CERIAS Tech Report 2003-49
Cropping-Resilient Segmented Multiple Watermarking
by Mikhail J. Atallah
Center for Education and Research
Information Assurance and Security
Purdue University, West Lafayette, IN 47907-2086

Cropping-Resilient Segmented Multiple Watermarking*
(Extended Abstract)

Keith Frikken and Mikhail Atallah

Purdue University

Abstract. Watermarking is a frequently used tool for digital rights management. An example of this is using watermarks to place ownership information into an object. There are many instances where placing multiple watermarks into the same object is desired. One mechanism that has been proposed for doing this is segmenting the data into a grid and placing watermarks into different regions of the grid. This is particularly suited for images and geographic information systems (GIS) databases as they already consist of a fine granularity grid (of pixels, geographic regions, etc.); a grid cell for watermarking is an aggregation of the original fine granularity cells. An attacker may be interested in only a subset of the watermarked data, and it is crucial that the watermarks survive in the subset selected by the attacker. In the kind of data mentioned above (images, GIS, etc.) such an attack typically consists of cropping, e.g. selecting a geographic region between two latitudes and longitudes (in the GIS case) or a rectangular region of pixels (in an image). The contribution of this paper is a set of schemes and their analysis for multiple watermark placement that maximizes resilience to the above mentioned cropping attack. This involves the definition of various performance metrics and their use in evaluating and comparing various placement schemes.

1 Introduction

Watermarking is a frequently used tool in digital rights management. For example, watermarking can be used for copyright protection [14]; this is done by placing an ownership watermark into the object. Another example is a digital VCR, where watermarks are placed into the object to convey what commands the user is allowed to perform on the object (read only, read and copy, etc.) [14]. Placing multiple watermarks into data has many applications; several examples appear in [13]. One digital rights management application of multiple watermarking is collaborative watermarking. In collaborative watermarking several

* Portions of this work were supported by Grants EIA-9903545 and ISS-0219560 from the National Science Foundation, Contract N00014-02-1-0364 from the Office of Naval Research, by sponsors of the Center for Education and Research in Information Assurance and Security, by Purdue Discovery Park's e-enterprise Center, and by the GAANN fellowship.

F. Dehne, J.-R. Sack, M. Smid (Eds.): WADS 2003, LNCS 2748, pp. 231–242, 2003.

organizations may have partial ownership of an object, and each organization wants to place ownership watermarks into the object. A single organization may choose to place multiple watermarks into the same object for various reasons. For example, defense in depth can be achieved by using different watermarking schemes that have different strengths and weaknesses.

Several techniques have been proposed for inserting multiple watermarks into an object including rewatermarking, segmented watermarking, interleaved watermarking, and composite watermarking [18]. Segmented watermarking divides the object into regions and places each watermark into a set of these regions. A scheme for determining regions is given in [4], but in this paper we assume the regions are equal sized rectangles as in [18]. However, we assume that each of these regions contains enough information to hide a single watermark. An attack against the segmented watermarking scheme would be to take a rectangular subset (a cropping) of the data to remove some of the watermarks. A watermark will survive a cropping if that watermark is contained in a region which is fully enclosed within the cropping. The purpose of the work in this paper is to maximize the number of recoverable watermarks for random croppings. For simplicity, we assume that all croppings are equally likely. The rest of this paper does not depend on the exact nature of the object being watermarked (image, GIS, NASA spatial data, etc.), as long as the object can be naturally partitioned into a grid, and is useful if an adversary may find a rectangular subset of the grid of value for stealing.

In the collaborative watermarking application mentioned above, the cropping attack can be carried out by an outsider or by any of the watermarking organizations. We introduce two performance metrics that are important to this application: (i) Maximum Non-Complete Area(MNCA) and (ii) Minimum Non-Full Area(MNFA). The MNCA is the maximum number of tiles that can be in a cropping which does not contain all watermarks; the MNCA provides a bound on the largest area that can be stolen such that one of the watermarks cannot be recovered. Obviously, minimizing the MNCA is a goal for a placement scheme. As a motivation for MNFA, observe that a cropping that is lacking a watermark yet contains more than one copy of another watermark is "bad". Ideally, no such croppings would exist, but short of this it is desirable to maximize the area of such croppings. The MNFA is the minimum number of tiles that can be in a cropping that does not contain all watermarks, but contains at least one duplicate watermark. The motivation for MNFA is that it is the minimum cropping that will allow an attacker to get away with something (i.e. have less watermarks than there are tiles); for any cropping with less tiles than the MNFA the number of watermarks will be the number of tiles, which is the best any placement can do. A placement scheme should attempt to maximize the MNFA. If a single organization uses multiple ownership watermarks then it is possible that only a subset of the watermarks need to be recovered for proof of ownership. If only t watermarks need to be recovered, the placement scheme should minimize the maximum area that does not contain at least t watermarks.

If we treat the watermarks as colors and the data as a grid, watermark placement can be viewed as grid coloring; in this paper we use the term color when discussing placement schemes and we use the terms tile and region interchangeably. This watermark placement problem is similar to a grid coloring problem used for declustering data in a database among multiple disks to parallelize I/O (see Section 2). For simplicity we restrict this work to data tiled along two dimensions. Furthermore, we only consider croppings of the data on tile boundaries, since every cropping contains a subcropping on tile boundaries. We define the area of a cropping to be the number of complete tiles contained in the cropping.

The results of our work include a formal definition of this problem and a formal definition of the above mentioned comparison heuristics (MNCA and MNFA). A scheme is given that colors any grid with M colors so that the MNCA is $O(M)$, and a scheme is given where the MNFA is $\Omega(M)$. Also in the case where only half of the watermarks need to be recovered, we provide a scheme that colors any grid with M colors in such a way that any area containing M tiles contains half of the watermarks when M is a power of 2. Furthermore, a set of experiments were performed to evaluate the performance of several schemes using these two comparison metrics.

The layout of the rest of this paper is as follows. In Section 2, we discuss the distributed database retrieval problem, which is similar to this watermarking placement problem, but has some key differences. In Section 3, we present a formal definition of this problem along with several results about MNCA, MNFA, and other constraints. In Section 4, we briefly discuss the results of our experimental analysis, and we summarize our contributions in Section 5. Due to space limitations, we often give a proof sketch of a claim; the details of these proofs will be given in the full paper.

2 Related Work

A problem that is similar to the watermark placement problem outlined in the previous section is the distributed database declustering problem. Given an n dimensional database divide each dimension uniformly to get tiles. By placing the tiles on different disks the retrieval of records during query processing can be parallelized, which reduces the I/O time to the time that it takes to retrieve the maximum number of tiles stored on the same disk. The problem of placing the records so that the response times for range queries is minimized has been well studied.

Given k disks and m tiles in a range query, an optimal tile placement would require an I/O time of $\lceil \frac{m}{k} \rceil$. It was shown in [1] that this bound is unachievable for all range queries in a grid except in a few limited circumstances. Since there are many cases where no scheme can achieve this optimal bound, several schemes have been developed to achieve performance that is close to optimal. These schemes include Disk Modulo DM [6], CMD [12], Fieldwise eXclusive or [11], and the HCAM approach [7]. These are just a subset of the techniques that have been proposed for declustering.

Suppose we are given k colors. The DM approach [6] assigns tile (x, y) to $(x + y) \bmod k$. The FX approach [11] assigns tile (x, y) to $(x \oplus y) \bmod k$. Cyclic allocation schemes [15] choose a skip value s such that $gcd(k, s) = 1$ and assigns tile (x, y) to $(x + sy) \bmod k$. The choice of the skip value is what defines the scheme. In RPHM (Relatively Prime Half Modulo), the shift value is defined to be the number nearest to $\frac{M}{2}$ that is relatively prime to M. The EXH (Exhaustive) scheme takes all values of s where $gcd(s, M) = 1$ and finds the one that optimizes a certain criterion. Another class of schemes are the permutation schemes [3], in these schemes a permutation ϕ of the numbers in $\{0, ..., k-1\}$ is chosen and then tile (x, y) is assigned color $(x - \phi^{-1}((y) \bmod k))$. Examples of permutation schemes are DM, the cyclic schemes, and GRS. In the GRS scheme [3] the permutation is computed as follows:

1. $\forall i \in \{0, ..., k-1\}$ compute the fractional part of $\frac{2i}{1+\sqrt{5}}$, and call it k_i.
2. Sort the values k_i and use this to define the permutation.

In [2], a coloring scheme was presented that was later found in [16] to be equivalent to $(x \oplus y^R) \bmod k$, where y^R is the $(\lceil \log k \rceil)$-bit reversal of y; in this paper we will call this scheme RFX (Reverse Fieldwise eXclusive-or). Recently, two new directions have been explored: i) the relation between this area and discrepancy theory [5,16], and ii) the use of redundancy [8,17,19], i.e. placing each record on multiple disks.

The database declustering problem appears similar to that of the watermarking representation problem defined in the previous section, but there are key differences:

1. In the database declustering problem the multiplicity of a color is of central importance, whereas in the watermarking placement problem multiplicity of a color in a cropping is irrelevant (as long as it is nonzero).
2. Given a coloring for k colors it is possible to construct a coloring for $k-1$ colors that will have the same MNCA by ignoring the kth color. In the database problem you cannot ignore a color since that tile may need to be retrieved.
3. Given a coloring for k colors it is possible to construct a coloring for $k+1$ colors that will have the same MNFA by ignoring the $(k+1)$st color. In the database problem this is like not using certain disks, which may improve the additive error from an optimal solution, but will not improve overall query performance (there may be a few cases where it does, but these are very limited).

3 Theoretical Results

3.1 Definitions and Basic Properties

Given M watermarks labeled $\{0, ..., M-1\}$ to place into a two dimensional data, which is tiled into a grid with dimension sizes $d_1 \in \aleph$ and $d_2 \in \aleph$, a *coloring* maps a grid location to a watermark and is defined by a function $C : \aleph \times \aleph \rightarrow$

$\{0, ..., M-1\}$. A coloring C is said to be *periodic* with period p if and only if $C(x,y) = C(x+p,y)$ and $C(x,y) = C(x,y+p)$ for all grid locations (x,y). Furthermore, if each watermark is represented every p tiles (in both dimensions) then the coloring is *completely periodic*. More formally, a coloring C is completely periodic with period p if and only if it is periodic with period p and $\forall w \in \{0,1,2,...,M-2,M-1\}, \forall (x,y) \in \aleph \times \aleph, \exists s_x, s_y$ such that $0 \leq s_x < p, 0 \leq s_y < p$ where $C(x+s_x,y) = w$ and $C(x,y+s_y) = w$.

A coloring works for a specific number of watermarks, but a family of colorings can be grouped together to create a *coloring scheme*. A coloring scheme $\{C_M\}_{M=1}^{\infty}$ is a set of colorings indexed by M, where C_M is a coloring for M watermarks. A coloring scheme $\{C_M\}_{M=1}^{\infty}$ is *completely periodic* with period $\{p_M\}_{M=1}^{\infty}$ if and only if the coloring C_M is completely periodic with period p_M for all $M \geq 1$. It is worth noting that the complete period of many coloring schemes is the number of colors itself; these schemes include: DM, the Cyclic schemes, and GRS; this is also true for the FX and RFX schemes when the number of colors is a power of two.

In what follows, whenever we say "rectangular subsection" of a grid, we implicitly include wraparound, e.g. in a 3×5 grid, the region $[2,0] \times [1,3]$ is considered to be rectangular (the reason for allowing wraparound will become apparent after reading Lemma 3–1). Given a coloring C and a rectangular subsection R, define a function W that computes the set of watermarks present in R, note that $W(R,C) = \{C(i,j), \forall (i,j) \in R\}$.

A watermarking entity will have certain desired constraints for a watermark placement scheme. Given an area threshold a and a watermark threshold b then a possible constraint on a scheme is that any cropping containing a or more tiles contains at least b distinct watermarks. More formally, given an area threshold a and a watermark threshold b a *constraint* (a,b) is *satisfied* for a grid G and coloring C if and only if for any rectangular subsection R in G, if $(|R| \geq a) \rightarrow (|W(R,C)| \geq b)$. A constraint (a,b) is said to be *universally satisfiable* if there is a coloring C such that for any grid G, C satisfies (a,b) for G. We consider only constraints (a,b) with $a \geq b$ and $b \leq M$, since it is trivial to prove that other constraints are unsatisfiable. Define a satisfiability function $S(C, M, (d_1, d_2), (a,b))$ that is true if and only if C satisfies the constraint (a,b) in a $d_1 \times d_2$ grid. Define a universally satisfiable function $US(C, M, (a,b))$ which is true if and only if the C universally satisfies constraint (a,b).

Lemma 3–1.: Given M watermarks, a coloring C that has complete period p, and a reasonable constraint (a,b) such that $S(C, M, (p,p), (a,b))$ is true, then $US(C, M, (a,b))$ is also true.

Proof: Suppose we are given an arbitrary grid and a rectangular subsection of that grid, call it R, of size $s_1 \times s_2$, where $s_1 s_2 \geq a$. We must show that $|W(R,C)| \geq b$. If s_1 or s_2 is greater than or equal to p then it is trivial since C has complete period p and thus contains all M watermarks. Assume $s_1 < p$ and $s_2 < p$, thus R fits in a $p \times p$ grid. Now R is a wraparound cropping in some $p \times p$ grid, and since $S(C_M, M, (p,p), (a,b))$ this area contains b watermarks. Therefore, the constraint is satisfied. □

A consequence of this lemma is that for the colorings defined for the database declustering problem, we need only to look at grids the size of the complete period for that coloring to determine if constraints are universally satisfiable. The following lemma shows how constraints that are universally satisfiable imply weaker constraints that are universally satisfiable.

Lemma 3–2.: If $US(C, M, (a, b))$, then: i) $US(C, M+1, (a, b))$, ii) $US(C, M, (a+1, b))$, and iii) $US(C, M, (a, b-1))$

Proof: The first part states that if a constraint can be universally satisfied for M watermarks, then it is universally satisfiable for $M+1$ watermarks. This is obvious since the $(M+1)$st watermark can be ignored, and the same constraint will still be satisfiable. Since any cropping containing $a+1$ tiles must contain a tiles, and likewise any cropping containing b watermarks must contain at least $b-1$ watermarks, the second and third parts are trivially true. □

3.2 Maximum Non-complete Area

Suppose an organization watermarks some data with the tiling method outlined previously; it would be desirable for this organization to know the largest rectangular subsection that does not contain its watermark as a measure of resilience to cropping of the placement scheme. There is such a subsection for every watermark; define the maximum area of all of these subsections as the *Maximum Non-Complete Area(MNCA)*. Formally, the MNCA of a coloring C for M colors is the value k such that $\neg US(C, M, (k, M))$ and $US(C, M, (k+1, M))$. Obviously, it is desirable to minimize the MNCA for a set of watermarks; note that a strictly optimal declustering would have a MNCA of $(M-1)$.

Theorem 3–3. The best achievable MNCA for any coloring of M watermarks, labeled $\{0, \cdots, M-1\}$ is $M-1$ (i.e. optimal) if and only if $M = 1, 2, 3,$ or 5.

Proof Sketch: For M=1, 2, or 3 the DM coloring scheme has optimal MNCA. For $M = 5$ the RPHM coloring has optimal MNCA. To show that the other cases cannot be done optimally, there are two cases to consider, M is even and M is odd.

Case 1: Suppose $M = 2k$ for some k (and ≥ 4), construct a $4 \times M$ grid (4 columns and M rows). BWOC, suppose that this can be colored optimally. The first column must contain all M colors, WLOG color them in sequential order top down as $(0, \cdots, 2k-1)$. Consider $2 \times k$ sections (which must contain all M colors) that have tiles in the first and second columns of the grid.

0	2	0	2
1	3	1	3
2	0	2	0
3	1	3	1

Diagram 1

From these it can be determined that the second column must be colored in the order $(k, \cdots, 2k-1, 0, \cdots, k-1)$. By similar reasoning, the third column must be $(0, \cdots, 2k-1)$ and the fourth column must be $(k, \cdots, 2k-1, 0, \cdots, k-1)$; the above construction is shown in Diagram 1 for a 4×4 grid colored with $M = 4$ colors. But this implies that a $4 \times \lceil \frac{M}{4} \rceil$ cropping only contains $2\lceil \frac{M}{4} \rceil < M$ colors and thus contradicts our assumption that the grid can be colored optimally.

Case 2: We omit this case of $M = 2k+1$ for some k, but it will be contained in the full version of the paper. However, the proof is similar to Case 1, but it is slightly more complicated. □

The previous theorem states that we cannot obtain optimal MNCA for most values of M. In this section we establish an upper bound on the best achievable MNCA of $O(M)$ for M colors. This is done by proving that the MNCA for GRS is $O(M)$ if M is a Fibonacci number, and this is generalized to any number of colors using a smoothing process that is defined after the next theorem.

Theorem 3–4. If a coloring C for M colors has a MNCA of r, then given $k \leq M$ colors it is possible to construct a coloring C' for k colors that has a MNCA no larger than r.

Proof: Suppose coloring C has a MNCA of r for M colors, which implies that $US(C, M, (r+1, M))$. Define a coloring C', where $C'(x, y) = (C(x, y) \bmod k)$. We must show $US(C', k, (r + 1, k))$. Suppose we are given a rectangular subsection R with area at least $r + 1$, and an arbitrary watermark $w \in \{0, 1, 2, ..., k - 1\}$. There must be a tile (x, y) in R, with $C(x, y) = w$ (since $US(C, M, (r + 1, M))$ and $k \leq M$), which implies $C'(x, y) = w$ and thus $US(C', k, (r + 1, k))$. □

The previous theorem implies that the best achievable MNCA for $M - 1$ colors can be no worse than the best achievable MNCA for M colors, or equivalently that the best achievable MNCA for a specific number of colors is a nondecreasing function of M. A coloring scheme that satisfies this property is called *MNCA-smooth*. Many coloring schemes are not MNCA-smooth (EXH, GRS, and FX), but we can modify these schemes so that this property will hold. Define a function MA that given a coloring returns the MNCA of the coloring. Given a coloring scheme $\{C_M\}_{M=1}^{\infty}$, define a new coloring scheme $\{D_M\}_{M=1}^{\infty}$ where $D_M = (C_k \bmod M)$ where k is chosen such that $MA(C_k) = \min_{M \leq j \leq MA(C_M)}(MA(C_j))$. This process creates a MNCA-smooth coloring scheme, which has MNCA no larger than $\{C_M\}_{M=1}^{\infty}$ for all values of M.

When the number of watermarks is a Fibonacci number (recall that they satisfy the recurrence $F_1 = 1, F_2 = 1$ and $F_k = F_{k-1} + F_{k-2}$), the GRS coloring scheme has a MNCA no larger than double the number of colors (see Theorem 3–5). Using Theorem 3–4, we can get a general bound of $\frac{10}{3}$ times the number of watermarks for any number of watermarks, see Corollary 3–6. Thus the GRS coloring scheme has a MNCA which is $O(M)$.

Theorem 3–5. The GRS coloring has a MNCA of no more than $2 * F_k$ for $M = F_k$ colors where F_k is the kth Fibonacci number.

Proof Sketch: We need only to consider croppings of an $M \times M$ grid with wraparound since the complete period of GRS is M. Suppose we are given such a cropping. To finish the proof we need the concept of gaps that has been defined for permutation schemes [3]. Given r consecutive rows there will be r instances of any color (one per row); the set of distances between these values (including the wraparound distance) will be the same for any color, and these distances are called the gaps of these rows (See Diagram 2 on the next page for more information on gaps). If an area is non-complete then it must have less columns than the maximum gap. It was shown in [3] and [10] that the maximum gap

for $r(= F_i + s)$ rows where $0 \leq s < F_{i-1}$ is F_{k-i+2}. It can be shown that $(F_i + s)(F_{k-i+2} - 1) < 2F_k$. Thus, given any number of rows the maximum area of a non-complete cropping is less than $2F_k$, hence we have proven that MNCA will be no larger than $2F_k$. □

1	2	3	4	5	0
4	5	0	1	2	3
0	1	2	3	4	5

Diagram 2

Diagram 2 shows a permutation coloring for $(M = 6)$ colors with 3 rows. The gaps between 0's are 2, 3, and 1. Notice that the gaps are the same (not necessarily in the same order) for any color.

Corollary 3–6. For M watermarks the MNCA-smoothed GRS scheme has a MNCA no more than $\frac{10}{3}M$.

Proof: If M is a Fibonacci number, then this bound is clearly true. Suppose M is not a Fibonacci number (note $M \geq 4$) then let F be the next Fibonacci number larger than M, note that $F \leq \frac{5}{3}M$, which is easy to verify with induction. Now we can use GRS for F colors to obtain a coloring for M colors that has a MNCA no larger than $2F$ (by theorem 3-4 and theorem 3-5). So the MNCA will be no larger than $\frac{10}{3}M$. □

3.3 Minimum Non-full Area

Another desirable trait of a watermark placement scheme is for small areas to have unique colors. For a coloring there is a minimum area that does not contain unique colors, call this area the *Minimum Non-Full Area(MNFA)*. Formally, the MNFA of a coloring C for M colors is the value k such that $\neg US(C, M, (k, \min\{M, k\}))$ and $US(C, M, (k-1, \min\{M, k-1\}))$. The MNFA is useful since it is the minimum area for which an attacker can attempt to "get away with something", i.e. a cropping that could contain more watermarks than it actually does. It is desirable to maximize the MNFA of a coloring, and the MNFA for a strictly optimal placement is ∞.

Lemma 3–7. If a coloring has a MNFA that is optimal for M colors, then the coloring will be optimal for MNCA as well.

Proof: Since the MNFA of C is optimal we know that $\forall k, US(C, M, (k, \min\{M, k\}))$, so this must be true for $k = M$, and so $US(C, M, (M, M))$. However, this implies that the MNCA is optimal. □

Theorem 3–8. The MNFA for any coloring of M watermarks is ∞ (i.e. optimal) if and only if $M = 1, 2, 3$, or 5.

Proof: For M=1, 2, or 3 the DM coloring scheme has optimal MNFA. For $M = 5$ the RPHM coloring has optimal MNFA. If for other values of M there was an optimal coloring for MNFA then this coloring would be optimal for MNCA (by lemma 3-7), but this contradicts theorem 3-4. □

Theorem 3–9. If a coloring C for M colors has a MNFA of r, then given $k \geq M$ colors C has a MNFA $\geq r$ for k colors.

Proof: Since C has a MNFA of r we know that $US(C, M, (r-1, r-1))$, but by applying the first part of lemma 3-2 repeatedly we get $US(C, k, (r-1, r-1))$. □

The previous theorem implies that the best achievable MNFA for $M+1$ colors can be no worse then the best MNFA for M colors, i.e. the best achievable MNFA is a nondecreasing function of M. A coloring scheme that satisfies this property is called *MNFA-smooth*. Many coloring schemes are not MNFA-smooth (EXH, GRS, and FX), but we can modify these schemes so that this property will hold. Like the MNCA, we can define a MNFA-smoothing process. Define a function *MNFA* that given a coloring returns the MNFA of the coloring. Given a coloring scheme $\{C_M\}_{M=1}^{\infty}$, define a new coloring scheme $\{D_M\}_{M=1}^{\infty}$ such that $D_M = C_k$ where k is chosen such that $MNFA(C_k) = \max_{1 \le j \le M}(MNFA(C_j))$. This process creates a MNFA-smooth coloring scheme, which has MNFA no worse than $\{C_M\}_{M=1}^{\infty}$ for all values of M. However, this transformation has a drawback; if this smoothing process is used then some colors will not be used, which means that some watermarks will not be contained in the data. However, this problem can be fixed by treating each color in the smoothed scheme as a group of colors and whenever a tile is assigned to a group it is randomly assigned a watermark from that group. In Theorem 3–10 and Corollary 3–11 we prove a lower bound of $\Omega(M)$ for the best achievable MNFA for any number of colors M. Like the proof for the upper bound on MNCA, we use the GRS coloring scheme to prove this lower bound on MNFA.

Theorem 3–10. The GRS coloring scheme has a MNFA larger than $\frac{3}{7}F_k$ for $M = F_k$ colors where F_k is the kth Fibonacci number.

Proof Sketch: We only need to consider croppings of an $M \times M$ grid with wraparound since complete period of GRS is M. Suppose we are given a such a cropping. we will use the same concept of gaps as in the proof of Theorem 3–5. If an area is non-full then it must have more columns than the minimum gap. It was shown in [3] and [10] that the minimum gap for $r(= F_i + s)$ rows where $0 \le s < F_{i-1}$ is at least F_{k-i}. It can be shown that $(r)(F_{k-i}+1) \ge (F_i)(F_{k-i}+1) > \frac{3}{7}F_k$. Thus given any number of rows there must be at least $\frac{3}{7}F_k$ tiles before there is a duplicate. Hence, the MNFA will be no less than $\frac{3}{7}M$ □

Corollary 3–11. For M watermarks there is a coloring where the MNFA is no less than $\frac{9}{35}M$.

Proof: If M is a Fibonacci number, then this bound is clearly true. Suppose M is not a Fibonacci number (note $M \ge 4$) then let F be the largest Fibonacci number smaller than M, an easy induction shows that $F \ge \frac{3}{5}M$. Now we can use GRS for F colors to obtain a coloring for M colors that has a MNFA no smaller than $\frac{3}{7}F$ (by theorem 3-9 and theorem 3-10). So the MNFA of the MNFA-smoothed scheme will be no smaller than $\frac{9}{35}M$. □

3.4 Other Satisfiability Properties

Suppose that to prove ownership of an item an entity only has to recover about half of its watermarks. The question becomes how much area is needed so that about half of the colors are represented. Theorem 3–12 states that it is possible to color a grid with $M = 2^k$ colors in such a way that any area containing M

tiles has at least $\frac{M}{2} = 2^{k-1}$ distinct colors. Corollary 3–13 generalizes this result for non-powers of two.

Theorem 3–12. Given $M = 2^k$ colors, there is a coloring C such that $US(C, M, (M, \frac{M}{2}))$.

Proof Sketch : Use the RFX coloring scheme for M colors. We only need to consider wraparound croppings in an $M \times M$ grid since the complete period for RFX is M when M is a power of 2. It can be shown that if you partition the columns into 2^s groups each with 2^{k-s} columns (that have a common prefix of size s), then given any column partition and any 2^s consecutive rows (including wraparound), the $2^{k-s} \times 2^s$ cropping defined by the intersection of the column partition and the rows will contain unique colors (and hence all colors). Furthermore, any cropping containing M tiles must have at least $\frac{M}{2}$ tiles in one of these regions, hence there must be at least $\frac{M}{2}$ colors. □.

Corollary 3–13. Given M colors, there is a coloring C such that $US(C, M, (2^{\lfloor \log(M) \rfloor}, 2^{\lfloor \log(M) \rfloor - 1}))$.

Proof: By theorem 3-12, we know that there is a coloring C such that $US(C, 2^{\lfloor \log(M) \rfloor}, (2^{\lfloor \log(M) \rfloor}, 2^{\lfloor \log(M) \rfloor - 1}))$. But since $M \geq 2^{\lfloor \log(M) \rfloor}$, by Lemma 3–2 we can conclude that $US(C, M, (2^{\lfloor \log(M) \rfloor}, 2^{\lfloor \log(M) \rfloor - 1}))$. □

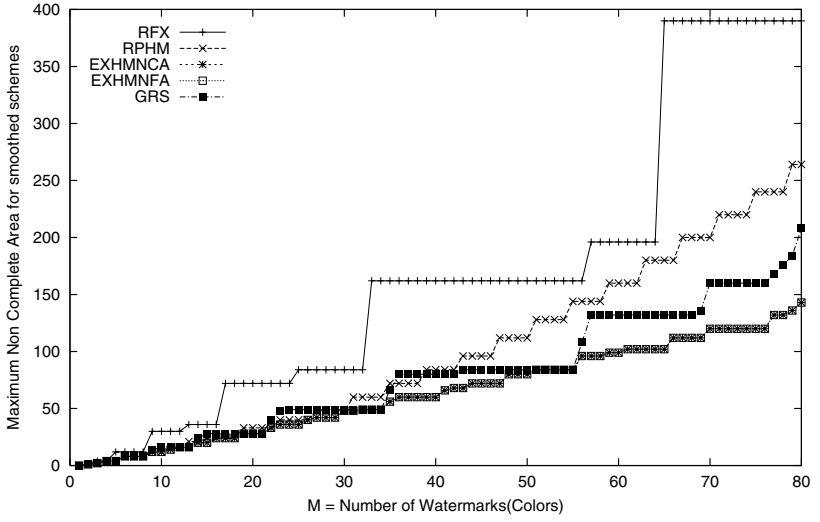

Fig. 1. MNCA of various MNCA-smoothed schemes

4 Experimental Results

To compare colorings we looked at the performance of various schemes with regards to their MNFA and MNCA. The colorings that were examined are: DM,

FX, RFX, RPHM, EXH (optimized for MNCA), EXH (optimized for MNFA), and GRS. Due to page constraints we only include the MNFA of these MNFA-smoothed schemes for up to 80 colors (Figure 1). Note that DM and FX are omitted due to poor performance.

Figure 1 shows that the stronger schemes are EXH and GRS, with EXH slightly outperforming GRS. When smoothing is used the criterion used to optimize EXH appear to have little effect on the performance of the scheme. Similar results occur when the performance criterion is MNFA.

5 Conclusion

Watermarking is a tool for digital rights management, and inserting multiple watermarks into the same data is an important application. A scheme for inserting multiple watermarks into an object consists of tiling the data into uniform rectangles and placing each watermark into a set of tiles; placement of the watermarks in such an environment effects the resilience of the object to croppings. This problem is relates to the distributed database declustering problem, but differs from the latter in significant aspects.

We propose two orthogonal heuristics to compare schemes: MNCA and MNFA. Other than in very limited cases, it is impossible to have optimal performance for either heuristic for every cropping in a grid. Given M colors to place in a grid, the GRS scheme that is smoothed for MNCA has a MNCA of $O(M)$ for any grid, and the GRS scheme that is smoothed for MNFA has a MNFA of $\Omega(M)$. Furthermore, if M is a Fibonacci number then the GRS scheme will achieve both of these bounds; extending both bounds to any number of colors is left for future work. Also, the RFX scheme was proven to have good properties if only half of the watermarks need to be recovered. Furthermore, we performed experiments to evaluate the performance of various schemes with regards to MNCA and MNFA and found that the GRS and EXH schemes have the strongest performance among the colorings schemes that were analyzed.

Acknowledgments. The authors would like to thank Dr. Rei Safavi-Naini for introducing us to this area and Dr. Sunil Prabhakar for his help with the distributed database declustering background.

References

1. K.A.S. Abdel-Ghaffar and A. El Abbadi. Optimal allocation of two-dimension data. In *Int. Conf. on Database Theory*, pages 409–418, Delphi, Greece, Jan. 1997.
2. M. J. Atallah and S. Prabhakar. (Almost) optimal parallel block access for range queries. In *Proc. of the 19th ACM Symposium on Principles of Database Systems (PODS)*, Dallas, Texas, May 2000.
3. R. Bhatia, R. K. Sinha, and C.-M. Chen. Declustering using golden ratio sequences. In *Proc. of Int'l. Conference On Data Engineering (ICDE)*, San Diego, California, March 2000.

4. G. Brisbane, R. Safavi-Naini, and P. Ogunbona. Region-based Watermarking for Images. In *Proceedings of Information Security Workshop (ISW)*, LNCS 1729, 1999, pages 154–166.
5. C. Chen and C. Cheng. From Discrepancy to Declustering: Near-optimal multidimensional declustering strategies for range queries. In *A CM Symposium On Principles of Database Systems (PODS)* 2002 pages 29–38.
6. H. C. Du and J. S. Sobolewski. Disk allocation for cartesian product files on multiple-disk Systems. *ACM Trans of Database Systems*, 7(1):82–101, 1982.
7. C. Faloutsos and P. Bhagwat. Declustering using fractals. In *Proc. of the 2nd Int. Conf. On Parallel und Distributed Information Systems*, pages 18–25, San Diego, CA, Jan 1993.
8. K. Frikken, M. Atallah,S. Prabhakar,R. Safavi-Naini. Optimal Parallel 1/0 for Range Queries through Replication. In *Proc. of DEXA*, LNCS 2453, pages 669–678, 2002.
9. J. Gray, B. Horst, and M. Walker. Parity striping of disc arrays: Low-cost reliable Storage with acceptable throughput. In *Proceedings of the Int. Conf. On Very Large Data Bases*, pages 148-161, Washington DC., August 1990.
10. A. Itai and Z.Rosberg. A golden ratio control policy for a multiple-access channel. In *IEEE Transactions On Automatic Control*, AC-29:712–718, 1984.
11. M. H. Kim and S. Pramanik. Optimal file distribution for partial match retrieval. In *Proc. ACM SIGMOD Int. Conf. On Management of Data*, pages 173–182, Chicago, 1988.
12. J. Li, J. Srivastava, and D. Rotem. CMD: a multidimensional declustering method for parallel database Systems. In *Proceedings of the Int. Conf. On Very Large Data Bases*, pages 3–14, Vancouver, Canada, August 1992.
13. F. Mintzer and G. Braudaway. If one watermark is good, are more better? Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, vol. 4, Phoenix, Arizona, May 1999.
14. F. Mintzer, G. Braudaway, and M. Yeung. Effective and Ineffective Digital Watermarks. In *IEEE ICIP*, volume 111, pages 9–12, Santa-Barbara, Cal, October 1997.
15. S. Prabhakar, K. Abdel-Ghaffar, D. Agrawal, and A. E1 Abbadi. Cyclic allocation of two-dimensional dat a. In *Proc. of the International Conference On Data Engineering (ICDE'98)*, pages 94–101, Orlando, Florida, Feb 1998.
16. Rakesh K. Sinha, Randeep Bhatia, and Chung-Min Chen. Asymptotically optimal declustering schemes for range queries. In *Proc. of 8th International Conference On Database Theory (ICDT)*, pages 144–158, London, UK, January 2001.
17. P. Sanders, S. Egner, and J . Korst. Fast concurrent access to parallel disks. In *11th ACM-SIAM Symposium On Discrete Algorithms*, 2000.
18. N. Sheppard, R. Safavi-Naini, P. Ogunbona. On multiple watermarking. In *ACM Multimedia Conference*, ACM Multimedia 2001, pp. 3–6.
19. A. Tosun and H. Ferhatosmanoglu. Optimal Parallel 1/0 Using Replication. OSU Technical Report OSU-CISRC-ll/Ol-TR26, 2001.