Privacy-Preserving Multimedia Retrieval

Wenjun Lu, Avinash Varna, Ashwin Swaminathan, and Min Wu

A. JAMES CLARK
SCHOOL OF ENGINEERING

Media And Security Team

Protecting Online Private Data

- Web services are used to store personal emails, share private pictures, and manage financial accounts.
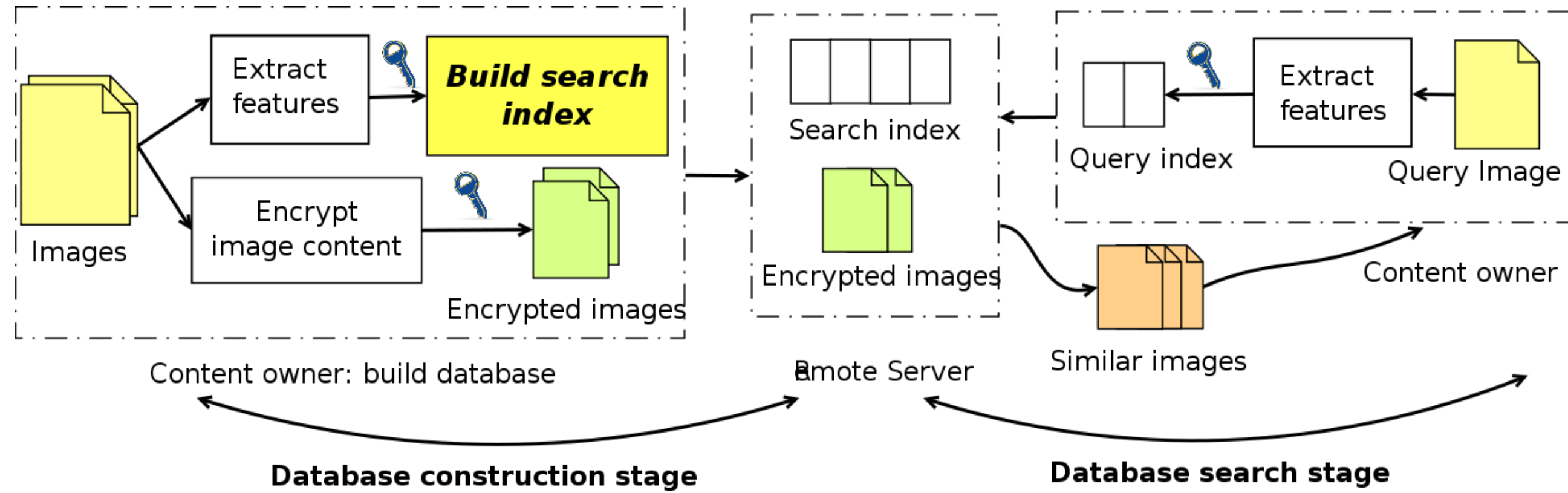- Personal info. management are increasingly carried out from the "cloud".

Gmail   Flickr

Dropbox   Mint.com

Is Our Online Data Really Safe?

Research goal:
  enable content-based search capability over encrypted multimedia databases for privacy protection

Framework for Secure Search

Extract features — Build search index — Search index — Query index — Extract features — Query Image

Images — Encrypt image content — Encrypted images — Encrypted images — Similar images — Content owner

Content owner: build database   Remote Server   Query Image

Database construction stage   **Database search stage**

User encrypts images and indexes; **Server** searches using encrypted indexes
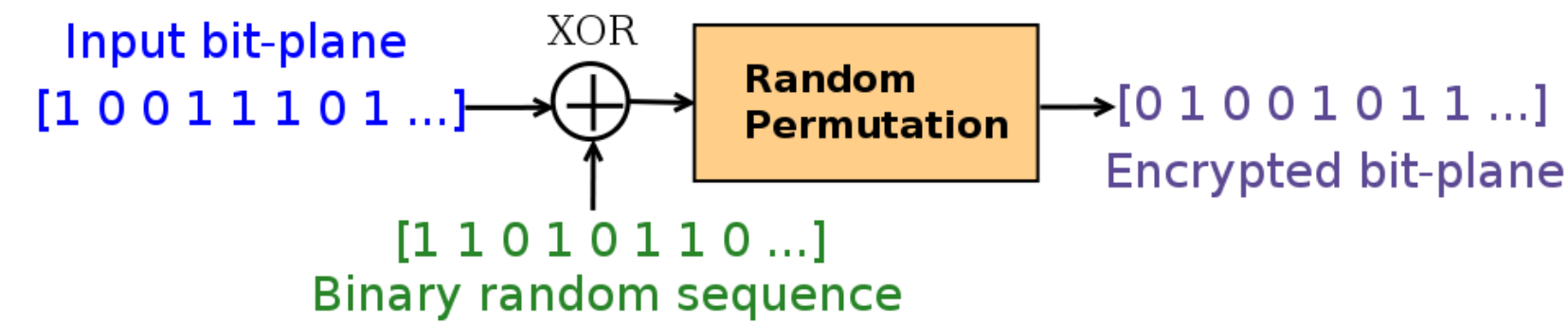
Visual Feature Protection

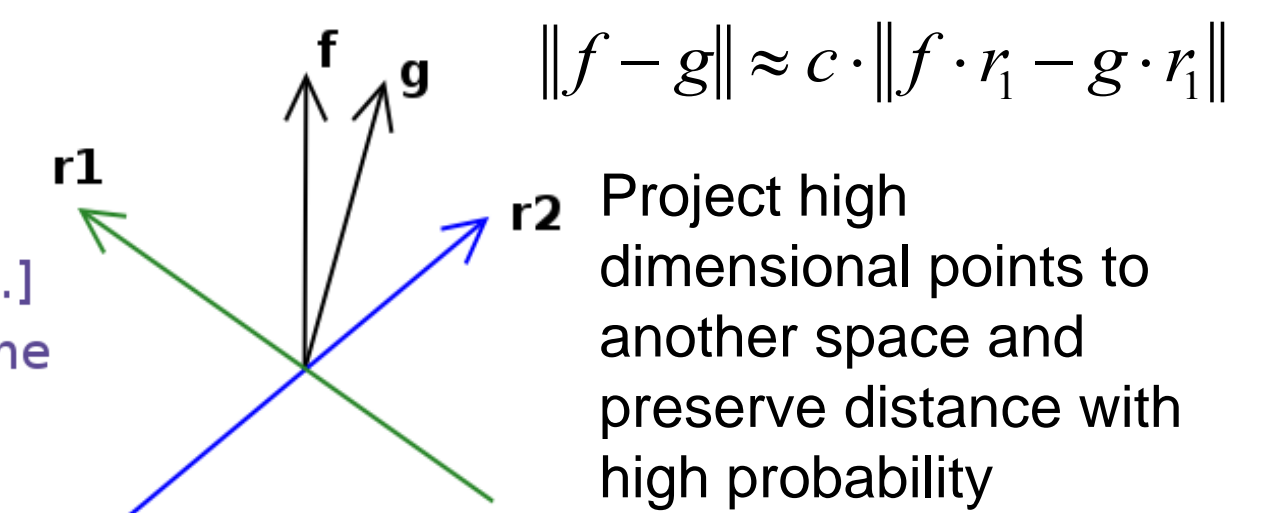Jointly apply techniques from:
- **Signal processing**: bit-plane representation, random projection, unary encoding
- **Cryptography**: pseudorandom permutation, xor pad, pseudorandom sequence

Bit-plane randomization:

Scramble bit-plane patterns of feature vectors while preserving their Hamming distance.
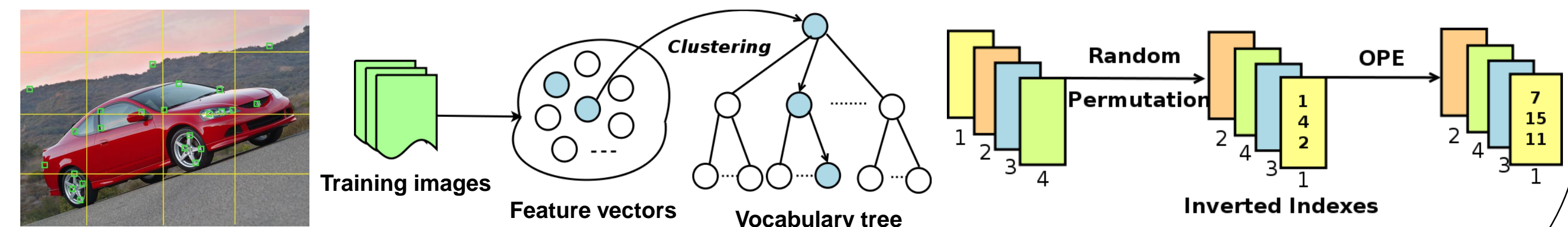
Input bit-plane
[1 0 0 1 1 1 0 1 …]   XOR   Random Permutation   [0 1 0 0 1 0 1 1 …]
Encrypted bit-plane

[1 1 0 1 0 1 1 0 …]
Binary random sequence

Random projection:

$$\|f - g\| \approx c \cdot \|f \cdot r_1 - g \cdot r_1\|$$

Project high dimensional points to another space and preserve distance with high probability
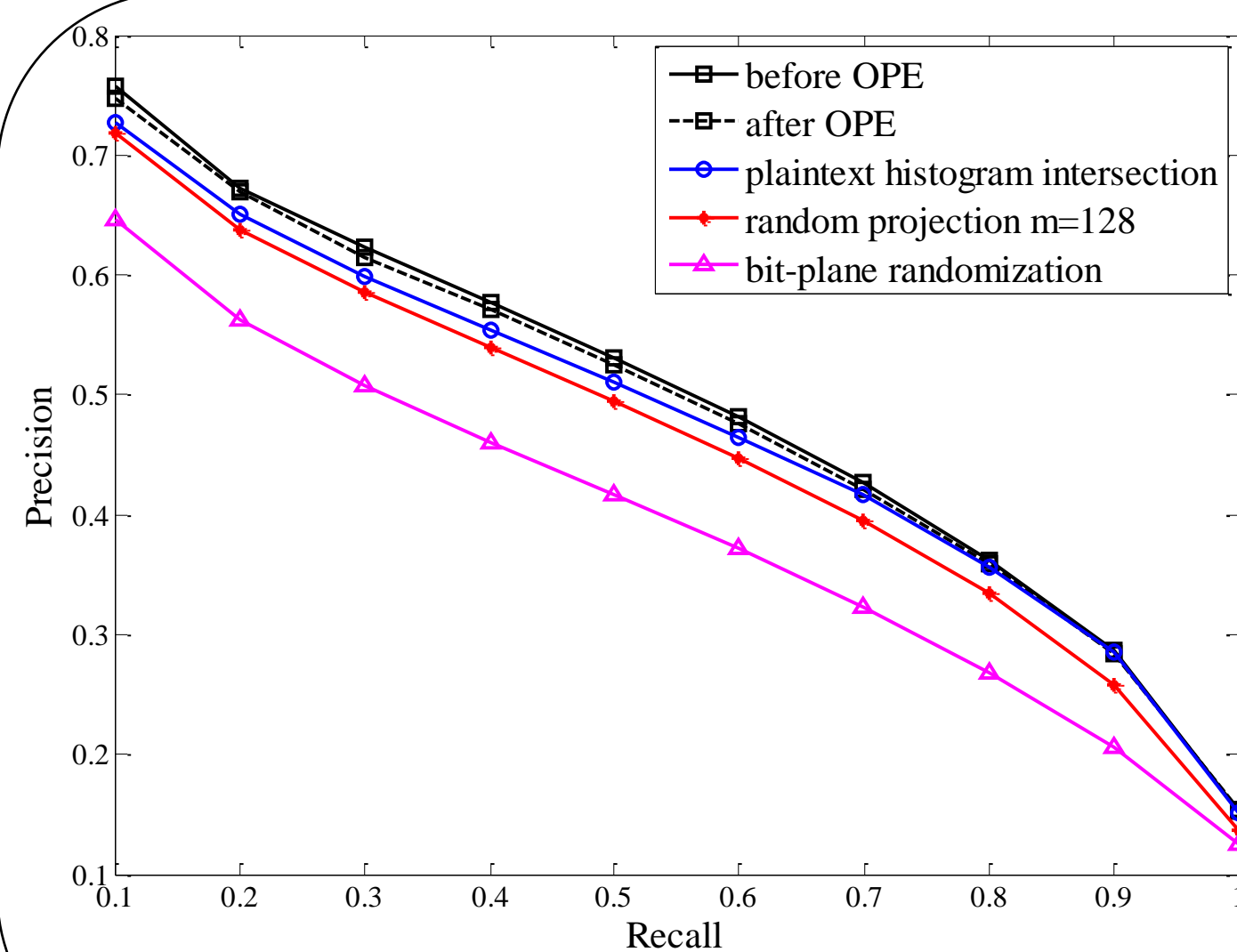
Securing Multimedia Search Indexes

Combine techniques from:
- **Information retrieval**: inverted index, min-hash, bag of words representation
- **Cryptography**: order preserving encryption, trapdoor function, pseudorandom permutation
- **Computer Vision**: feature vocabulary tree, visual-words representation of images

Training images   Feature vectors   Clustering   Vocabulary tree   Random Permutation   OPE   Inverted Indexes

Local features → Cluster and generate inverted index → Order preserving encryption
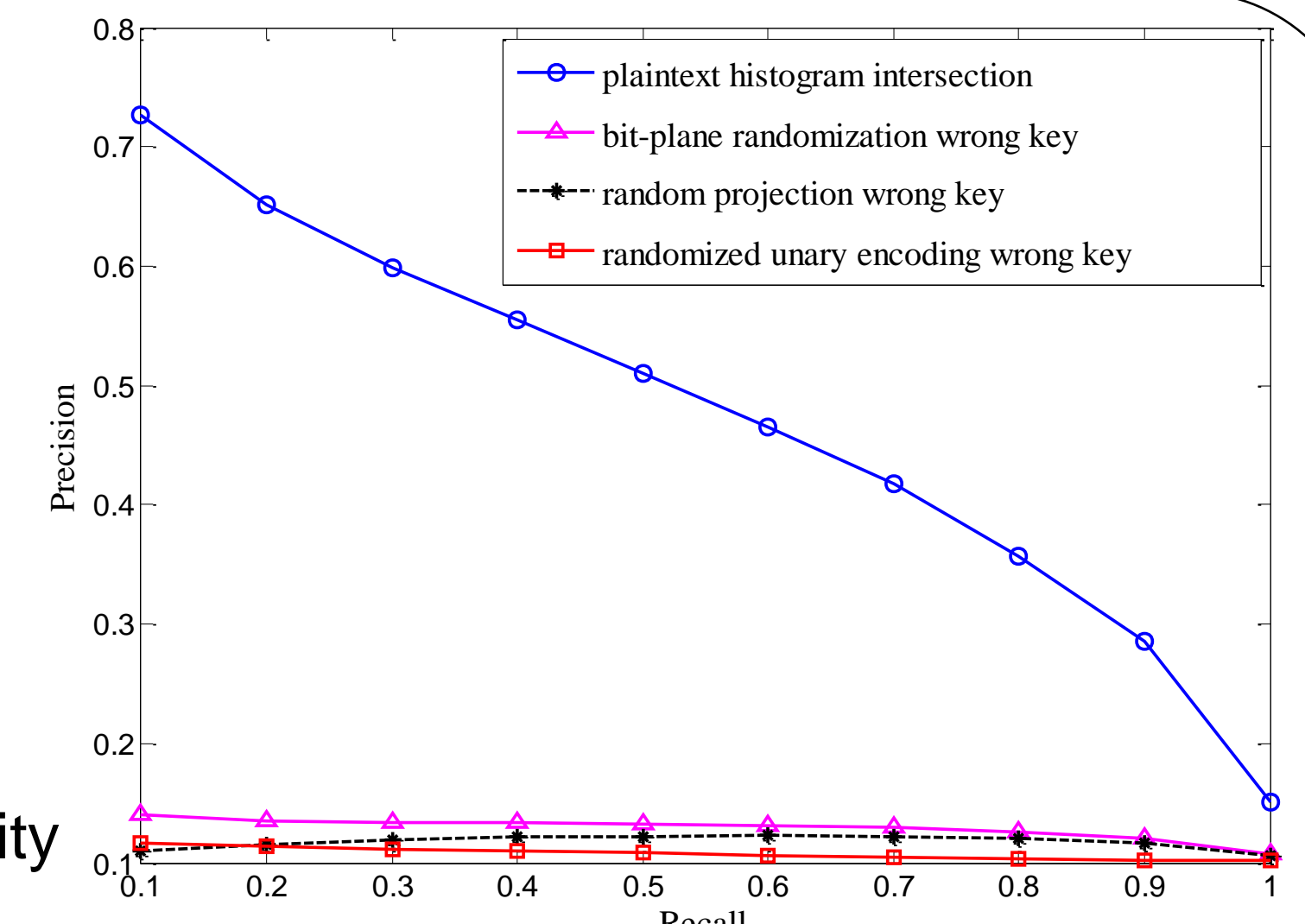
Retrieval and Security Performance

- Performed retrieval on a color image database using HSV color histograms
- Vocabulary tree with 1000 visual words: trained with 1000 images
- Multimedia data are encrypted by strong cryptography ciphers
- Feature/index protection schemes prevent information leak to adversaries
- Conducted security proofs and experimental validation

Summary

- Achieved content based multimedia retrieval over encrypted database
- Efficiently protects multimedia data privacy without affecting retrieval capability
- Provable security to protect privacy for secure online media management

before OPE; after OPE; plaintext histogram intersection; random projection m=128; bit-plane randomization

Precision / Recall

Retrieval Performance

plaintext histogram intersection; bit-plane randomization wrong key; random projection wrong key; randomized unary encoding wrong key

Precision / Recall

Wrong key retrieval