of M . Hence there is a $\{0, 1, *\}$ addressing B of Q itself, say of length c . If A is a Q -addressing of M of length t then each integer a_{ik} can be replaced by the string

$$b_{a_{ik},1} b_{a_{ik},2} \cdots b_{a_{ik},c}$$

and the result is a $\{0, 1, *\}$ addressing of M of length ct . □

It may seem strange that the symbols $0, 1, *$ themselves do not satisfy the triangle inequality, since digraphs do. An addressing of a digraph G is equivalent to an isometric embedding in the Cartesian product of t copies of $\{0, 1, *\}$. The next theorem justifies this shortcoming.

Theorem 3. Let Q be an $n \times n$ matrix of non-negative integers which satisfies the triangle inequality, i.e. $q_{ik} \leq q_{ij} + q_{jk}$. Then there is a digraph (in fact, a graph) G such that G has no Q -addressing of any length.

Proof. Suppose not and let A be a Q -addressing of the bipartite graph $K_{2,n+1}$, where v_1 and v_2 are the vertices in the part of size 2. For each $i, j > 2$ we have $d_{ij} = d_{i1} + d_{1j}$, hence $q_{a_{ik},a_{jk}} = q_{a_{ik},a_{1k}} + q_{a_{1k},a_{jk}}$ for every k between 1 and the address length; the same holds for v_1 replaced by v_2 . Now for each k there must be a pair $i, j > 2$ such that $a_{ik} = a_{jk}$, since the entries in A range over only n values; then $q_{a_{ik},a_{jk}} = 0$ since $d_{ii} = 0$. Thus

$$q_{a_{1k},a_{jk}} = q_{a_{1k},a_{ik}} = 0 = q_{a_{ik},a_{2k}}$$

and therefore $q_{a_{1k},a_{2k}} = 0$; since this is true for every k we have $d_{12} = 0$, a contradiction. □

Corollary. There is no finite metric space X with the property that for every graph G there is a Cartesian power of X into which G is isometrically embeddable.

3. The Complete Digraph

The complete graph K_n was an interesting special case in [3], as an (undirected) addressing of K_n amounts to a partition of its edges into complete bipartite subgraphs. The necessity of address length $n - 1$ (and therefore of at least $n - 1$ subgraphs in the partition) has been shown in several ways, all involving linear algebra.

Let K_n^* be the complete digraph on n vertices, i.e. for every $i \neq j$ there is an arc from v_i to v_j . Theorem 1 addresses K_n^* by the $\{0, 1\}$ -complement of the identity

matrix; any permutation matrix or its complement will do. There are also length- n addressings with stars, as the following example (suggested by R. Roth) shows in the $n = 5$ case:

$$\begin{pmatrix} * & 0 & 0 & 1 & 1 \\ 0 & * & 1 & 0 & 1 \\ 0 & 1 & * & 1 & 0 \\ 1 & 0 & 1 & * & 0 \\ 1 & 1 & 0 & 0 & * \end{pmatrix}$$

Any addressing A amounts to a partition of the arcs of K_n^* into complete bipartite graphs all of whose edges are directed from the first part to the second. If A is regarded as undirected it becomes an addressing of the matrix

$$\begin{pmatrix} 0 & 2 & 2 & \cdots & 2 & 2 \\ 2 & 0 & 2 & \cdots & 2 & 2 \\ 2 & 2 & 0 & \cdots & 2 & 2 \\ & & & \vdots & & \\ 2 & 2 & 2 & \cdots & 2 & 0 \end{pmatrix}$$

which has $n - 1$ negative eigenvalues, and the argument in [3] shows that A must have length at least $n - 1$. Is length $n - 1$ possible? Again, linear algebra gives the answer.

Theorem 4. *An $n \times n$ matrix M with non-negative integer entries and zero diagonal has an addressing of length t if and only if there are $n \times t$ and $t \times n$ $\{0, 1\}$ -matrices B and C such that $M = BC$.*

Proof. If A is an addressing of M of length t , let $b_{ik} = 1$ if $a_{ik} = 0$, and $b_{ik} = 0$ otherwise; let $c_{ki} = 1$ if $a_{ik} = 1$, and $c_{ki} = 0$ otherwise. Then

$$m_{ij} = \sum_{k=1}^t b_{ik}c_{kj} = |\{k: a_{ik} = 0 \text{ and } a_{kj} = 1\}|$$

as desired.

For the converse note that we cannot have $b_{ik} = c_{ki} = 1$ since otherwise $m_{ii} \neq 0$. Hence we can define

$$a_{ik} = \begin{cases} 0 & \text{if } b_{ik} = 1 \\ 1 & \text{if } c_{ki} = 1 \\ * & \text{otherwise.} \end{cases} \quad \square$$

Corollary. K_n^* has no addressing of length less than n .

Proof. Since the distance matrix of K_n^* is non-singular, it cannot be factored into matrices of rank less than n . □

The factorization argument turns out not to be useful in the best-case analysis (below), since there are $n \times n$ matrices M with zero diagonal and positive integers everywhere else such that $\text{rank}(M)$ is only 3. Instead we use combinatorial analysis to show that any digraph on n vertices requires addressing length more than $\log_2 n$.

4. The Best Case

Instead of asking which digraph on n vertices has the shortest addressing, it is more convenient to look for the largest digraph which can be addressed in length t . In the undirected case [3] the answer is easily seen to be the hypercube of dimension t , which has 2^t vertices; the directed case is only slightly more subtle.

Let H_t be the (symmetric) digraph whose vertices u_1, \dots, u_m are the subsets of $\{1, 2, \dots, t\}$ of size $\lfloor t/2 \rfloor$, with u_i and u_j connected by arcs in both directions just when $|u_i \cap u_j| = \lfloor t/2 \rfloor - 1$.

Theorem 5. H_t has an addressing of length t and is the unique largest such digraph, i.e. if G has at least $\binom{t}{\lfloor t/2 \rfloor}$ vertices and is addressable in length at most t , then G is isomorphic to H_t .

Proof. The addressing of H_t is straightforward: set a_{ik} equal to 0 if $k \in u_i$, and 1 otherwise.

On the other hand, let $B = \{b_{ik}\}$ be an addressing of the digraph G on $n \geq \binom{t}{\lfloor t/2 \rfloor}$ vertices, and set $S_i = \{k: b_{ik} = 1\}$. We cannot have $S_i \supset S_j$ for $i \neq j$, else d_{ij} would be zero; hence the S_i 's form an antichain in the Boolean lattice of subsets of $\{1, \dots, t\}$. The strong form of Sperner's Theorem [6] says that when t is even, the unique largest such antichain is the set T of subsets of size $t/2$; thus $n = \binom{t}{t/2}$, $\{S_i\} = T$ and no stars may appear in B without collapsing a distance to zero. The map $v_i \mapsto S_i$ thus induces an isomorphism from G to H_t .

If t is odd there are two largest such antichains, the sets of size $\lfloor t/2 \rfloor$ and the sets of size $\lfloor t/2 \rfloor + 1$; in the latter case the isomorphism from G to H_t takes v_i to the complement of S_i . \square

5. The Case of the Directed Cycle

The directed cycle C_n has vertices v_1, \dots, v_n and arcs $(v_1, v_2), (v_2, v_3), \dots, (v_{n-1}, v_n), (v_n, v_1)$. As a strongly connected digraph on n vertices it has the fewest arcs and the greatest distances, and is the extremal case for the addressing given in Theorem 1. The addressing problem for C_n may be formulated as an abstract problem in extremal set theory as follows: Let $\{(A_i, B_i)\}$, $1 \leq i \leq n$, be a family of pairs of sets such that for all i and j , $|A_i \cap B_j|$ is less than n and equal to $j - i$ modulo n ; now minimize the cardinality of the union of the A_i 's and B_i 's.

Theorem 6. There is a constant $c > 0$ such that C_n has no addressing of length less than $cn^{3/2}$.

Proof. Let A be an addressing of C_n of length t , and let U_k be the k^{th} column of A . U_k is said to reduce the pair (i, j) if $(a_{ik}, a_{jk}) = (0, 1)$; for each i , $1 \leq i \leq n$, let $S_i = \{k: U_k \text{ reduces } (i + 1, i)\}$. Since $d_{i+1, i} = n - 1$, each S_i has cardinality $n - 1$; but $|S_i \cap S_{i+j}| \leq j - 1$ because if $k \in S_i \cap S_{i+j}$ then U_k also reduces $(i + 1, i + j)$ and $d_{i+1, i+j}$

is only $j - 1$. Fix $m = \lfloor (2n)^{1/2} \rfloor$ and note:

$$\begin{aligned} t &\geq |S_1 \cup S_2 \cup \dots \cup S_m| \\ &\geq \sum_{i \leq m} |S_i| - \sum_{i < j \leq m} |S_i \cap S_j| \\ &\geq m(n - 1) - \sum_{i=1}^{m-1} i(m - 1 - i) \\ &= m(n - 1) - (1/6)m^3 + m^2 - (5/6)m - 1 \end{aligned}$$

which is asymptotic to $(8/9)^{1/2}n^{3/2}$. □

It is in fact possible to take care of all the $(n - 1)$'s in the distance matrix in order $n^{3/2}$, using a simplification of the first procedure used in the proof of Theorem 7 below. Somewhat greater address length seems to be required to take care of the whole matrix, but not as much as $n(n - 1)$.

Theorem 7. *There is a constant c such that C_n can be addressed in length $cn^{5/3}(\ln n)^{1/3}$.*

Proof. The required addressing of C_n is obtained by a carefully tuned amalgamation of three procedures, details of which will be selectively omitted.

For convenience let $q = \lfloor n^{1/3}(\ln n)^{-1/3} \rfloor$, $r = \lfloor n^{1/2}(\ln n)^{1/2} \rfloor$ and $s = \lfloor n^{1/6}(\ln n)^{-1/6} \rfloor$.

The first procedure reduces all entries in the distance matrix D of size at least $n - q$, while doing minimal damage to the rest of the entries. Let U_{jk} be the column (of the addressing matrix A , potentially) whose i^{th} entry u_i is given by

$$u_i = \begin{cases} 0 & \text{if } i \equiv k \pmod{j} \\ 1 & \text{if } i \in \{k - 1, k - 2, \dots, k - q\} \pmod{j} \\ * & \text{otherwise.} \end{cases}$$

Let V_j be the block $\{U_{jk} : k = 0, 1, \dots, j - 1\}$. Then V_j reduces every D -entry of size $\geq n - q$ exactly once, and every entry of size $h < n - q$ is reduced once for each multiple of j between $h + 1$ and $h + q$. Let W_j be the union of r copies of V_j .

Now let P be the set of all primes p between r and $2r$, so that $|P|$ is asymptotically $n^{1/2}(\ln n)^{-1/2}$ by the Prime Number Theorem [4]. Let $J = \{sp : p \in P\}$ and let W be the union of $\{W_j : j \in J\}$. Then W reduces each large entry $r \cdot |J| \sim n$ times, i.e. completely; and the number of columns in W is about $r \cdot |J| \cdot rs \sim n^{5/3}(\ln n)^{1/3}$. Further, if $r < h < n - q$ then only $q/s \sim s$ numbers between $h + 1$ and $h + q$ can be multiples of s and none can have more than one factor in P , so no entry of size h can be reduced by more than $rs \sim n^{2/3}(\ln n)^{1/3}$; entries of size less than r will not be reduced at all.

The entries of size less than $n - q$ are reduced by a second procedure. Let X_{jk} be the column whose i^{th} entry x_i is given by

$$x_i = \begin{cases} 1 & \text{if } i + k \in \{1, 2, \dots, j\} \pmod{n} \\ 0 & \text{if } i + k \in \{j + 1, j + 2, \dots, 2j\} \pmod{n} \\ * & \text{otherwise.} \end{cases}$$

Let Y_j be the block $\{X_{jk} : 1 \leq k \leq n\}$, for $j \leq n/2$, and let Z_j consist of $n/2j$ copies of Y_j . If $f_j(h)$ is the amount by which Z_j reduces each D -entry of size h , then $f_j(n - 1) = 1 = f_j(n - 2j + 1)$ and $f_j(n - 2j) = n/2$; f_j is linear in the intervals between these

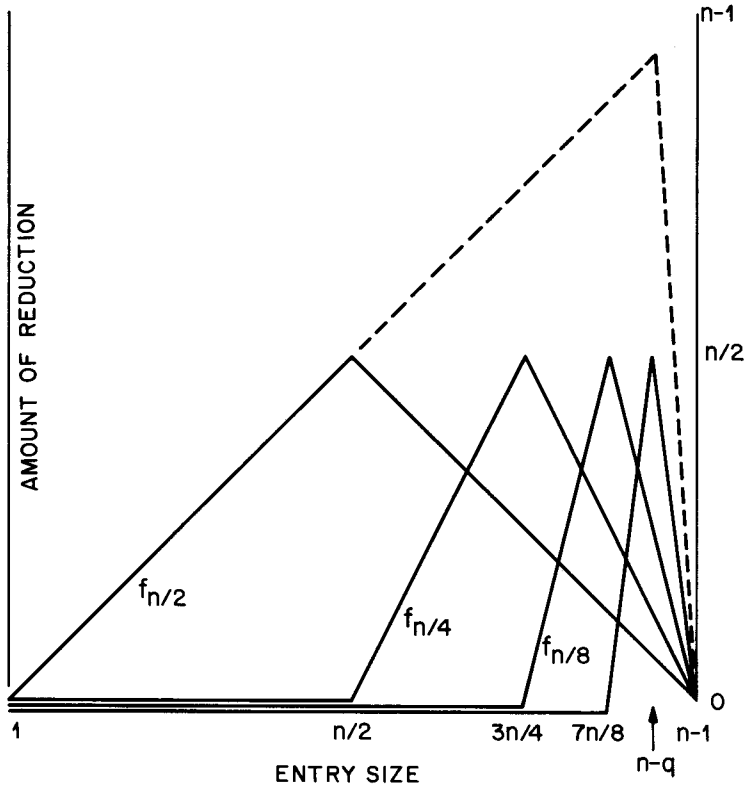


Fig. 1

values, and $f_j(h) = 0$ for $h < n - 2j + 1$. Finally, let

$$Z = \bigcup \{Z_{(n/2^i)}; i = 1, 2, \dots, [\log_2(n/q)]\}.$$

The effect of Z is indicated by the dotted line in Fig. 1; essentially, it wipes out every D -entry below $n - q$ and helps out somewhat with the large entries as well. The number of columns in Z is $n \sum (n/(2n/2^i)) = n \sum 2^{i-1} = n(2^{[\log_2(n/q)]} - 1) \sim n^2/q \sim n^{5/3}(\ln n)^{1/3}$.

Now the two large blocks W and Z defined above cannot immediately be combined, as they would overkill some entries of the distance matrix. Instead, W is modified to allow for the assistance given by Z on the large entries; and Z is shrunk slightly to allow for the incidental reduction of small entries by W . When finally combined, W and Z reduce every entry of D to a number between 0 and $n^{2/3}(\ln n)^{1/3}$.

Finally an application of Theorem 1 completes the addressing with one more block of $n^{5/3}(\ln n)^{1/3}$ columns. □

6. The Worst Case

Several heuristic algorithms for addressing digraphs have been devised by Dent [1] and they appear to work fairly well (giving addressings of length not much more than linear in n) on random digraphs dense enough to be strongly connected.

However, the bottom line on digraph addressing is that it does take lengths of order n^2 in the worst case.

Theorem 8. *There is a constant c such that for any n , there is a digraph G on n vertices which requires addressing length at least cn^2 .*

Proof. Let G have vertices $u_1, \dots, u_r; v_1, \dots, v_r$; and w_1, \dots, w_s . The arcs of G consist of $(w_s, v_1); (v_1, u_1); (v_r, u_r); (u_r, w_1)$; all (w_i, w_{i+1}) for $i = 1, 2, \dots, s - 1$; and all (u_i, v_j) for $1 \leq i < j \leq r$. (See Fig. 2.)

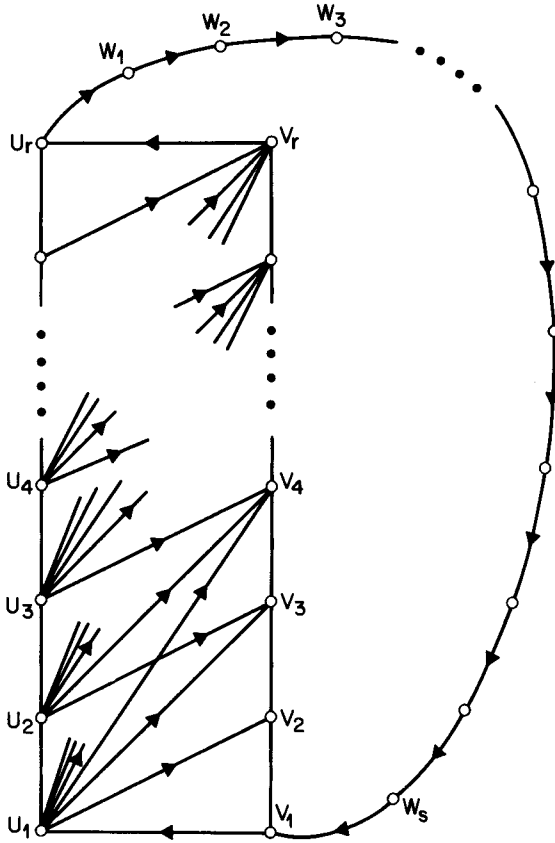


Fig. 2

We then have that $d(u_i, v_j) = 1$ for $i < j$ but $d(u_i, v_i) = s + 5$ for $1 < i < r$; at the ends, $d(u_1, v_1) = s + 3 = d(u_r, v_r)$. If A is an addressing of G of length t and U_k is the k^{th} column of A , let

$$S_i = \{k: U_k \text{ reduces the entry } d(u_i, v_i)\}.$$

If U_k reduces both $d(u_i, v_i)$ and $d(u_j, v_j)$ for $i < j$, then it must also reduce $d(u_i, v_j)$; hence $|S_i \cap S_j| \leq 1$ for $i \neq j$. We then have

$$\begin{aligned}
 t &\geq |S_1 \cup S_2 \cup \dots \cup S_r| \\
 &\geq r(s + 5) - 4 - r(r - 1)/2 \\
 &= \frac{1}{10}n^2 + \frac{11}{10}n - 4 \quad \text{for } r = \frac{1}{5}n, \quad s = \frac{3}{5}n, \quad \square
 \end{aligned}$$

Noga Alon has found a more complex construction which requires addressing length asymptotic to $n^2/8$. In the other direction, Alon has shown that any digraph G can be addressed in length at most $(7n^2/8) + O(n)$. We sketch the proof below.

We first construct a new undirected graph \bar{G} from G as follows:

$$\begin{aligned}
 V(\bar{G}) &= V(G), \\
 E(\bar{G}) &= \{\{a, b\}: d_G(a, b) > \frac{3}{4}n \text{ or } d_G(b, a) > \frac{3}{4}n\}.
 \end{aligned}$$

Since G is (strongly) connected, then the maximum degree $\bar{\Delta}$ of \bar{G} is at most $(n/2)$. Thus, the chromatic number $\bar{\chi}$ of \bar{G} is at most $\bar{\Delta} + 1 \leq (n/2) + 1$. Let $V(\bar{G}) = V_1 \cup \dots \cup V_{\bar{\chi}}$ be a partition of the vertex set of \bar{G} into $\bar{\chi}$ independent sets. Note that if $x \in V(\bar{G})$ and $a, b \in V_i$ for some i with

$$d_G(x, a) = \frac{3}{4}n + t$$

for some $t > 0$ then

$$d_G(x, b) \geq t.$$

This follows from the triangle inequality for d_G and the fact that by the definition of V_i ,

$$d_G(a, b) \leq \frac{3}{4}n, \quad d_G(b, a) \leq \frac{3}{4}n.$$

Also note that we always have $t < n/4$. Thus, we can reserve a set of $n/4$ coordinates in the addresses to take care of all distances from some $x \in V(\bar{G})$ to some $a \in V_i$ which exceed $3n/4$. We do this by putting all 1's in these positions for each $u \in V_i$ and exactly $t(x)$ 0's and $n/4 - t(x)$ *'s in these positions for each $x \in V(\bar{G}) \setminus V_i$ where

$$t(x) = \max_{u \in V_i} \{d_G(x, u) - \frac{3}{4}n\}$$

provided this is positive (otherwise define $t(x)$ to be 0).

Doing this for all $V_i, 1 \leq i \leq \bar{\chi}$, requires at most $(n/4)(1 + (n/2))$ coordinate positions, and has the effect of reducing all distances to be at most $3n/4$. The remaining distances can now be taken care of in the usual (trivial) way, using at most $3n^2/4$ more coordinates. Thus, G has been addressed with length at most $(7n^2/8) + O(n)$ as claimed.

We point out that by applying an iterated form of this argument, it is possible to reduce the bound to $(3n^2/4) + o(n^2)$ (we omit the details). At present, we have no clear idea as to what true constant should be.

Acknowledgement. The authors have had valuable discussions on the preceding topic with Noga Alon, Eric Dent, and Robert Roth.

References

1. Dent, E.: Digraph Addressing. Master's Thesis in Mathematics and Computer Science, Emory U., May 1983

2. Graham, R.L., Pollak, H.O.: On the addressing problem for loop switching. *Bell System Tech. J.* **50**, 2495–2519 (1971)
3. Graham, R.L., Pollak, H.O.: On embedding graphs in squashed cubes. *Graph Theory and Applications (Lecture Notes in Mathematics 303)*. Berlin-Heidelberg-New York: Springer 1979
4. Hardy, G.H., Wright, E.M.: *Introduction to the Theory of Numbers* (5th ed.). Oxford: Oxford University Press 1979
5. Pierce, J.R.: Network for block switching of data. *Bell System Tech. J.* **51**, 1133–1145 (1972)
6. Sperner, E.: Ein Satz über Untermengen einer endlichen Menge. *Math. Z.* **27**, 544–548 (1928)
7. Winkler, P.M.: Proof of the squashed cube conjecture. *Combinatorica* **3**, 135–139 (1983)