Fingerprinting, traitor tracing, marking assumption

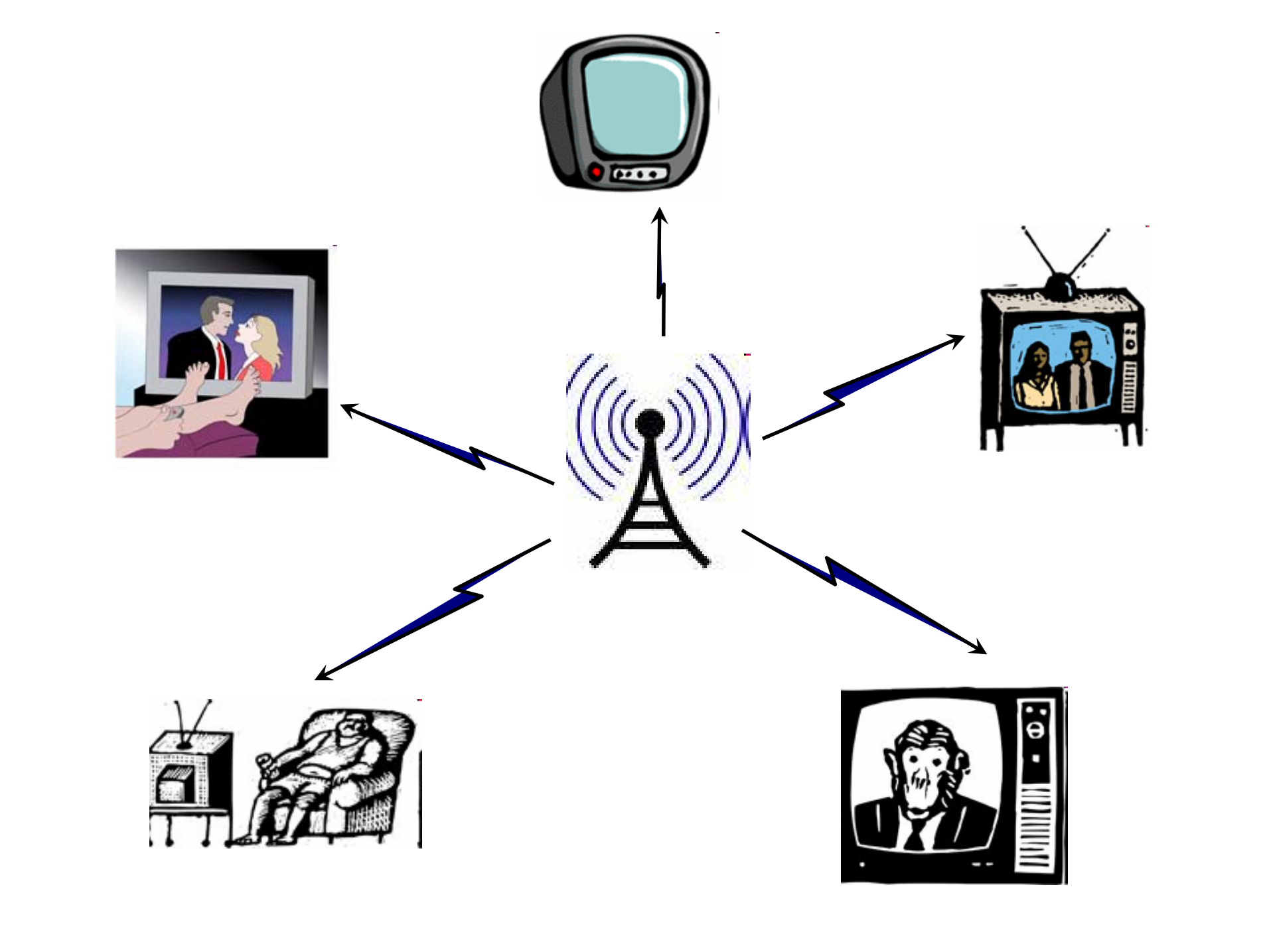
Alexander Barg
University of Maryland, College Park

ICITS 2011, Amsterdam

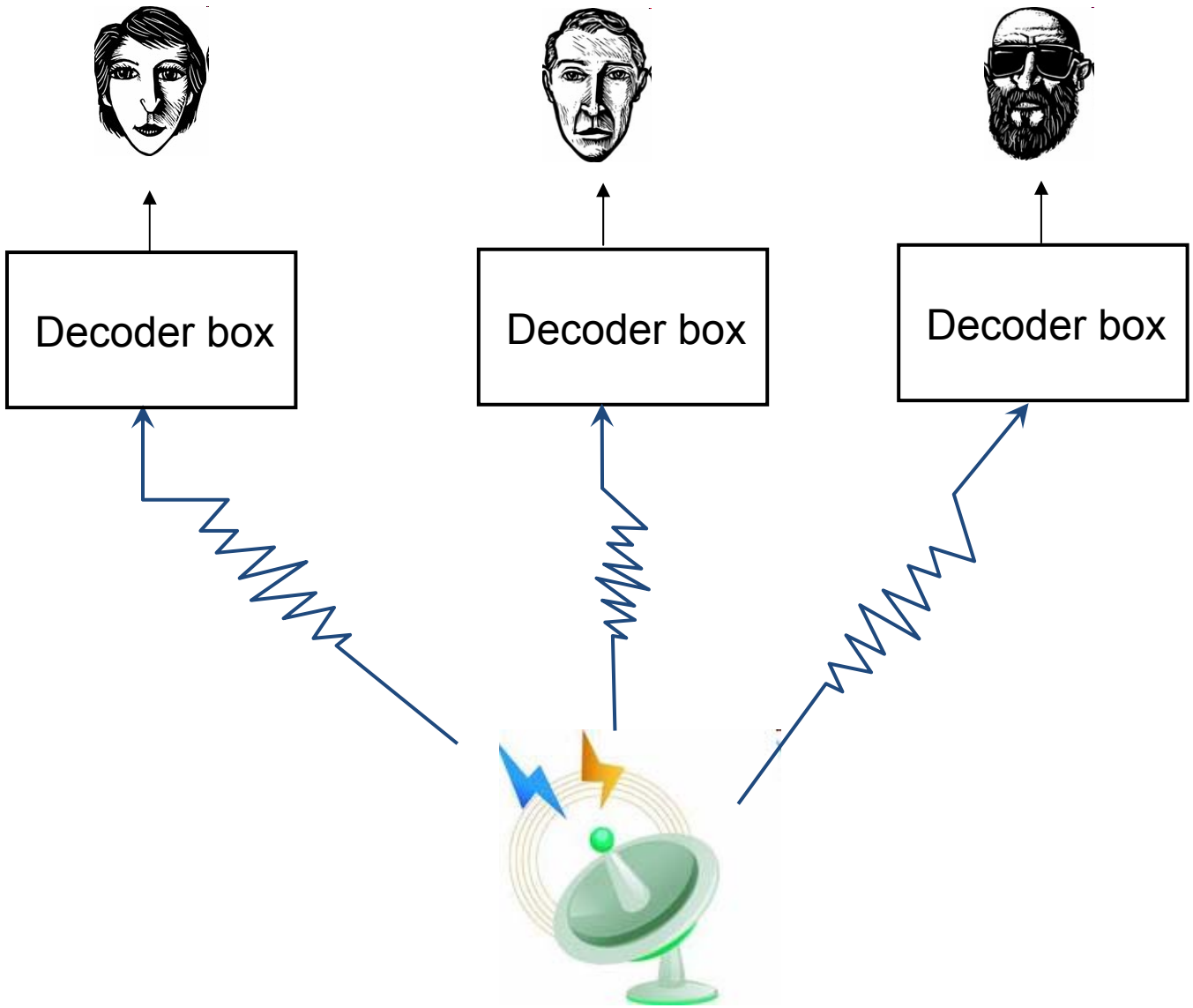
Acknowledgment:

Based in part of joint works with

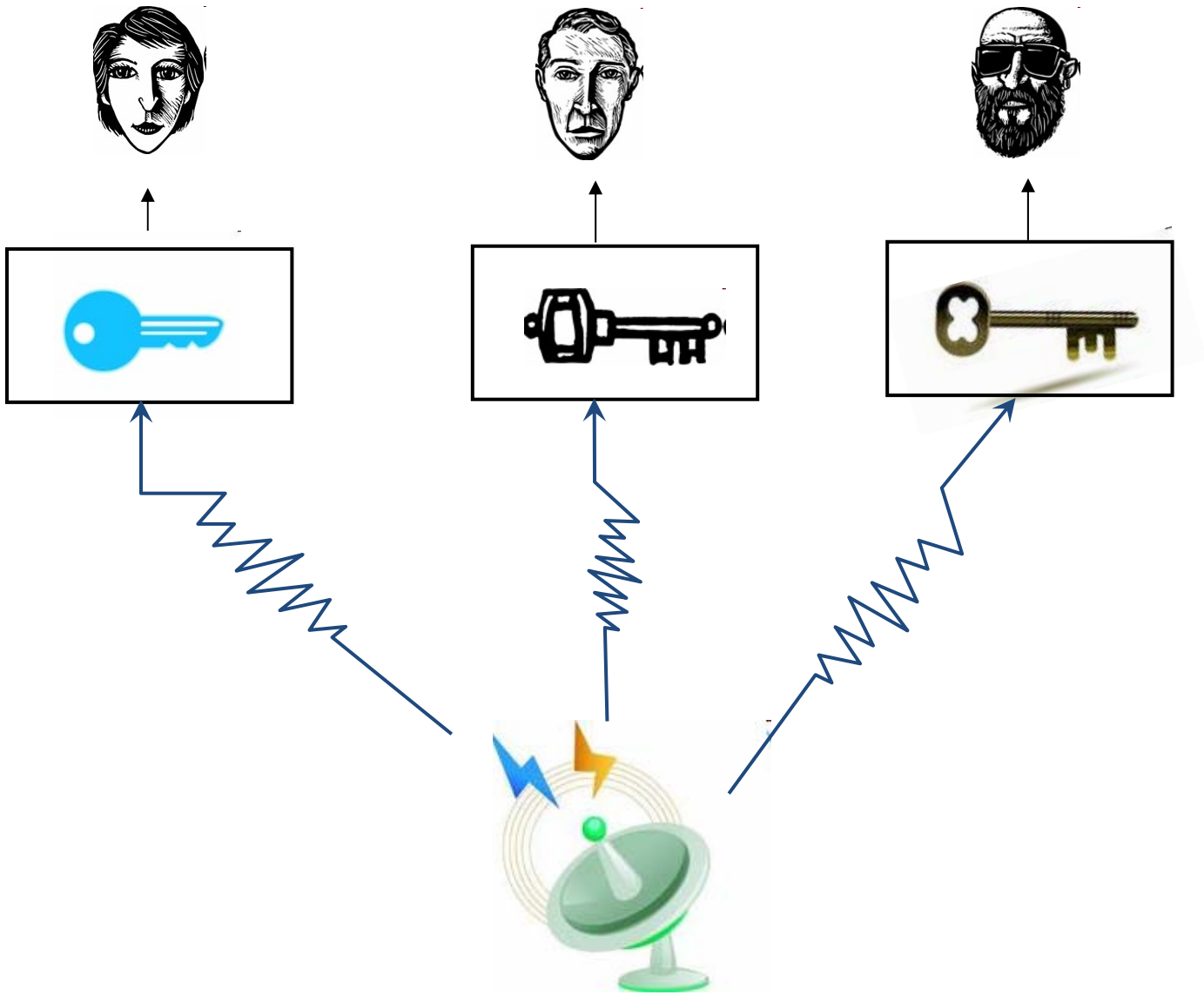
Prasanth A. (2009, '10)
Prasanth A. and I. Dumer (2006-8)
Grigory Kabatiansky (2001; 2010-11)



Pay-per-view subscription system



Pay-per-view subscription system



Open *t*-resilient traceability scheme

The access key S is transmitted over the public channel to the users in an encrypted form.

They use their personal keys  to decrypt S and access the contents.

The personal keys are represented by strings

$k^j=(k_1^j,\ldots,k_n^j)$, $j=1,\ldots, M$

Collusion attack: a group of users attempt to create to a pirate decoder with an untraceable personal key

B. Chor, A. Fiat, M. Naor, Tracing traitors, Crypto'94

Open *t*-resilient traceability scheme

The access key S is transmitted over the public channel to the users in an encrypted form.

They use their personal keys  to decrypt S and access the contents.

The personal keys are represented by strings

$k^j=(k_1^j,\ldots,k_n^j)$, $j=1,\ldots, M$

Collusion attack: a group of users attempt to create to a pirate decoder with an untraceable personal key

Other applications: digital fingerprinting, media fingerprinting

K_1 = 1 3 2 0 7 9 8 3 8

K_2 = 1 2 5 9 1 9 8 2 5

K_3 = 4 5 6 0 4 9 8 7 8

Unregistered Key

Y = 1 3 5 0 7 9 8 2 5

(1,2) 0 0 0 0 0 0 0 0 0

(1,3) 0 0 X 0 0 0 0 X X

(2,3) 0 X 0 0 X 0 0 0 0

K_1 = 1 3 2 0 7 9 8 3 8

K_2 = 1 2 5 9 1 9 8 2 5

K_3 = 4 5 6 0 4 9 8 7 8

Unregistered Key

Y = 1 3 5 0 7 9 8 2 5

(1,2) 0 0 0 0 0 0 0 0 0

(1,3) 0 0 X 0 0 0 0 X X

(2,3) 0 X 0 0 X 0 0 0 0

Parents of Y: {1,2}

Call coordinate i ***detectable*** for coalition X if $|\{x_{1,i}, x_{2,i}\}|>1$

Marking assumption (Boneh-Shaw '98):

The pirates cannot change the contents of undetectable coordinates.

Objective of system designer: identify pirates exactly or with low error

Fingerprinting

$\mathcal{M}$ = {1,2,…,M} - the set of users
U$\subset\mathcal{M}$, |U|$\leq$t a coalition

Fingerprinting code:

$f_k: \mathcal{M} \to \mathcal{Q}^n$ (assignment)
$\phi_k: \mathcal{Q}^n \to \mathcal{M} \cup \{0\}$ (identification)
$k \in \mathcal{K}$ randomization $P_K(k)=\pi(k)$

Let $U=\{u_1,\ldots,u_t\}$, $f_k(u_i)=x_i$

Collusion attack:

$V(y|x_1,\ldots,x_t) > 0$ only if y follows the marking assumption

Fingerprinting capacity

Goal of system designer: maximize the number of supported users $M=q^{Rn}$

Error probability of identification:

$$e(U,F,\Phi,V)=E_K \sum_{y:\ \phi_K(y)\notin U} V(y|f_K(U))$$

A randomized code (F,Φ) is t-fingerprinting with ε-error if

$$\max_{V\in \mathcal{V}_r} \max_{U:\ |U|\le t} e(U,F,\Phi,V) < \varepsilon$$

Rate $R\geq 0$ is ε-achievable for t-secure fingerprinitng if for every $\delta>0$ and every (sufficiently large) n there exists a q-ary code (F,Φ) of length n with rate

$$(1/n)\log_q M > R-\delta$$

$C_{t,q}(\varepsilon)$=sup of ε-achievable rates

Capacity $C_{t,q}=\lim_{\varepsilon\to 0} C_{t,q}(\varepsilon)$

Fingerprinting capacity

A.B., G.R. Blakley, G. Kabatiansky (ISIT 2001)
For any constant t

$C_{t,q}>0$, $\varepsilon=\exp(-\Theta\, n)$

separating arrays, list decoding

G. Tardos (FOCS '03) $C_{t,q} \geq \Omega(t^{-2})$

time-varying randomized encoding map

A.B., Prasanth A., I. Dumer (ISIT '07)

lower bounds: $C_{2,2} \geq 1/4$; $C_{3,2} \geq 1/12$;

upper bounds: $C_{2,2} \leq 0.322$, $C_{3,2} \leq 0.199$

$\Omega(1/t^2) \leq C_{t,2} \leq O(1/t)$

Fingerprinting capacity

A general upper bound (A.B., Prasanth A., I. Dumer ('07)):

$$C_{t,q}(\epsilon) \leq \min_{V \in \mathcal{V}_t} \max_{P_{X_1,\ldots X_t}} \max_{1 \leq i \leq t} I(X_i; Y | X_1^{i-1}, X_{i+1}^t)$$

where $X_1, \ldots, X_t, Y$ q-ary rv's

$P_{Y|X_1 \ldots X_t} = V$ s.t. $X_1, \ldots, X_t$ are independent.

E. Amiri and G. Tardos (SODA'09) computed the asymptotics of this bound:

$C_{t,2}=\Theta(t^{-2})$

Y.-W. Huang and P. Moulin (ISIT'09): $1/(2t^2 \ln 2) \leq C_{t,2} \leq 1/(t^2 \ln 2)$
general results on fingerprinting capacity

Constructions of capacity-approaching fingerprinting codes

Two-level fingerprinting

The set of users $\mathcal{M}=\mathcal{M}_1\times\mathcal{M}_2$ (M_1 groups of M_2 users each)

Encoder f_K: $\mathcal{M}_1\times\mathcal{M}_2 \to \mathcal{Q}^n$

Tracing ϕ_K: $\mathcal{Q}^n \to (\mathcal{M}_1\cup 0)\times(\mathcal{M}_2\cup 0)$

If coalition satisfies $|U| \leq t_2$, $\phi_K(y) \cap U\neq\emptyset$
If $t_2<t\leq t_1$, then ϕ_K identifies correctly the group that contains some of the pirates

Existence of two-level fingerprinting codes such that
$M_1=q^{nR1}$, $M_2=q^{nR2}$,
$R_1>0$, $R_2>0$
Prasanth A., A.B. (ISIT2010)

K_1 = 1 3 5 0 7 9 8 2 8

K_2 = 1 2 5 9 1 9 8 2 5

K_3 = 4 5 6 0 4 9 8 7 5

Unregistered Key

Y = 1 ? 5 0 ? 9 8 2 5

(1,2) 0 0 0 0 0 0 0 0 0

(1,3) 0 0 0 0 0 0 0 0 0

(2,3) 0 0 0 0 0 0 0 0 0

Parents of Y: {1,2} or {2,3} or {1,3}

identification impossible

More broadly, the pirates may deviate from the marking assumption in a certain number of coordinates

To what extent can we relax the marking assumption?

Parent identifying codes

C a subset of $\mathcal{Q}^n$, where $\mathcal{Q}$ is a finite set of cardinality q (alphabet)

$U=\{x_1, x_2,\ldots,x_t\} \subset C$ - pirate coalition, a set of t pirates

$y=f(U) \in \mathcal{Q}^n$ – collusion attack

$\langle U \rangle$ – set of descendants of U (with or without the marking assumption)

$\langle C \rangle_t = \bigcup_{U \subset C,\ |U| \leq t} \langle U \rangle$ - set of all possible attack vectors y

Definition: C has a ***t-IPP property*** if for all $y \in \langle C \rangle_t$

$$\bigcap_{U \subset C, |U| \leq t, y \in \langle U \rangle} U \neq \emptyset$$

H.D.L.Hollmann, J.H.vanLint, J.-P.Linnartz, L.M.G.M.Tolhuizen, JCTA 1998, no. 2 (case t=2)

Collusion attacks

$U=\{x_1, x_2,\ldots,x_t\}$

Narrow attack rule: $y_i \in \{x_{1,i},x_{2,i},\ldots,x_{t,i}\}$

For narrow-sense attack, it is possible to construct large-size IPP codes if (and only if) $t \leq q-1$ (nonzero rate; exact identification)

A. B., G. Cohen, S. Encheva, G. Kabatiansky, G. Zémor,
SIAM J. Discrete Math, 14, 2001.

Intermediate case

$U=\{x_1, x_2,\ldots,x_t\}$

$y=(y_1,y_2,\ldots,y_n) \in \mathcal{Q}^n \cup ?$ (erasure)

Narrow attack rule: $y_i \in \{x_{1,i},x_{2,i},\ldots,x_{t,i}\}$

What happens for more powerful attacks? –

Suppose there are εn coordinates that deviate from the above rule while the remaining $(1-\varepsilon)n$ coordinates obey it

Call such εn coordinates mutant

H.-J. Gutz and B. Pfitzmann, '99
T. Sirvent, '07
D. Boneh and M. Naor '08
O Bliiet and D. Phan '08.

Problem statement

A (t,ε)-IPP code (**robust t-IPP code**) $C \subset \mathcal{Q}^n$ guarantees exact identification of at least one member of the pirate coalition U, $|U| \leq t$ for any collusion attack with at most εn mutations.

Define $$R_q(n,t,\epsilon) = \max\{R(\mathcal{C}) : \mathcal{C} \subset \mathcal{Q}^n \text{ is } (t,\epsilon)\text{-IPP}\}$$
$$R_q(t,\epsilon) = \liminf_{n\to\infty} R_q(n,t,\epsilon).$$

Find $$\epsilon_{\text{crit}} = \epsilon_{\text{crit}}(q,t) := \sup(\epsilon : R_q(t,\epsilon) > 0)$$

Call coordinate i ***detectable*** for coalition U if $|\{x_{1,i}, x_{2,i}, \ldots, x_{t,i}\}| \geq 2$

(i) only detectable coordinates can mutate, always erasure: $\epsilon^{*,D}_{\text{crit}}$

(ii) only detectable coordinates can mutate: $\epsilon^{D}_{\text{crit}}$

(iii) any coordinate can be erased $\epsilon^{*}_{\text{crit}}$

(iv) any coordinate can mutate to any letter in $\mathcal{Q}$: ϵ_{crit}

Problem statement

A (t,ε)-IPP code **(robust t-IPP code)** $C \subset \mathcal{Q}^n$ guarantees ***exact identification*** of at least one member of the pirate coalition U, $|U| \leq t$ for any collusion attack with at most εn mutations.

Find $$\epsilon_{\text{crit}} = \epsilon_{\text{crit}}(q,t) := \sup(\epsilon : R_q(t,\epsilon) > 0).$$

(i) only detectable coordinates can mutate, always erasure: $\epsilon^{*,D}_{\text{crit}}$

(ii) only detectable coordinates can mutate: $\epsilon^{D}_{\text{crit}}$

(iii) any coordinate can be erased $\epsilon^{*}_{\text{crit}}$

(iv) any coordinate can mutate to any letter in $\mathcal{Q}$: ϵ_{crit}

$$\epsilon_{\text{crit}}(q,t) \leq \left\{ \begin{array}{c} \epsilon^{*}_{\text{crit}}(q,t) \\ \epsilon^{D}_{\text{crit}}(q,t) \end{array} \right\} \leq \epsilon^{*,D}_{\text{crit}}(q,t)$$

Existence of robust IPP codes

(t,ε) Traceability Code: (t,ε)-IPP code that permits pirate identification by the minimum Hamming distance to y

Proposition: For $q > t^2/(1-\varepsilon(t+1))$ there exist infinite sequences of (t,ε)-TA codes with positive code rate.

Proof: Coalition $U=\{u^1,\ldots,u^t\}$, $y \in \langle U \rangle_t$

$$\sum_{u \in X} s_H(y,u) \geq (1-\epsilon)n,$$

$$s_H(y,c) \leq n\epsilon + \sum_{u \in U} s_H(u,c) \leq n\epsilon + t(n-d)$$
$$< (1-\epsilon)nt^{-1}$$

This connects the Hamming distance to q,ε.

Corollary: For $q>t^2$, $\varepsilon_{crit}(q,t) \geq 1/(t+1)-t^2/(q(t+1))$

B. Chor, A. Fiat, M. Naor, Tracing traitors, Crypto'94

Existence of robust IPP codes

Traceability yields existence of IPP codes for $q \geq t^2$

Theorem: $\varepsilon_{crit}(q,t)>0$ for $q \geq t+1$

Proof idea:
(t,u)-hashing families, $u=\lfloor (t/2+1)^2 \rfloor$

$C \subset \mathcal{Q}^n$ is ***(t,u)-hash*** if $\forall\, T \subset U \subset C$, $|T|=t, |U|=u$

$\exists\, i$: $\forall\, x \in T, y \in U, y \neq x$: $x_i \neq y_i$

(t,u)-hash distance = #hash coordinates $\geq 2t+1$

Upper bounds for robust IPP codes

1. **Hash codes.** A code $C \subset \mathcal{Q}^n$ is called a hash code (s-PHF) if for any s codewords there is a hash coordinate,i.e., a coordinate that separates them: $x_{1,i} \neq x_{2,i} \neq \ldots \neq x_{s,i}$.
   For two s-subsets U_1, U_2, $U_1 \cap U_2 = \emptyset$, the number of coordinates that separates them is called s-hash distance $d_s(U_1, U_2)$

Proposition:

(L. A. Bassalygo, M.Burmester, A.G.Dyachkov, G.A.Kabatiansky, Hash codes ISIT'97)

Let C be a code with $d_s(C)=d$

$$|\mathcal{C}| \leq \binom{s}{2} \frac{d}{d - n\pi_{s,q}} \qquad (d > n\pi_{s,q})$$

$$\pi_{s,q} \triangleq \prod_{i=1}^{s-1} (1 - iq^{-1})$$

Upper bounds for robust IPP codes

1. **Hash codes.** A code $C \subset \mathcal{Q}^n$ is called a hash code (s-PHF) if for any s codewords there is a hash coordinate,i.e., a coordinate that separates them: $x_{1,i} \neq x_{2,i} \neq \ldots \neq x_{s,i}$.
   For two s-subsets U_1, U_2, $U_1 \cap U_2 = \emptyset$, the number of coordinates that separates them is called s-hash distance $d_s(U_1, U_2)$

2. **Upper bound**

Theorem:

$$\epsilon^{*,D}_{\text{crit}}(q,t) < \pi_{t+1,q},$$

$$\epsilon^{D}_{\text{crit}}(q,t) < \pi_{t+1,q}/(t+1).$$

$$\pi_{s,q} \triangleq \prod_{i=1}^{s-1}(1 - iq^{-1})$$

Proof idea: Let C be a t-IPP code, and let d_{t+1} be its hash distance
Take the t+1 codewords that realize d_{t+1}. Form

y_i=? if i is a hash coordinate

y_i=maj($x_{1,i}$,…,$x_{t+1,i}$) otherwise

Robust 2-IPP codes

In the case t=2 we can find exact answers

Theorem: q $\geq$ 3:

$$\epsilon^{*,D}_{\mathsf{crit}}(q,2) = \pi_{3,q} \triangleq \left(1-\frac{1}{q}\right)\left(1-\frac{2}{q}\right),$$

$$\epsilon^{D}_{\mathsf{crit}}(q,2) = 1/3\pi_{3,q}.$$

$$\epsilon^{*}_{\mathsf{crit}}(q,2) = \delta_{2,2} \triangleq (1-q^{-1})(1-3q^{-1}+3q^{-2})$$

$$\epsilon_{\mathsf{crit}}(q,2) = 1/3\pi_{3,q}$$

2-IPP codes

A code $C \subset \mathcal{Q}^n$ is (2,2)-separating if every (x_1,x_2), $(x_3,x_4) \in C \times C$ are separated by some coordinate

C is 3-hash if every x_1,x_2,x_3 are separated by a coordinate

Lemma (Hollmann et al. '98)
C is 2-IPP iff C is (2,2)-separating and 3-hash.

Robust 2-IPP codes

Exact answers are available

Theorem:

$$\epsilon^{*,D}_{\text{crit}}(q,2) = \pi_{3,q} \triangleq \left(1-\frac{1}{q}\right)\left(1-\frac{2}{q}\right),$$

$$\epsilon^{D}_{\text{crit}}(q,2) = 1/3\pi_{3,q}.$$

$$\epsilon^{*}_{\text{crit}}(q,2) = \delta_{2,2} \triangleq (1-q^{-1})(1-3q^{-1}+3q^{-2})$$

$$\epsilon_{\text{crit}}(q,2) = 1/3\pi_{3,q}$$

Proof idea: C is (2,2) separating if every distinct $x_1,x_2,x_3,x_4 \in C$ satisfy
$\exists$ i: $\{x_{1,i},x_{2,i}\}\cap\{x_{3,i},x_{4,i}\}=\emptyset$.
If in addition $x_{1,i}=x_{2,i}$ and $x_{3,i}=x_{4,i}$, we say that C has a restricted (2,2) separating property

Robust 2-IPP codes

Theorem:

$$\epsilon^{*,D}_{\mathsf{crit}}(3,2) = \epsilon^{*}_{\mathsf{crit}}(3,2) = 2/9$$

$$\epsilon^{D}_{\mathsf{crit}}(3,2) = \epsilon_{\mathsf{crit}}(3,2) = 2/27$$

Summary

1. Fingerprinting codes. Pirates are not restricted in their detectable coordinates (but follow the marking assumption).
   Exact identification impossible.
   Families of randomized codes; fingerprinting capacity.

2. Parent identifying codes. Pirates must follow their assigned keys in both detectable and undetectable coordinates.
   Exact identification possible

3. Robust parent identifying codes. Pirates are not restricted in detectable coordinates or do not follow the marking assumption, or both.
   Under some restrictions exact identification is still possible.
   t-IPP codes with distance
   Preprint: IACR eprint archive 2011/227