

the secure embedding scheme (see Section II), it is hard to generate a watermarked version of c which has a low correlation with k' . An estimation attack usually yields a watermarked object that still correlates well with k' ; a judge will reject the accusation on such an object, as it can only originate from a malicious seller (k' is only available to the seller).

- Finally, S can attempt to cheat in step 6 by submitting the customer a wrongly encrypted watermark $E_K(n||w \oplus k)$. However, this is detected by the client in step 7 by checking the integrity of the transaction number contained therein.

IV. CONCLUSION

In this correspondence, we proposed a buyer–seller protocol that utilizes the concepts of secure watermark embedding. In contrast to the known solutions, which use homomorphic public-key encryption on the content and impose unpractical constraints on computational resources and transmission bandwidth, our protocol is efficient due to the use of recent secure embedding algorithms.

REFERENCES

- [1] N. Memon and P. Wang, "A buyer-seller watermarking protocol," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 643–649, Apr. 2001.
- [2] H. Ju, H.-Y. Kim, D. Lee, and J. Lim, "An anonymous buyer-seller watermarking protocol with anonymity control," in *Information Security and Cryptology*, ser. Lect. Notes Comput. Sci. Berlin, Germany: Springer, 2002, vol. 2587, pp. 421–432.
- [3] J.-J. Choi, K. Sakurai, and J.-H. Park, "Does it need trusted third party? Design of buyer-seller watermarking protocol without trusted third party," in *Proc. 1st Int. Conf. Applied Cryptography and Network Security*, ser. Lect. Notes Comput. Sci. Berlin, Germany: Springer, 2003, vol. 2846, pp. 265–279.
- [4] J.-G. Choi and J.-H. Park, "A generalization of an anonymous buyer-seller watermarking protocol and its application to mobile communications," in *Proc. 3rd Int. Workshop Digital Watermarking*, ser. Lect. Notes Comput. Sci. Berlin, Germany: Springer, 2004, vol. 3304, pp. 232–243.
- [5] C. Lei, P. Yu, P. Tsai, and M. Chan, "An efficient and anonymous buyer-seller watermarking protocol," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 643–649, Apr. 2004.
- [6] M. Kuribayashi and H. Tanaka, "Fingerprinting protocol for images based on additive homomorphic property," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2129–2139, Dec. 2005.
- [7] F. Frattolillo, "Watermarking protocol for web context," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 350–363, Sep. 2007.
- [8] D. Kundur, "Video fingerprinting and encryption principles for digital rights management," *Proc. IEEE*, vol. 92, no. 6, pp. 918–932, Jun. 2004.
- [9] A. Adelsbach, U. Huber, and A.-R. Sadeghi, "Finger casting—joint fingerprinting and decryption of broadcast messages," in *Proc. 11th Australasian Conf. Information Security Privacy*, ser. Lect. Notes Comput. Sci. Berlin, Germany: Springer, 2006, vol. 4058, pp. 136–147.
- [10] A. Lemma, S. Katzenbeisser, M. Celik, and M. van der Veen, "Secure watermark embedding through partial encryption," in *Proc. 5th Int. Workshop Digital Watermarking*, ser. Lect. Notes Comput. Sci. Berlin, Germany: Springer, 2006, vol. 4283, pp. 433–445.
- [11] M. Celik, A. Lemma, S. Katzenbeisser, and M. van der Veen, "Secure embedding of spread spectrum watermarks using look-up-tables," in *Proc. IEEE Int. Conf. Acoustics, Speech Signal Processing*, 2007, pp. 153–156.
- [12] S. Emmanuel and M. Kankanhalli, "Copyright protection for MPEG-2 compressed broadcast video," in *Proc. IEEE Int. Conf. Multimedia Expo.*, 2001, pp. 206–209.
- [13] D. Kirovski, H. Malvar, and Y. Yacobi, "A dual watermark-fingerprint system," *IEEE Multimedia*, vol. 11, no. 3, pp. 59–73, Jul.–Sep. 2004, (also see U.S. patent application 2005/0 660 550).

Cryptographic Secrecy of Steganographic Matrix Embedding

Phillip A. Regalia

Abstract—Some recent information-hiding schemes are scrutinized in terms of their cryptographic secrecy. The schemes under study appeal to the so-called matrix embedding strategy, designed to optimize embedding capacity under distortion constraints, as opposed to any cryptographic measure. Nonetheless, we establish conditions under which a key equivocation function is optimal, and show that under reasonable key generation models, a perfect secrecy property is nearly satisfied, limited by a mutual information measure that decreases exponentially with the block length.

Index Terms—Key equivocation, message equivocation, perfect secrecy, steganography, wet paper coding.

I. INTRODUCTION

INFORMATION embedding encompasses watermarking and steganography, in which the former places emphasis on robustness of the embedded payload to a subsequent attack, while the latter aims instead for the presence of a payload to go undetected. Many of the techniques of information embedding have analogues in multiuser communications, particularly spread-spectrum techniques and more recently, the revival of dirty paper coding [1] and similarly inspired strategies [2]–[4]. Such techniques have independently met with information-theoretic analyses [5]–[7] that prove an important optimality property: The embedding capacity attains the theoretical maximum subject to distortion constraints and robustness to attack.

While such results solve the capacity issue from a communications perspective, they do not directly address secrecy issues which are fundamental from cryptographic considerations [8] that underlie steganography. Owing to this heritage, various information-theoretic security frameworks have been advanced in the context of steganography and watermarking. Perhaps the earliest is Mittelholzer [9], who proposed an information-theoretic framework for distortion constraints, detectability, and robustness to attack; some of these problems have since met with solutions in [5]–[7], [10]. Further developments by Cachin [11] emphasize active versus passive warden models, distinguish perfect secrecy from perfect security and, interestingly, lead to an early version of binning codes [11, Theor. 2], now known to underlie optimal embedding strategies [6], [12]. A different approach is taken in [13], using an oracle-based scheme rather than one relying on ensemble averages, such as entropy or mutual information. Parallel developments addressing watermarking security include [14] and [15] whose results on addressing message obfuscation (via "perfect covering" as a proposed counterpart to Shannon's perfect secrecy condition [8]) have some relevance to steganography, in the sense that satisfaction of certain criteria ensure that no information on the embedded message is leaked from the watermarked/stego signal. The

Manuscript received January 16, 2008; revised July 15, 2008. Current version published November 19, 2008. This work was supported by the National Science Foundation under Grant CCF 0634757. Parts of this work were presented at the International Conference Acoustics, Speech, and Signal Processing (ICASSP-2008), Las Vegas, NV, April 2008. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Fernando Perez-Gonzalez.

The author is with the Department of Electrical Engineering and Computer Science, Catholic University of America, Washington, DC 20064 USA, and also with the Institut Telecom, Department CITI, Evry 91011, France (e-mail: regalia@cua.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIFS.2008.2002940

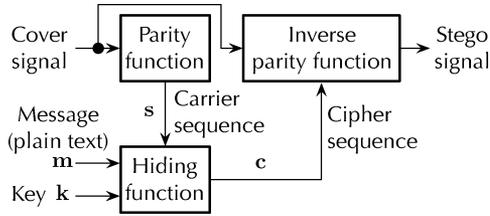


Fig. 1. Information hiding setup with parity-function data.

more basic question of hypothesis testing for steganography, in which a warden tests the hypothesis of whether an intercepted message contains a hidden message, is treated in [16]–[21].

This paper focuses on traditional secrecy aspects with message and key equivocation functions, for a class of matrix embedding algorithms. These are reviewed in Section II and, as noted before, were designed from the viewpoint of optimizing payload capacity as opposed to any cryptographic consideration. Nonetheless, the schemes under study are shown to exhibit unexpectedly good cryptographic secrecy due, in essence, to the additional randomness injected by the cover signal. This may remove the need for further message obfuscation stages in some applications.

Section II reviews the hiding schemes under study to present matrix embedding and wet paper coding in a common framework. Our main results detailing the (near) optimality of message equivocation and key equivocation functions are collected in Section III, with concluding remarks in Section IV.

II. PROBLEM SETUP

We begin with the basic setup of Fig. 1. A cover signal (image, audio, video, etc.) generates a binary carrier sequence by way of a parity function that maps quasicontinuous amplitude samples to zeros and ones, and is assumed publicly known [2], [22]. The carrier sequence is then modified to embed a given message (or plain text), producing a cipher text signal. (The terms “plain text” and “cipher text” identify the role such signals would play in standard cryptography; the hiding function would correspond to an “encryption” function, although we avoid that term since the hiding functions that will be reviewed were not professed from an encryption standpoint.) The original cover signal is then modified into the stego signal whose parity function output agrees with the cipher signal [2], [22], as indicated by the inverse parity function block. The sender and receiver agree on a private key.

Random variables are denoted by uppercase italic letters, with lowercase bold letters denoting particular realizations. Thus, S denotes the carrier sequence from the parity function (with s a particular realization), M denotes the message to embed, K denotes the key, and C denotes the cipher signal produced by the hiding function. We treat all signals as vectors of bits (each 0 or 1), using componentwise modulo-2 addition over the Galois field \mathcal{F}_2 . In particular:

- the carrier sequence contains n bits, derived from n parity checks that comprise the parity function, so that $S \in \mathcal{F}_2^n$; we verify below that a “good” parity function should render all 2^n realizations of S equally probable;
- the (plain text) message M collects q bits ($M \in \mathcal{F}_2^q$) with $q < n$; as in traditional cryptography, ideally all 2^q configurations of M would be equally probable, which becomes feasible through compression (e.g., [23]);
- the key $K \in \mathcal{F}_2^{q \times n}$ is a $q \times n$ parity check matrix. For a particular realization \mathbf{k} of this matrix and a particular message \mathbf{m} , the chosen cipher signal $\mathbf{c} \in \mathcal{F}_2^n$ satisfies

$$\mathbf{m} \equiv \mathbf{k}\mathbf{c} \pmod{2}$$

among other constraints that will be detailed. Thus, if the receiver knows the key \mathbf{k} , the hidden message \mathbf{m} can be recovered from the cipher text \mathbf{c} .

Example 1: The simplest example of a parity function is one which extracts the least-significant bit plane(s) of an image (e.g., [24]–[26]), although modifying the least-significant bit is vulnerable to detection [16]–[18], [20]. A more advanced version would calculate each bit in the carrier sequence as the parity (exclusive-OR) of a number of bits drawn from spatially separated pixels [2], [22], which is more likely to yield a high entropy sequence. Since the carrier and cipher sequences are binary, the inverse parity function is realized by flipping a single bit in the cover signal for each bit in the carrier sequence which changes when forming the cipher sequence; the choice of which bit to flip may be randomized. \diamond

Entropy is denoted by $H(\cdot)$ and mutual information by $I(\cdot; \cdot)$ as in

$$H(S) = - \sum_{\mathbf{s} \in \mathcal{F}_2^n} \Pr(\mathbf{s}) \log_2 \Pr(\mathbf{s})$$

$$I(M; C) = \sum_{\mathbf{m} \in \mathcal{F}_2^q} \sum_{\mathbf{c} \in \mathcal{F}_2^n} \Pr(\mathbf{m}, \mathbf{c}) \log_2 \frac{\Pr(\mathbf{m}, \mathbf{c})}{\Pr(\mathbf{m}) \Pr(\mathbf{c})}$$

For a given key \mathbf{k} , assumed of full rank q , the set of binary vectors $\mathbf{b} \in \mathcal{F}_2^n$ that lie in its null space defines a code of rate $r = 1 - (q/n)$, denoted $G_{\mathbf{k}}(\mathbf{0})$

$$G_{\mathbf{k}}(\mathbf{0}) = \{\mathbf{b} \in \mathcal{F}_2^n : \mathbf{0} \equiv \mathbf{k}\mathbf{b} \pmod{2}\}$$

The set of binary vectors which produce instead a given “syndrome” \mathbf{m} defines a coset [12] $G_{\mathbf{k}}(\mathbf{m})$ for that syndrome

$$G_{\mathbf{k}}(\mathbf{m}) = \{\mathbf{b} \in \mathcal{F}_2^n : \mathbf{m} \equiv \mathbf{k}\mathbf{b} \pmod{2}\}$$

The member of $G_{\mathbf{k}}(\mathbf{m})$ of the lowest Hamming weight (smallest number of 1 s) is the coset leader for the syndrome \mathbf{m} .¹

A. Matrix Embedding

We consider the transcription of the optimal information embedding techniques of [5]–[7] to the binary case, which is closely allied with the matrix embedding approach of [4]. This method seeks a cipher signal \mathbf{c} minimizing the Hamming distance $d(\mathbf{s}, \mathbf{c})$ subject to the constraint $\mathbf{m} \equiv \mathbf{k}\mathbf{c} \pmod{2}$. The amounts to “quantizing” the carrier sequence \mathbf{s} to the coset $G_{\mathbf{k}}(\mathbf{m})$. If \mathbf{e} denotes the coset leader for the syndrome $\mathbf{m} - \mathbf{k}\mathbf{s} \pmod{2}$, then $\mathbf{c} \equiv \mathbf{s} + \mathbf{e} \pmod{2}$ is the closest member of $G_{\mathbf{k}}(\mathbf{m})$ to \mathbf{s} in the Hamming distance. For any \mathbf{m} , define the average distortion (per bit) as

$$D = \frac{1}{n} \sum_{\mathbf{s} \in \mathcal{F}_2^n} \Pr(\mathbf{s}) \min_{\mathbf{c} \in G_{\mathbf{k}}(\mathbf{m})} d(\mathbf{s}, \mathbf{c}).$$

Since each \mathbf{s} is quantized to a code of rate $r = 1 - (q/n)$, the average distortion is lower bounded through the rate-distortion function [27, Theor. 10.3.1] as

$$H_2(D) \geq H'(S) - r = 1 - r = \frac{q}{n}$$

in which $H_2(D) = -D \log_2 D - (1 - D) \log_2 (1 - D)$ is the binary entropy function, and $H'(S)$ is the per-bit entropy rate of the carrier sequence S [with $H'(S) = H(S)/n$ for large enough n]; note that $H'(S) = 1$ if S is uniformly distributed, as a “good” parity function would ideally provide. The lower bound on D (where $H_2(D) = 1 - r$) is achieved if the (error correction) code $G_{\mathbf{k}}(\mathbf{0})$ achieves channel capacity over a binary symmetric channel with error probability D (e.g., [27] and [28]). Design methods for such codes [29], [30] may thus

¹In case of nonuniqueness, a particular lowest-weight member is arbitrarily selected as the unique coset leader.

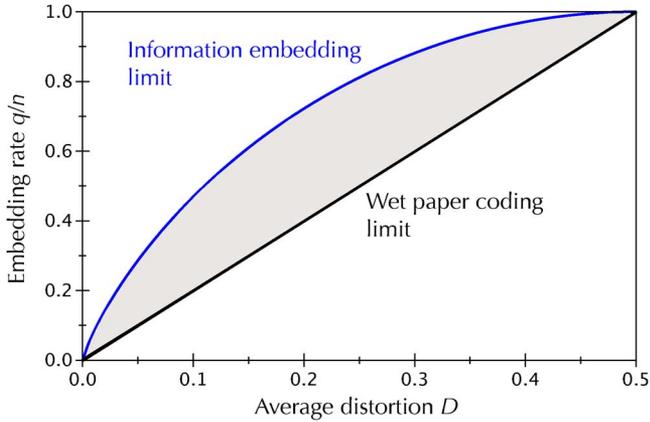


Fig. 2. Achievable embedding rate versus distortion as the shaded area, delineated by the information embedding and wet paper coding curves.

be used to generate “low distortion” keys. The “covering code” definition from [31] may also be recognized as formulating an optimal source code for a uniform source via the Gilbert–Varshamov bound; the parity-check matrix of any such “covering code” would thus also furnish a suitable key.

B. Wet Paper Encoding

Wet paper coding [2] likewise produces a cipher sequence \mathbf{c} fulfilling $\mathbf{m} \equiv \mathbf{k}\mathbf{c} \pmod{2}$, but with different constraints. An index set—call it \mathbf{t} —collects q integers from $\{1, 2, \dots, n\}$ and permits \mathbf{c} to differ from the carrier sequence \mathbf{s} only in the positions comprising \mathbf{t} (with the remaining positions “locked” [32]). This index set gives an “information set” [33] provided a $q \times q$ submatrix of \mathbf{k} —built by retaining columns whose indices are in \mathbf{t} —has full rank. This holds with high probability [2], and a unique solution for \mathbf{c} results. The index set is randomly selected, and need not be known to the receiver, who can still recover the message as $\mathbf{m} \equiv \mathbf{k}\mathbf{c} \pmod{2}$. The idea is to randomize the modification positions in order to better evade detection by a warden or eavesdropper [2], [22]. We thus introduce T (a “tool” which complements the key) as a random variable comprising the indices used in the hiding stage, with \mathbf{t} denoting a particular outcome. (This fulfills the role of a private random source in the formulation of [19]). The average distortion from this method is $D = 0.5q/n$ [2], with q/n being the embedding rate. This is larger than the average distortion attainable using the construction of Section II-A, as plotted in Fig. 2.

An intermediary between the two curves of Fig. 2 is obtained by allowing \mathbf{c} and \mathbf{s} to differ in l positions, with $q \leq l \leq n$, giving 2^{l-q} possibilities for \mathbf{c} and, thus, generally lower distortion [3] than the formulation of [2]. Specifically, let \mathbf{t} now contain l indices, and let $\mathbf{k}' \in \mathcal{F}_2^{q \times l}$ be the $q \times l$ matrix formed from the columns of \mathbf{k} whose indices are contained in the set \mathbf{t} and, likewise, let $\mathbf{s}' \in \mathcal{F}_2^l$ be the vector formed from the entries of \mathbf{s} , whose indices are in \mathbf{t} . Choosing \mathbf{e}' as the coset leader of $G_{\mathbf{k}'}(\mathbf{m} - \mathbf{k}'\mathbf{s}')$ (with respect to \mathbf{k}') for the syndrome $\mathbf{m} - \mathbf{k}'\mathbf{s}' \pmod{2}$, the cipher sequence becomes $\mathbf{c} \equiv \mathbf{s} + \mathbf{e}' \pmod{2}$, with \mathbf{e}' being a zero-padded version of \mathbf{e}' , obtained by inserting zeros into the “locked” positions. The per-bit Hamming distortion measure becomes

$$D = \frac{1}{n} \sum_{\mathbf{s}' \in \mathcal{F}_2^l} \Pr(\mathbf{s}') d(\mathbf{s}', \mathbf{c}') = \frac{1}{l} \sum_{\mathbf{s}' \in \mathcal{F}_2^l} \Pr(\mathbf{s}') d(\mathbf{s}, \mathbf{c})$$

if \mathbf{s} is uniformly distributed, since $d(\mathbf{s}', \mathbf{c}') = d(\mathbf{s}, \mathbf{c})$ as \mathbf{s} and \mathbf{c} differ only in the positions comprising the index set \mathbf{t} . Akin to the optimal embedding construct of the previous subsection, the distortion is then

lower bounded by the inequality $H_2(nD/l) \geq q/l$. At the upper extreme $l = n$, the formulation reverts to the construct of Section II-A, while the lower extreme $l = q$ gives the original wet paper coding construct of [2]. The results of Section III will apply for all values of l between these extremes, depicted by the shaded zone of Fig. 2, allowing the optimal embedding and wet paper constructs to be treated in a common framework.

C. Watermarking

Watermarking generally seeks to allow the message M to be recovered even if the cipher text C suffers further distortion. The maximum embedding capacity q/n subject to embedding distortion and robustness to attack is solved in [5]–[7], and given for the binary case as the convex envelope of $H_2(D) - H_2(p)$, in which the attack channel is modeled as a binary symmetric channel with crossover probability p . A practical construct to achieve this embedding capacity results by partitioning the key \mathbf{k} row-wise and choosing the closest \mathbf{c} to \mathbf{s} that satisfies

$$q \left\{ \begin{bmatrix} \mathbf{m} \\ \mathbf{0} \end{bmatrix} \right\} \equiv \underbrace{\begin{bmatrix} \mathbf{k}_1 \\ \mathbf{k}_2 \end{bmatrix}}_{\mathbf{k}} \mathbf{c} \pmod{2}.$$

Here, the null space of \mathbf{k}_2 gives a “good” error correction code (specifically, a capacity-approaching code for a binary symmetric channel with crossover probability p , thus imposing $q'/n \geq H_2(p)$), and that of \mathbf{k} a “good” quantization code (specifically, attaining the rate-distortion bound for compressing a uniform source to rate $1 - (q + q')/n$) [6], [12]. Note that the nullspace of \mathbf{k} is contained in that of \mathbf{k}_2 , giving a nested structure. If a sufficient number of cipher texts \mathbf{c} are observed for the same key \mathbf{k} , the linear subspace they span builds the orthogonal complement to \mathbf{k}_2 . This reveals information on the key \mathbf{k} and, more seriously from a steganographic viewpoint, alerts an observer that \mathbf{c} may contain a hidden message.

We should note that the corresponding construction using instead a mean-square distortion measure, Gaussian signals, and the nested lattice codes of [12], has had various security holes exposed in [34] and [35]. It is thus not surprising that the binary transcription considered here should likewise suffer security weaknesses. This reflects how secrecy properties of optimal (with respect to payload capacity) watermarking prove deficient. This is best viewed as yet another instance of how watermarking and steganography can differ, despite considerable overlap in their formulations; further details on watermarking security and secrecy are surveyed in [14] and [15].

III. CRYPTOGRAPHIC SECRECY MEASURES

Cryptographic secrecy will be assessed here using two traditional measures. We treat first the key equivocation function [36] $I(K; C) = H(K) - H(K|C)$ which measures how much information may be revealed about the key K from observations of the cipher text C , and then study the message equivocation function $I(M; C)$ underlying the perfect secrecy [8], [36], [37] condition.

Lemma 1: The key equivocation function is given by

$$I(K; C) = H(C) - H(M) - I(T; K, C) - I(S; K, C, T).$$

This differs from a standard result [36, Theor. 2.10] by the inclusion of mutual information terms involving the carrier sequence S and tool set T , which are absent in classical cryptography. For the proof, expand the joint entropy as

$$\begin{aligned} H(C, K, M, T, S) &= H(K, M, T, S) + \underbrace{H(C|K, M, T, S)}_{=0} \\ &= H(K) + H(M) + H(T) + H(S) \end{aligned}$$

in which $H(C|K, M, T, S) = 0$ since the key, message, carrier, and tool together determine the cipher text; the key, message, carrier signal, and tool are likewise assumed to be mutually independent. By a separate expansion, we also have

$$H(C, K, M, T, S) = H(C) + H(K|C) + \underbrace{H(M|K, C)}_{=0} + H(T|C, K, M) + H(S|C, K, M, T)$$

in which $H(M|K, C) = 0$, $H(T|C, K, M) = H(T|C, K)$, and $H(S|C, K, M, T) = H(S|C, K, T)$ since the key and cipher text together determine the message. Equating the expansions and isolating $H(K) - H(K|C) = I(K; C)$ gives the lemma statement. \diamond

The following theorem gives conditions that ensure that the key is not revealed by the cipher text, and that the cipher text has maximum entropy.

Theorem 1: For matrix embedding and wet paper hiding, the key equivocation function is bounded as

$$I(K; C) \leq [q - H(M)] + [H(C) - H(S)].$$

In particular, if $H(S) = n$ (all carrier sequences equally probable) and $H(M) = q$ (all messages equally probable), then

$$I(K; C) = 0 \quad \text{and} \quad H(C) = n.$$

For the proof, insert $I(S; K, C, T) = H(S) - H(S|K, C, T)$ into the expression of lemma 1, and isolate $H(S|K, C, T)$ as

$$H(S|K, C, T) = I(K; C) + I(T; K, C) + H(M) + H(S) - H(C). \quad (1)$$

We claim now that $H(S|K, C, T) \leq q$. To verify, consider any realization $(\mathbf{k}, \mathbf{c}, \mathbf{t})$, and set $\mathbf{m} \equiv \mathbf{k}\mathbf{c} \pmod{2}$. To construct a candidate carrier sequence \mathbf{s} , we note that $\mathbf{s} - \mathbf{c} \pmod{2}$ must be a coset leader, of which there are 2^q . [For the matrix embedding technique of Section II-A, pick any $\mathbf{d} \in \mathcal{F}_2^q$ and let \mathbf{e} be the coset leader with respect to \mathbf{k} for $\mathbf{m} - \mathbf{d} \pmod{2}$, to get $\mathbf{s} \equiv \mathbf{c} + \mathbf{e} \pmod{2}$]; the same reasoning carries through for the wet paper scheme by restricting attention to the unlocked positions contained in \mathbf{t} . This limits \mathbf{s} to one of the 2^q configurations. As such, the entropy of S , given any fixed triple $(\mathbf{k}, \mathbf{c}, \mathbf{t})$ is bounded as $H(S|\mathbf{k}, \mathbf{c}, \mathbf{t}) \leq q$. By averaging over the joint probability $\Pr(\mathbf{k}, \mathbf{c}, \mathbf{t})$, the conditional entropy is likewise bounded

$$H(S|K, C, T) = \sum_{\mathbf{k}, \mathbf{c}, \mathbf{t}} \Pr(\mathbf{k}, \mathbf{c}, \mathbf{t}) H(S|\mathbf{k}, \mathbf{c}, \mathbf{t}) \leq q.$$

This gives, via (1) $I(K; C) + I(T; K, C) + H(M) + H(S) - H(C) \leq q$ or

$$I(K; C) + I(T; K, C) \leq [q - H(M)] + [H(C) - H(S)].$$

Now, if $H(M) = q$ and $H(S) = n$, then $I(K; C) + I(T; K, C) \leq H(C) - n$. But as the cipher sequence has n bits, necessarily $H(C) \leq n$, giving $I(K; C) + I(T; K, C) \leq 0$. Since mutual information is nonnegative, this must give $I(K; C) = 0$ and $I(T; K, C) = 0$ and, therefore, $H(C) = n$ and $H(S|K, C) = q$ as well. \diamond

We note in passing that the condition $I(T; K, C) = 0$ implies that knowledge of the key and cipher signal does not assist in deducing the tool set T . For the watermarking scheme, by contrast, the cipher text leaks information on the key, as noted in Section II-C, and exploited in [34] and [35].

We consider next the message equivocation function $I(M; C)$ that measures how much message information leaks through the cipher

signal C , if the key K is unknown. By rearranging two expansions of the joint entropy $H(M, C, K, T, S)$, we can show

$$\begin{aligned} I(M; C) &= H(M) - H(M|C) \\ &= H(C) - H(K) - H(T) - H(S) + H(K|M, C) \\ &\quad + H(T|M, K, C) + H(S|M, K, C, T) \\ &= q - I(K; M, C) \end{aligned}$$

using $H(C) = H(S) = n$ as well as $I(T; M, C, K) = I(T; C, K) = 0$ and $H(S|M, K, C, T) = H(S|K, C, T) = q$ from Theorem 1. Now, the nonnegativity of $I(M; C)$ implies $I(K; M, C) \leq q$. But knowledge of the message M and cipher signal C reveals q parity constraints on the key; if a full q bits of information are imparted so that $I(K; M, C) = q$, then the perfect secrecy condition [8], [36], [37] $I(M; C) = 0$ will be satisfied.²

In this direction, we may consider two models for a randomly selected key:

- 1) If \mathbf{k} is a ‘‘good’’ parity check matrix, then so is $\mathbf{P}_q \mathbf{k} \mathbf{P}_n$, where \mathbf{P}_q and \mathbf{P}_n are $q \times q$ and $n \times n$ permutation matrices, respectively. This gives up to $q!n!$ keys to choose from. (In practice, fewer distinct keys will result, since many of these permuted keys will coincide.) A key from this space may be selected with uniform probability.
- 2) The elements K_{ij} of the key may be modeled as i.i.d. Bernoulli random variables, with

$$\Pr(K_{ij} = 1) = 1 - \Pr(K_{ij} = 0) = p.$$

The value p is small, and diminishes as $1/n$, for a good low density parity check matrix:

Example 2: Families of low-density parity check matrices may be described by their degree distribution polynomials [29]

$$\lambda(z) = \sum_{i \geq 2} \lambda_i z^{i-1}, \quad \rho(z) = \sum_{i \geq 2} \rho_i z^{i-1}$$

in which λ_i (respectively, ρ_i) is the fraction of edges which emanate from a variable (respectively, check) node of degree i in a factor graph. (Stated otherwise, $\lambda_i/[i \int_0^1 \lambda(z) dz]$ is the fraction of columns of the parity check matrix having Hamming weight i , and $\rho_i/[i \int_0^1 \rho(z) dz]$ is the fraction of rows having weight i). The code rate r is determined through [29]

$$r = 1 - \frac{q}{n} = 1 - \frac{\int_0^1 \rho(z) dz}{\int_0^1 \lambda(z) dz}$$

and the number of ones in the parity-check matrix becomes

$$\frac{n}{\int_0^1 \lambda(z) dz} = \frac{q}{\int_0^1 \rho(z) dz}.$$

The density is the number of ones divided by the number of elements of the matrix, or

$$p = \frac{1}{nq} \frac{q}{\int_0^1 \rho(z) dz} = \frac{1}{n} \frac{1}{\int_0^1 \rho(z) dz}$$

which, for the fixed-degree distribution polynomials, diminishes as $1/n$. \diamond

We then claim:

²The perfect secrecy definition here concerns zero message equivocation, as used in cryptography [8], [36], [37]. The term ‘‘perfect secrecy’’ has also been used in watermarking to denote instead zero key equivocation [14], [15], as used in Theorem 1.

Theorem 2: If $H(S) = n$ and $H(M) = q$, then: 1) For the uniform permutation key model

$$I(M; C) = 2^{-n}q.$$

2) For the Bernoulli key model, with sufficiently large nq

$$I(M; C) \approx 2^{-n}nqH_2(p)$$

where $H_2(p) = -p \log_2 p - (1-p) \log_2 (1-p)$ is the binary entropy function.

For the verification, given that $I(M; C) = q - I(K; M, C)$, we examine $I(K; M, C) = H(K) - H(K|M, C)$.

Consider first the permutation model of key selection, and let $|\mathcal{K}|$ denote the number of distinct permutations of a ‘‘good’’ key. By assigning a uniform probability to each key, we have $H(K) = \log_2 |\mathcal{K}|$. Consider now fixing a message \mathbf{m} and a ciphertext \mathbf{c} , and finding the subset of keys satisfying $\mathbf{m} \equiv \mathbf{k}\mathbf{c} \pmod{2}$. If $\mathbf{c} \neq \mathbf{0}$, then for any \mathbf{m} this introduces q parity constraints, so that a fraction $1/2^q$ of the available keys will satisfy the constraints (giving cardinality $|\mathcal{K}|/2^q$). Thus, $H(K|\mathbf{m}, \mathbf{c} \neq \mathbf{0}) = \log_2 |\mathcal{K}| - q$. For the event $\mathbf{c} = \mathbf{0}$, necessarily $\mathbf{m} = \mathbf{0}$ and all keys satisfy $\mathbf{0} \equiv \mathbf{k}\mathbf{0} \pmod{2}$. Thus, $H(K|\mathbf{m} = \mathbf{0}, \mathbf{c} = \mathbf{0}) = H(K)$. By averaging over the joint probability $\Pr(\mathbf{m}, \mathbf{c})$, the conditional entropy becomes

$$\begin{aligned} H(K|M, C) &= H(K|\mathbf{m} = \mathbf{0}, \mathbf{c} = \mathbf{0}) \Pr(\mathbf{m} = \mathbf{0}, \mathbf{c} = \mathbf{0}) \\ &\quad + \sum_{\mathbf{c} \neq \mathbf{0}} H(K|\mathbf{m}, \mathbf{c}) \Pr(\mathbf{m}, \mathbf{c}) \\ &= H(K)2^{-n} + [H(K) - q](1 - 2^{-n}) \\ &= H(K) - (1 - 2^{-n})q \end{aligned}$$

in which $\Pr(\mathbf{m} = \mathbf{0}, \mathbf{c} = \mathbf{0}) = 2^{-n}$, because $H(C) = n$ implies $\Pr(\mathbf{c}) = 2^{-n}$ for each \mathbf{c} , and

$$\Pr(\mathbf{c} = \mathbf{0}) = \Pr(\mathbf{m} = \mathbf{0}, \mathbf{c} = \mathbf{0}) + \underbrace{\sum_{\mathbf{m} \neq \mathbf{0}} \Pr(\mathbf{m}, \mathbf{c} = \mathbf{0})}_{=0}$$

as $\mathbf{c} = \mathbf{0}$ implies $\mathbf{m} = \mathbf{0}$. Thus, $I(K; M, C) = H(K) - H(K|M, C) = (1 - 2^{-n})q$, giving $I(M; C) = 2^{-n}q$ for this case.

For the case in which the nq elements of the key K are i.i.d. and Bernoulli, we instead have $H(K) = nqH_2(p)$. We again expand the conditional entropy $H(K|M, C)$ as

$$H(K|M, C) = \sum_{\mathbf{m}, \mathbf{c}} H(K|\mathbf{m}, \mathbf{c}) \Pr(\mathbf{m}, \mathbf{c}).$$

As before, the case $\mathbf{c} = \mathbf{0}$ implies $\mathbf{m} = \mathbf{0}$, and all keys satisfy $\mathbf{0} = \mathbf{k}\mathbf{0}$. Thus, $H(K|\mathbf{0}, \mathbf{0}) = nqH_2(p)$, and we still have $\Pr(\mathbf{m} = \mathbf{0}, \mathbf{c} = \mathbf{0}) = 2^{-n}$, giving $H(K|\mathbf{0}, \mathbf{0}) \Pr(\mathbf{0}, \mathbf{0}) = 2^{-n}nqH_2(p)$.

For the events $\mathbf{c} \neq \mathbf{0}$, introduce the typical set of keys

$$A_{nq}^\epsilon = \{\mathbf{k} : nq(H_2(p) - \epsilon) \leq -\log_2 \Pr(\mathbf{k}) \leq nq(H_2(p) + \epsilon)\}.$$

For any fixed ϵ , the probability mass of the typical set A_{nq}^ϵ approaches 1 arbitrarily closely as nq grows and, for large nq , its cardinality satisfies [27]

$$(1 - \epsilon)2^{nq[H_2(p) - \epsilon]} \leq |A_{nq}^\epsilon| \leq 2^{nq[H_2(p) + \epsilon]}.$$

Now, for any cipher text $\mathbf{c} \neq \mathbf{0}$ and any message \mathbf{m} , denote

$$A_{nq}^\epsilon(\mathbf{m}, \mathbf{c}) = \{\mathbf{k} \in A_{nq}^\epsilon : \mathbf{m} \equiv \mathbf{k}\mathbf{c} \pmod{2}\}$$

as the subset of typical keys consistent with the given $(\mathbf{m}, \mathbf{c} \neq \mathbf{0})$. As the equation $\mathbf{m} \equiv \mathbf{k}\mathbf{c} \pmod{2}$ introduces q parity constraints, a fraction $1/2^q$ of the typical keys will satisfy them, so that $|A_{nq}^\epsilon(\mathbf{m}, \mathbf{c})| = 2^{-q}|A_{nq}^\epsilon|$. More precisely, the probability mass of any such subset is lower bounded by

$$\begin{aligned} \Pr[A_{nq}^\epsilon(\mathbf{m}, \mathbf{c})] &\geq 2^{-nq[H_2(p) + \epsilon]} \times 2^{-q}(1 - \epsilon)2^{nq[H_2(p) - \epsilon]} \\ &= (1 - \epsilon)2^{-2nq\epsilon - q} \end{aligned}$$

in which the first term on the right-hand side of the first line is a lower bound on the probability of each element \mathbf{k} from the set A_{nq}^ϵ , while the second term is a lower bound on the cardinality of the subset $A_{nq}^\epsilon(\mathbf{m}, \mathbf{c})$. In the same way, this follows an upper bound:

$$\begin{aligned} \Pr[A_{nq}^\epsilon(\mathbf{m}, \mathbf{c})] &\leq 2^{-nq[H_2(p) - \epsilon]} \times 2^{-q}2^{nq[H_2(p) + \epsilon]} \\ &= 2^{2nq\epsilon - q} \end{aligned}$$

Consider now the conditioned entropy

$$\begin{aligned} H(K|\mathbf{m}, \mathbf{c}) &= - \sum_{\mathbf{k} \in A_{nq}^\epsilon(\mathbf{m}, \mathbf{c})} \frac{\Pr(\mathbf{k})}{\Pr[A_{nq}^\epsilon(\mathbf{m}, \mathbf{c})]} \\ &\quad \times \log_2 \frac{\Pr(\mathbf{k})}{\Pr[A_{nq}^\epsilon(\mathbf{m}, \mathbf{c})]}. \end{aligned}$$

By normalization

$$\sum_{\mathbf{k} \in A_{nq}^\epsilon(\mathbf{m}, \mathbf{c})} \frac{\Pr(\mathbf{k})}{\Pr[A_{nq}^\epsilon(\mathbf{m}, \mathbf{c})]} = 1$$

and, thus, upper and lower bounds on $-\log_2(\Pr(\mathbf{k})/\Pr[A_{nq}^\epsilon(\mathbf{m}, \mathbf{c})])$ provide upper and lower bounds on $H(K|\mathbf{m}, \mathbf{c})$. Now, the typical set bounds on $\Pr(\mathbf{k})$ and $\Pr[A_{nq}^\epsilon(\mathbf{m}, \mathbf{c})]$ can be invoked to verify that

$$\begin{aligned} nq[H_2(p) - 3\epsilon] - q &\leq -\log_2 \frac{\Pr(\mathbf{k})}{\Pr[A_{nq}^\epsilon(\mathbf{m}, \mathbf{c})]} \\ &\leq nq[H_2(p) + 3\epsilon] - q + \log_2(1 - \epsilon) \end{aligned}$$

where the bounds therefore apply to $H(K|\mathbf{m}, \mathbf{c})$ as well as to $\sum_{\mathbf{m}, \mathbf{c} \neq \mathbf{0}} H(K|\mathbf{m}, \mathbf{c}) \Pr(\mathbf{m}, \mathbf{c})$. Combining the pieces

$$\begin{aligned} I(K; M, C) &= H(K) - H(K|M, C) \\ &= nqH_2(p) - 2^{-n}nqH_2(p) \\ &\quad - nqH_2(p) + q + \mathcal{O}(\epsilon) \end{aligned}$$

in which $\mathcal{O}(\epsilon)$ collects terms that vanish as $\epsilon \rightarrow 0$. We thus have $I(M; C) = q - I(K; M, C) = 2^{-n}nqH_2(p) + \mathcal{O}(\epsilon)$, and part 2) of the theorem follows by letting $\epsilon \rightarrow 0$. \diamond

Thus, perfect secrecy is approached exponentially fast in the block length n , under either key generation model. A ‘‘perfect stego system’’ is defined in [9] as one with zero message equivocation and minimum embedding distortion. The matrix embedding scheme of Section II-A is thus nearly a perfect stegosystem, as it achieves the minimum embedding distortion according to the rate-distortion bound, and has near-zero message equivocation. Other schemes from [9], such as cyclic shift modulation and stream ciphers, attain zero message equivocation, but fall short of the embedding rate-distortion bound.

IV. CONCLUDING REMARKS

The steganographic matrix embedding schemes under study have been shown to exhibit favorable cryptographic secrecy concerning the message equivocation and key equivocation functions. The main results, however, are conditioned on maximal entropy in the carrier and message sequences. In practice, this may not be a severe limitation,

since proper design of the parity function can reasonably be expected to yield high entropy realizations, and the main results show that such a property is indeed desirable.

REFERENCES

- [1] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.
- [2] J. Fridrich, M. Goljan, P. Lisoněk, and D. Soukal, "Writing on wet paper," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3923–3935, Oct. 2005.
- [3] J. Fridrich, M. Goljan, and D. Soukal, "Wet paper codes with improved coding efficiency," *IEEE Trans. Inf. Forensics Security*, vol. 1, no. 1, pp. 102–110, Mar. 2006.
- [4] J. Fridrich and D. Soukal, "Matrix embedding for large payloads," *IEEE Trans. Inf. Forensics Security*, vol. 1, no. 3, pp. 390–394, Sep. 2006.
- [5] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 563–593, Mar. 2003.
- [6] R. J. Barron, B. Chen, and G. W. Wornell, "The duality between information embedding and source coding with side information and some applications," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1159–1180, May 2003.
- [7] S. S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1181–1203, May 2003.
- [8] C. E. Shannon, "Communication theory of secrecy systems," *Bell Syst. Tech. J.*, vol. 28, pp. 656–715, 1949.
- [9] T. Mittelholzer, "An information-theoretic approach to steganography and watermarking," in *Proc. 3rd Int. Workshop Information Hiding*, ser. Lect. Notes Comput. Sci., A. Pfitzmann, Ed. London, U.K.: Springer, Oct. 1999, vol. 1768, pp. 1–16.
- [10] A. S. Cohen and A. Lapidot, "The Gaussian watermarking game," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1639–1667, Jun. 2002.
- [11] C. Cachin, "An information-theoretic model for steganography," in *Second Int. Workshop on Information Hiding*, D. Aucsmith, Ed., 1998, vol. 1525, pp. 306–318, ser. Lect. Notes Comput. Sci., Springer.
- [12] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1250–1276, Jun. 2002.
- [13] S. Katzenbeisser and F. A. P. Petitcolas, "Defining security in steganographic systems," in *Security and Watermarking of Multimedia Contents IV*, E. J. Delp and P. W. Won, Eds. Philadelphia, PA: SPIE, 2002, vol. 4765, pp. 260–268, Proc. SPIE.
- [14] F. Cayre, C. Fontaine, and T. Furon, "Watermarking security: Theory and practice," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3976–3987, Oct. 2005.
- [15] L. Pérez-Freire, P. Comesaña, J. R. Troncoso-Pastoriza, and F. Pérez-Gonzales, "Watermarking security: A survey," *Trans. DHMS I*, vol. 41–72, 2006.
- [16] J. Fridrich, M. Goljan, and R. Du, "Detecting LSB steganography in color and gray-scale images," *IEEE Multimedia*, vol. 8, no. 4, pp. 22–28, Oct./Nov. 2001.
- [17] H. Farid, "Detecting hidden messages using higher-order statistical models," in *Proc. Int. Conf. Image Processing*, Rochester, NY, 2002, pp. 905–908.
- [18] S. Dumitrescu, X. Wu, and Z. Wang, "Detection of LSB steganography via sample pair analysis," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1995–2007, Jul. 2003.
- [19] C. Cachin, "An information-theoretic model for steganography," *Inf. Comput.*, vol. 192, no. 1, pp. 41–56, July 2004.
- [20] O. Dabeer, K. Sullivan, U. Madhow, S. Chandrasekaran, and B. S. Manjunath, "Detection of hiding in the least significant bit," *IEEE Trans. Signal Process.*, vol. 52, no. 10, pp. 3046–3068, Oct. 2004.
- [21] Y. Wang and P. Moulin, "Steganalysis of block-structured stegotext," in *Proc. SPIE*, E. J. Delp and P. W. Wong, Eds., San Jose, CA, 2004, vol. 5306, no. 1, pp. 477–488, SPIE.
- [22] R. J. Anderson and F. A. P. Petitcolas, "On the limits of steganography," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 4, pp. 474–481, Apr. 1998.
- [23] M. E. Hellman, "An extension of the Shannon theory approach to cryptography," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 3, pp. 289–294, May 1977.
- [24] C. Kurak and J. McHugh, "A cautionary note on image downgrading," in *Proc. IEEE Annu. Computer Security Applications Conf.*, 1992, pp. 153–159.
- [25] N. F. Johnson and S. Jajodia, "Exploring steganography: Seeing the unseen," *IEEE Comput. Mag.*, vol. 31, no. 2, pp. 26–34, Feb. 1998.
- [26] A. Whitehead, "Towards eliminating steganographic communication," in *Proc. 3rd Annual Conf. Privacy, Security Trust*, 2005.
- [27] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: Wiley, 2007.
- [28] E. Martinian and M. J. Wainwright, "Low density codes achieve the rate-distortion bound," presented at the Data Compression Conf., Snow Bird, UT, Mar. 2006.
- [29] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.
- [30] A. W. Eckford, F. R. Kschischang, and S. Pasupathy, "On designing good LDPC codes for Markov channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 5–21, Jan. 2007.
- [31] F. Galand and G. Kabatiansky, "Information hiding by coverings," in *Proc. Information Theory Workshop*, 2003, pp. 151–154.
- [32] C. Fontaine and F. Galand, "How can Reed-Solomon codes improve steganographic schemes?," in *Proc. 9th Int. Workshop Information Hiding*, ser. Lect. Notes Comput. Sci. Berlin, Germany: Springer, 2008, vol. 4567, pp. 130–134.
- [33] T. Johansson and F. Jönsson, "On the complexity of some cryptographic problems based on the general decoding problem," *IEEE Trans. Inf. Theory*, vol. 48, no. 10, pp. 2669–2678, Oct. 2002.
- [34] L. Pérez-Freire, F. Pérez-Gonzales, T. Furon, and P. Comesaña, "Security of lattice-based data hiding against the known message attack," *IEEE Trans. Inf. Forensics Security*, vol. 1, no. 4, pp. 421–439, Dec. 2006.
- [35] L. Pérez-Freire and F. Pérez-Gonzales, "Exploiting security holes in lattice data hiding," in *Proc. Ninth Int. Workshop Information Hiding*, 2007, vol. 4567, Lect. Notes Comput. Sci., Springer.
- [36] D. R. Stinson, *Cryptography: Theory and Practice*, 3rd ed. Boca Raton, FL: CRC, 2006.
- [37] W. Trappe and L. C. Washington, *Introduction to Cryptography with Coding Theory*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 2006.