Order-$(t - 1)$ Balanced Multiple-Valued Filing Scheme Using Difference Vectors*

Lois Wright Hawkes

Department of Mathematics and Computer Science, Florida State University, Tallahassee, Florida 32306

A balanced multiple-valued filing scheme of order $(t - 1)$ is constructed based on the finite projective geometry $PG(t, q)$, $(t + 1)$ not prime, $t$ odd. The attributes and attribute values are defined using a spread of flats from $PG(t, q)$ as in Yamamoto et al. (Inform. Contr. 21 (1972), 72-91) and Berman (Inform. Contr. 32 (1976), 128-138). Certain structural properties of $PG(t, q)$ are exploited, resulting in an easily constructed bucket structure. The $(t - 2)$-flats, which represent the buckets, are partitioned into orbits each of which is uniquely identifiable by its difference vector. The storage and retrieval algorithms are based on these difference vectors. There is no scanning of buckets or sub-buckets. The concepts in this paper are illustrated by an example based on $PG(5, 2)$.

1. INTRODUCTION

The concept of filing and storing information is not a new one. However, only very recently has the amount of data to be stored become so voluminous as to antequate earlier filing procedures. Now, large, computerized files of information are becoming commonplace in a rapidly increasing number of fields. New techniques to retrieve information from these computer data bases are needed.

The typical computerized filing system has the actual data record stored in comparatively slow permanent memory. The address at which the record is stored is called the accession number of the record. A set of addresses, $M$, of relatively fast memory are used to store these accession numbers. A filing scheme is defined by a file $F$ of records, a storing rule, and a retrieval rule for queries. The storing rule gives those addresses in $M$ where the accession number of the record will be stored. The retrieval rule involves obtaining the addresses in $M$ where accession numbers of records pertaining to the query are stored. Each record can be represented by a set of identifiers or attributes which have previously been determined as characteristic of the record. In general, a record can be represented as an $m$-vector $(a_{i_1}, ..., a_{i_m})$, where $a_{i_j}$

* This work was supported in part by the National Science Foundation under Grant MCS7906664.

0019-9958/81/070034-24$02.00/0
Copyright © 1981 by Academic Press, Inc.
All rights of reproduction in any form reserved.
refers to the $j_i$th value of the $i$th attribute $A_i$, $j_i = 0, 1, ..., (n_i - 1)$, $i = 1, 2, ..., m$, and each record has $m$ attributes. Those records which have portions of their identifiers in common will have their accession numbers allocated a common storage area called a bucket. Subsets within a bucket which further partition the accession numbers using more restrictive sets of common factors, are referred to as sub-buckets.

A filing system is of order $k$ if those records having any $k$ values of $k$ different attributes appear in one and only one bucket.

A query $Q$ to the system is defined as a subset of attribute values. Those records with attribute values matching those of the query will be retrieved.

Some of the first retrieval schemes, indeed a good number of those now in use, are based on matching operations. The simplest retrieval method, first order inverted filing, has one bucket corresponding to each attribute value. A record is retrieved by scanning the buckets for the appropriate value. When a query involves two attributes, the retrieval process involves obtaining all records pertaining to both the first attribute and to the second, and then finding those records common to the two groups. The matching problem of a two-attribute query can be avoided by using a second order inverted filing scheme, i.e., each bucket corresponds to a pair of values of two attributes. This idea can be extended to queries involving more than two attributes, however, the number of buckets associated with the scheme can become unwieldy. Moreover, often redundancy, or the storing of accession numbers in more than one bucket or sub-bucket, will increase as the order $k$ increases.

In an attempt to overcome some of the aforementioned problems, several researchers, notably Abraham et al. (1968), Bose and Koch (1969), Ghosh (1969), Ghosh and Abraham (1968), Ray-Chaudhuri (1968), Yamamoto et al. (1972), Lum (1970) and Berman (1976), have applied combinatorial mathematics to the problem of information retrieval. Representative of this approach is the balanced filing scheme of Abraham et al. (1968) for binary-valued attributes. The more general balanced multiple-valued filing scheme of order $k$ ($BMFS_k$) was defined by Ghosh and Abraham (1968). It has parameters $(k, n_0, n_1, ..., n_{m-1}, b)$, where $k$ denotes the number of attributes in a query and $n_i$, $i = 0, 1, ..., m - 1$, is the number of values the $i$th attribute can take on; each record has $m$ attributes, and $b$ is the number of buckets, not necessarily mutually exclusive. The scheme satisfies:

(i) the number of records in a bucket is less than or equal to the total number of records in the whole file;

(ii) records associated with $k$ given values of $k$ different attributes appear in one and only one bucket;

(iii) to every bucket there is an algebraic method of identification.

Yamamoto et al. (1972) (YTF) developed a new type of $BMFS_2$ by
deleting all lines lying on a cyclically generated spread of a finite projective geometry, \( PG(t, q) \) for \( (t + 1) \) not prime. The generation method follows from the earlier work of Yamamoto \textit{et al.} (1966) (YFH), where a spread of \( r \)-flats was generated from an initial \( r \)-flat. Consequently, this BMFS\(_2\) is easily implemented.

Berman (1976) constructed a BMFS\(_2\) with the bucket structure of YTF but with storage and retrieval algorithms based on a difference set representation. This allows the buckets to be identified without scanning the whole list of buckets as in YTF.

In this paper we present an order \( (t - 1) \) filing scheme based on the finite projective geometry \( PG(t, q) \), \( (t + 1) \) not prime, \( t \) odd. The definition of attributes and values are those of YTF and Berman. Certain structural properties of \( PG(t, q) \) are exploited resulting in an easily implemented BMFS\(_{t-2}\). The \( (t - 2) \)-flats, which represent the buckets, are partitioned into orbits, each of which is uniquely identifiable by its difference vector. The storage and retrieval algorithms are based on these difference vectors.

An example using \( PG(5, 2) \) illustrates the concepts presented in this paper.

2. Finite Projective Geometries

Both Rao (1944) and Yamamoto \textit{et al.} (1966) studied the cyclic property of points and flats in a finite projective geometry, \( PG(t, q) \), of dimension \( t \), based on the Galois Field \( GF(q) \), \( q \) a power of a prime. We shall elaborate on their results and present the theory necessary to develop our bucket structure.

If \( (t + 1) \) is a multiple of \( (i + 1) \), then so also is \( (q^{t+1} - 1) \) a multiple of \( (q^{i+1} - 1) \), \( q \) a power of a prime. Then, for \( a \) a primitive element of the Galois Field \( GF(q^{i+1} - 1) \), \( (a^w)^u = 1 \) for \( w = (q^{i+1} - 1)/(q^{i+1} - 1) \) and \( u = (q^{i+1} - 1) \) the least integer such that \( (a^w)^u = 1 \). Thus \( a^w \) is a primitive element of \( GF(q^{i+1}) \) and can be used to give the following representation of \( GF(q^{i+1}) \):

\[
GF(q^{i+1}) = (0, a^0, a^w, \ldots, a^{(q^{i+1} - 2)w}).
\]

A finite projective geometry, \( PG(t, q) \) of dimension \( t \) is the set of points such that:

(i) a point \( \alpha \) in \( PG(t, q) \) is a non-zero element of \( GF(q^{t+1}) \);

(ii) two points \( \alpha \) and \( \beta \) represent the same point if and only if there is a non-zero element \( a \) of \( GF(q) \) such that \( \alpha = a\beta \);

(iii) a \( r \)-flat in \( PG(t, q) \) is a set of points spanned by \( (r + 1) \) linearly independent points \( \alpha_0, \alpha_1, \ldots, \alpha_r \) over \( GF(q) \).
The projective geometry $PG(i, q)$ corresponding to $GF(q^{i+1})$ is:

$$PG(i, q) = \{(a^0), (a^w), \ldots, (a^{(q^{i+1}-1)/(q-1)w})\} \tag{2.2}$$

or equivalently,

$$\{0, w, 2w, \ldots, ((q^{i+1}-1)/(q-1))w\}. \tag{2.3}$$

In particular,

$$GF(q) = (0, a^0, a^w, \ldots, a^{(q-2)w}) \tag{2.4}$$

and,

$$PG(t, q) = \{(0), (a), \ldots, (a^{v-1})\} \tag{2.5}$$

or

$$PG(t, q) = \{0, 1, \ldots, (v-1)\}, \tag{2.6}$$

where $v = (q^{t+1}-1)/(q-1)$ is the number of points in $PG(t, q)$. We shall use interchangeably $a^w$ and $w$ as representatives of a point; this form is termed the multiplicative representation.

The Galois Field, $GF(q^{t+1})$ can also be represented as residue classes of polynomials over $GF(q)$ modulo $f(x)$, an irreducible polynomial of degree $(t+1)$ over $GF(q)$. This polynomial is chosen as the minimal function of the primitive element $a$, above. This gives the additive form of the elements of $GF(q^{t+1})$, i.e.,

$$GF(q^{t+1}) = \{a_0 + a_1x + \cdots + a_tx^t; a_i \in GF(q), i = 0, 1, \ldots, t\} \tag{2.7}$$

$$= \{a_0, a_1, \ldots, a_t; a_i \in GF(q), i = 0, 1, \ldots, t\}.$$

If $V_r(0)$ denotes a $r$-flat in $PG(t, q)$ passing through $(r+1)$ linearly independent points $(a^0, b^0, \ldots, a^r, b^r)$, then $V_r(0)$ is the set of points:

$$V_r(0) = \{(a_0a^0 + a_1a^{b_0} + \cdots + a_ra^{b_r})\} \tag{2.8}$$

and the $r$-flat $V_r(c)$ is

$$V_r(c) = \{(a_0a^{b_0+c} + a_1a^{b_1+c} + \cdots + a_ra^{b_r+c})\}, \tag{2.9}$$

the $a_i$ from $GF(q)$. For some positive integer $c$, $V_r(c) = V_r(0)$. This integer is called a cycle of the initial $r$-flat $V_r(0)$, (Rao). The least value of the cycles of any $r$-flat $V_r(0)$ is called the minimum cycle (YFH).

A spread is a set of flats in $PG(t, q)$ such that each point of the geometry appears in one and only one member of the set.
The number of the $r$-flats in $PG(t, q)$ is:

$$\phi(t, r, q) = \frac{(q^{m+1}-1) \cdots (q^{m-r+1}-1)}{(q^{r+1}-1) \cdots (q-1)}.$$  \hspace{1cm} (2.10)

If $(t+1)$ and $(r+1)$ are relatively prime, then all the $r$-flats of $PG(t, q)$ have minimum cycle $\nu$ and can be generated from the $\nu = \phi(t, r, q)/\nu$ initial $r$-flats, $\nu = \phi(t, o, q)$. If the highest common factor (HCF) of $(t+1)$ and $(r+1)$ is

$$p_1^{\alpha_1}p_2^{\alpha_2} \cdots p_f^{\alpha_f} > 1$$

with the $p_i$ primes such that $p_i < p_{i+1}$, $i = 1, 2, \ldots, f-1$, then the number of distinct minimum cycles is:

$$\prod_{i=1}^{f} (1 + \beta_i).$$  \hspace{1cm} (2.11)

Let

$$\theta(x_1, \ldots, x_f) = (q^{t+1}-1)/(q^{p_1^{\alpha_1} \cdots p_f^{\alpha_f}} - 1)$$

$$t(x_1, \ldots, x_f) = ((t+1)/(p_1^{\alpha_1} \cdots p_f^{\alpha_f})) - 1$$

$$r(x_1, \ldots, x_f) = ((r+1)/(p_1^{\alpha_1} \cdots p_f^{\alpha_f})) - 1$$

$$q(x_1, \ldots, x_f) = q^{p_1^{\alpha_1} \cdots p_f^{\alpha_f}}.$$  

Then the number of $r(p_1^{\alpha_1} \cdots p_f^{\alpha_f})$-flats with the cycle $\theta(x_1, \ldots, x_f)$ and minimum cycle $\theta(x_1, \ldots, x_f)$ are, respectively:

$$n(x_1, \ldots, x_f) = \phi(t(x_1, \ldots, x_f), r(x_1, \ldots, x_f), q(x_1, \ldots, x_f)),$$

$$n^*(\beta_1, \ldots, \beta_f) = n(\beta_1, \ldots, \beta_f),$$

$$n^*(x_1, \ldots, x_f) = n(x_1, \ldots, x_f) - \sum n^*(y_1, \ldots, y_f),$$

the summation over those $y_i$ such that $x_i \leq y_i \leq \beta_i$, and there is a $j$ such that $x_j < y_j$. The number of initial $r$-flats of any minimum cycle $\theta(x_1, \ldots, x_f)$ is

$$\nu(x_1, \ldots, x_f) = n^*(x_1, \ldots, x_f)/\theta(x_1, \ldots, x_f),$$  \hspace{1cm} (2.12)

from which the totality of $r$-flats with minimum cycle $\theta(x_1, \ldots, x_f)$ can be generated (YFH).

The filing scheme to be defined uses the 1-flats and the $(t-2)$-flats of $PG(t, q)$, $t$ odd.

**Lemma 1.** Both the 1-flats and the $(t-2)$-flats of $PG(t, q)$, $t$ odd, have
the two minimum cycles \( \theta(1) = \frac{(q^{t+1} - 1)}{(q^2 - 1)} \) and \( \theta(0) = \frac{(q^{t+1} - 1)}{(q - 1)} \).

Proof. The HCF of \( (t + 1) \) and \( (t - 1) \) is 2; the HCF of \( (t + 1) \) and \( (1 + 1) \) is also 2 since \( t \) is odd. The rest follows by applying the above theory of YFH.

We define \( Z \) as the transformation which takes the multiplicative point representation of an \( i \)-flat and adds 1 to each point \( \mod(q^{t+1} - 1)/(q - 1) \), i.e.,

\[
Z(a_0, a_1, \ldots, a_s) = (a_0 + 1, a_1 + 1, \ldots, a_s + 1),
\]

\( s = \phi(i, o, q) - 1, \mod(q^{t+1} - 1)/(q - 1) \).

(2.13)

For \( V_i \) an \( i \)-field we define \( Z^n \) as

\[
Z^n(V_i) = Z(Z^{n-1}(V_i)), n > 1.
\]

(2.14)

Each initial \( i \)-flat with minimal cycle \( \theta(j) \) and the \( \theta(j) - 1 \) \( i \)-flats generated from it by \( Z \) can be considered an orbit of \( i \)-flats in the partition of the \( i \)-flats induced by the group of transformations \( (Z, Z^2, \ldots, Z^{(t-1)}) \), \( e \) the identity, \( j = 1, 0; i = 1, (t - 2) \). Thus each initial \( i \)-flat is associated with a distinct orbit. The number of \( i \)-flats in an orbit is the value of the minimum cycle.

A second transformation, \( g \), which takes the point \( j \) to the point \( qj, \mod(q^{t+1} - 1)/(q - 1) \) partitions the orbits into equivalence classes.

\[
g: j \rightarrow qj \mod(q^{t+1} - 1)/(q - 1).
\]

(2.15)

This follows since the mapping \( a \rightarrow a^q \) is a collineation.

**Lemma 2.** The number of orbits of minimal cycle \( \theta(1) = \frac{(q^{t+1} - 1)}{(q^2 - 1)} \) is 1, and the number of orbits of minimal cycle \( \theta(0) = \frac{(q^{t+1} - 1)}{(q - 1)} \) is \( q(q^{t-1} - 1)/(q^2 - 1) \) for both the 1-flats and the \((t-2)\)-flats of \( PG(t, q) \), \( t \) odd.

Proof. The HCF\((t + 1, 1 + 1) = 2 \) and HCF\((t + 1, t - 1) = 2 \) since \( t \) is odd.

Thus, applying the theory of YFH, we obtain \( v(1) = 1 \) and \( v(0) = q(q^{t-1} - 1)/(q^2 - 1) \) for both the 1-flats and the \((t-2)\)-flats. The orbit of 1-flats with minimal cycle \( \theta(1) = \frac{(q^{t+1} - 1)}{(q^2 - 1)} \) is a spread of 1-flats with \( \phi(1, 0, q) = \frac{(q^2 - 1)}{(q - 1)} \) points in each 1-flat (YTF).

We label the initial \((t-2)\)-flats as \( B(j, 0) \) and each transformation of \( B(j, 0) \) by \( Z^i \) as \( B(j, i) \). These flats will correspond to the buckets of the filing scheme.
THEOREM 1. There is a one to one correspondence between the 1-flats of minimal cycle \( \theta(j) \) and the \((t-2)\)-flats of minimum cycle \( \theta(j) \) in \( \text{PG}(t,q) \), \( t \) odd, \( j = 1, 0 \). In particular, for \( \alpha \) a primitive element of \( \text{GF}(q^{t+1}) \) and \( V_1(0) = \{(a_0\alpha^0 + a_1\alpha^d)\} \) a 1-flat in \( \text{PG}(t,q) \), \( a_i \in \text{GF}(q) \), then

(i) if \( \alpha^d \) is not a primitive element of \( \text{GF}(q^{t+1}) \) for some \( i < t \), then there is a one-to-one correspondence between the 1-flat generated by the points \( (o, d) \) and the \((t-2)\)-flat generated by the \((t-1)\) points \( (o, d, 2d, 3d, \ldots, (t-2) d) \) and their orbits;

(ii) if \( \alpha^d \) is a primitive element of \( \text{GF}(q^{t+1}) \) for some \( i < t \), then there is a one-to-one correspondence between the 1-flat generated by the points \( (o, d) \) and the \((t-2)\)-flat generated by the \((i + 1)\) linearly independent points \( (o, d, 2d, \ldots, id) \) and \( ((t - 2) - i) \) other linearly independent points;

(iii) if \( d = \theta(1) \), there is a one-to-one correspondence between the 1-flat \( V_1(0) \) and the \((t-2)\)-flat generated from a subset of the \( \theta(1) \) 1-flats \( V_1(0) \), \( V_1(1), \ldots, V_1(\theta(1) - 1) \).

Proof. Lemma 2 established that the number of 1-flats of minimum cycle \( \theta(j) \) is the same as the number of \((t-2)\)-flats of minimum cycle \( \theta(j) \) in \( \text{PG}(t,q) \), \( t \) odd, \( j = 1, 0 \). We must now show the one to one mapping. Let \( V_1(0) = \{(a_0\alpha^0 + a_1\alpha^d)\}, a_i \in \text{GF}(q), \alpha \) a primitive element of \( \text{GF}(q^{t+1}) \), be an 1-flat in \( \text{PG}(t,q) \). Then, either

(i) \( \alpha^d \) is not a primitive element of \( \text{GF}(q^{t+1}) \), for some \( i < t \) or

(ii) \( \alpha^d \) is a primitive element of \( \text{GF}(q^{t+1}) \), for some \( i < t \), \( d \neq \theta(1) \) or

(iii) \( d = \theta(1) \).

Case (i). Since \( \alpha^d \) is not a primitive element of \( \text{GF}(q^{t+1}) \) for some \( i < t \), \( \alpha^d \) has a minimal polynomial of degree \( (t+1) \). Thus, \( (\alpha^0, \alpha^d, \alpha^{2d}, \ldots, \alpha^{(t-2)d}) \) are linearly independent and

\[ V_{t-2}(0) = \{(a_0\alpha^0 + a_1\alpha^d + \cdots + a_{t-2}\alpha^{(t-2)d})\} \]

is a \((t-2)\)-flat generating the orbit \( j \). Let \( \Lambda_j \) be the set of all possible differences between pairs of points in \( V_{t-2}(0) \), i.e., \( d \in \Lambda_j \) if there are points \( \alpha^a, \alpha^b \) in \( V_{t-2}(0) \) such that \( (a-b) = d \mod(q^{t+1} - 1)/(q-1) \). Since \( V_{t-2}(0) \) can be generated by the \((t-1)\) points of the form \( id, i = 0, 1, \ldots, t-2 \), the value \( d \) must be in \( \Lambda_j \) more than once. In particular, we obtain \( d \) in \( \Lambda_j \) by selecting two points \( \alpha^a \) and \( \alpha^b \) from \( V_{t-2}(0) \) such that \( \alpha^a = \alpha^d \alpha^b \), i.e., \( \alpha^b = a_0\alpha^0 + a_1\alpha^d + \cdots + a_{t-3}\alpha^{(t-3)d} \), \( a_i \in \text{GF}(q) \) and not all the \( a_i \) are 0. The number of times \( d \) in \( \Lambda_j \) is equivalent to the number of ways \( \alpha^b \) can be chosen, i.e., \( (q^{t+2} - 1)/(q-1) \) ways. (Note: those points containing \( \alpha^{(t-2)d} \) are considered in the points represented by \( \alpha^a \).)

Now \( V_{t-2}(i), i = 1, 2, \ldots, \theta(0) \), has the same difference set \( \Lambda_j \) as \( V_{t-2}(0) \) since \( V_{t-2}(i) \) simply adds \( i \) to each point in \( V_{t-2}(0) \) and hence must maintain the differences. Assume now that \( d \) occurs \( (q^{t-2} - 1)/(q-1) \) times
in some \((t-2)\)-flat \(V_{t-2}(s)\) in orbit \(j', j \neq j'\). Then there are points \(\alpha^e\) and \(\alpha^f\) in \(V_{t-2}(s)\) such that \(e - f = d\). Moreover, there must also be points \(\alpha^{e + 2d}, \alpha^{e + 3d}, \ldots, \alpha^{e + (t-2)d}\), i.e., \(V_{t-2}(s) = \{(c_0 \alpha^e + c_1 \alpha^{e + d} + \cdots + c_{t-2} \alpha^{e + (t-2)d)}\}\) and hence, \(V_{t-2}(s - f) = \{(c_0 \alpha^0 + \cdots + c_{t-2} \alpha^{(t-2)d)}\}\), \(c_i \in GF(q)\). This implies \(V_{t-2}(0) = V_{t-2}(s - f)\), a contradiction since then the two orbits would be the same. Thus there is a one to one correspondence between the 1-flat generated by the points \((o, d)\) and the \((t-2)\)-flat generated by the \((t-1)\) points \((o, d, 2d, 3d, \ldots, (t-2)d)\) and between their orbits.

**Case (ii).** The points \((o, d, 2d, 3d, \ldots, i \alpha^d)\) are linearly independent since \(\alpha^d\) is a primitive element of \(GF(q^{i+1})\). By adding \(((t-2) - i)\) other points such that the \((t-1)\) points are linearly independent, we generate a \((t-2)\) flat which, by similar arguments to case (i), corresponds to the 1-flat generated by \((o, d)\).

**Case (iii).** From YFH, the \((t-2)\)-flat corresponding to the 1-flat generated by \(V_{t-2}(0) = \{(a_0 + a_1 \alpha^d), d = \theta(1)\}\), is the \((t-2)\)-flat \(V_{t-2}(0)\) composed of \((q^{t-1} - 1)/(q^2 - 1)\) 1-flats, each of which belongs to the set of \(\theta(1)\) 1-flats \(V_1(0), V_1(1), \ldots, V_1(\theta(1) - 1)\) generated from the 1-flat \(V_1(0)\) of minimum cycle \(\theta(1)\).

**Corollary.** Let \(V_{t-2}(0) = (a_0, a_1, \ldots, a_s), s = \phi(t-2, o, q) - 1, a_i < a_{i+1}, i = 0, 1, \ldots, s - 1\) be the initial \((t-2)\) flat in orbit \(j\). Let the vector \(D_j \subset A_j\) be the set of \((s+1)\) differences \(d_i = a_{i+1} - a_i, \mod(q^{t+1} - 1)/(q - 1), i = 0, 1, \ldots, s; s + 1 \equiv 0\). Then the elements of the difference vector for each \((t-2)\)-flat \(V_{t-2}(i), i = 0, 1, \ldots, \theta(k)\), in \(j\) are equal to those in \(D_j, k = 1, 0; 1 \leq j \leq \phi(1)\) for \(k = 1, j = \phi(1) + 1\) for \(k = 0\).

**Proof.** The set of all possible differences, \(A_j\) is the same for any \((t-2)\)-flat in orbit \(j\). Thus, the pair-wise ascending differences of \(D_j\) must also be the same for each \((t-2)\)-flat in orbit \(j\).

Each flat \(V_{t-2}(i)\) in orbit \(j\) has its points listed in ascending order of magnitude. Thus the lowest magnitude point in \(V_{t-2}(i)\) may not be \((a_0 + i)\) for \(a_0\) the lowest magnitude point in \(V_{t-2}(0)\). Consequently, \(D'\), the difference vector of \(V_{t-2}(i)\) may be a cyclically shifted version of \(D_j\). The storage and retrieval algorithms compensate for this by either shifting the flat and \(D'\) until \(D'\) and \(D_j\) coincide or, if each \(D_j\) has an uniquely identifiable point \(d_b\), by locating the corresponding point \(v_b\) in \(D'\). Then the difference \((b^* - b)\) allows the point position of \((a_0 + i)\) in \(D'\) to be located.

**Example.** The concepts presented in this paper will be illustrated using \(PG(5, 2)\). Let \(\alpha\) be a primitive element of \(GF(2^6)\) with minimal polynomial \(x^6 = x + 1\) over \(GF(2)\). The elements of \(GF(2^6)\), i.e., the points of \(PG(5, 2)\) can be represented by powers of \(\alpha\) (multiplicative form (2.6)) or as coef-
coefficients of a degree 5 polynomial (additive form (2.7)). Both are listed in Table I, which is used to obtain linear combinations of points when forming flats.

The algorithm given in Appendix A is used to generate the 1-flats and \((t - 2)\)-flats. It is based on the transformation \(g (2.15)\), the transformation \(Z (2.13)\), and Theorem 1. The advantage of this method is that most of the flats are generated by integer addition (obtaining other flats in the orbit) or integer multiplication (obtaining other initial flats in the class) rather than using Table I to determine the points in a flat.

Following the algorithm, we determine that \(GF(2^6)\) has 2 sub-fields, \(GF(2^3)\) and \(GF(2^2)\), i.e., \(i = 2\) and \(i = 1\). For \(GF(2^3)\), \(L = \{1, 3\}\). We select \(d_0 = 1\) and obtain:

\[
(1, 1): (0, 1, 5). \tag{2.16}
\]

The subscript denotes the dimension of the flat; we list the points in

<table>
<thead>
<tr>
<th>Table I</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 = (100000)</td>
</tr>
<tr>
<td>1 = (010000)</td>
</tr>
<tr>
<td>2 = (001000)</td>
</tr>
<tr>
<td>3 = (000100)</td>
</tr>
<tr>
<td>4 = (000010)</td>
</tr>
<tr>
<td>5 = (000001)</td>
</tr>
<tr>
<td>6 = (110000)</td>
</tr>
<tr>
<td>7 = (011000)</td>
</tr>
<tr>
<td>8 = (001100)</td>
</tr>
<tr>
<td>9 = (000110)</td>
</tr>
<tr>
<td>10 = (000011)</td>
</tr>
<tr>
<td>11 = (110001)</td>
</tr>
<tr>
<td>12 = (101000)</td>
</tr>
<tr>
<td>13 = (010100)</td>
</tr>
<tr>
<td>14 = (001010)</td>
</tr>
<tr>
<td>15 = (000101)</td>
</tr>
<tr>
<td>16 = (110010)</td>
</tr>
<tr>
<td>17 = (011001)</td>
</tr>
<tr>
<td>18 = (111000)</td>
</tr>
<tr>
<td>19 = (011100)</td>
</tr>
<tr>
<td>20 = (001110)</td>
</tr>
</tbody>
</table>
ascending order of magnitude. Using step (3) of the algorithm and the fact \((11) = (0, 1, 5)\), the set of all possible differences, mod 7, is \(N_1 = \{1, 2, 3, 4, 5, 6\}\). Hence \(L\) becomes empty, i.e., \(L = \emptyset\). Thus \(GF(2^3)\) has only one orbit. And, as \(GF(2^3)\) has no subfield we next return to \(GF(2^6)\) for which we have, from step (2), \(L = \{1, 3, 5, 7, 11\}\). We select \(d_0 = 1\) and take \((0, 1)\) as two independent points and form the initial 1-flat \((11)\) using Table I.

\[(11): (0, 1, 6)\]

We obtain \(N_1 = \{1, 6, 5, 62, 57, 58\}\) as above and find \(\rho = 6\), since by step (4), there is no power \(\rho < 6\) of \(q = 2\), such that \(n = 2^\rho n' \mod 63\), \(n, n' \in (11)\), i.e., \(1 \cdot 2^6 = 1 \mod 63\). The remaining initial 1-flats in class I_1 are obtained simply by multiplying \((11)\) by \(2^s\), \(s = 1, 2, \ldots, 5\) (step (5)). Similarly the sets \(N_{1+s}\) are formed.

\[
\begin{align*}
I_1 & \quad (11): (0, 1, 6) \quad (41): (0, 8, 48) \\
(21) & \quad (0, 2, 12) \quad (51): (0, 16, 33) \\
(31) & \quad (0, 4, 24) \quad (61): (0, 3, 32)
\end{align*}
\]

We remove from \(L\) those points in \(N = \bigcup_{s=0}^{5} N_{1+s}\), leaving \(L = \{7, 11\}\). As \(L\) is not empty, we return to step (3). We now select another point \(d_0 = 7\). We obtain \(N_1 = \{7, 26, 19, 56, 37, 44\}\) and, as \(7 \times 2^3 = 56, \rho = 3\). We list class II_1 with its three initial 1-flats with minimum cycle \(\theta(0) = 63\), obtained by multiplying \((71)\) by \(2^s\), \(s = 1, 2\).

\[
\begin{align*}
(71) & \quad (0, 7, 26) \\
(81) & \quad (0, 14, 52) \\
(91) & \quad (0, 28, 41)
\end{align*}
\]

We have \(N = \{7, 26, 19, 56, 37, 44, 14, 52, 38, 49, 11, 25, 28, 41, 13, 35, 22, 50\}\), so \(L = \emptyset\). Next, having exhausted \(L\), we proceed to step (6) and form \(L' = \{9, 27\}\). We take \(d_0 = 9\), which is a primitive element in \(GF(2^3)\) (step (7)). Thus, as this subfield is isomorphic to (2.16) we simply replace each point \(d'\) in (2.16) by \(9d'\), i.e.,

\[
III_1 \quad (101): (0, 9, 45)
\]

\(N = \{9, 45, 36, 54, 18, 27\}\) and hence \(L' = \emptyset\). Since (2.16) had only one orbit we did not need to form \(N\) as \(L'\) had to be empty. Finally, we have \(d = 21 = \theta(1)\), giving the one initial 1-flat of minimum cycle \(\theta(1) = 21\), using step (8).

\[
IV_1 \quad (111): (0, 21, 42)
\]

See Table II for a complete list of initial 1-flats.
TABLE II

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>(1)</td>
<td>(0, 1, 6)</td>
<td>(4)</td>
</tr>
<tr>
<td></td>
<td>(2)</td>
<td>(0, 2, 12)</td>
<td>(5)</td>
</tr>
<tr>
<td></td>
<td>(3)</td>
<td>(0, 4, 24)</td>
<td>(6)</td>
</tr>
<tr>
<td>II</td>
<td>(7)</td>
<td>(0, 7, 26)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(8)</td>
<td>(0, 14, 52)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(9)</td>
<td>(0, 28, 41)</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>(10)</td>
<td>(0, 9, 45)</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>(11)</td>
<td>(0, 21, 42)</td>
<td></td>
</tr>
</tbody>
</table>

Each of the initial 1-flats with minimum cycle \( \theta(0) = 63 \) (Lemma 1), has 62 remaining flats in the orbit with initial 1-flat \((0, a_0, a_1)\), i.e., \((a + i, a_0 + i, a_1 + i)\), \(i = 1, 2, \ldots, 62\). Similarly, the 1-flats with minimal cycle \( \theta(1) = 21 \) are formed by adding \(i\), \(i = 1, 2, \ldots, 20\).

The initial \((t - 2) = 3\)-flats are easily found due to the one to one correspondence between the 1-flats and the \((t - 2)\)-flats, described in part II of the algorithm. We label the initial 3-flats as \((j_3)\), \(j = 1, 2, \ldots, 11\), to exhibit the correspondence. We begin with class I_3 using step (9). Since the independent points of \((1_3)\) are \((0, 1)\), i.e., \(d_0 = 1\), the independent points for \((1_3)\) are \((0, d_0, 2d_0, 3d_0) = (0, 1, 2, 3)\), from which using Table I, we have:

\[(1_3) : (0, 1, 2, 3, 6, 7, 8, 12, 13, 18, 26, 27, 32, 35, 48).\]

The remaining initial 3-flats in class I_3 are obtained by multiplication by \(2^s\), mod 63, \(s = 1, 2, \ldots, 5\). Class II_3 3-flats are obtained similarly. Class III_3 corresponds to the sub-geometry on \(GF(2^3)\), so we select, using step (11), as independent points \((0, d, 2d) = (0, 9, 18)\) since \(i = 2\). We arbitrarily select 10 as a fourth point giving

\[III_3 (10_3) : (0, 9, 10, 15, 17, 18, 23, 27, 36, 45, 47, 54, 57, 58, 61).\]

The remaining class of 3-flats has minimum cycle \( \theta(1) = 21 \) and the initial 3-flat is formed from the set of 1-flats \(V_1 (i)\), \(i = 0, 1, \ldots, 20\), where

\[V_1(0) = (0, 21, 42).\]

Using step (13), we first take the points 0, 21, 42 and one other point, say 1. Then, 22 and 43 must also be in the 3-flat. Using Table I, we find all linear combinations of \((0, 21, 42, 1, 22, 43)\). Continuing in this manner we obtain \((11_3) = V_1(0) + V_1(1) + V_1(6) + V_1(8) + V_1(18)\), i.e.,

\[(11_3) : (0, 1, 6, 8, 18, 21, 22, 27, 29, 39, 42, 43, 48, 50, 60).\]
See Table III for the complete list of initial \((t - 2)\)-flats. The remaining flats in the orbit are obtained by adding \(i\) to each point, mod 63, \(i = 1, 2, \ldots, 20\). The 10 initial 3-flats of minimum cycle 63 are used to generate the remaining flats in each orbit \(j\) by adding \(i\) to each point mod 63, \(i = 1, 2, \ldots, 62; j = 1, 2, \ldots, 10\).

Each initial 3-flat will correspond to a bucket labeled \(B(j, 0)\) in the filing scheme, \(j = 1, 2, \ldots, 11\). We now have Table II and Table III. The ease with
which the 3-flats are formed is obvious with only four of the 3-flats being calculated using Table I, the remainder obtained through integer multiplication by 2 or integer addition.

For each orbit \( j \) with initial 3-flat \((j_3) = (a_0, a_1, ..., a_{14})\) there is an unique difference vector \( D_j = (d_0, d_1, ..., d_{14}) \), where \( d_i = a_{i+1} - a_i, i = 0, 1, ..., 14, a_{15} = a_0, j = 1, 2, ..., 11 \). Since \((113)\) consists of five sets of points \((0 + i, 21 + i, 42 + i), i = 0, 1, 6, 8, 18\), \( D_{11} \) has five values repeated three times. We select for each \( D_j \), the largest difference \( d_b \) (if there is a tie, select the largest ordered pair \( d_{b-1}, d_b \)) and list it and \( d_{b-1} \). These are used to simplify the identification procedure. The difference vectors and \( d_b, d_{b-1} \) appear in Table IV.

3. Design of Bucket Structure

3.1. Attributes and Values

Following YTF, we use a spread of \( r \)-flats, in \( PG(t, q) \), \( t + 1 \) not prime, \( t \) odd, to define the attributes and attribute values. The \( v \) points in \( PG(t, q) \) are partitioned into \( m = (q^{t+1} - 1)/(q^{t+1} - 1) \) sets of \( n = (q^{t+1} - 1)/(q - 1) \) points each. Then if \( V_r(0) = (0, m, 2m, ..., (n-1)m) \), we let \( V_r(i) \) correspond to attribute \( A_i \) and let the point \((i + jm), i = 1, 2, ..., m; j = 0, 1, ..., (n-1) \) in \( V_r(i) \) represent \( a_{ij} \), the \( j \)th value of the \( i \)th attribute. A record \( f \in F \) is then given as \((a_{ij_1}, a_{ij_2}, ..., a_{ij_m}), j_i \in \{0, 1, ..., n-1\}, i = 1, 2, ..., m \). For any point \( s \) we can immediately determine its attribute and value, i.e., \( i = s \mod m \) and \( j = (s - i)/m \). Using this representation every point in a record \( f \) of the file \( F \) of records lies on a different \( r \)-flat in the spread \((A_1, A_2, ..., A_m)\). An admissible query of order \( k \) to this filing scheme is \((s_1, s_2, ..., s_k)\), where each \( s_i, i = 1, 2, ..., k \), is a point of \( PG(t, q) \) and \((s_i - s_j) \neq 0 \mod m \) for any \( i, j \in \{1, 2, ..., k\} \).

Example. In \( PG(5, 2) \), we take \( r = 1 \) and \( m = \phi(1) = 21 \) and \( n = 3 \), i.e., we use the 1-flats in the orbit with minimum cycle 21. Thus, we obtain Table V.

3.2. Bucket Construction

In this filing scheme every \((t-2)\)-flat in \( PG(t, q) \) corresponds to a bucket. Any \((t-1)\) linearly independent points uniquely determine a \((t-2)\)-flat, i.e., a bucket. Moreover, if only \((t-1-i)\) of the points are linearly independent, \( i = 1, 2, ..., c \), where \( c = (t-1) - (j + 2) \) for \( j \) the largest integer such that \( \phi(j, o, q) < (t-1) \), then \( i \) points can be determined such that the set of \((t-1+i)\) points always yields an unique \((t-2)\)-flat for the original \((t-1)\) points. For example, we can choose the \( i \) largest magnitude points less than
the original \((t - 1)\) points \(\mod \phi(t, o, q)\) such that we have \((t - 1)\) linearly independent points. This always yields an unique \((t - 2)\)-flat, given the original \((t - 1)\) dependent points. Thus, the set of buckets satisfies condition (ii) of the \(\text{BMFS}_{t-1}\).

The total number of buckets \(b\) is then

\[
b = \begin{cases} 
\phi(t, r, q) & \text{if } (t - 2) \neq r \\
\phi(t, r, q) - m & \text{if } (t - 2) = r.
\end{cases}
\]

We determine the bucket for the \((t - 1)\) points \(a_0, a_1, \ldots, a_{t-2}\), \(a_i < a_j\) for \(i < j\), in the following way.

1. If \(a_0, a_1, \ldots, a_{t-2}\) are linearly independent, go to step (2). For \((t - 1 - i)\) linearly independent points, select the \(i\) greatest magnitude points \(\mod \phi(t, o, q)\) such that we obtain \((t - 1)\) independent points.

2. Generate the flat \(e = \{a_0, a_1, \ldots, a_s\}, s = \phi(t - 2, o, q) - 1\), from the \((t - 1)\) independent points and order the points such that \(a_i < a_{i+1}\), \(i = 0, 1, \ldots, s - 1\).

3. Form the difference vector \(V = (v_0, v_1, \ldots, v_s)\), such that \(v_i = a_{i+1} - a_i\), \(i = 0, 1, \ldots, s; s + 1 \equiv 0\, \mod \phi(t, o, q)\).

4. Determine the vector \(D_j\) such that the differences in \(V\) are identical with those in \(D_j\).

5. Find position \(b^*\) in \(V\) such that \(v_{b^*} = d_b, d_b \in D_j\). Then \(k = (b^* - b) \mod (s + 1)\). (If there is no \(d_b\), then shift \(e\) and \(V\) until \(V\) coincides with \(D_j\) and set \(k = 0\).) The bucket is then \(B(j, a_k)\).

Thus, our scheme satisfies condition (iii) of \(\text{BMFS}_{t-1}\).

**Example.** The set of initial 3-flats for \(PG(5, 2)\) given in Table III constitute the buckets \(B(j, o), j = 1, 2, \ldots, 11\). The remaining \((\theta(0) - 1) = 62\) 3-flats in the first 10 orbits are represented as \(B(j, i), i = 1, 2, \ldots, 62\), where if \(a\) is a point in \(B(j, o)\), then \(Z^j(a) = (a + i) \mod 63\), is the corresponding point in \(B(j, i)\). The eleventh orbit corresponds to \(B(11, 0)\) and the 20 remaining buckets are \(B(11, i), i = 1, 2, \ldots, 20\).

### 3.3. Storing on \((t - 1)\) Attribute Values

The sub-bucket labels are defined as follows. Each sub-bucket corresponds to a distinct set of \((t - 1)\) of the \((s + 1)\) points \((a_0, a_1, \ldots, a_s)\) in the \((t - 2)\)-flat \(B(j, a_k)\), \(s = \phi(t - 2, o, q) - 1\). The positions in any \((t - 2)\)-flat are labeled \(0, 1, 2, \ldots, s\). The sets of \((t - 1)\) of these positions can be listed in increasing order of magnitude with the highest order position changing most
rapidly. The sub-buckets are numbered sequentially corresponding to this list. If the \((t-1)\) points are \(a_0, a_1, \ldots, a_{t-2}\), then we can define

\[
u_i = c \quad \text{for} \quad a_i = a_t, \quad i = 0, 1, \ldots, t-2, \quad 0 \leq c \leq s, \quad (3.2)
\]
i.e., \(u_i\) indicates the position of \(a_i\) in \(B(j, a_k)\). This ordering allows the sub-bucket label \(L(u_0, \ldots, u_{t-2})\) (i.e., the position in the list) to be calculated algebraically. This numbering is illustrated below, where there are \((\frac{s}{t-1})\) sub-buckets in all, some of which may be unused (if two points \(i, j\) are such that \((i-j) = 0 \mod m\), then the corresponding sub-bucket is unused; also as each set \(U\) of linearly dependent points is stored in only one bucket, those sub-buckets corresponding to \(U\) are unused in all other buckets.):

<table>
<thead>
<tr>
<th>label (L(u_0, u_1, \ldots, u_{t-2}))</th>
<th>positions (u_0, u_1, \ldots, u_{t-2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0, 1, 2, 3, \ldots, (t-3, t-2)</td>
</tr>
<tr>
<td>2</td>
<td>0, 1, 2, 3, \ldots, (t-3, t-1)</td>
</tr>
<tr>
<td>3</td>
<td>0, 1, 2, 3, \ldots, (t-3, t)</td>
</tr>
<tr>
<td>4</td>
<td>0, 1, 2, 3, \ldots, (t-3, t+1)</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>(s-(t-3))</td>
<td>0, 1, 2, 3, \ldots, (t-3, s)</td>
</tr>
<tr>
<td>(s-(t-2))</td>
<td>0, 1, 2, 3, \ldots, (t-2, t-1)</td>
</tr>
<tr>
<td>(s-(t-1))</td>
<td>0, 1, 2, 3, \ldots, (t-2, t)</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>(\binom{s+1}{t-1})</td>
<td>(s-(t-2), s-(t-3), \ldots, s-1, s)</td>
</tr>
</tbody>
</table>

i.e.,

\[
L(u_0, u_1, \ldots, u_{t-2}) = \sum_{l_0=1}^{u_0} \sum_{l_1=1}^{w_{u_0}} \sum_{l_2=1}^{w_{u_0}} \cdots \sum_{l_{t-3}=1}^{w_{u_0}} \sum_{l_{t-2}=1}^{t} \sum_{i=1}^{w_{u_0}-j+1} \sum_{k=1}^{i} (k)
\]

\[
+ \sum_{l_1=1}^{u_1-u_0-1} \sum_{l_2=1}^{w_{u_1}} \sum_{l_3=1}^{w_{u_1}} \cdots \sum_{j=l_{t-5}}^{w_{u_1}-j+1} \sum_{i=1}^{j} \sum_{k=1}^{i} (k)
\]

\[
+ \sum_{l_2=1}^{u_2-u_1-1} \sum_{l_3=1}^{w_{u_2}} \sum_{l_4=1}^{w_{u_2}} \cdots \sum_{j=l_{t-5}}^{w_{u_2}-j+1} \sum_{i=1}^{j} \sum_{k=1}^{i} (k) + \cdots
\]

\[
+ \sum_{l_{t-4}=1}^{u_{t-4}-u_{t-5}-1} \sum_{l_{t-3}=1}^{s-u_{t-4}-1-i} \sum_{i=1}^{s-u_{t-4}-1} \sum_{k=1}^{i} (k) + \sum_{k=s-u_{t-3}+1}^{s-u_{t-4}-1} (k) + (u_{t-2}-u_{t-3}), \quad (3.3)
\]
BMFS\(_{t-1}\) USING DIFFERENCE VECTORS

### TABLE V

<table>
<thead>
<tr>
<th>(A_1)</th>
<th>(A_6)</th>
<th>(A_{15})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 22, 43)</td>
<td>(8, 29, 50)</td>
<td>(15, 36, 57)</td>
</tr>
<tr>
<td>(A_2)</td>
<td>(A_7)</td>
<td>(A_{16})</td>
</tr>
<tr>
<td>(2, 23, 44)</td>
<td>(9, 30, 51)</td>
<td>(16, 37, 58)</td>
</tr>
<tr>
<td>(A_3)</td>
<td>(A_8)</td>
<td>(A_{17})</td>
</tr>
<tr>
<td>(3, 24, 45)</td>
<td>(10, 31, 52)</td>
<td>(17, 38, 59)</td>
</tr>
<tr>
<td>(A_4)</td>
<td>(A_9)</td>
<td>(A_{18})</td>
</tr>
<tr>
<td>(4, 25, 46)</td>
<td>(11, 32, 53)</td>
<td>(18, 39, 60)</td>
</tr>
<tr>
<td>(A_5)</td>
<td>(A_{10})</td>
<td>(A_{19})</td>
</tr>
<tr>
<td>(5, 26, 47)</td>
<td>(12, 33, 54)</td>
<td>(19, 40, 61)</td>
</tr>
<tr>
<td>(A_6)</td>
<td>(A_{11})</td>
<td>(A_{20})</td>
</tr>
<tr>
<td>(6, 27, 48)</td>
<td>(13, 34, 55)</td>
<td>(20, 41, 62)</td>
</tr>
<tr>
<td>(A_7)</td>
<td>(A_{12})</td>
<td>(A_{21})</td>
</tr>
<tr>
<td>(7, 28, 49)</td>
<td>(14, 35, 56)</td>
<td>(21, 42, 0)</td>
</tr>
</tbody>
</table>

where \(w_i = s - u_{i-1} - (t - i) + 2\), \(u_{(-1)} = -1\), \(s = \phi(t - 2, 0, q) - 1\). Given \(t\), the various sums can be expressed in tabular form (see (3.4) and Tables VI and VII).

The procedure for storing the record \(f = (a_0, a_1, ..., a_{m-1})\) consists of the following:

1. For each set of \((t-1)\) points \(a_{i_0}, a_{i_1}, ..., a_{i_{t-2}}\), determine the bucket \(B(j, \alpha_k)\).
2. The sub-bucket in which the accession number of the record is stored has label \(L(u_{i_0}, u_{i_1}, ..., u_{i_{t-2}})\).

As we assume that the frequency distribution of the different types of records in the file is uniform, the filing scheme satisfies condition (i) of BMFS\(_{t-2}\).

**EXAMPLE.** Let \(f \in F\) be the record \(f = (1, 23, 45, 4, 26, 48, 7, 29, 51, 10, 32, 33, 13, 36, 16, 38, 60, 19, 41, 0)\), for \(f_i \in A_i\) in Table V. To store the 4-tuple \((1, 23, 45, 4)\) we order it in ascending order as \((1, 4, 23, 45)\). The four points are linearly independent and we generate the 3-flat

\[ e = (1, 4, 5, 10, 21, 23, 25, 32, 33, 35, 38, 45, 46, 51, 60), \]

where the 4-tuple values are underlined. We now form the difference vector \(V = (3, 1, 5, 11, 2, 2, 7, 1, 2, 3, 7, 1, 5, 9, 4)\). By comparing the largest magnitude pair of elements in \(V\) to the corresponding values in Table IV, \((d_b, d_{b-1})\) we find \(j = 9\). (Since the largest magnitude pair is distinct in each vector, we need only compare these rather than the whole vector.) The largest value in \(V\) is in position \(b^* = 3\); in \(D_9\), the corresponding position is \(b = 2\). Thus \(k = (b^* - b) \mod 15 = 1\), and \(B(f, \alpha_k) = B(9, \alpha_1) = B(9, 4)\), i.e., the 3-flat \(e\) is the initial 3-flat \(B(9, 0)\) with 4 added to each point, \(\mod 63\).

To determine the sub-bucket we require the formula \(L(u_0, u_1, u_2, u_3)\) where the four possible positions are selected from \((0, 1, 2, 3, ..., 14)\), i.e.,
Simplifying (3.3), we have

\[ L(u_0, u_1, u_2, u_3) = f_1(u_0) + f_2(u_1, u_0) + (u_2 - u_1 - 1)(13 - u_1) \]
\[ - \frac{(u_2 - u_1 - 2)(u_2 - u_1 - 1)}{2} + (u_3 - u_2), \quad (3.4) \]

where

\[ f_1(u_0) = \sum_{j=1}^{u_0} \sum_{i=1}^{13-j} \sum_{k=1}^{l} (k) \]

and

\[ f_2(u_1, u_0) = \sum_{j=1}^{u_1-u_0-1} \sum_{i=1}^{13-u_0-j} (i), \quad u_1 > u_0. \]

The values of \( f_1(u_0) \) and \( f_2(u_1, u_0) \) are calculated once and stored. They are given in Tables VI and VII.

We have a total of \( \binom{15}{4} = 1365 \) sub-buckets in each bucket; those which have points \( i, j \) such that \( (i - j) = 0 \) (mod 21) are unused.

Returning to the example, the position vector for the query \( (1, 4, 23, 45) \) is \( (0, 1, 5, 11) \) and

\[ L(0, 1, 5, 11) = f_1(0) + f_2(1, 0) + (3)(12) - \frac{(2)(3)}{2} + 6 \]
\[ = 0 + 0 + 36 - 3 + 6 = 39. \]
BMFS\textsubscript{t−1} USING DIFFERENCE VECTORS

<table>
<thead>
<tr>
<th>Table VI</th>
</tr>
</thead>
<tbody>
<tr>
<td>(u_0)</td>
</tr>
<tr>
<td>(0)</td>
</tr>
<tr>
<td>(1)</td>
</tr>
<tr>
<td>(2)</td>
</tr>
<tr>
<td>(3)</td>
</tr>
</tbody>
</table>

Thus the record with 4-tuple \((1, 4, 23, 45)\) has its accession number stored in sub-bucket 39 of bucket \(B(9, 4)\).

To store the 4-tuple \((1, 4, 7, 36)\) we first note that the points \((4, 7, 36)\) are linearly dependent. Thus, we choose the greatest magnitude point less than 1, i.e. 0 and form the 3-flat

\[ e = (0, 1, 2, 4, 6, 7, 12, 14, 16, 24, 26, 33, 36, 52, 54) \]

based on the independent points \((0, 1, 4, 7)\); the original four points are underlined. The difference vector is

\[ V = (1, 1, 2, 2, 1, 5, 2, 8, 2, 7, 3, 16, 2, 9) \]

Referring to Table IV, \(j = 2\). Moreover, we have \(b^* = 12, b = 12\) and thus, \(k = (b^* - b) = 0\) and \(a_0 = 0\). Hence the bucket is \(B(2, a_0) = B(2, 0)\). The sub-bucket label is calculated for the original 4-tuple \((1, 4, 7, 36)\); the position vector is \((1, 3, 5, 12)\) and

\[ L(1, 3, 5, 12) = f_1(1) + f_3(3, 1) + (1)(10) - 0 + 7 = 364 + 66 + 10 + 7 = 447. \]

Thus the accession number is stored in sub-bucket 447 of bucket \(B(2, 0)\).

3.4. Retrieval on \((t−1)\) Query

To retrieve for a \((t−1)\)-query \((a_0, a_1, ..., a_{t−1})\);

1. Determine the bucket \(B(j, a_k)\) as for the storage algorithm.
2. Calculate the sub-bucket label \(L(u_0, u_1, ..., u_{t−1})\).
3. All accession numbers stored in sub-bucket \(L(u_0, u_0, ..., u_{t−1})\) of bucket \(B(j, a_k)\) are retrieved.

Example. Given the query \((29, 51, 16, 19)\) we first calculate the 3-flat

\[ e = (13, 14, 16, 17, 19, 22, 26, 29, 37, 45, 46, 48, 51, 58, 62) \]
\begin{table}
\centering
\begin{tabular}{cccccccccc}
\hline
$(u_1, u_0)$ & $f_2(u_1, u_0)$ & $(u_1, u_0)$ & $f_2(u_1, u_0)$ & $(u_1, u_0)$ & $f_2(u_1, u_0)$ & $(u_1, u_0)$ & $f_2(u_1, u_0)$ & $(u_1, u_0)$ & $f_2(u_1, u_0)$ \\
\hline
$(1, 0)$ & 0 & (6, 1) & 202 & (12, 2) & 219 & (11, 4) & 116 & (9, 7) & 15 \\
$(2, 0)$ & 78 & (7, 1) & 230 & (4, 3) & 0 & (12, 4) & 119 & (10, 7) & 25 \\
$(3, 0)$ & 144 & (8, 1) & 251 & (5, 3) & 45 & (6, 5) & 0 & (11, 7) & 31 \\
$(4, 0)$ & 199 & (9, 1) & 266 & (6, 3) & 81 & (7, 5) & 28 & (12, 7) & 34 \\
$(5, 0)$ & 244 & (10, 1) & 276 & (7, 3) & 109 & (8, 5) & 49 & (9, 8) & 0 \\
$(6, 0)$ & 280 & (11, 1) & 282 & (8, 3) & 130 & (9, 5) & 64 & (10, 8) & 10 \\
$(7, 0)$ & 308 & (12, 1) & 285 & (9, 3) & 145 & (10, 5) & 74 & (11, 8) & 16 \\
$(8, 0)$ & 329 & (3, 2) & 0 & (10, 3) & 155 & (11, 5) & 80 & (12, 8) & 19 \\
$(9, 0)$ & 344 & (4, 2) & 55 & (11, 3) & 161 & (12, 5) & 83 & (10, 9) & 0 \\
$(10, 0)$ & 354 & (5, 2) & 100 & (12, 3) & 164 & (7, 6) & 0 & (11, 9) & 6 \\
$(11, 0)$ & 360 & (6, 2) & 136 & (5, 4) & 0 & (8, 6) & 21 & (12, 9) & 9 \\
$(12, 0)$ & 363 & (7, 2) & 164 & (6, 4) & 36 & (9, 6) & 36 & (11, 10) & 0 \\
$(2, 1)$ & 0 & (8, 2) & 185 & (7, 4) & 64 & (10, 6) & 46 & (12, 10) & 3 \\
$(3, 1)$ & 66 & (9, 2) & 200 & (8, 4) & 85 & (11, 6) & 52 & (12, 11) & 0 \\
$(4, 1)$ & 121 & (10, 2) & 210 & (9, 4) & 100 & (12, 6) & 55 & & \\
$(5, 1)$ & 166 & (11, 2) & 216 & (10, 4) & 110 & (8, 7) & 0 & & \\
\hline
\end{tabular}
\end{table}
from which the difference vector \( V = (1, 2, 1, 2, 3, 4, 3, 8, 8, 1, 2, 3, 7, 4, 14) \) is calculated. The largest value in \( V \) is \( v_{14} = 14 \). Thus, referring to Table IV, we identify the orbit as \( j = 6 \). We have \( b^* = 14 \) and \( b = 14 \), so \( k = (b^* - b) = 0 \). Hence the bucket is \( B(6, a_0) = B(6, 13) \). To obtain the sub-bucket label, we first determine the position vector \((2, 4, 7, 12)\). The sub-bucket label is then

\[
L(2, 4, 7, 12) = f_1(2) + f_2(4, 2) + (2)(9) - \frac{(1)(2)}{2} + 5
\]

\[
= 650 + 55 + 18 - 1 + 5 = 727.
\]

We then retrieve all accession numbers in sub-bucket 727 of bucket \( B(6, 13) \).

3.5. Storage and Retrieval on \( k \) Attribute Values, \( k < (t - 1) \)

By the addition of extra sub-buckets to the buckets, the filing scheme can be used for queries of order \( k, k < (t - 1) \). The definition of attributes and attribute values is unchanged.

Due to the one to one correspondence between the 1-flats and the \((t - 2)\)-flats any queries on two points can be stored using Berman's method for 2-queries. For any two points \( a_0, a_1 \) such that \( e = a_0 - a_1 \neq 0 \) (mod \( m \)), \( a_0 \) and \( a_1 \) are in the unique bucket \( B(\lambda(e), a_0 - \mu(e)) \), where the values of the functions \( \lambda \) and \( \mu \) are given in a table. Berman's buckets are in one to one correspondence with the \((t - 2)\)-flat buckets defined above. Sub-bucket labels \( L(u_0, u_1) \) are added to the buckets to accommodate the 2-queries. Similarly, the singleton query \( a_0 \) is stored, as in Berman, in bucket \( B(1, a_0) \) and an additional sub-bucket with label \( L(u_0) \) is added to each bucket in the first orbit.

We can accommodate queries of order \(-k, 2 < k < (t - 1)\) in two different ways. In the first method, we maintain the bucket structure defined above and treat the \( k \)-uple as a set of \(< (t - 1)\) independent points. We augment this set to \((t - 1)\) linearly independent points and store in (retrieve from) the corresponding bucket. Additional sub-buckets must be added to differentiate between \( k \)-value and \((t - 1)\)-value sub-buckets. The redundancy problem can be mitigated by chaining.

Alternately, we can apply YFH's theory to the \( k \)-flats and establish tables of differences for the \( k \)-orbits. This adds more buckets and sub-buckets but makes for a neater and more efficient retrieval scheme which handles all \( k \)-queries, \( 1 \leq k \leq (t - 1) \). Isomorphism between \( k \)-orbits and \( k' \)-orbits, (as for the 1-orbits and \((t - 2)\)-orbits) reduces the number of buckets.
4. Conclusions

The major advantages of this filing system are the ease with which the buckets are constructed, and the simple calculation required to determine the sub-bucket associated with a query. The disadvantage is the redundancy involved in storing each record in \((m-1)\) sub-buckets. This can be alleviated somewhat through chaining. Also, for many large data bases, a substantial number of records have fewer than \(m\) attributes, say \(m' < m\). Thus the accession number is only stored in \((m'-1) < (m-1)\) sub-buckets since the filing system does not depend on the parameter \(m\).

Redundancy could be reduced by storing records using Berman’s sub-bucket structure, however, then, both the number of sub-buckets per bucket \((2^{s+1}-1)\), and the number of sub-buckets checked per each retrieval, \((2^{s-t+2}-1)\), would become prohibitive. For example, in the scheme discussed above, these values would be \((2^{15}-1)\) and \((2^{11}-1)\), respectively.

APPENDIX A: Algorithm to Generate 1-flats and \((t-2)\)-flats of \(PG(t, q), t \text{ odd}, \text{partitioned into orbits and classes}\)

Note: All calculation \(\mod(q^{t+1}-1)/(q-1)\)

I. 1-flats, Orbits, Classes

1. Find subfields: Determine if \(GF(q^{t+1})\) has any subfields, i.e., if there is an integer \(i\) such that \((i + 1)|(t + 1)\). (The element \(d = (q^{t+1}-1)/(q^{i+1}-1)\) corresponds to a primitive element of the subfield \(GF(q^{i+1})\)). If so, use the algorithm to determine the orbits and classes of \(GF(q^{i+1})\), for all \(i \neq 1\), where calculations are \(\mod(q^{t+1}-1)/(q-1)\).

2. Roots of degree \((t + 1)\) minimal polynomial: Form a list \(L\) consisting of the least magnitude root of each minimal polynomial of degree \((t + 1)\) associated with the field (see tables in Appendix C of Peterson and Weldon (1972)). Set \(j \leftarrow 1; k \leftarrow 1\).

3. Form initial 1-flat and difference sets \(N_j\): Select an element \(d_0\) from \(L\). Form the initial 1-flat \(j_1 = \{(a_0a_0^0 + a_1a_1^0)\}, a_0, a_1 \in GF(q)\). Form the set of non-zero differences \(N_j\) such that \(n \in N_j\) if \(n = s - s', a_s, a_{s'} \in j_1, s \neq s'\).

4. Number of class members: Determine the least integer \(\rho\), such that \(n = q^\rho n' \mod(q^{t+1}-1)/(q-1), 0 < \rho < (t + 1), n, n' \in N_j\). If there is no such \(\rho\), set \(\rho = (t + 1)\). Class \(k\) has \(\rho\) members.

5. Generate class members: Generate the remaining \((\rho - 1)\) initial 1-flats in class \(k\) by multiplying the points of \(j_1\) by \(q^s\) to give the initial 1-flat \((j + s)_1, s = 1, 2, ..., (\rho - 1)\). Form the corresponding sets \(N_{j+s}\) similarly, i.e.,
multiply \( N_j \) by \( q^s \) to give \( N_{j+s} \). Set \( j \leftarrow j + \rho; k \leftarrow k + 1 \); remove from \( L \) any points that occur in \( N = \bigcup_{s=0}^{p-1} N_{j+s} \). If \( L \) is not empty, go to (3).

(6) Roots of degree \((t + 1)\) minimal polynomial: Form the list \( L' \) consisting of the least magnitude root of each minimal polynomial of degree \((t + 1)\) associated with the field and such that no element is equal to \( \theta(1) \).

(7) Isomorphic subfields: If \( L' \) is empty, go to (8). Otherwise select an element \( d_0 \) from \( L' \) (this corresponds to a primitive element in the sub-field \( GF(q^{i+1}) \)). The structure of these initial 1-flats and orbits is isomorphic to that of \( GF(q^{i+1}) \) where each point \( d' \) of the subfield is replaced by \( (d_0)(d') \mod(q^{i+1}-1)/(q-1) \). The initial 1-flats are labeled \((j + w)\) for \( w \) the label of the subfield initial 1-flat. Form the sets \( N_j \) and \( N = \bigcup_{s=0}^{p-1} N_{j+s} \). The classes are labeled \((k + x)\) for \( x \) the label of the corresponding subfield class. Set \( j \leftarrow j + w*; k \leftarrow k + x* \) for \( w*, x* \) the number of initial 1-flats and classes in the subfield, respectively. Remove those elements from \( L' \) which occur in the set \( N \). Repeat step (7).

(8) \( \theta(1) \) initial 1-flat: We now have \( v(0) = q(q^{t-1}-1)/(q^2-1) \) initial 1-flats divided into classes. For minimal cycle \( \theta(1) \) here is one initial 1-flat given as \((0, \theta(1), 2\theta(1), ..., q\theta(1))\). Label this as initial 1-flat \( v(0) + 1 \) and class \( k + 1 \).

Remaining 1-flats: The remaining 1-flats in each orbit are obtained by adding 1 to each point \( \theta(0) \) times for the first \( v(0) \) initial 1-flats and \( \theta(1) \) times for the last initial 1-flat. This completes the generation of the 1-flats, orbits and classes. Set \( j \leftarrow 1; k \leftarrow 1 \).

II. \((t-2)\)-flats, Orbits, Classes

(9) \((t-2)\)-flats corresponding to steps (3)-(5): Form the \((t-2)\)-flats corresponding to those 1-flats formed in steps (3)-(5). For the initial 1-flat \( j_1 \) with independent points \((0, d_0)\), generate the \((t-2)\)-flat with linearly independent points \((0, d_0, 2d_0, 3d_0, ..., (t-2) d_0)\). (Note: these points are linearly independent since the corresponding minimal polynomial is of degree \((t + 1)\)). Label this initial \((t-2)\)-flat \( j_{t-2} \). To obtain the remaining \((\rho-1)\) initial \((t-2)\)-flats in class \( k \), multiply the initial \((t-2)\)-flat \( j_{t-2} \) by \( q^i \) to obtain the initial \((t-2)\)-flat \((j+s)_{t-2}, s = 1, 2, ..., (\rho-1) \). Set \( j \leftarrow j + \rho; k \leftarrow k + 1 \).

(10) \((t-2)\)-flats for each class formed by (3)-(5): Repeat step (9) for each corresponding class generated in steps (3)-(5) above.

(11) \((t-2)\)-flats corresponding to step (7): For the initial 1-flats of step (7), the corresponding initial \((t-2)\)-flats are formed as follows. For the initial 1-flat \( j_1 \) corresponding to the subfield \( GF(q^{i+1}) \), the associated \((t-2)\)-flat has linearly independent points \((0, d_0, 2d_0, 3d_0, ..., id_0)\) plus \((t-2-i)\) other distinct points such that the \((t-1)\) points are linearly independent.
The remaining initial \((t-2)\)-flats in the class are formed as in step (9) for all the initial \((t-2)\)-flats corresponding to this subfield.

(12) \((t-2)\)-flats for each class formed in (7): Repeat (11) for each subfield of step (7).

(13) \((t-2)\)-flats of minimal cycle \(\theta(1)\): For the remaining initial 1-flat of minimum cycle \(\theta(1)\) we use YFH’s method to form the corresponding initial \((t-2)\)-flat. The points \((0, \theta(1), 2\theta(1), \ldots, q\theta(1))\) are in the required \((t-2)\)-flat. Select any other linearly independent point \(c_0\). Then \((0 + c_0, \theta(1) + c_0, \ldots, q\theta(1) + c_0)\) are in the flat plus all linear combinations of the \(2(q+1)\) points. Select a point \(c_1\) not in this set and continue in the same way until \((q^{t-1} - 1)/(q - 1)\) points are obtained.

Remaining \((t-2)\)-flats: The remaining \((t-2)\)-flats in each orbit are obtained by adding 1 to each point in the flat, \(\theta(0)\) times for the first \(\nu(0)\) orbits, and \(\theta(1)\) times for the last orbit.

REFERENCES


