Definition of IPP Code

- an (ℓ, n, q)-code is a w-IPP code if and only if

$$\bigcap_{\mathcal{C}_0 \in \mathsf{susp}_w(\mathbf{f})} \mathcal{C}_0 \neq \emptyset$$

  for all $\mathbf{f} \in \mathsf{desc}_w(\mathcal{C})$

- given any hybrid fingerprint $\mathbf{f}$ which was created by a coalition of size at most w in a w-IPP code, any codeword in the above-defined intersection is an *identifiable parent*

- the w-IPP property means that at least one parent is identifiable

- $\mathcal{C}$ is a 2-IPP code *if and only if* $A(\mathcal{C})$ is simultaneously a PHF$(\ell; n, q, 3)$ and an SHF$(\ell; n, q, \{2, 2\})$

Results on 2-IPP Codes From [HvLLT]

- let $n^*(\ell, q, w)$ denote the maximum integer n such that an (ℓ, n, q) w-IPP code exists
- $n^*(1, q, 2) = q$
- $n^*(2, q, 2) = q$
- $3q - 12\sqrt{q} \le n^*(3, q, 2) \le 3q - 1$
- $n^*(4, q, 2)$ is $\Omega(q\sqrt{q})$ and $O(q^2)$ (note: a recent result by Alon, Fischer and Szegedy shows that n^* is $\Omega(q^{2-\epsilon})$ for all $\epsilon > 0$)
- $n^*(5, q, 2)$ is $\Theta(q^2)$
- $n^*(\ell, q, 2)$ is $\Omega(q^{\lceil \ell/4 \rceil})$
- $n^*(\ell, q, 2)$ is $O(q^{\lceil \ell/3 \rceil})$
- $n^*(\ell, q, 2)$ is an exponential function of ℓ (for fixed q)

w-IPP CODES

- there is no known "nice" characterization of a w-IPP code for arbitrary $w > 2$
- some partial results are known:
  - an (ℓ, n, q)-code is a w-IPP code *only if* the $n \times \ell$ array of codewords is simultaneously a PHF$(\ell; n, q, w+1)$ and an SHF$(\ell; n, q, \{w, w\})$
  - an (ℓ, n, q)-code is w-IPP *if* the $n \times \ell$ array of codewords is a PHF$(\ell; n, q, \lfloor (w+2)^2/4 \rfloor)$

Tracing Illegally Redistributed Keys

- suppose that every user in a network is given a "decoder box" that allows encrypted broadcasts to be decrypted
- i.e., we have a BES in which every user can decrypt the broadcast
- every decoder box contains a different collection of keys
- a coalition of w bad guys might create a "pirate decoder" by combining keys from their decoder boxes
- the keys in each decoder box can be thought of as a codeword in a certain code, and the keys in a pirate decoder can be thought of as a codeword in the w-descendant code
- if the code is traceable (e.g., if it satisfies the w-IPP property), then a pirate decoder can be traced back to at least one member of the coalition that created it

The Broadcast Encryption Scheme

- the TA chooses ℓ sets of keys, denoted $\mathcal{K}_1, \ldots, \mathcal{K}_\ell$, where each $\mathcal{K}_i$ consists of q keys chosen from $\mathbb{Z}_m$, for some fixed m
- for $1 \leq i \leq \ell$, let $\mathcal{K}_i = \{k_{i,j} : 1 \leq j \leq q\}$
- a decoder box contains ℓ keys, one from each set $\mathcal{K}_i$
- the secret key $K \in \mathbb{Z}_p$ (which is used to encrypt the broadcast content, M) is split into ℓ shares using an (ℓ, ℓ) threshold scheme
- the shares are $s_1, \ldots, s_\ell$, where $s_1 + \cdots + s_\ell \equiv K \pmod{m}$
- K is used to encrypt M, and for $1 \leq i \leq \ell$, every $k_{i,j}$ is used to encrypt s_i
- the entire broadcast consists of

$$e_K(M) \quad \text{and} \quad (e_{k_{i,j}}(s_i) : 1 \leq i \leq \ell, 1 \leq j \leq q)$$

Decoder Boxes and Codes

- every user can perform the following operations:
  1. decrypt all ℓ shares of K,
  2. reconstruct K, and
  3. decrypt M
- each decoder box corresponds to a codeword $\mathbf{c} \in Q^\ell$, where $Q = \{1, \ldots, q\}$, in an obvious way:

$$\{k_{1,j_1}, k_{2,j_2}, \ldots, k_{\ell,j_\ell}\} \leftrightarrow (j_1, j_2, \ldots, j_\ell)$$

- the keys in a pirate decoder form a codeword in the w-descendant code

Variations on a Theme: Efficient Tracing

- we now study a special class of w-IPP codes that have very efficient tracing algorithms
- let $\mathsf{dist}(\cdot,\cdot)$ denote the hamming distance between two vectors
- for $\mathbf{f} \in \mathsf{desc}_w(\mathcal{C})$, the *nearest neighbour* to $\mathbf{f}$ is a codeword $\mathbf{c} \in \mathcal{C}$ such that $\mathsf{dist}(\mathbf{f}, \mathbf{c})$ is minimized
- denote the nearest neighbour to $\mathbf{f}$ by $\mathsf{neighbour}(\mathbf{f})$
- $\mathsf{neighbour}(\mathbf{f})$ can be found in time $O(n)$ by exhaustive search
- the code is said to be a *w-TA code* if

$$\mathsf{neighbour}(\mathbf{f}) \in \bigcap_{\mathcal{C}_0 \in \mathsf{susp}_w(\mathbf{f})} \mathcal{C}_0$$

for all $\mathbf{f} \in \mathsf{desc}_w(\mathcal{C})$

An Example

We present a $(5, 16, 4)$ 2-TA code:

$$\begin{array}{ll}
\mathbf{c}_1 = (1,1,1,1,1) & \mathbf{c}_2 = (1,2,2,2,2) \\
\mathbf{c}_3 = (1,3,3,3,3) & \mathbf{c}_4 = (1,4,4,4,4) \\
\mathbf{c}_5 = (2,1,2,3,4) & \mathbf{c}_6 = (2,2,1,4,3) \\
\mathbf{c}_7 = (2,3,4,1,2) & \mathbf{c}_8 = (2,4,3,2,1) \\
\mathbf{c}_9 = (3,1,4,2,3) & \mathbf{c}_{10} = (3,2,3,1,4) \\
\mathbf{c}_{11} = (3,3,2,4,1) & \mathbf{c}_{12} = (3,4,1,3,2) \\
\mathbf{c}_{13} = (4,1,3,4,2) & \mathbf{c}_{14} = (4,2,4,3,1) \\
\mathbf{c}_{15} = (4,3,1,2,4) & \mathbf{c}_{16} = (4,4,2,1,3)
\end{array}$$

An Example (cont.)

- suppose that the coalition $\mathcal{C}_0 = \{\mathbf{c}_5, \mathbf{c}_{11}\}$
- we have

$$\begin{aligned} \mathbf{c}_5 &= (2,1,2,3,4), \\ \mathbf{c}_{11} &= (3,3,2,4,1), \quad \text{and} \\ \mathbf{f} &= (2,3,2,4,4) \end{aligned}$$

  is in $\mathsf{desc}(\mathcal{C}_0)$
- $\mathsf{dist}(\mathbf{f}, \mathbf{c}_5) = \mathsf{dist}(\mathbf{f}, \mathbf{c}_{11}) = 2$
- $\mathsf{dist}(\mathbf{f}, \mathbf{c}_i) \geq 3$ for all $i \neq 5, 11$
- hence $\mathbf{c}_5$ and $\mathbf{c}_{11}$ are both identifiable parents of $\mathbf{f}$

Codes with Large Hamming Distance are TA Codes

- suppose that $\mathcal{C}$ is an (ℓ, n, q)-code having minimum hamming distance $d > \ell(1 - 1/w^2)$
- define $\mathsf{match}(\mathbf{c}_i, \mathbf{c}_j) = \ell - \mathsf{dist}(\mathbf{c}_i, \mathbf{c}_j)$
- let $\mathcal{C}_0 = \{\mathbf{c}_1, \ldots, \mathbf{c}_w\}$ and let $\mathbf{f} \in \mathsf{desc}(\mathcal{C}_0)$
- there exists $\mathbf{c}_i \in \mathcal{C}_0$ such that $\mathsf{match}(\mathbf{f}, \mathbf{c}_i) \geq \ell/w$
- hence, $\mathsf{dist}(\mathbf{f}, \mathbf{c}_i) \leq \ell - \ell/w$
- let $\mathbf{b} \in \mathcal{C} \backslash \mathcal{C}_0$
- then
$$\mathsf{match}(\mathbf{f}, \mathbf{b}) \leq \sum \mathsf{match}(\mathbf{c}_j, \mathbf{b}) \leq w(\ell - d)$$
- hence, $\mathsf{dist}(\mathbf{f}, \mathbf{b}) \geq \ell - w(\ell - d)$
- $w(\ell - d) < \ell/w$, so the code is a w-TA code

A Construction for TA Codes

- suppose q is prime, $\ell \leq q$ and $t < q$
- define $\mathcal{P}(q, \ell, t)$ to consist of all polynomials $a(x) \in \mathbb{Z}_q[x]$ having degree at most $t-1$
- define

$$\mathcal{C}(q, \ell, t) = \{(a(0), a(1), \ldots, a(\ell-1)) : a(x) \in \mathcal{P}(q, \ell, t)\}$$

- then $\mathcal{C}(q, \ell, t)$ is an (ℓ, q^t, q)-code having minimum hamming distance $d = \ell - t + 1$
- define $t = \lceil \frac{\ell}{w^2} \rceil$; then $\frac{\ell}{w^2} \leq \lceil \frac{\ell}{w^2} \rceil < \frac{\ell}{w^2} + 1$
- then $d > \ell(1 - 1/w^2)$, and hence we have a w-TA code with $n = q^{\lceil \frac{\ell}{w^2} \rceil}$
- when $w = 2$, we get a w-TA code with $n = q^{\lceil \frac{\ell}{4} \rceil}$

Variations on a Theme: Preventing Framing

- an (ℓ, n, q)-code, $\mathcal{C}$, is a *w-frameproof code* (w-FP code) if no coalition of size w can "frame" a user not in the coalition (i.e., if $\mathbf{f} \in \mathcal{C}$, then $\mathbf{f}$ is a member of every suspect coalition)
- more formally, the code is a w-FP code if $\mathbf{c} \in \mathcal{C}_0$ for all $\mathbf{c} \in \mathcal{C}$ and all $\mathcal{C}_0 \in \mathsf{susp}_w(\mathbf{c})$
- an (ℓ, n, q)-code is a *w-secure frameproof code* (w-SFP code) if no coalition of size w can "frame" a disjoint coalition of size w (i.e., any suspect coalition contains at least one guilty user)
- more formally, the code is a w-SFP code if $\mathcal{C}_0 \cap \mathcal{C}_1 \neq \emptyset$ for all $\mathbf{f} \in \mathsf{desc}_w(\mathcal{C})$ and all $\mathcal{C}_0, \mathcal{C}_1 \in \mathsf{susp}_w(\mathbf{f})$
- for example, in a 2-SFP code, it could happen that $\mathsf{susp}_w(\mathbf{f}) = \{\{\mathbf{c}_i, \mathbf{c}_j\}, \{\mathbf{c}_i, \mathbf{c}_k\}, \{\mathbf{c}_j, \mathbf{c}_k\}\}$

Preventing Framing (cont.)

- w-FP and w-SFP codes can be characterized in terms of hash families
- an (ℓ, n, q)-code is a w-FP code *if and only if* the $n \times \ell$ array of codewords is an $\mathrm{SHF}(\ell; n, q, \{w, 1\})$
- an (ℓ, n, q)-code is a w-SFP code *if and only if* the $n \times \ell$ array of codewords is an $\mathrm{SHF}(\ell; n, q, \{w, w\})$
- it is clear from the definitions (and the characterizations we have provided) that

$$w\text{-IPP} \Rightarrow w\text{-SFP} \Rightarrow w\text{-FP}$$

- w-FP and w-SFP codes are "easier" to construct than w-IPP codes because they satisfy weaker conditions