MIX-SPLIT: CONTROLLED MIXING OF SECRETS AND TRACEABLE PSEUDONYM GENERATION USING CODEBOOKS

*Kannan Karthik**

Indian Institute of Technology Guwahati
Dept. of Electronics and Electrical Engg.
Guwahati, Assam, India 781039

Dimitrios Hatzinakos

University of Toronto
Dept. of Electrical and Computer Engg.
Toronto, ON, Canada M5S3G4

ABSTRACT

A non-perfect secret sharing scheme called MIX-SPLIT is a substitution cipher created by mixing two statistically similar binary sequences (secrets) through a codebook. At the heart of the algorithm are the hidden partitions which define the identity of the shares generated. By imposing certain constraints on the codebook these partitions can be made invisible, opening up the possibility of constructing traceable pseudonyms which are inherently frameproof. These codes by virtue of their parental dependency (inheritance) can be applied towards both content authentication as well as tracking.

Index Terms— MIX-SPLIT, non-perfect, frameproof code, secret sharing

1. INTRODUCTION

In non-perfect secret sharing schemes the notion of an access structure is loosely defined. In such schemes subsets of shares derived from a particular parent secret have a tendency to leak out some information regarding the secret. So far non-perfect secret sharing schemes have been mostly of academic interest focussing primarily on the development of a generalized framework for representing such schemes. Ramp schemes were first proposed by Blakley and Meadows [1] in which three different types of sets were identified within the access structure: access sets which reveal full information about the secret, partial access sets which leak out some finite information about the secret and non-access sets which do not reveal anything about the secret. Since then, there has been a sufficient body of literature [2][3][4] directed towards the structure of non-perfect schemes but very little towards finding a set of feasible applications.

In our earlier work originating from [5], we were intent on positioning an interesting substitution cipher called MIX-SPLIT. The cipher created mixed shares of two different but statistically similar parent secrets by controlling the mixing through a codebook. Since the shares inherit the properties of the codebook, it was observed in [6] and [7] that one can construct anti-collusion codes, authentication codes and joint access schemes, opening up the possibility for a unified approach towards designing traitor tracing and non-perfect secret sharing schemes. Conditional entropy can used to quantify the extent of information leakage but, since, it is a scalar quantity it does not tell us what portion of the secret is revealed by the coalition of shares. For this a geometric view was constructed in [8] which allowed us to assign a direction to this leakage. This view is important since if each coalition reveals a unique portion of the secret it can be used to either construct a group authentication code or generate keys for selective access. In this paper we open up the MIX-SPLIT algorithm and study its application towards the construction of frameproof codes. The concept of a c-frameproof code originated in the classic paper by Boneh and Shah [9]. Here we propose an alternative but simplistic construction using MIX-SPLIT where every share generated represents a fingerprint which serves as a traceable pseudonym. Two parent secrets which are mixed to form the shares, are broken down into v hidden subsequences (also known as partitions). The identity of a share is buried in these partitions through a carefully concealed inheritance directed by a secret codebook. Without an unlocking sequence the partitions remain invisible and can thus be used to form a frameproof code. These codes later on will be shown to have a parental connection, allowing them to be used for content authentication also.

Note that in sharp contrast, a perfect secret sharing scheme is all *white*: i.e. the individual shares appear as white noise both with respect to the secret and also in relation to one another. None of the illegitimate coalitions can be affiliated to the parent secret. Hence there is no scope for pursuing traitor tracing within that framework.

In Section II, the basics of the MIX-SPLIT algorithm, some definitions and properties are presented. The notion called a hidden partition and conditions governing the visibility of these partitions is discussed as a simple collection of three rules in Section III. These rules are extended towards the construction of a short frameproof code in Section IV. Finally in Section V we close this paper with a collage of

*Corresponding author.

properties and potential applications for MIX-SPLIT based constructions.

2. MIX-SPLIT

Two compressed sources $Src(\vec{X})$ and $Src(\vec{Y})$ produce independent sequences of binary random variables, $\vec{X} = (x_1, x_2, .., x_L)$ and $\vec{Y} = (y_1, y_2, .., y_L)$ respectively. Since the sources are assumed to be fully compressed, $x_i \perp x_j$ and $y_i \perp y_j$, where $\perp$ implies orthogonality. Further we state that the symbols x_i and y_j are identically distributed, i.e., $\Pr(x_i = 1) = \Pr(y_j = 1) = 0.5$ and $\Pr(x_i = 0) = \Pr(y_j = 0) = 0.5$ $\forall i, j \in \{1, 2, .., L\}$. A simple substitution cipher has been proposed for generating a set of shares $S_i, i = 1, 2, ..n$ by mixing the information from $\vec{X}$ and $\vec{Y}$. Share generation comprises of two steps:

Step1: Partitioning

v hidden partitions of $\vec{X}$ are created, i.e. $\vec{X} = [\vec{X}_1||\vec{X}_2||...||\vec{X}_v]$, where, '||' is the string concatenation cum mix operator. Corresponding to the L bits in $\vec{X}$ there are L unique bit positions which are captured by the set $P = \{1, 2, 3, ...L\}$. This set P is subdivided into v disjoint sets $P_1, P_2, ..P_v$ such that $P = P_1 \cup P_2 \cup ..P_v$, where $P_j \subset P$. For a given set $P_j, j = 1, 2, ..v$, $\vec{X}_j = \vec{X}(P_j)$ represents the information corresponding to the bit positions specified in P_j. $P_1, P_2, ...$ are chosen randomly to ensure good mixing of information.

Example for partitioning:

$\vec{X} = [x_1, x_2, x_3, x_4, x_5, x_6]$ is the chosen secret, with $x_i \in \{0, 1\}$ and the set of all possible bit positions as, $P = \{1, 2, 3, 4, 5, 6\}$. Let $v = 3$ be the number of partitions. Choice of random partitions: $P_1 = \{1, 5\}$, $P_2 = \{3, 6\}$ and $P_3 = \{2, 4\}$. The corresponding segments of $\vec{X}$ are: $\vec{X}_1 = \vec{X}(P_1) = [x_1, x_5]$, $\vec{X}_2 = \vec{X}(P_2) = [x_3, x_6]$ and $\vec{X}_3 = \vec{X}(P_3) = [x_2, x_4]$. Thus we shall define the string mix cum concatenation operator '||', as, $\vec{X} = [\vec{X}_1||\vec{X}_2||\vec{X}_3]$.

Similarly, v partitions of $\vec{Y}$ is created as $\vec{Y}_j = \vec{Y}(P_j)$ for $j = 1..v$.

Step2: Mixing $\vec{X}$ and $\vec{Y}$ and splitting into n shares

Each of the n shares can be written as,

$$S_i = (S_{i1}||S_{i2}||...||S_{iv}) \tag{1}$$

where, the sequence S_{ij} is chosen according to a pre-designed codebook.

$$\begin{aligned} S_{ij} &= \vec{X}_j \text{ if } c_{ij} = 1 \\ S_{ij} &= \vec{Y}_j \text{ if } c_{ij} = 0 \end{aligned} \tag{2}$$

The binary value $c_{ij} \in \{0, 1\}$ is a part of the codebook,

$$\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1v} \\ c_{21} & c_{22} & \cdots & c_{2v} \\ \cdots & \cdots & \cdots & \cdots \\ c_{n1} & c_{n2} & \cdots & c_{nv} \end{pmatrix} \tag{3}$$

where, n represents the number of users and v the number of partitions. If w is the hamming weight of the codeword $[c_{i1}, c_{i2}, ..., c_{iv}]$ for some row i in $\mathbf{C}$, the proportion in which the two secrets $\vec{X}$ and $\vec{Y}$ are mixed is determined by the ratio $w : (v - w)$ or the parameter $\alpha = w/v$.

2.1. Hidden partitions and some definitions

The secrets which are concealed comprise of v hidden subsequences, $\vec{X}(P_j)$ and $\vec{Y}(P_j)$, where P_j are the partitions created at the start of the share generation process. As long as these partitions are hidden, each of the subsequences will also remain hidden. It is proposed to reveal some of these partitions through a share collusion process. But first we have to examine the circumstances under which these partitions can be kept invisible. For the next two sections we will enforce a condition that the binary sequence $\vec{Y}$ is derived from $\vec{X}$ by taking its bit complement. For instance if $\vec{X} = [1, 0, 0, 1, 1]$, $\vec{Y} = BIT\_CMP(\vec{X}) = [0, 1, 1, 0, 0]$, where the $BIT\_CMP(.)$ is the bit-complement function. This constraint allows us to control the visibility of the partitions through a plain share collusion, a feature that is important for authentication and fingerprinting.

Every bit string contained in the partitions is mapped to a real value in the interval $[0, 1]$ and is referred to as a *co-ordinate*. Note here that the knowledge of a particular partition completely specifies the co-ordinate. Hence the hidden components of the secret are essentially the *partitions* of the subsequences. Let $\vec{X}_j = \vec{X}(P_j) = [x_{j_1}, x_{j_2}, x_{j_3}, ..., x_{j_{L_p}}]$, where $x_{j_k} \in \{0, 1\}$. The co-ordinate corresponding to the partition P_j is,

$$\begin{aligned} a_j &= \text{String2Point}(\vec{X}_j) \\ &= \sum_{r=1}^{L_p} x_{j_r} \cdot \left[\frac{1}{2}\right]^r \end{aligned} \tag{4}$$

where, $L_p = L/v$ is the length of each partition. The ordered set of co-ordinates $[a_1, a_2, .., a_v]$ is a point inside a v-dimensional UNIT hypercube. Note that an arbitrary collusion of shares may reveal any $k < v$ co-ordinates which implies that the secret could lie anywhere amongst a set of discrete points on a $(v-k)$ dimensional hyperplane bound by the unit hypercube. As the number of shares in the coalition is increased the region of uncertainty will shrink. This is a perfect example of a non-perfect secret sharing scheme in which the conditional entropy $H(\vec{X}|B_t) < H(\vec{X})$ where, B_t is a coalition of some t shares. Unfortunately conditional entropy is a

scalar quantity and does not tell us which portion of the secret is leaked out by a particular coalition. For this we may have to construct a geometric view discussed in [8].

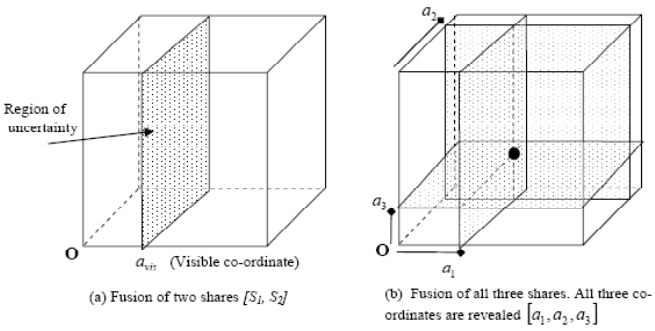
Example: For a 3-out-of-3 sharing scheme with a $\vec{X} : \vec{Y}$ mix ratio of 2:1, we can use the codebook,

$$\mathbf{C} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \tag{5}$$

The secret point $[a_1, a_2, a_3]$ becomes visible when all three shares are stacked and a three-way bit comparison across columns of the binary sequences is executed to extract the bit-locations of the subsequences. Note that each of the three columns in the codebook indicates a unique association between the three shares which leaks out the partitions P_A, P_B and P_C by virtue of a columnwise bit comparison as shown below.

$$\begin{aligned} P_A = P_1 &= \{r_A\}, \text{ s.t. } [(S_1(r_A) = S_2(r_A)] \neq S_3(r_A) \\ P_B = P_2 &= \{r_B\}, \text{ s.t. } [S_1(r_B) = S_3(r_B)] \neq S_2(r_B) \\ P_C = P_3 &= \{r_C\}, \text{ s.t. } [S_2(r_C) = S_3(r_C)] \neq S_1(r_C) \end{aligned} \tag{6}$$

On the other hand a coalition of any two shares will reveal exactly one co-ordinate a_i which is called the *visible* co-ordinate confining the secret to a plane parallel to one of the axes as shown in Fig. 1(a). Thus the same example also represents



(a) Fusion of two shares $[S_i, S_j]$

(b) Fusion of all three shares. All three co-ordinates are revealed $[a_1, a_2, a_3]$

Fig. 1. Geometric interpretation of MIX-SPLIT. Illustration of selective leakage of information through share fusion. When two shares are fused, the secret is confined to a plane parallel to the X-axis. When all three shares are stacked, the point having co-ordinates $[a_1, a_2, a_3]$ is revealed [8].

a (2,3)-selective access code. In such applications the partitions can be destroyed immediately after the shares are created. Simply stacking and performing a bitwise comparison will reveal the hidden partitions.

3. CONDITIONS FOR PARTITION INVISIBILITY

A codebook (or a portion of it) may be designed in such a way that no matter which subset of shares are stacked, virtually no information is revealed regarding the partitions. Let $\mathbf{A}$ be a subdesign matrix obtained by row-sampling the original codebook $\mathbf{C}$, representing a subset of shares $B_t = \{S_{i_1}, S_{i_2}, .., S_{i_t}\}$, where $i_k \in \{1, 2, .., n\}$. Each block matrix $\mathbf{A}$ can in turn be decomposed into columns, $\mathbf{A} = [\bar{w}_1, \bar{w}_2, .., \bar{w}_v]$, where $\bar{w}_j$ is a $t \times 1$ binary column vector. We may define the set $\mathbf{VC}(B_t(\mathbf{A}))$ as the set of visible co-ordinates (or partitions) obtained by stacking and comparing the coalition of t shares defined by matrix $\mathbf{A}$. The following three rules govern the invisibility of the partitions.

Rule 1: Complementary columns lead to inseparable partitions

$\mathbf{VC}[B_t(\mathbf{A})] = \emptyset$ IF for every $\bar{w}_j \in \mathbf{A}$, $BIT\_CMP[\bar{w}_j] \in \mathbf{A}$ even though $\bar{w}_j$ may be distinct.

Example 1:

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix} \tag{7}$$

In the above example, when users stack their three shares $S_{i_1}, S_{i_2}, S_{i_3}$, two sets of positions can be disclosed defined by the sets P_A and P_B,

$$\begin{aligned} P_A &= \{r_A\}, \text{ s.t. } [(S_{i_1}(r_A) = S_{i_2}(r_A) = S_{i_3}(r_A)] \\ P_B &= \{r_B\}, \text{ s.t. } S_{i_1}(r_B) \neq [S_{i_2}(r_B) = S_{i_3}(r_B)] \end{aligned}$$

where, $P_A = P_1 \cup P_2$ and $P_B = P_3 \cup P_4$. Since one cannot separate P_1 and P_2 from $P_1 \cup P_2$ this leaves the two partitions mixed. A similar argument can be constructed for P_3 and P_4. In general depending on particular column $\bar{w}_j$ (combination of ones and zeroes) one can form a unique relation between the shares in the stack. Complementing this column does not alter the relationship. Thus a pair of complementary columns result in a mixed set of partitions. In this example, there are no visible co-ordinates, i.e. $\mathbf{VC}[B_t(\mathbf{A})] = \emptyset$. Since each stack comparison operation tends to narrow down the search for the partitions $P_i, i = 1, 2, 3, 4$ this is a non-perfect scheme. However for a sufficiently large set $P_A = P_1 \cup P_2$ it is very difficult to split this into two constituent partitions P_1, P_2 without prior information. Thus no co-ordinates are visible from complementary patterns.

Rule 2: Rowsampling of a complementary pattern is complementary

IF $\mathbf{VC}[B_t(\mathbf{A})] = \emptyset$, THEN $\mathbf{VC}[B_{t-1}(\mathbf{B})] = \emptyset$
where, $\mathbf{B}$ is a sub-block obtained by rowsampling of $\mathbf{A}$.

Example 2:

$$\mathbf{B} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \tag{8}$$

By observation for coalitions of size $t \geq 3$, if the parent coalition $\mathbf{A}$ comprises of complementary columns (Example 1), the rowsampled matrix $\mathbf{B}$, derived from $\mathbf{A}$, will also be made up of complementary columns. Thus it follows from Rule 1, that $\mathbf{VC}[B_{t=2}(\mathbf{B})] = \emptyset$. There will be no visible co-ordinates.

Rule 3: Single share is always mixed (no partitions visible)

$\mathbf{VC}[B_{t=1}(\mathbf{A})] = \emptyset$ irrespective of the choice of $\mathbf{A}$.
Since a single share is a mixture of $\vec{X}$ and $\vec{Y} = BIT\_CMP(\vec{X})$, which are statistically similar, there is no way one can separate the partitions P_j from the mixture. As a result the co-ordinates will always remain invisible for a single share.

4. FRAMEPROOF CODE CONSTRUCTION: THE UNLOCKING SEQUENCE

We consider an application of the above three rules. First note that in a MIX-SPLIT based construction a share may comprise of the following two components:

- Subsequences $\vec{X}(P_j)$ and $\vec{Y}(P_j)$, $j = 1, 2, .., v$.
- Codebook which decides the exact choice of subsequences and inheritance.

These components together make up the signature of a particular share. The codebook and partitions are destroyed immediately after the co-ordinates $a_j = \text{String2Point}(\vec{X}(P_j))$ are determined prior to share distribution. These co-ordinates are retained for verification. If the shares $S_i, i = 1, 2, ..n$ represent fingerprints of n users, it is expected that no point should it be possible for one user (say user i) to frame another (user j) even if he/she manages to acquire his/her exact fingerprint code sequence S_i. To generalize this to a coalition of $t < n$ users, it should not be possible for a coalition t or fewer traitors to frame a user outside the group. Since the heart of a MIX-SPLIT share is a partition, any illegitimate coalition should not be able to determine the partitions. However these partitions (some or all) should be revealed when a suitable unlocking sequence is provided and this can be used not only for tracing but also for content authentication as the shares possess an identifiable parent property (IPP).

We present a n-frameproof code with $n = 4$. Note that the this codebook satisfies rules 1-3, i.e, every column has a complementary counterpart. The codebook is,

$$\mathbf{C}_4 = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \end{pmatrix} \tag{9}$$

and the corresponding frameproof fingerprints are,

$$\begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \end{pmatrix} = \begin{pmatrix} \vec{X}(P_1) & \vec{X}(P_2) & \vec{X}(P_3) & \vec{Y}(P_4) & \vec{Y}(P_5) & \vec{Y}(P_6) \\ \vec{X}(P_1) & \vec{Y}(P_2) & \vec{Y}(P_3) & \vec{X}(P_4) & \vec{X}(P_5) & \vec{Y}(P_6) \\ \vec{Y}(P_1) & \vec{Y}(P_2) & \vec{X}(P_3) & \vec{X}(P_4) & \vec{Y}(P_5) & \vec{X}(P_6) \\ \vec{Y}(P_1) & \vec{X}(P_2) & \vec{Y}(P_3) & \vec{Y}(P_4) & \vec{X}(P_5) & \vec{X}(P_6) \end{pmatrix} \tag{10}$$

where, $\vec{Y} = BIT\_CMP[\vec{X}]$. Note that this incidentally is also a ($K = 3$, $n = 4$) anti-collusion code useful for detecting three or fewer colluders amongst a small user group of four users [7]. Take any coalition, say of some three shares $B_{t=3} = \{S_1, S_3, S_4\}$, the corresponding block matrix $\mathbf{A}$ also satisfies Rule 1. Note that every column in $\mathbf{A}$ has a complementary counterpart e.g. *col-1* with *col-6*, *col-2* with *col-4* and *col-3* with *col-5*.

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \end{pmatrix} \tag{11}$$

This holds true even for coalitions of size $t = 1, 2$, i.e. $\mathbf{VC}[B_t(\mathbf{A})] = \emptyset$, $t \leq n$. However if one appends a code $[1, 1, 1, 1, 1, 1]$ to $\mathbf{A}$, the three rules are no longer applicable and one can extract the partitions by stacking the four sequences. Note here that the fourth sequence which serves as the unlocking sequence is the parent $\vec{X}$ itself (which lies with the distributor or source).

$$\mathbf{B}_{app} = \begin{pmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \end{pmatrix} \tag{12}$$

Observe that every column in the new matrix $\mathbf{B}_{app}$ is unique, hence six different stack relations will yield all the six partitions $P_1, P_2, .., P_6$. This holds true for any three out of four shares. On the other hand when the unlocking sequence is applied to a stack of two shares exactly two partitions are revealed. This unlocking aspect is illustrated using Table. 1. Notice that with the help of the unlocking sequence $\vec{X} \equiv [1, 1, 1, 1, 1, 1]$ the identity of each one of the individual shares can be confirmed as they can be uniquely linked with $\vec{X}$ (uncertainty region is a unique cluster of points within a 6-dimensional unit hypercube). The first row in Table. 1 shows that standalone coalitions of shares (without the unlocking sequence $\vec{X}$) will not release any partition (Rules 1-3).

Table 1. Visible partitions for the $\mathbf{C}_4$ frameproof code.

Information stacked	Visible partitions $\mathbf{VC}(B_t)$
$B_t \subset \{S_1, S_2, S_3, S_4\}$ ($t \leq 4$)	$\emptyset$ (Rules 1-3)
$\{S_1, S_2, \vec{X}\}$	P_1, P_6
$\{S_1, S_3, \vec{X}\}$	P_3, P_5
$\{S_1, S_4, \vec{X}\}$	P_2, P_4
$\{S_2, S_3, \vec{X}\}$	P_4, P_2
$\{S_2, S_4, \vec{X}\}$	P_5, P_3
$\{S_3, S_4, \vec{X}\}$	P_6, P_1
$B_{t=3}, \vec{X}$	$P_1, P_2, P_3, P_4, P_5, P_6$

5. CONCEALED INHERITANCE AND OTHER PROPERTIES OF MIX-SPLIT

In a MIX-SPLIT construction, the shares and the parent secrets form a dependence graph as shown in Fig. 2. The codebook which is used to mix $\vec{X}$ and $\vec{Y}$ creates different realizations of the parents. These realizations can be termed as

children as they inherit the characteristics of the parents. We revert to the original definition in Section. 2, which starts with the assumption that $\vec{X}$ and $\vec{Y}$ are I.I.D. sequences of random variables (analogy: referred to as traits here). The partitions $P_j \subset \{1, 2, 3, .., L\}$ define a group of traits which are collectively inherited by a share (or child node). These traits are concealed at the time of birth (i.e. while creating the shares). Consequently the exact identity of the parents are hidden from the shareholders or children (anonymity). This is a special case of concealed inheritance where every standalone share is an anonymous representation of two unknown parents. However in the presence of the parents one can establish a link between the shares and the secrets.

Given the pairs $[\vec{X}, S]$ and $[\vec{Y}, S]$ and the bi-polar one-to-one mappings $\vec{X}_b = 2\vec{X} - \bar{1}, \vec{Y}_b = 2\vec{Y} - \bar{1}$ and $S_b = 2S - \bar{1}$,

$$\begin{aligned} \text{Corr}(\vec{X}_b, S_b) &= \frac{\vec{X}_b \cdot [\vec{S}_b]^T}{L} \approx \alpha \\ \text{Corr}(\vec{Y}_b, S_b) &= \frac{\vec{Y}_b \cdot [\vec{S}_b]^T}{L} \approx 1 - \alpha \end{aligned}$$

S can be easily linked to its parents $\vec{X}$ and $\vec{Y}$ since the codebook controls the correlation between the derived share S and the secret. Here, $\bar{1}$ represents the all one vector. However, a stand-alone share S does not reveal its heritage as it is a mixture of two statistically similar sequences, i.e. parents remain anonymous. This duality associated with the MIX-SPLIT constructions opens up applications such as e-Voting.

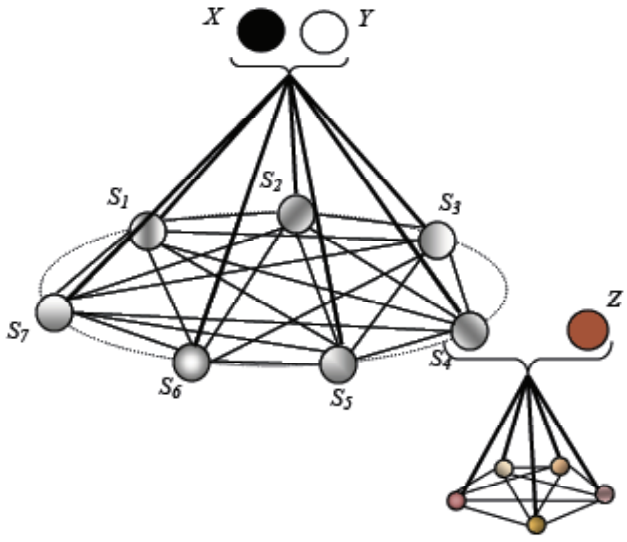


Fig. 2. Multicolored shares and dependencies.

Fig. 2 *encompasses* some of the following application scenarios:

- **Selective access tokens:** Each node S_i may represent a unique colored sequence from $\vec{X}, \vec{Y}$. S_i and its children can be used to generate keys for selective access [6]. An example is the (2,3) selective access code discussed in Section. 2.1 where every 2-coalition (out of a total of three users) leaks out a unique component of the secret (i.e. a specific partition becomes visible). Thus each share becomes an access token and a subset of such tokens can be used for releasing a key to a portion of a secured sensitive record. Fine grained access control can thus be a potential application.
- **Anti-collusion codes:** Since every child node (share) is a unique realization obtained by mixing two parents they can represent fingerprints for traitor tracing. A careful association can be built between the child shares and this inter-connection becomes essential for constructing anti-collusion codes. This aspect is discussed in brief in Section. 5.1.
- **Authentication with traceability:** Parental dependency on the other hand allows authentication and cross-linking. Since the child nodes inherit some of the parent's traits, each share can be used both as an authentication watermark as well as a fingerprint for tracking a particular document.
- **Anonymity, joint authentication and tracing:** The simultaneous mixing and splitting of two disparate data sets can be used for anonymous fingerprinting in which the [creator, buyer] pair is imprinted in copies of some digital portrait without revealing either identity [6]. Here we apply the property that a share S which is derived by mixing two statistically similar secrets $\vec{X}$ and $\vec{Y}$ becomes a traceable pseudonym.
- **Joint access with tracing:** Since MIX-SPLIT was originally designed as a n-out-of-n non-perfect secret sharing scheme, one imminent application is that of joint access with tracking of all illegitimate share fusions B_t s.t. $t < n$. For images, which cannot tolerate even moderate distortion levels, this scheme provides dual protection. A 3-out-of-3 or 5-out-of-5 "access control + tracing scheme" can be implemented using the codebooks [7],

$$\mathbf{C}_{3,3} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \tag{13}$$

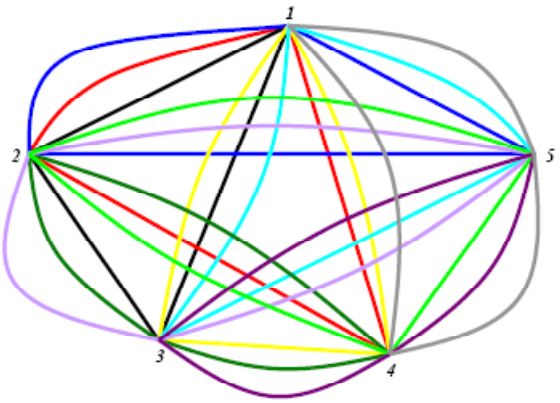
$$\mathbf{C}_{5,5} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \tag{14}$$

To reconstruct retrieve the secrets, all the n shares need to stacked and a columnwise majority bit vote evaluated to extract the parent $\vec{X}$ and a minority vote to extract $\vec{Y}$.

5.1. Associations between shares

Contrary to a perfect secret sharing scheme, through MIX-SPLIT it is possible to induce dependencies between shares. It is possible to create a unique association between any subset of shares through a specific codebook, which can be utilized for traitor tracing. The 5-out-of-5 codebook (Eqn. 14) [7] can be depicted as an edge colored graph (Fig. 3) in which each share $S_i, i \in \{1,2,3,4,5\}$ is represented by a vertex and each edge is an *association* between two shares. The linking may take place through any of the $v = 10$ different dimensions (corresponds to the 10 colors used in the graph). Corresponding to these v colors, the secrets $\vec{X}$ and $\vec{Y}$ are subdivided into v disjoint parts. Any share S_i comprises of 6 parts from $\vec{X}$ and 4 parts from $\vec{Y}$. Each part of $\vec{X}$ (represented as color $V_j, j = 1,2,..10$) is mapped as '1' and that of $\vec{Y}$ as '0' which in turn can be treated as the absence of $\vec{X}$.

It is easy to see that this codebook mimics an anti-collusion code and by design a majority vote of any subset of codewords will result in a unique binary pattern. Illegitimate coalitions can be represented by a subgraph $G(ST)$, induced by the vertices $ST \subset \{1,2,3,4,5\}$.



	Dimensions or Marks or Colors									
Shares	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}
S_1	1	1	1	1	1					1
S_2	1	1	1			1	1	1		
S_3	1			1	1	1	1		1	
S_4		1		1		1		1	1	1
S_5			1		1		1	1	1	1

Each dimension or color V_i, connects exactly 3 different shares $[S_i, S_j, S_k]$ where $i, j, k \in \{1,2,3,4,5\}$. The blank spots are '0'.

Fig. 3. Graph indicating associations between different shares for a 5-out-of-5 non-perfect secret sharing scheme. The construction also mimics an anti-collusion code for tracing all subsets of traitors within a group of 5 users.

6. CONCLUSIONS

When perfect secrecy is compromised in secret sharing, this opens up a pandora's box which can be used to create a variety of schemes such as anti-collusion codes, authentication codes, joint access schemes and traceable pseudonyms. The focal point of this paper is in the construction of traceable pseudonyms for fingerprinting through a non-perfect secret sharing scheme called MIX-SPLIT. A set of rules are discussed to make the construction frameproof.

7. REFERENCES

[1] G. R. Blakley and C. Meadows, "Security of ramp schemes," in *Proceedings of CRYPTO 84 on Advances in cryptology*, New York, NY, USA, 1985, pp. 242–268, Springer-Verlag New York, Inc.

[2] K. Kurosawa, K. Okada, K. Sakano, W. Ogata, and S. Tsujii, "Nonperfect secret sharing schemes and matroids," *Lecture Notes in Computer Science*, vol. 765, pp. 126–141, 1994.

[3] W. Ogata and K. Kurosawa, "Some basic properties of general nonperfect secret sharing schemes," *Jour. of Universal Comp. Science*, vol. 4, no. 8, pp. 690–704, 1998.

[4] M. Iwamoto and H. Yamamoto, "Strongly secure ramp secret sharing schemes for general access structures," *Info. Processing Letters*, vol. 97, no. 2, pp. 52–57, 2006.

[5] K. Karthik, "Methodologies for access control and fingerprinting of multimedia," *Doctoral Thesis, University of Toronto, Department of Electrical and Computer Engineering*, 2006.

[6] K. Karthik and D. Hatzinakos, "A Unified approach to construct Non-perfect Secret Sharing and Traitor Tracing schemes," *Proceedings of International Conference on Security and Management (SAM'07)*, June 25-28, Las Vegas, Nevada, 2007.

[7] K. Karthik and D. Hatzinakos, "Multimedia Encoding for Access Control with Traitor Tracing: Balancing Secrecy, Privacy and Traceability," *VDM Verlag Dr. Muller*, 2008, ISBN: 978-3-8364-3638-0.

[8] K. Karthik and D. Hatzinakos, "Secure group authentication using a non-perfect secret sharing scheme based on controlled mixing," in *Proceedings of INDICON'09*, Gandhinagar, India, 2009.

[9] D. Boneh and J. Shaw, "Collusion secure fingerprinting for digital data," *IEEE Transactions on Information Theory*, vol. 44, pp. 1897–1905, Sept 1998.