SECURITY ANALYSIS FOR PRIVACY PRESERVING SEARCH OF MULTIMEDIA

Wenjun Lu, Avinash L. Varna, and Min Wu
Department of Electrical and Computer Engineering,
University of Maryland, College Park

ABSTRACT

With the increasing popularity of digital multimedia such as images and videos and the advent of the cloud computing paradigm, a fast growing amount of private and sensitive multimedia data are being stored and managed over the network cloud. To provide enhanced security and privacy protection beyond traditional access control techniques, privacy preserving multimedia retrieval techniques have been proposed recently to allow content-based multimedia retrieval directly over encrypted databases and achieve accurate retrieval comparable to conventional retrieval schemes. In this paper, we introduce a security definition for the privacy preserving retrieval scenario and show that the recently proposed schemes are secure under the proposed security definition.

Index Terms— Content based multimedia retrieval, privacy preserving search, security analysis

1. INTRODUCTION

The advancement of information technology has been rapidly integrating the physical world where we live and the online world that we rely on for retrieving, sharing, and managing information. With convenient web services, such as social networking sites and the suite of Google applications emerging rapidly, more and more sensitive personal data are being stored and managed over the network cloud. The paradigm of cloud computing brings convenient information access but also raises concerns regarding the privacy of sensitive information stored online, which calls for new technologies that can provide increased security and privacy protection for online information management.

Traditional privacy protection for online personal data focuses on access control and secure data transmission, which ensure that the data can be securely transmitted to the server and no unauthorized people can access the data. However, once the data arrives at the server, the server decrypts the data and operates on plaintext in order to provide services to users, such as categorization, search, and data analysis. This makes the user’s private information vulnerable to untrustworthy service providers and malicious intruders. To address this problem, recent techniques for information retrieval over the encrypted domain aim at providing efficient and accurate search capability directly over encrypted documents without decrypting them first. Early works [1–3] in this area have brought search capabilities to encrypted text documents.

Due to the widespread use of digital cameras and media-oriented smartphones, multimedia constitute a significant part of today’s personal data collections. Securely storing and managing this large volume of multimedia data online is becoming a desirable option for convenient data access anywhere anytime. Recent work in [4, 5] extends the area of information retrieval over encrypted domain to allow content-based search over encrypted multimedia databases and therefore offers flexible approaches to manage private multimedia collections online. In [4], visual features extracted from image documents are encrypted in a distance-preserving fashion so that encrypted features can be directly compared for similarity evaluation. In [5], state-of-the-art search indexes for multimedia data are encrypted while their efficient search capabilities are preserved. It has been shown that secure multimedia retrieval schemes proposed in [4, 5] can achieve performance comparable to traditional unsecured retrieval schemes.

In this paper, we provide rigorous security analysis for privacy preserving multimedia retrieval schemes proposed in [4, 5]. By analyzing threat models of real-world applications, we propose a security definition formulated using the notion of indistinguishability and show that the privacy preserving retrieval schemes are secure under the proposed security definition. The concept of indistinguishability is broadly used in modern cryptanalysis for provable security in practical systems [6]. Our security analysis is carried out on two representative retrieval schemes and can be applied to other schemes in [4, 5] straightforwardly. The security analysis presented in this paper not only demonstrates the feasibility of using privacy preserving retrieval techniques in practical online multimedia management, but also provides valuable guideline on selecting different retrieval schemes for different application scenarios. In the remainder of the paper, we briefly review the two representative secure multimedia retrieval schemes in Section 2 and then provide security analysis in Section 3. Conclusions are drawn in Section 4.

Email: {wenjunlu, varna, minwu}@umd.edu
2. PRIVACY PRESERVING MULTIMEDIA RETRIEVAL TECHNIQUES

In modern multimedia search and retrieval, content similarity is typically evaluated using search indexes or visual features, such as color histograms and salient points [7]. In the framework of privacy preserving multimedia retrieval, images are encrypted using standard ciphers and visual features or search indexes extracted from the images will be protected by the encryption schemes proposed in [4, 5]. The remote server stores only the encrypted images and encrypted search indexes. During retrieval, the search indexes from the query image will be encrypted in a similar way to allow similarity comparison in the encrypted domain. Below we briefly review representative encryption schemes for visual features and search indexes, respectively. Readers are referred to [4, 5] for further details.

2.1. Searchable encryption of visual features

In content-based image retrieval, the distance between visual features has been widely used to evaluate the content similarity between two images [7]. For secure multimedia retrieval, the idea of searchable encryption was proposed in [4] to encrypt visual features in a distance-preserving fashion. Denoting image features as vectors in \( \mathbb{R}^n \), for a searchable encryption \( E(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^m \), given two vectors \( f \) and \( g \), \( d_E(f, g) \approx c \cdot d(f, g) \), where \( d_E(\cdot, \cdot) \) and \( d(\cdot, \cdot) \) are appropriate distance measures on ciphertext and plaintext spaces, respectively, and \( c \) is a constant scaling factor.

**Random projection** One implementation for searchable encryption is based on random projection, which linearly maps a vector into a lower dimensional space. Since close points in a high dimensional space are very likely to remain close after projection onto a lower dimensional space, random projection is commonly used to reduce the size of database entries and enable efficient search of approximate nearest neighbors over a large database.

During encryption, a secret key is used to generate an \( m \times n \) projection matrix \( R \) with independent Gaussian components. The projection operation is expressed as \( y = Rx \), where \( x \) is the input vector and \( y \) is the projected vector in the transformed subspace. Projections onto the same secret subspace will approximately preserve the \( L_2 \) distance, whereas the distance between projections onto two different subspaces will appear random. Security of random projection relies on the fact that it is computationally infeasible to determine the vectors used for the projection in a brute-force attack. To further improve security, the feature vector can also be first encoded into a binary string using unary encoding and then pseudorandomly permuted and XORed with a pseudorandom binary string as an additional encryption stage.

2.2. Secure search indexes for multimedia

Efficiency and scalability are critical for retrieval over large multimedia databases and are achieved using search indexes. An alternative technique [5] for privacy preserving multimedia retrieval is to encrypt state-of-the-art search indexes on the user side while retaining their efficient search capability. As indexes are typically more compact than visual features, secure indexes provide a different trade-off between user-side computational complexity and communication overhead from the searchable encryptions discussed above.

The search indexes considered in [5] are built upon the visual words representation [8]. As shown in Fig. 1, visual features are first extracted from the multimedia document, hierarchically clustered and assigned to a representative feature called “visual word”. The entire multimedia document is then represented as a set of keywords, analogous to the bag of words representation in the text retrieval literature.

**Secure inverted index** Given the visual words representations of a collection of multimedia documents, inverted indexes can be constructed for each visual word capturing its frequency of occurrence in each multimedia document. For privacy protection, the visual word IDs are encrypted using pseudorandom permutations, and the word frequency values are protected using order preserving encryption (OPE). Order preserving encryption preserves the relative order of two values after encryption, i.e. \( E(x) < E(y) \) if \( x < y \). The use of OPE can prevent adversarial attacks that exploit statistical information of word frequency values. The similarity of two documents is then compared directly in the encrypted domain by using the Jaccard similarity metric:

\[
\text{Sim}(I_1, I_2) = \frac{\sum \min(W_1(i), W_2(i))}{\sum \max(W_1(i), W_2(i))},
\]

where \( W_1(i) \) and \( W_2(i) \) are the encrypted word frequency values of the \( i \)th visual word that appears in documents \( I_1 \) and \( I_2 \), respectively. Use of inverted indexes improves the retrieval efficiency, because only documents that share visual words with the query need to be examined.

3. SECURITY ANALYSIS

The security analysis presented in this section is under the Ciphertext Only Attack model (COA), which assumes that the adversary has access only to the ciphertext, i.e., the encrypted images and indexes. Furthermore, we assume that the adversary is semi-honest, i.e., the adversary will follow the execution requirement of the protocol but may use what they see during the execution to compute more than what they need to know. Semi-honest model is a reasonable assumption for adversaries such as third-party service providers. Security anal-
ysis is carried out on two representative schemes from [4, 5], and the analysis is applicable and extensible to other retrieval schemes in a straightforward manner.

For secure online multimedia management, the encrypted images and their encrypted features or indexes are stored on the remote server and thus accessible by system administrators and any potential intruders. In the COA model, a secure retrieval scheme should be able to protect the following items from the adversary: (1) plaintext content of the encrypted database images \( \{ I_i \} \) and the query image \( Q \); (2) the secret key \( K \) used in the encryption; and (3) any function of the images \( f(I_i) \) and \( f(Q) \), which can be plaintext features of the images. We provide a security definition formulated using the concept of content indistinguishability, which is adapted from the indistinguishability definition [9] widely used in cryptography analysis. We show that most of the schemes proposed in [4, 5] are provably secure under this definition.

**Content indistinguishability:** Consider the following security experiment: an adversary chooses two images \( I_0 \) and \( I_1 \). The content owner chooses a secret key \( K_0 \) and an encryption scheme \( \mathcal{E} \) to obtain the encrypted indexes \( \mathcal{E}(I_0, K_0) \) and \( \mathcal{E}(I_1, K_0) \). The content owner then randomly chooses a value \( b \) from \( \{0, 1\} \) with equal probability and sends the encrypted index \( \mathcal{E}(I_b, K_0) \) to the adversary. Using any probabilistic polynomial time algorithm, the adversary outputs a number \( b' \) as an estimate of \( b \). The retrieval scheme satisfies **content indistinguishability** if

\[
|\Pr[b' = b] - \frac{1}{2}| \leq \text{negl}(n),
\]

for every choice of \( I_0, I_1 \) by the adversary, where \( \text{negl}(n) \) is a negligible function of the security parameter \( n \). The probability is taken over all possible random choices by the adversary and in the experiment, i.e. the secret key \( K_0 \) and the value \( b \).

Here, a function \( f(\cdot) \) is negligible if for every polynomial \( p(\cdot) \), there exists an \( N \) such that for all integers \( n > N \) it holds that \( f(n) < 1/p(n) \), implying that \( f \) decays faster than any inverse polynomial function. The parameter \( n \) typically determines the length of the secret key and the security strength of the encryption scheme.

The content indistinguishability definition essentially states that a computationally bounded adversary cannot distinguish two encrypted features or indexes even if he/she has knowledge of the plaintext features or indexes. In the COA model, the attacker is assumed to have knowledge only of the ciphertext. Therefore, if a feature/index encryption scheme satisfies the above definition, the adversary will not be able to distinguish any two encrypted images in terms of their content, and confidentiality is preserved under the COA model.

A general approach to prove security is to reduce the security of the entire scheme to the security of basic cryptographic building blocks, such as pseudorandom functions and pseudorandom permutations [6]. As these cryptographic primitives are considered hard to break by any probabilistic polynomial time algorithm, we can prove the security of a retrieval scheme by showing that breaking the scheme is equivalent to breaking the cryptographic primitives. Next, we outline the security proof for the feature protection scheme based on random projection under the security definition of content indistinguishability. The proof for other schemes can be carried out in a similar way.

**Proof for random projection.** Given two images \( I_0 \) and \( I_1 \), we denote their feature vectors as \( f_b, b \in \{0, 1\} \), which are normalized so that \( \|f_b\|_2 = c \) for some constant \( c \). For the random projection based scheme, the content owner chooses a secret key \( K_0 \) and generates a pseudorandom matrix \( R \) whose elements are independent Gaussian variables from \( \mathcal{N}(0, 1) \). Encryption by random projection is denoted as

\[
f_{ib} \equiv \mathcal{E}(f_b, K_0) = R \cdot f_b = (r_1 \cdot f_{ib}, r_2 \cdot f_{ib}, \ldots, r_m \cdot f_{ib}),
\]

where \( r_i \) is the \( i \)-th row of the matrix \( R \).

Assuming that the matrix \( R \) is truly random, with each element drawn independently from a standard Gaussian distribution, the encrypted features \( f_{ib}, f_{1b} \) are vectors whose components are independent Gaussian variables \( \mathcal{N}(0, c^2) \), as \( f_{ib}, f_{1b} \) are independent Gaussian \( \mathcal{N}(0, c^2) \) for any \( i \in \{1, \ldots, m\} \) and \( b \in \{0, 1\} \). The conditional probability of the encrypted feature given the plaintext feature only depends on the value of \( c \). Since all plaintext features are normalized to have the same value of \( c \), we have \( \Pr[f_{ib} | f_{ib}] = \Pr[f_{ib} | f_{1b}] \), for any \( f_{ib} \), which satisfies Shannon’s definition of perfect secrecy. A cryptosystem satisfies perfect secrecy if the posterior probability of the ciphertext given the plaintext is exactly the same as the prior probability of the ciphertext, for all ciphertexts and plaintexts. Therefore, the probability that any probabilistic polynomial time algorithm can distinguish \( f_0 \) and \( f_1 \) is exactly 1/2.

After replacing the truly random matrix \( R \) with a pseudorandom matrix, the probability that \( f_0 \) and \( f_1 \) are distinguished by a probabilistic polynomial time attacker is denoted as \( \Pr(b' = b) = \frac{1}{2} + \epsilon(n) \). If the function \( \epsilon(n) \) is negligibly small, it will imply that there exists a polynomial time algorithm to distinguish a truly random sequence from a pseudorandom sequence, which contradicts the definition of cryptographically secure pseudorandom sequences [6]. Thus \( \epsilon(n) \) must be negligibly small implying that feature protection schemes based on random projection satisfy the content indistinguishability.

**Security analysis of inverted index based scheme.** The above security analysis can be applied to most of the secure retrieval schemes proposed in [4, 5] and the only scheme that does not satisfy the definition of content indistinguishability is the inverted index based scheme. In this part, we provide security analysis to this secure retrieval scheme and discuss its implications on practical application of such scheme. The inverted index based scheme encrypts the inverted index by pseudorandom permutation and order preserving encryption. Although the order statistics are scrambled by pseudorandom permutation, the variance of the index is preserved by order
preserving encryption. Therefore, in the security experiment of content indistinguishability, an adversary can choose two images whose bag-of-words representations have different variance and such variance will be preserved after encryption, allowing the adversary to distinguish the two images. In the COA model where the adversary has no knowledge about the plaintext images, the variance of the encrypted features alone provide very limited information about the image content to the adversary.

The use of order preserving encryption (OPE) reveals some information about the plaintext in terms of its variance and makes the secure retrieval scheme based on OPE not secure under the content indistinguishability definition. A proper security definition for OPE with clear practical implications would be desirable but is a challenging task. Only recently, Boldyreva et al. [10] attempted a formal security definition for OPE and provided an OPE construction that can achieve the best possible security. It is still not clear what kind of information will be leaked by such a secure OPE construction in a practical scenario. Despite the security limitations, order preserving encryption can enable efficient indexing and range query. Such a trade-off is acknowledged by many other work which use OPE for range query in the encrypted domain, which would otherwise be much more difficult. We refer readers to [10] for detailed security analysis of OPE and recommend that OPE based scheme should be used with caution in practice.

The security of privacy-preserving retrieval schemes under the COA model has been experimentally verified in [4, 5]. The experiment is carried out over a subset of Corel database which contains 1000 images equally divided into 10 categories. Query indexes encrypted using both correct and wrong keys are used to search the database. Search performance is shown in Fig. 2. The precision value around 0.1 for the wrong key retrieval verifies that query indexes encrypted using a wrong key are equally likely to be closest to any encrypted index in the database and such retrieval is equivalent to picking images randomly from the database. The security of the retrieval schemes under the Known-Plaintext Attack model is also briefly discussed in [4, 5], where different schemes exhibit different trade-offs between computational complexity and security in KPA model. It should be noted that none of the schemes can be secure under the more stringent Chosen-Plaintext Attack model, because then the adversary can choose any plaintext image as query to learn the content of the entire image database.

4. CONCLUSIONS

The recently proposed techniques of privacy preserving search for multimedia achieve content-based retrieval directly over encrypted databases and offer important tools for securely managing private multimedia information online. In this paper, we provide rigorous security analysis for privacy-preserving multimedia retrieval techniques. The security definition is proposed for the ciphertext only attack model and proofs for two representative retrieval schemes are provided. The security analysis can be carried out to other schemes in a straightforward way and offers the necessary support for their applications in secure online multimedia management.

5. REFERENCES