CDMA-Based Fingerprinting Technique

Minoru Kuribayashi

Assistant Professor

Graduate School of Engineering

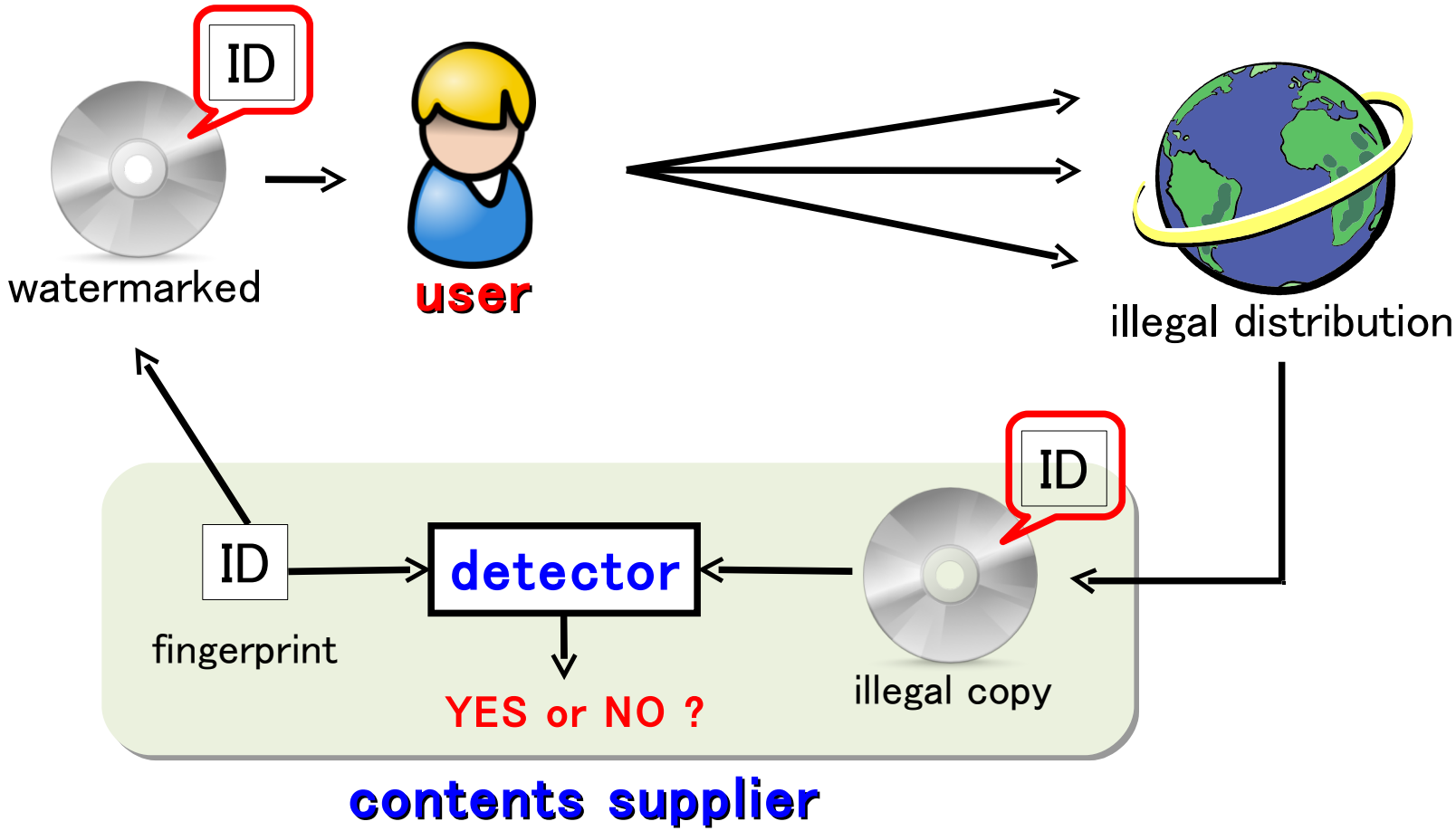
Kobe University, Japan

Contents

Digital Fingerprinting

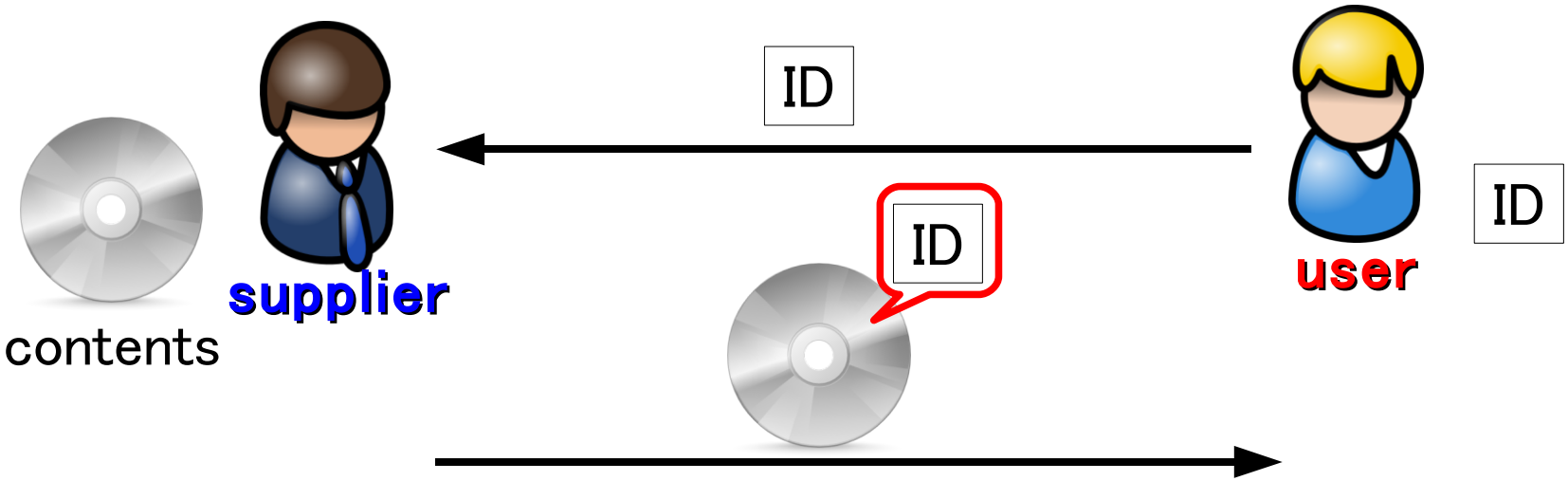
Insert user's ID into digital contents

- Identify the illegal users
- Protect multimedia contents from unauthorized distribution



Problem in Symmetric Protocol

If a contents supplier embeds user's ID information into the contents, the user may be framed by the supplier.

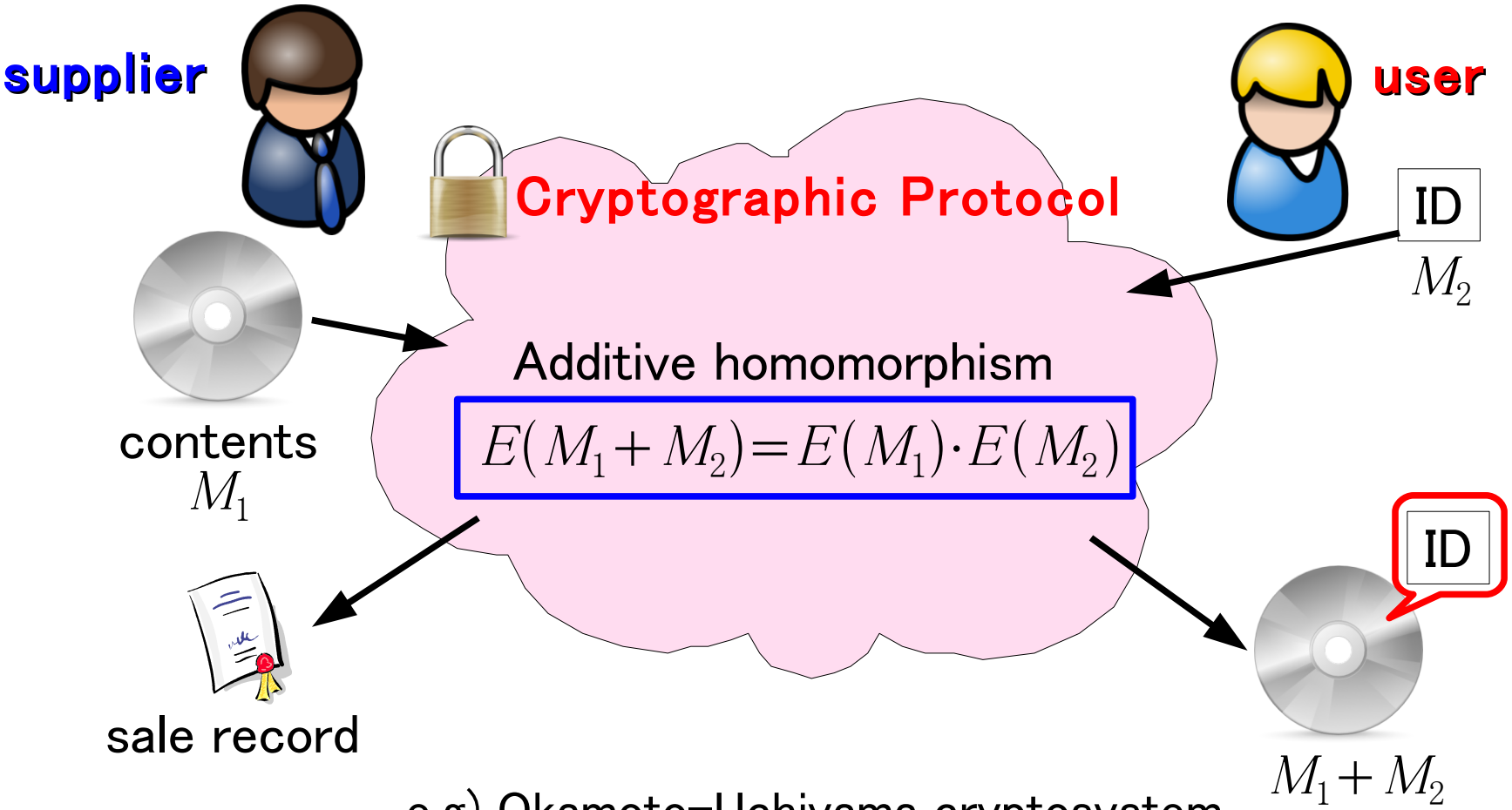


Both party have the fingerprinted contents in this model.

A malicious supplier may distribute it by himself in order to claim that the copy is come from the user.

Asymmetric Protocol

Using the homomorphic property of public-key cryptosystem, only the buyer can obtain the fingerprinted contents.



e.g) Okamoto-Uchiyama cryptosystem
Paillier cryptosystem

Robustness againt Attack

- **Fingerprinting Technique**

  Unique fingerprint signal is embedded in digital contents using a watermarking technique.

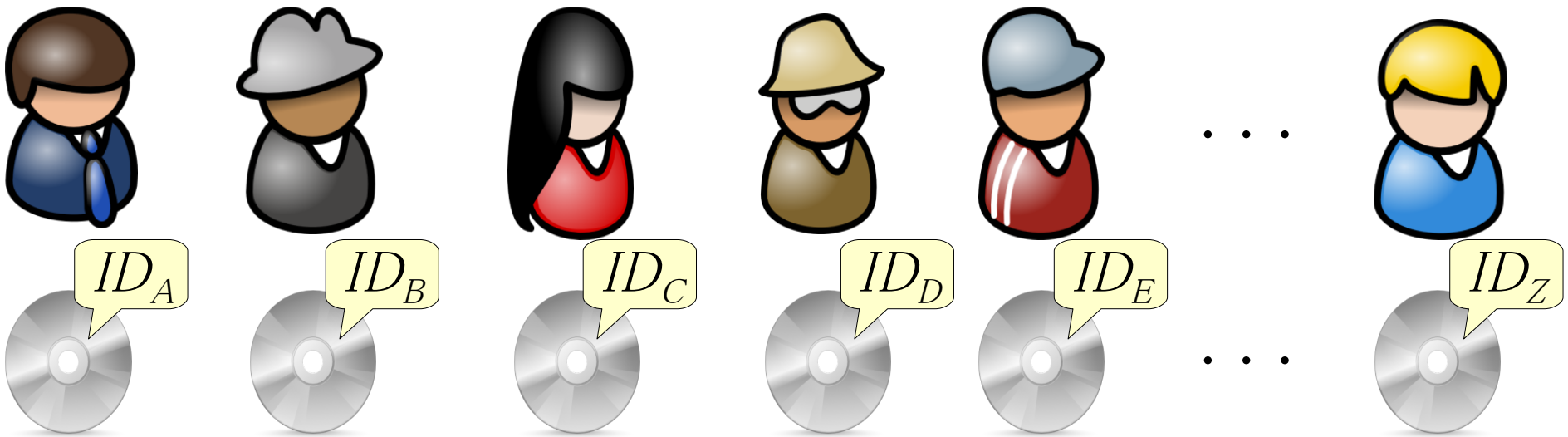
- **Required Property**

  Robustness against attack

  - Attack for the watermarking technique.
  - Attack by some users → Collusion Attack

Collusion Attack

Each user owns uniquely fingerprinted contents

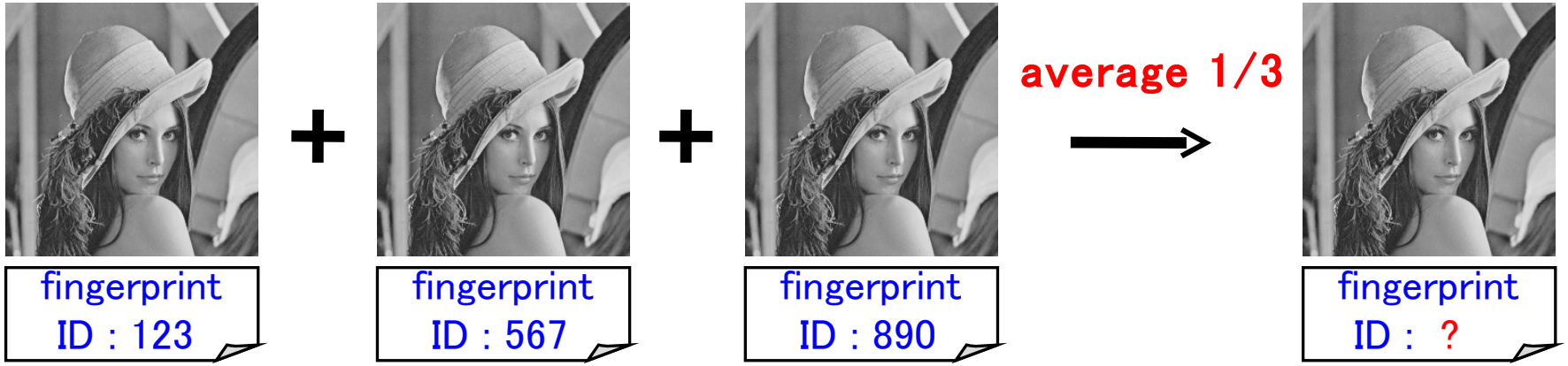


Some users combine their copies of a same contents to delete/modify the embedded fingerprint.

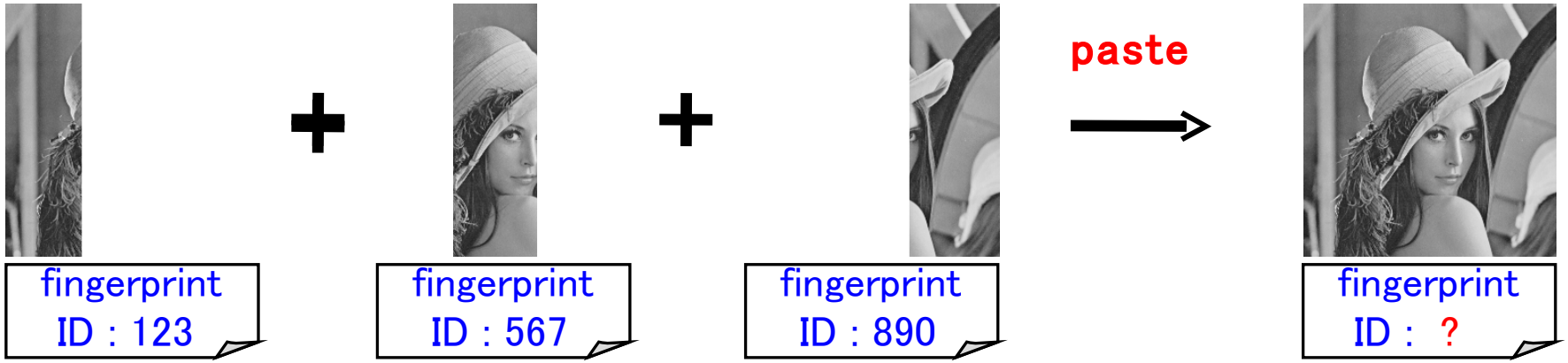
eg.) averaging, interleaving, etc.

Example

- **Averaging attack**



- **Interleaving attack**



Approach

Spread spectrum fingerprinting

Mutually independent signals are assigned as fingerprints.

eg.) I. J. Cox and J. Kilian and F. T. Leighton and T. Shamson
"Secure Spread Spectrum Watermarking for Multimedia,"
IEEE Trans. Image Processing, vol.6, no.12, pp.1673-1687, 1997.

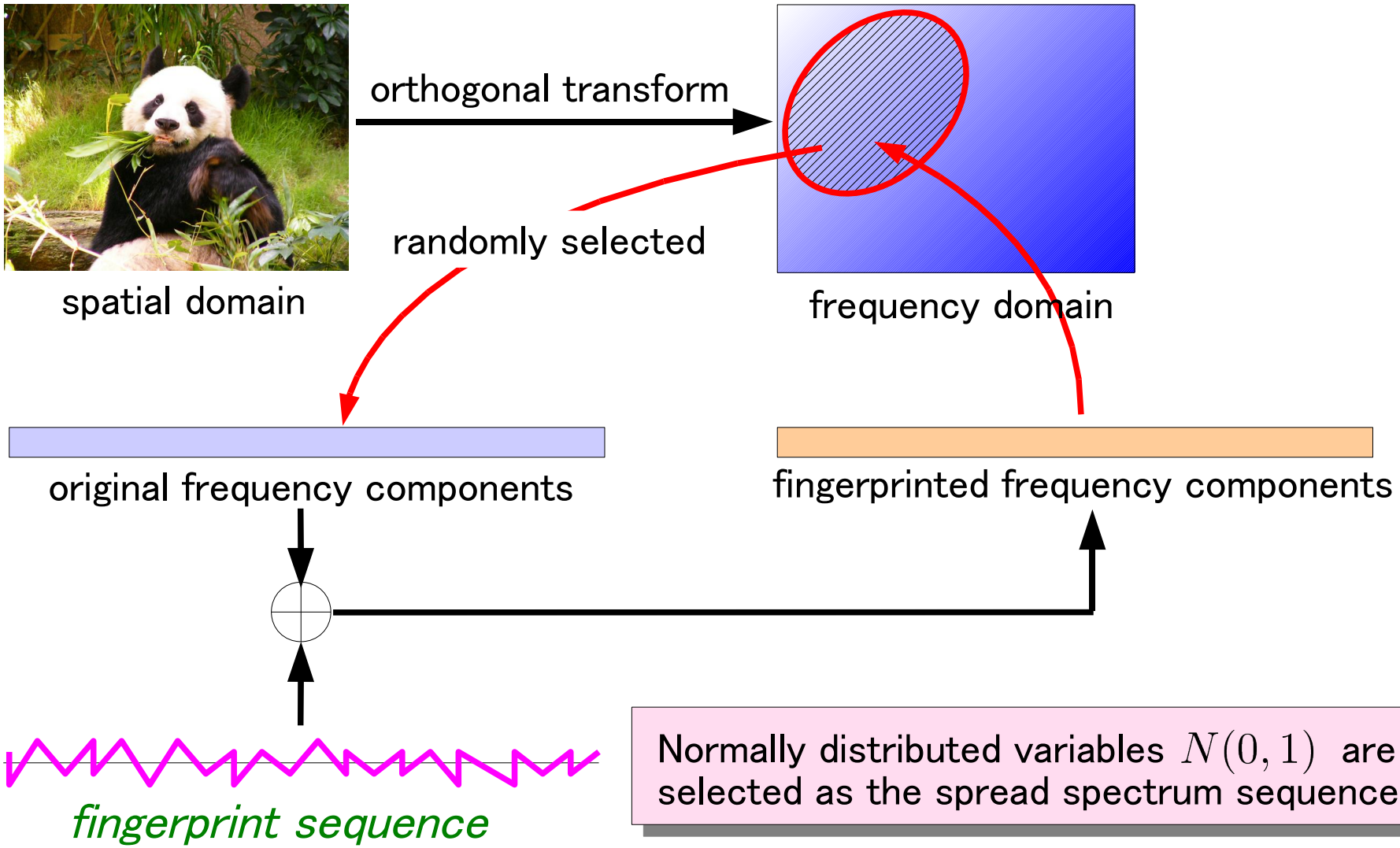
- Collusion-secure fingerprinting code

  eg.) c-secure code, anti-collusion code, tardos code, etc.

The code-length is extremely long.

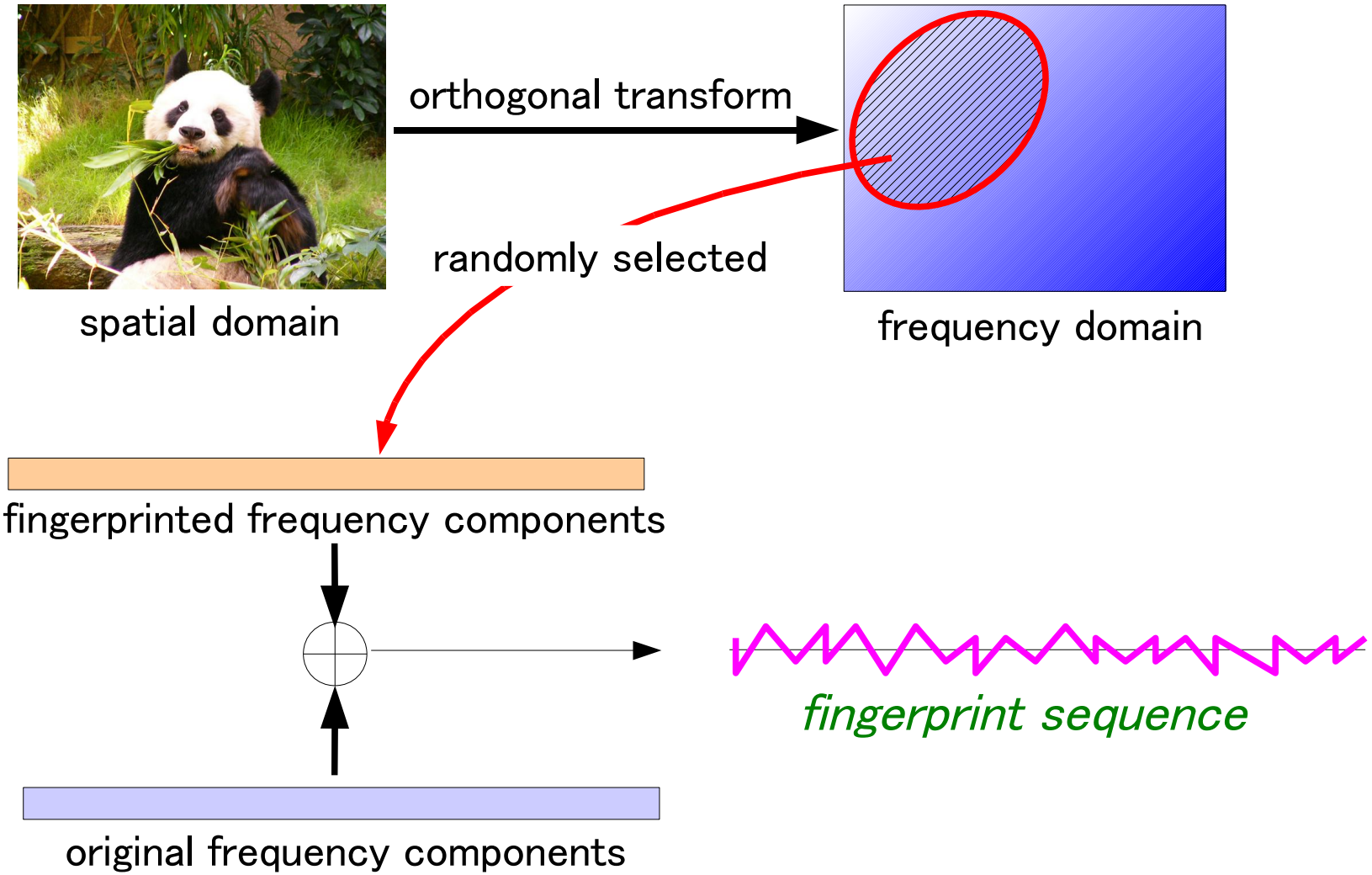
Spread Spectrum Fingerprinting (Cox's Method)

Embedding Procedure



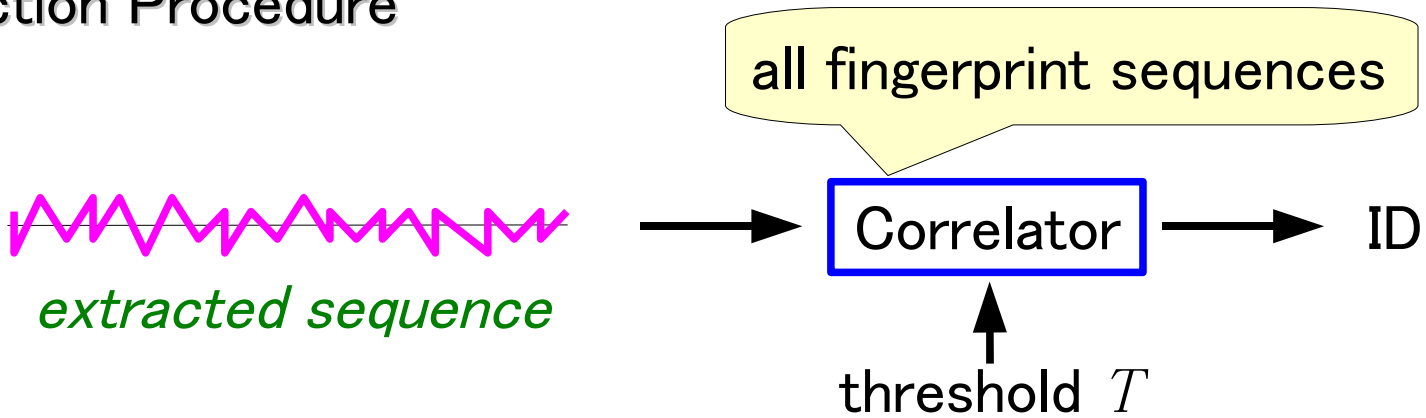
Normally distributed variables $N(0,1)$ are selected as the spread spectrum sequence.

Extraction Procedure

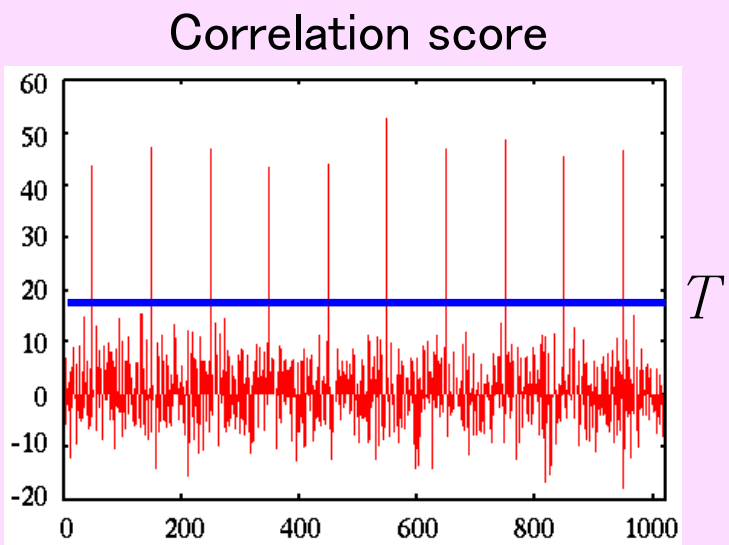


Spread Spectrum Fingerprinting (Cox's Method)

Detection Procedure



The users are regarded as guilty if the correlation scores exceed a threshold.

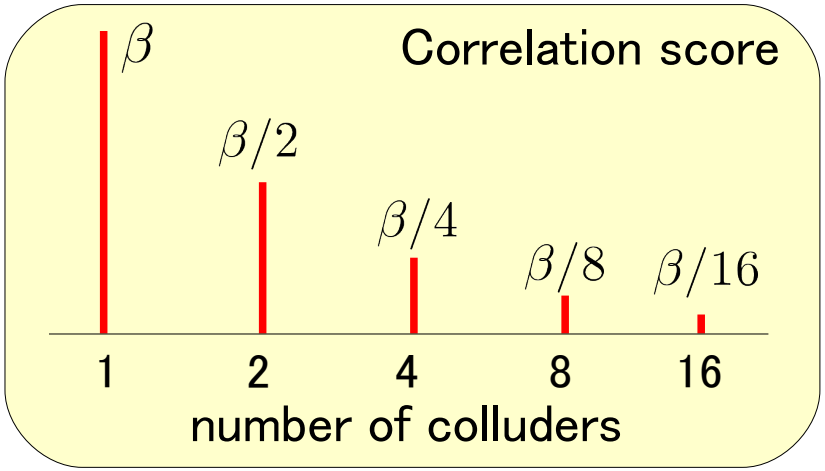


Properties

- The method retains high robustness against any other attacks such as filtering, lossy compression, additive noise, etc.

- The attenuation of the correlation score is linear with respect to the number of colluders.

If the number of colluder is small, the detector identifies them.



- The required computational costs is too high.

Detection procedure : $O(NL)$

N : number of users

L : length of sequnce

Objective

High robustness & low computational costs

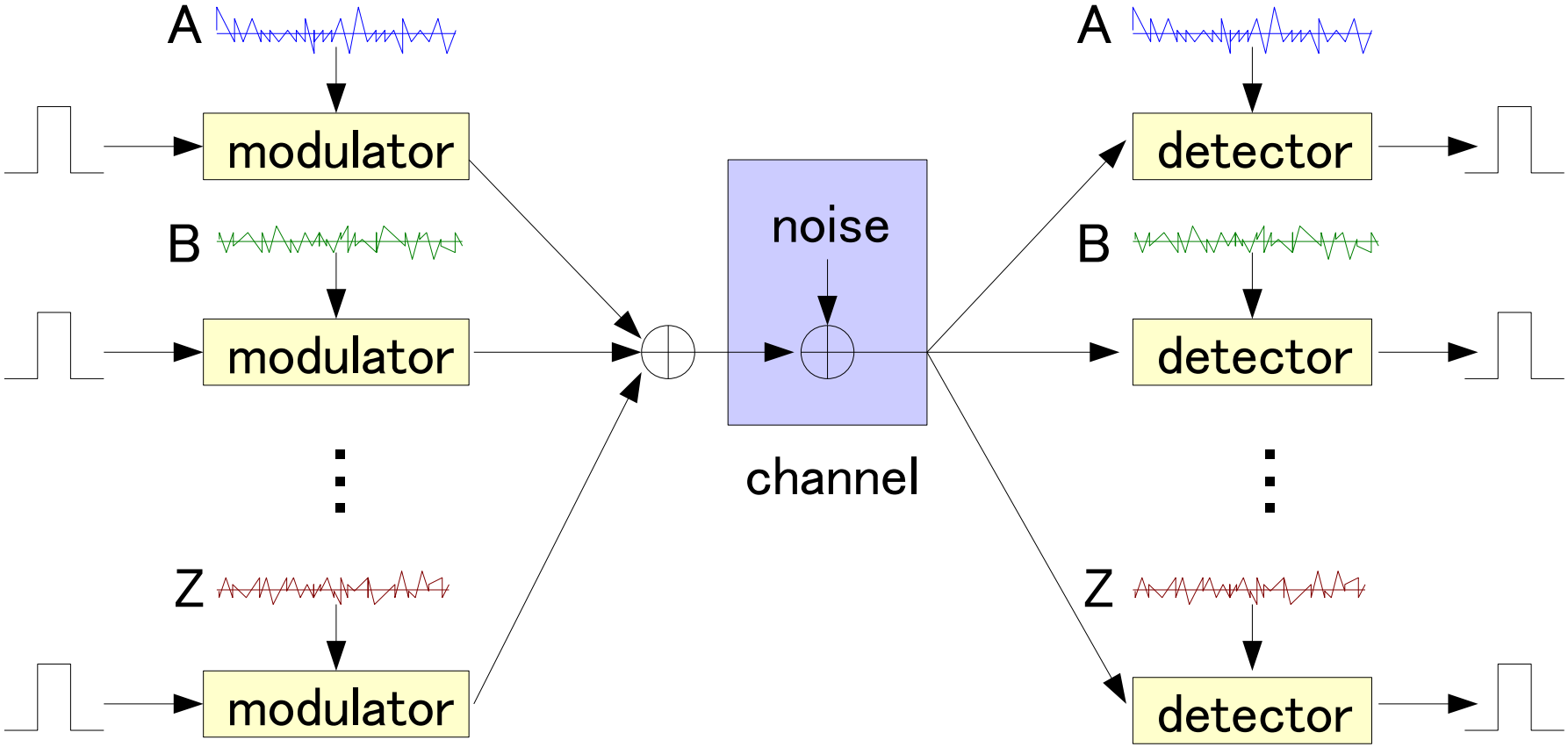
The idea of Cox's method is derived from the spread spectrum technique

The other techniques adapted in signal processing could be employed.

The Cox's method can be enhanced by **CDMA technique**.

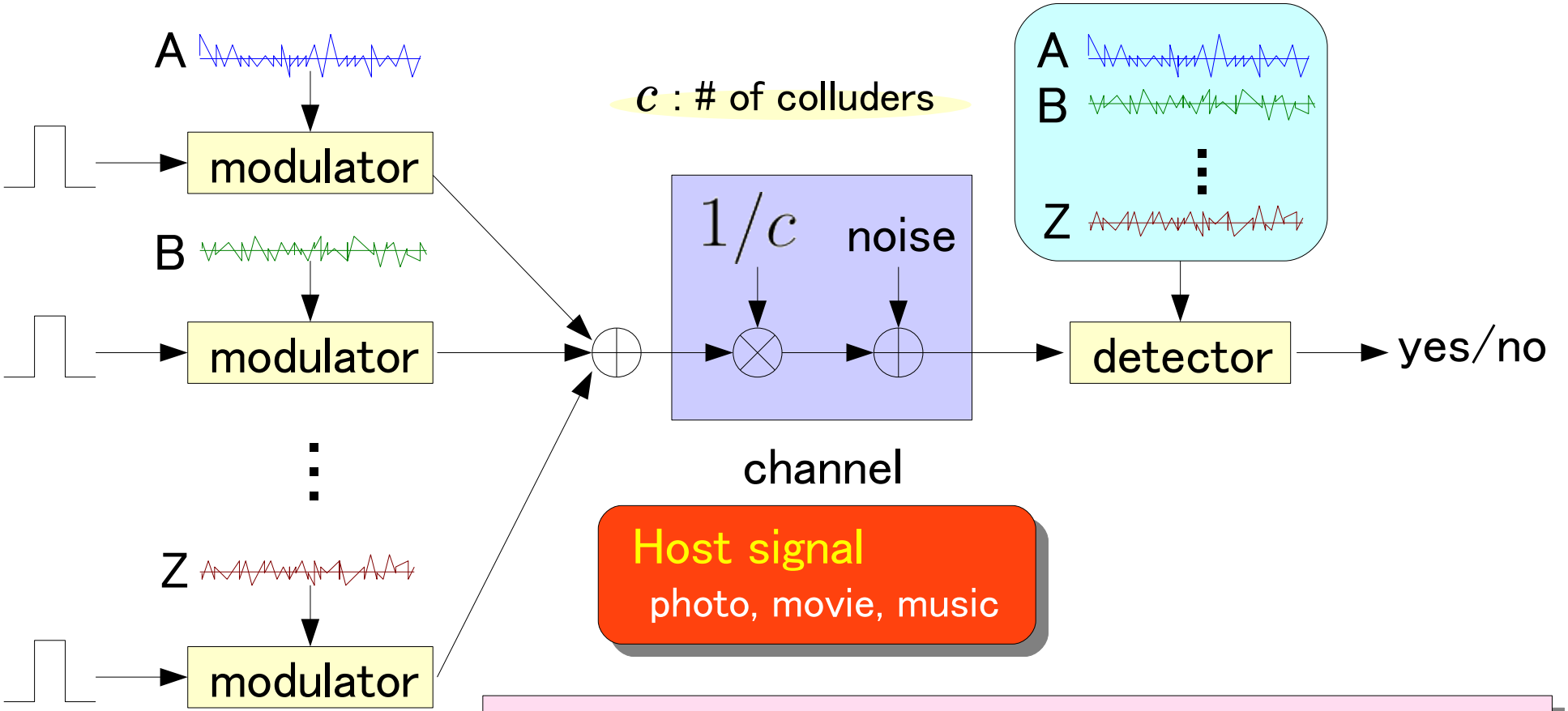
CDMA Communication Model

- Signals of some users are multiplexed in one communication channel
- Each detector checks the correlation with own PN sequence



CDMA-Based Fingerprinting Model

It follows similar channel model except for the number of signals stored in a detector



Averaging collusion = Some fingerprint sequences are multiplexed in the channel

CDMA-Based Fingerprinting Scheme

[IWSEC2007]

N. Hayashi, M. Kuribayashi, M. Morii
"Collusion-Resistant Fingerprinting Scheme Based on the CDMA-Technique"
Proc. IWSEC2007, LNCS 4752, pp.28-43, Springer, 2007.

- Based on the quasi-orthogonality, hierarchical structure is produced using two kinds of SS sequences.

group ID user ID

Fingerprint information (i_g, i_u)

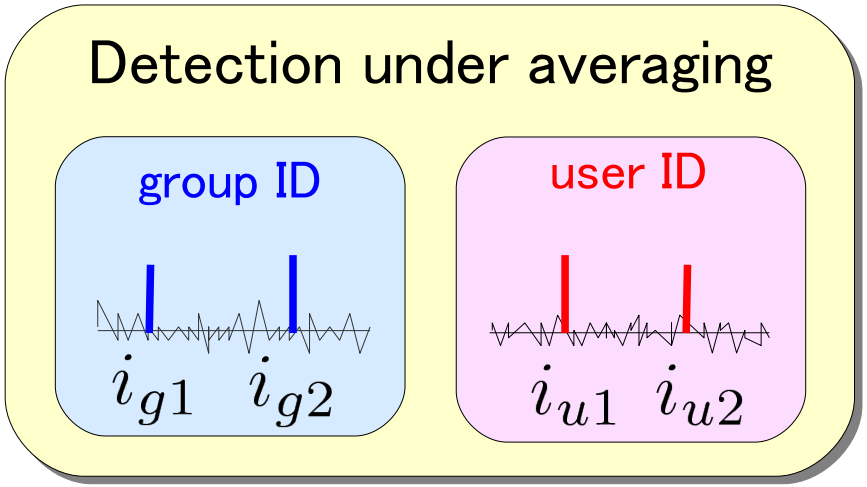
- Theoretically quasi-orthogonal sequences are designed using a PN sequence combined with orthogonal transform.

DCT

- A design of threshold is based on a given false-positive probability

Hierarchical Construction

If two sequences are simply applied ...



Possible patterns of colluders's ID

Case1: (i_{g1}, i_{u1}) and (i_{g2}, i_{u2})

Case2: (i_{g1}, i_{u2}) and (i_{g2}, i_{u1})

It is impossible to identify the combination.

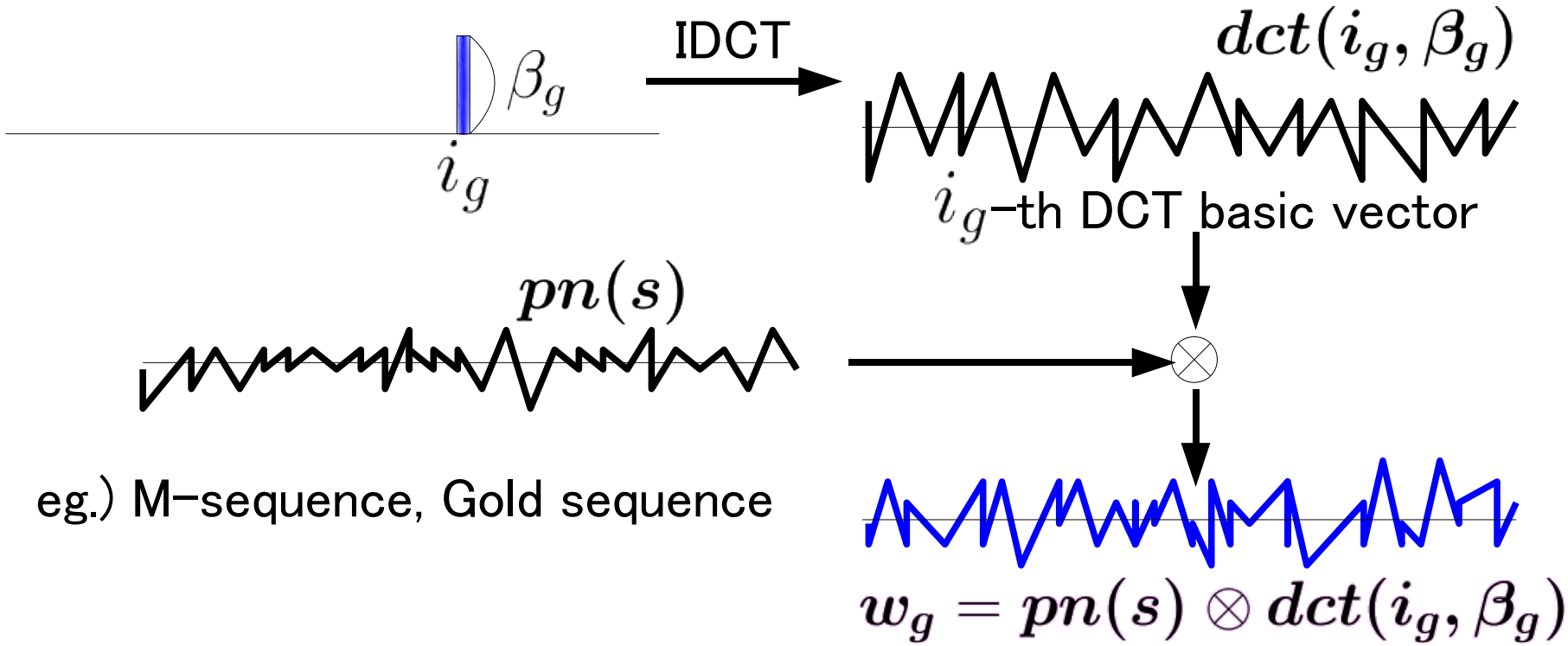
Our approach

We produce the dependency between group ID and user ID

How to Generate Fingerprint Sequence

Sequence for group ID

Mutually orthogonal



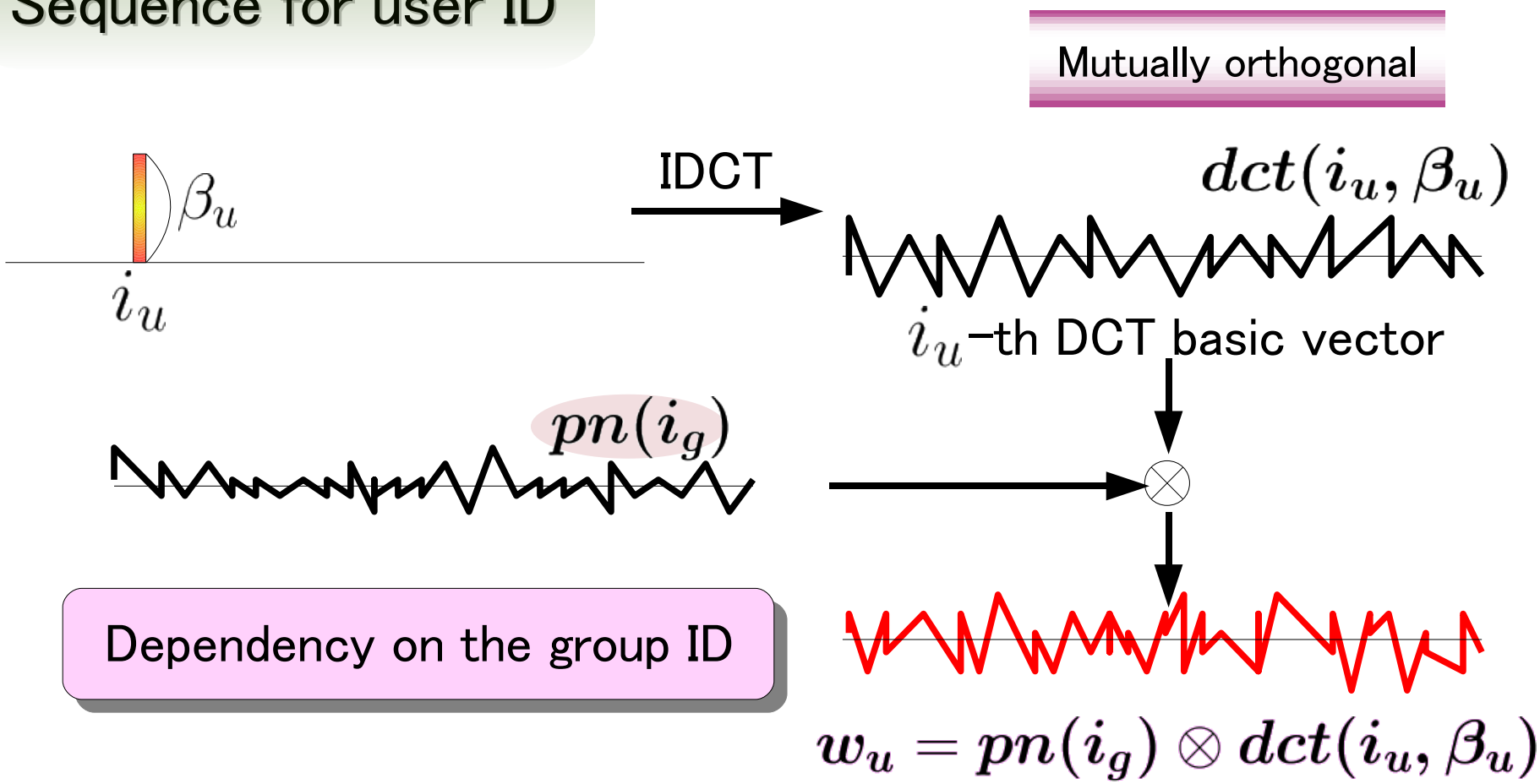
eg.) M−sequence, Gold sequence

Without the secret key s,

the estimation of the sequence is difficult.

How to Generate Fingerprint Sequence

Sequence for user ID



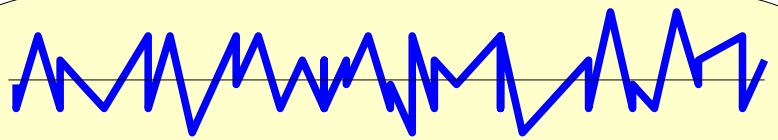
Dependency on the group ID

Without the detection of group ID, the detection of user ID is impossible.

How to Generate Fingerprint Sequence

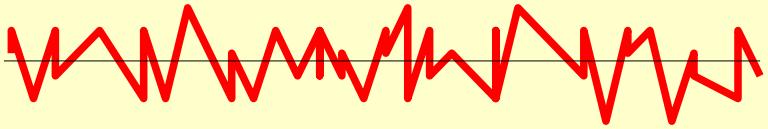
Fingerprint information (i_g, i_u)

Sequence for group ID



$$w_g = pn(s) \otimes dct(i_g, \beta_g)$$

Sequence for user ID



$$w_u = pn(i_g) \otimes dct(i_u, \beta_u)$$

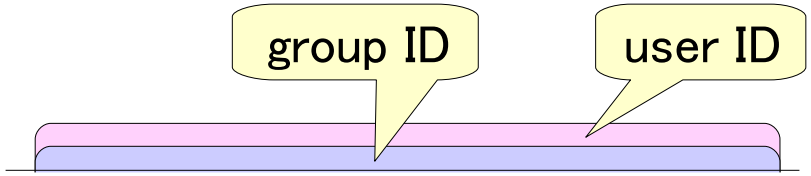
s : secret key

β_g : embedding strength

β_u: embedding strength

fingerprint sequence

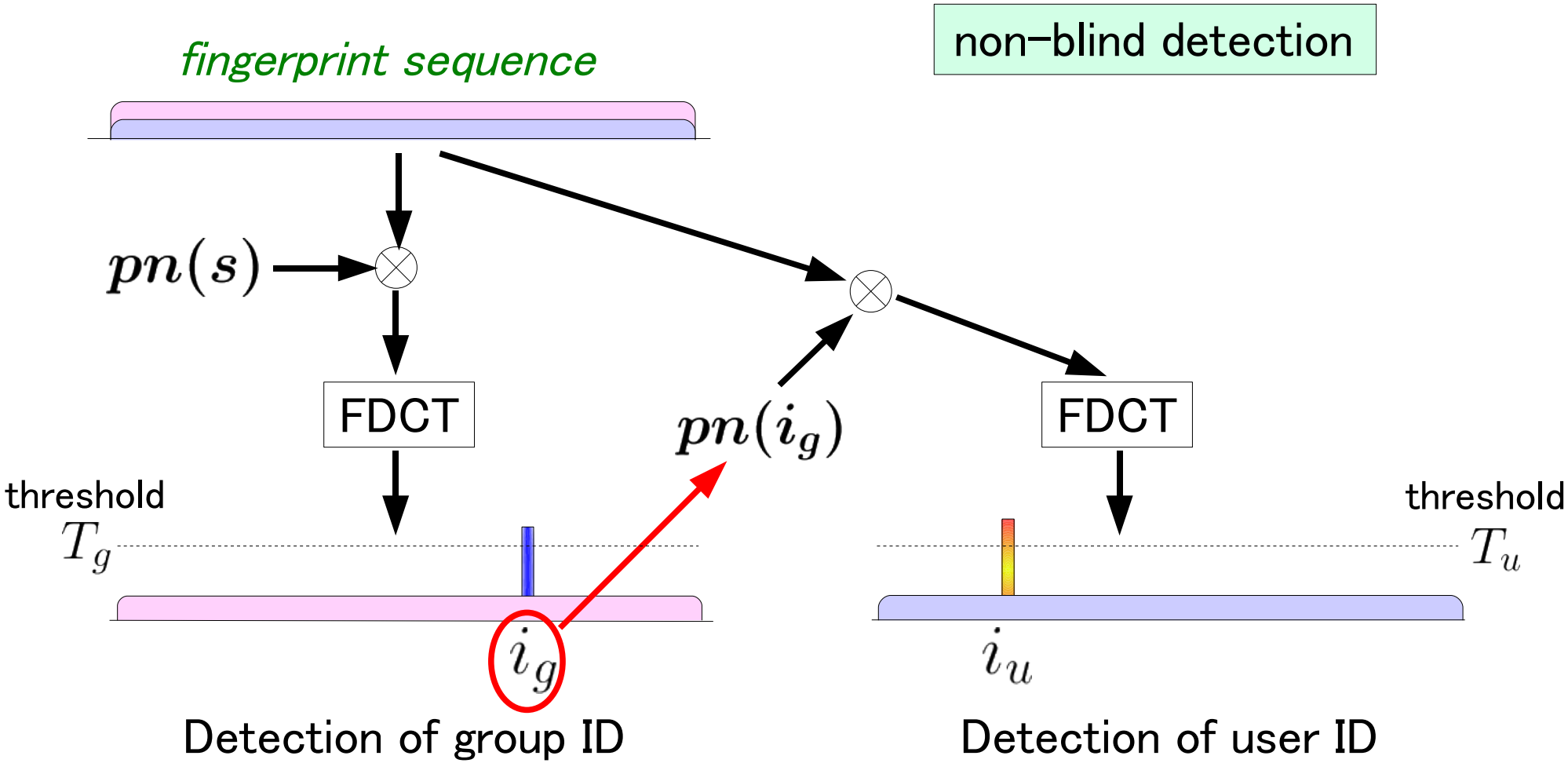
Two kinds of signals are multiplexed.



The length of sequence is ℓ

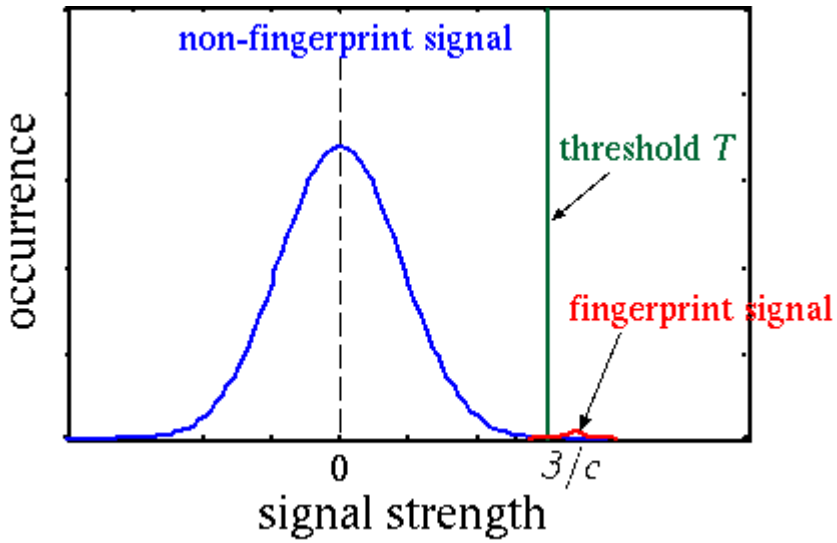
Detection Procedure

From the DCT coefficients of fingerprinted image, the fingerprint sequence is obtained by subtracting from the original coefficients.



Design of Proper Threshold

Considering the characteristic of detected signals, we can find that non-fingerprint signal follows Gaussian distribution with mean zero.



The statistical analysis gives

$$Pe = \frac{1}{2}\mathrm{erfc}\left(\frac{T}{\sqrt{2\sigma^2}}\right)$$

Pe : false-positive probability

σ^2 : variance

When a false-positive probability Pe is given, the corresponding threshold T can be calculated.

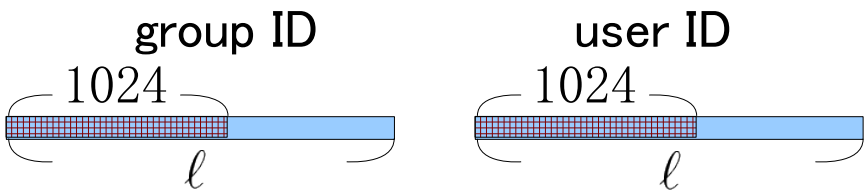
Simulation Results

image : “Lena” (512×512 pixel, 256-level gray scale)

length ℓ : $1024, 2048, 4096, 8192$

# of user : 2^{20} (1 million)

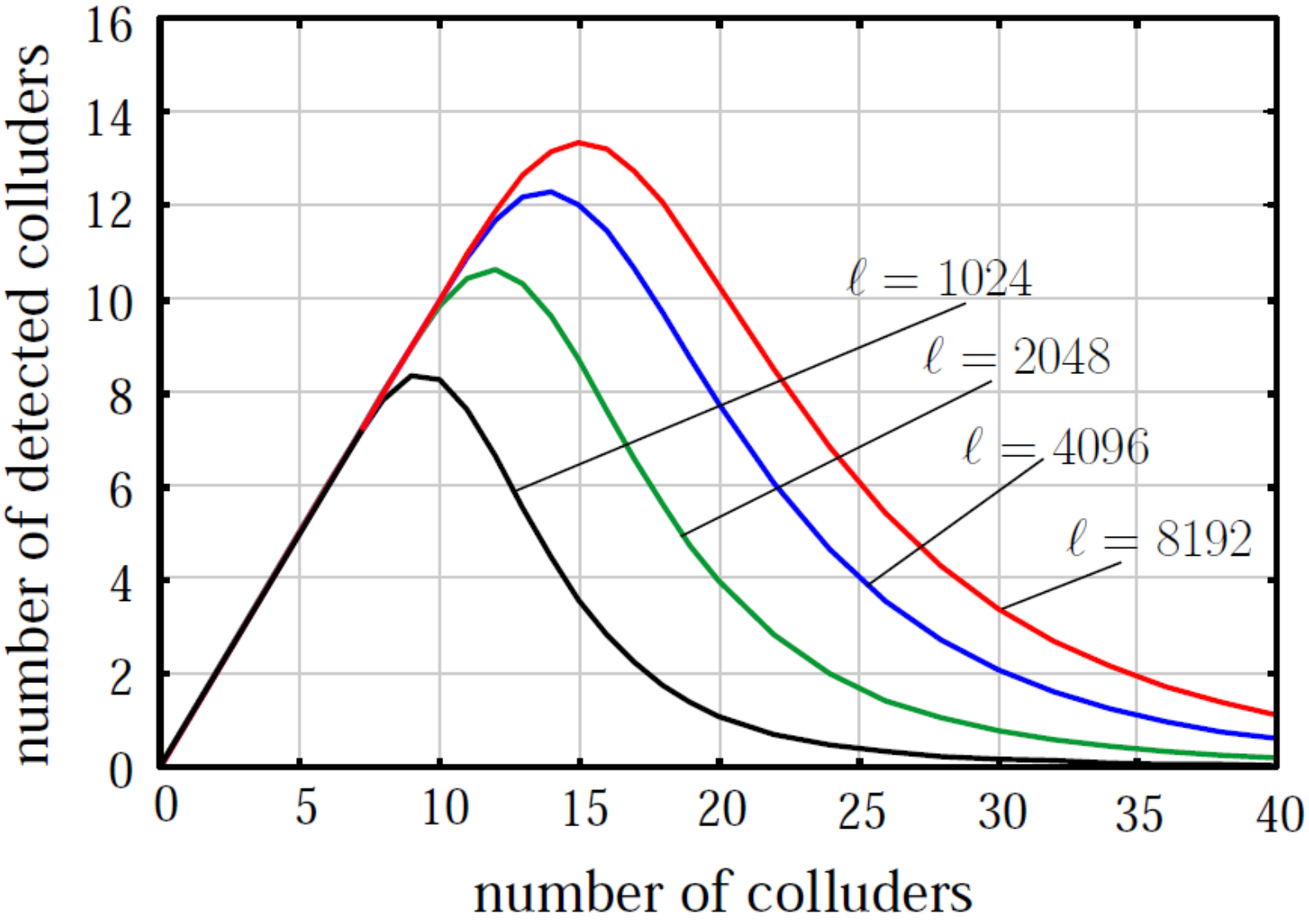
1024x1024



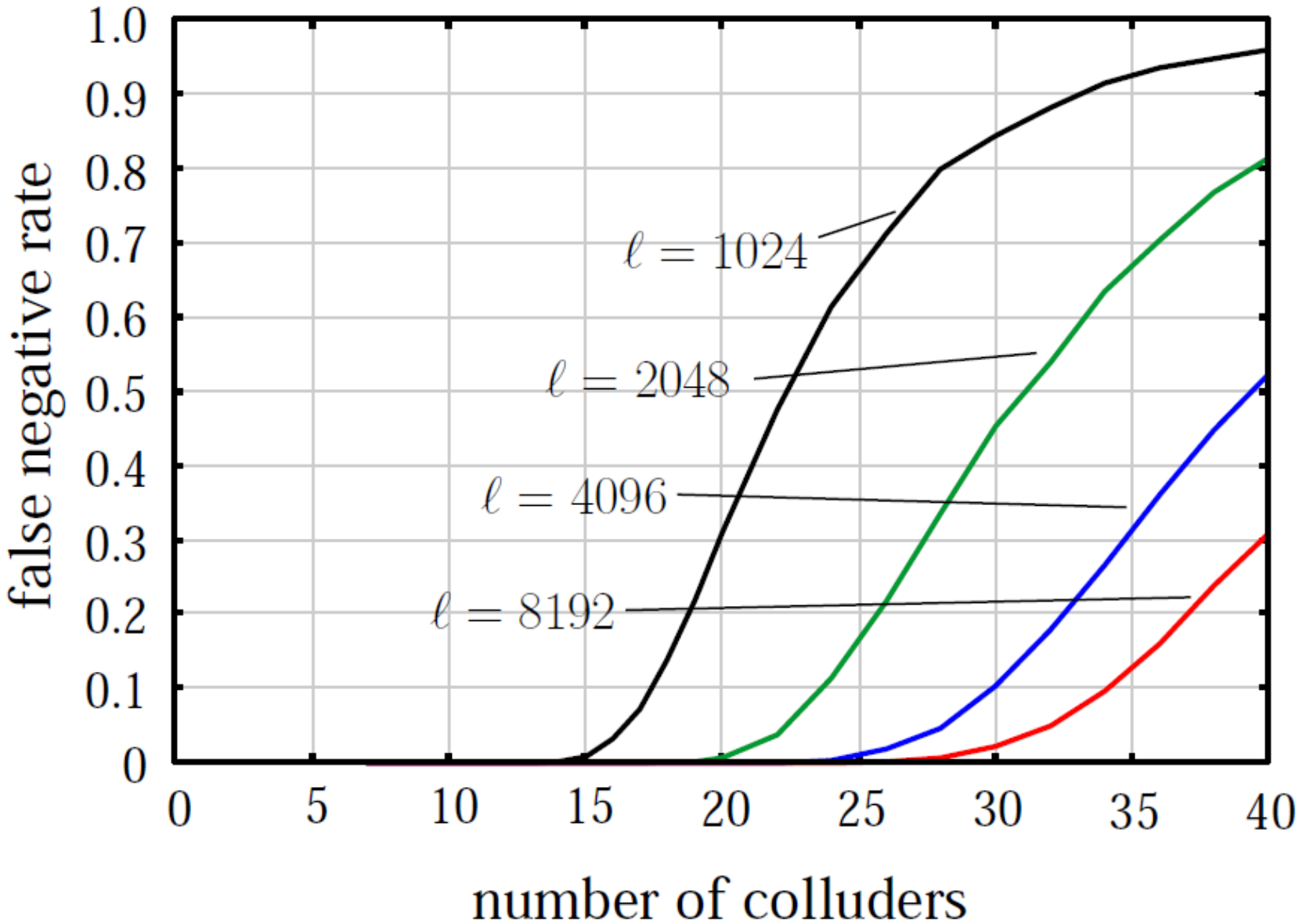
quality of fingerprinted image : PSNR = 45 [dB]

attack : averaging collusion + JPEG compression (35%)

True Positive Detection

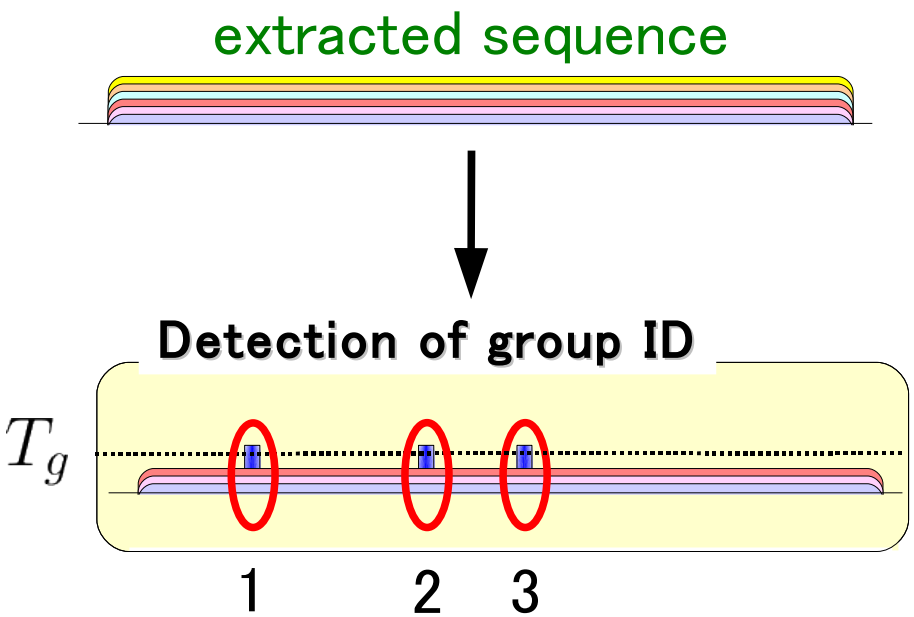


False-Negative Probability



Interference

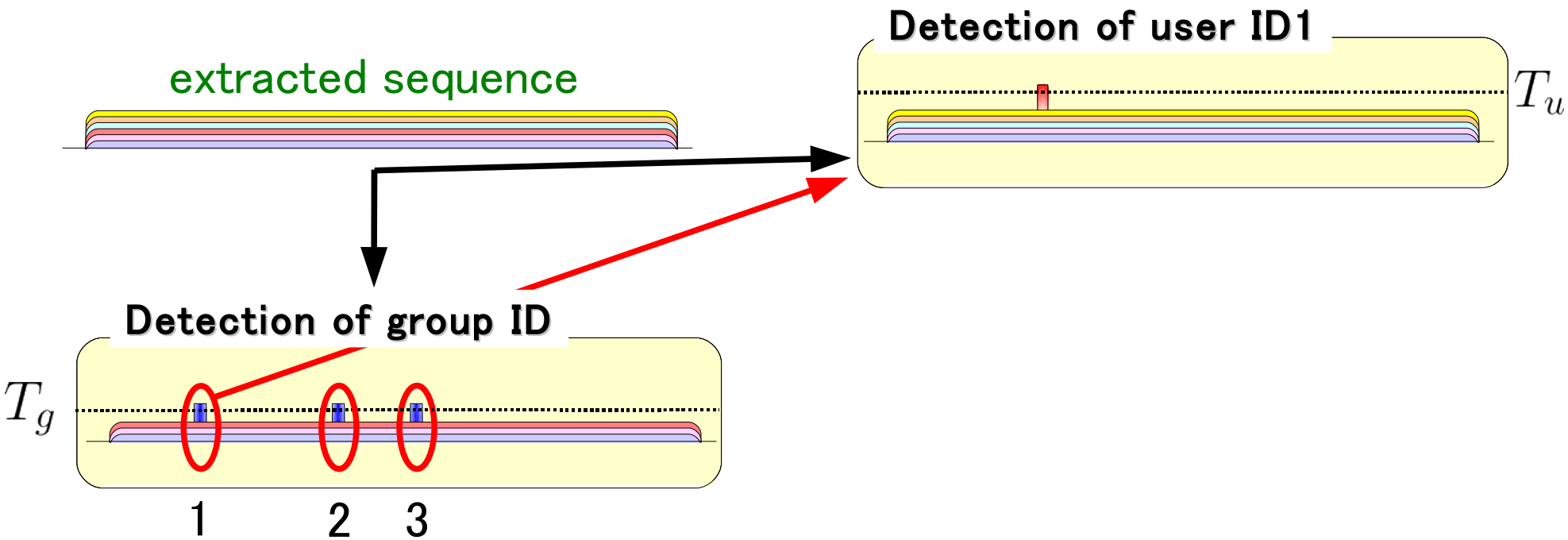
- When 3 users are colluded and their contents are averaged, 6 sequences are multiplexed.



- ▷ Embedded signal energy is attenuated by 1/3.
- ▷ Each fingerprint sequence remains as interference of other sequences.

Interference

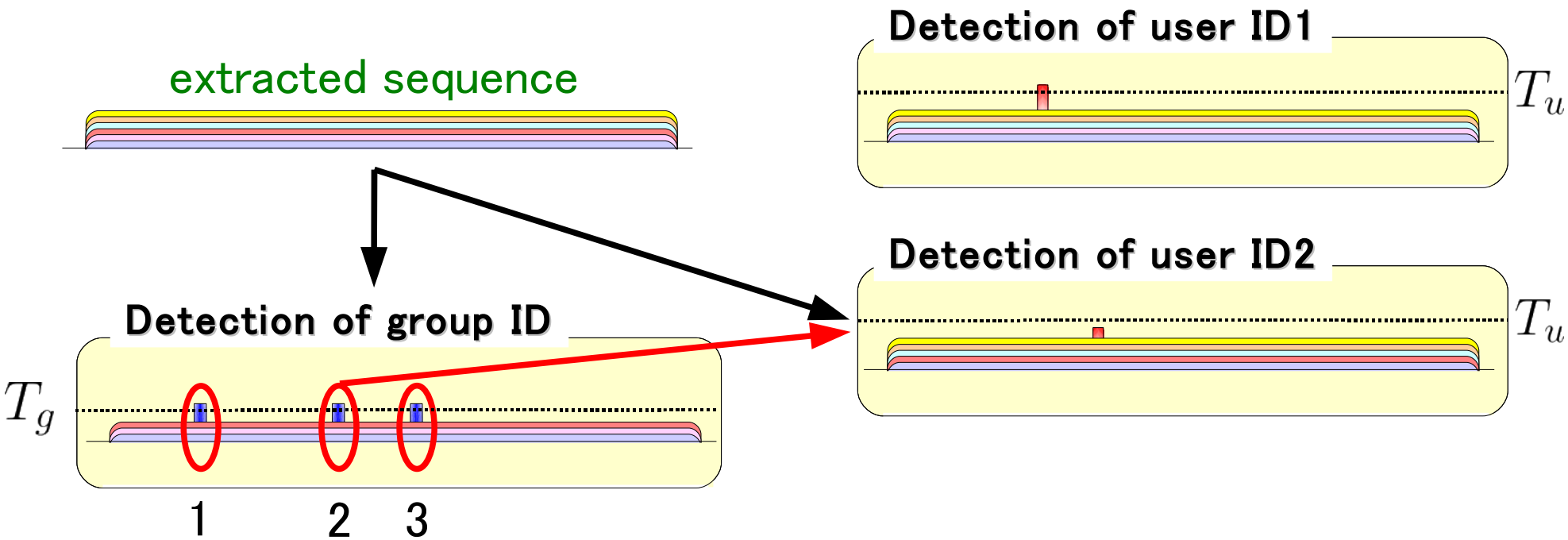
- When 3 users are colluded and their contents are averaged, 6 sequences are multiplexed.



- ▷ Embedded signal energy is attenuated by 1/3.
- ▷ Each fingerprint sequence remains as interference of other sequences.

Interference

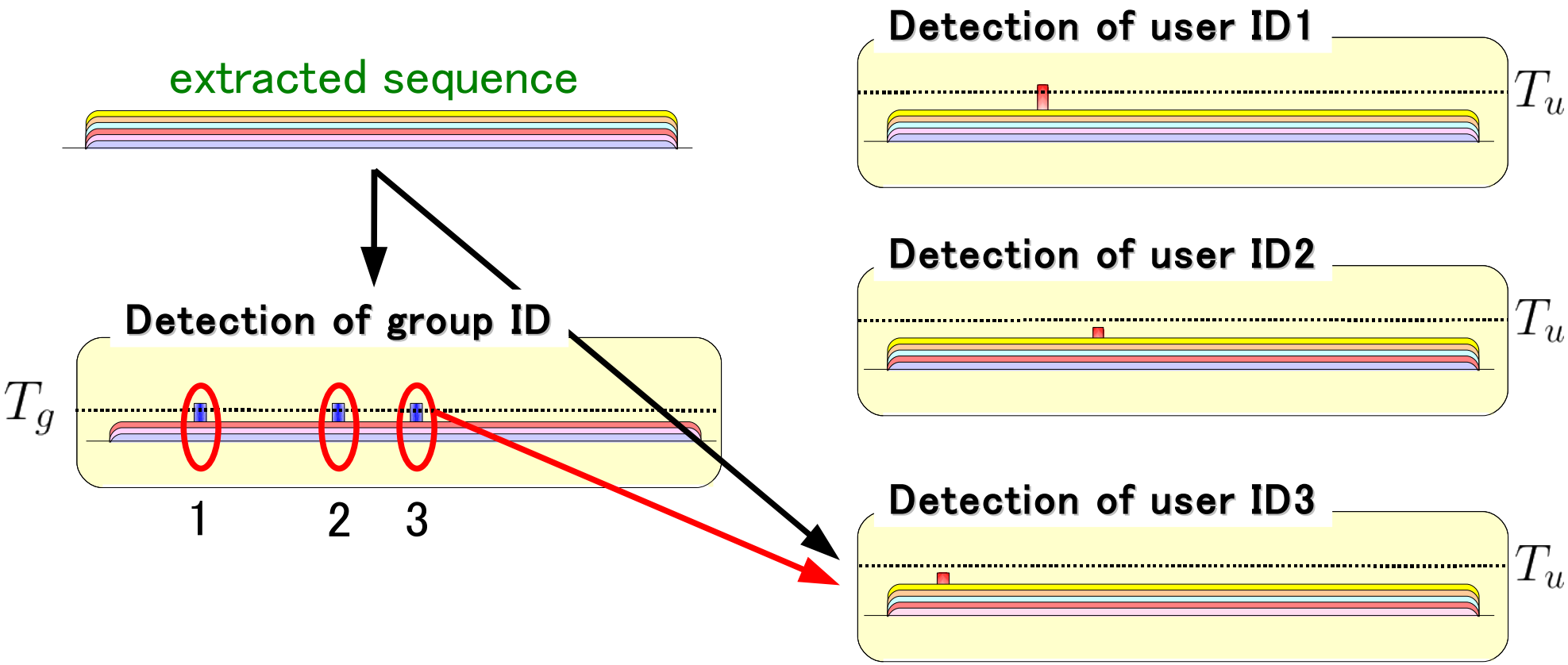
- When 3 users are colluded and their contents are averaged, 6 sequences are multiplexed.



- Embedded signal energy is attenuated by 1/3.
- Each fingerprint sequence remains as interference of other sequences.

Interference

- When 3 users are colluded and their contents are averaged, 6 sequences are multiplexed.



- Embedded signal energy is attenuated by 1/3.
- Each fingerprint sequence remains as interference of other sequences.

Effective Detection Method

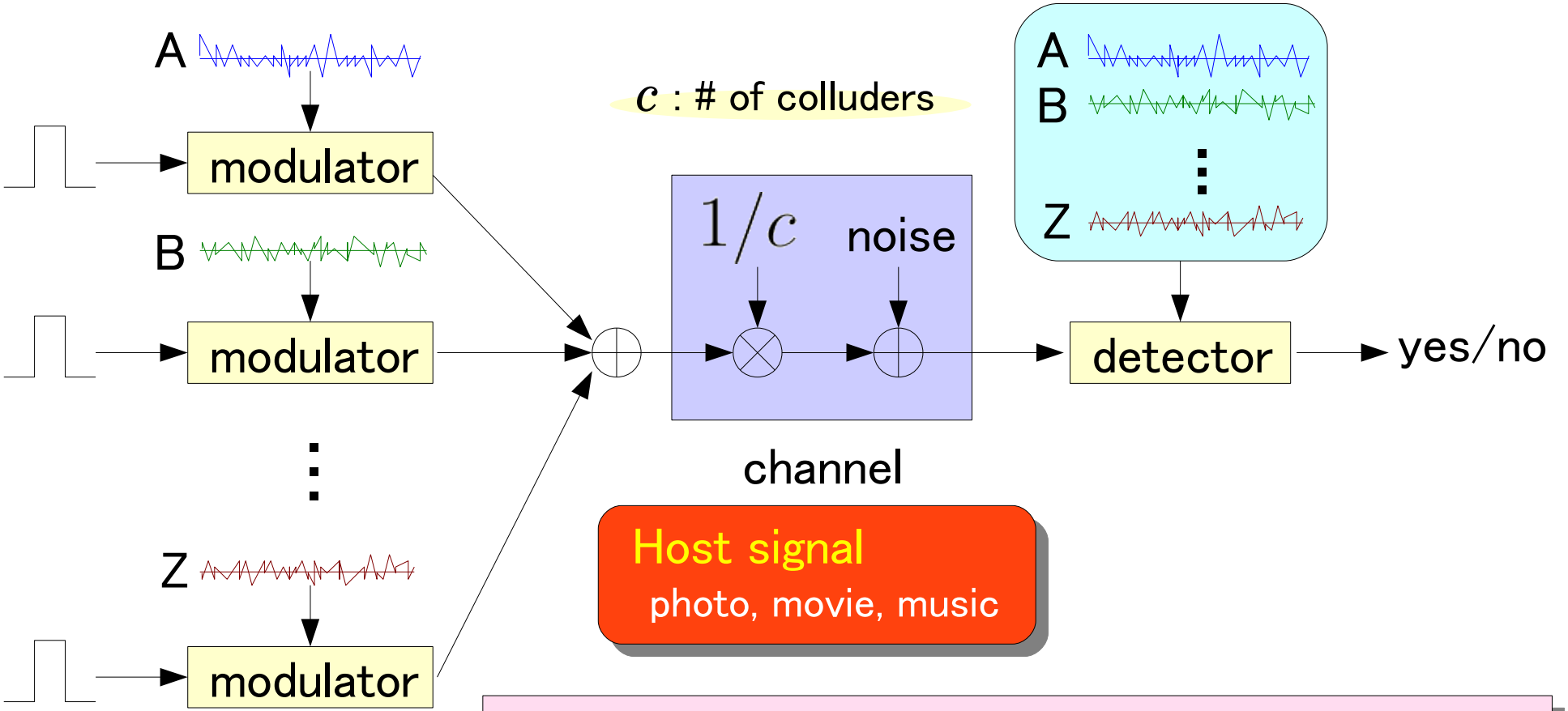
[IH2008]

M. Kuribayashi and M. Morii
"Iterative Detection Method for CDMA-Based Fingerprinting Scheme,"
Proc. IH2008, LNCS 5284, pp.357-371, Springer, 2008.

- Removal operation
  - Interference among fingerprint sequences is effectively reduced.
- Iterative detection
  - Until no signal can be found, the detection operation is iteratively performed after the removal of interference.
- Two kinds of thresholds
  - The removal operation is adaptively performed for detected signals.

CDMA-Based Fingerprinting Model

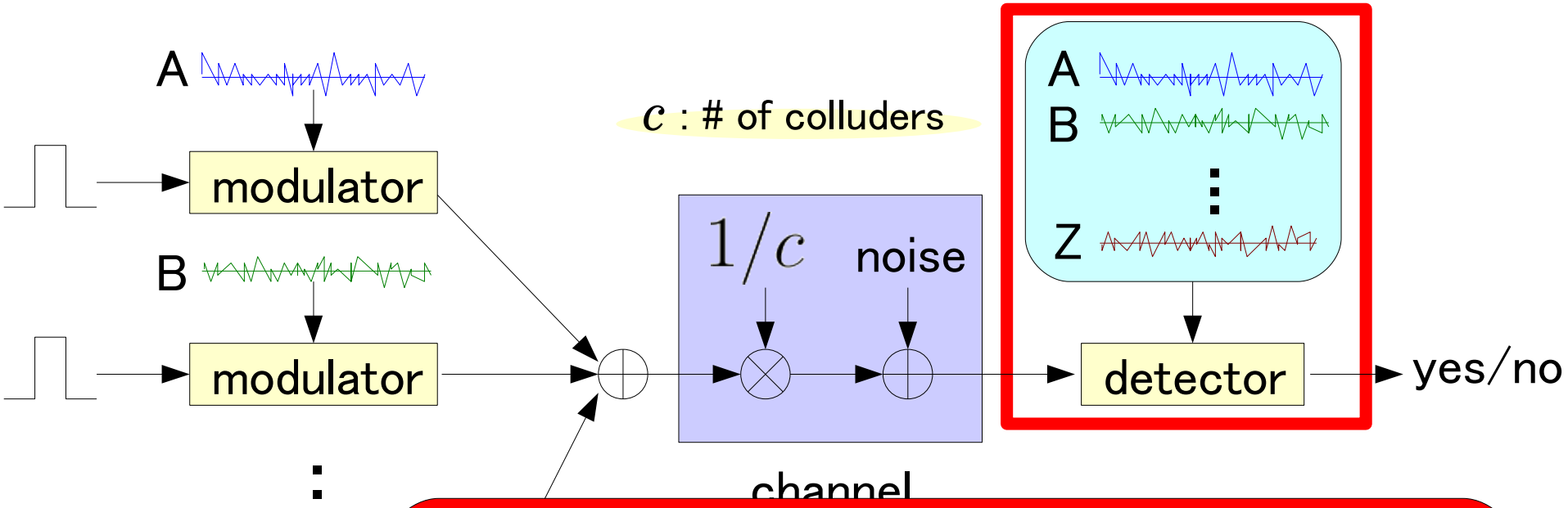
It follows similar channel model except for the number of signals stored in a detector



Averaging collusion = Some fingerprint sequences are multiplexed in the channel

CDMA-Based Fingerprinting Model

It follows similar channel model except for the number of signals stored in a detector

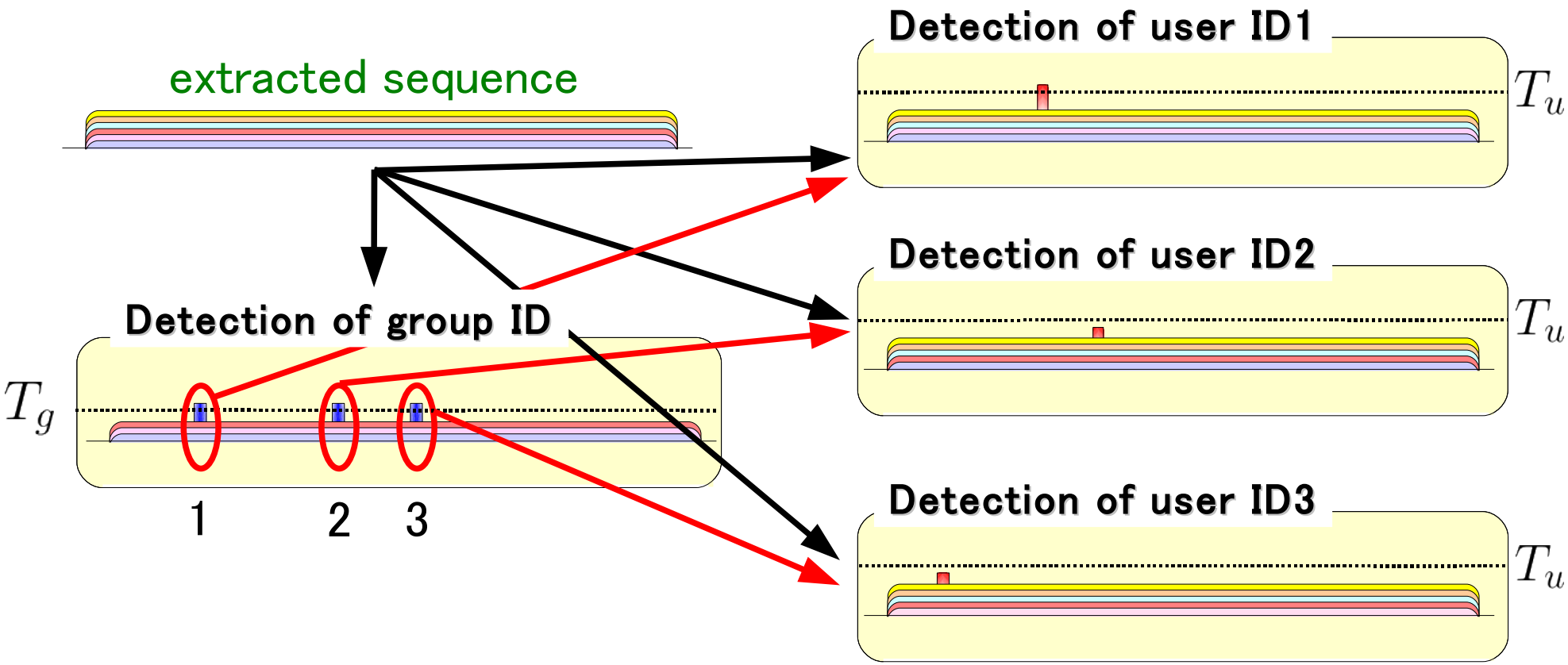


The detector stores all candidates

By removing detected signals, the interference among fingerprints can be reduced.

Interference

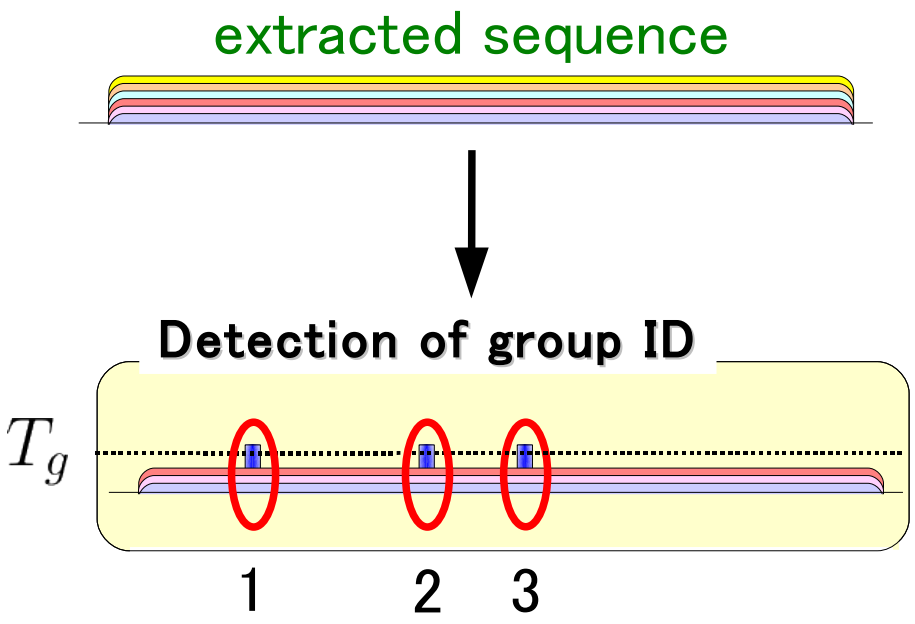
- When 3 users are colluded and their contents are averaged, 6 sequences are multiplexed.



- Embedded signal energy is attenuated by 1/3.
- Each fingerprint sequence remains as interference of other sequences.

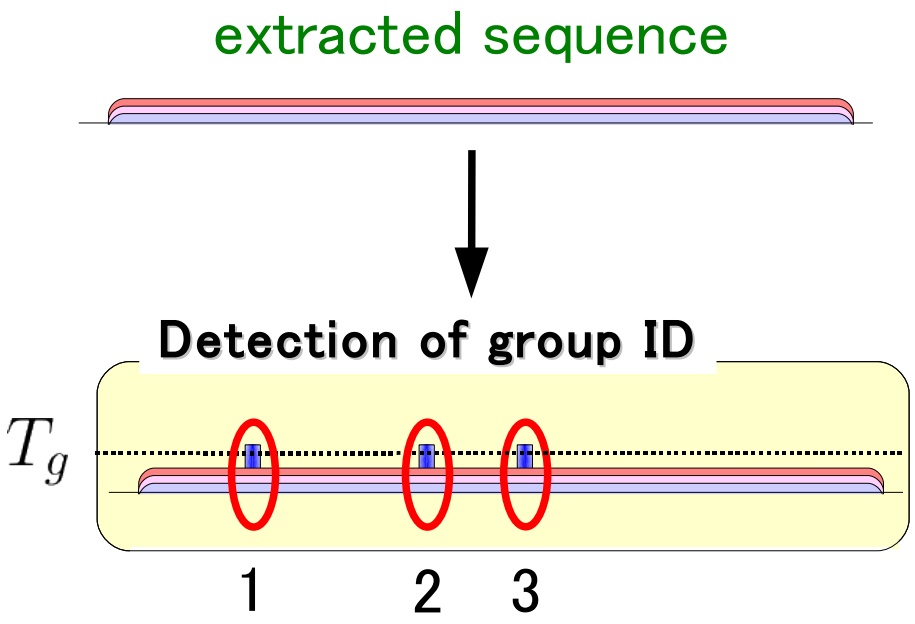
Removal of interference

- When fingerprints are detected,
  the signals are removed from the detected sequence to reduce the interference.



Removal of interference

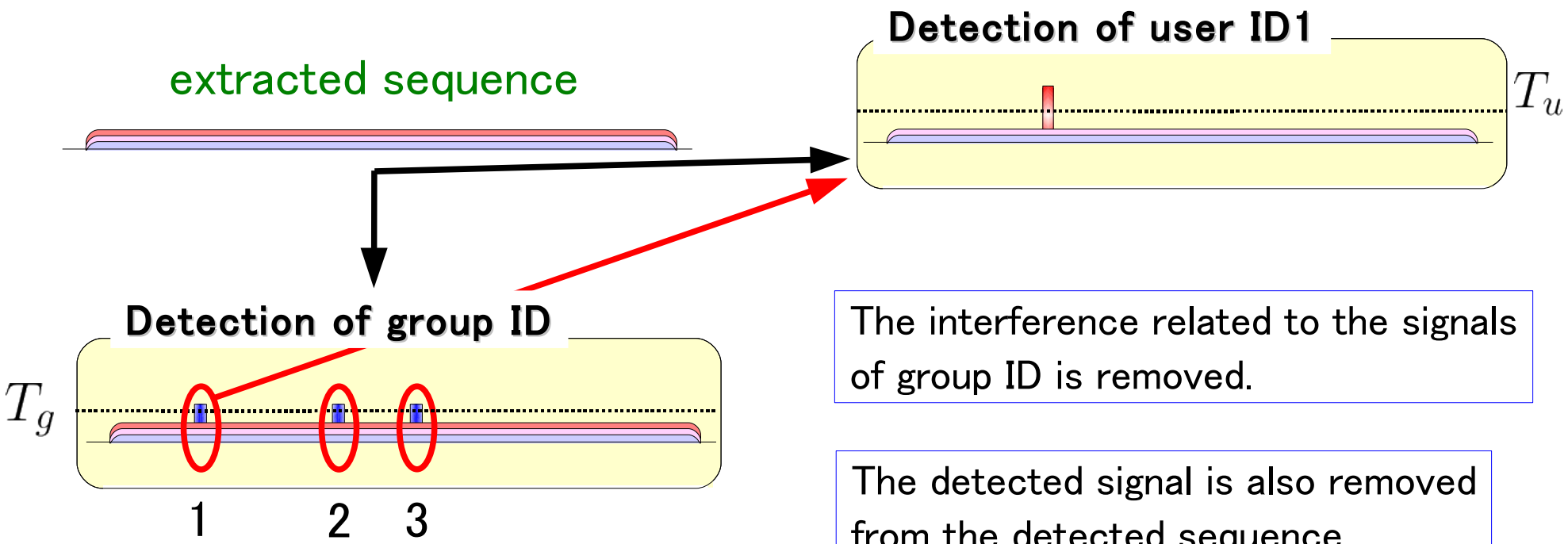
- When fingerprints are detected,
  the signals are removed from the detected sequence to reduce the interference.



Remove the 3 detected signals

Removal of interference

- When fingerprints are detected,
  the signals are removed from the detected sequence to reduce the interference.

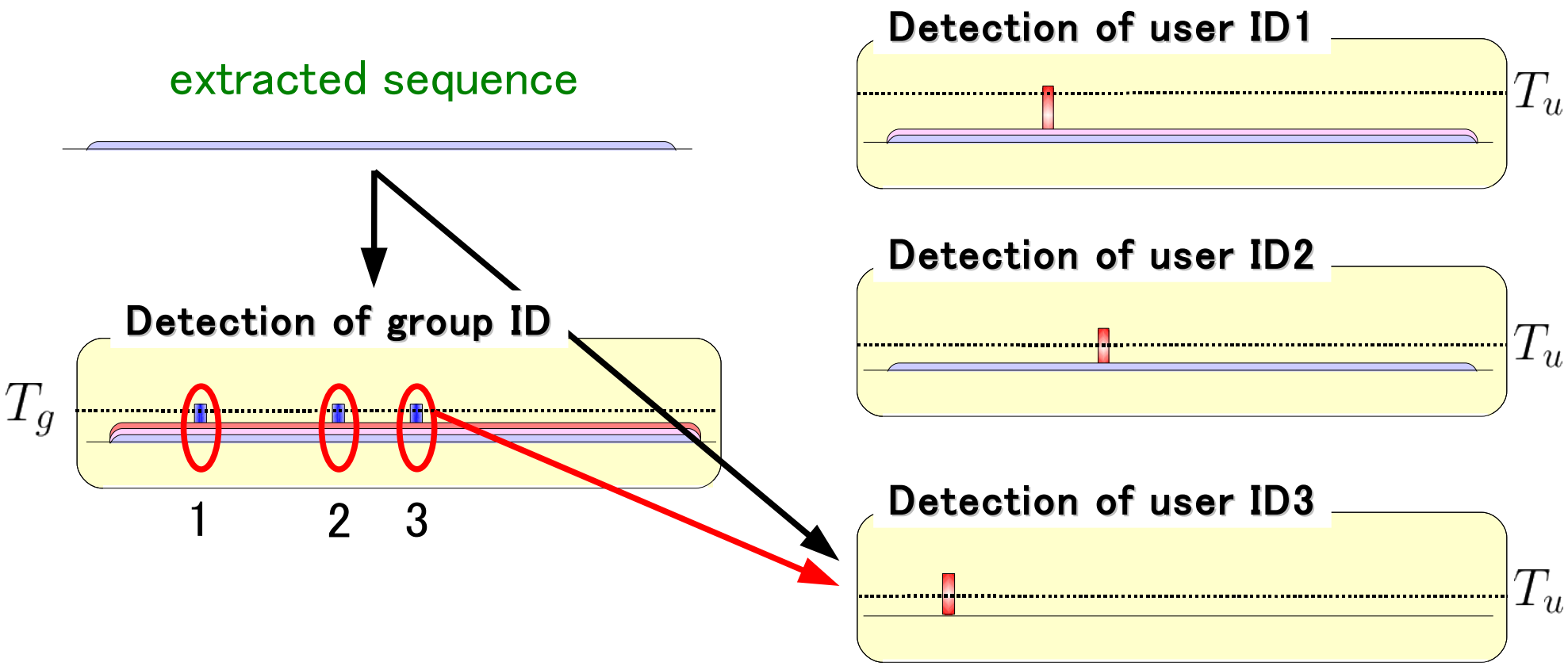


The interference related to the signals of group ID is removed.

The detected signal is also removed from the detected sequence.

Removal of interference

- When fingerprints are detected,
  the signals are removed from the detected sequence to reduce the interference.

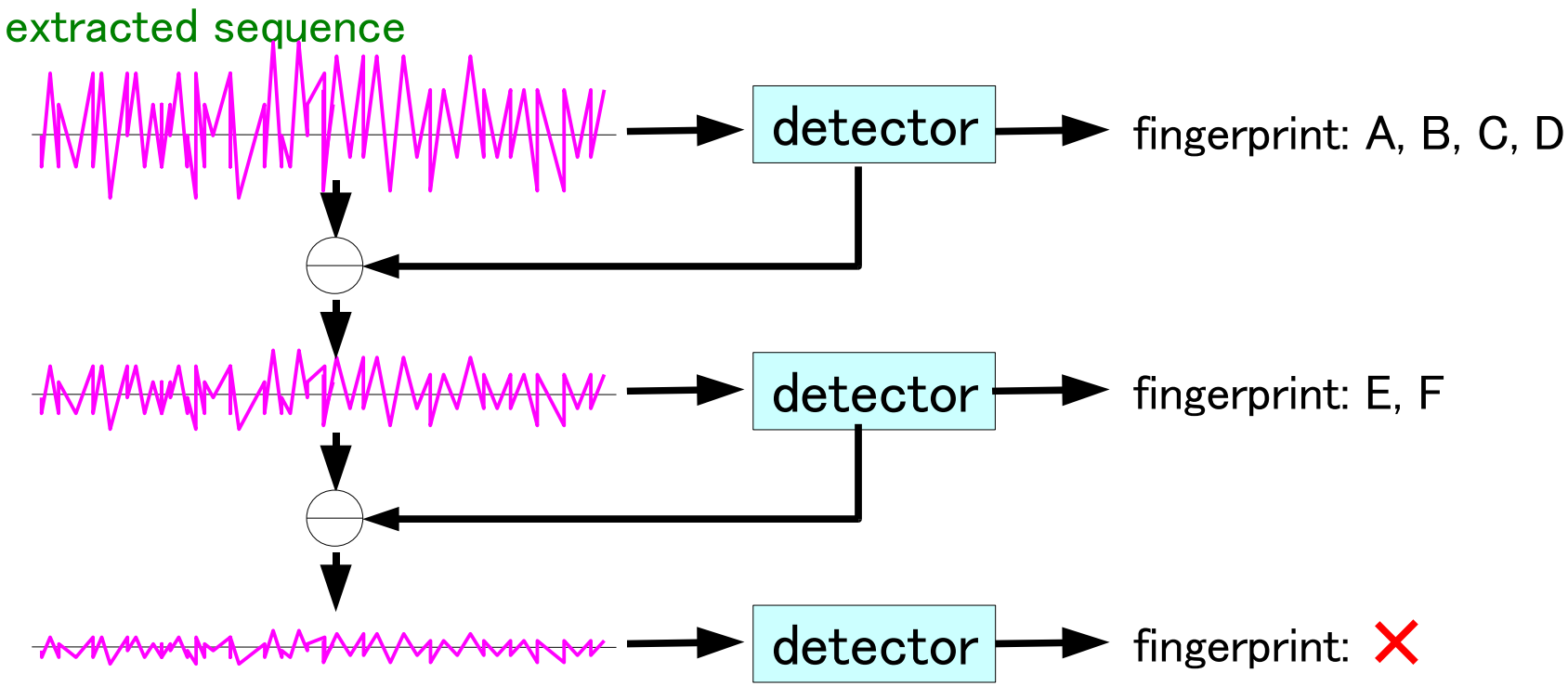


The interference of fingerprint signals is completely removed.

Iterative Detection

Because the number of colluders are unknown at the detector, the iterative detection with removal operation improves the traceability.

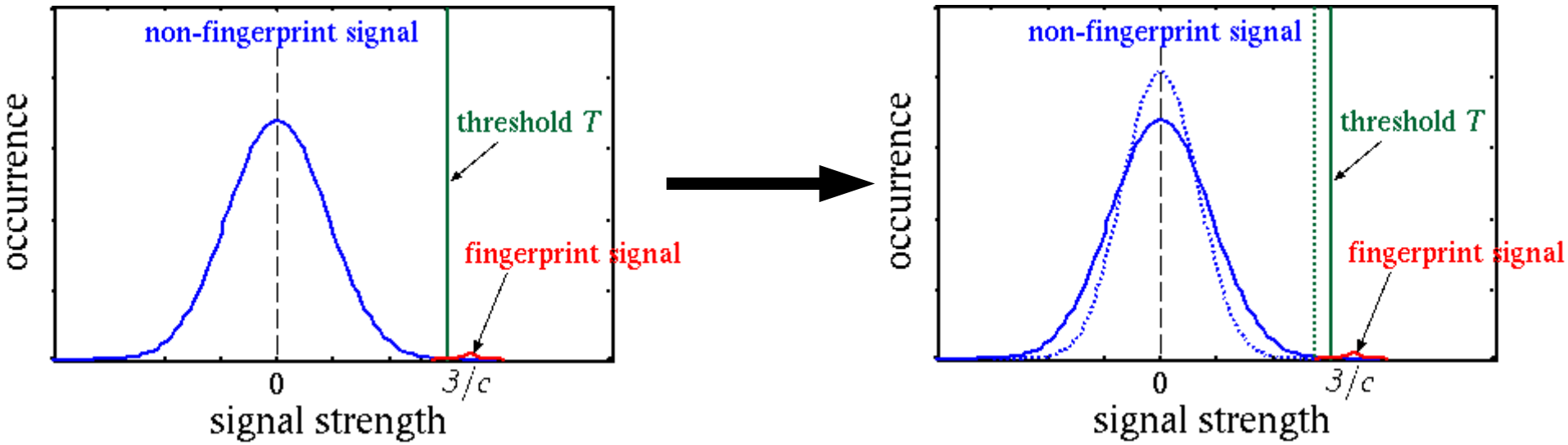
e.x) 6 colluders (A, B, C, D, E, F)



When no fingerprint is detected, the iteration is stopped.

Analysis of Removal Operation

The variance σ^2 of interference is decreased by the removal operation.

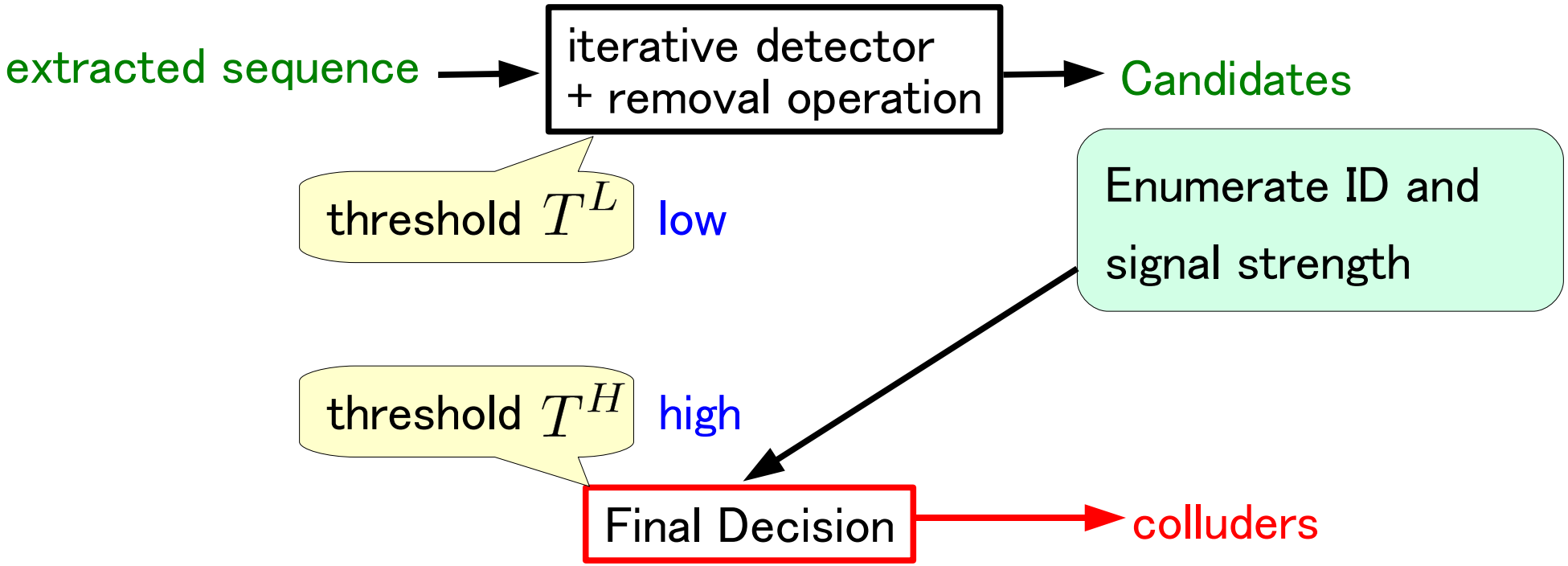


Remember that the false-positive probability is given by

$$\boxed{Pe = \frac{1}{2}\mathrm{erfc}\left(\frac{T}{\sqrt{2\sigma^2}}\right)}$$

Under the equal false-positive probability, the threshold can be set lower.

Two Kinds of Thresholds

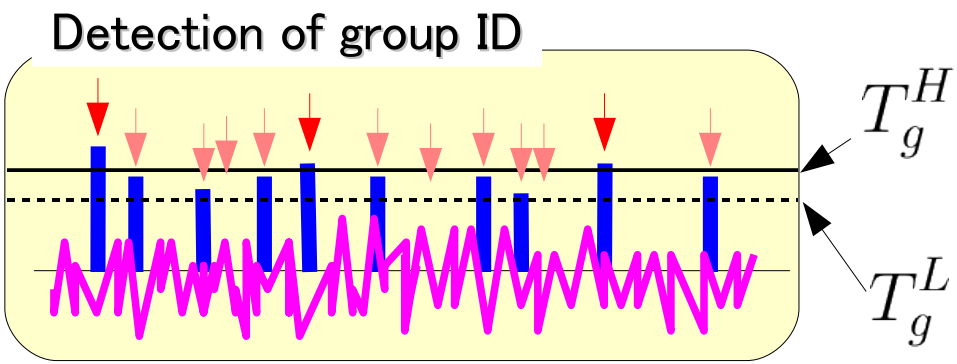


lower threshold T^L : detect candidates as many as possible

higher threshold T^H : decrease the false-positive detection

Two Kinds of Thresholds

- In order to remove the detected signals adaptively, two kinds of thresholds are introduced for the detection.



- If signal energy exceeds T_g^H,

  it is obviously determined by colluders' group ID.

  ⬇

  Detected signals are removed for the reduction of interference.

- If signal energy is within the range $[T_g^L, T_g^H]$,

  it may be a colluders group ID.

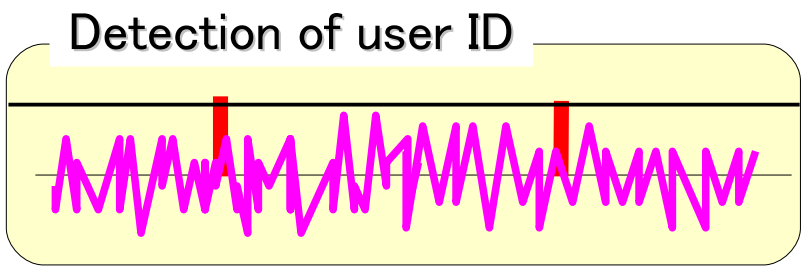
  ⬇

  It is detected as a potential candidates.

  The signal is removed only when the corresponding user ID is detected.

Two Kinds of Thresholds

- For each candidate (group ID), the detection of user ID is performed.



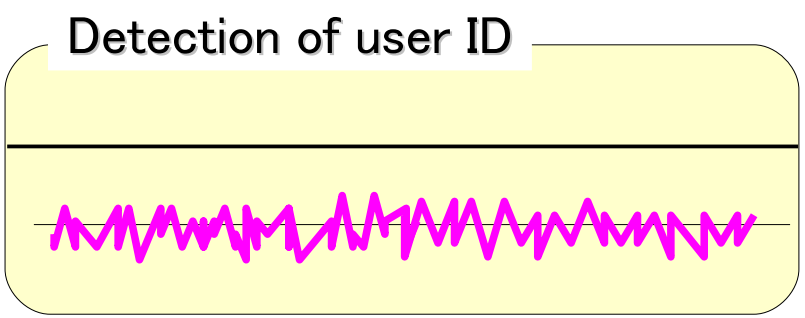
T_u^L

At first, we use a lower threshold T_u^L to avoid the false-negative.

The strength of detected signal is stored

- After the iterative detection, some pairs of fingerprint information are listed.

The detection operation is performed once again to calculate the higher threshold T_u^H.

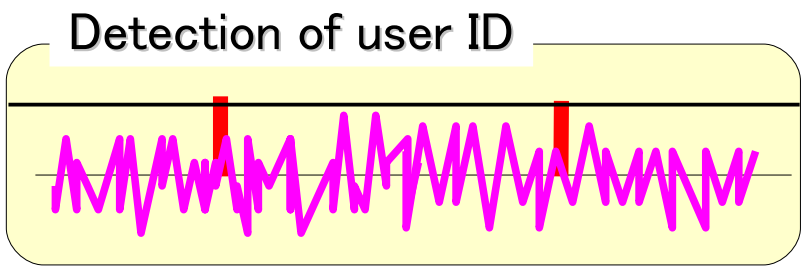


T_u^H

Wrongly detected signals will be decreased after the removal of interference..

Two Kinds of Thresholds

- For each candidate (group ID), the detection of user ID is performed.



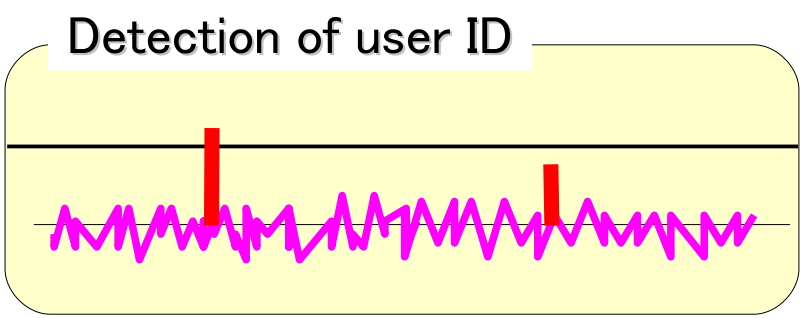
T_u^L

At first, we use a lower threshold T_u^L to avoid the false-negative.

The strength of detected signal is stored

- After the iterative detection, some pairs of fingerprint information are listed.

The detection operation is performed once again to calculate the higher threshold T_u^H.



T_u^H

By adding the stored signal on it, more accurate decision is possible.

Wrongly detected signals will be decreased after the removal of interference..

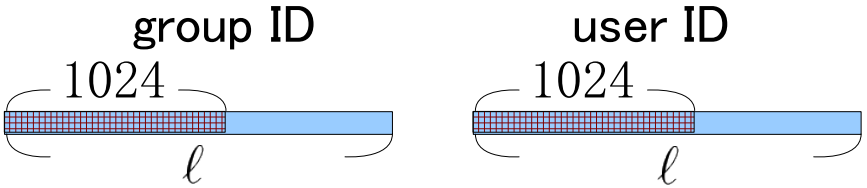
Simulation Results

image : “Lena” (512×512 pixel, 256-level gray scale)

length ℓ : $1024, 2048, 4096, 8192$

# of user : 2^{20} (1 million)

1024x1024



proposed I : iterative detection + removal operation

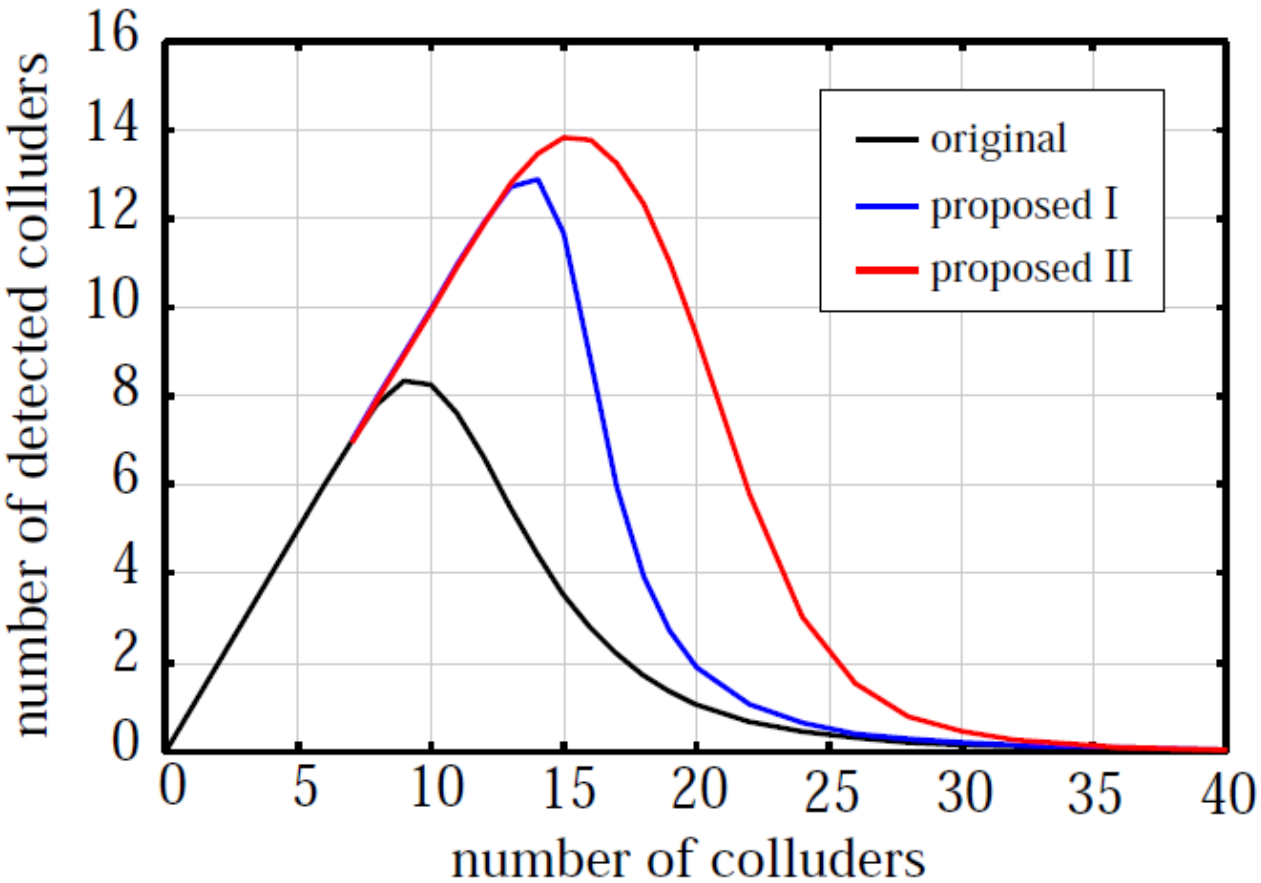
proposed II : proposed I + two kinds of thresholds

quality of fingerprinted image : PSNR = 45 [dB]

attack : averaging collusion + JPEG compression (35%)

True Positive Detection

$\ell = 1024$



original, proposed I

$Pe_g = 10^{-3}$

$Pe_u = 10^{-8}$

proposed II

$Pe_g^H = 5\times10^{-3}$

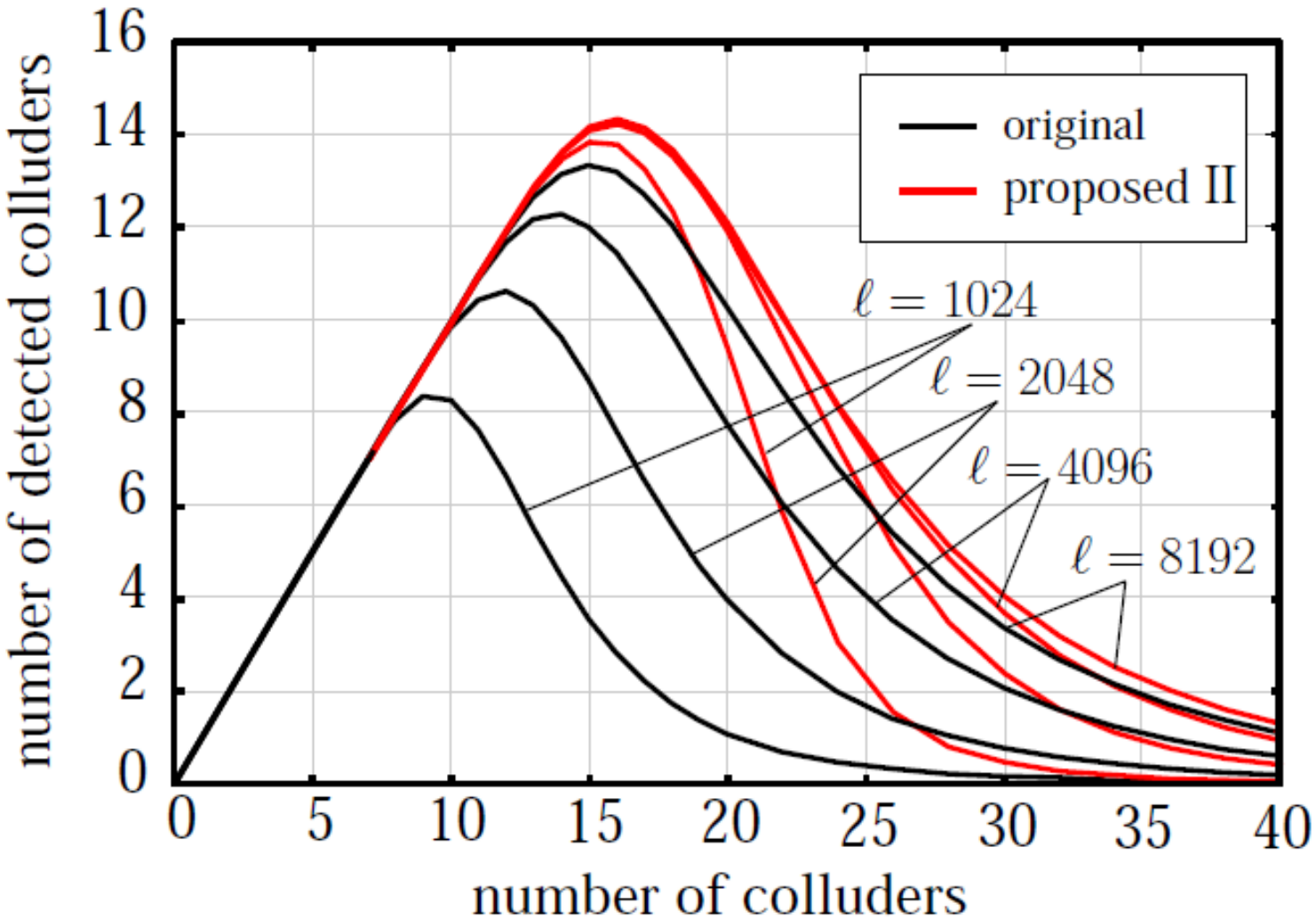
$Pe_g^L = 10^{-4}$

$Pe_u^H = 2.5\times10^{-9}$

$Pe_u^L = 10^{-5}$

Our detector improves the traceability.

True Positive Detection



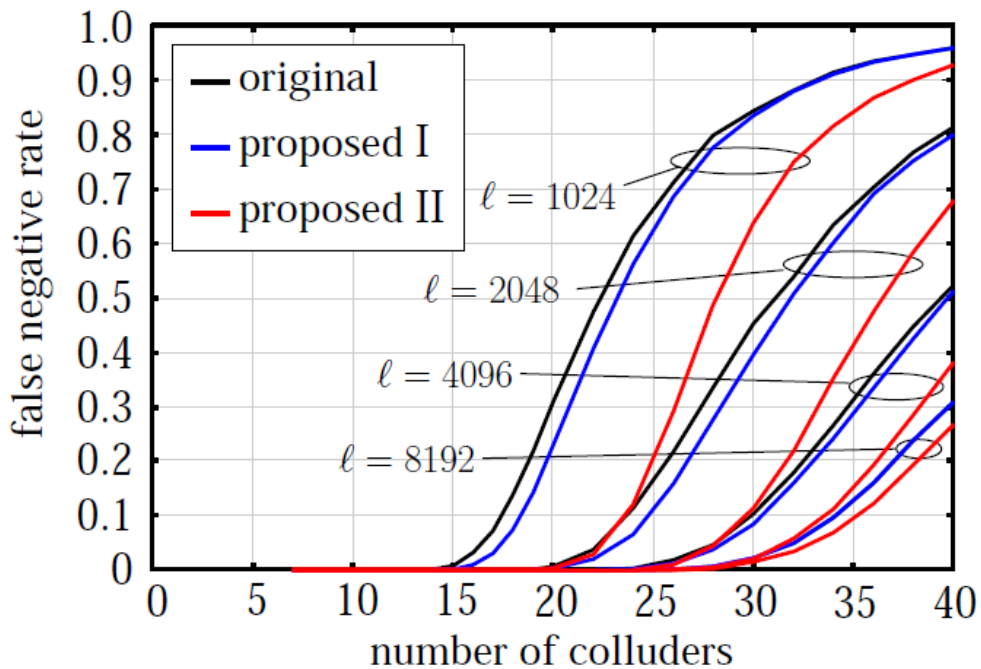
It confirms that our detector improves the true positive detection.

False Probabilities

- False Positive Detection [$\times 10^{-4}$]

length ℓ	original	proposed I	proposed II
1024	2.00	2.63	3.04
2048	2.08	3.08	1.08
4096	1.54	3.17	1.08
8192	3.83	4.38	1.58

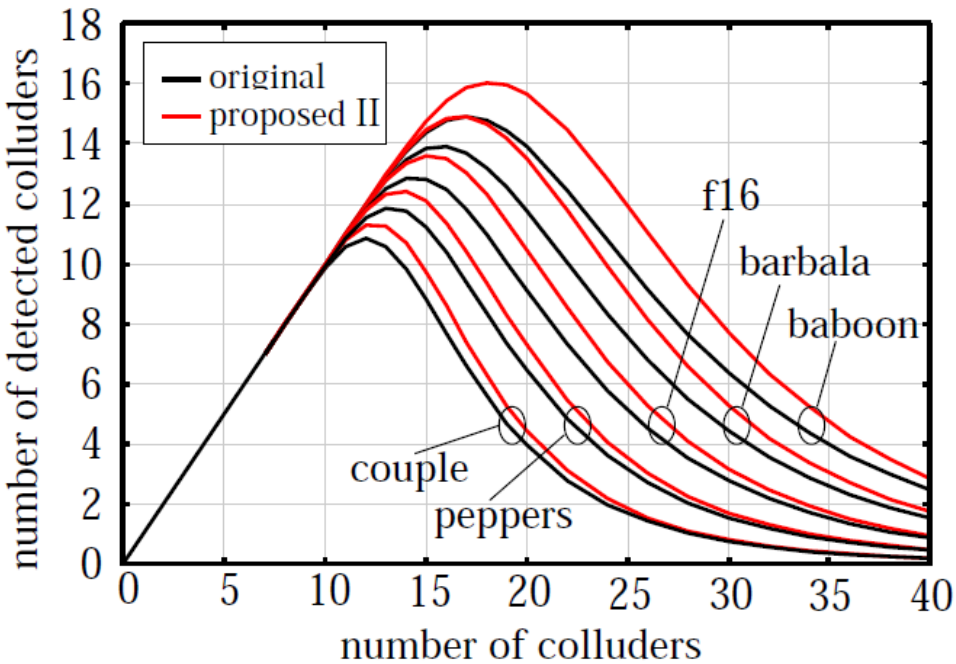
- False Negative Rate



Results of Other Images

Similar results are obtained for other images.

- True positive detection with $\ell = 8192$

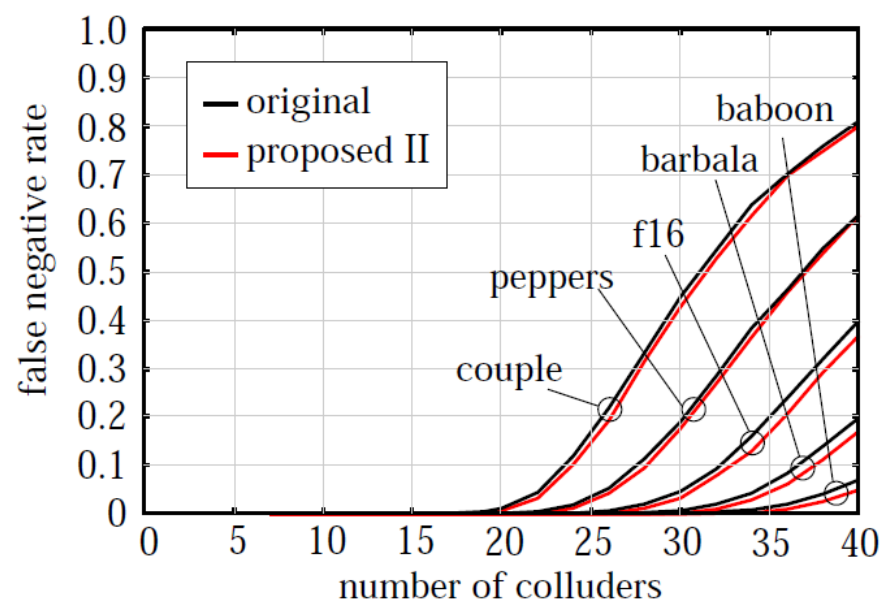


The noise caused by JPEG compression changes the performance of traceability.

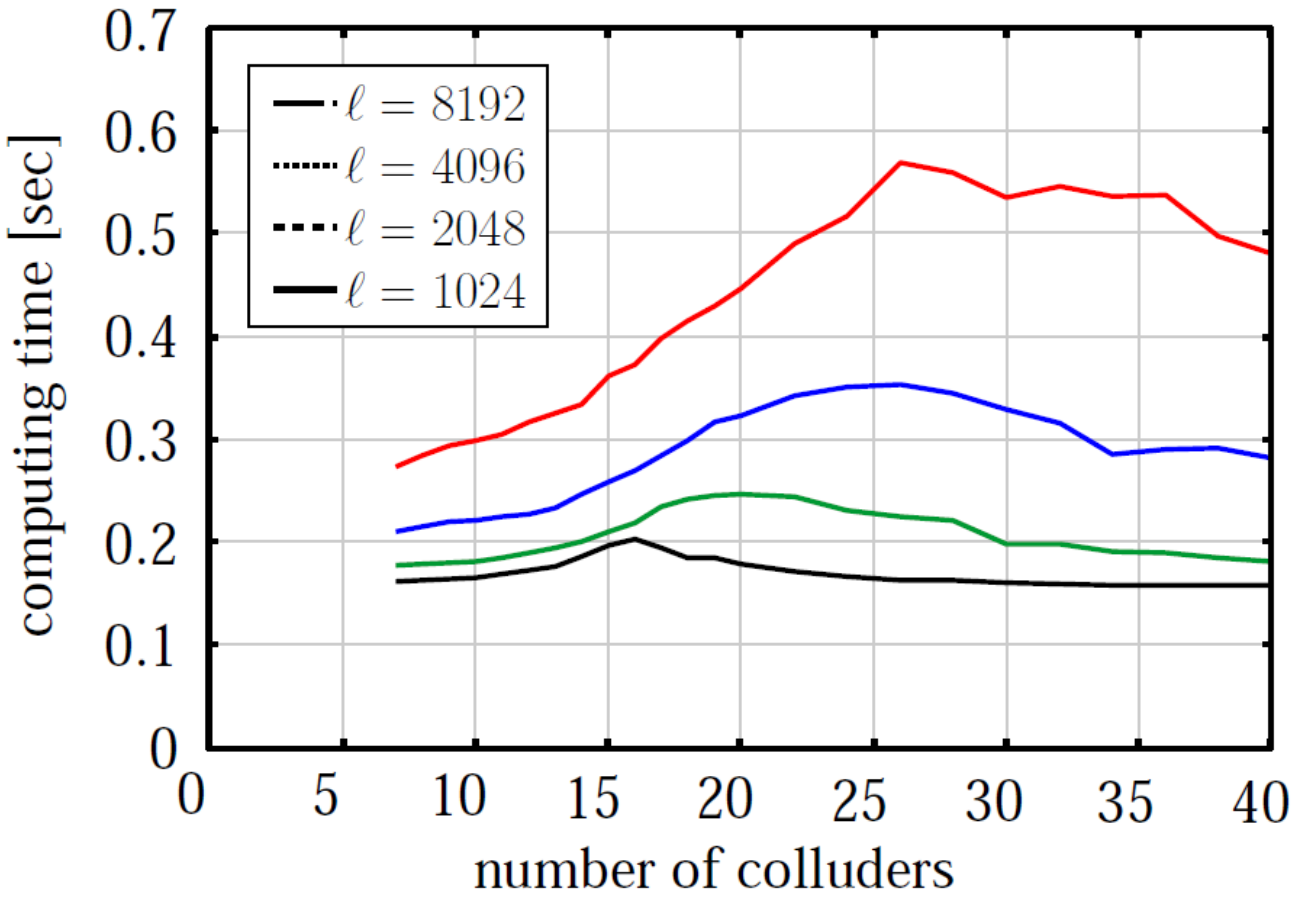
- False Positive Detection [$\times 10^{-4}$]

	original	proposed II
baboon	4.92	1.21
barbala	3.54	0.75
couple	3.13	1.21
f16	2.83	0.67
peppers	3.00	1.25

- False Negative Rate



Time Consumption



Hardware

CPU : Intel Core2Quad Q6700 (2.67GHz)
RAM : 8GB
OS : CentOS for X86-64 version 5.2

Conclusion

Spread Spectrum Fingerprinting based on CDMA Technique

- Assign a pair of spectrum components to each user
  - A combination of PN sequence and orthogonal transform
- Construct a hierarchical structure in the sequences
- Design a proper threshold for a given false-positive probability
- Improve the detector to detect more colluders and less innocent
  - remove operation + iterative detection + two kinds of thresholds

High robustness & low computational costs

Thank you for your attention

Any Question ?

Minoru Kuribayashi

E-mail: kminoru@kobe-u.ac.jp