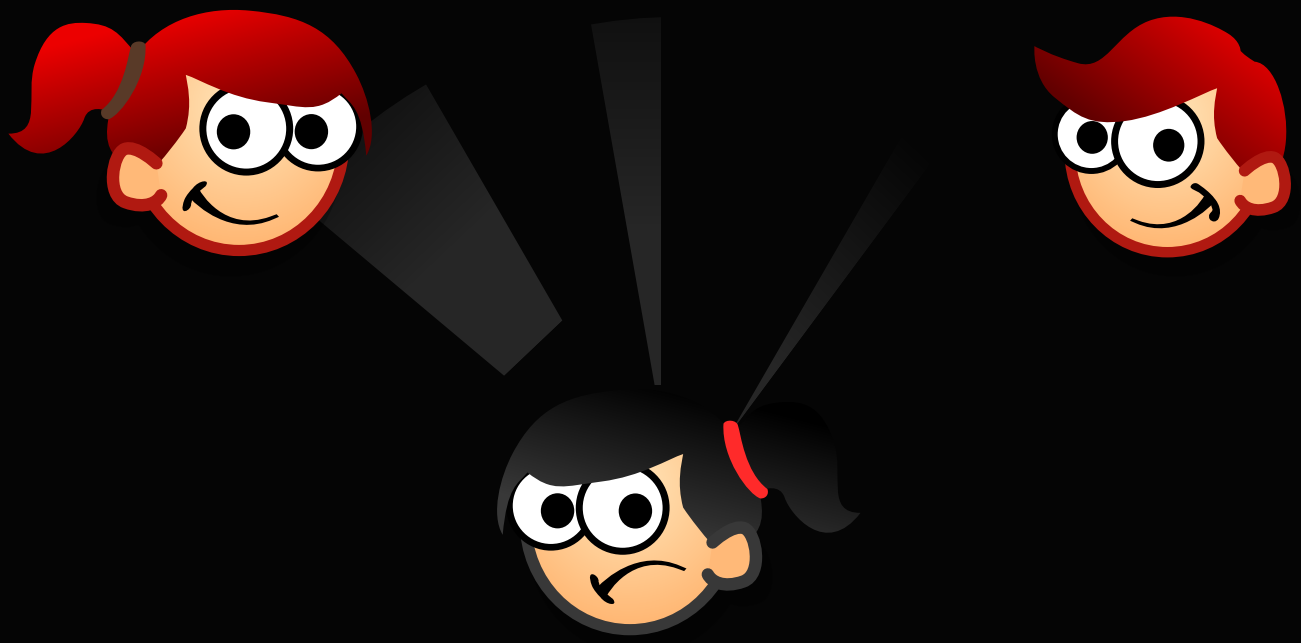
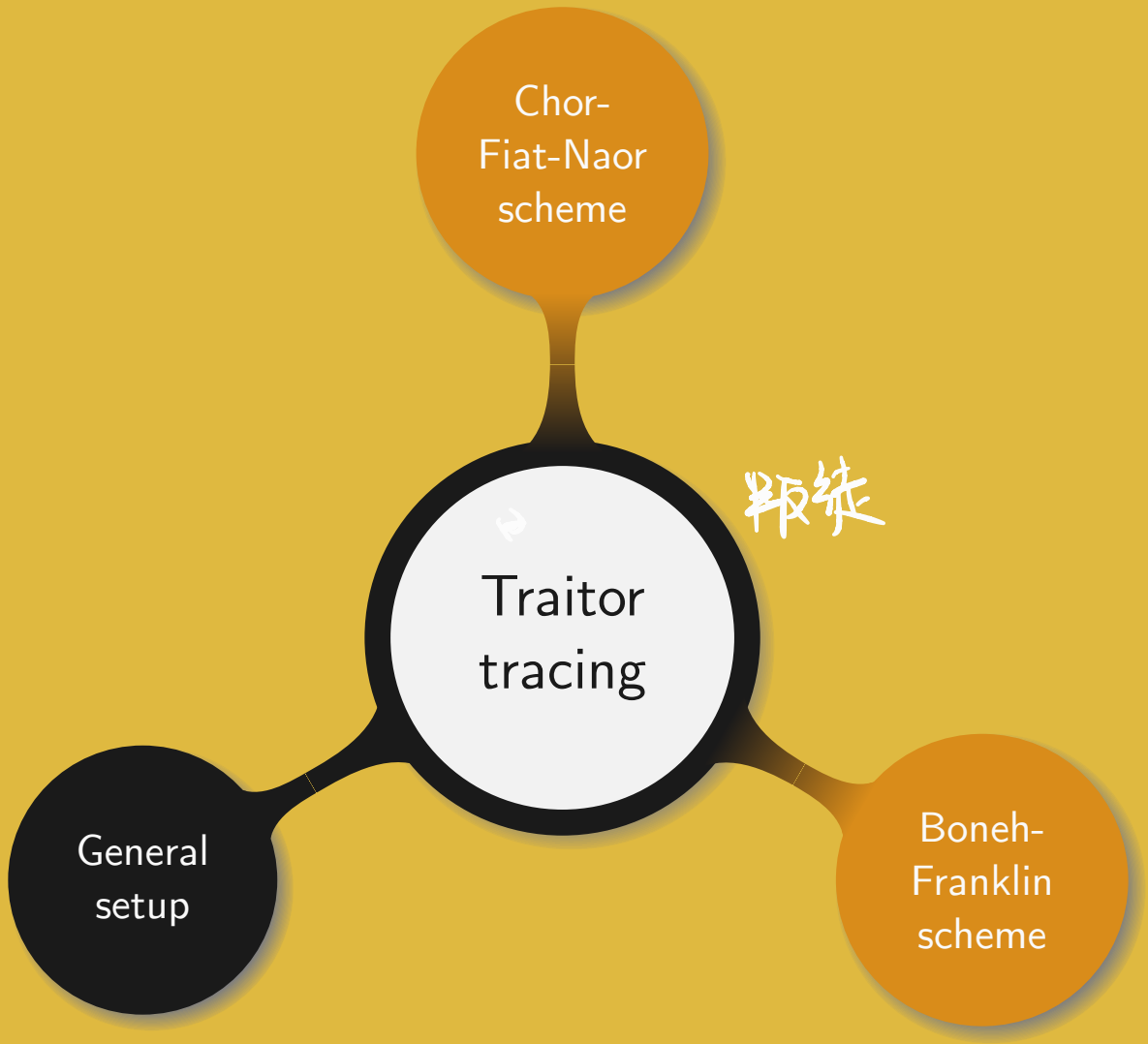


Introduction to Cryptography

7. Traitor tracing

Manuel – Summer 2021

Chapter organisation



A basic PayTV setup

A (too) simple solution

In this setup each user:

- Generates a pair of public/private keys
- Shares his public key with the service provider
- Receives the broadcast encrypted using his public key
- Decrypts and enjoys the program

Drawback: inefficient due to the amount of bandwidth required

Improving the solution: design a scheme where the encrypted information can be decrypted using different secret keys

Possible attacks

Three general types of attack:

- Decrypt the broadcast and share it
- Record the encrypted broadcast and share the decryption key with other people such that they can watch it
- Create a new secret key from several secret keys collected from various users

The two first cases are not traceable. The third scenario allows the construction of pirate decoders which can be sold at a large scale. The goal is to construct a scheme where tracing the "traitors" who shared their secret key is possible.

Tracing goals

Desirable properties of the scheme:

- Allows to trace the piracy
- Prevent legitimate users from being incriminated or framed by the traitors
- Allows the disconnection of illegitimate users
- Supplies legal evidence of the pirate's identity

Generalities

Types of schemes:

- Symmetric vs. asymmetric: how is encryption done
- Static vs. dynamic: keys changes at certain intervals
- Alternate approach: include credit card number in the user's key or use watermarking

Components of a Traitor Tracing scheme:

- Key generation and distribution
- Encryption and decryption methods
- Tracing algorithm

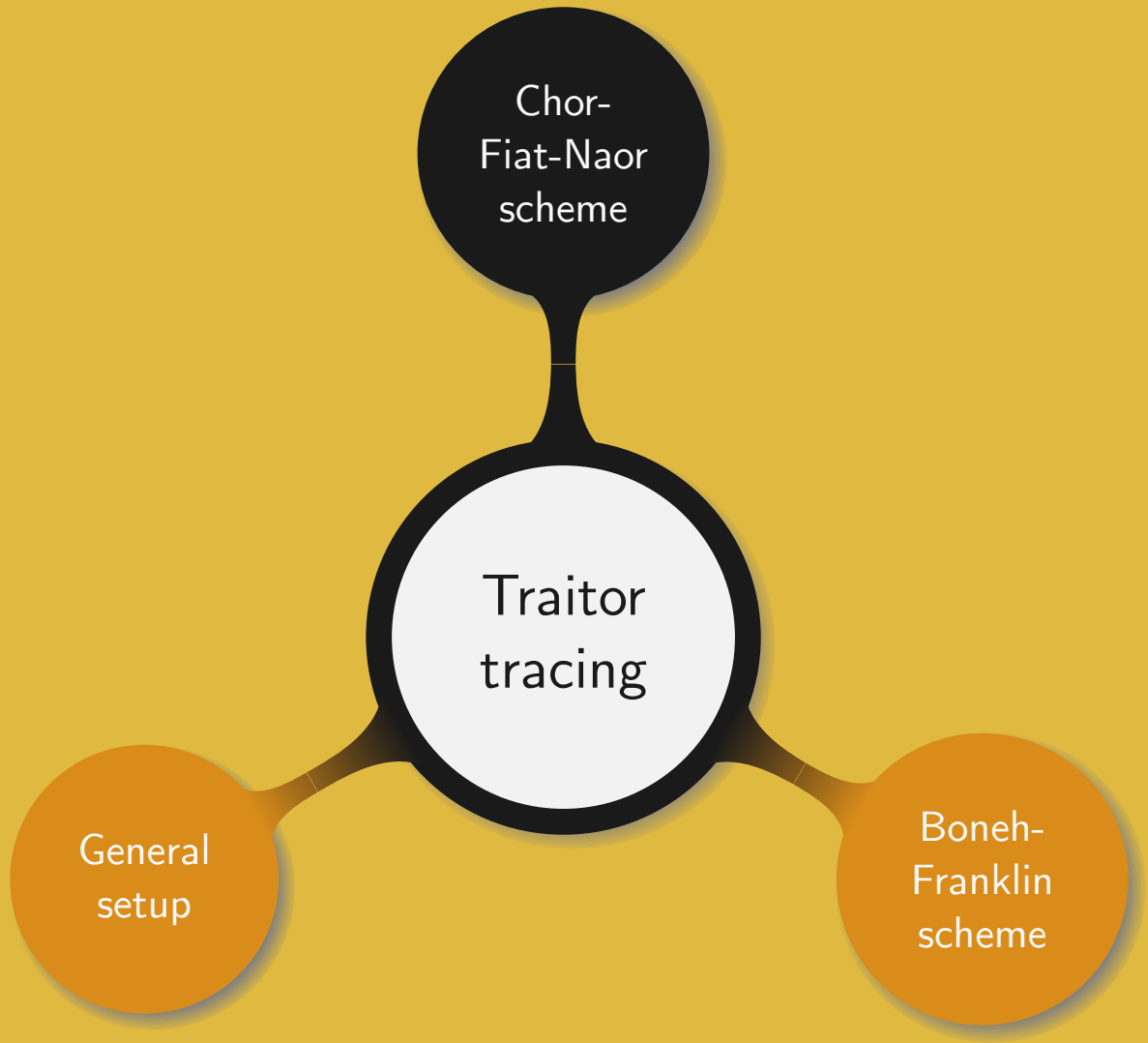
Resilience

When several users collude to generate a new pirate decoder, they use some personal information. The goal of the tracing algorithm is then to spot at least one traitor.

Definition

Assuming the underlying encryption to be secure, a scheme is said to be *m-resilient* if it can trace at least one traitor from a coalition of at most m malicious users.

Chapter organisation



Method

$b_i \in \{0,1\}$

Given n users $u_1, \ldots, u_n$ and $2 \log n$ keys

$$k_{1,0}, k_{1,1}, k_{2,0}, \ldots, k_{\log n,0}, k_{\log n,1},$$

define the key K_i of user u_i by 每个 user 都有自己的 K_i

$$K_i = \langle k_{1,b_{i,1}}, k_{2,b_{i,2}}, \ldots, k_{\log n, b_{i,\log n}} \rangle,$$

where $b_{i,j}$ is the j-th bit in the binary representation of i.

Applying this strategy, minimizes the number of keys as well as the bandwidth necessary to transmit the encrypted program to all the users.
Example. For eight users six keys $k_{1,0}, k_{1,1}, k_{2,0}, \ldots, k_{3,1}$ are defined. Since $(5)_{10} = (101)_2$, user u_5 has key $K_5 = \langle k_{1,1}, k_{2,0}, k_{3,1} \rangle$.

Method

Given some information m to broadcast, it is encrypted using a symmetric encryption protocol E with a secret key S. Then proceed as follows.

- Choose s_i, $1 \leq i \leq \log n$ such that

$$S = s_1 \oplus s_2 \oplus \cdots \oplus s_{\log n}$$

- Encrypt s_i using E and $k_{i,0}, k_{i,1}$
- Broadcast both the encrypted version of m and of the secret key S.

As each user u_i, $1 \leq i \leq n$, knows either $k_{i,0}$ or $k_{i,1}$, everybody can recover the secret key S and then decrypt the information m contained in $E_S(m)$.

用k_{i0}, k_{i1}来解密s_i

how come

begin to track

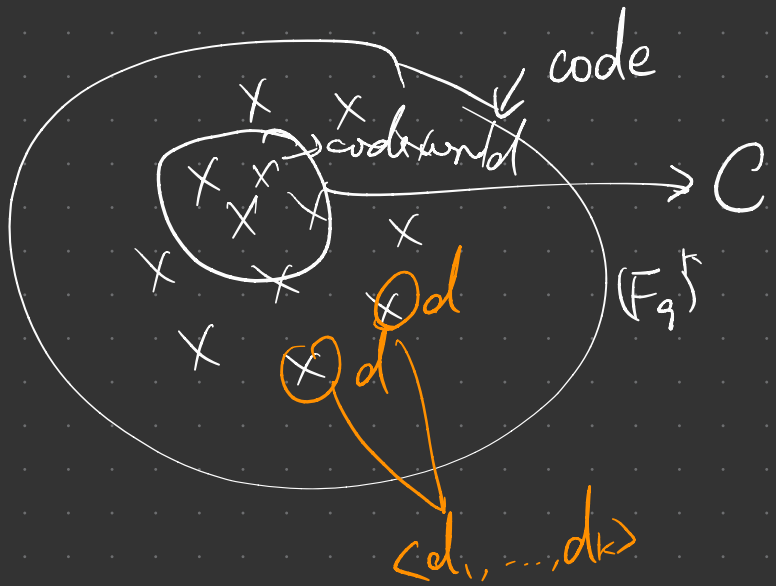
Definitions

Let E be a symmetric encryption protocol with keys of size l.

1. A *codeword* is a k-tuple of elements from $\mathbb{F}_q$, where $q = 2^l$
2. A set of codewords is called a *code*
3. Let $\mathcal{C} \subset \left(\mathbb{F}_q\right)^k$ be a code and $d = \langle d_1, \ldots, d_k \rangle$ be a codeword that is not in $\mathcal{C}$. If for all $1 \leq i \leq k$ there exists a codeword $c = \langle c_1, \ldots, c_k \rangle$ in $\mathcal{C}$ such that $d_i = c_i$, then d is called a *descendant* of $\mathcal{C}$. All the descendants of $\mathcal{C}$ form a *descendant code* of $\mathcal{C}$, denoted $\mathrm{desc}(\mathcal{C})$
4. Let d be a descendant of $\mathcal{C}$ and $\mathcal{S}_d = \left\{\mathcal{C}_p \subseteq \mathcal{C} : \ d \in \mathrm{desc}(\mathcal{C}_p)\right\}$. A codeword $c \in \mathcal{C}$ is an *identifiable parent* for d if

$$c \in \bigcap_{\mathcal{C}_p \in \mathcal{S}_d}$$

k elements from x extensions


code
codewords
C
$(F_q)^n$
d
d'
$\langle d_1, \ldots, d_k \rangle$

Explanations

In the context of a PayTV the previous definitions can be interpreted by identifying each codeword to a decoder.

The idea is then to define a code $\mathcal{C}$ by assigning a codeword to each decoder, in such a way that $\text{desc}(\mathcal{C}) \bigcap \mathcal{C}$ is empty.

The key used in a pirate decoder being constructed from elements of $\mathcal{C}$, it is a descendant of $\mathcal{C}$. Then $\mathcal{S}_d$ defines the set of suspects who could be involved in the generation of d.

An *identifiable parent* c from $\mathcal{S}_d$ is a suspect decoder which can be identified as guilty, since d is derived from c.

Explanations

Example. Let $\mathcal{C}$ be the code defined by

$$c_1 = \langle 0,0,0\rangle, \quad c_2 = \langle 0,1,1\rangle, \quad c_3 = \langle 0,2,2\rangle, \quad c_4 = \langle 1,0,3\rangle,$$
$$c_5 = \langle 2,0,4\rangle, \quad c_6 = \langle 3,3,0\rangle, \quad c_7 = \langle 4,4,0\rangle.$$

Assume that among the c_i, $1 \leq i \leq 7$, two traitors collude to construct a codeword $d = \langle d_1, d_2, d_3\rangle$. If any coordinate of d is non-zero then at least one parent can be identified:

$$d_1 = 1 \to c_4, \quad d_1 = 2 \to c_5, \quad d_1 = 3 \to c_6, \quad d_1 = 4 \to c_7,$$
$$d_2 = 1 \to c_2, \quad d_2 = 2 \to c_3, \quad d_2 = 3 \to c_6, \quad d_1 = 4 \to c_7,$$
$$d_3 = 1 \to c_2, \quad d_3 = 2 \to c_3, \quad d_3 = 3 \to c_4, \quad d_3 = 4 \to c_5.$$

Finally if $d = \langle 0,0,0\rangle$, then c_1 is an identifiable parent.

Definitions

1. The *hamming distance* between two elements a and b of $\left(\mathbb{F}_q\right)^k$ is defined as $\text{dist}(a, b) = |\{i : \ a_i \neq b_i, 1 \leq i \leq k\}|$ 不相等的个数
2. Let $\mathcal{C}$ be a code, then the minimal distance of $\mathcal{C}$ is

$$\text{dist}(\mathcal{C}) = \min\left\{\text{dist}(a, b) : \ a, b \in \mathcal{C}, a \neq b\right\}$$

Example. Reusing the code from example 7.14 we note that $\text{dist}(c_1, c_i)$, $2 \leq i \leq 7$, is 2, while $\text{dist}(c_2, c_4) = 3$. We can observe that no distance is smaller than 2 such that $\text{dist}(\mathcal{C}) = 2$.

Attributing the codewords

We now introduce a result which provides some hint on how to choose the distance in order to be able to identify at least one parent of an illegal decoder.

Theorem

Let $\mathcal{C} \subset \left(\mathbb{F}_q\right)^k$ be a code of length k and minimal distance D. If $D > k(1 - 1/w^2)$, where w is the size of the coalition, then it is possible to identify a parent of a descendant of $\mathcal{C}$.

Proof. For any a, b in $\left(\mathbb{F}_q\right)^k$, we define $\text{match}(a, b) = k - \text{dist}(a, b)$. Let $\mathcal{S}_d = \left\{\mathcal{C}_p \subseteq \mathcal{C} : \ d \in \text{desc}(\mathcal{C}_p)\right\}$ denote the set of suspects and d be a descendant of $\mathcal{C}_p \subset \mathcal{S}_d$. Let c be the closest element from d. We will now prove that c belongs to $\mathcal{C}_p$.

Attributing the codewords

Proof (continued). First note that since d is a descendant of $\mathcal{C}_p$, it follows that

$$\sum_{c' \in \mathcal{C}_p} \text{match}(d, c') \geq k.$$

Then as the coalition features w users it means that $|\mathcal{C}_p| \leq w$, and we can find a codeword c' in $\mathcal{C}_p$ such that

$$\text{match}(d, c') \geq \frac{k}{w}.$$

Recalling that c is the closest element from d we get

$$\text{match}(d, c) \geq \frac{k}{w}.$$

Attributing the codewords

Proof (continued). Finally we consider the number of common coordinates between $b \in \mathcal{C} \backslash \mathcal{C}_p$ and $d \in \mathrm{desc}(\mathcal{C}_p)$

$$\begin{aligned} \mathrm{match}(d, b) &\leq \sum_{c' \in \mathcal{C}_p} \mathrm{match}(c', b) \\ &\leq w(k - D). \end{aligned}$$

If $D > k(1 - 1/w^2)$, then clearly $\mathrm{match}(d, b) < \mathrm{match}(d, c)$. Since this is true for any $b \notin \mathcal{C}_p$, this means that c belongs to $\mathcal{C}_p$. □

This result is extremely useful as it provides information on how to construct the code and appropriately select the distance in order to trace traitors. As a general rule, the larger the minimum distance between two codewords, the easier to trace. On the other hand, having a large minimum distance will decrease the number of possible codewords in the code.

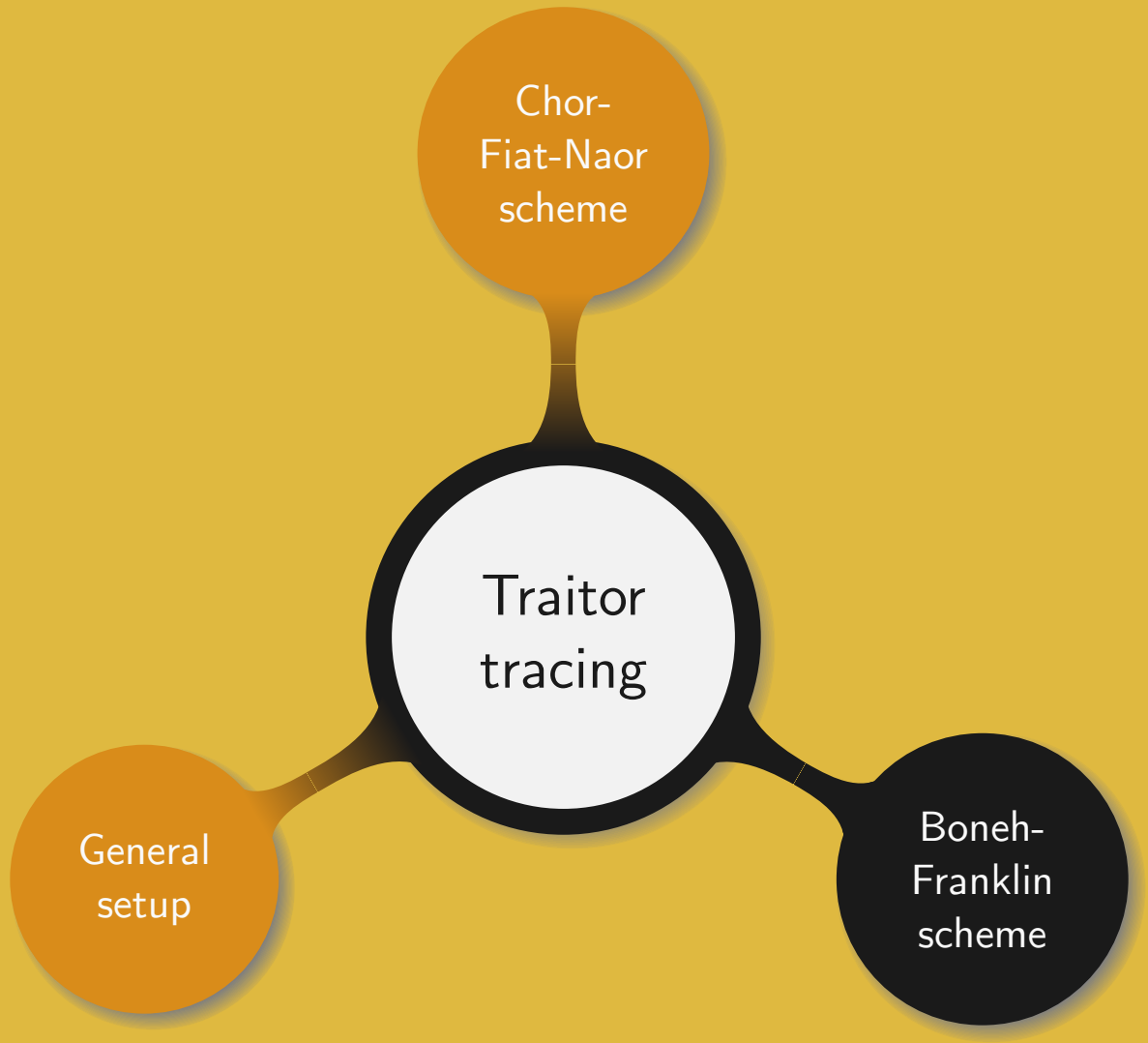
Back to Chor-Fiat-Nahor scheme

We notice the following properties:

- The scheme is using symmetric cryptography
- The number of decoders is n
- Each decoder is represented by a k-tuple of $\mathbb{F}_q$, with $k = \log n$
- The scheme is 1-resilient

The key aspect of this method is to choose a "good code"

Chapter organisation



Problem (Representation Problem)

Let G be a cyclic group of order n and $g_1, \cdots, g_m$, be m distinct generators of G. Then any element $y \in G$ can be expressed as $\prod_{i=1}^{m} g_i^{e_i}$, for some $0 \leq e_i \leq \varphi(n)$. We say that $(e_1, \cdots, e_m)$ is a representation of y in the base $(g_1, \cdots, g_m)$. Given G, y and a base $(g_1, \cdots, g_m)$, find the representation of y.

This problem can be seen as a generalisation of the DLP (3.71). Moreover when the generators are chosen randomly, finding two different representations of a given element is as hard as solving the DLP.

Setup

Simple description:

- p is prime
- G is a subgroup of prime order q
- g is a generator of G
- m is the maximal size of the coalition the scheme can trace
- $l \geq 2m + 2$ is the number of private keys
- $\mathcal{C} = \{c_1, \cdots, c_l\}$ is a code of $\mathbb{Z}^{2m}$

The scheme now described is CPA-1 secure but it can be extended into an enhanced CCA-2 version. This has the effect of more closely mirroring a real life context.

Key generation

The public and private keys are generated as follows:

1. Choose $2m$ random elements r_i, $1 \leq i \leq 2m$, in $\mathbb{F}_q$ and for each r_i compute $g_i = g^{r_i}$
2. Set the public key to $\langle y, g_1, \cdots, g_{2m} \rangle$, where $y = \prod_{i=1}^{2m} g_i^{\alpha_i}$, with the α_i being random elements from $\mathbb{F}_q$
3. Set the private key $k_i \in \mathbb{F}_q$ such that $k_i c_i$ is a representation of y in the base $(g_1, \cdots, g_{2m})$. That is

$$k_i = \frac{\sum_{j=1}^{2m} r_j \alpha_{i_j}}{\sum_{j=1}^{2m} r_j c_{i_j}} \bmod q$$

Encryption and decryption

Encryption:

- A message M in G
- Generate a random a in $\mathbb{F}_q$
- Define the ciphertext as $C = \langle My^a, g_1^a, \cdots, g_{2m}^a \rangle$

Decryption:

- A ciphertext $C = \langle My^a, g_1^a, \cdots, g_{2m}^a \rangle$
- Use the i-th secret key k_i to compute $U = \left(\prod_{j=1}^{2m} \left(g_j^a \right)^{c_{i_j}} \right)^{k_i}$

$$
\begin{aligned}
U &= \left(g^{\sum_{j=1}^{2m} r_j c_{i_j}} \right)^{k_i a} \\
&= \left(g^{\sum_{j=1}^{2m} r_j \alpha_{i_j}} \right)^{a} \\
&= y^a
\end{aligned}
$$

- Recover $My^a / U = M$

Tracing

The tracing algorithm being more advanced we do not detail it here but only highlight the main ideas.

The key principle behind the tracing ability is related to the difficulty of finding new representations. In fact if several users collude they are able to construct a new representation y. However this construction leads to a so called "convex combination" of the traitor's keys. It can be proved that if one can find a new representation that is not a convex combination of already known representations then one can solve the DLP.

By analysing the newly generated representation it is then possible to trace at least one traitor.

Revocation

Traitor tracing in practice:

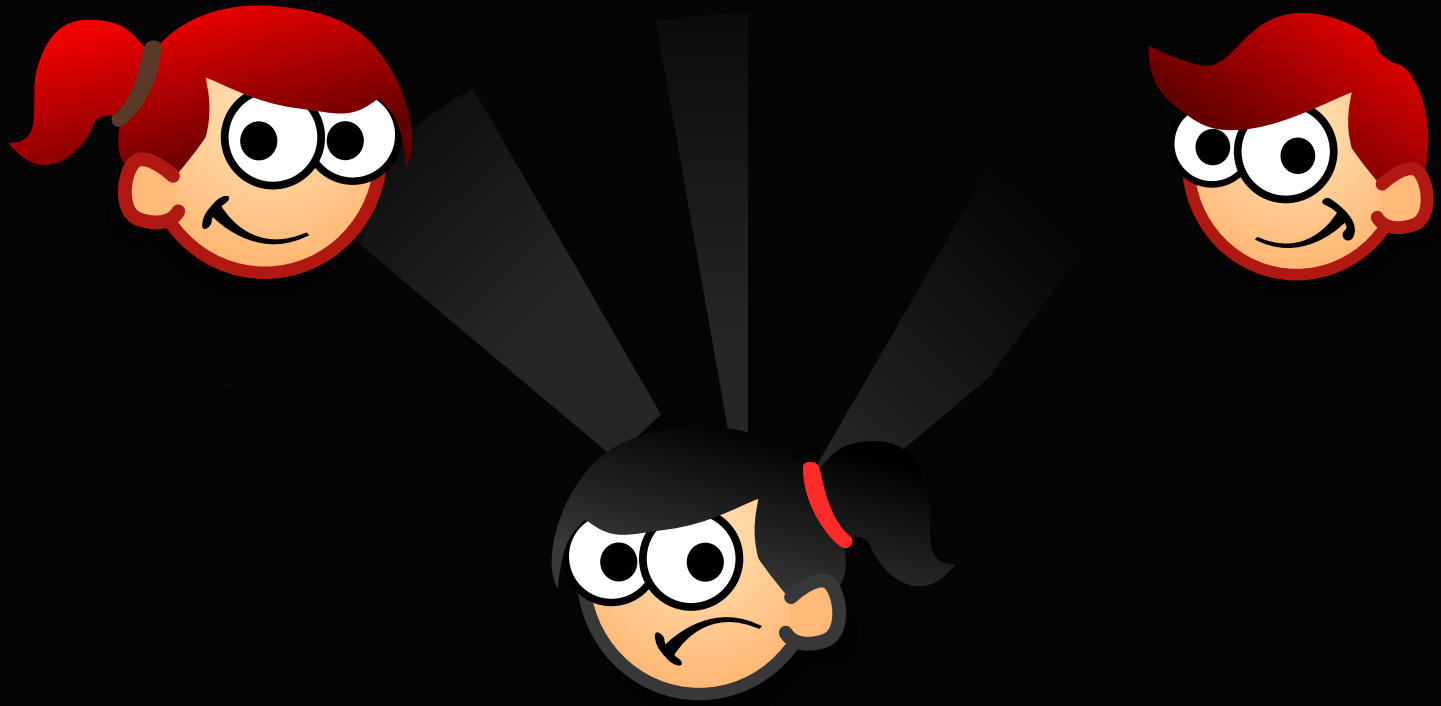
- Simple case: the value of the constructed k is known
- Harder case:
  - Only a card containing the key k is available
  - Not possible to directly read k
  - Blackbox traitor tracing

Key revocation:

- Black list: too complex
- Alternative:
  - Define the set of all the secret keys $\mathcal{K} = \{k_1, \cdots, k_m\}$
  - For each subset $\mathfrak{K}$ of $\mathcal{K}$ generate a public key whose encryption can only be decrypted by the keys in $\mathfrak{K}$
  - To revoke a secret key change the encryption key

Key points

- List the type of attacks a multicast system can undergo
- What is the resilience?
- Explain the basic idea to trace traitors
- How to handle key revocation?



Thank you!