Private content identification based on soft fingerprinting

Sviatoslav Voloshynovskiy*, Taras Holotyak, Oleksiy Koval, Fokko Beekhof and Farzad Farhadzadeh
University of Geneva, Department of Computer Science,
7 route de Drize, CH 1227, Geneva, Switzerland

ABSTRACT

In many problems such as biometrics, multimedia search, retrieval, recommendation systems requiring privacy-preserving similarity computations and identification, some binary features are stored in the public domain or outsourced to third parties that might raise certain privacy concerns about the original data. To avoid this privacy leak, privacy protection is used. In most cases, privacy protection is uniformly applied to all binary features resulting in data degradation and corresponding loss of performance. To avoid this undesirable effect we propose a new privacy amplification technique that is based on data hiding principles and benefits from side information about bit reliability a.k.a. *soft fingerprinting*. In this paper, we investigate the identification-rate vs privacy-leak trade-off. The analysis is performed for the case of a perfect match between side information shared between the encoder and decoder as well as for the case of partial side information.

1. INTRODUCTION

Content identification systems are widely used in various emerging applications ranging from identification of physical objects and humans to multimedia management (content filtering, content tagging) and security (copyright protection, broadcast monitoring, etc.). Most identification techniques are based on binary digital fingerprinting. A digital fingerprint represents a short, robust and distinctive content description allowing fast and privacy-preserving operations. In this case, all operations are performed on the fingerprint instead of on the original large and privacy-sensitive data, thus allowing introduction of crypto-based security into the analog or noisy digital world.[1] These new techniques are able to overcome the fundamental sensitivity of classical cryptographic encryption and one-way functions to minor noise in input data by trading-off the security and robustness.

This paper is an extension of our previous work.[2,3] We have previously considered the rate-privacy-complexity trade-off for identification applications.[2] This approach is based on global privacy amplification, where all bits of stored fingerprints are randomized with identical probability disregarding their reliabilities. This approach is similar in spirit to a compression-based approach.[4] However, contrarily to the previous approach a concept of bit reliability was introduced to reduce the identification complexity based on a bounded distance decoder (BDD).[5] Obviously, such a construction does not fully benefit from the fact that the information about the reliable bits can be present at the encoder and decoder, which can be used not only for the efficient decoding but also for the enhanced privacy amplification.

Therefore, we introduced an information-theoretic framework for the analysis of private content identification based on finite length fingerprinting with bit reliability side information.[3] In this paper, contrary to previous works in content authentication based on *helper data*,[1,4,6,7] we propose and extend a privacy amplification mechanism, which is adaptive to the bit reliability, and demonstrate its advantages over the state-of-the-art privacy amplification in the identification problem. We present and analyze a privacy-preserving technique, which asymptotically achieves the theoretical identification performance limits in terms of identification rate. The analysis is performed for the case of a perfect match between the side information shared between the encoder and decoder as well as for the case of partial side information.

Notations. We use capital letters to denote scalar random variables X, bold capital letters to denote vector random variables $\mathbf{X}$, corresponding small letters x and small bold letters $\mathbf{x}$ to denote the realizations of scalar and vector random variables, respectively, i.e., $\mathbf{x} = \{x(1), x(2), ..., x(N)\}$. $\mathbf{b_x}$ is used to denote the binary version of $\mathbf{x}$. We use $X \sim p(x)$ to indicate that a random variable X follows $p_X(x)$.

*The contact author is S. Voloshynovskiy (email: svolos@unige.ch). http://sip.unige.ch

2. IDENTIFICATION PROBLEM FORMULATION

We will assume that the *data owner* has M entries in the database indexed by an index m, i.e., $\mathbf{x}(m) \in \mathbb{R}^N$, $1 \leq m \leq M$, where $M = 2^{LR}$ with R to be the identification rate of an (M, L)-fingerprinting code and L stands for the fingerprint length. The index m is associated with all identification information (ownership, time of creation, distribution channel, etc.) and the data $\mathbf{x}(m)$ is some privacy sensitive part of the database represented by images, video, audio, biometrics, physical unclonable functions (PUFs), etc. The *data user* has a query data $\mathbf{y} \in \mathbb{R}^N$ that can be in some relationship with $\mathbf{x}(m)$ via a probabilistic model $p(\mathbf{y}|\mathbf{x})$ or can represent some irrelevant input $\mathbf{x}'$. The data user wishes to retrieve the identification information of $\mathbf{x}(m)$ that is the closest to the query $\mathbf{y}$ or reject the query, if no relevant database entry is found. For complexity and privacy reasons, the above identification is performed in the domain of digital fingerprints $\mathbf{b_x} \in \{-1, +1\}^L$ and $\mathbf{b_y} \in \{-1, +1\}^L$ that are short length, secure and robust counterparts of $\mathbf{x}$ and $\mathbf{y}$, respectively. Moreover, to ensure adequate privacy protection of digital fingerprints, the data owner applies privacy amplification (PA) to produce a protected version $\mathbf{b_u}(m)$ of $\mathbf{b_x}(m)$. The resulting fingerprints can be shared with third parties for various security and management services. In particular, the storage of the resulting codebook/database of protected fingerprints $\mathbf{b_u}(m)$, $1 \leq m \leq M$, and the content identification can be performed on a remote server that can be honest in terms of claimed functionalities but curious in terms of observing, analyzing or leaking the stored data. The result of identification should be an estimate of index $\hat{m}$ of the corresponding closest entry or the erasure, i.e., null hypothesis. If the query is properly identified, the corresponding encrypted content $\mathbf{x}(m)$ or associated identification information is delivered to the data user using the predefined data exchange protocol. At the same time, the attacker can observe the entire database and analyze query.[8]

In the scope of this paper, we will assume that the binary fingerprints are obtained by a dimensionality reduction transform $\mathbf{W}$ and binarization Q. The projected vectors of lower dimensionality $\tilde{\mathbf{x}}(m) \in \mathbb{R}^L$ and $\tilde{\mathbf{y}} \in \mathbb{R}^L$ are obtained from $\mathbf{x}(m)$ and $\mathbf{y}$ based on the dimensionality reduction transform:

$$\tilde{\mathbf{x}}(m) = \mathbf{W}\mathbf{x}(m), \tag{1}$$

$$\tilde{\mathbf{y}} = \mathbf{W}\mathbf{y}, \tag{2}$$

where $\mathbf{W} \in \mathbb{R}^{L \times N}$ with $L \leq N$ and $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_L)^T$ consists of a set of projection basis vectors $\mathbf{w}_i \in \mathbb{R}^N$ with $1 \leq i \leq L$. The dimensionality reduction transform is based on any randomized orthogonal matrix $\mathbf{W}$ (random projection transform) whose elements $w_{i,j}$ are generated from some specified distribution.

The binarization is performed as:

$$b_{\mathbf{x}_i} = sign(\mathbf{w}_i^T \mathbf{x}), \tag{3}$$

where $b_{\mathbf{x}_i} \in \{-1, +1\}$, with $1 \leq i \leq L$ and $sign(a) = +1$, if $a \geq 0$ and -1, otherwise. Since all projections are independent, it can be assumed that all bits in $\mathbf{b_x}$ will be independent and equiprobable for sufficiently large L.[†]

The mismatch between the data owner fingerprint $\mathbf{b_x}$ and data user query fingerprint $\mathbf{b_y}$ can be modeled based on the memoryless binary symmetric channel (BSC) model with a probability of bit error P_b. It was shown that $P_b = \frac{1}{\pi} \arccos(\rho_{\tilde{X}\tilde{Y}})$, where $\rho_{\tilde{X}\tilde{Y}}$ is a correlation coefficient between $\tilde{X}$ and $\tilde{Y}$.[2]

In,[2] we formulated the identification problem with the global privacy amplification as a composite hypothesis test:

$$\begin{cases} H_0: & p(\mathbf{b_y}|H_0) = p(\mathbf{b_y}|\mathbf{b_x}'), \\ H_m: & p(\mathbf{b_y}|H_m) = p(\mathbf{b_y}|\mathbf{b_u}(m)), m = 1, \cdots, M. \end{cases} \tag{4}$$

In the binary fingerprinting domain, the link between $\mathbf{b_x}$ and between $\mathbf{b_y}$ and $\mathbf{b_x}$ and $\mathbf{b_u}$ can be considered based on the BSC models with corresponding bit error probabilities P_b and λ, respectively. The parameter λ corresponds to the BSC serving as a test channel for the compressed version $\mathbf{b_u}$ considered as the privacy amplification. Under the above assumption, these two BSCs $\mathbf{b_x} \to \mathbf{b_u}$ and $\mathbf{b_x} \to \mathbf{b_y}$ can be considered as

[†]This assumption is only possible for independent input data. Since the transformed vectors will closely follow the Gaussian pdf but will not necessarily be decorrelated, one can apply the principle component analysis to decorrelate them, that, for the case of Gaussian data, will also provide their independence.

an equivalent channel $\mathbf{b_u} \to \mathbf{b_y}$ obtained by their concatenation with the cross-probability P_{b_e} equals to the convolution $P_{b_e} = P_b * \lambda = P_b(1-\lambda) + \lambda(1-P_b)$. Under these conditions, the corresponding hypothesis (4) are:

$$\begin{cases} H_0: & p(\mathbf{b_y}|\mathbf{b_x}') = \frac{1}{2^L}, \\ H_m: & p(\mathbf{b_y}|\mathbf{b_u}(m)) = \left(\frac{P_{b_e}}{1-P_{b_e}}\right)^{d^H(\mathbf{b_y},\mathbf{b_u}(m))} (1-P_{b_e})^L, \end{cases} \tag{5}$$

where $d^H(.,.)$ denotes the Hamming distance.

Under (5), the corresponding test can be written as:

$$d^H(\mathbf{b_y}, \mathbf{b_u}(m)) \leq L\gamma, \tag{6}$$

where $\gamma = \frac{-\tau + \log_2(1-P_{b_e})}{\log_2 \frac{1-P_{b_e}}{P_{b_e}}}$. We will refer to this decision rule as the BDD that produces a unique $\hat{m}$. To minimize the overall identification error and to achieve the identification capacity that coincides with the capacity of the corresponding BSC, it was shown that the threshold should satisfy $\gamma_{\text{opt}} = \frac{1-R+\log_2(1-P_{b_e})-1/L}{\log_2\left(\frac{1-P_{b_e}}{P_{b_e}}\right)}$.[2] The efficient implementation of identification search strategy based on the Hamming sphere-based BDD interpretation was demonstrated in the same paper along the conditions on P_{b_e} under which this search outperforms the exhaustive one. For the case of asymptotically large L, it was also shown that:

Proposition 1. For $P_{b_e} \leq \gamma \leq \frac{1}{2}$ and if $H_2(\gamma) \leq 1 - R$ there exist codes with rate R and error probability P_e such that:

$$\lim_{L\to\infty} P_e = 0. \tag{7}$$

As soon as γ is arbitrarily close to P_{b_e}, the rate $R = 1 - H_2(P_{b_e})$ is achievable, and it is referred to as private identification capacity:

$$C_{id} = 1 - H_2(P_{b_e}). \tag{8}$$

The privacy leak L_p about $\mathbf{B_x}$ from the public $\mathbf{B_u}$ is defined by the mutual information between them[‡]:

$$L_p = I(B_{\text{u}}; B_{\text{x}}) = 1 - H_2(\lambda). \tag{9}$$

The trade-off between the identification rate and privacy leak is achieved by the selection of parameter λ. This parameter is applied to all bits disregarding their actual bit reliability shown to be equal to[2]:

$$P_{b|\tilde{x}} = Q\left(\frac{|\tilde{x}|}{\sigma_Z}\right), \tag{10}$$

which stands for the bit error probability for a given projection coefficient $\tilde{x}$ under the assumption that $p(\tilde{x}, \tilde{y})$ corresponds to jointly Gaussian distribution in the random projection domain and σ_Z denotes a standard deviation of additive Gaussian noise in the random projection domain.

3. PRIVACY PROTECTION STRATEGIES

Having introduced the performance and privacy leakage measures in the identification setup, we will characterize the existing privacy preserving strategies and demonstrate their shortcomings. The analysis will be focused on two known techniques, i.e., encryption and randomization/compression, and a new alternative technique that we refer to as privacy protection based on data hiding.

[‡]A more conservative definition of privacy leak would be $I(B_{\text{u}}; X)$.

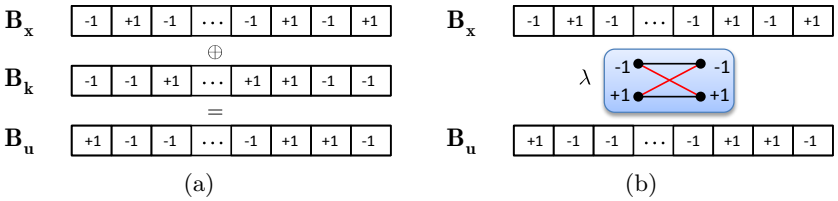


Figure 1: Privacy protection based on: (a) encryption and (b) randomization/compression.

3.1. Privacy protection based on encryption

Proposition 2. (*protection based on encryption*): The entire database $\mathbf{b_x}(m)$, $1 \leq m \leq M$ containing all binary codewords is encrypted by the same key $\mathbf{b_k} \in \{\pm 1\}^L$ thus resulting in $\mathbf{b_u}(m) = \mathbf{b_x}(m) \oplus \mathbf{b_k}$ (Fig. 1a). Then the identification rate under the known key and identification BSC with P_b and privacy leak for an uninformed attacker are:

$$R_{id}^{\text{Crypto}} = I(B_{\text{u}}; B_{\text{y}}, B_{\text{k}}) = 1 - H_2(P_b), \tag{11}$$

$$L_p^{\text{Crypto}} = I(B_{\text{x}}; B_{\text{u}}) = \frac{(M-1)L}{ML}, \tag{12}$$

which tends to 1 for large M.

Proof: The rate R_{id}^{Crypto} is defined by the cross-over probability of the identification BSC. Therefore, it is necessary to demonstrate only the part of L_p^{Crypto} when M entries are encrypted by the same key $\mathbf{B}_{\text{k}}$:

$$\begin{aligned} I(\mathbf{B}_{\text{u}}(1), \cdots, \mathbf{B}_{\text{u}}(M); \mathbf{B}_{\text{x}}(1), \cdots, \mathbf{B}_{\text{x}}(M)) &= I(\mathbf{B}_{\text{x}}(1) \oplus \mathbf{B}_{\text{k}}, \cdots, \mathbf{B}_{\text{x}}(M) \oplus \mathbf{B}_{\text{k}}; \mathbf{B}_{\text{x}}(1), \cdots, \mathbf{B}_{\text{x}}(M)) && (13)\\ &= H(\mathbf{B}_{\text{x}}(1) \oplus \mathbf{B}_{\text{k}}, \cdots, \mathbf{B}_{\text{x}}(M) \oplus \mathbf{B}_{\text{k}}) && (14)\\ &- H(\mathbf{B}_{\text{x}}(1) \oplus \mathbf{B}_{\text{k}}, \cdots, \mathbf{B}_{\text{x}}(1) \oplus \mathbf{B}_{\text{k}} | \mathbf{B}_{\text{x}}(1), \cdots, \mathbf{B}_{\text{x}}(M)) && (15)\\ &= (M-1)L, && (16) \end{aligned}$$

for the assumed i.i.d. uniform fingerprints. This corresponds to the leak about the entire database that can be estimated per symbol of fingerprint as $\frac{(M-1)L}{ML}$. $\frac{(M-1)L}{ML} \to 1$ for large M.

This result confirms the pessimistic result of Shannon that the entropy of key should be larger or equal to the entropy of plain text to be encrypted. If the entropy of key were at least the same as the entire database, i.e, each individual codeword had its own key, the above privacy leak would be equal to zero.[§] This would provide perfect privacy. In the above case, the entropy of the key is just L that leads to the asymptotic disclosure of the database.

Therefore, such a protection has several severe security issues related to key management. The key should be known at the decoder and it can not be used twice. However, according to the condition of Proposition 2, the entire database is encrypted by the same key. At the same time, the usage of individual key $B_{\text{k}}(m)$ for each particular database entry $B_{\text{x}}(m)$ is not practical, since it is equivalent to the identification itself.

3.2. Privacy protection based on randomization/compression

Proposition 3. (*protection based on randomization/compression*): Each codeword $\mathbf{b_x}(m)$, $1 \leq m \leq M$ of the database is randomized or compressed with the factor λ as considered in section 2. This is equivalent to sending the database over the BSC with the cross-over probability λ (Fig. 1b). Then the identification rate under the identification BSC with P_b and privacy leak are:

$$R_{id}^{\text{Rand}} = I(B_{\text{u}}; B_{\text{y}}) = 1 - H_2(\lambda * P_b), \tag{17}$$

$$L_p^{\text{Rand}} = I(B_{\text{x}}; B_{\text{u}}) = 1 - H_2(\lambda). \tag{18}$$

[§]The term $H(\mathbf{B}_{\text{x}}(1) \oplus \mathbf{B}_{\text{k}}, \cdots, \mathbf{B}_{\text{x}}(1) \oplus \mathbf{B}_{\text{k}} | \mathbf{B}_{\text{x}}(1), \cdots, \mathbf{B}_{\text{x}}(M))$ would be equal to ML in this case.

This privacy protection technique does not require any key at the decoder and the effect of privacy protection is solely based on data degradation. Therefore, the decoder is not aware about the data degradations introduced at the encoder. However, the lack of a shared secret between the encoder and decoder leads to a considerable degradation of performance. The condition of perfect privacy, when $L_p^{\text{Rand}} \to 0$, can be only achieved when $\lambda \to 0.5$. However, in this case $R_{id}^{\text{Rand}} \to 0$ and better privacy is achieved at the price of significant rate loss. This conclusion can be also generalized to content authentication,[4] where the privacy protection is implicitly based on the Wyner-Ziv framework with the compressed data in the codebook construction.

3.3. Privacy protection based on data hiding

In this section we consider an alternative technique that benefits from the common secret $\mathbf{s}$ shared between the encoder and decoder. This shared secret has certain properties that differ from the conditions of proposition 2:

(a) the secret $\mathbf{s}$ is unique for each content, i.e., in total one has $\mathbf{s}(m)$, $1 \leq m \leq M$ secrets at the encoder;

(b) the secret $\mathbf{s}$ is independent of the binary fingerprint $\mathbf{b_x}$ but it is extracted from $\mathbf{x}$ or its projected version $\tilde{\mathbf{x}}$;

(c) there might be a mismatch between the secret at the encoder $\mathbf{s}^e$ and the secret at the decoder $\mathbf{s}^d$;

(d) the presence of shared secret or side information converts the binary identification channel with cross-over probability P_b into a BSC with the secret dependent cross-over probability $P_{b|s}$;

(e) the secrets $\mathbf{s}^e$ and $\mathbf{s}^d$ can be considered as the magnitudes of the projected vectors $\mathbf{s}^e = |\tilde{\mathbf{x}}|$ and $\mathbf{s}^d = |\tilde{\mathbf{y}}|$ or positions of certain bits with predefined properties $\mathbf{s}^e = \mathbf{p_x}$ and $\mathbf{s}^d = \mathbf{p_y}$, where $\mathbf{p_x} \in \{0,1\}^L$ and $\mathbf{p_y} \in \{0,1\}^L$ with 1s indicating the positions of these bits.

Proposition 4. (*identification rate with shared secret*): The identification rate with shared secrets $\mathbf{s}^e$ and $\mathbf{s}^d$ independent with $\mathbf{b_x}$ and $\mathbf{b_y}$ is:

$$R_{id}^{s^e,s^d} = I(B_{\text{x}}, S^e; B_{\text{y}}, S^d) = I(S^e; S^d) + I(B_{\text{x}}; B_{\text{y}}|S^e). \tag{19}$$

Proof:

$$I(B_{\text{x}}, S^e; B_{\text{y}}, S^d) \overset{(a)}{=} I(B_{\text{y}}, S^d; S^e) + I(B_{\text{y}}, S^d; B_{\text{x}}|S^e) \tag{20}$$

$$= I(S^e; S^d) + \underbrace{I(B_{\text{y}}; S^e|S^d)}_{(b)=0} + I(B_{\text{x}}; B_{\text{y}}|S^e) + \underbrace{I(S^d; B_{\text{x}}|S^e, B_{\text{y}})}_{(c)=0}, \tag{21}$$

where (a) follows from the chain rule for mutual information, (b) results from $I(B_{\text{y}}; S^e|S^d) = H(B_{\text{y}}|S^d) - H(B_{\text{y}}|S^e, S^d) = 0$ and (c) from $I(S^d; B_{\text{x}}|S^e, B_{\text{y}}) = H(B_{\text{x}}|S^e, B_{\text{y}}) - H(B_{\text{x}}|S^e, B_{\text{y}}, S^d) = 0$.

In this paper, we do not address the possible rate gain due to the side information $I(S^e; S^d)$ since S^e will not be stored in the database due to the privacy and memory storage reasons. However, this option should be noticed as a very promising one due to the possible significant rate gain. Therefore, in the following we only concentrate on the term $I(B_{\text{x}}; B_{\text{y}}|S^e)$. However, since one observes only S^d instead of S^e and B_{u} is stored in the database instead of B_{x}, we will consider the protected counterpart $I(B_{\text{u}}; B_{\text{y}}|S^d)$ while the term $I(B_{\text{u}}; B_{\text{y}}|S^e)$ will be analyzed for the comparison reasons.

The main idea behind the privacy protection based on data hiding consists in secret placement of L fingerprint bits into the random vector of length $J > L$ keeping the positions of fingerprint bits only known to authorized parties sharing the secrets $\mathbf{s}^e$ and $\mathbf{s}^d$. For purposes of analysis and assuming perfect secret sharing $\mathbf{s}^e = \mathbf{s}^d = \mathbf{s}$, the considered setup can be converted into two equivalent parallel channels as shown in Fig. 2.

It is assumed that L fingerprint bits are grouped together and communicated via the same channel or stored with the same randomization factor λ_1 and the remaining $J - L$ bits are random masking bits communicated

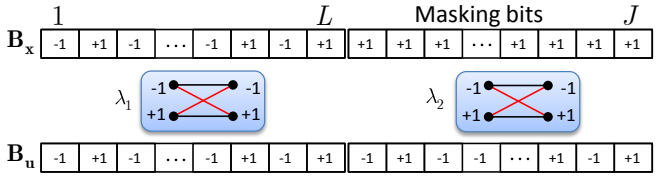


Figure 2: Privacy protection based on data hiding: equivalent communication model.

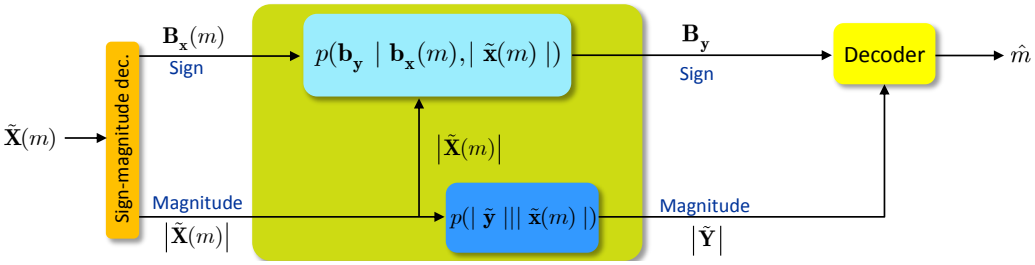


Figure 3: Fingerprinting model of communications based on sign-magnitude decomposition.

or stored with the randomization factor $\lambda_2 > \lambda_1$. Alternatively, it is also possible to consider $\mathbf{b_x}$ to consists of J bits, where L bits are considered as "useful" or informative bits and $J - L$ bits are interpreted as "host" or masking bits for hiding. If $\lambda_2 = \lambda_1$, the considered scheme is equivalent to one from section 3.2. Another particular case consists in the complete randomization of the masking bits with $\lambda_2 = 0.5$ with the simultaneous preservation of useful bits, i.e., $\lambda_1 = 0$.

Proposition 5. (*protection based on data hiding*): Assuming the secret placement of useful bits with perfect secret sharing $\mathbf{s}^e = \mathbf{s}^d = \mathbf{s}$ and $\lambda_1 = 0$ and $\lambda_2 = 0.5$, the rate-privacy pair is:

$$R_{id}^{\mathrm{DH}} = I(B_\mathrm{u}; B_\mathrm{y}|S) = \frac{L}{J}\left(1 - H_2(P_{b|s})\right), \tag{22}$$

$$L_p^{\mathrm{DH}} = I(B_\mathrm{x}; B_\mathrm{u}) = 1 - H_2\left(\frac{J-L}{2J}\right). \tag{23}$$

Remark 1. The privacy L_p^{DH} reflects the fact that common information between B_u and B_x might be exploited by the attacker to reconstruct B_x from B_u. At the same time, the attacker might try to guess the positions of L fingerprint bits using combinatoric analysis. In this case, the number of guesses can be bounded as:

$$\binom{J}{L} \leq 2^{JH_2\left(\frac{L}{J}\right)}, \tag{24}$$

trials which results for example for $J = 1024$ and $L = 32$ into 2^{205} trials that makes this strategy unfeasible.

Being very attractive, the considered strategy in practice requires the precise definition of a secret sharing mechanism between $\mathbf{s}^e$ and $\mathbf{s}^d$ and computation of conditions for which $P_{b|s} < P_b$. This analysis will be performed based on a sign-magnitude decomposition that is introduced in the next section.

4. SIGN-MAGNITUDE DECOMPOSITION

Without loss of information one can decompose the projected coefficient as $\tilde{x} = sign(\tilde{x})|\tilde{x}| = b_\mathrm{x} s^e$, where $b_\mathrm{x} = sign(\tilde{x})$ and $s^e = |\tilde{x}|$ as shown in Fig. 3. Assuming the independence between the sign and magnitude, the general identification channel $p(\tilde{y}|\tilde{x})$ can be decomposed into two channels $b_\mathrm{x} \to b_\mathrm{y}$ and $|\tilde{x}| \to |\tilde{y}|$. In the general case, the channel $|\tilde{x}| \to |\tilde{y}|$ is also the information transmission channel. However, we will consider it as a channel for solely common secret sharing. Thus, the main identification channel is considered in part of BSC $b_\mathrm{x} \to b_\mathrm{y}$. At the same time, this channel has a state determined by $|\tilde{x}|$.

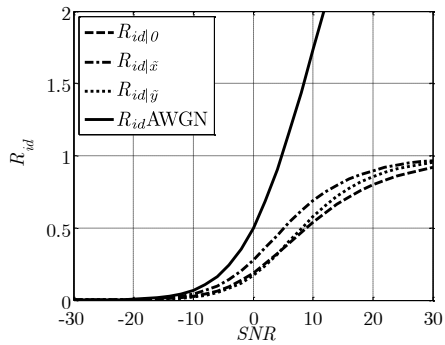


Figure 4: The identification rates under hard fingerprinting and soft fingerprinting with perfect and partial CSI.

4.1. Theoretical model of channel decomposition

Assuming the additive Gaussian noise observation model $\tilde{y} = \tilde{x} + \tilde{z}$, where $\tilde{z}$ represents the zero-mean Gaussian noise with variance σ_Z^2 in the projected domain, one can characterize the behavior of BSC $b_{\mathrm{x}} \to b_{\mathrm{y}}$ by the cross-over probability (10), which is determined by the state $|\tilde{x}|$.[2] Under such a setting, many existing fingerprinting systems can be considered as *hard* fingerprinting, i.e., those that do not use information about the channel state, and *soft* fingerprinting, i.e., those that benefit from this knowledge. Therefore, depending on the availability of *channel state information* (CSI), one can distinguish 3 major cases: (a) *no CSI* ($S^e = \oslash, S^d = \oslash$) where only $p(\tilde{x})$ is known (hard fingerprinting), (b) *perfect CSI* ($S^e = S^d = S = |\tilde{X}|$) (soft fingerprinting with perfect CSI) and (c) *partial CSI* ($S^e = |\tilde{X}|$, $S^d = |\tilde{Y}|$) (soft fingerprinting with partial CSI). At this stage, we assume that no privacy amplification is applied and $B_{\mathrm{u}} = B_{\mathrm{x}}$.

Remark 2. The identification rate under the hard fingerprinting is:

$$R_{id|0} = I(B_{\mathrm{u}}; B_{\mathrm{y}}|\oslash) = 1 - H_2(P_b), \tag{25}$$

where $P_b = E_{p(\tilde{x})}[P_{b|\tilde{x}}]$ is the average probability of bit error.

Remark 3. The identification rate under the soft fingerprinting with perfect CSI is:

$$R_{id|\tilde{x}} = I(B_{\mathrm{u}}; B_{\mathrm{y}}||\tilde{X}|) = \int_{-\infty}^{+\infty} I(B_{\mathrm{u}}; B_{\mathrm{y}}||\tilde{X}| = |\tilde{x}|)p(\tilde{x})d\tilde{x} = 1 - 2\int_{0}^{+\infty} H_2\left(Q\left(\frac{\tilde{x}}{\sigma_Z}\right)\right)p(\tilde{x})d\tilde{x}. \tag{26}$$

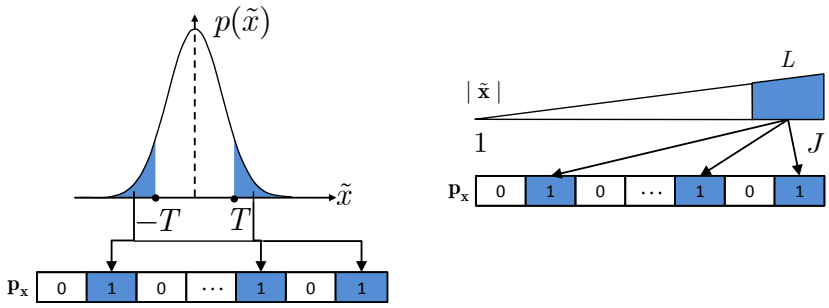
Remark 4. The identification rate under the soft fingerprinting with partial CSI is:

$$R_{id|\tilde{y}} = I(B_{\mathrm{u}}; B_{\mathrm{y}}||\tilde{Y}|) = 1 - 2\int_{0}^{+\infty} H_2\left(\int_{-\infty}^{+\infty} Q\left(\frac{\tilde{y}}{\sigma_Z}\right)p(\tilde{y}|\tilde{x})d\tilde{y}\right)p(\tilde{x})d\tilde{x}. \tag{27}$$

The identification rates under hard fingerprinting, soft fingerprinting with perfect and partial CSIs are shown in Fig. 4. The observation model is considered in terms of the signal-to-noise ratio (SNR) defined as $\mathrm{SNR} = 10\log_{10}\frac{\sigma_X^2}{\sigma_Z^2}$. For the comparison reasons the capacity of AWGN identification channel $C = I(\tilde{X}; \tilde{Y}) = \frac{1}{2}\log_2\left(1 + \frac{\sigma_X^2}{\sigma_Z^2}\right)$is also shown in Fig. 4.

Remark 5. The impact of CSI on the identification rates:

- the presence of perfect CSI at the decoder enhances the identification rate with respect to the hard fingerprinting;
- the partial CSI at the decoder enhances the identification rate for the high SNRs, i.e., in the region where it is not severely corrupted by the observation noise, and contrarily slightly degrades the rate with respect to the hard fingerprinting for the low SNRs;



(a) (b)

Figure 5: Channel splitting based on: (a) thresholding and (b) sorting.

- the achievable identification rate for all considered fingerprinting techniques is saturated at 1 for the high SNR. The gap between the identification rate of the AWGN identification and fingerprinting channels is in part of $I(|\tilde{X}|;|\tilde{Y}|)$, in accordance with Proposition 4.

Remark 6. Since the rate $I(|\tilde{X}|;|\tilde{Y}|) > I(B_{\mathrm{u}};B_{\mathrm{y}}||\tilde{Y}|)$ at the high SNR, one can reverse the consideration and analyze the identification setup with respect to the channel $|\tilde{X}| \to |\tilde{Y}|$ instead of $B_{\mathrm{u}} \to B_{\mathrm{y}}$. In this case, B_{u} is considered as the helper data.

4.2. Practical model of channel decomposition

In this part, we introduce the practical model that achieves the above theoretical limits based on a *channel splitting*. The channel splitting model assumes that the fingerprint bits are transmitted via several BSCs with parameters defined by the CSI. In the most simple case of 2-channel splitting, two BSCs are considered. Without loss of generality the channel splitting can be performed based on the available side information. At this moment, we assume that $\mathbf{s}^e = \mathbf{s}^d = |\tilde{\mathbf{x}}|$. The channel splitting can be performed based on the thresholding of magnitude coefficients with the threshold T (Fig. 5a). Equivalently, the same proportion of coefficients can be chosen based on sorting of magnitudes $|\tilde{\mathbf{x}}|$ of J coefficients and selecting L strongest ones (Fig. 5b). The positions of selected coefficients correspond to another form of secret sharing via vector $\mathbf{p_x}$ containing 1s in the positions of reliable components.

The L fingerprint bits related to the large magnitude coefficients are considered as those belonging the good BSC with the cross-over probability P_b^G and the remaining to the bad one with P_b^B.

Remark 7. The cross-over probabilities for good and bad channels based on the perfect CSI ($\mathbf{s}^e = \mathbf{s}^d = |\tilde{\mathbf{x}}|$) are:

$$P_{b|\tilde{x}}^G = \frac{2}{\mathrm{Pr}^G} \int_T^{+\infty} Q\left(\frac{\tilde{x}}{\sigma_Z}\right) p(\tilde{x}) d\tilde{x}, \tag{28}$$

$$P_{b|\tilde{x}}^B = \frac{2}{\mathrm{Pr}^B} \int_0^{T} Q\left(\frac{\tilde{x}}{\sigma_Z}\right) p(\tilde{x}) d\tilde{x}, \tag{29}$$

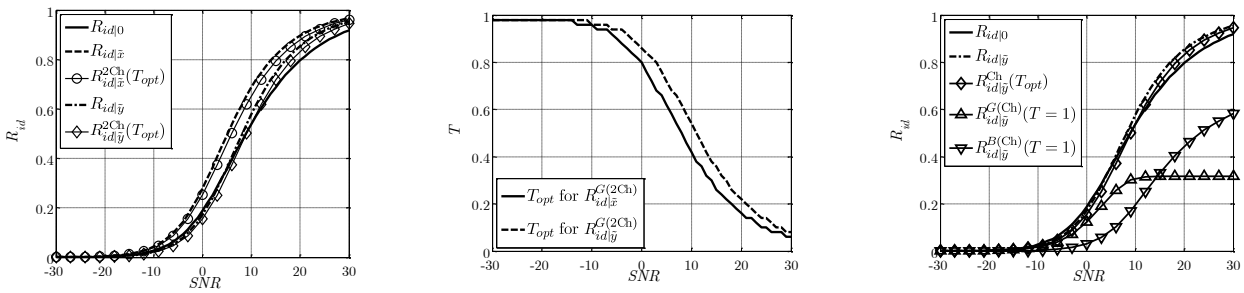
where $\mathrm{Pr}^G = 2\int_T^{+\infty} p(\tilde{x})d\tilde{x}$ and $\mathrm{Pr}^B = 2\int_0^T p(\tilde{x})d\tilde{x}$ correspond to probabilities of observing the good and bad channels, respectively. The corresponding identification rates are:

$$R_{id|\tilde{x}}^G = \mathrm{Pr}^G \left(1 - H_2\left(P_{b|\tilde{x}}^G\right)\right), \tag{30}$$

$$R_{id|\tilde{x}}^B = \mathrm{Pr}^B \left(1 - H_2\left(P_{b|\tilde{x}}^B\right)\right), \tag{31}$$

and the total rate is:

$$R_{id|\tilde{x}}^{2\mathrm{Ch}} = R_{id|\tilde{x}}^G + R_{id|\tilde{x}}^B. \tag{32}$$



(a) (b) (c)

Figure 6: Approaching theoretical rates based on 2-channel splitting model for the optimal threshold selection: (a) achievable identification rates under different CSIs (b) optimal thresholds for perfect and partial CSI (c) rates for good and bad channels for $T = 1$.

Remark 8. The cross-over probabilities for good and bad channels based on the partial CSI ($\mathbf{s}^e = |\tilde{\mathbf{x}}|$ and $\mathbf{s}^d = |\tilde{\mathbf{y}}|$) are:

$$P^G_{b|\tilde{y}} = \frac{2}{\mathrm{Pr}^G} \int_T^{+\infty} \int_{-\infty}^{+\infty} Q\left(\frac{\tilde{y}}{\sigma_Z}\right) p(\tilde{y}|\tilde{x}) p(\tilde{x}) d\tilde{y} d\tilde{x}, \tag{33}$$

$$P^B_{b|\tilde{y}} = \frac{2}{\mathrm{Pr}^B} \int_0^{T} \int_{-\infty}^{+\infty} Q\left(\frac{\tilde{y}}{\sigma_Z}\right) p(\tilde{y}|\tilde{x}) p(\tilde{x}) d\tilde{y} d\tilde{x}. \tag{34}$$

The corresponding identification rates are:

$$R^G_{id|\tilde{y}} = \mathrm{Pr}^G \left(1 - H_2\left(P^G_{b|\tilde{y}}\right)\right), \tag{35}$$

$$R^B_{id|\tilde{y}} = \mathrm{Pr}^B \left(1 - H_2\left(P^B_{b|\tilde{y}}\right)\right), \tag{36}$$

and the total rate is:

$$R^{2\mathrm{Ch}}_{id|\tilde{y}} = R^G_{id|\tilde{y}} + R^B_{id|\tilde{y}}. \tag{37}$$

Remark 9. The total cross-over probability P_b remains the same as for the hard fingerprinting, if no CSI is used:

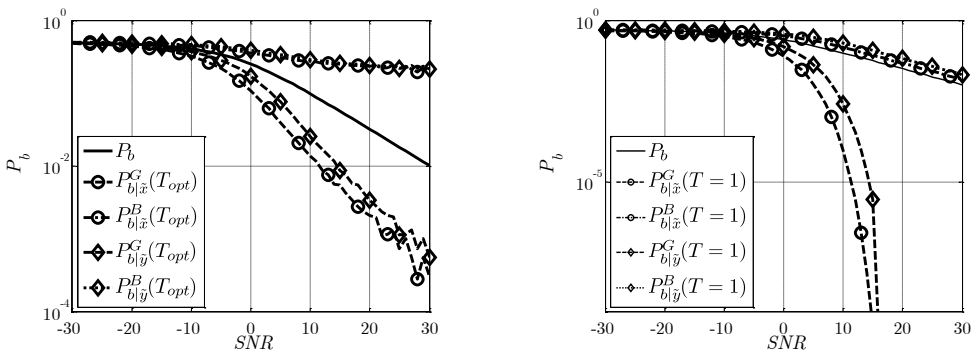
$$P_b = 2 \int_0^{+\infty} Q\left(\frac{\tilde{x}}{\sigma_Z}\right) p(\tilde{x}) d\tilde{x} = \mathrm{Pr}^G P^G_{b|\tilde{x}} + \mathrm{Pr}^B P^B_{b|\tilde{x}}. \tag{38}$$

The channel splitting by the selection of threshold T can be performed according to the several strategies:

- **Strategy 1**: maximize the total rate $R^{2\mathrm{Ch}}_{id|\tilde{x}}$ or $R^{2\mathrm{Ch}}_{id|\tilde{y}}$ to approach upper theoretical limits $R_{id|\tilde{x}}$ or $R_{id|\tilde{y}}$, respectively, that gives optimal values of thresholds $T_{\mathrm{opt}|\tilde{x}}$ and $T_{\mathrm{opt}|\tilde{y}}$ for each SNR;
- **Strategy 2**: minimize probabilities $P^G_{b|\tilde{x}}$ or $P^G_{b|\tilde{y}}$ for search complexity reasons[9] or privacy amplification considered in this paper that creates a sort of *channel polarization* after some SNR when certain number of bits can be communicated errorless.

According to strategy 1, the binary channel splitting approaches theoretical performance limits under the optimal threshold selection as shown in Fig. 6a. The remaining gap is easily compensated by more accurate models using more channels in splitting. The optimal thresholds $T_{\mathrm{opt}|\tilde{x}}$ and $T_{\mathrm{opt}|\tilde{y}}$ are shown in Fig. 6b. The individual rates $R^G_{id|\tilde{y}}$ and $R^B_{id|\tilde{y}}$ obtained for optimal $T_{\mathrm{opt}|\tilde{y}}$ with respect to the total rates are shown in Fig. 6c.

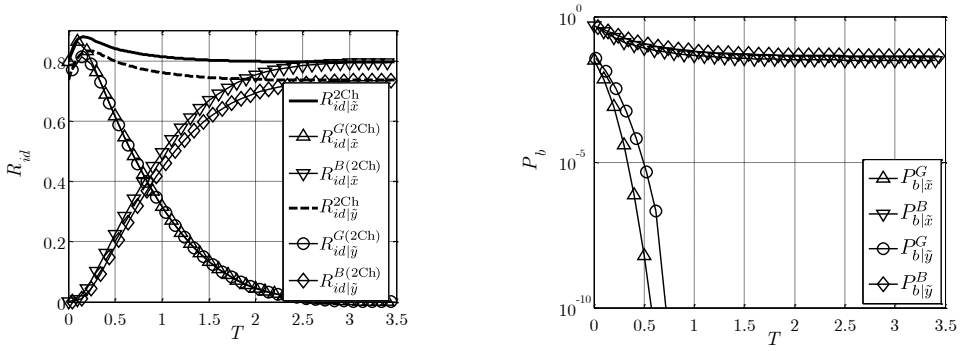
To exemplify strategy 2 we show in Fig. 7a the pairs of cross-over probabilities for good and bad channels under perfect and partial CSI that resulted from the strategy 1. Fig. 7b shows the same pairs for the fixed threshold T, where one can clearly observe the significant reduction of $P^G_{b|\tilde{x}}$ or $P^G_{b|\tilde{y}}$ that asymptotically goes to zero for SNR$>$ 15 dB.



(a) (b)

Figure 7: Probabilities of bit errors for: (a) T_{opt} and (b) $T = 1$.

Obviously, this strategy is not optimal in terms of total rate maximization. However, the expected loss while approaching the total rate with the minimization of $P_{b|\tilde{y}}^G$ is relatively small in comparison to the factor of decrease of $P_{b|\tilde{y}}^G$. The dependences of achievable rates and cross-over probabilities on the threshold T are shown in Fig. 8a and Fig. 8b for SNR= 20 dB, respectively. The rate of convergence of $P_{b|\tilde{y}}^G$ to zero is exponential with respect to the loss in the identification rate. The plots also clearly demonstrate the polarization effect when almost all useful rate is concentrated in the good channel and the bad channel can be completely disregarded from the data transmission/storage point of view.



(a) (b)

Figure 8: Achievable rates and cross-over probabilities for channel splitting at SNR=20dB.

5. PRIVACY PROTECTION BASED ON DATA HIDING AND CHANNEL SPLITTING

In this section, we combine the results of privacy protection based on data hiding with the practical framework of channel splitting allowing establishing the shared secret and trading-off the cross-over probabilities and identification rates in the good and bad channels.

Proposition 6. (*privacy protection based on data hiding and 2 channel splitting*): Redefining the variables according to the 2 channel splitting model, i.e., $\lambda_1 = \lambda^G$, $\lambda_2 = \lambda^B$ and $P_{b|s} = P_b^G$, and denoting $P_{b_e}^G = P_b^G * \lambda^G$ and $P_{b_e}^B = P_b^B * \lambda^B$ with the indices for the corresponding $\tilde{x}$ and $\tilde{y}$ types of CSI, the identification rate-privacy leak pair for the perfect CSI yields:

$$R_{id|\tilde{x}}^{2\ \mathrm{Ch}} = \Pr^G \left(1 - H_2\left(P_{b_e|\tilde{x}}^G\right)\right) + \Pr^B \left(1 - H_2\left(P_{b_e|\tilde{x}}^B\right)\right), \tag{39}$$

$$L_p^{2\ \mathrm{Ch}} = 1 - H_2\left(\bar{\lambda}_{\tilde{x}}\right), \tag{40}$$

where $\bar{\lambda}_{\tilde{x}} = \Pr^G \lambda_{\tilde{x}}^G + \Pr^B \lambda_{\tilde{x}}^B$, this model is schematically shown in Fig. 9a, and for the partial CSI:

$$R_{id|\tilde{y}}^{2\ \mathrm{Ch}} = \Pr^G \left(1 - H_2\left(P_{b_e|\tilde{y}}^G\right)\right) + \Pr^B \left(1 - H_2\left(P_{b_e|\tilde{y}}^B\right)\right), \tag{41}$$

$$L_p^{2\ \mathrm{Ch}} = 1 - H_2\left(\bar{\lambda}_{\tilde{y}}\right), \tag{42}$$

where $\bar{\lambda}_{\tilde{y}} = \Pr^G \lambda_{\tilde{y}}^G + \Pr^B \lambda_{\tilde{y}}^B$.

Remark 10. One possible strategy for trading-off the identification rate-privacy leak is based on the constraint optimization similar to the distortion allocation in the rate-distortion theory that consists in fixing the desired L_p resulting in $\bar{\lambda}$ and maximizing R_{id} by allocating different λ^G and λ^B distortions to each channel.

Remark 11. Selecting the parameters $\lambda^G = 0$ and $\lambda^B = 0.5$, the results coincide with proposition 5.

If one assumes the perfect CSI $|\tilde{x}|$ is shared between the encoder and decoder via a public channel, it unavoidably leads to information leakage about $\tilde{x}$. To overcome this shortcoming, we propose to use only partial CSI $|\tilde{y}|$ extracted directly for the noisy observation $\tilde{y}$ thus avoiding any additional storage of private data in the public domain. Unfortunately, this leads to a certain rate loss in $R_{id|\tilde{y}}$ with respect to $R_{id|\tilde{x}}$ and mismatch in the shared secret used for privacy amplification, which can be reduced based on *buffering*.

Proposition 7. *(buffering):* Assuming the encoder has access to the side information based on $|\tilde{x}|$ and the decoder to the noisy counterpart $|\tilde{y}|$, the buffering is equivalent to the 3-channel splitting. In this case, one chooses L the most reliable bits representing the actual fingerprint communicated via the BSC with $P_{b|\tilde{y}}^G$ (good channel), which appears with the probability $\Pr^G = \frac{L}{J}$ and stored with the factor $\lambda_{\tilde{y}}^G$ in the database. Then, $B - L$ bits represent the buffer, where B bits are assigned to the buffer and reliable components. The buffer bits are communicated via the BSC with $P_{b|\tilde{y}}^{BF}$ (buffer channel), which appears with the probability $\Pr^{BF} = \frac{B-L}{J}$, and stored with the factor $\lambda_{\tilde{y}}^{BF}$ in the database. The remaining $J - B$ bits are communicated via the third BSC with $P_{b|\tilde{y}}^B$ (bad channel), which appears with the probability $\Pr^B = \frac{J-B}{J}$, that are stored with the factor $\lambda_{\tilde{y}}^B$. Let $P_{b_e|\tilde{x}}^G = P_{b|\tilde{y}}^G * \lambda_{\tilde{y}}^G$, $P_{b_e|\tilde{x}}^{BF} = P_{b|\tilde{y}}^{BF} * \lambda_{\tilde{y}}^{BF}$ and $P_{b_e|\tilde{x}}^B = P_{b|\tilde{y}}^B * \lambda_{\tilde{y}}^B$, the identification rate-privacy leak pair yields:

$$\begin{aligned} R_{id|\tilde{y}}^{2\ \mathrm{Ch\ +\ BF}} &= \Pr^G \left(1 - H_2\left(P_{b_e|\tilde{y}}^G\right)\right) + \Pr^{BF} \left(1 - H_2\left(P_{b_e|\tilde{y}}^{BF}\right)\right) + \Pr^B \left(1 - H_2\left(P_{b_e|\tilde{y}}^B\right)\right), \\ L_p^{2\ \mathrm{Ch\ +\ BF}} &= 1 - H_2\left(\bar{\lambda}_{\tilde{y}}^{BF}\right), \end{aligned} \tag{43}$$

with $\bar{\lambda}_{\tilde{y}}^{BF} = \Pr^G \lambda_{\tilde{y}}^G + \Pr^{BF} \lambda_{\tilde{y}}^{BF} + \Pr^B \lambda_{\tilde{y}}^B$. This model is schematically shown in Fig. 9b.

Remark 12. One practical strategy to the selection of privacy parameters consists in selection $\lambda_{\tilde{y}}^G = \lambda_{\tilde{y}}^{BF} = 0$ and $\lambda_{\tilde{y}}^B = 0.5$ that yields:

$$R_{id|\tilde{y}}^{2\ \mathrm{Ch\ +\ BF}} = \Pr^G \left(1 - H_2\left(P_{b_e|\tilde{y}}^G\right)\right) + \Pr^{BF} \left(1 - H_2\left(P_{b_e|\tilde{y}}^{BF}\right)\right), \tag{44}$$

$$L_p^{2\ \mathrm{Ch\ +\ BF}} = 1 - H_2\left(0.5\Pr^B\right). \tag{45}$$

6. RESULTS OF COMPUTER SIMULATION

In this section, we present the results of computer simulations for the considered identification rate-privacy leak formulation under the Gaussian observation and binarized setup. All results are obtained for 10000 noise realizations and 100 input vectors.

To demonstrate the relationship between different considered privacy preserving strategies, we have selected SNR = 10 dB and SNR = 20 dB that corresponds to $P_b = 0.1$ and $P_b = 0.02$ and performed the simulation for the cases of non-adaptive privacy amplification (17)-(18), 2-channel splitting model with the perfect CSI

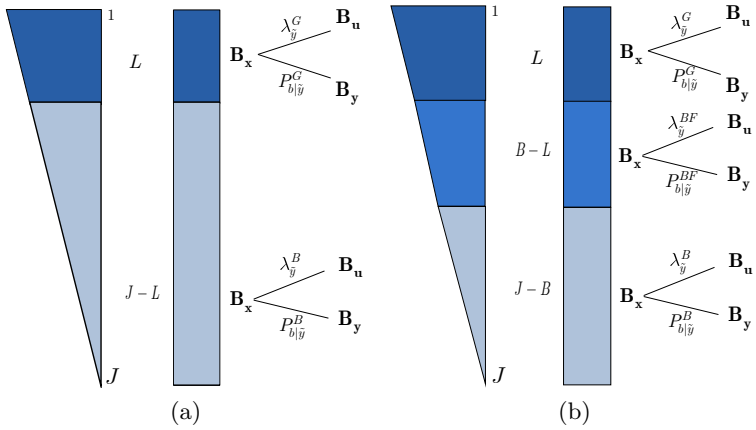


Figure 9: Channel splitting model: (a) 2-channel and (b) 3 channel splitting with buffering.

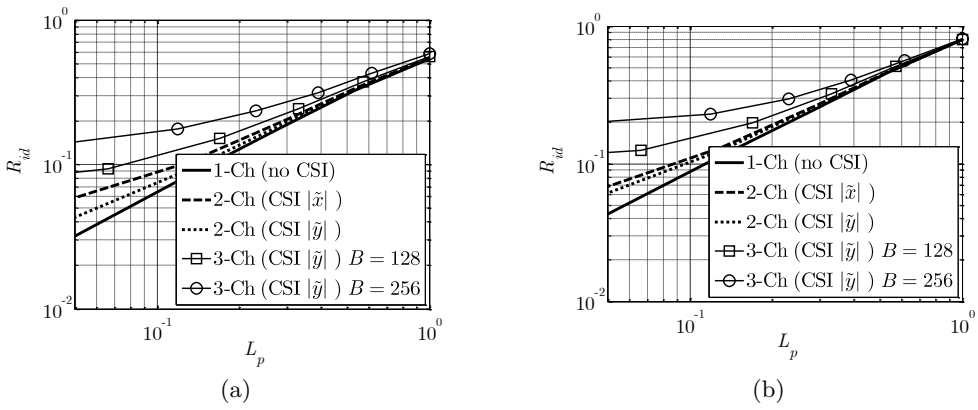


Figure 10: Identification rate-privacy leak trade-off for $J = 1024$ and $L = 32$ for different channel splitting models: no splitting (1-channel model), 2-channel splitting and different side information and 3-channel splitting based on buffering: (a) SNR=10dB and (b) SNR=20dB.

(39)-(40) and partial CSI (41)-(42), 3-channel model (44)-(45) based on the partial CSI and buffer lengths of $B = 128$ and $B = 256$. The results of modeling are shown in Fig. 10.

The presence of shared secret clearly enhances the identification rate-privacy trade-off for all cases. The impact of side information mismatch in a simple 2-channel splitting model degrades the performance with respect to the perfect side information in the region of small privacy-leak rates. The increase of the amount of channels in the 3-channel splitting model leads to better approximation of real channel model (27). The results obtained even for the partial side information clearly indicate an increase in performance that is especially important for the region of small privacy leak rates.

7. CONCLUSIONS

We considered the privacy amplification mechanism based on the bit reliability. Several techniques are analyzed for the case of perfect and imperfect side information shared between the encoder and decoder. In particular, we established that one can achieve considerable privacy amplification using even imperfect side information without the identification rate loss. We demonstrated that the privacy amplification can be solved without any publicly stored information about reliable bits contrary to the state-of-the-art methods.

8. ACKNOWLEDGMENTS

This work is supported by SNF projects 1119770 and 132337.

REFERENCES

1. Tuyls, P., Skoric, B., and (Eds.), T. K., [*Security with Noisy Data: On Private Biometrics, Secure Key Storage and Anti-Counterfeiting*], Springer (2007).
2. Voloshynovskiy, S., Koval, O., Beekhof, F., Farhadzadeh, F., and Holotyak, T., "Information-theoretical analysis of private content identification," in [*IEEE Information Theory Workshop, ITW2010*], (Aug.30-Sep.3 2010).
3. Voloshynovskiy, S., Koval, O., Holotyak, T., Beekhof, F., and Farhadzadeh, F., "Privacy amplification of content identification based on fingerprint bit reliability," in [*Proceedings of IEEE International Workshop on Information Forensics and Security*], (December 12–15 2010).
4. Ignatenko, T. and Willems, F., "Privacy leakage in biometric secrecy systems," in [*46th Annual Allerton Conference on Communication, Control, and Computing*], 850–857 (23-26 Sept. 2008).
5. Beekhof, F., Voloshynovskiy, S., Koval, O., and Holotyak, T., "Fast identification algorithms for forensic applications," in [*Proceedings of IEEE International Workshop on Information Forensics and Security*], (December 6–9 2009).
6. Martinian, E., Yekhanin, S., and Yedidia, J., "Secure biometrics via syndromes," in [*43rd Annual Allerton Conference on Communications, Control, and Computing*], (October 2005).
7. Sutcu, Y., Li, Q., and Memon, N., "Protecting biometric templates with sketch: Theory and practice," *IEEE Transactions on Information Forensics and Security* **2**, 503–512 (2007).
8. Voloshynovskiy, S., Beekhof, F., Koval, O., and Holotyak, T., "On privacy preserving search in large scale distributed systems: a signal processing view on searchable encryption," in [*Proceedings of the International Workshop on Signal Processing in the EncryptEd Domain*], (2009).
9. Holotyak, T., Voloshynovskiy, S., Beekhof, F., and Koval, O., "Fast identification of highly distorted images," in [*Proceedings of SPIE Photonics West, Electronic Imaging 2010 / Media Forensics and Security XII*], (January 21–24 2010).