

# Short c-secure fingerprinting codes

6th Information Security Conference  
Oct 1--3, 2003, Bristol UK

Tri Van Le, Mike Burmester, Jianghi Hu  
Florida State University

# Outline

- Motivation
- Definitions
- A probabilistic fingerprint code.
- Lemmas and main Theorem.
- Optimality
- Discussion

# Motivation

- Boneh-Shaw proved that  $\Omega(c \log(1/\epsilon))$  is a lower bound for the length of  $c$ -secure fingerprint codes.
- No fingerprint codes meeting this lower bound were known until recently.

# Definitions

- Digital fingerprinting is a watermarking technique that protects intellectual property.
- Each copy of a document is fingerprinted by marking it in appropriate places in such a way that it is hard to forge/remove the marks.
- A collusion of buyers can always detect those marked positions where their marks differ.

# Definitions

A *c-secure fingerprinting code with error  $\epsilon$*  will detect at least one colluder from a *c-collusion* with probability at least  $1-\epsilon$ .

# A probabilistic fingerprint code

- $n$  = number of buyers
- $c$  = upper bound on number of colluders ( $c > 1$ ),
- $p = 1/c$

A fingerprint codeword is of type:

$$x = x_1 x_2 \dots x_l \text{ where } \text{Prob}[x_i = 1] = p$$

# The tracing algorithm

Let

- $C$  be a coalition, with  $z$  their **forged** fingerprint
- $u$  be the fingerprint of an **innocent buyer**
- $H_0(u, z) = \#$  places  $i$  such that  $u_i = 1$  and  $z_i = 0$

The tracing algorithm selects a buyer whose fingerprint  $u^*$  minimizes  $H_0(u, z)$ , that is:

$$H_0(u^*, z) \leq H_0(u, z) \text{ for all fingerprints } u.$$

# Lemmas and main Theorem

## Lemma 1

$$\text{Prob}[u^* \text{ is not in } C] = e^{-O(l) + \ln n}$$

# Lemmas

## Lemma 2 [Chernoff bound]

Let  $X_1, \dots, X_t$  be  $t$  independent identically distributed random variables with the same expected value  $x$  and range  $[0, 1]$ .

Let  $0 < \delta < 1$  be any constant. Then:

$$\text{Prob}[1/t \sum_{i=1,t} X_i \leq (1-\delta)x] \leq \exp(-t\delta^2x/3).$$

## Lemma 3 [Extended Chernoff bound]

Let  $X_1, \dots, X_t$  be  $t$  independent identically distributed random variables with the same expected value  $x$  and range  $[a, b]$ .

Let  $0 < \delta < 1$  be any constant. Then:

$$\text{Prob}[1/t \sum_{i=1,t} X_i \leq (1-\delta)x + \delta a] \leq \exp(-t\delta^2(x-a)/3(b-a)).$$

# Proof of Lemma 1

- Let  $d(1,0)=1$  and  $d(x,y)=0$  otherwise
- Let  $\Delta_i(u, C) = d(u_i, z_i) - 1/c \sum_{v \in C} d(v_i, z_i)$

d		
0	x	y
1	1	0

be the **advantage** of an innocent buyer  $u$  against coalition  $C$  at **position  $i$** .

Then  $\sum_{i=1, n} \Delta_i(u, C) = H_0(u, z) - 1/c \sum_{v \in C} H_0(v, z)$  is the **advantage** of an innocent buyer  $u$  against coalition  $C$ . It is enough to show that:

$$\text{Prob}[\sum_{i=1, n} \Delta_i(u, C) \leq 0] \leq \varepsilon/n.$$

# Proof of Lemma 1, continued

- Let  $i$  be a position such that  $z_i=0$ .
- Let  $k = \#\{v \in C \mid u[i]=0\}$  be the number of 0 that  $C$  has at position  $i$ .
- Given such  $i$  and  $k$ , the probability distribution of  $\Delta_i(u, C)$  is:

$u[i]$	$\Delta_i(u, C)$	probability
1	$k/c$	$p$
0	$k/c-1$	$q$

# Proof of Lemma 1, continued

- Let  $D_i = d(u[i], 0) - 1/c \sum_{v \in C} d(v_i, 0)$  if  $k > 0$ ,  
and  $D_i = 0$  otherwise.
- Then  $D_i$  are independent random variables with distribution:

$u[i]$	$D_i$	probability
1	$k/c$	$p \binom{c}{k} q^k p^{c-k}, 1 \leq k \leq c.$
0	$k/c-1$	$q \binom{c}{k} q^k p^{c-k}, 1 \leq k \leq c.$
	0	otherwise.

# Proof of Lemma 1, continued

$$\begin{aligned} E(D_i) &= \sum_{k=1}^c \binom{c}{k} q^k p^{c-k} \left( q \left( \frac{k}{c} - 1 \right) + p \frac{k}{c} \right) \\ &= \sum_{k=1}^c \binom{c}{k} q^k p^{c-k} \left( \frac{k}{c} - q \right) \\ &= \frac{1}{c} \sum_{k=1}^c \binom{c}{k} q^k p^{c-k} k - q \sum_{k=1}^c \binom{c}{k} q^k p^{c-k} \\ &= \frac{1}{c} cq - q(1 - p^c) \\ &= qp^c. \end{aligned}$$

# Proof of Lemma 1, continued

- Observe that  $\Delta_i(u, C) = (1-z_i)D_i$  and thus:

$$\sum_{i=1, \dots, l} \Delta_i(u, C) = \sum_{z_i=0} D_i$$

- Let  $n_0 = \#\{i \mid z_i=0\}$ . Apply Lemma 3 to the  $n_0$  independent random variables  $D_{ij}$  ( $j = 1, n_0; z_{ij} = 0$ ) with expected value  $qp^c$  and range  $[-1, 1]$ , and  $\delta = qp^c/(1+qp^c)$  we obtain:

$$\text{Prob}[\sum_{i=1, \dots, l} \Delta_i(u, C) \leq 0] \leq \exp(-n_0 q^2 p^{2c} / 6(1+qp^c)).$$

- Substitute  $\Delta_i(u, C) = d(u_i, z_i) - 1/c \sum_{v \in C} d(v_i, z_i)$  we get:

$$\text{Prob}[H_0(u, z) - 1/c \sum_{v \in C} H_0(v, C) \leq 0] \leq \exp(-n_0 q^2 p^{2c} / 6(1+qp^c)).$$

# Proof of Lemma 1, continued

- Now by the definition of  $u^*$ ,  $H_0(u^*, z) \leq H_0(v, z)$  for all  $v \in C$ .  
Then:  $H_0(u^*, z) \leq 1/c \sum_{v \in C} H_0(v, C)$ .

- Thus:

$$\begin{aligned} \text{Prob}[u^* \notin C] &\leq \sum_{u \notin C} \text{Prob}[H_0(u, z) \leq 1/c \sum_{v \in C} H_0(v, C)] \\ &\leq (n-c) \exp(-n_0 q^2 p^{2c} / 6(1+qp^c)) \\ &< \exp(-n_0 q^2 p^{2c} / 6(1+qp^c) + \ln(n)). \end{aligned}$$

- Let  $m_0$  be the number of all zero columns of  $C$ . Then  $m_0 \leq n_0$ .

By Lemma 2, taking  $\delta = 1/2$ ,  $x = q^c$ ,

$$\text{Prob}[m_0/l \leq q^c/2] \leq \exp(-l/12).$$

Therefore:  $\text{Prob}[n_0 \leq q^c l/2] \leq \exp(-l/12)$

# Proof of Lemma 1, conclusion

Then

$$\begin{aligned}\text{Prob}[u^* \notin C] &< \exp(-q^2 p^{2c} l / 12(1+qp^c) + \ln(n)) + \exp(-l/12) \\ &< e^{-al + \ln(n)}\end{aligned}$$

where  $a = q^2 p^{2c} / 12(1+qp^c)$ .

It follows that:  $l = 1/a \ln(n/\varepsilon) = O(\ln(n/\varepsilon))$ .

# Theorem

## Theorem

Let  $c > 1$  be constant (or sublogarithmic). Then for all  $\varepsilon > 0$ ,  $n > 1$ , there is a  $(l, n)$ -code with  $l = O(\ln(n/\varepsilon))$  that is  $c$ -secure with  $\varepsilon$  error.

## Proof

We will catch any buyer with fingerprint  $u$  whose  $H_0(u, w)$  is minimal. Then apply the Lemma to get  $\text{Prob}[u \text{ is not in } C] < \varepsilon$ .

# Asymptotic Optimality

Boneh-Shaw proved that  $\Omega(c \log(1/\epsilon))$  is a lower bound for the length of  $c$ -secure fingerprint code.

This implies that for constant  $c$  our codes with length  $l = O(\ln(1/\epsilon))$  are asymptotically optimal, provided that  $\epsilon < n^{-a}$  for some constant  $a > 0$ .

In general, for constant  $c$  and arbitrary  $\epsilon > 0$  our codes are more efficient than those of Boneh Shaw by a factor of  $O(1/\epsilon)$ .

# Discussion

Recent results (done independently):

- Pickert-Shelat-Smith (SODA 2003)
  - Lower bound  $\Omega(c^2 (\ln(n/\epsilon)))$
- Gabor Tardos (STOC 03)
  - Optimal Probabilistic Fingerprint Codes which meet the bound  $\Omega(c^2 (\ln(n/\epsilon)))$