Combinatorial Batch Codes

Maura B. Paterson[1] Douglas R. Stinson[2] Ruizhong Wei[3]

[1]Information Security Group
Royal Holloway, University of London

[2]David R. Cheriton School of Computer Science
University of Waterloo

[3]Department of Computer Science
Lakehead University

Combinatorial Configurations and their Applications
August 5, 2009

combinatorial batch codes

Batch codes were introduced by Ishai, Kushilevitz, Ostrovsky and Sahai at STOC 2004. We study a special case of batch codes that we call combinatorial batch codes.

- n items

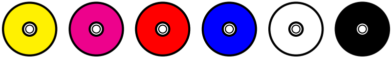
- m servers

- N:=total number of items stored
- Goal: retrieve any k items by reading at most one from each server
- Here $n = 6$, $m = 3$, $N = 9$ and $k = 2$

combinatorial batch codes

Batch codes were introduced by Ishai, Kushilevitz, Ostrovsky and Sahai at STOC 2004. We study a special case of batch codes that we call combinatorial batch codes.

- n items

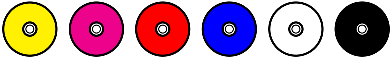


- m servers
- N:=total number of items stored
- Goal: retrieve any k items by reading at most one from each server
- Here $n = 6$, $m = 3$, $N = 9$ and $k = 2$

combinatorial batch codes

Batch codes were introduced by Ishai, Kushilevitz, Ostrovsky and Sahai at STOC 2004. We study a special case of batch codes that we call combinatorial batch codes.

- n items



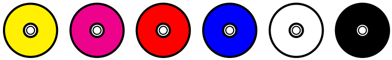
- m servers

- N:=total number of items stored
- Goal: retrieve any k items by reading at most one from each server
- Here $n = 6$, $m = 3$, $N = 9$ and $k = 2$

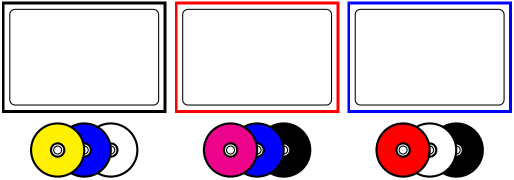
combinatorial batch codes

Batch codes were introduced by Ishai, Kushilevitz, Ostrovsky and Sahai at STOC 2004. We study a special case of batch codes that we call combinatorial batch codes.

- n items



- m servers

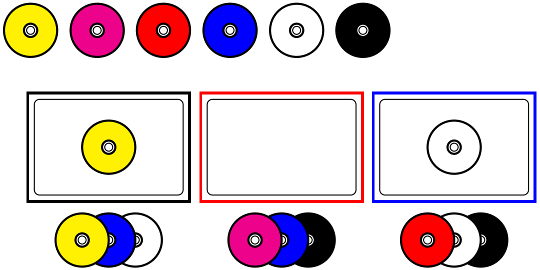


- N:=total number of items stored
- Goal: retrieve any k items by reading at most one from each server
- Here $n = 6$, $m = 3$, $N = 9$ and $k = 2$

combinatorial batch codes

Batch codes were introduced by Ishai, Kushilevitz, Ostrovsky and Sahai at STOC 2004. We study a special case of batch codes that we call combinatorial batch codes.

- n items
- m servers

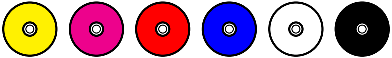


- N:=total number of items stored
- Goal: retrieve any k items by reading at most one from each server
- Here $n = 6$, $m = 3$, $N = 9$ and $k = 2$

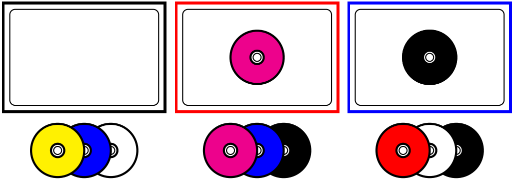
combinatorial batch codes

Batch codes were introduced by Ishai, Kushilevitz, Ostrovsky and Sahai at STOC 2004. We study a special case of batch codes that we call combinatorial batch codes.

- n items



- m servers



- N:=total number of items stored
- Goal: retrieve any k items by reading at most one from each server
- Here $n = 6$, $m = 3$, $N = 9$ and $k = 2$

questions

Notation: $(n, N, k, m) - CBC$
n items, total storage N, m servers, k items read

Given n, m, k, what is the minimum possible value of N? Denote this value by $N(n, k, m)$.

For a fixed rate $\frac{n}{N}$, and fixed k, what is the largest possible value of n (as a function of m)?

incidence matrix representation

Lemma

An $m \times n$ 0-1 matrix containing exactly N ones is an incidence matrix of an $(n, N, k, m) - CBC \Leftrightarrow$ any k columns contain a $k \times k$ submatrix with a transversal containing k ones.

items

servers						
	1	0	0	1	1	0
	0	1	0	1	0	1
	0	0	1	0	1	1

We view the incidence matrix as representing a set system:

point $\leftrightarrow$ server

block $\leftrightarrow$ set of servers containing a particular item

Hall's marriage theorem

Theorem

Suppose $(X, \mathcal{B})$ is a set system. Then any subcollection of k blocks $B_1, B_2, \ldots, B_k$ has a system of distinct representatives $\Leftrightarrow$ for all i, $1 \leq i \leq k$ it holds that

SDR(i) *for any subcollection of i blocks $B_{j_1}, B_{j_2}, \ldots, B_{j_i} \in \mathcal{B}$ it holds that $\left|\bigcup_{l=1}^{i} B_{j_l}\right| \geq i$.*

The previous example corresponds to the set system containing blocks $\{1\}$, $\{2\}$, $\{3\}$, $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$.

trivial examples

- Each server stores all items: $m = k$, $N = kn$
- Each server stores one item: $m = n = N$. Here we have $N(n, k, n) = n$.

Therefore we are only interested in examples with $n < N < kn$. When $k = m$, we have $N(n, k, k) = kn - k(k-1)$.

Example: $k = 4$, $n = 7$

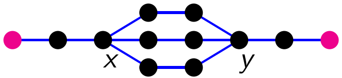
1	0	0	0	1	1	1
0	1	0	0	1	1	1
0	0	1	0	1	1	1
0	0	0	1	1	1	1

Next, we present a construction for batch codes where n is a bit larger than m.

(k,p)-flying saucer

Suppose $k \equiv 2 \pmod 3$

- Two vertices x and y are joined by p (disjoint) paths of length $\frac{k+1}{3}$.
- Paths of length $\frac{k-2}{3}$ are joined to x and y.



$k = 8,\ p = 3$

- number of vertices is $V = \frac{(p+2)(k-2)}{3} + 2$.
- number of edges is $E = \frac{p(k+1)}{3} + \frac{2(k-2)}{3}$.

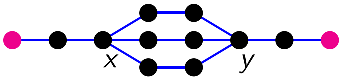
constructing a CBC from a flying-saucer

Theorem

Let k, p be positive integers with $k \equiv 2 \pmod 3$ and suppose $m \geq V$. Then there exists an $(m+p, m+p+E, k, m) - CBC$.

- Construct a (k,p)-FS and add isolated vertices until there are m vertices.
- Add a loop to each isolated vertex, and to each of the vertices of degree 1.
- Construct an incidence matrix whose rows are labelled by the vertices and whose columns are labelled by edges.

constructing a CBC from a flying-saucer



$k = 8, \ p = 3$

1	0	0	0	0	0	0	0	0	0	0	0	0	1	0
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	1	1	1	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	1	0	0	1	1	0	0	0
0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	1	1	0	0

$(15, 28, 8, 12)$-CBC

the special case $n = m + 1$

Theorem

For any positive integer k, we have $N(m+1, k, m) = m + k$.

- A $(k, 1)$-FS is just a path of length $k - 1$
- The construction produces an $(m+1, m+k, k, m)$-CBC.
- The pigeon-hole principle can be used to show this is the best you can do.

an optimal construction for large n

Theorem
For $n \geq (k-1)\binom{m}{k-1}$, we have

$$N(n,k,m) = kn - (k-1)\binom{m}{k-1}.$$

Example: $n = 23,\ m = 5,\ k = 3$

1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
1	1	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	1	1	1
0	0	1	1	0	0	0	0	1	1	0	0	0	0	1	1	1	1	0	0	1	1	1
0	0	0	0	1	1	0	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	0
0	0	0	0	0	0	1	1	0	0	0	0	1	1	0	0	1	1	1	1	0	0	0

uniform batch codes with rate $\frac{1}{c}$

A batch code is uniform if every server stores the same number of items.

We would like to determine the maximum n for which there exists a uniform (n, cn, k, m)-CBC; denote this number by $n(m, c, k)$.

Theorem

$n(m, c, k) \leq \frac{(k-1)\binom{m}{c}}{\binom{k-1}{c}}$.

We have equality in the following cases:

- $n(m, c, c+1) = c\binom{m}{c}$
- $n(m, c, c+2) = \binom{m}{c}$

batch codes with rate $\frac{1}{2}$

- Each block has two points $\rightarrow$ we can represent the set system by a multigraph.
- For each $i \leq k$, the graph contains no subgraphs with i edges but fewer than i vertices.

Lemma

If there is a graph G with m vertices, n edges and girth g, then there is a uniform $(n, 2n, k, m)$-CBC with $k = 2g - \lfloor g/2 \rfloor - 1$ and rate $1/2$.

constructions from bipartite graphs

- A bipartite graph has girth at least 4.
- The complete bipartite graph $K_{\lceil \frac{m}{2} \rceil, \lfloor \frac{m}{2} \rfloor}$ yields a uniform $(\left\lceil (m^2-1)/4 \right\rceil, 2\left\lceil (m^2-1/4 \right\rceil, 5, m)$-CBC with rate $1/2$.

Theorem

$$\left\lceil \frac{m^2-1}{4} \right\rceil \leq n(m,2,5) \leq \left\lceil \frac{m^2+2m-3}{4} \right\rceil .$$

The proof of the upper bound uses a result of Dirac on graphs that contain no subgraph isomorphic to $K_4 - e$.

d-regular graphs of large girth

Margulis 1984, Lubotzky *et al.* 1988 constructed d-regular graphs with girth

$$g \geq \frac{4}{3}\frac{\log m}{\log(d-1)} - \frac{\log 4}{\log(d-1)},$$

where $d-1$ is any prime $p \equiv 1 \pmod 4$ and m is the number of vertices.

Theorem

There exists a uniform $(dm/2, dm, 2\log m/\log(d-1), m)$-CBC when $d-1 \equiv 1 \pmod 4$ is prime.

(This CBC has $n = \Omega(m^{(k+2)/k})$.)

probabilistic construction for arbitrary c

Theorem

For integers $c \geq 2$, $k \geq 2$ there exists a constant $a_{c,k}$ such that there exists a uniform (n, cn, k, m)-CBC with $n \geq a_{c,k} m^{ck/(k-1)-1}$, having rate $1/c$.

Here $n = \Omega(m^{ck/(k-1)-1})$. This improves the result in Ishai *et al.* who showed that $n = \Omega(m^{c-1})$.

open problems

1. How close to being optimal are the constructions using flying saucers? In particular, is it true that

$$N(m+p,k,m) - N(m+p-1,k,m) \approx \frac{k}{3}$$

when $p > 1$ and m is sufficiently large as a function of p and k?

2. Are there explicit constructions for "good" uniform batch codes with fixed rate $1/c$, where $c > 2$ is an integer?
3. Can $N(n,k,m)$ be computed for a range of values of n, where $n < (k-1)\binom{m}{k-1}$?

thank you for your attention!

Y. Ishai, E. Kushilevitz, R. Ostrovsky and A. Sahai. *Batch codes and their applications*, in Proceedings of the 36th Annual ACM Symposium on Theory of Computing, ACM Press, New York, 262–271.

M. B. Paterson, D. R. Stinson and R. Wei. *Combinatorial batch codes*, Advances in Mathematics of Communications, 3(1) 13–27, 2009. `http://dx.doi.org/10.3934/amc.2009.3.13`