Sequences of radius k: how to fetch many huge objects into small memory for pairwise computations.

Jerzy W. Jaromczyk[1**] and Zbigniew Lonc[2]

[1] University of Kentucky, Lexington, KY, USA. <jurek@cs.uky.edu>,
[2] Warsaw University of Technology, Warsaw, Poland. <zblonc@mini.pw.edu.pl>

Abstract. Let $a_1, a_2, \ldots, a_m$ be a sequence over $[n] = \{1, \ldots n\}$. We say that a sequence $a_1, a_2, \ldots a_m$ has the *k-radius* property if every pair of different elements in $[n]$ occurs at least once within distance at most k; the *distance* $\mathrm{d}(a_i, a_j) = |i - j|$. We demonstrate lower and (asymptotically) matching upper bounds for sequences with the *k-radius* property. Such sequences are applicable, for example, in computations of two-argument functions for all $\binom{n}{2}$ pairs of large objects such as medical images, bitmaps or matrices, when processing occurs in a memory of size capable of storing $k + 1$ objects, $k < n$. We focus on the model when elements are read into the memory in a FIFO fashion that correspond to streaming the data or a special type of caching. We present asymptotically optimal constructions; they are based on the Euler totient theorem and recursion.

1 Introduction

The problem that we study originated in the context of computing a two-argument function (which we denote by g) for all pairs of n large objects, such as medical images, bitmaps or matrices [GJ02]. The memory size is too small to store all the objects at once; we assume it can store at most $k + 1$ objects at the same time. For that reason, the simple two-loop algorithm that iterates through the pairs of objects is not useful, as most of the objects are not readily available. The task is to provide the shortest possible sequence of *read* operations that will ensure that, for all pairs (i, j), at some point in time both o_i and o_j will reside in memory and $g(o_i, o_j)$ (as well as $g(o_j, o_i)$, if g is non-symmetric) can be evaluated.

The *read* operation assumes that, if memory is full, the next element takes the place of one of the elements currently residing in memory. The particular selection of the element to be replaced is a part of the strategy. The replacement strategy based on a FIFO queue of size $k + 1$ (that is, remove the first element and append a new element to the end of the sequence) is particularly interesting because of its applications. For example, it appears in processing a

** This work was supported in part by the University of Kentucky subcontract of grants 5P20RR016481-03 and 2P20RR016481-04 from NCRR-NIH, and by KY NSF EPSCOR grant EPS-0132295.

large number of huge objects located in a remote database when locally storing fetched data may be either impossible or impractical and where limited bandwidth and time necessitate efficient scheduling of the data requests. Moreover, FIFO-like processing allows for read-ahead requests to preemptively fetch data. There are other applications of k-radius sequences; e.g., they may be viewed as a systematic and efficient scheduling for the caching process [SChD02].

The problem can be formalized as follows.

Let $a_1, a_2, \ldots, a_m$ be a sequence over $[n]$. The *distance* $\mathrm{d}(a_i, a_j) = |i - j|$. We say that a sequence $a_1, a_2, \ldots a_m$ has the *k-radius* property if every pair of different elements in $[n]$ occurs at least once within distance at most k. In other words, each pair of objects will appear at least once inside a window of size $k+1$ that slides along the sequence.

In $\boxed{1}\,\boxed{2}\,\boxed{3}\,\boxed{4}\,\boxed{5}\,\boxed{6}\,\boxed{1}\,\boxed{2}\,\boxed{4}\,\boxed{5}\,\boxed{3}\,\boxed{6}$ of length 12, each of 15 pairs of numbers from $\{1, 2, 3, 4, 5, 6\}$ occurs within a distance of at most 2. This 2-radius sequence for $n = 6$ is a shortest (although not unique) such sequence, as we will see later. Some objects, such as $\boxed{1}$ and $\boxed{2}$ appear in this sequence within distance not larger than 2 more than once. We place numbers in boxes to indicate that each element in the sequence consists of potentially huge satellite data.

We study the following questions:

Question 1. What is the least length $f_k(n)$ of a sequence over $[n] = \{1, 2, \ldots, n\}$ that has the *k-radius* property?

Question 2. How can we construct short sequences with the *k-radius* property?

We prove that

$$f_1(n) = \frac{1}{2}n^2\left(1 + o(1)\right), \qquad f_2(n) = \frac{1}{4}n^2\left(1 + o(1)\right),$$

and

$$\frac{1}{2k}n^2\left(1 + o(1)\right) \le f_k(n) \le \frac{1}{4 \cdot \lfloor (k+1)/3 \rfloor}n^2\left(1 + o(1)\right)$$

(for a fixed $k > 2$). We also show a construction of 2-radius sequences and use it as a ground condition in a recursive algorithm for constructing asymptotically optimal sequences with the k-radius property for $k > 2$. The correctness of the construction is based on the Euler totient theorem for congruences (mod n) of powers of the totient function $\phi(n)$: the number of numbers relatively prime with n [GKP89].

2 Simple cases, lower bounds and examples

A result for all pairs appearing consecutively in a sequence (in our terminology, the 1-radius property) was obtained by Ghosh [G75]; see also [LTT81]. It was formulated in the context of databases and the consecutive 1 property.

Theorem 1. *(Ghosh 1975)*

$$f_1(n) = \begin{cases} \binom{n}{2} + 1, & \textit{for } n \textit{ odd;} \\ \binom{n}{2} + \frac{1}{2}n, & \textit{for } n \textit{ even.} \end{cases}$$

For n objects, the length of the shortest sequence with the *k-radius* property depends on k and its length is bounded from below as follows:

Theorem 2.

$$f_k(n) \geq \left\lceil \frac{1}{k}\binom{n}{2} + \frac{k+1}{2} \right\rceil$$

Proof. There are $m-i$, $i = 1, 2, \ldots, k$, pairs of objects in a sequence $a_1, a_2, \ldots, a_m$ for which the distance is i. Hence we have

$$\sum_{i=1}^{k}(m-i) = mk - (1 + 2 + \cdots + k) = mk - \frac{(k+1)k}{2}$$

pairs for which the distance is at most k. Thus, if the sequence $a_1, a_2, \ldots, a_m$ contains all pairs over $[n]$ within distance at most k, then

$$\binom{n}{2} \leq mk - \frac{(k+1)k}{2}$$
$$f_k(m) \geq m \geq \frac{1}{k}\binom{n}{2} + \frac{k+1}{2}.$$

■

The above lower bound can be slightly improved.

Theorem 3.

$$f_k(n) \geq \left\lceil \frac{n-1}{2k} \right\rceil n + \sum_{j=1}^{k}\left(\left\lceil \frac{n+k-j}{2k} \right\rceil - \left\lceil \frac{n-1}{2k} \right\rceil\right)$$

Proof. Let $a_1, a_2, \ldots, a_m$ be a sequence over $[n]$ with the k-radius property and $m = f_k(n)$ (i.e. the sequence is of the shortest possible length). We claim that the first k objects $a_1, a_2, \ldots, a_k$ are pairwise different. Indeed, if $a_i = a_j$, $1 \leq i < j \leq k$, then we delete a_i. The resulting sequence still has the k-radius property, in contradiction with the definition of m.

Let $r(i)$ be the number of occurrences of i in the sequence $a_1, a_2, \ldots, a_m$, $i = 1, 2, \ldots, n$. Obviously, for every $i \in [n]$, $2kr(i) \geq n - 1$, so $r(i) \geq \lceil \frac{n-1}{2k} \rceil$.

For $j = 1, \ldots, k$, object a_j has only $j-1+k$ objects within distance at most k. Hence $j - 1 + k + (r(a_j) - 1) \cdot 2k \geq n - 1$
so

$$r(a_j) \geq \left\lceil \frac{n+k-j}{2k} \right\rceil, \qquad j = 1, 2, \ldots, k$$

As $a_1, a_2, \ldots, a_k$ are pairwise different,

$$\begin{aligned} f_k(n) = m = \sum_{i=1}^{n} r(i) &\geq (n-k) \left\lceil \frac{n-1}{2k} \right\rceil + \sum_{j=1}^{k} \left\lceil \frac{n+k-j}{2k} \right\rceil \\ &= \left\lceil \frac{n-1}{2k} \right\rceil n + \sum_{j=1}^{k} \left(\left\lceil \frac{n+k-j}{2k} \right\rceil - \left\lceil \frac{n-1}{2k} \right\rceil \right). \end{aligned}$$

■

Corollary 1.

$$f_2(n) \geq \begin{cases} \frac{1}{2}\binom{n}{2} + \frac{1}{4}n + 1, & n \equiv 0 \pmod 4 \\ \frac{1}{2}\binom{n}{2} + 2, & n \equiv 1 \pmod 4 \\ \frac{1}{2}\binom{n}{2} + \frac{3}{4}n, & n \equiv 2 \pmod 4 \\ \frac{1}{2}\binom{n}{2} + \frac{1}{2}n, & n \equiv 3 \pmod 4 \end{cases}$$

Remark 1. The bound given by Theorem 3 is for $k = 2$ not worse than the bound in Theorem 2.

For small values of n we can directly compare the lower and upper bounds.

$f_2(2) = 2$	12
$f_2(3) = 3$	123
$f_2(4) = 5$	12341
$f_2(5) = 7$	1234512
$f_2(6) = 12$	123456124536
$f_2(7) = 14$	12345632756147
$f_2(8) \geq 17$	

The lower bounds follow from the Corollary 1.

3 Construction for $k = 2$

The presented construction of a sequence with the *2-radius* property for n elements is based on the divisibility properties of numbers and their divisors, and Euler's theorem [GKP89].

We first prove several lemmas.

Let p be a positive divisor of an odd positive integer n. Define

$$[n]_p = \left\{p, 2p, \ldots, \left(\frac{n}{p} - 1\right) p\right\} \qquad [n]_p^* = \left\{ip \in [n]_p : \left(i, \frac{n}{p}\right) = 1\right\}$$

where (a, b) is the greatest common divisor of a and b.

Lemma 1. *The family* $\left\{[n]_p^* : \ p \mid n \text{ and } 1 \leq p < n\right\}$ *is a partition of* $[n-1]$.

Proof. First we show that

$$[n-1] = \bigcup_{p \mid n} [n]_p^*$$

Let $m \in [n-1]$ and define $d = (m, n)$. Then $\left(\frac{m}{d}, \frac{n}{d}\right) = 1$, so $m = \frac{m}{d} \cdot d \in [n]_d^*$.

It suffices to show that $[n]_p^* \cap [n]_q^* = \emptyset$, for two different positive divisors p and q of n. Suppose it is not true and let $m \in [n]_p^* \cap [n]_q^*$. Then $m = ip = jq$, where $(i, n/p) = 1$ and $(j, n/q) = 1$. Denote $d = (p, q)$. Clearly, there are positive integers p_1 and q_1 such that $p = dp_1$, $q = dq_1$, and $(p_1, q_1) = 1$. As $ip = jq$, $idp_1 = jdq_1$, so $ip_1 = jq_1$. Since $(p_1, q_1) = 1$, $q_1 \mid i$. On the other hand, $q \mid n$, so $n = qn_1$ for some positive integer n_1. Consequently,

$$\frac{n}{p} = \frac{qn_1}{dp_1} = \frac{dq_1 n_1}{dp_1} = q_1 \frac{n_1}{p_1}$$

(as n/p is an integer and $(p_1, q_1) = 1$, n_1/p_1 is itself an integer). Hence $q_1 \mid n/p$ and $q_1 \mid i$, so $q_1 = 1$ because $(i, n/p) = 1$. For a similar reason $p_1 = 1$ and consequently $p = d = q$, a contradiction. ∎

We define a directed graph G_n on $[n-1]$. A pair (i, j) is an edge in G_n if $j \equiv 2i \pmod{n}$, for $i, j \in [n-1]$. Notice that G_n is well-defined. Indeed, suppose that $2i \equiv 0 \pmod{n}$ for some $i \in [n-1]$. As n is odd, it implies $i \equiv 0 \pmod{n}$, a contradiction. Clearly, for each $i \in [n-1]$, there is a unique $j \in [n-1]$ such that (i, j) is an edge in G_n. Hence the outdegrees of vertices in G_n are all equal to 1.

As n is odd, by Euler's theorem [GKP89] we have $2^{\varphi(n)} \equiv 1 \pmod{n}$, where $\varphi(n)$ is Euler's function [GKP89]. Let $j \in [n-1]$ and define $i \equiv 2^{\varphi(n)-1} \cdot j \pmod{n}$. Then

$$2i = 2 \cdot \left(2^{\varphi(n)-1} \cdot j\right) = 2^{\varphi(n)} \cdot j \equiv j \pmod{n}$$

Hence the indegree of each vertex in G_n is at least 1. Since the sum of the indegrees of all vertices in a directed graph is equal to the sum of the outdegrees, the indegree of each vertex is 1. Therefore the components of G_n are cycles.

Lemma 2. *Let* $p \mid n$, $1 \leq p < n$, *and let* t_p *be the least integer* $t > 1$ *such that* $2^t \equiv 1 \pmod{\frac{n}{p}}$. *The graph induced in* G_n *by vertices from* $[n]_p^*$ *is a family of cycles of length* t_p.

Proof. Let $s \in [n]_p^*$. Then $s = ip$, $1 \le i < \frac{n}{p}$ and $(i, n/p) = 1$.

If $1 \le i \le \frac{n}{2p}$, then $2s = 2ip < n$ (because n is odd), so $2s \in [n]_p$. If $\frac{n}{2p} < i < \frac{n}{p}$ then $2s = 2ip \equiv 2ip - n = (2i - n/p)p \pmod{n}$ and $1 \le 2i - \frac{n}{p} < 2\frac{n}{p} - \frac{n}{p} = \frac{n}{p}$. Hence $2s - n \in [n]_p$. We have shown that $2s \pmod{n}$ belongs to $[n]_p$. Let $d = (2i, n/p)$, when $1 \le i \le \frac{n}{2p}$ and $d = (2i - n/p, n/p)$, when $\frac{n}{2p} < i < \frac{n}{p}$. Since n/p is odd and $(i, n/p) = 1$ we get $d = 1$.

We proved that the cycle in G_n containing s has all vertices in $[n]_p^*$. The length of this cycle is the smallest integer t such that $s \equiv 2^t \cdot s \pmod{n}$. This is equivalent to $ip \equiv 2^t \cdot ip \pmod{n}$, for some $1 \le i < n/p$, which in turn is equivalent to $(2^t - 1)\, i \equiv 0 \pmod{n/p}$. Since $(i, n/p) = 1$, the last condition is equivalent to $2^t \equiv 1 \pmod{n/p}$. Hence the length of the cycle containing s, and consequently the length of any cycle in the subgraph of G_n induced by the vertices of $[n]_p^*$, is t_p. ■

Lemma 3. *The number of cycles in G_n is at most $\frac{5n}{\log_2 n}$.*

Proof. Let $p \mid n$, $1 \le p < n$. Observe first that, as $2^{t_p} \equiv 1 \pmod{n/p}$ and $t_p > 1$, $2^{t_p} \ge n/p + 1$. Hence $t_p \ge \log_2(n/p)$. Thus the number of cycles in the subgraph of G_n induced by the set of vertices $[n]_p^*$ is at most $\frac{1}{log_2(n/p)} \cdot \varphi\,(n/p)$. Consequently, the total number of cycles in G_n is at most

$$\begin{aligned}
\sum_{\substack{p|n \\ 1\le p<n}} \frac{1}{\log_2 \frac{n}{p}} \varphi\left(\frac{n}{p}\right) &= \sum_{\substack{p|n \\ 1<p\le n}} \frac{\varphi(p)}{\log_2 p} = \sum_{\substack{p|n \\ 1<p<n^{\frac{1}{3}}}} \frac{\varphi(p)}{\log_2 p} + \sum_{\substack{p|n \\ n^{\frac{1}{3}}\le p\le n}} \frac{\varphi(p)}{\log_2 p} \\
&\le \sum_{\substack{p|n \\ 1<p<n^{\frac{1}{3}}}} n^{\frac{1}{3}} + \sum_{\substack{p|n \\ n^{\frac{1}{3}}\le p\le n}} \frac{\varphi(p)}{\log_2 n^{\frac{1}{3}}} \le n^{\frac{2}{3}} + \frac{1}{\frac{1}{3}\log_2 n} \cdot \sum_{\substack{p|n \\ 1<p\le n}} \varphi(p) \\
&= n^{\frac{2}{3}} + \frac{3}{\log_2 n} \cdot n \le \frac{5n}{\log_2 n}.
\end{aligned}$$

■

Let $\overline{G}_n$ be a directed graph on $[(n-1)/2]$ such that, for $i, j \in [(n-1)/2]$, (i, j) is an edge if

$$j = \begin{cases} 2i, & \text{if } 2i \le \frac{n-1}{2} \\ n - 2i, & \text{if } 2i > \frac{n-1}{2}. \end{cases}$$

Clearly, the outdegree of every vertex in $\overline{G}_n$ is 1. Let $j \in [(n-1)/2]$. If j is even, i.e. $j = 2i$ for some $i \in [(n-1)/2]$, then (i, j) is an edge in $\overline{G}_n$. If j is odd then let $i = (n-j)/2$. Since $2i = n - j > (n-1)/2$ and $n - 2i = j$, (i, j) is an edge in $\overline{G}_n$. We have shown that every vertex in $\overline{G}_n$ has outdegree and indegree equal to 1.

Let $s \in [(n-1)/2]$ and denote by $\overline{C}$ the cycle in $\overline{G}_n$ containing s. Let $C = \{c_0, c_1, \ldots, c_{t-1}\}$ be a cycle in G_n containing s, $c_j \equiv 2^j s \pmod{n}$. We shall show by induction on j that $\overline{c}_j \in \overline{C}$, where $\overline{c}_j = \min(c_j, n - c_j)$, $j = 0, 1, \ldots, t-1$. Obviously, $\overline{c}_0 = s \in \overline{C}$. Assume that $\overline{c}_0, \overline{c}_1, \ldots, \overline{c}_{j-1} \in \overline{C}$, $j < t$. We shall show that $\overline{c}_j \in \overline{C}$. Let us consider four cases.

Case 1: $c_{j-1} \le (n-1)/2$ and $c_j \le (n-1)/2$

Then $\overline{c}_{j-1} = c_{j-1}$, $\overline{c}_j = c_j$, and $2\overline{c}_{j-1} = \overline{c}_j \le (n-1)/2$, so $(\overline{c}_{j-1}, \overline{c}_j)$ is an edge in $\overline{G}_n$. Consequently, $\overline{c}_j \in \overline{C}$ (as $\overline{c}_{j-1} \in \overline{C}$).

Case 2: $c_{j-1} \le (n-1)/2$ and $c_j > (n-1)/2$

Then $\overline{c}_{j-1} = c_{j-1}$ and $\overline{c}_j = n - c_j$. Observe that $2\overline{c}_{j-1} = 2c_{j-1} = c_j > (n-1)/2$ and $n - 2\overline{c}_{j-1} = n - c_j = \overline{c}_j$, so $(\overline{c}_{j-1}, \overline{c}_j)$ is an edge in $\overline{G}_n$. Consequently, $\overline{c}_j \in \overline{C}$.

Case 3: $c_{j-1} > (n-1)/2$ and $c_j \le (n-1)/2$

Then $\overline{c}_{j-1} = n - c_{j-1}$ and $\overline{c}_j = c_j$. Moreover $2c_{j-1} = n + c_j$. Observe that

$$\begin{aligned} 2\overline{c}_{j-1} &= 2(n - c_{j-1}) = 2n - 2c_{j-1} = 2n - (n + c_j) = n - c_j \\ &\ge n - \frac{n-1}{2} = \frac{n+1}{2} \end{aligned}$$

and $n - 2\overline{c}_{j-1} = c_j = \overline{c}_j$, so $(\overline{c}_{j-1}, \overline{c}_j)$ is an edge in $\overline{G}_n$. Consequently, $\overline{c}_j \in \overline{C}$.

Case 4: $c_{j-1} > (n-1)/2$ and $c_j > (n-1)/2$

Then $\overline{c}_{j-1} = n - c_{j-1}$, $\overline{c}_j = n - c_j$, and $2c_{j-1} = n + c_j$. Moreover,

$$2\overline{c}_{j-1} = 2n - 2c_{j-1} = 2n - n - c_j < n - \frac{n-1}{2} = \frac{n+1}{2}$$

so $2\overline{c}_{j-1} \le \frac{n-1}{2}$. Since $2\overline{c}_{j-1} = n - c_j = \overline{c}_j$, we have that $(\overline{c}_{j-1}, \overline{c}_j)$ is an edge in $\overline{G}_n$, and thus $\overline{c}_j \in \overline{C}$.

We proved that $\{\overline{c}_0, \overline{c}_1, \ldots, \overline{c}_{t-1}\} \subseteq \overline{C}$ and that $(\overline{c}_{j-1}, \overline{c}_j)$ is an edge in $\overline{G}_n$, for $j = 1, 2, \ldots, t-1$. It is easy to show (proceeding as in cases 1 and 3) that $(\overline{c}_{t-1}, \overline{c}_0)$ is an edge in $\overline{G}_n$. Thus

$$\{\overline{c}_0, \overline{c}_1, \ldots, \overline{c}_{t-1}\} = \overline{C}$$

We have proved that the number of cycles in $\overline{G}_n$ is not larger than in G_n. By Lemma 3 we get the next lemma.

Lemma 4. *The number of cycles in $\overline{G}_n$ is at most $\frac{5n}{\log_2 n}$.*

Now a sequence with the *2-radius* property can be defined as follows.

We choose from each cycle C in $\overline{G}_n$ $\left\lfloor \frac{1}{2}\ell_c \right\rfloor$ pairs of vertices joined by an edge, where ℓ_c is the length of C. This way, letting C range over the cycles of $\overline{G}_n$, we get at least

$$\sum_C \left\lfloor \frac{1}{2}\ell_c \right\rfloor \geq \sum_C \left(\frac{1}{2}\ell_c - \frac{1}{2}\right) \geq \frac{1}{2}\sum_C \ell_c - \frac{1}{2}\frac{5n}{\log_2 n} = \frac{1}{2}\frac{n-1}{2} - \frac{1}{2}\frac{5n}{\log_2 n}$$

pairwise disjoint pairs (by Lemma 4). Denote by $\mathcal{A}$ the set of is in the chosen pairs.

For each such pair (i,j) we create a sequence s_i, which is a concatenation of the following d sequences, where $d = d_i = (i,n)$:

$0, i, 2i, \ldots, \left(\frac{n}{d} - 1\right) i, 0, i \pmod n$
$1, i+1, 2i+1, \ldots, \left(\frac{n}{d} - 1\right) i+1, 1, i+1 \pmod n$
$2, i+2, 2i+2, \ldots, \left(\frac{n}{d} - 1\right) i+2, 2, i+2 \pmod n$
$\ldots$
$d-1, i+d-1, 2i+d-1, \ldots, \left(\frac{n}{d} - 1\right) i+d-1, d-1, i+d-1 \pmod n$.

For $\alpha, \beta \in \{0, 1, \ldots, n-1\}$, define the distance

$$\operatorname{dist}(\alpha,\beta) = \operatorname{dist}(\beta,\alpha) = \min\left(|\beta-\alpha|, n - |\beta-\alpha|\right).$$

Observe that every pair in $\{0, 1, \ldots, n-1\}$ for which the distance is i occurs in s_i as consecutive objects; and each pair for which the distance is j occurs in s_i within a distance of 2. Moreover, the length of s_i is $n + 2d$.

The number of vertices in $\overline{G}_n$ which are not in any of the chosen pairs (i,j) is at most

$$\frac{n-1}{2} - \left(\frac{n-1}{2} - \frac{5n}{\log_2 n}\right) = \frac{5n}{\log_2 n}.$$

For each such vertex ℓ we create a sequence s_ℓ given by the same definition as s_i (replacing i with ℓ). Denote the set of these vertices ℓ by $\mathcal{B}$.

It is easily seen that a concatenation s of all sequences s_i and s_ℓ has the 2-radius property. The length of s is

$$\begin{aligned}\sum_{i\in\mathcal{A}\cup\mathcal{B}} (n+2d_i) &= \sum_{i\in\mathcal{A}}(n+2d_i) + \sum_{i\in\mathcal{B}}(n+2d_i) \leq |\mathcal{A}|\cdot n + 2\sum_{i\in\mathcal{A}} d_i + |\mathcal{B}|\cdot 2n \\ &\leq \frac{1}{2}\frac{n-1}{2}n + \frac{5n}{\log_2 n}\cdot 2n + 2\sum_{i=1}^{n-1} d_i = \frac{1}{2}\binom{n}{2} + \frac{10n^2}{\log_2 n} + 2\sum_{i=1}^{n-1} d_i.\end{aligned}$$

Let us estimate the sum $\sum_{i=1}^{n-1} d_i$.

$$\sum_{i=1}^{n-1} d_i = \sum_{\substack{p|n \\ 1\leq p<n}} p\cdot|\{i : (i,n) = p\}| = \sum_{\substack{p|n \\ 1\leq p<n}} p\cdot\varphi\left(\frac{n}{p}\right) \leq \sum_{\substack{p|n \\ 1\leq p<n}} p\cdot\frac{n}{p} = \sum_{\substack{p|n \\ 1\leq p<n}} n.$$

Let $n = p_1 p_2 \cdots p_r$ be a factorization of n into (not necessarily distinct) primes. The number of divisors of n is then not larger than 2^r. As n is odd, $n = p_1 p_2 \cdots p_r \geq 3^r$, so $2^r \leq 2^{\log_3 n} = 2^{\log_2 n \cdot \log_3 2} = n^{\log_3 2} \leq n^{0.64}$. Consequently $\sum_{i=1}^{n-1} d_i \leq n^{1.64}$ and the length of s is at most

$$\frac{1}{2}\binom{n}{2} + \frac{10n^2}{log_2 n} + 2n^{1.64} = \frac{1}{2}\binom{n}{2}(1 + o(1)).$$

If n is even then we construct the sequence s for $n-1$ and concatenate it with the sequence $1, 2, n, 3, 4, 5, 6, n, 7, 8, 9, 10, n, \ldots$ whose length is not larger than $n - 1 + \lceil \frac{n-1}{4} \rceil$.

4 A general construction using a 1-radius sequence.

Let $M = \left\lceil \frac{n}{\lfloor (k+1)/2 \rfloor} \right\rceil$ and $X = \{x_1, x_2, \ldots, x_M\}$. Consider a sequence p of length $f_1(M)$ in which each pair of elements from X occurs consecutively. We partition the set $[n]$ into M disjoint subsets of cardinality $\left\lfloor \frac{k+1}{2} \right\rfloor$ except for at most one of a smaller cardinality. Denote these subsets by $A_1, A_2, \ldots, A_M$.

Denote by s_i any sequence (permutation) of elements of A_i. Define s to be the sequence obtained from p by replacing every occurrence of x_i by the sequence s_i (for every $i = 1, 2, \ldots, M$). Observe that each two elements $r, t \in [n]$ occur within distance at most k in the sequence s. Indeed, let $r \in A_p \subseteq [n]$ and $t \in A_q \subseteq [n]$. The elements $x_p, x_q \in X$ are at least once neighbors in the sequence p. Since $|s_p| \leq \left\lfloor \frac{k+1}{2} \right\rfloor$ and $|s_q| \leq \left\lfloor \frac{k+1}{2} \right\rfloor$, r and t are in s within distance at most $2\left\lfloor \frac{k+1}{2} \right\rfloor - 1 \leq k$.

We proved that $f_k(n) \leq |s|$. Let us compute $|s|$. If M is odd then

$$\begin{aligned}
|s| &\leq f_1(M) \cdot \left\lfloor \frac{k+1}{2} \right\rfloor \leq \left(\binom{M}{2} + 1\right) \left\lfloor \frac{k+1}{2} \right\rfloor = \left(\frac{1}{2}M^2 - \frac{1}{2}M + 1\right) \left\lfloor \frac{k+1}{2} \right\rfloor \\
&\leq \left(\frac{1}{2} \cdot \frac{(n + \lfloor (k+1)/2 \rfloor)^2}{\lfloor (k+1)/2 \rfloor^2} - \frac{1}{2} \cdot \frac{n + \lfloor (k+1)/2 \rfloor}{\lfloor (k+1)/2 \rfloor} + 1\right) \cdot \left\lfloor \frac{k+1}{2} \right\rfloor \\
&= \frac{1}{2} \frac{n^2}{\lfloor (k+1)/2 \rfloor} + \frac{1}{2} n + \left\lfloor \frac{k+1}{2} \right\rfloor.
\end{aligned}$$

If M is even, we have

$$\begin{aligned}
|s| &\leq \left(\binom{M}{2} + \frac{1}{2}M\right) \left\lfloor \frac{k+1}{2} \right\rfloor = \frac{1}{2} \left\lceil \frac{n}{\lfloor (k+1)/2 \rfloor} \right\rceil^2 \cdot \left\lfloor \frac{k+1}{2} \right\rfloor \\
&\leq \frac{1}{2} \cdot \frac{(n + \lfloor (k+1)/2 \rfloor)^2}{\lfloor (k+1)/2 \rfloor^2} \cdot \left\lfloor \frac{k+1}{2} \right\rfloor = \frac{1}{2} \frac{n^2}{\lfloor (k+1)/2 \rfloor} + n + \frac{1}{2} \left\lfloor \frac{k+1}{2} \right\rfloor.
\end{aligned}$$

Consequently, we have the following:

Theorem 4. $f_k(n) \leq \frac{n^2}{2\lfloor (k+1)/2 \rfloor} + n + \frac{1}{2} \left\lfloor \frac{k+1}{2} \right\rfloor$.

5 A general construction using a 2-radius sequence

Let $M = \left\lceil \frac{n}{\lfloor (k+1)/3 \rfloor} \right\rceil$ and $X = \{x_1, \ldots, x_M\}$. Consider the sequence p of length $f_2(M)$ in which two elements of X occur within distance at most 2. We partition the set $[n]$ into M disjoint subsets of cardinality $\lfloor \frac{k+1}{3} \rfloor$, except for at most one of a smaller cardinality. Call these subsets $A_1, A_2, \ldots, A_M$.

As in the previous construction we denote by s_i any sequence of elements of A_i. The sequence s is defined as in the previous construction.

Observe that each two elements $r, t \in [n]$ occur within distance at most k in the sequence s. Indeed, let $r \in A_p \subseteq [n]$ and $t \in A_q \subseteq [n]$. As x_p and $x_q$occur within distance at most 2 in p, r and t are in s within a distance of $3 \lfloor \frac{k+1}{3} \rfloor - 1 \leq k$ of each other.

We have proved that

$$f_k(n) \leq f_2(M) \cdot \left\lfloor \frac{k+1}{3} \right\rfloor = |s| \, .$$

Let us estimate $|s|$, for a fixed k.

$$|s| = \left\lfloor \frac{k+1}{3} \right\rfloor \cdot \frac{1}{4} \left(\frac{n}{\lfloor (k+1)/3 \rfloor} \right)^2 (1 + o(1)) = \frac{1}{4} \frac{n^2}{\lfloor (k+1)/3 \rfloor} (1 + o(1)) \, .$$

We have the following theorem.

Theorem 5.

$$f_k(n) \leq \frac{n^2}{4 \lfloor (k+1)/3 \rfloor} (1 + o(1))$$

To illustrate the algorithm, consider a set $A = \{a, b, c, d, e, f, g, h, i, j, k, l\}$ of cardinality $n = 12$. Take $k = 5$. To construct a sequence with the 5-radius property, we can use a sequence $\boxed{1}\,\boxed{2}\,\boxed{3}\,\boxed{4}\,\boxed{5}\,\boxed{6}\,\boxed{1}\,\boxed{2}\,\boxed{4}\,\boxed{5}\,\boxed{3}\,\boxed{6}$ with the 2-radius property for $M = 6$ objects. Since $k = 5$, we partition A into groups of size $\frac{6}{3} = 2$ each, for example by taking consecutive pairs of letters, such as $A_1 = \{a, b\}$, $A_2 = \{c, d\}$,..., $A_6 = \{k, l\}$. Following the algorithm, we replace each number i with A_i. This results in $\boxed{a, b}\,\boxed{c, d}\,\boxed{e, f}\,\boxed{g, h}\,\boxed{i, j}\,\boxed{k, l}\,\boxed{a, b}\,\boxed{c, d}\,\boxed{g, h}\,\boxed{i, j}\,\boxed{e, f}\,\boxed{k, l}$ (i.e., $abcdefghijklabcdghijefkl$), which has the 5-radius property.

The reasoning applied in the two constructions can be easily generalized to prove the following proposition.

Proposition 1. *For $k \geq m$,*

$$f_k(n) \leq f_m \left(\left\lceil \frac{n}{\lfloor (k+1)/(m+1) \rfloor} \right\rceil \right) \cdot \left\lfloor \frac{k+1}{m+1} \right\rfloor$$

.

Corollary 2.

$$f_1(n) = \frac{1}{2}n^2\left(1 + o(1)\right), \qquad f_2(n) = \frac{1}{4}n^2\left(1 + o(1)\right),$$

and finally

$$\frac{1}{2k}n^2\left(1 + o(1)\right) \le f_k(n) \le \frac{1}{4 \cdot \lfloor (k+1)/3 \rfloor}n^2\left(1 + o(1)\right),$$

for a fixed $k > 2$.

6 Other strategies

Above we have focused on a particular strategy based on a FIFO (of size $k + 1$) that removes the oldest element and places the new element at the end of the sequence. While this strategy seems to be natural for streaming and transmission of data, other strategies exist that can be useful for other applications. In the general case, we could allow the removal of any element from the current sequence of $k+1$ objects, placing the new element in an arbitrary position in this sequence. Such a problem is of interest in caching when costs may depend on what locations in the cache are affected. In a more specific case, the replacement could follow the LIFO strategy, which replaces elements by popping one or more of the most recently inserted objects and then pushing new elements. Such a strategy lends itself to a simple construction of a sequence: divide the set into disjoint groups of k elements (except perhaps for the last group that can be smaller); place a group in the memory and add as the $(k+1)$st element, consecutively, all the remaining elements from not yet processed groups; when finished, replace the group with the next group and repeat the process until all the groups have been in the memory. This leads to a sequence of the length similar to the FIFO strategy, but occasionally all the elements in the memory must be replaced.

7 Conclusions

Sequences with the k-radius property are useful in scheduling the fetching of data to process each pair of objects—e.g., to compute two-argument functions—when the memory is limited and the objects are huge. As such they are applicable to streaming and caching images or arrays and they help optimize the transmission time or the number of memory rewrites when the application requires access to every pair of objects. The presented results demonstrate asymptotically optimal constructions for sequences with the k-property and demonstrate that the length of such sequences approximately halves when the memory size doubles. Additional constructions based on Steiner systems and finite geometries are presented in [JL04]. One natural generalization of the problem is to consider shortest sequences where all the triples, or quadruples (etc.) appear within radius k at least once.

Acknowledgement

The first author would like to thank J. Gilkerson for discussions that let to the formulation of the problem, N. Imam for experimental implementations for small-radius sequencies, and N. Moore for help with typesetting and proofreading.

Bibliography

[CR99] Colbourn, Ch., J., Rosa, A., Triple Systems, Clarendon Press - Oxford, 1999.

[GJ02] Gilkerson, J. W., Jaromczyk, J. W., Restoring the order of images in a sequence of MRI slices, Manuscript, 2002.

[G75] Gosh, S.P., Consecutive storage of relevant records with redundancy, *Communications of the ACM* 18(8):464-471, 1975.

[GGL95] Graham, R. L., Grötschel, M., Lovász, L., Handbook of Combinatorics, The MIT Press - North Holland, 1995.

[GKP89] Graham, R. L., Knuth, D. E., Patashnik, O., Concrete Mathematics, Addison-Wesley, 1989.

[JL04] Jaromczyk, J. W., Lonc, Z., Sequences of radius k. *Technical Report TR 417-04, Computer Science - University of Kentucky*, 2004.

[LTT81] Lonc Z., Traczyk T. Truszczyński, M., Optimal f-graphs for the family of all k-subsets of an n-set. Data base file organization (Warsaw, 1981), 247–270, Notes Rep. Comput. Sci. Appl. Math., 6, Academic Press, New York, 1983.

[SChD02] Sen, S., Chatterjee, S., Dumir, N., Towards a theory of cache-efficient algorithms, Journal of the ACM, 49(6):828-858, 2002.