Security analysis of robust perceptual hashing

Oleksiy Koval, Sviatoslav Voloshynovskiy, Fokko Beekhof and Thierry Pun*

ABSTRACT

In this paper we considered the problem of security analysis of robust perceptual hashing in authentication application. The main goal of our analysis was to estimate the amount of trial efforts of the attacker, who is acting within the Kerckhoffs security principle, to reveal a secret key. For this purpose, we proposed to use Shannon equivocation that provides an estimate of complexity of the key search performed based on all available prior information and presented its application to security evaluation of particular robust perceptual hashing algorithms.

1. INTRODUCTION

A problem of prevention of malicious illegal actions like violation of authors rights, unauthorized prohibited usage, unauthorized duplication and redistribution of digital media, high fidelity counterfeiting of digital and analog content as well as goods and products has become a central one in modern society nowadays. The reason is due to the high level of development of modern copying and reproducing equipment as well as communications networks. Existence of such a problem justifies an urgent demand for computationally efficient methods and algorithms for reliable document, product and person *identification* able to withstand various attacking strategies and simultaneously being privacy preserving as well as *authentication* methods able to establish the originality of the media file content. Another potential application of these methods is accurate information retrieval in large databases under accuracy and complexity constraints.[1]

A classical solution to the latter problem is based on the use of cryptographic hash functions[2] that being acceptable for security are not entirely free from certain drawbacks. One of the most evident disadvantages that is related to the cryptographic hash application for protection of multimedia files concerns randomized hash generation for the data that differ at least by a single bit. Thus, such a hash output will be generated from the inputs represented by perceptually or semantically identical content stored in different formats. Besides that, cryptographic hash functions are mostly developed for authenticating of data in binary format that is not usually used to represent multimedia files.

In order to overcome the drawbacks and weaknesses of a classical approach, robust perceptual hashing has been recently proposed[3] and have constituted the core of a challenging and dynamically developing research area. The conflicting requirements one should address while designing a robust perceptual hash include:

- **robustness**, i.e., the ability of the hash to produce asymptotically equivalent output based on inputs that differ by a certain distortion level introduced due to some non-malicious manipulations with multimedia data as some signal processing and/or desynchronization transformations;
- **security**, i.e., the ability of the attacker to learn the keyed hash function or the key k without the knowledge of this key based on the observed data y^N and structure of hash codebook;
- **universality**, i.e., optimal or asymptotically optimal performance of the hash in the case of lack of prior knowledge about the statistics of input source distribution and channel that is related to the machine learning framework and universal hypothesis testing.

*O. Koval, S. Voloshynovskiy, F. Beekhof and T. Pun are with CUI-University of Geneva, Stochastic Information Processing Group, Battelle Batiment A, 7 route de Drize, 1227 Carouge, Switzerland. The contact author is S. Voloshynovskiy (email: svolos@cui.unige.ch). http://sip.unige.ch

Thus, a *robust perceptual hash* can be defined as a one-way function, which takes multimedia objects as inputs, and generates sufficiently short binary strings approximately invariant under perceptual-quality-preserving modifications.

Robustness/invariance of geometric hashing was extensively studied in computer vision (the interested reader is pointed to[4] and the references therein). Most of the existing results obtained in the domain of perceptual hashing in data-hiding community[5–7] are dedicated to the investigation of resistance of practical hashing algorithms to various attacks. An attempt to introduce an information-theoretic framework for design and analysis of robust perceptual hashing as well as to consider the problem of its universality was recently performed[8, 9] where the problem of robust hashing is considered as a multiple hypothesis testing and it is demonstrated that under specific constraints on the cardinality of the set of parameters of geometrical desyncronization transform and host data statistics, generalized likelihood ratio test satisfies the conditions of asymptotic universality.

Measurement of the security of robust perceptual hashing is usually performed within the Kerckhoffs principle[10] and is related with information that can be obtained from the analysis of hashing algorithm design particularities as well as using all the available prior information (like an original image and/or a hash vector) about the secret key used to produce the signature.[11] The motivation of such an analysis is related to the attacker's capabilities of acting with the power of a legitimate user in the case he/she has an exact knowledge of the secret key. The difficulty of revealing such an information can be measured by diffusion/confusion capabilities of robust perceptual hash.[12] Firstly introduced by Shannon,[13] confusion and diffusion are used to evaluate the security of cryptographic systems. The measure of security in this case is provided by the complexity of relation between the secret key and the hash value (confusion) and between the plain text and the hash value (diffusion). By other words, the output of the hashing algorithm should be totally unpredictable in the case a single bit modification of the secret key (input data) is performed. The performed tests[12] of some practical robust perceptual hashing algorithms allowed to conclude about low diffusion sensitivity of these algorithms that can serve as a prohibitive argument of their application for multimedia data authentication.

Another attempt of security analysis of robust perceptual hashing algorithms based on information-theoretic argument was performed by Swaminathan *et al.*[14] The authors proposed to measure the efforts of the adversary to reveal the secret key in terms of the differential entropy of the hash. They performed investigation of security performance of certain practical robust perceptual hashing algorithms as well as initiate the analysis of robustness/security trade-off. However, one should admit that there is no direct relation between the secret key and hashing algorithm output randomness. The key can be seen rather as a source of common randomness shared by the legitimate users and similarly to data-hiding might be distributed uniformly and independently of the statistics of robust perceptual hashing output that was suppose by Swaminathan *et al.*[14] to be Gaussian.

In order to incorporate all available prior information in set-ups where multiple copies of plain text/hash can be observed by the opponent, Swaminathan *et al.* extended the analysis performed in[14] and suggested to use as a security measure of robust perceptual hashing unicity distance.[15] This concept, firstly introduced by Shannon for a two-letter substitution cipher[13] is referred to a number of encrypted/plain text pairs produced with the same secret key that allows a unique determination of this secret. The performed analysis revealed the requirement to increase a hash vector length in order to provide the enhanced security level.

In this paper we address the issue of estimation of the attacker efforts while revealing the secret key within the Kerckhoffs security framework[10] when only a single pair of original plain text/robust perceptual hash is available.

The remaining part of the paper has the following structure. The problem of data identification and authentication based on robust perceptual hashing is formulated in Section 2. Security analysis of robust perceptual hashing is performed in Section 3. Finally, Section 4 concludes the paper and draws some future perspectives.

Notations We use capital letters to denote scalar random variables X and corresponding small letters x to denote their realizations. The superscript N is used to designate length-N vectors $x^N = [x[1], x[2], ..., x[N]]$ with k^{th} element $x[k]$. An i^{th} block of a vector x^N of dimensionality P is denoted $x_i^P = [x_i[1], x_i[2], ..., x_i[P]]$. We use $X \sim p_X(x)$ or simply $X \sim p(x)$ to indicate that a random variable X is distributed according to $p_X(x)$. The variance of X is denoted by σ_X^2 denotes the variance of X. We use Σ to denote a covariance matrix. Calligraphic fonts $\mathcal{X}$ denote sets $X \in \mathcal{X}$ and $|\mathcal{X}|$ denotes the cardinality of set $\mathcal{X}$. Superscript T stays for matrix transposition.

2. PROBLEM FORMULATION

Definition. Robust perceptual hash (Figure 1) consists of a hash set $\mathcal{M} = \{1, 2, ..., |\mathcal{M}|\}$ or hash index set $\{0, 1\}^L$ and a hash function:

$$\psi_K : \mathcal{X}^N \times \mathcal{K} \to \mathcal{M}, \tag{1}$$

$$\psi_K : \mathcal{X}^N \times \mathcal{K} \to \{0, 1\}^L, \tag{2}$$

where $L = \log_2 |\mathcal{M}|$ and $\mathcal{X}^N$ stays for the set of input data assumed to be generated by a discrete memoryless source (DMC) and $\mathcal{K} = \{1, 2, ..., |\mathcal{K}|\}$ denotes a set of secret keys. The mapping (1) should satisfy the following requirements.

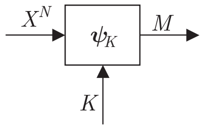

Figure 1. Robust perceptual hashing as a key-dependent mapping.

Randomization:

$$Pr[\psi_K(X^N) = \alpha] \approx \frac{1}{2^L}, \alpha \in \{0, 1\}^{L\dagger}.$$

Pairwise independence for perceptually different inputs:

$$Pr[\psi_K(X^N) = \alpha | \psi_K(Y^N) = \beta] \approx Pr[\psi_K(X^N) = \alpha], \alpha, \beta \in \{0, 1\}^L.$$

Invariance under perceptual similarity:

$$Pr[\psi_K(X^N) = \psi_K(\hat{X^N})] \approx 1,$$

where $X^N, \hat{X^N}$ are perceptually similar data vectors with respect to a certain distance measure: $d^N(X^N, \hat{X^N}) \leq D^{adm}$.

Hash generation according to (1) can be considered as output of a deterministic identification code since the source of randomness, the secret key $K = k$, is symmetrically known to both encoder and decoder.[16] On the other hand, as it was multiply reported in the literature, one can consider a problem of construction of the hash codebook as a similar to a rate distortion, where one is targeting representing X^N using a set of i.i.d. distortion typical sequences $\hat{\mathcal{X}}^N$ of cardinality 2^{NR} such that $d^N(x^N, \hat{x}(i)^N) \leq D^{adm}$ and to declare the corresponding codeword index i as a hash. At the verification stage, given the output of the channel $p(y^N|x^N)$ that can be considered as consisting of a constrained desynchronization part and discrete memoryless part,[8, 9] it is necessary to solve a hypothesis testing problem in order to identify y^N. A type of a performed test and a particular codebook construction are defined by a particular application scenario.

Codebook construction: In the scope of this paper we assume a multiuser scenario, where $|\mathcal{K}|$ distinctive users exist identified by a particular secret key $k \in \mathcal{K}$. The key K is used to define a user hash codebook $\hat{\mathcal{X}}^N(K)$. The generation of the codebook is performed by producing $|\mathcal{K}|2^{NR}$ distinct distortion typical sequences independently at random from a distribution $p_{\hat{X}^N|X^N}(\hat{x}^N|x^N)$ and randomly distributing them in $|\mathcal{K}|$ groups by 2^{NR} sequences in each one.

Hash-aided identification vs. hash-aided authentication: At this moment we would like to distinguish two different applications that are of particular interest for this paper. In the first scenario (identification) one

$^\dagger\alpha$ is a scalar value given in the binary representation.

assumes as the main task to find the answer to the following question: "What is the index that represents a given channel output in the codebook defined by a secret key $K = k$?"(Figure 2,a). Contrarily, in the case of data authentication (Figure 2,b), the main goal is to certify if the index produced by the hash code (1) coincides with a reference one. Thus, in the former formulation the answer is provided by means of a multiple hypothesis testing while in the latter case a binary hypothesis testing η_B should be performed.

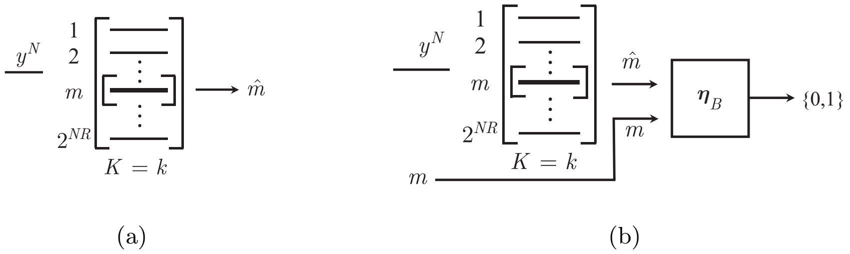

Figure 2. Robust perceptual hashing-based identification (a) vs authentication (b) scenario.

In the following Sections of this paper we will mostly concentrate on the security analysis of robust perceptual hashing-based authentication.

3. PERCEPTUAL HASH SECURITY ANALYSIS: EQUIVOCATION

As it was already mentioned in the introductory part of this paper, in order to comply with Kerckhoffs security principle, we assume a complete transparency of the authentication protocol, i.e., all user codebooks as well as explicit technicalities of the exploited robust perceptual hashing algorithm are assumed to be available to the attacker. The only unknown parameter that defines the randomness of the scheme is represented by a secret key $K = k$ that is known to come from a given set $\mathcal{K} = \{1, 2, ..., |\mathcal{K}|\}$. Therefore, the ultimate goal of our analysis is, given all the available prior knowledge, to estimate the complexity to reveal a particular secret $K = k$ that will allow the eavesdropper to act with a complete power of a legitimate user and to produce fakes that will be interpreted by the system as authentic originals.

Following similar strategy that was already used in data-hiding community to measure the secrecy of watermarking schemes,[17,18] we propose to formulate the problem in terms of equivocation. Firstly introduced by Shannon,[13] this quantity in cryptography is defined as the ambiguity about the secret that remains after observing the cyphertexts. For the case of security analysis of robust perceptual hashing, equivocation can be redefined as follows:

$$H(K|\psi_K(X^N)) = H(K) - I(K; \psi_K(X^N)), \tag{3}$$

where $H(K)$ and $H(K|\psi_K(X^N))$ are entropy of the secret key K and conditional entropy of K given a particular media file X^N and its hash $\psi_K(X^N)$, respectively, and $I(K; \psi_K(X^N))$ stays for mutual information between K and $\psi_K(X^N)$.[19] All information theoretic quantities are computed under assumption of public availability of X^N.

3.1. Security evaluation of practical robust perceptual hashing algorithms

In this subsection of the paper we present an equivocation-based security analysis of certain robust perceptual hashing algorithms.

3.1.1. Robust perceptual hashing based on random projections

A block diagram of the first method[20] inspired by an early development of J.Fridrich[3] is presented in Figure 3.

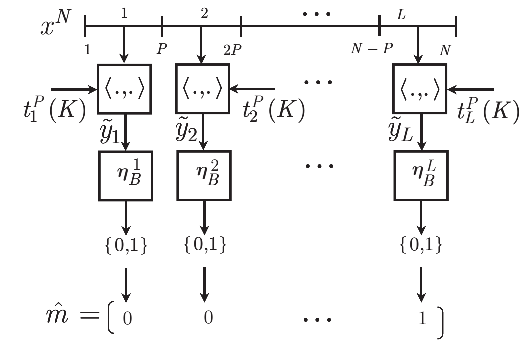

Figure 3. A practical robust perceptual hashing algorithm based on block random projections.

Here, it is assumed that the realization of the data to be authenticated x^N is block wise randomly projected on a set of random patterns $t_i^P(K), i \in \{1, 2, ..., L\}$ generated from the key K. The projection output $\tilde{y}_i, i \in \{1, 2, ..., L\}$, is acting as an input to a set of binary hypothesis testing blocks $\eta_B^i, i \in \{1, 2, ..., L\}$:

$$\tilde{y}^L = T^{LP}(K)x^N, \tag{4}$$

where $T^{LP}(K)$ is a random projection operator that has a block structure:

$$T^{PL} = \begin{pmatrix} T_1^{P\times 1}(K) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & T_2^{P\times 1}(K) & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & T_L^{P\times 1}(K) \end{pmatrix}, \tag{5}$$

where $\mathbf{0}$ is a $P \times 1$ vector of zeros.

It is important to note for the further analysis that there is a unique correspondence between the secret key $K = k$ and a random projection operator $T^{LP}(K)$. Thus, we can redefine the mutual information term in (3) as follows:

$$I(K; \psi_K(X^N)) = I(T^{LP}(K); \psi_K(X^N))). \tag{6}$$

The main goal of the ongoing security analysis is to estimate how much information can be obtained about $T_{i,j}(K)$, an j^{th} random component of a particular $T_i^{P\times 1}(K) = [T_{i,1}(K), T_{i,2}(K), ..., T_{i,j}(K), ..., T_{i,P}(K)]$, from observing a hash value $\{0, 1\}$. It will be performed under the following assumptions. It is supposed in the scope of this paper that the keys are uniformly distributed over the key-set $\mathcal{K}$; thus, the entropy of K is given by $\log_2 |\mathcal{K}|$. Moreover, we assume that the projection operators $T_i^{P\times 1}(K)$ acting on the block of dimensionality P level are distributed i.i.d. accordingly to a multivariate zero mean Gaussian distribution with a covariance matrix $\sigma_T^2 \mathbb{I}^{P\times P}$. Hash input vectors on the block level $x_i^P = [x_i[1], x_i[2], ..., x_i[P]]$ are considered to be composed of i.i.d. realizations of a uniform distribution supported on the interval $[0, 1]$.

Then, an equivalent protocol of a lossy transmission of $T_{i,j}(K)$ can be presented as it is shown in Figure 4. According to this protocol, the binary channel output is obtained by a comparison with a decision threshold of $(x_i[j]T_{i,j}(K) + Z_{i,j})$, where $Z_{i,j} = \sum_{l=1, l\neq j}^{P} x_i[l]T_{i,l}(K)$ is a Gaussian random variable with a zero mean and variance $\sigma_{\tilde{Y}_i}^2 = \sigma_T^2 \sum_{j=1}^{P} (x_i[j])^2$, where we assume $\sigma_{\tilde{Y}_i}^2 = \sigma_{\tilde{Y}}^2$ for all blocks. Such a statistical behavior of $Z_{i,j}$ is justified using a based on the Central Limit Theorem[19] argument similar to one exploited by Swaminathan *et al.*[14] It is important mentioning that the output of a binary test can be considered as an index produced by a one-bit minimum distortion scalar quantization of $(x_i[j]T_{i,j}(K) + Z_{i,j})$. According to the hash randomization property and assumed statistical model of the binary channel input, one can explicitly assume that the threshold value is equal to zero.

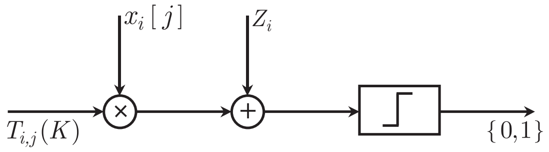

Figure 4. An equivalent protocol for a lossy transmission of a random projection component $T_{i,j}(K)$.

Therefore, one can write the mutual information between the random projector component $T_{i,j}(K)$ and $\hat{M}_i$, where $\hat{M}_i = \psi^i_K(X^N)$ as follows:

$$I(T_{i,j}(K); \hat{M}_i) = H(\hat{M}_i) - H(\hat{M}_i|T_{i,j}(K)) = H(\hat{M}_i) - H(Q(x_i[j]T_{i,j}(K) + Z_{i,j})|T_{i,j}(K))$$
$$= 1 - H(Q(x_i[j]T_{i,j}(K) + Z_{i,j}) - x_i[j]T_{i,j}(K)|T_{i,j}(K)), \quad (7)$$

where $H(\hat{M}_i) = 1$ due to the hash randomization, $Q(\cdot)$ is a scalar minimum distortion one-bit quantizer. It is interesting to note that $(Q(x_i[J]T_{i,j}(K) + Z_{i,j}) - x_i[j]T_{i,j}(K))$ coincides with the definition of the subtractive dither quantization.[21] It is well known that such a quantizer possesses certain useful properties like providing pair wise independence between the quantization input, quantization noise and dither. However, these properties are achieved in the case of specific limitations that constrain characteristic functions of three mentioned signals and basically can be satisfied only using quantization at high rates.[22,23] That is definitely not the case due to the structure of our quantizer that complicates and makes further analysis of (7) intractable.

Instead, we will try to upper bound $I(T_{i,j}(K); \hat{M}_i)$ by $I(T_{i,j}(K); \sum_{j=1}^{P} x_i[j]T_j(K))$ using the following Markov chain[19] between $T_{i,j}(K) \rightarrow \sum_{j=1}^{P} x_i[j]T_{i,j}(K) \rightarrow Q(\sum_{j=1}^{P} x_i[j]T_{i,j}(K))$. Let $\tilde{Y}_i = \sum_{j=1}^{P} x_i[j]T_{i,j}(K)$. Then, according to the above mentioned justification, it will be distributed with the Gaussian density with zero mean and variance $\sigma^2_{\tilde{Y}} = \sigma^2_{\tilde{Y}_i} = \sigma^2_T \sum_{j=1}^{P}(x_i[j])^2$, where $x_i[j]$ is the j^{th} element of the i^{th} hash block x_i^P. Then, a joint distribution of $\{T_{i,j}(K), \tilde{Y}_i\}$ will be bivariate Gaussian with zero mean and a 2×2 covariance matrix $\Sigma_{T_{i,j}\tilde{Y}_i}$, where

$$\Sigma_{T_i\tilde{Y}} = \begin{pmatrix} \sigma^2_T & x_i[j]\sigma^2_T \\ x_i[j]\sigma^2_T & \sigma^2_T \sum_{j=1}^{P}(x_i[j])^2 \end{pmatrix}. \quad (8)$$

Finally, using a definition of the mutual information between two Gaussian vectors[19] one has:

$$I(T_{i,j}(K); \tilde{Y}) = \frac{1}{2}\log_2\Big(\frac{\det(\Sigma_{TT})\det(\Sigma_{\tilde{Y}\tilde{Y}})}{\det(\Sigma_{T\tilde{Y}})}\Big) = \frac{1}{2}\log_2\Big(\frac{\sigma^2_T\sigma^2_T\sum_{l=1}^{P}(x_i[j])^2}{\sigma^4_T\sum_{l=1,l\neq j}^{P}(x_i[l])^2}\Big)$$
$$= \frac{1}{2}\log_2\Big(\frac{\sum_{l=1}^{P}(x_i[j])^2}{\sum_{l=1,l\neq j}^{P}(x_i[l])^2}\Big), \quad (9)$$

where $\det(\cdot)$ denotes a matrix determinant.

In order to measure the information leakage about the secret key measured in the projection domain, a set of experiments was accomplished summarized in Fig. 5. Mutual information $I(T_{i,j}(K); \tilde{Y}_i)$ as a function of a block size P was measured under the following setup:

- hashing input on a block level X^P is an i.i.d. vector of dimensionality $P, P = \{32, 64, 128, 256, 512, 1024, 2048, 4096, 8192, 16384, 32768, 65536, 262144\}$, such that X is uniformly distributed on the interval $[0, 1]$;
- random projection operator $T_i^{1\times P}(K)$ is an i.i.d. zero mean unit variance Gaussian vector of dimensionality P.

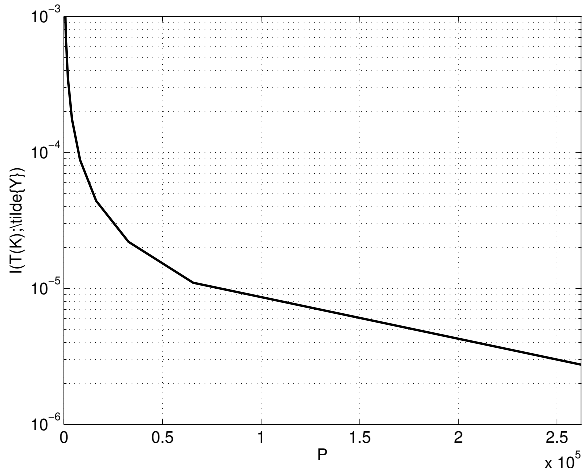

Figure 5. Information leakage about the secret key (per component) measured in the projected domain.

The obtained experimental results allow to conclude that the analyzed scheme is not perfectly secure. Information leakage is upper bounded by a non-negative value that depends on the hashing input block dimensionality. Thus, an evident suggestion to keep the security of the scheme at the maximally achievable level, one can propose to produce individual hash bits by projecting an entire vector x^N on a corresponding set of projector operators $T^{LN}(K)$ (Figure 6).

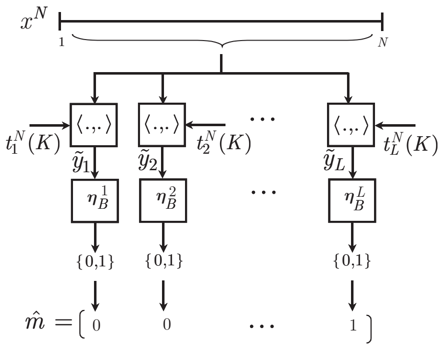

Figure 6. A practical robust perceptual hashing algorithm based on entire vector random projections.

3.1.2. Robust perceptual hashing based on dithering and distributed source coding

The second method under the analysis based on dithering of the input of the distributed source coding encoder was recently proposed by M. Johnsson and K. Ramchandran[24] (Figure 7).

The main idea of this robust perceptual hashing technique is to compress an input signal X^N previously dithered by a secret key-dependent sequence $T^{1\times N}(K)$ close to the rate-distortion limit using Wyner-Ziv encoding[25] under the assumption of the symmetric secret availability at the authenticity check. In this case one can note that the following Markov chain relationship is valid: $T^{1\times N}(K) \rightarrow (T^{1\times N}(K) + x^N) \rightarrow Q_{WZ}(T^{1\times N}(K) + x^N)$, where $Q_{WZ}(T^{1\times N}(K) + x^N)$ is the Wyner-Ziv reconstruction of $(T^{1\times N}(K) + x^N)$ uniquely relevant to the hash value that is the index of the bin in the Wyner-Ziv codebook to which this reconstruction codeword belongs.

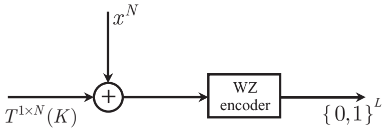

Figure 7. A practical robust perceptual hashing algorithm based on dithering and distributed coding framework.

Analysis assumptions adopted in the previous subsection for the security analysis of robust perceptual hashing based on random projections will not provide reasonable results, since availability of x^N under the additive interaction with secret $T^{1\times N}(K)$ allows to conclude that the exact realization of the dither can be obtained by the attacker at the Wyner-Ziv encoder input. Instead, in order to estimate the real security leakage, one needs to characterize $I((T^{1\times N}(K)+X^N);Q_{WZ}(T^{1\times N}(K)+X^N))$, the Wyner-Ziv rate distortion function,[25, 26] that is upper bounding the term of interest $I(T^{1\times N}(K);Q_{WZ}(T^{1\times N}(K)+X^N))$. However, a particular moment of this analysis is relevant to different kinds of side information available to legitimate user and the attacker. While, as it was already mentioned above, the hash scheme is designed supposing the availability of $T^{1\times N}(K)$ at the verification stage, the attacker is trying to benefit from the availability of the original vector x^N.

Thus, in this consideration the actual security leakage will be upper bounded by the per-sample mutual information term of the Wyner-Ziv rate distortion function[25, 26] that from the perspective of the attacker can be represented as follows:

$$I(U;X+T)-I(U;X)=H(U|X)-H(U|X+T)=H(U|X)-H(U|X+T,X)=I(U;X+T|X), \qquad (10)$$

where the second equality follows from the fact of conditional independence of U and X, given $X+T$, $U=X+T+Z$ is the auxiliary random variable used for the Wyner-Ziv random-binning-based codebook generation and Z is a zero-mean Gaussian random variable with the variance σ_Z^2 selected in order to guarantee the Wyner-Ziv distortion constraint $E\left[((X+T)-E(X+T|U,X))^2\right]\leq D$.

Furthermore, using the structure of the Wyner-Ziv codebook and denoting by $h(X),h(X|Y)$ differential entropy of a continuous random variable X and conditional differential entropy of a continuous random variable X given a realization of continuous random variable Y, one obtains:

$$\begin{aligned} I(U;X+T|X) &= h(U|X)-h(U|X+T,X) \\ &= h(X+T+Z|X)-h(X+T+Z|X+T,X) \\ &= h(T+Z|X)-h(T+Z|T,X) \\ &= h(T+Z)-h(Z) \\ &= \frac{1}{2}\log_2\left(\frac{\sigma_T^2+\sigma_Z^2}{\sigma_Z^2}\right) \\ &= \frac{1}{2}\log_2\left(1+\frac{\sigma_T^2}{\sigma_Z^2}\right) \qquad (11) \end{aligned}$$

that is nothing else as the information that can be revealed by the attacker about the secret at the output of the AWGN channel with channel noise $Z\sim\mathcal{N}(0,\sigma_Z^2)$. Evidently, in order to provide zero information leakage about the secret to the attacker, the variance σ_T^2 should be equal to zero or $\sigma_Z^2\to\infty$. Otherwise, as in the case of random projection-based hashing considered in the previous subsection, one can claim that dither-based hashing is not absolutely secure.

In order to estimate security leakage measured in terms of $I(U;X+T|X)$ in a more realistic scenario, we considered this mutual information as a function of SNR$=10\log_{10}\left(\frac{\sigma_T^2}{\sigma_Z^2}\right)$. We assumed that the dither is an i.i.d. Gaussian random variable with zero mean and unite variance. The results of our simulations are presented in Figure 8.

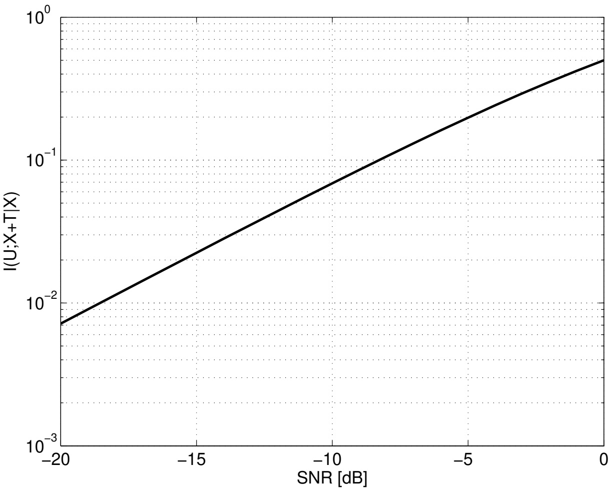

Figure 8. Information leakage about the secret key/dither.

4. CONCLUSIONS AND FUTURE RESEARCH PERSPECTIVES

In this paper we considered the problem of security analysis of robust perceptual hashing in authentication application. The conducted analysis was performed within Kerckhoffs' security framework meaning that all information about the hashing algorithm design including the original and robust hash signature are available to the opponent. The only remaining secret limiting the power of the attacker to act as a legal user is the secret key. We formulated the problem of secret key estimation by the attacker based on all available prior knowledge in terms of a number of computational efforts he requires to reach the goal. In order to see how far the opponent can advance from the exhaustive search, we proposed to measure this complexity in terms of Shannon equivocation. We performed experimental analysis of two particular robust perceptual hashing algorithms and demonstrated that both techniques provide certain auxiliary information to the attacker.

Possible ways of extending the obtained results are twofold. First, we would like to consider in more details the estimation of the actual information leakage of random projection-based techniques. Secondly, we are planning to elaborate optimal attacking scenarios in terms of trial efforts aiming at the most efficient secret key estimation strategy based on the theory of guessing.[27,28]

5. ACKNOWLEDGEMENTS

This paper was partially supported by SNF Professeur Boursier grants PP002–68653, 114613 and 200021–111643, by the European Commission through the IST Programme under contract IST-2002-507932-ECRYPT, FP6-507609-SIMILAR and Swiss IM2 projects.

REFERENCES

1. I. K. Sethin and R. C. Jain, eds., *Storage and Retrieval for Image and Video Database V*, **3022**, (San Jose, California, U.S.A.), The Society for Imaging Science and Technology (IS&T) and the International Society for Optical Engineering (SPIE), SPIE, Feb. 1997.
2. W. Mao, *Modern Cryptography: Theory and Practice*, Prentice Hall Professional Technical Reference, 2003.
3. J. Fridrich, "Robust bit extraction from images," in *Proceedings ICMCS'99*, **2**, pp. 536–540, (Florence, Italy), June 1999.
4. M. Lifshits, I. Blayvas, R. Goldenberg, E. Rivlin, and M. Rudzsky, "Rehashing for bayesian geometric hashing," in *ICPR 2004*, August 2004.
5. M. K. Mihcak, R. Venkatesan, and T. Liu, "Watermarking via optimization algorithms for quantizing randomized semi-global image features,"

6. V. Monga and B. L. Evans, "Robust perceptual image hashing using feature points," in *ICIP 2004*, **3**, pp. 677–680, (Singapore), October 2004.
7. R. Venkatesanan, S. Koon, M. Jacubowski, and P. Moulin, "Robust image hashing," in *ICIP 2000*, (Vancouver, BC, Canada), September 2000.
8. O. Koval, S. Voloshynovskiy, F. Beekhof, and T. Pun, "Decision-theoretic consideration of robust hashing: link to practical algorithms," in *WaCha 2007*, (Saint-Malo, France), June 2007.
9. O. Koval, S. Voloshynovskiy, F. Beekhof, and T. Pun, "Robust perceptual hashing as classification problem: decision-theoretic and practical considerations," in *MMSP 2007*, (Chania, Crete, Greece), October 2007.
10. A. Kerckhoffs, "La cryptographie militaire," *Journal des sciences militaires* **IX**, 1883.
11. S. S. Kozat, R. Venkatesan, and M. K. Mihcak, "Robust hashing via matrix invariances," in *ICIP 2004*, **3**, pp. 677–680, (Singapore), October 2004.
12. B. Coskun and N. Memon, "Robust image hashing," in *ICIP 2000*, (Vancouver, BC, Canada), September 2000.
13. C. E. Shannon, "Communication theory of secrecy systems," *Bell System Technical Journal* **28**, pp. 656–715, 1949.
14. A. Swaminathan, Y. Mao, and M. Wu, "Robust and secure hashing for images," *IEEE Transactions on Information Forensics and Security* **1**, pp. 215–230, June 2006.
15. Y. Mao and M. Wu, "Unicity distance of robust image hashing," *IEEE Trans. on Information Forensics and Security* **2**, pp. 462–467, September 2007.
16. P. Moulin and R. Koetter, "A framework for the design of good watermark identification codes," in *IS&T/SPIE Electronic Imaging'06, Session: Security and Watermarking of Multimedia Contents*, (San Jose, CA, USA), January 2006.
17. F. Cayre, C. Fontaine, and T. Furon, "Pwatermarking security: Theory and practice," *IEEE Transactions on Signal Processing* **special issue "Supplement on Secure Media II"**, 2005.
18. E. Topak, S. Voloshynovskiy, O. Koval, M. Mihcak, and T. Pun, "Security analysis of robust data hiding with geometrically structured codebooks," in *Proceedings of the SPIE International Conference on Security and Watermarking of Multimedia Contents*, (San Jose, CA, USA), January 2005.
19. T. Cover and J. Thomas, *Elements of Information Theory.*, Wiley and Sons, New York, 1991.
20. R. Villan, S. Voloshynovskiy, O. Koval, F. Deguillaume, and T. Pun, "Tamper-proofing of electronic and printed text documents via robust hashing and data-hiding," in *IS&T/SPIE Electronic Imaging'7, Session: Security and Watermarking of Multimedia Contents*, (San Jose, CA, USA), January 2007.
21. N. Jayant and P. Noll, *Digital coding of waveforms: principles and applications to speech and video*, Prentice-Hall, 1984.
22. L. Schuchman, "Dither signals and their effect on quantization noise," in *IEEE Trans. on Commun. Technol.*, **COM-12**, pp. 162–165, December 1964.
23. B. Widrow and I. Kollar, *Quantization Noise in Digital Computation, Signal Processing, and Control*, New York, -:Cambridge University Press, 2007.
24. M. Johnson and K. Ramchandran, "Dither-based secure image hashing using distributed source coding," in *Proceedings of the IEEE Int. Conf. on Image Processing ICIP-03*, (Barcelona, Spain), september 2003.
25. A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Information Theory* **22**, pp. 1–10, January 1976.
26. A. Wyner, "The rate-distortion function for source coding with side information at the decoder-ii: General sources," *Information and Control* **38**, pp. 60–80, 1978.
27. J. L. Massey, "Guessing and entropy," in *IEEE International Symposium on Information Theory ISIT 1994*, **3**, p. 204, (Trondheim, Norway), June 1994.
28. E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Trans. on Information Theory* **42**, January 1996.