Maximum Identifiable Parent Property Codes for Watermarking

Wen Jiang and Xingxing Yu
School of Mathematics
Georgia Institute of Technology, Atlanta, Georgia, 30332
Email: wjiang@math.gatech.edu, yu@math.gatech.edu

Ye (Geoffrey) Li
School of Electrical and Computer Engineering
Georgia Institute of Technology, Atlanta, Georgia, 30332
Email: liye@ece.gatech.edu

***Abstract*–In order to provide copyright protection for digital materials, a distributor embeds different watermarks (codewords) into different products before sending them to customers. The watermark in each product can be used to identify the customer who buys that product; thereby, redistributing the product is equivalent to exposing the customer's identity. However, a group of customers can collude and compare their products to detect their watermarks, and then create a new product with a pirate watermark. Codes with *identifiable parent property* (IPP) can provide means of traceability in the presence of a collusion attack. Many results have been obtained on codes with IPP, including those on constructions and sizes. In particular, an estimation of the maximum size of IPP codes of length 3 is given in [5]. With the help from graph theory, we devise an algoriuthm which determines the maximum size of IPP codes of length 3. We also use techniques from graph theory and nonlinear optimization to give a precise formula for the maximum size of IPP codes of length 3 when the size of the alphabet can be written as $r^2 + 2r$ for some integer r. Moreover, our arguments allow us to construct classes of maximum size IPP codes of length 3, and these codes allow for efficient tracing.**

***Keywords*–Watermark, IPP code, edge colored graph, components, Kuhn-Tuker Conditions**.

This work is partially supported by the NSA under Grant MDA904-03-1-0052 and NSF under Grant DMS-0245530.

I. Introduction

Fingerprinting, first introduced by Wagner [1], is a technique for identifying users who use digital materials for unintended purposes, such as redistribution. In order to control the redistribution of digital materials, a distributor embeds a watermark (codeword) into each product through a variety of watermarking techniques before sending it to customers. Using different watermarks for different copies makes each copy unique. The watermark in each product can be used to identify the customer who buys that product and, thereby, redistributing the product is equivalent to exposing the customer's identity. However, a group of customers can collude and compare their products to detect their watermarks, and then create a new product with a pirate watermark. The problem of designing fingerprints that can withstand collusion and allow for the identification of colluders has been extensively studied in recent years. Several schemes for tracing the customers who use their content for unitended purposes are designed in [2] and [3]. Additive embedding technique against collusion and a new class of codes, called anti-collusion codes, are proposed in [4].

Codes with *identifiable parent property* (IPP), first introduced in [5], can provide means of traceability in the presence of a collusion attack (for example, tracing the source of an unauthorized distribution). Clearly, an IPP code with large size can mark more digital products.

Constructions of maximum size IPP codes have been studied (for example, [5]) and decoding algorithms of IPP codes have been proposed (for example, [7] and [8]). Bounds on the maximum size of IPP codes of length n are obtained in [5]. Such bounds are improved in [6] for IPP codes of length 4. For IPP codes of length three over an alphabet with q elements, it is proved in [5] that the maximum size is bounded above by $3q-1$.

In this paper, we design an algorithm which determines the maximum size of IPP codes of length 3 over any given alphabet Q. In fact, when $|Q|$ can be written in the form r^2+2r, we show that a lower bound given in [5] in fact gives a precise formula for the maximum size of IPP codes of length 3. We conjetcure that a similar result holds when $|Q| = r^2+2r+k$ with $1 \leq k \leq 2r+2$. We also show how to construct maximum size IPP codes of length 3. It turns out that such codes have very low decoding complexity and hence allow for efficient tracing.

To obtain our results, we associate to a code C an edge colored graph as in [5], in such a way that the IPP of C is equivalent to certain structural conditions on the graph. We then use IPP graphs to construct a class of IPP codes of length 3, which improves a lower bound on the maximum size of IPP codes of length 3. This is done in Section II. We show in Section III that there is always a maximum size IPP code of length 3 whose associated graph has a special structure, which is used in Section IV to reduce the maximum size problem to a nonlinear programming problem. We then design a MATLAB program which determines the maximum size of IPP codes of length 3 over Q, and construct maximum size IPP codes from the outcomes of the program. In Section V, we show that the strucure of our constructed codes can be used for efficient tracing. In Section VI, we solve the nonlinear programming problem when $|Q|$ can be written as r^2+2r for some integer r, thereby, giving a precise formular for the maximum size of IPP codes of length 3 when $|Q| = r^2+2r$.

Notations and terminology from graph theory and nonlinear programming are given in the Appendix, as well as detailed proofs of some technical lemmas.

II. Codes and Associated Graphs

In this section, we define graphs associated with codes. We shall see that a code is IPP if and only if its associated graph satisfies certain structural properties. We then prove a lower bound on the maximum size of an IPP code of length 3 by constructing IPP codes from IPP graphs, which improves the lower boud given in [5].

Throughout the rest of the paper, we fix the alphabet $Q = \{\alpha_1, \alpha_2, \cdots, \alpha_q\}$. For any $\mathbf{x} \in Q^n$, we also write $\mathbf{x} = (x_1, x_2, \cdots, x_n)$. Let $C \subseteq Q^n$ be a code. For any two words $\mathbf{a}$, $\mathbf{b} \in Q^n$, the *descendant* set of $\mathbf{a}$ and $\mathbf{b}$ is defined as

$$D(\mathbf{a}, \mathbf{b}) := \{\mathbf{x} \in Q^n | x_i \in \{a_i, b_i\}, i = 1, 2, \cdots, n\}.$$

If $\mathbf{c} \in D(\mathbf{a}, \mathbf{b})$, then $\mathbf{a}$ and $\mathbf{b}$ are *parents* of $\mathbf{c}$. For a code C, we define its *descendant code* to be

$$C^* := \bigcup_{\mathbf{a},\mathbf{b} \in \mathbf{C}} D(\mathbf{a}, \mathbf{b}).$$

For example, if $C = \{0000, 1111\}$, then $C^* = F_2^4$.

We say that a code C has *IPP* if for every word in C^*, at least one of its parents can be identified. An IPP code $C \subseteq Q^n$ is said to be *maximum* if its size is the largest among all IPP codes of length n over Q.

The following result from [5] gives a necessary and sufficient condition for a code to have IPP.

***Lemma** 2.1: A code $C \subseteq Q^n$ has IPP if and only if the following two conditions hold:*
(IPP1) for any three distinct codewords **a, b, c** *in C, there exists some $1 \leq i \leq n$ such that a_i, b_i, c_i are all distinct, and*
(IPP2) for any four distinct codewords **a, b, c, d** *in C, there exists some $1 \leq i \leq n$ such that $\{a_i, b_i\} \cap \{c_i, d_i\} = \phi$.*

We shall use graphs to study maximum size IPP codes. See Appendix A for graph theory terminology. For each code $C \subseteq Q^n$, we define a graph G such that the vertices of G represent the codewords in C, and two vertices of G are joined by an edge of color i if their corresponding codewords have the same ith coordinate. Clearly $|C| = |V(G)|$. For each $1 \leq i \leq n$, let $G(i)$ denote the *spanning subgraph* of G whose edges are those edges of G with color i. We say G is an *IPP*

graph if C is an IPP code. With the above definition and notations, Lemma 2.1 can be stated as follows,

***Lemma** 2.2: Let $C \subseteq Q^n$ and G be its associated graph. Then G is an IPP graph if and only if the following two conditions hold:*
(IPP1) for any three distinct vertices u, v, w of G, there exists some color $1 \leq i \leq n$ such that u, v, w belong to three different components of $G(i)$, and
(IPP2) for any four distinct vertices u, v, w, x of G, there exists some color $1 \leq i \leq n$ such that any component of $G(i)$ containing u or v contains neither w nor x.

The following result is a direct consequence of Lemma 2.2.

***Lemma** 2.3: Let G be an IPP graph and S be a union of components of G. Then S is an IPP graph.*

For graphs associated with some IPP codes of length 3, the following useful result proved in [5] gives structure information.

***Lemma** 2.4: Let $C \subseteq Q^3$ be an IPP code and G be its associated graph. Then*
(i) if $|C| > q$, no two vertices of G are joined by more than one edge,
(ii) G contains no triangle whose edges use three different colors, and
(iii) G contains no path $v_1v_2v_3v_4$ whose edges v_1v_2, v_2v_3, v_3v_4 use three different colors.

By applying Lemma 2.2 and Lemma 2.3, we can establishe the converse of Lemma 2.3, its proof can be found in Appendix B.

***Lemma** 2.5: Let $C \subseteq Q^3$ and G be its associated graph, and let S, T be unions of components of G such that $S \cap T = \emptyset$ and $S \cup T = G$. If S and T are IPP graphs, then so is G.*

Let S be a component of an edge colored graph G. If edges of S use only one color, then S is called a *uni-color component*; if edges of S use only two colors, then S is called a *bi-color component*; if edges of S use exactly three colors, then S is called a *tri-color component*.

***Lemma** 2.6: Let $C \subseteq Q^3$ and G be its associated graph, and let S be a component of G.*

(i) If S is a uni-color component or a bi-color component in G then S is an IPP graph.

(ii) If there are a vertex z of S and complete subgraphs S_1, S_2, S_3 of S such that $S_1 \cup S_2 \cup S_3 = S$, all edges of S_i are colored with color i, $1 \leq i \leq 3$, and $V(S_i \cap S_j) = \{z\}$ for all $1 \leq i \neq j \leq 3$, then S is an IPP graph.

The proof of Lemma 2.6 is presented in Appendix C.

Notice that for each integer $q \geq 24$ there exist integers r and k such that $r \geq 4$, $0 \leq k \leq 2r + 2$, and $q = r^2 + 2r + k$.

In the following theorem, we use IPP graphs to give a lower bound on the size of a maximum IPP code of length 3 over Q. In our proof, we actually give an explicit construction of several classes of IPP codes, including one class constructed in [5]. Such codes provide means for efficient tracing, as will be shown in Section V.

***Theorem** 2.7: Let $q = r^2 + 2r + k$ where $r \geq 4$ and $0 \leq k \leq 2r + 2$. Let G be the graph associated with a maximum IPP code $C \subseteq Q^3$. Then $|V(G)| \geq h(q)$, where*

$$h(q) = \begin{cases} 3q - 6r, & \text{if } k = 0 \\ 3q - 6r - 2, & \text{if } k = 1 \\ 3q - 6r - 3, & \text{if } 2 \leq k \leq r + 1 \\ 3q - 6r - 5, & \text{if } k = r + 2 \\ 3q - 6r - 6, & \text{if } r + 3 \leq k \leq 2r + 2 \ . \end{cases} \tag{1}$$

Proof. It suffices to construct an IPP graph G with $|V(G)| = h(q)$. We distingush five cases according to k.

Case 1. $k = 0$ and $q = r^2 + 2r$

First, we construct an edge colored graph B_m for each $1 \leq m \leq 3$, as follows. Take r disjoint complete graphs R_s with $|V(R_s)| = r$, $1 \leq s \leq r$, and label the vertices of each R_s by $v^m_{s,1}, v^m_{s,2}, \cdots, v^m_{s,r}$. Let $\{i, j\} = \{1, 2, 3\} - \{m\}$. Color all edges of each R_s with color i. For each $1 \leq t \leq r$, join every pair of vertices from $\{v^m_{1,t}, v^m_{2,t}, \cdots, v^m_{r,t}\}$ by edges of color j. Let B_m denote the resulting edge colored graph, $1 \leq m \leq 3$. Note that $|V(B_m)| = r^2$ and the edges of B_m do not use color m. The components of $B_m(i)$ are the graphs R_s, $1 \leq s \leq r$, the components of $B_m(j)$ are the complete graphs with vertex set $\{v^m_{1,t}, v^m_{2,t}, \cdots, v^m_{r,t}\}$ $(1 \leq t \leq r)$, and the components of $B_m(m)$ are the isolated vertices $v^m_{s,t}$ $(1 \leq s, t \leq r)$.

Now let G denote the edge colored graph which is the

$$C_0 = \begin{pmatrix} (\alpha_1,\alpha_1,\alpha_1), & (\alpha_2,\alpha_1,\alpha_{r+1}), & \cdots & (\alpha_r,\alpha_1,\alpha_{r^2-r+1}), \\ (\alpha_1,\alpha_2,\alpha_2), & (\alpha_2,\alpha_2,\alpha_{r+2}), & \cdots & (\alpha_r,\alpha_2,\alpha_{r^2-r+2}), \\ \vdots & \vdots & \vdots & \vdots \\ (\alpha_1,\alpha_r,\alpha_r), & (\alpha_2,\alpha_r,\alpha_{2r}), & \cdots & (\alpha_r,\alpha_r,\alpha_{r^2}), \\ \\ (\alpha_{r+1},\alpha_{r+1},\alpha_{r^2+1}), & (\alpha_{2r+1},\alpha_{r+2},\alpha_{r^2+1}), & \cdots & (\alpha_{r^2+1},\alpha_{2r},\alpha_{r^2+1}), \\ (\alpha_{r+2},\alpha_{r+1},\alpha_{r^2+2}), & (\alpha_{2r+2},\alpha_{r+2},\alpha_{r^2+2}), & \cdots & (\alpha_{r^2+2},\alpha_{2r},\alpha_{r^2+2}), \\ \vdots & \vdots & \vdots & \vdots \\ (\alpha_{r+r},\alpha_{r+1},\alpha_{r^2+r}), & (\alpha_{2r+r},\alpha_{r+2},\alpha_{r^2+r}), & \cdots & (\alpha_{r^2+r},\alpha_{2r},\alpha_{r^2+r}), \\ \\ (\alpha_{r^2+r+1},\alpha_{2r+1},\alpha_{r^2+r+1}), & (\alpha_{r^2+r+1},\alpha_{3r+1},\alpha_{r^2+r+2}), & \cdots & (\alpha_{r^2+r+1},\alpha_{r^2+r+1},\alpha_{r^2+2r}), \\ (\alpha_{r^2+r+2},\alpha_{2r+2},\alpha_{r^2+r+1}), & (\alpha_{r^2+r+2},\alpha_{3r+2},\alpha_{r^2+r+2}), & \cdots & (\alpha_{r^2+r+2},\alpha_{r^2+r+2},\alpha_{r^2+2r}), \\ \vdots & \vdots & \vdots & \vdots \\ (\alpha_{r^2+2r},\alpha_{2r+r},\alpha_{r^2+r+1}), & (\alpha_{r^2+r+r},\alpha_{3r+r},\alpha_{r^2+r+2}), & \cdots & (\alpha_{r^2+2r},\alpha_{r^2+2r},\alpha_{r^2+2r}). \end{pmatrix} \quad (2)$$

disjoint union of B_1, B_2 and B_3. That is, B_1, B_2 and B_3 are the three bi-color components of G. By Lemma 2.6, B_1, B_2 and B_3 are IPP graphs. In view of Lemma 2.5, G is an IPP graph. Note that $|V(G)| = 3r^2 = 3q - 6r = h(q)$.

We will show that G is associated with the code $C \subseteq Q^3$ given in array (2). For each $1 \le s, t \le r$, let

$$(\alpha_s, \alpha_t, \alpha_{(s-1)r+t})$$

be the codeword in C corresponding to the vertex $v^3_{s,t}$ of B_3, let

$$(\alpha_{sr+t}, \alpha_{r+s}, \alpha_{r^2+t})$$

be the codeword in C corresponding to the vertex $v^1_{s,t}$ of B_1, and let

$$(\alpha_{r^2+r+t}, \alpha_{(s+1)r+t}, \alpha_{r^2+r+s})$$

be the codeword in C corresponding to the vertex $v^2_{s,t}$ of B_2. Then $C = C_0$. It is easy to verify that G is associated with C.

For the other cases when $k \ge 1$, we can construct IPP graphs similar to Case 1.

Case 2. $k = 1$ and $q = r^2 + 2r + 1$

Construct B_1, B_2, B_3 as exactly in Case 1. Add one more vertex $v^3_{1,r+1}$ to B_3, for each $1 \le t \le r$, join $v^3_{1,r+1}$ and $v^3_{1,t}$ by an edge of color 1, now we obtain a new component B'_3. Let G denote the disjoint union of these three bi-color components B_1, B_2, B'_3, then G is an IPP graph by Lemma 2.5 and Lemma 2.6. Notice that $|V(G)| = 3r^2 + 1 = 3q - 6r - 2$.

Let

$$(\alpha_1, \alpha_{r^2+2r+1}, \alpha_{r^2+2r+1})$$

be the codeword in C corresponding to the added vertex $v^3_{1,r+1}$, and let

$$C = C_0 \cup \{(\alpha_1, \alpha_{r^2+2r+k}, \alpha_{r^2+2r+k})\}. \quad (3)$$

Then G is associated with the code C in (3).

Case 3. $2 \le k \le r + 1$ and $q = r^2 + 2r + k$

Construct B_1, B_2, B_3 as exactly in Case 1. For each $1 \le m \le 3$, add $k - 1$ vertices $v^m_{s,r+1}$ $(1 \le s \le k - 1)$ to B_m. Let $\{i, j\} = \{1, 2, 3\} - \{m\}$. For each $1 \le s \le k-1$, join $v^m_{s,r+1}$ and $v^m_{s,t}$ $(1 \le t \le r)$ by an edge of color i. Join every pair of $\{v^m_{1,r+1}, v^m_{2,r+1}, \cdots, v^m_{k-1,r+1}$ by edges of color j. Let B'_m be the resulting bi-color graph, $1 \le m \le 3$. Then B'_m is an IPP graph by Lemma 2.6. Let G denote the disjoint union of these three bi-color components B'_1, B'_2, B'_3, then G is an IPP graph by Lemma 2.5. Note that $|V(G)| = 3r^2 + 3(k - 1) = 3q - 6r - 3$.

Let X_m, $1 \le m \le 3$, denote the codeword set

corresponding to the $k-1$ vertices added to B_m, where

$$X_3 = \{(\alpha_s, \alpha_{r^2+2r+1}, \alpha_{r^2+2r+s}), 1 \le s \le k-1\},$$
$$X_1 = \{(\alpha_{r^2+2r+s}, \alpha_{r+s}, \alpha_{r^2+2r+k}), 1 \le s \le k-1\},$$
$$X_2 = \{(\alpha_{r^2+2r+k}, \alpha_{r^2+2r+1+s}, \alpha_{r^2+r+s}), 1 \le s \le k-1\}.$$

Let

$$C = C_0 \cup X_3 \cup X_1 \cup X_2. \tag{4}$$

Then G is associated with the code C in (4).

Case 4. $k = r+2$ and $q = r^2+2r+k$

Construct B_1', B_2', B_3' exactly as in Case 3. Add one more vertex $v^3_{1,r+2}$ to B_3', for each $1 \le t \le r+1$, join $v^3_{1,r+2}$ and $v^3_{1,t}$ by an edge of color 1, now we obtain a new component B_3''. Let G denote the disjoint union of these three bi-color components B_1', B_2', B_3'', then G is an IPP graph by Lemma 2.5 and Lemma 2.6. Notice that $|V(G)| = 3r^2+3r+1 = 3q-6r-5$.

Let

$$(\alpha_1, \alpha_{r^2+3r+2}, \alpha_{r^2+3r+2})$$

be the codeword in C corresponding to the added vertex $v^3_{1,r+2}$, and let

$$C = C_0 \cup X_3 \cup X_1 \cup X_2 \cup \{(\alpha_1, \alpha_{r^2+3r+2}, \alpha_{r^2+3r+2})\}. \tag{5}$$

Then G is associated with the code C in (3).

Case 5. $r+3 \le k \le 2r+2$ and $q = r^2+2r+k$

Construct B_1', B_2', B_3' exactly as in Case 3. For each $1 \le m \le 3$, add $k-r-2$ vertices $v^m_{s,r+1}$ $(1 \le s \le k-1)$ to B_m'. Let $\{i,j\} = \{1,2,3\} - \{m\}$. For each $1 \le s \le k-r-2$, join $v^m_{s,r+2}$ and $v^m_{s,t}$ $(1 \le t \le r+1)$ by an edge of color i. Join every pair of $\{v^m_{1,r+2}, v^m_{2,r+2}, \cdots, v^m_{k-r-2,r+2}$ by edges of color j. Let B_m'' be the resulting bi-color graph, $1 \le m \le 3$. Then B_m'' is an IPP graph by Lemma 2.6. Let G denote the disjoint union of these three bi-color components B_1'', B_2'', B_3'', then G is an IPP graph by Lemma 2.5. Note that $|V(G)| = 3r^2+3r+3(k-r-2) = 3q-6r-6$.

Let Y_m, $1 \le m \le 3$, denote the codeword set corresponding to the $k-r-2$ vertices added to B_m', where

$$Y_3 = \{(\alpha_s, \alpha_{r^2+3r+2}, \alpha_{r^2+3r+1+s}), 1 \le s \le k-r-2\},$$
$$Y_1 = \{(\alpha_{r^2+3r+1+s}, \alpha_{r+s}, \alpha_{r^2+2r+k}), 1 \le s \le k-r-2\},$$
$$Y_2 = \{(\alpha_{r^2+2r+k}, \alpha_{r^2+3r+2+s}, \alpha_{r^2+r+s}), 1 \le s \le k-r-2\}.$$

Let

$$C = C_0 \cup X_3 \cup X_1 \cup X_2 \cup Y_3 \cup Y_1 \cup Y_2. \tag{6}$$

Then G is associated with the code C in (6). □

III. Maximum IPP Graphs

The main result of this section gives structural conditions on a graph associated with a maximum IPP code of length 3. This is crucial for constructing maximum IPP codes which allow for efficient tracing. First, we need some notation.

An edge colored graph G is a *proper* if it consists of exactly three bi-color components S_1, S_2, S_3 such that S_i does not use color i, for all $1 \le i \le 3$. For an IPP code $C \subseteq Q^3$, let $Q_i(C)$ $(1 \le i \le 3)$ denote the set of elements of Q, each of which occurs as the ith coordinate of some codeword in C. For a subgraph H of G associated with $C \subseteq Q^3$, $Q_i(H)$ denotes the set of elements of Q, each of which occurs as the ith coordinate of some codeword corresponding to a vertex of H. The following result is proved in [5].

***Lemma** 3.1: Let $C \subseteq Q^3$ be an IPP code, let G be its associated graph, and let S be a component of G. Then one of the following holds.*
(i) S is a uni-color component, and if all edges of S use color i for some $1 \le i \le 3$ then $|Q_i(S)| = 1$ and $|Q_j(S)| = |V(S)|$ for all $j \in \{1,2,3\} - \{i\}$.
(ii) S is a bi-color component, and if the edges of S use colors i and j for some $1 \le i \ne j \le 3$ then $|Q_k(S)| = |V(S)|$ for $k \in \{1,2,3\} - \{i,j\}$.
(iii) S is a tri-color component, there exist a vertex v of G and three complete subgraphs S_1, S_2, S_3 of G such that all edges of S_i use color i for $1 \le i \le 3$, $V(S_i \cap S_j) = \{v\}$ for all $1 \le i \ne j \le 3$, and

$$|V(S)| = \frac{|Q_1(S)| + |Q_2(S)| + |Q_3(S)| - 1}{2}.$$

Using the above lemma, we can prove the following result.

***Theorem** 3.2: Let $C \subseteq Q^3$ be an IPP code and G be its associated graph. If G contains no bi-color components, then $|V(G)| < \frac{3q}{2}$.*

Proof. Let $S_1, S_2, \cdots, S_m$ be the components of G, which are either uni-color or tri-color. Then by (i) and

(iii) of Lemma 3.1, we have

$$|V(S_i)| = \frac{|Q_1(S_i)| + |Q_2(S_i)| + |Q_3(S_i)| - 1}{2}.$$

Since $\sum_{i=1}^{m} |Q_j(S_i)| \leq q$ for all $1 \leq j \leq 3$, we have

$$\begin{aligned} |V(G)| &= \textstyle\sum_{i=1}^{m} |V(S_i)| \\ &= \textstyle\sum_{i=1}^{m} \frac{1}{2}(|Q_1(S_i)| + |Q_2(S_i)| + |Q_3(S_i)| - 1) \\ &\leq (3q - m)/2. \quad \square \end{aligned}$$

Theorem 3.2 tells us that if the associated edge graph of an IPP code of length 3 contains no bi-color components, then this code is not maximum.

Next, we prove a series of lemmas concernning the structure of IPP graphs.

***Lemma** 3.3: Let $C \subseteq Q^3$ be a maximum IPP code and let G be its associated graph. Then G has at least three components and one of these is a bi-color component.*

Proof. Suppose G has at most two components. If G has exactly one component, then by (i) of Lemma 2.4, $|V(G)| \leq q$. If G has exactly two components, then again by (i) of Lemma 2.4 we have $|V(G)| \leq 2q - 2$. In both cases, we see that $|V(G)| < h(q)$. (Recall the function $h(q)$ in (1)). Hence, by Theroem 2.7, $|C|$ is not maximum, a contradiction.

Now assume G has no bi-color components. Then by Theorem 3.2, $|V(G)| < 3q/2 < h(q)$. In view of Theorem 2.7, $|C|$ is not maximum, a contradiction. $\square$

The following three Lemmas are proved in Appendix D, E, F.

***Lemma** 3.4: Suppose C is a maximum IPP code which is chosen so that its associated graph G has the minimum number of components. Then no two components of G use exactly the same colors.*

***Lemma** 3.5: Suppose C is a maximum IPP code which is chosen so that its associated graph G has the minimum number of components. Then G has no uni-color components.*

***Lemma** 3.6: Suppose C is a maximum IPP code which is chosen so that its associated graph G has the minimum number of components. Then G has no tri-color components.*

Now we can prove the main result of this section.

***Theorem** 3.7: There exists a maximum IPP code $C \subseteq Q^3$ such that its associated graph is a proper graph.*

Proof. Choose a maximum IPP code C such that its associated graph G has the minimum number of components. By Lemma 3.5 and Lemma 3.6, all components of G are bi-color. Therefore, by Lemma 3.4, G has at most three components. It follows from Lemma 3.3 that G has exactly three bi-color components. Hence, G is proper. $\square$

Theorem 3.7 tells us the explicit structure of a class of maximum IPP codes of length 3. We can use this theorem to construct maximum IPP codes which allow for efficient tracing.

IV. A Nonlinear Programming and Maximum IPP Codes Construction

In this section, we reduce the problem of determining the size of maximum IPP codes of length 3 to a nonlinear programming problem. Based on this nonlinear programming, we design an algorithm which determine the size of a maximum IPP code. From the outcome of the algorithm, we show how to construct maximum size IPP codes of length 3.

A. Nonlinear Programming Formulation

First, we need a result on IPP graphs to derive constraints for the nonlinear programming.

***Lemma** 4.1: Let $C \subseteq Q^3$ be an IPP code and G be its associated graph. Let S be a component of G whose edges use colors i and j for some $1 \leq i \neq j \leq 3$, and let $\{k\} = \{1, 2, 3\} - \{i, j\}$. Then*

$$|Q_i(S)| + |Q_j(S)| - 1 \leq |Q_k(S)| \leq |Q_i(S)||Q_j(S)|.$$

Proof. Let $R_1, \cdots, R_{n_1}$ denote the components of $S(i)$, and let $T_1, \cdots, T_{n_2}$ be the components of $S(j)$. Then those codewords in C corresponding to vertices of R_s (respectively, T_t) have the same ith (respectively, jth) coordinate, and R_s (respectively, T_t) is a complete graph. Moreover, by (i) of Lemma 2.4, $|V(R_s) \cap V(T_t)| \leq 1$.

Define an auxiliery graph H as follows. The vertices of H are $R_1, \cdots, R_{n_1}$ and $T_1, \cdots, T_{n_2}$. For any $1 \leq s \leq n_1$ and $1 \leq t \leq n_2$, R_s and T_t are joined with an edge in H when $|V(R_s) \cap V(T_t)| = 1$. Let m denote the number of edges in H. Since S is connected, H is connected, and hence, $m \geq n_1 + n_2 - 1$. Note that

m represents the number of pairs R_s and T_t such that $|V(R_s) \cap V(T_t)| \neq 0$.

We now count $|Q_k(S)|$ for $1 \leq k \leq 3$. Since there is no edge of color k, and because $\cup_{s=1}^{n_1} R_s$ and $\cup_{t=1}^{n_2} T_t$ have m vertices in common, we have

$$\begin{aligned}
&|Q_i(S)| = n_1 + (\textstyle\sum_{t=1}^{n_2} |V(T_t)|) - m,\\
&|Q_j(S)| = n_2 + (\textstyle\sum_{s=1}^{n_1} |V(R_s)|) - m, \text{ and}\\
&|Q_k(S)| = (\textstyle\sum_{s=1}^{n_1} |V(R_s)|) + (\textstyle\sum_{t=1}^{n_2} |V(T_t)|) - m.
\end{aligned}$$

Hence

$$|Q_i(S)| + |Q_j(S)| = |Q_k(S)| + n_1 + n_2 - m.$$

Since $m \geq n_1 + n_2 - 1$, we have

$$|Q_k(S)| \geq |Q_i(S)| + |Q_j(S)| - 1.$$

Next, we show $|Q_k(S)| \leq |Q_i(S)||Q_j(S)|$. Suppose on the contrary $|Q_k(S)| \geq |Q_i(S)||Q_j(S)| + 1$. Then there must exist some symbol $a \in Q_i(S)$ such that a occurs as the first coordinate of at least $|Q_j(S)| + 1$ codewords. By pigeonhole principle, there exist two vertices of S joined by two edges, contradicting (i) of Lemma 2.4. $\square$

Remark. Let $C \subseteq Q^3$ be an IPP code, let G be its associated graph, and let S be a bi-color component of G whose edges use color i and color j. Then $|Q_k(S)| = |Q_i(S)||Q_j(S)|$ only if for any $a \in Q_i(S)$ (respectively, $a \in Q_j(S)$), a occurs as the ith (respectively, jth) coordinate in exactly $|Q_j(S)|$ (respectively, $|Q_i(S)|$) codewords. Therefore, each component of $S(i)$ (respectively, $S(j)$) is a complete graph on exactly $|Q_j(S)|$ (respectively, $|Q_i(S)|$) vertices. It is easy to see that this necessary condition is also sufficient for $|Q_k(S)| = |Q_i(S)||Q_j(S)|$.

Let $C \subseteq Q^3$ be a maximum IPP code and G be its associated graph. By Theorem 3.7, we may choose C so that G has exactly three components B_1, B_2, B_3 which are all bi-color components and color i doesn't occur in B_i for $1 \leq i \leq 3$. Then

$$|C| = |V(G)| = |Q_1(B_1)| + |Q_2(B_2)| + |Q_3(B_3)| \quad (7)$$

and

$$\sum_{i=1}^{3} |Q_j(B_i)| \leq q, \ \ j = 1, 2, 3. \quad (8)$$

Moreover, it follows from Lemma 4.1 that

$$\begin{aligned}
&|Q_2(B_1)| + |Q_3(B_1)| - 1 \leq |Q_1(B_1)| \leq |Q_2(B_1)||Q_3(B_1)|,\\
&|Q_1(B_2)| + |Q_3(B_2)| - 1 \leq |Q_2(B_2)| \leq |Q_1(B_2)||Q_3(B_2)|,\\
&|Q_1(B_3)| + |Q_2(B_3)| - 1 \leq |Q_3(B_3)| \leq |Q_1(B_3)||Q_2(B_3)|.
\end{aligned} \quad (9)$$

Next, we translate the problem of finding the size of a maximum IPP code of length 3 over Q to a non-linear programming problem. Since B_1, B_2, B_3 are three bi-color components, $|Q_j(B_i)| \geq 2$, $1 \leq i, j, \leq 3$. Combine this with (8), we obtain $2 \leq |Q_j(B_i)| \leq q-4$, $1 \leq i, j, \leq 3$. For $1 \leq i, j \leq 3$, let $|Q_j(B_3)| = x_j, |Q_j(B_1)| = y_j, |Q_j(B_2)| = z_j$. Let $N = \{2, 3, \cdots, q-4\}$. Then Finding the maximum $|V(G)|$ in (7) subject to (8) and (9) is equivalent to solving the following nonlinear optimization problem:
maximize the function

$$f(\mathbf{x}) = x_3 + y_1 + z_2 \quad (10)$$

subject to

$$\begin{aligned}
&\mathbf{x} = (x_1, x_2, x_3, y_1, y_2, y_3, z_1, z_2, z_3) \in N^9,\\
&\text{(i)} \quad g_1(\mathbf{x}) = x_3 - x_1 x_2 \leq 0,\\
&\text{(ii)} \quad g_2(\mathbf{x}) = y_1 - y_2 y_3 \leq 0,\\
&\text{(iii)} \quad g_3(\mathbf{x}) = z_2 - z_1 z_3 \leq 0,\\
&\text{(iv)} \quad g_4(\mathbf{x}) = x_1 + y_1 + z_1 - q \leq 0,\\
&\text{(v)} \quad g_5(\mathbf{x}) = x_2 + y_2 + z_2 - q \leq 0,\\
&\text{(vi)} \quad g_6(\mathbf{x}) = x_3 + y_3 + z_3 - q \leq 0,\\
&\text{(vii)} \quad g_7(\mathbf{x}) = x_1 + x_2 - 1 - x_3 \leq 0,\\
&\text{(viii)} \quad g_8(\mathbf{x}) = y_2 + y_3 - 1 - y_1 \leq 0,\\
&\text{(ix)} \quad g_9(\mathbf{x}) = z_1 + z_3 - 1 - z_2 \leq 0.
\end{aligned} \quad (11)$$

B. An Algorithm

By analyzing the the constraints in (11), we can obtain a smaller interval for each coordinate of $\mathbf{x}$. For example, from the constraint $g_7(\mathbf{x})$, we get $x_1 + x_2 \leq x_3 + 1$, so $x_1 \leq x_3 - 1 \leq q - 5$ and $x_2 \leq x_3 - 1 \leq q - 5$. Again by $g_7(\mathbf{x})$, we have $x_1 x_2 \leq (\frac{x_1 + x_2}{2})^2 \leq \frac{(x_3 + 1)^2}{4}$. Similarly, we can get a smaller interval for other coordinates of $\mathbf{x}$.

With the above formulation, the following MATLAB program can be used to determine the maximum size

of IPP codes of length 3. This program searches for an optimal solution to maximize (10) under (11).

function [maximum, **x**]=$f(q)$
maximum=0; **x**=zeros(1,9);
for $x_1 = 2 : (q-5)$
for x_2 = 2:min(floor($(q-3)^2/4x_1$), $q-3-x_1$)
for $y_2 = 2 : (q-5)$
for y_3 = 2:min(floor($(q-3)^2/4y_2$), $q-3-y_2$)
for $z_1 = 2 : (q-5)$
for z_3 = 2:min(floor($(q-3)^2/4z_1$), $q-3-z_1$)
for $x_3 = (x_1+x_2-1)$:min(min($x_1x_2, q-4$), $q-y_3-z_3$)
for $y_1 = (y_2+y_3-1)$:min(min($y_2y_3, q-4$), $q-x_1-z_1$)
for $z_2 = (z_1+z_3-1)$:min(min($z_1z_3, q-4$), $q-x_2-y_2$)
Temp=$x_3+y_1+z_2$;
if Temp>maximum
maximum=Temp;
$\mathbf{x} = [x_1, x_2, x_3, y_1, y_2, y_3, z_1, z_2, z_3]$;
end;
end; end; end; end; end; end; end; end; end;
return;

C. Maximum IPP Code Construction

In this subsection, we describe the construction of a maximum IPP code $C \subseteq Q^3$ based on IPP graphs. Let

$$\mathbf{x} = (x_1, x_2, x_3, y_1, y_2, y_3, z_1, z_2, z_3) \in N^9$$

be the output of the above algorithm.

Recall that $Q = \{\alpha_1, \alpha_2, \cdots, \alpha_q\}$. First, we construct an edge colored IPP graph which contains three bi-color components B_1, B_2, B_3, where the edges of B_i, $1 \le i \le 3$, do not use color i.

We construct B_3 as follows. Take x_1 disjoint complete graphs R_s, $1 \le s \le x_1$ such that

$$\max_{1 \le s \le x_1} |V(R_s)| = x_2$$

and

$$\sum_{s=1}^{x_1} |V(R_s)| = x_3.$$

Label the vertices of each R_s by $v_{s,1}, v_{s,2}, \cdots, v_{s,|V(R_s)|}$. Color all edges of each R_s with color 1. For each $1 \le t \le x_2$, let $J_t = \{s||V(R_s)| \ge t\}$. Join every pair of vertices from $\{v_{s,t}, s \in J_t\}$ by edges of color 2. Let B_3 denote the resulting edge colored graph. Note that $|V(B_3)| = x_3$ and the edges of B_3 do not use color 3. The components of $B_3(1)$ are the graphs R_s, $1 \le s \le x_1$, the components of $B_3(2)$ are the complete graphs with vertex set $\{v_{s,t}, s \in J_t\}$, $1 \le t \le x_2$, and the components of $B_3(3)$ are the isolated vertices $\{v_{s,1}, v_{s,2}, \cdots, v_{s,|V(R_s)|}, 1 \le s \le x_1\}$.

We construct B_1 and B_2 similar to B_3.

Take y_2 disjoint complete graphs R_s, $1 \le s \le y_2$ such that

$$\max_{1 \le s \le y_2} |V(R_s)| = y_3$$

and

$$\sum_{s=1}^{y_2} |V(R_s)| = y_1.$$

Label the vertices of each R_s by $v_{s,1}, v_{s,2}, \cdots, v_{s,|V(R_s)|}$. Color all edges of each R_s with color 2. For each $1 \le t \le y_3$, let $J_t = \{s||V(R_s)| \ge t\}$. Join every pair of vertices from $\{v_{s,t}, s \in J_t\}$ by edges of color 3. Let B_1 denote the resulting edge colored graph. Note that $|V(B_1)| = y_1$ and the edges of B_1 do not use color 1. The components of $B_1(2)$ are the graphs R_s, $1 \le s \le y_2$, the components of $B_1(3)$ are the complete graphs with vertex set $\{v_{s,t}, s \in J_t\}$, $1 \le t \le y_3$, and the components of $B_1(1)$ are the isolated vertices $\{v_{s,1}, v_{s,2}, \cdots, v_{s,|V(R_s)|}, 1 \le s \le y_2\}$.

Take z_3 disjoint complete graphs R_s, $1 \le s \le z_2$ such that

$$\max_{1 \le s \le z_3} |V(R_s)| = z_1$$

and

$$\sum_{s=1}^{z_3} |V(R_s)| = z_2.$$

Label the vertices of each R_s by $v_{s,1}, v_{s,2}, \cdots, v_{s,|V(R_s)|}$. Color all edges of each R_s with color 3. For each $1 \le t \le z_1$, let $J_t = \{s||V(R_s)| \ge t\}$. Join every pair of vertices from $\{v_{s,t}, s \in J_t\}$ by edges of color 1. Let B_2 denote the resulting edge colored graph. Note that $|V(B_2)| = z_2$ and the edges of B_2 do not use color 2. The components of $B_2(3)$ are the graphs R_s, $1 \le s \le z_3$, the components of $B_2(1)$ are the complete graphs with vertex set $\{v_{s,t}, s \in J_t\}$, $1 \le t \le z_1$, and the components of $B_1(1)$ are the isolated vertices $\{v_{s,1}, v_{s,2}, \cdots, v_{s,|V(R_s)|}, 1 \le s \le z_3\}$.

Now let G denote the edge colored graph which is the disjoint union of B_1, B_2 and B_3. That is, B_1, B_2 and B_3 are the components of G. Hence, G is an IPP graph in view of Lemma 2.6 and Lemma 2.5. Note that $|V(G) = x_3 + y_1 + z_2$.

Based on the IPP graph G, we can construct an IPP code C as follows.

Let the codeword in C corresponding to the vertex $v_{i,j}$ $(1 \le i \le x_1, 1 \le j \le |V(R_i)|)$ in B_3 be

$$(\alpha_i, \alpha_j, \alpha_{(\sum_{s=1}^{i-1}|V(R_s)|)+j}), \tag{12}$$

let the codeword in C corresponding to the vertex $v_{i,j}$ $(1 \le i \le y_2, 1 \le j \le |V(R_i)|)$ in B_1 be

$$(\alpha_{x_1+(\sum_{s=1}^{i-1}|V(R_s)|)+j}, \alpha_{x_2+i}, \alpha_{x_3+j}), \tag{13}$$

and let the codeword in C corresponding to the vertex $v_{i,j}$ $(1 \le i \le z_3, 1 \le j \le |V(R_i)|)$ in B_2 be

$$(\alpha_{x_1+y_1+j}, \alpha_{x_2+y_2+(\sum_{s=1}^{i-1}|V(R_s)|)+j}, \alpha_{x_3+y_3+i}). \tag{14}$$

It is easy to verify that G is the the associated graph of the code C.

V. Code Efficient Tracing

In this section, we will show the maximum IPP code C constructed in Section IV.C allows efficient tracing. Notice that any codeword of C are one of these three versions from (12), (13), (14),

(I) $(\alpha_i, \alpha_j, \alpha_{(\sum_{s=1}^{i-1}|V(R_s)|)+j})$,
(II) $(\alpha_{x_1+(\sum_{s=1}^{i-1}|V(R_s)|)+j}, \alpha_{x_2+i}, \alpha_{x_3+j})$,
(III) $(\alpha_{x_1+y_1+j}, \alpha_{x_2+y_2+(\sum_{s=1}^{i-1}|V(R_s)|)+j}, \alpha_{x_3+y_3+i})$.

Hence, any two coordinates of a codeword uniquely determine the third coordinate of this codeword. For any descendant $\mathbf{x} = (x_1, x_2, x_3) \in C^*$, then at least two coordinates x_k, x_l $(1 \le k < l \le 3)$ are from the same codeword, and x_k, x_l uniquely determine the third coordinate by our encoding method. Hence, the parent of $\mathbf{x}$ that contributes two coordinates to the descendant $\mathbf{x}$ can be identified quickly.

VI. A Precise Formula

In this section, we shall use method from non-linear programming to determine the maximum size of an IPP code $C \subseteq Q^3$ when $q = r^2 + 2r$ for some integer $r \ge 4$.

First, we ignore the set constraint $\mathbf{x} \in N^9$, as is common with solving non-linear programming problems, hoping that an optimum solution will be in N^9. Second, since we aim to maximize $f(\mathbf{x})$, the inequality constraints $g_7(\mathbf{x}) \le 0$, $g_8(\mathbf{x}) \le 0$, and $g_8(\mathbf{x}) \le 0$ seem to be less likely to be active for an optimal solution. Let $\Omega = [2, q-4]^9 \subseteq R^9$. Hence, we proceed to solve the following nonlinear programming problem.

Maximize

$$f(\mathbf{x}) = x_3 + y_1 + z_2 \tag{15}$$

subject to

$$\begin{aligned}
&\mathbf{x} \in \Omega,\\
\text{(i)}\quad &g_1(\mathbf{x}) = x_3 - x_1x_2 \le 0,\\
\text{(ii)}\quad &g_2(\mathbf{x}) = y_1 - y_2y_3 \le 0,\\
\text{(iii)}\quad &g_3(\mathbf{x}) = z_2 - z_1z_3 \le 0,\\
\text{(iv)}\quad &g_4(\mathbf{x}) = x_1 + y_1 + z_1 - q \le 0,\\
\text{(v)}\quad &g_5(\mathbf{x}) = x_2 + y_2 + z_2 - q \le 0,\\
\text{(vi)}\quad &g_6(\mathbf{x}) = x_3 + y_3 + z_3 - q \le 0.
\end{aligned} \tag{16}$$

Note that $f(\mathbf{x})$ is a continuous function, and the domain of $\mathbf{x}$ is bounded. Thus, the maximum of $f(\mathbf{x})$ does exist.

Next, we show that every points $\mathbf{x} \in \Omega$ satisfies (16) is a regular point. To do this, we need to find gradient vectors of $g_i(\mathbf{x})$ at $\mathbf{x} = (x_1, \ldots, x_9)$. By a simple calculation, we see that

$$\begin{aligned}
\nabla g_1(\mathbf{x}) &= (-x_2, -x_1, 1, 0, 0, 0, 0, 0, 0),\\
\nabla g_2(\mathbf{x}) &= (0, 0, 0, 1, -y_3, -y_2, 0, 0, 0),\\
\nabla g_3(\mathbf{x}) &= (0, 0, 0, 0, 0, 0, -z_3, 1, -z_1),\\
\nabla g_4(\mathbf{x}) &= (1, 0, 0, 1, 0, 0, 1, 0, 0),\\
\nabla g_5(\mathbf{x}) &= (0, 1, 0, 0, 1, 0, 0, 1, 0),\\
\nabla g_6(\mathbf{x}) &= (0, 0, 1, 0, 0, 1, 0, 0, 1).
\end{aligned} \tag{17}$$

It is an easy excercise to show that if $\sum_{i=1}^{6} c_i \nabla g_i(\mathbf{x}) = \mathbf{0}$ then $c_i = 0$ for all $1 \le i \le 6$. Hence these six vectors $\nabla g_i(\mathbf{x})$ are linearly independent. Therefore, we have the following.

***Lemma** 6.1: Every point $\mathbf{x} \in \Omega$ satisfying (16) is a regular point.*

The Khun-Tuker conditions (in Appendix) are necessary for $f(\mathbf{x})$ to achieve local maximum at regular points. For the nonlinear programming in (15) and (16), it can be stated as follows.

***Lemma** 6.2: Suppose f and g_i, $i = 1, 2, \cdots, 6$, are given as in (15) and (16), and let $\mathbf{x} \in \Omega$. If f has local maximum at $\mathbf{x}$ then there is a vector $(\mu_1, \mu_2, \cdots, \mu_6)$ with $\mu_i \geq 0$ for all $1 \leq i \leq 6$ such that*

$$\frac{\partial f(\mathbf{x})}{\partial x_j} - \sum_{i=1}^{6} \mu_i \frac{\partial g_i(\mathbf{x})}{\partial x_j} = 0,$$
$$\mu_i g_i(\mathbf{x}) = 0, i = 1, 2, \cdots, 6.$$

Note that the conditions $\mu_i g_i(\mathbf{x}) = 0$ and $\mu_i \geq 0$ imply that if g_i is not active at $\mathbf{x}$ then $\mu_i = 0$. This shows that only active constraints will be used when we determining potential maximum points.

Next, we show that

***Theorem** 6.3: Let $q = r^2 + 2r$ for some $r \geq 4$. If $C \subseteq Q^3$ is a maximum IPP code, then $|C| = 3q - 6r$.*

Proof. By the above analysis, our objective is to find a solution $\mathbf{x} = (x_1, x_2, x_3, y_1, y_2, y_3, z_1, z_2, z_3) \in N^9$ to maximize (15) subject to (16) when $q = r^2 + 2r$. It is hoped that our optimal solution also satisfies the constraints (vii)-(ix) in (11).

Let $L(\mathbf{x}) = f(\mathbf{x}) - \sum_{i=1}^{6} \mu_i g_i(\mathbf{x})$, $\mu_i \geq 0$ for $1 \leq i \leq 6$. By Lemma 6.2, we have

$$\frac{\partial L}{\partial x_1} = \mu_1 x_2 - \mu_4 = 0, \quad \frac{\partial L}{\partial x_2} = \mu_1 x_1 - \mu_5 = 0,$$
$$\frac{\partial L}{\partial x_3} = 1 - \mu_1 - \mu_6 = 0, \quad \frac{\partial L}{\partial y_1} = 1 - \mu_2 - \mu_4 = 0,$$
$$\frac{\partial L}{\partial y_2} = \mu_2 y_3 - \mu_5 = 0, \quad \frac{\partial L}{\partial y_3} = \mu_2 y_2 - \mu_6 = 0,$$
$$\frac{\partial L}{\partial z_1} = \mu_3 z_3 - \mu_4 = 0, \quad \frac{\partial L}{\partial z_2} = 1 - \mu_3 - \mu_5 = 0,$$
$$\frac{\partial L}{\partial z_3} = \mu_3 z_1 - \mu_6 = 0, \quad \mu_i \geq 0, \mu_i g_i = 0, i = 1, 2, \cdots, 6.$$

For any feasible point $\mathbf{x}$, if $g_i(\mathbf{x})$ is inactive for some $1 \leq i \leq 6$, then $\mu_i = 0$, and it is easy to show $f(\mathbf{x}) \leq 2q - 2 < 3q - 6r$.

Therefore, we only need to consider those feasible points such that all functional constraints $g_i(\mathbf{x})$, $1 \leq i \leq 6$, are active, which implies all μ_i 's are positive. Hence,

$$\mu_4 = 1 - \mu_2, \quad \mu_5 = 1 - \mu_3, \quad \mu_6 = 1 - \mu_1,$$
$$x_1 = \frac{1 - \mu_3}{\mu_1}, \quad x_2 = \frac{1 - \mu_2}{\mu_1}, \quad y_2 = \frac{1 - \mu_1}{\mu_2},$$
$$y_3 = \frac{1 - \mu_3}{\mu_2}, \quad z_1 = \frac{1 - \mu_1}{\mu_3}, \quad z_3 = \frac{1 - \mu_2}{\mu_3}.$$

Since all $g_i(\mathbf{x})$, $1 \leq i \leq 6$, are active, we have

$$r^2 + 2r - \frac{1 - \mu_3}{\mu_2} - \frac{1 - \mu_2}{\mu_3} - \frac{(1 - \mu_2)(1 - \mu_3)}{\mu_1^2} = 0,$$
$$r^2 + 2r - \frac{1 - \mu_3}{\mu_1} - \frac{1 - \mu_1}{\mu_3} - \frac{(1 - \mu_3)(1 - \mu_1)}{\mu_2^2} = 0,$$
$$r^2 + 2r - \frac{1 - \mu_2}{\mu_1} - \frac{1 - \mu_1}{\mu_2} - \frac{(1 - \mu_1)(1 - \mu_2)}{\mu_3^2} = 0.$$

By running the computation on Maple 9, we obtain a positive solution

$$\mu_1 = \mu_2 = \mu_3 = \frac{1}{r+1}, \ \mu_4 = \mu_5 = \mu_6 = \frac{r}{r+1}.$$

It follows that

$$\mathbf{x} = (r, r, r^2, r^2, r, r, r, r^2, r) \tag{18}$$

is the unique solution that satisfies Lemma 6.2, and this solution also satisfies constraints (vii)-(ix) as well. Hence, (18) is a local maximum point, and it is the globle maximum point by the uniqueness. Therefore, $|C| = \max_{\mathbf{x} \in N^9} f(\mathbf{x}) = 3r^2 = 3q - 6r$. $\square$

VII. Conclusion

In this paper, we design IPP codes for digital fingerprinting. Using techniques from graph theory and nonlinear optimization, we derive the maximum size of IPP codes of length 3. Based on the stuctural information from IPP graphs, we design maximum IPP codes. Compared to some previous tracability codes, the advantage is that of our IPP codes allow for very efficient tracing.

VIII. Acknowledgment

We would like to thank Professor Ray Liu for introducing this topic to us.

Appendix

A. Some Concepts from Graph Theory

A *graph* G consists of a vertex set $V(G)$ and an edge set $E(G)$, and each edge joins two vertices of G. A graph H is a *subgraph* of G if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. Let H be a subgraph of G, $G - H$ is the subgraph of G obtained by deleting the vertices of H and all edges of G incident with the vertices in $V(H)$. We say a graph is a *complete graph* if there is exactly one edge between every pair of distinct vertices. A complete

graph with 3 vertices is also called a *triangle*. A *path* in a graph is a sequence of distinct vertices $v_1v_2\cdots v_m$ such that there is an edge joining v_i and v_{i+1} for all $1 \leq i \leq m-1$; in this case, we say that the path is between v_1 and v_m. A graph is *connected* if there is a path between every pair of distinct vertices. A *component* of a graph is a maximal connected subgraph. We assume an isolated vertex of a graph is also a component.

B. Proof of Lemma 2.5

It suffices to prove that G satisfies (IPP1) and (IPP2) of Lemma 2.2.

To prove that G satisfies (IPP1) of Lemma 2.2, let u, v, w be three distinct vertices of G. We need to show that there exists some $1 \leq i \leq 3$ such that u, v, w belong to three different components of $G(i)$. First, assume $\{u, v, w\} \subseteq V(S)$. Since S is an IPP graph, there exists some $1 \leq i \leq 3$ such that u, v, w belong to three different components of $S(i)$. Since S is a union of components of G, any component of $S(i)$ is also a component of $G(i)$. Hence, u, v, w belong to three different components of $G(i)$. So we may assume that $\{u, v, w\} \not\subseteq V(S)$. Similarly, we may assume that $\{u, v, w\} \not\subseteq V(T)$. Then by symmetry, we may assume that $u, v \in V(S)$ and $w \in V(T)$. Since S is an IPP graph, there exists some color $1 \leq i \leq 3$ such that u and v belong to two different components of $S(i)$. Since $S \cap T = \emptyset$, the component of $T(i)$ containing w is disjoint from $S(i)$. Since S and T are unions of components of G, each component of $S(i)$ or $T(i)$ is also a component of $G(i)$. Hence, u, v, w belong to three different components of $G(i)$. So G satisfies (IPP1) of Lemma 2.2.

To prove that G satisfies (IPP2) of Lemma 2.2, let u, v, w, x be four distinct vertices of G. We need to show that no component of $G(i)$ containing u or v contains w or x.

First, assume $\{u, v, w, x\} \subseteq V(S)$. Since S is an IPP graph, there exists some $1 \leq i \leq 3$ such that no component of $S(i)$ containing u or v contains w or x. Since S is a union of components of G, the components of $G(i)$ containing one of $\{u, v, w, x\}$ is also a component of $S(i)$. Hence, no compoent of $G(i)$ containing u or v contains w or x.

So we may assume that $\{u, v, w, x\} \not\subseteq V(S)$. Similarly, we may assume that $\{u, v, w, x\} \not\subseteq V(T)$.

Next, assume that one of S and T contain three of $\{u, v, w, x\}$, and the other contains one of $\{u, v, w, x\}$. By symmetry, we may assume that $u, v, w \in V(S)$ and $w \in V(T)$. Since S is an IPP graph, there exists some $1 \leq i \leq 3$ such that u, v, w belong to three different components of $S(i)$. Since $S \cap T = \emptyset$, the component of $T(i)$ containing x is disjoint from $S(i)$. Also since S and T are unions of components of G, each component of $G(i)$ containing one of $\{u, v, w, x\}$ is also a component of $S(i)$ or $T(i)$. Therefore, no component of $G(i)$ containing u or v contains w or x.

So we may assume that each of S and T contains exactly two vertices from $\{u, v, w, x\}$. We need to consider two more cases.

First, suppose one of S and T contains $\{u, v\}$ and the other contains $\{w, x\}$. By symmetry, assume that $\{u, v\} \subseteq V(S)$ and $\{w, x\} \subseteq V(T)$. Since S is an IPP graph, there exists some $1 \leq i \leq 3$ such that u, v belong to different components of $S(i)$. Note that any component of $T(i)$ containing w or x is contained in T and, hence, is disjoint from $S(i)$. As before, any component of $G(i)$ containing one of $\{u, v, w, x\}$ is a component of $S(i)$ or $T(i)$. Hence no component of $G(i)$ containing u or v contains w or x.

Therefore, the remaining case to be considered is when neither S nor T contains $\{u, v\}$ or $\{w, x\}$. By symmetry, we may assume that $\{u, w\} \subseteq V(S)$ and $\{v, x\} \subseteq V(T)$. Again, since S is an IPP graph, there exists some $1 \leq i \leq 3$ such that u, w belong to different components of $S(i)$. Similarly, since T is an IPP graph, there exists some $1 \leq j \leq 3$ such trhat v, x belong to different components of $T(j)$. Note that any component of $G(i)$ containing one of $\{u, v, w, x\}$ is a component of $S(i)$ or $T(j)$. If v, x belong to different components of $T(i)$, then we see that u, v, w, x belong to four different components of $G(i)$. So we may assume that v, x belong to the same component of $T(i)$. Then $i \neq j$ and by (i) of Lemma 2.4, v, x belong to different components of $T(k)$, where $\{k\} = \{1, 2, 3\} - \{i, j\}$. Similarly, if u, w belong to different components of $S(j)$, then u, v, w, x belong

to four different components of $G(j)$. So we may assume u, w belong to the same component of $S(j)$. Hence, by (i) of Lemma 2.4, u, w belong to different components of $S(k)$. Again, since S and T are unions of components, any component of $G(k)$ containing one of $\{u, v, w, x\}$ is a component of $S(k)$ or $T(k)$. Therefore, u, v, w, x belong to four different components of $G(k)$. □

C. Proof of Lemma 2.6

It suffices to show that (i) and (ii) of Lemma 2.2 hold for S.

To prove (i), we let i be the color not used by edges of S. Then every component of $S(i)$ is an isolated vertex. Hence, (IPP1) and (IPP2) of Lemma 2.2 hold. Since G is associated with C, S is also associated with a code (whose codewords are the codewords in C corresponding to the vertices of S).

Next, we prove (ii). Let u, v, w be distinct vertices of S. If $\{u, v, w\} \subseteq V(S_i) \cup V(S_j)$ for some $1 \leq i \neq j \leq 3$, then we see that u, v, w belong to three different components of $S(k)$ for $k \in \{1, 2, 3\} - \{i, j\}$. If no S_i contains two of $\{u, v, w\}$, then clearly, u, v, w belong to different components of $S(1)$. So (i) of Lemma 2.2 holds.

Now let u, v, w, x be four distinct vertices of S. If $\{u, v, w, x\} \subseteq V(S_i \cup S_j)$ for some $1 \leq i \neq j \leq 3$, then we see that u, v, w, x belong to four different components of $S(k)$, where $k \in \{1, 2, 3\} - \{i, j\}$. So assume that we may assume by symmetry that S_1 contains two of $\{u, v, w, x\}$ and each of $S_2 - \{z\}$ contains exactly one of $\{u, v, w, x\}$. First, assume $\{u, v\} \subseteq V(S_1)$. Then S_1 is a component of $S(1)$, and so, no component of $S(1)$ containing u or v contains w or x. Similarly, if $\{w, x\} \subseteq V(S_1)$, then no component of $S(1)$ containing u or v contains w or x. So by symmetry, assume that $\{u, w\} \subseteq V(S_1)$, $v \in V(S_2) - \{z\}$, and $x \in V(S_3) - \{z\}$. If $u = z$ then $\{u, v\} \subseteq V(S_2)$ and, as in the previous case, no component of $S(1)$ containing u or v contains w or x. So assume $u \neq z$. Then we see that the component of $S(2)$ containing u is an isolated vertex, and the component of $S(2)$ containing v is S_2. Hence, no component of $S(1)$ containing u or v contains w or x. □

D. Proof of Lemma 3.4

First, assume that there are two uni-color components S and T of G whose edges use the same color i for some $1 \leq i \leq 3$. Let G' be the graph obtained from G by adding edges uv for all $u \in V(S)$ and $v \in V(T)$. Let H denote the component of G' containing S and T. Note that all other components of G' are components of G. It is easy to see that G' is the edge colored associated with a code C', where C' is obtained from C by changing the ith coordinate of each codeword in C corresponding to a vertex of T to the ith coordinate of the codewords corresponding to the vertices of S. Since H is a uni-color component, H is an IPP graph (by (i) of Lemma 2.6). Since $G - V(S \cup T)$ is a union of components of G, $G - V(S \cup T)$ is an IPP graph. Therefore, by Lemma 2.5, $G' = (G - V(S \cup T)) \cup H$ is also an IPP graph. However, $|V(G')| = |V(G)$ and the number of the components of G' is less than that of G, contradicting the choice of C and G.

Now assume that there are two bi-color components S and T of G whose edges use the same colors i and j for some $1 \leq i \neq j \leq 3$. Let S' be a component of $S(i)$ and T' be a component of $T(i)$. Let G' be the graph obtained from G by adding edges uv of color i for all $u \in V(S')$ and $v \in V(T')$. Let H denote the component of G' containing S and T. Note that all other components of G' are components of G. It is easy to see that G' is the edge colored graph associated with a code C', where C' is obtained from C by changing the ith coordinate of each codeword in C corresponding to a vertex of T' to the ith coordinate of the codewords corresponding to the vertices of S'. Since H is a bi-color component, H is an IPP graph by (ii) of Lemma 2.6. Since $G - V(S \cup T)$ is a union of components of G, $G - V(S \cup T)$ is an IPP graph. Therefore, by Lemma 2.5, $G' = (G - V(S \cup T)) \cup H$ is also an IPP graph. However, $|V(G')| = |V(G)$ and the number of the components of G' is less than that of G, contradicting the choice of C and G.

Finally, assume that there are two tri-color components S and T of G. By (iii) of Lemma 3.1, there exist a vertex v of S (respectively, w of T) and three complete subgraphs S_1, S_2, S_3 of S (respectively, T_1, T_2, T_3 of T) such that all edges of each S_i (respectively, T_i) use color

i for $1 \leq i \leq 3$ and $V(S_i \cap S_j) = \{v\}$ (respectively, $V(T_i \cap T_j) = \{w\}$) for all $1 \leq i \neq j \leq 3$. Let G' be the graph obtained from G by adding edges xy of color 1 for all $x \in V(S_1)$ and $y \in V(T_1)$, adding edges xy of color 2 for all $x \in V(S_2)$ and $y \in V(T_2) - \{w\}$, adding edges xy of color 3 for all $x \in V(S_3)$ and $y \in V(T_3) - \{w\}$, and deleting edges between w and $V(T_2 \cup T_3) - \{w\}$. Let H denote the component of G' obtained from S and T. Note that there are three complete subgraphs H_1, H_2 and H_3 of H such that all edges of each H_i are colored by i for $1 \leq i \leq 3$ and $V(H_i \cap H_j) = \{v\}$ for all $1 \leq i \neq j \leq 3$. Hence, it follows from (iii) of Lemma 2.6 that H is an IPP graph. Note that all other components of G' are components of G. It is easy to see that G' is the edge colored graph associated with a code C', where C' is obtained from C by changing the ith coordinate of each codeword in C corresponding to a vertex of $T_i - w$ to the ith coordinate of the codewords in C corresponding to the vertices of S_i $(1 \leq i \leq 3)$, and changing the 1st coordinate of the codeword corresponding to w to the 1st coordinate of the codewords corresponding to the vertices in S_1. Since $G - V(S \cup T)$ is a union of components of G, $G - V(S \cup T)$ is an IPP graph. Therefore, by Lemma 2.5, $G' = (G - V(S \cup T)) \cup H$ is also an IPP graph. However, $|V(G')| = |V(G)$ and the number of the components of G' is less than that of G, contradicting the choice of C and G. $\square$

E. Proof of Lemma 3.5

Suppose on the contrary that G contains a uni-color component S, whose edges are colored with i for some $1 \leq i \leq 3$.

First, we show that we may choose S so that the color used in S is also used in another component T of G. Suppose this is not true. Then all other components of G are uni-color or bi-color components. By Lemma 3.3, let T be a bi-color component. Then the edges of T use colors from $\{1, 2, 3\} - \{i\}$. By Lemma 3.3, G has a component U other than S and T. If U is a uni-color component, then we see from Lemma 3.4 that the edges of U use a color from $\{1, 2, 3\} - \{i\}$, and therefore, U, T would give the desired choice. So we may assume that U is also a bi-color component. Then by Lemma 3.4 again, U and T cannot use the same two colors. Hence, color i is used in U. Therefore, S and U give the desired choice.

Let T' be a component of $T(i)$. Let G' be the graph obtained from G by adding edges uv of color i for all $u \in V(S)$ and $v \in V(T')$. Clearly, G' is the graph associated with a code $C' \subseteq Q^3$ obtained from C by changing the ith coordinate of those codewords in C corresponding to vertices of T' to the ith coordinate of the codewords in C corresponding to vertices of S. Let H be the component of G' containing $S \cup T$. Note that $G - V(H)$ consists of components of G, and hence, is an IPP graph.

When T is a bi-color component, we see from (iii) of Lemma 2.6 that H is an IPP graph. Now assume T is a tri-color component. Then there exist a vertex v of T and complete subgraphs T_1, T_2, T_3 of T such that all edges of T_s use color s for $1 \leq s \leq 3$ and $V(T_s \cap T_t) = \{v\}$ for all $1 \leq s \neq t \leq 3$. In this case, $T' = T_i$, and we see that H has three complete subgraphs H_1, H_2, H_3 such that $H_1 \cup H_2 \cup H_3 = H$, all edges of each H_s use color s for $1 \leq s \leq 3$, and $V(H_s \cap H_t) = \{v\}$ for all $1 \leq s \neq t \leq 3$. (In fact, $V(H_i) = V(S) \cup V(T_i)$.) By (iii) of Lemma 2.6, H is an IPP graph.

Since both H and $G - V(H)$ are IPP graphs, it follows from Lemma 2.5 that G' is an IPP graph. However, $|V(G')| = |V(G)|$ and G' has fewer components than G, contradicting the choice of C and G. $\square$

F. Proof of Lemma 3.6

Suppose G contains a tri-color component S. Then there exist a vertex v of S and complete subgraphs S_1, S_2, S_3 of S such that $S_1 \cup S_2 \cup S_3 = S$, all edges of each S_i use color i for $1 \leq i \leq 3$, and $V(S_i \cap S_j) = \{v\}$ for $1 \leq i \neq j \leq 3$. By Lemma 3.4, S is the only tri-component of G. By Lemma 3.5, all components of G other than S are bi-color components. Therefore, it follows from Lemma 3.3 that there are two bi-color components in G, say T and U. By Lemma 3.4, we may assume that T uses colors 1 and 2, and U uses colors 2 and 3.

Next, we construct new graph G'. Let T' be a component of $T(1)$, let T'' be a component of $T(2)$, and let U' be a component of $U(3)$. Let G' be obtained from G by adding edges xy of color 1 for all $x \in V(T')$ and $y \in V(S_1)$, adding edges xy of color 2 for all $x \in V(T'')$ and $y \in V(S_2) - \{v\}$, adding all edges of color 3 for all $x \in V(U')$ and $y \in V(S_3) - \{v\}$, and deleting all edges of S_2 and S_3 incident to v. Let H_1 denote the component of G' containing S_1, S_2 and T, and let H_2 denote the component of G' containing $S_3 - \{v\}$ and U. Note that both H_1 and H_2 are bi-color components of G'.

Clearly, G' is the graph associated with a code $C' \subseteq Q^3$, where C' is obtained from C by changing the first coordinate of the codewords in C corresponding to vertices of S_1 to the fisrt coordinate of codewords in C corresponding to vertices of T', changing the second coordinate of the codewords in C corresponding to vertices of $S_2 - \{v\}$ to the second coordinate of codewords in C corresponding to vertices of T'', changing the third coordinate of the codewords in C corresponding to vertices of $S_3 - \{v\}$ to the third coordinate of codewords in C corresponding to vertices of U'.

Since H_1 and H_2 are bi-color components, H_1 and H_2 are IPP graphs by (i) and (ii) of Lemma 2.6. Since G is an IPP graph, $G - V(S \cup T \cup U)$ (if non-empty) is also an IPP graph. So by Lemma 2.5, G' is an IPP graph. Clearly, $|V(G')| = |V(G)|$. However, G' has fewer components than G, contradicting the choice of C and G. $\square$

G. Some Results from Nonlinear Programing

Let $\Omega \subseteq R^n$ and $g_i(\mathbf{x}) \leq 0$ be funtional constraints. A point $\mathbf{x} \in \Omega$ that satisfies all functional constraints is said to be *feasible*. An constraint $g_i(\mathbf{x}) \leq 0$ is said to be *active* at a feasible point $\mathbf{x}$ if $g_i(\mathbf{x}) = 0$, and *inactive* if $g_i(\mathbf{x}) < 0$.

Suppose $\mathbf{x}$ is a feasible point, and let J be the set of indices j for which $g_j(\mathbf{x}) = 0$. Then $\mathbf{x}$ is said to be a *regular point* of the constraints if the gradient vectors $\{\nabla g_j(\mathbf{x}), j \in J\}$ are linearly independent .

Khun-Tuker Conditions: *Suppose f and g_i, $i = 1, 2, \cdots, m$, are continuous and possess continuous second partial derivates. Let $\mathbf{x} \in \Omega$ be a local regular maximum point for the problem*

$$\begin{aligned} &\textit{maximize} \quad f(\mathbf{x}) \\ &\textit{subject to} \quad g_i(\mathbf{x}) \leq 0, i = 1, 2, \cdots, m. \end{aligned}$$

Then there is a vector $(\mu_1, \mu_2, \cdots, \mu_m)$ with $\mu_i \geq 0$ such that

$$\frac{\partial f(\mathbf{x})}{\partial x_j} - \sum_{i=1}^{m} \mu_i \frac{\partial g_i(\mathbf{x})}{\partial x_j} = 0,$$
$$\mu_i g_i(\mathbf{x}) = 0, i = 1, 2, \cdots, m.$$

REFERENCES

[1] N. Wagner, "Fingerprinting," *Proceedings of the 1983 IEEE Symposium on Security and Privacy*, pp. 18-22, April 1983.

[2] B. Chor, A. Fiat, M. Naor, and B. Pinkas, "Tracing traitors," *IEEE Trans. Inform. Theory*, vol. 46, pp. 893-910, May 2000.

[3] D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1897-1905, Sep. 1998.

[4] Wade Trappe, Min Wu, Z. Jane Wang, and K. J. Ray Liu, "Anti-collusion Fingerprinting for Multimedia," *IEEE Trans. Signal Processing.*, vol. 51, pp. 1069-1086, April 2003.

[5] Henk D. L. Hollmann, Jack H. Van Lint, Jean-Paul Linnartz and Ludo M. G. M. Tolhuizen, "On Codes with the Identifiable Parent Property," *J. Combi. Theory, Series A*, vol. 82, pp. 121-133, 1998.

[6] Noga Alon, Eldar Fischer and Mario Szegedy, "Parent-Identifying Codes," *J. Combi.Theory, Series A*, vol. 95, pp. 349-359, 2001.

[7] Marcel Fernandz, Miguel Soriano, "Decoding Codes with the Identifiable Parent Property," *Proceedings of the Seventh IEEE International Symposium on Computers and Communications*, pp. 1-6, 2002.

[8] J. Kerner and K. Marton, "New bounds for perfect hashing via information theory," *Europ. J. Combi.*, vol. 9, pp. 523-530, 1986.

[9] J. A. Bondy and U. S. R. Murty, *Graph theory with applications*, Macmillan Press Ltd, 1976.

[10] J. Korner and M. Lucertini, "Compressing inconsistent data," *IEEE Trans. Information. Theory* vol. 40, pp. 706-715, May, 1994.

[11] J. F. J. MacWilliams and N. J. Sloane, *The Theory of Error-Correcting Codes*, Elsevier, Amsterdam, 1997.

[12] David G. Luenberger, *Linear and Nonlinear Programming*, Second Edition, Kluwer Academic Publishers, 1987.