Semantic Security for the Wiretap Channel

Mihir Bellare[1], Stefano Tessaro[2], and Alexander Vardy[3]

[1] Department of Computer Science & Engineering, University of California San Diego, cseweb.ucsd.edu/~mihir/
[2] CSAIL, Massachusetts Institute of Technology, people.csail.mit.edu/tessaro/
[3] Department of Electrical & Computer Engineering, University of California San Diego, http://www.ece.ucsd.edu/~avardy/

Abstract. The wiretap channel is a setting where one aims to provide information-theoretic privacy of communicated data based solely on the assumption that the channel from sender to adversary is "noisier" than the channel from sender to receiver. It has developed in the Information and Coding (I&C) community over the last 30 years largely divorced from the parallel development of modern cryptography. This paper aims to bridge the gap with a cryptographic treatment involving advances on two fronts, namely definitions and schemes. On the first front (definitions), we explain that the mis-r definition in current use is weak and propose two alternatives: mis (based on mutual information) and ss (based on the classical notion of semantic security). We prove them equivalent, thereby connecting two fundamentally different ways of defining privacy and providing a new, strong and well-founded target for constructions. On the second front (schemes), we provide the first explicit scheme with all the following characteristics: it is proven to achieve both security (ss and mis, not just mis-r) and decodability; it has optimal rate; and both the encryption and decryption algorithms are proven to be polynomial-time.

1 Introduction

The wiretap channel is a setting where one aims to obtain information-theoretic privacy under the sole assumption that the channel from sender to adversary is "noisier" than the channel from sender to receiver. Introduced by Wyner, Csiszár and Körner in the late seventies [41, 14], it has developed in the Information and Coding (I&C) community largely divorced from the parallel development of modern cryptography. This paper aims to bridge the gap with a cryptographic treatment involving advances on two fronts.

The first is *definitions*. We explain that the security definition in current use, that we call mis-r (mutual-information security for random messages) is weak and insufficient to provide security of applications. We suggest strong, new definitions. One, that we call mis (mutual-information security), is an extension of mis-r and thus rooted in the I&C tradition and intuition. Another, semantic security (ss), adapts the cryptographic "gold standard" emanating from [19] and

universally accepted in the cryptographic community. We prove the two equivalent, thereby connecting two fundamentally different ways of defining privacy and providing a new, strong and well-founded target for constructions.

The second is *schemes*. Classically, the focus of the I&C community has been proofs of existence of mis-r schemes of optimal rate. The proofs are non-constructive so that the schemes may not even be explicit let alone polynomial time. Recently, there has been progress towards explicit mis-r schemes [30, 29, 21]. We take this effort to its full culmination by providing the first explicit construction of a scheme with all the following characteristics: it is proven to achieve security (not just mis-r but ss and mis) over the adversary channel; it is proven to achieve decodabilty over the receiver channel; it has optimal rate; and both the encryption and decryption algorithms are proven to be polynomial-time.

Today the possibility of realizing the wiretap setting in wireless networks is receiving practical attention and fueling a fresh look at this area. Our work helps guide the choice of security goals and schemes. Let us now look at all this in more detail.

THE WIRETAP MODEL. The setting is depicted in Fig. 1. The sender applies to her message M a randomized encryption algorithm $\mathcal{E}\colon \{0,1\}^m \to \{0,1\}^c$ to get what we call the *sender ciphertext* $\mathsf{X} \leftarrow\!\!\$\ \mathcal{E}(\mathsf{M})$. This is transmitted to the receiver over the receiver channel ChR so that the latter gets a *receiver ciphertext* $\mathsf{Y} \leftarrow\!\!\$\ \mathsf{ChR}(\mathsf{X})$ which he decrypts via algorithm $\mathcal{D}$ to recover the message. The adversary's wiretap is modeled as another channel ChA and she accordingly gets an *adversary ciphertext* $\mathsf{Z} \leftarrow\!\!\$\ \mathsf{ChA}(\mathsf{X})$ from which she tries to glean whatever she can about the message.

A *channel* is a randomized function specified by a transition probability matrix W where $W[x, y]$ is the probability that input x results in output y. Here x, y are strings. The canonical example is the Binary Symmetric Channel BSC_p with crossover probability $p \leq 1/2$ taking a binary string x of any length and returning the string y of the same length formed by flipping each bit of x independently with probability p. For concreteness and simplicity of exposition we will often phrase discussions in the setting where $\mathsf{ChR}, \mathsf{ChA}$ are BSCs with crossover probabilities $p_R, p_A \leq 1/2$ respectively, but our results apply in much greater generality. In this case the assumption that ChA is "noisier" than ChR corresponds to the assumption that $p_R < p_A$. This is the only assumption made: the adversary is computationally unbounded, and the scheme is keyless, meaning sender and receiver are not assumed to a priori share any information not known to the adversary.

The setting now has a literature encompassing hundreds of papers. (See the survey [28] or the book [6].) Schemes must satisfy two conditions, namely *decoding* and *security*. The *decoding* condition asks that the scheme provide error-correction over the receiver channel, namely the limit as $k \to \infty$ of the maximum, over all M of length m, of $\Pr[\mathcal{D}(\mathsf{ChR}(\mathcal{E}(M))) \neq M]$, is 0, where k is an underlying security parameter of which m, c are functions. The original security condition of [41] was that $\lim_{k\to\infty} \mathbf{I}(\mathsf{M}; \mathsf{ChA}(\mathcal{E}(\mathsf{M}))/m = 0$ where the message random vari-

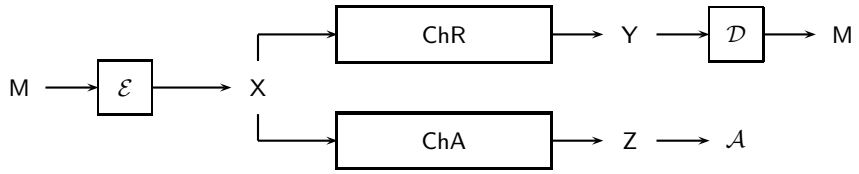


Fig. 1. Wiretap channel model. See text for explanations.

able M is uniformly distributed over $\{0,1\}^m$ and $\mathbf{I}(\mathsf{M};\mathsf{Z}) = \mathbf{H}(\mathsf{M}) - \mathbf{H}(\mathsf{M} \mid \mathsf{Z})$ is the mutual information. This was critiqued by [31, 32] who put forth the stronger security condition now in use, namely that $\lim_{k\to\infty} \mathbf{I}(\mathsf{M};\mathsf{ChA}(\mathcal{E}(\mathsf{M})) = 0$, the random variable M continuing to be uniformly distributed over $\{0,1\}^m$. The *rate* $\mathbf{Rate}(\mathcal{E})$ of a scheme is m/c.

The literature has focused on determining the maximum possible rate. (We stress that the maximum is over all possible schemes, not just ones that are explicitly given or polynomial time.) Shannon's seminal result [37] says that if we ignore security and merely consider achieving the decoding condition then this optimal rate is the receiver channel capacity, which in the BSC case is $1-h_2(p_R)$ where h_2 is the binary entropy function $h_2(p) = -p\lg(p) - (1-p)\lg(1-p)$. He gave non-constructive proofs of existence of schemes meeting capacity. Coming in with this background and the added security condition, it was natural for the wiretap community to follow Shannon's lead and begin by asking what is the maximum achievable rate subject to both the security and decoding conditions. This optimal rate is called the secrecy capacity and, in the case of BSCs, equals the difference $(1 - h_2(p_R)) - (1 - h_2(p_A)) = h_2(p_A) - h_2(p_R)$ in capacities of the receiver and adversary channels. Non-constructive proofs of the existence of schemes with this optimal rate were given in [41, 14, 7]. Efforts to obtain explicit, secure schemes of optimal rate followed [40, 33, 30, 29, 21].

Context. Practical interest in the wiretap setting is escalating. Its proponents note two striking benefits over conventional cryptography: (1) no computational assumptions, and (2) no keys and hence no key distribution. Item (1) is attractive to governments who are concerned with long-term security and worried about quantum computing. Item (2) is attractive in a world where vulnerable, low-power devices are proliferating and key-distribution and key-management are unsurmountable obstacles to security. The practical challenge is to realize a secrecy capacity, meaning ensure by physical means that the adversary channel is noisier than the receiver one. The degradation with distance of radio communication signal quality is the basis of several approaches being investigated for wireless settings. Government-sponsored Ziva Corporation [42] is using optical techniques to build a receiver channel in such a way that wiretapping results in a degraded channel. A program called Physical Layer Security aimed at practical realization of the wiretap channel is the subject of books [6] and conferences [24]. All this activity means that schemes are being sought for implementation. If so, we need privacy definitions that yield security in applications, and we need con-

structive results yielding practical schemes achieving privacy under these definitions. This is what we aim to supply.

DEFINITIONS. A security *metric* xs associates to encryption function $\mathcal{E}\colon \{0,1\}^m \to \{0,1\}^c$ and adversary channel ChA a number $\mathbf{Adv}^{\mathrm{xs}}(\mathcal{E};\mathsf{ChA})$ that measures the maximum "advantage" of an adversary in breaking the scheme under metric xs. For example, the metric underlying the current, above-mentioned security condition is $\mathbf{Adv}^{\text{mis-r}}(\mathcal{E};\mathsf{ChA}) = \mathbf{I}(\mathsf{M};\mathsf{ChA}(\mathcal{E}(\mathsf{M})))$ where M is uniformly distributed over $\{0,1\}^m$. We call this the mis-r (mutual-information security for random messages) metric because messages are assumed to be random. From the cryptographic perspective, this is extraordinarily weak, for we know that real messages are not random. (They may be files, votes or any type of structured data, often with low entropy. Contrary to a view in the I&C community, compression does *not* render data random, as can be seen from the case of votes, where the message space has very low entropy.) This leads us to suggest a stronger metric that we call *mutual-information security*, defined via $\mathbf{Adv}^{\mathrm{mis}}(\mathcal{E};\mathsf{ChA}) = \max_{\mathsf{M}} \mathbf{I}(\mathsf{M};\mathsf{ChA}(\mathcal{E}(\mathsf{M})))$ where the maximum is over all random variables M over $\{0,1\}^m$, regardless of their distribution.

At this point, we have a legitimate metric, but how does it capture privacy? The intuition is that it is measuring the difference in the number of bits required to encode the message before and after seeing the ciphertext. This intuition is alien to cryptographers, whose metrics are based on much more direct and usage-driven privacy requirements. Cryptographers understand since [19] that encryption must hide all partial information about the message, meaning the adversary should have little advantage in computing a function of the message given the ciphertext. (Examples of partial information about a message include its first bit or even the XOR of the first and second bits.) The mis-r and mis metrics ask for nothing like this and are based on entirely different intuition. We extend Goldwasser and Micali's semantic security [19] definition to the wiretap setting, defining $\mathbf{Adv}^{\mathrm{ss}}(\mathcal{E};\mathsf{ChA})$ as

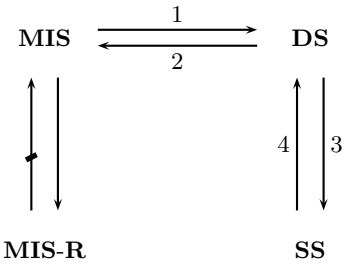
$$\max_{f,\mathsf{M}} \left(\max_{\mathcal{A}} \Pr[\mathcal{A}(\mathsf{ChA}(\mathcal{E}(\mathsf{M}))) = f(\mathsf{M})] - \max_{\mathcal{S}} \Pr[\mathcal{S}(m) = f(\mathsf{M})] \right) .$$

Within the parentheses is the maximum probability that an adversary $\mathcal{A}$, given the adversary ciphertext, can compute the result of function f on the message, minus the maximum probability that a simulator $\mathcal{S}$ can do the same given only the length of the message. We also define a distinguishing security (ds) metric as an analog of indistinguishability [19] via

$$\mathbf{Adv}^{\mathrm{ds}}(\mathcal{E};\mathsf{ChA}) \;=\; \max_{\mathcal{A},M_0,M_1} 2\Pr[\mathcal{A}(M_0,M_1,\mathsf{ChA}(\mathcal{E}(M_{\mathsf{b}}))) = \mathsf{b}] - 1$$

where challenge bit b is uniformly distributed over $\{0,1\}$ and the maximum is over all m-bit messages M_0, M_1 and all adversaries $\mathcal{A}$. For any metric xs, we say $\mathcal{E}$ provides XS-security over ChA if $\lim_{k\to\infty} \mathbf{Adv}^{\mathrm{xs}}(\mathcal{E};\mathsf{ChA}) = 0$.

RELATIONS. The mutual information between message and ciphertext, as measured by mis, is, as noted above, the change in the number of bits needed to encode the message created by seeing the ciphertext. It is not clear why this should



1	$\epsilon_{\rm ds} \le \sqrt{2\epsilon_{\rm mis}}$	Theorem 5
2	$\epsilon_{\rm mis} \le 2\epsilon_{\rm ds} \lg \frac{2^c}{\epsilon_{\rm ds}}$	Theorem 8
3	$\epsilon_{\rm ss} \le \epsilon_{\rm ds}$	Theorem 1
4	$\epsilon_{\rm ds} \le 2\epsilon_{\rm ss}$	Theorem 1

Fig. 2. Relations between notions. An arrow $\mathbf{A} \to \mathbf{B}$ is an implication, meaning every scheme that is **A**-secure is also **B**-secure, while a barred arrow $\mathbf{A} \not\to \mathbf{B}$ is a separation, meaning that there is a **A**-secure scheme that is not **B**-secure. If $\mathcal{E}\colon \{0,1\}^m \to \{0,1\}^c$ is the encryption algorithm and ChA the adversary channel, we let $\epsilon_{\rm xs} = \mathbf{Adv}^{\rm xs}(\mathcal{E}; \mathsf{ChA})$. The table then shows the quantitative bounds underlying the annotated implications.

measure privacy in the sense of semantic security. Yet we are able to show that mutual-information security and semantic security are equivalent, meaning an encryption scheme is MIS-secure if and only if it is SS-secure. Fig. 2 summarizes this and other relations we establish that between them settle all possible relations. The equivalence between SS and DS is the information-theoretic analogue of the corresponding well-known equivalence in the computational setting [19, 3]. As there, however, it brings the important benefit that we can now work with the technically simpler DS. We then show that MIS implies DS by reducing to Pinsker's inequality. We show conversely that DS implies MIS via a general relation between mutual information and statistical distance. As Fig. 2 indicates, the asymptotic relations are all underlain by concrete quantitative and polynomial relations between the advantages. On the other hand, we show that in general MIS-R does not imply MIS, meaning the former is strictly weaker than the latter. We do this by exhibiting an encryption function $\mathcal{E}$ and channel ChA such that $\mathcal{E}$ is MIS-R-secure relative to ChA but MIS-insecure relative to ChA. Furthermore we do this for the case that ChA is a BSC.

Our scheme. We provide the first explicit scheme with all the following characteristics over a large class of adversary and receiver channels including BSCs: (1) It is DS (hence SS, MIS and MIS-R) secure (2) It is proven to satisfy the decoding condition (3) It has optimal rate (4) It is fully polynomial time, meaning both encryption and decryption algorithms run in polynomial time (5) the errors in the security and decoding conditions do not just vanish in the limit but at an exponential rate. Our scheme is based on three main ideas: (1) the use of *invertible* extractors (2) analysis via smooth min-entropy, and (3) a (surprising) result saying that for certain types of schemes, security on random messages implies security on all messages.

Recall that the secrecy capacity is the optimal rate for MIS-R schemes meeting the decoding condition and in the case of BSCs it equals $h_2(p_A) - h_2(p_R)$.

Since DS-security is stronger than MIS-R security, the optimal rate could in principle be smaller. Perhaps surprisingly, it isn't: for a broad class of channels including symmetric channels, the optimal rate is the same for DS and MIS-R security. This follows by applying our above-mentioned result showing that MIS-R implies MIS for certain types of schemes and channels, to known results on achieving the secrecy capacity for MIS-R. Thus, when we say, above, that our scheme achieves optimal rate, this rate is in fact the secrecy capacity.

A common misconception is to think that privacy and error-correction may be completely de-coupled, meaning one would first build a scheme that is secure when the receiver channel is noiseless and then add an ECC on top to meet the decoding condition with a noisy receiver channel. This does not work because the error-correction helps the adversary by reducing the noise over the adversary channel. The two requirements do need to be considered together. Nonetheless, our approach is modular, combining (invertible) extractors with existing ECCs in a blackbox way. As a consequence, any ECC of sufficient rate may be used. This is an attractive feature of our scheme from the practical perspective. In addition our scheme is simple and efficient.

Our claims (proven DS-security and decoding with optimal rate) hold not only for BSCs but for a wide range of receiver and adversary channels.

A CONCRETE INSTANTIATION. As a consequence of our general paradigm, we prove that the following scheme achieves secrecy capacity for the BSC setting with $p_R < p_A \leq 1/2$. For any ECC $\mathsf{E}\colon \{0,1\}^k \to \{0,1\}^n$ such that $k \approx (1 - h_2(p_R))\cdot n$ (such ECCs can be built e.g. from polar codes [2] or from concatenated codes [18]) and a parameter $t \geq 1$, our encryption function $\mathcal{E}$, on input an m-bit message M, where $m = b \cdot t$ and $b \approx (h_2(p_A) - h_2(p_R)) \cdot n$, first selects a random k-bit string $A \neq 0^k$ and t random $(k-b)$-bit strings $R[1], \ldots, R[t]$. It then splits M into t b-bit blocks $M[1], \ldots, M[t]$, and outputs

$$\mathcal{E}(M) = \mathsf{E}(A) \parallel \mathsf{E}(A \odot (M[1] \parallel R[1])) \parallel \cdots \parallel \mathsf{E}(A \odot (M[t] \parallel R[t])) \ ,$$

where $\odot$ is multiplication of k-bit strings interpreted as elements of the extension field $\mathrm{GF}(2^k)$.

RELATED WORK. This paper was formed by merging [5, 4] which together function as the full version of this paper. We refer there for all proofs omitted from this paper and also for full and comprehensive surveys of the large body of work related to wiretap security, and more broadly, to information-theoretically secure communication in a noisy setup. Here we discuss the most related work.

Relations between entropy- and distance-based security metrics have been explored in settings other than the wiretap one, using techniques similar to ours [13, 7, 15], the last in the context of statistically-private committment. Iwamoto and Ohta [25] relate different notions of indistinguishability for statistically secure symmetric encryption. In the context of key-agreement in the wiretap setting (a simpler problem than ours) Csiszár [13] relates MIS-R and RDS, the latter being DS for random messages.

Wyner's syndrome coding approach [41] and extensions by Cohen and Zémor [9, 10] only provide weak security. Hayashi and Matsumoto [21] replace the

matrix-multiplication in these schemes by a universal hash function and show MIS-R security of their scheme. Their result could be used to obtain an alternative to the proof of RDS security used in our scheme for the common case where the extractor is obtained from a universal hash function. However, the obtained security bound is not explicit, making their result unsuitable to applications. Moreover, our proof also yields as a special case a cleaner proof for the result of [21].

Syndrome coding could be viewed as a special case of coset coding which is based on an inner code and outer code. Instantiations of this approach have been considered in [40, 33, 38] using LDPC codes, but polynomial-time decoding is possible only if the adversary channel is a binary erasure channel or the receiver channel is noiseless.

Mahdavifar and Vardy [29] and Hof and Shamai [23] (similar ideas also appeared in [26, 1]) use polar codes [2] to build encryption schemes for the wiretap setting with binary-input symmetric channels. However, these schemes only provide weak security. The full version [30] of [29] provides a variant of the scheme achieving MIS-R security. They use results of the present paper (namely our above-mentioned result that MIS-R implies MIS for certain schemes, whose proof they re-produce for completeness), to conclude that their scheme is also MIS-secure. However there is no proof that decryption (decoding) is possible in their scheme, even in principle let alone in polynomial time. Also efficient generation of polar codes is an open research direction with first results only now appearing [39], and hence relying on this specific code family may be somewhat problematic. Our solution, in contrast, works for arbitrary codes.

As explained in [5], fuzzy extractors [17] can be used to build a DS-secure scheme with polynomial-time encoding and decoding, but the rate of this scheme is far from optimal and the approach is inherently limited to low-rate schemes. We note that (seedless) invertible extractors were previously used in [8] within schemes for the "wiretap channel II" model [34], where the adversary (adaptively) erases ciphertext bits. In contrast to our work, only random-message security was considered in [8].

2 Preliminaries

Basic notation and definitions. "PT" stands for "polynomial-time." If s is a binary string then $s[i]$ denotes its i-th bit and $|s|$ denotes its length. If S is a set then $|S|$ denotes its size. If x is a real number then $|x|$ denotes its absolute value. A function $f : \{0,1\}^m \to \{0,1\}^n$ is *linear* if $f(x \oplus y) = f(x) \oplus f(y)$ for all $x, y \in \{0,1\}^m$. A probability distribution is a function P that associates to each x a probability $P(x) \in [0,1]$. The support $\text{SUPP}(P)$ is the set of all x such that $P(x) > 0$. All probability distributions in this paper are discrete. Associate to random variable X and event E the probability distributions $P_X, P_{X|E}$ defined for all x by $P_{\mathsf{X}}(x) = \Pr[\mathsf{X} = x]$ and $P_{\mathsf{X}|E}(x) = \Pr[\mathsf{X} = x \mid E]$. We denote by $\lg(\cdot)$ the logarithm in base two, and by $\ln(\cdot)$ the natural logarithm. We adopt standard conventions such as $0 \lg 0 = 0 \lg \infty = 0$

and $\Pr[E_1|E_2] = 0$ when $\Pr[E_2] = 0$. The function $h\colon [0,1] \to [0,1]$ is defined by $h(x) = -x \lg x$. The (Shannon) entropy of probability distribution P is defined by $\mathbf{H}(P) = \sum_x h(P(x))$ and the statistical difference between probability distributions P, Q is defined by $\mathbf{SD}(P;Q) = 0.5 \cdot \sum_x |P(x) - Q(x)|$. If X, Y are random variables the (Shannon) entropy is defined by $\mathbf{H}(\mathsf{X}) = \mathbf{H}(P_{\mathsf{X}}) = \sum_x h(P_{\mathsf{X}}(x))$. The conditional entropy is defined via $\mathbf{H}(\mathsf{X} \mid \mathsf{Y} = y) = \sum_x h(P_{\mathsf{X}|\mathsf{Y}=y}(x))$ and $\mathbf{H}(\mathsf{X} \mid \mathsf{Y}) = \sum_y P_{\mathsf{Y}}(y) \cdot \mathbf{H}(\mathsf{X} \mid \mathsf{Y} = y)$. The mutual information is defined by $\mathbf{I}(\mathsf{X};\mathsf{Y}) = \mathbf{H}(\mathsf{X}) - \mathbf{H}(\mathsf{X} \mid \mathsf{Y})$. The statistical or variational distance between random variables $\mathsf{X}_1, \mathsf{X}_2$ is $\mathbf{SD}(\mathsf{X}_1;\mathsf{X}_2) = \mathbf{SD}(P_{\mathsf{X}_1}; P_{\mathsf{X}_2}) = 0.5 \cdot \sum_x |\Pr[\mathsf{X}_1 = x] - \Pr[\mathsf{X}_2 = x]|$. The min-entropy of random variable X is $\mathbf{H}_\infty(\mathsf{X}) = -\lg(\max_x \Pr[\mathsf{X} = x])$ and if Z is also a random variable the conditional min-entropy is $\mathbf{H}_\infty(\mathsf{X}|\mathsf{Z}) = -\lg(\sum_z \Pr[\mathsf{Z} = z] \max_x \Pr[\mathsf{X} = x|\mathsf{Z} = z])$.

Transforms, channels and algorithms. We say that T is a transform with domain D and range R, written $T\colon\ D \to R$, if $T(x)$ is a random variable over R for every $x \in D$. Thus, T is fully specified by a sequence $P = \{P_x\}_{x \in D}$ of probability distributions over R, where $P_x(y) = \Pr[T(x) = y]$ for all $x \in D$ and $y \in R$. We call P the distribution associated to T. This distribution can be specified by a $|D|$ by $|R|$ transition probability matrix W defined by $W[x,y] = P_x(y)$. A channel is simply a transform. This is a very general notion of a channel but it does mean channels are memoryless in the sense that two successive transmissions over the same channel are independent random variables. The transition probability matrix representation is the most common one in this case. A (randomized) algorithm is also a transform. Finally, an adversary too is a transform, and so is a simulator.

If $T\colon\ \{0,1\} \to R$ is a transform we may apply it to inputs of any length. The understanding is that T is applied independently to each bit of the input. For example BSC_p, classically defined as a 1-bit channel, is here viewed as taking inputs of arbitrary length and flipping each bit independently with probability p. Similarly, we apply a transform $T\colon\ \{0,1\}^l \to R$ to any input whose length is divisible by l. We say that a transform $T\colon\ D \to R$ with transition matrix W is *symmetric* if the there exists a partition of the range as $R = R_1 \cup \cdots \cup R_n$ such that for all i the sub-matix $W_i = W[\cdot, R_i]$ induced by the columns in R_i is strongly symmetric, i.e., all rows of W_i are permutations of each other, and all columns of W_i are permutations of each other.

3 Security metrics and relations

Encryption functions and schemes. An *encryption function* is a transform $\mathcal{E}\colon\ \{0,1\}^m \to \{0,1\}^c$ where m is the message length and c is the sender ciphertext length. The *rate* of $\mathcal{E}$ is $\mathbf{Rate}(\mathcal{E}) = m/c$. If $\mathsf{ChR}\colon\ \{0,1\}^c \to \{0,1\}^d$ is a receiver channel then a *decryption function* for $\mathcal{E}$ over ChR is a transform $\mathcal{D}\colon\ \{0,1\}^d \to \{0,1\}^m$ whose decryption error $\mathbf{DE}(\mathcal{E};\mathcal{D};\mathsf{ChR})$ is defined as the maximum, over all $M \in \{0,1\}^m$, of $\Pr[\mathcal{D}(\mathsf{ChR}(\mathcal{E}(M))) \neq M]$.

An *encryption scheme* $\overline{\mathcal{E}} = \{\mathcal{E}_k\}_{k \in \mathbb{N}}$ is a family of encryption functions where $\mathcal{E}_k\colon\ \{0,1\}^{m(k)} \to \{0,1\}^{c(k)}$ for functions $m, c\colon\ \mathbb{N} \to \mathbb{N}$ called the mes-

sage length and sender ciphertext length of the scheme. The *rate* of $\overline{\mathcal{E}}$ is the function $\mathbf{Rate}_{\overline{\mathcal{E}}}\colon \mathbb{N} \to \mathbb{R}$ defined by $\mathbf{Rate}_{\overline{\mathcal{E}}}(k) = \mathbf{Rate}(\mathcal{E}_k)$ for all $k \in \mathbb{N}$. Suppose $\overline{\mathsf{ChR}} = \{\mathsf{ChR}_k\}_{k\in\mathbb{N}}$ is a family of receiver channels where $\mathsf{ChR}_k\colon \{0,1\}^{c(k)} \to \{0,1\}^{d(k)}$. Then a *decryption scheme* for $\overline{\mathcal{E}}$ over $\overline{\mathsf{ChR}}$ is a family $\overline{\mathcal{D}} = \{\mathcal{D}_k\}_{k\in\mathbb{N}}$ where $\mathcal{D}_k\colon \{0,1\}^{d(k)} \to \{0,1\}^{m(k)}$ is a decryption function for $\mathcal{E}_k$ over ChR_k. We say that $\overline{\mathcal{D}}$ is a correct decryption scheme for $\overline{\mathcal{E}}$ relative to $\overline{\mathsf{ChR}}$ if the limit as $k \to \infty$ of $\mathbf{DE}(\mathcal{E}_k; \mathcal{D}_k; \mathsf{ChR}_k)$ is 0. We say that encryption scheme $\overline{\mathcal{E}}$ is decryptable relative to $\overline{\mathsf{ChR}}$ if there exists a correct decryption scheme for $\overline{\mathcal{E}}$ relative to $\overline{\mathsf{ChR}}$. A family $\{\mathcal{S}_k\}_{k\in\mathbb{N}}$ (eg. an encryption or decryption scheme) is PT if there is a polynomial time computable function which on input 1^k (the unary representation of k) and x returns $\mathcal{S}_k(x)$. Our constructs will provide PT encryption and decryption schemes.

Security metrics. Let $\mathcal{E}\colon \{0,1\}^m \to \{0,1\}^c$ be an encryption function and let $\mathsf{ChA}\colon \{0,1\}^c \to \{0,1\}^d$ be an adversary channel. Security depends only on these, not on the receiver channel. We now define semantic security (ss), distinguishing security (ds), random mutual-information security (mis-r) and mutual information security (mis). For each type of security $\mathrm{xs} \in \{\mathrm{ss}, \mathrm{ds}, \text{mis-r}, \mathrm{mis}\}$, we associate to $\mathcal{E}; \mathsf{ChA}$ a real number $\mathbf{Adv}^{\mathrm{xs}}(\mathcal{E}; \mathsf{ChA})$. The smaller this number, the more secure is $\mathcal{E}; \mathsf{ChA}$ according to the metric in question. The security of an encryption function is quantitative, as captured by the advantage. We might measure it in bits, saying that $\mathcal{E}; \mathsf{ChA}$ has s bits of xs-security if $\mathbf{Adv}^{\mathrm{xs}}(\mathcal{E}; \mathsf{ChA}) \le 2^{-s}$. A qualitative definition of "secure," meaning one under which something is secure or not secure, may only be made asymptotically, meaning for schemes. We say encryption scheme $\overline{\mathcal{E}} = \{\mathcal{E}_k\}_{k\in\mathbb{N}}$ is XS-secure relative to $\overline{\mathsf{ChA}} = \{\mathsf{ChA}_k\}_{k\in\mathbb{N}}$ if $\lim_{k\to\infty} \mathbf{Adv}^{\mathrm{xs}}(\mathcal{E}_k; \mathsf{ChA}_k) = 0$. This does not mandate any particular rate at which the advantage should vanish, but in our constructions this rate is exponentially vanishing with k. We define the ss advantage $\mathbf{Adv}^{\mathrm{ss}}(\mathcal{E}; \mathsf{ChA})$ as

$$\max_{f,\mathsf{M}} \left(\max_{\mathcal{A}} \Pr[\mathcal{A}(\mathsf{ChA}(\mathcal{E}(\mathsf{M}))) = f(\mathsf{M})] - \max_{\mathcal{S}} \Pr[\mathcal{S}(m) = f(\mathsf{M})] \right) . \qquad (1)$$

Here f is a transform with domain $\{0,1\}^m$ that represents partial information about the message. Examples include $f(M) = M$ or $f(M) = M[1]$ or $f(M) = M[1] \oplus \cdots \oplus M[m]$, where $M[i]$ is the i-th bit of M. But f could be a much more complex function, and could even be randomized. The adversary's goal is to compute $f(\mathsf{M})$ given an adversary ciphertext $\mathsf{ChA}(\mathcal{E}(\mathsf{M}))$ formed by encrypting message M. The probability that it does this is $\Pr[\mathcal{A}(\mathsf{ChA}(\mathcal{E}(\mathsf{M}))) = f(\mathsf{M})]$, then maximized over all adversaries $\mathcal{A}$ to achieve strategy independence. We then subtract the a priori probability of success, meaning the maximum possible probability of computing $f(\mathsf{M})$ if you are not given the adversary ciphertext. Finally, the outer max over all f, M ensures that the metric measures the extent to which *any* partial information leaks regardless of message distribution. We define the distinguishing advantage via

$$\mathbf{Adv}^{\mathrm{ds}}(\mathcal{E}; \mathsf{ChA}) = \max_{\mathcal{A}, M_0, M_1} 2\Pr[\mathcal{A}(M_0, M_1, \mathsf{ChA}(\mathcal{E}(M_{\mathsf{b}}))) = \mathsf{b}] - 1 \qquad (2)$$

$$= \max_{M_0, M_1} \mathbf{SD}(\mathsf{ChA}(\mathcal{E}(M_0)); \mathsf{ChA}(\mathcal{E}(M_1))) . \qquad (3)$$

In Eq. (2), $\Pr[\mathcal{A}(M_0, M_1, \mathsf{ChA}(\mathcal{E}(M_\mathsf{b}))) = \mathsf{b}]$ is the probability that adversary $\mathcal{A}$, given m-bit messages M_0, M_1 and an adversary ciphertext emanating from M_b, correctly identifies the random challenge bit b. The a priori success probability being $1/2$, the advantage is appropriately scaled. This advantage is equal to the statistical distance between the random variables $\mathsf{ChA}(\mathcal{E}(M_0))$ and $\mathsf{ChA}(\mathcal{E}(M_1))$. The mutual-information security advantages are defined via

$$\mathbf{Adv}^{\text{mir-r}}(\mathcal{E}; \mathsf{ChA}) = \mathbf{I}(\mathsf{U}; \mathsf{ChA}(\mathcal{E}(\mathsf{U}))) \tag{4}$$
$$\mathbf{Adv}^{\text{mis}}(\mathcal{E}; \mathsf{ChA}) = \max_{\mathsf{M}} \mathbf{I}(\mathsf{M}; \mathsf{ChA}(\mathcal{E}(\mathsf{M}))) \tag{5}$$

where the random variable U is uniformly distributed over $\{0,1\}^m$.

DS IS EQUIVALENT TO SS. Theorem 1 below says that SS and DS are equivalent up to a small constant factor in the advantage. This is helpful because DS is more analytically tractable than SS. The proof is an extension of the classical ones in computational cryptography and is given in [5].

Theorem 1. [DS $\leftrightarrow$ SS] *Let $\mathcal{E}$: $\{0,1\}^m \to \{0,1\}^c$ be an encryption function and ChA an adversary channel. Then* $\mathbf{Adv}^{\text{ss}}(\mathcal{E}; \mathsf{ChA}) \le \mathbf{Adv}^{\text{ds}}(\mathcal{E}; \mathsf{ChA}) \le 2 \cdot \mathbf{Adv}^{\text{ss}}(\mathcal{E}; \mathsf{ChA})$. ∎

MIS IMPLIES DS. The KL divergence is a distance measure for probability distributions P, Q defined by $\mathbf{D}(P; Q) = \sum_x P(x) \lg P(x)/Q(x)$. Let M, C be random variables. Probability distributions $J_{M,C}, I_{M,C}$ are defined for all M, C by $J_{\mathsf{M},\mathsf{C}}(M, C) = \Pr[\,\mathsf{M} = M, \mathsf{C} = C\,]$ and $I_{\mathsf{M},\mathsf{C}}(M, C) = \Pr[\,\mathsf{M} = M\,] \cdot \Pr[\,\mathsf{C} = C\,]$. Thus $J_{\mathsf{M},\mathsf{C}}$ is the joint distribution of M and C, while $I_{\mathsf{M},\mathsf{C}}$ is the "independent" or product distribution. The following is standard:

Lemma 2. *Let M, C be random variables. Then* $\mathbf{I}(\mathsf{M}; \mathsf{C}) = \mathbf{D}(J_{\mathsf{M},\mathsf{C}}; I_{\mathsf{M},\mathsf{C}})$. ∎

Pinsker's inequality —from [35] with the tight constant from [12]— lower bounds the KL divergence between two distributions in terms of their statistical distance:

Lemma 3. *Let P, Q be probability distributions. Then* $\mathbf{D}(P; Q) \ge 2 \cdot \mathbf{SD}(P; Q)^2$. ∎

To use the above we need the following, whose proof is in [5]:

Lemma 4. *Let M be uniformly distributed over $\{M_0, M_1\} \subseteq \{0,1\}^m$. Let g: $\{0,1\}^m \to \{0,1\}^c$ be a transform and let $\mathsf{C} = g(\mathsf{M})$. Then* $\mathbf{SD}(J_{\mathsf{M},\mathsf{C}}; I_{\mathsf{M},\mathsf{C}})$ *equals* $\mathbf{SD}(g(M_0); g(M_1))/2$. ∎

Combining the lemmas, we show the following in [5]:

Theorem 5. [MIS $\to$ DS] *Let $\mathcal{E}$: $\{0,1\}^m \to \{0,1\}^c$ be an encryption function and ChA an adversary channel. Then* $\mathbf{Adv}^{\text{ds}}(\mathcal{E}; \mathsf{ChA}) \le \sqrt{2 \cdot \mathbf{Adv}^{\text{mis}}(\mathcal{E}; \mathsf{ChA})}$. ∎

DS IMPLIES MIS. The following general lemma from [5] bounds the difference in entropy between two distributions in terms of their statistical distance. It is a slight strengthening of [11, Theorem 16.3.2]. Similar bounds are provided in [22].

Lemma 6. *Let P, Q be probability distributions. Let $N = |\text{SUPP}(P) \cup \text{SUPP}(Q)|$ and $\epsilon = \mathbf{SD}(P; Q)$. Then $\mathbf{H}(P) - \mathbf{H}(Q) \leq 2\epsilon \cdot \lg(N/\epsilon)$.* ∎

To exploit this, we define the *pairwise statistical distance* $\mathbf{PSD}(\mathsf{M}; \mathsf{C})$ between random variables M, C as the maximum, over all messages $M_0, M_1 \in \text{SUPP}(P_{\mathsf{M}})$, of $\mathbf{SD}(P_{\mathsf{C}|\mathsf{M}=M_0}; P_{\mathsf{C}|\mathsf{M}=M_1})$. The proof of the following is in [5].

Lemma 7. *Let M, C be random variables. Then $\mathbf{SD}(P_{\mathsf{C}}; P_{\mathsf{C}|\mathsf{M}=M}) \leq \mathbf{PSD}(\mathsf{M}; \mathsf{C})$ for any M.* ∎

From this we conclude in [5] that DS implies MIS:

Theorem 8. [DS → MIS] *Let $\mathcal{E}$: $\{0,1\}^m \to \{0,1\}^c$ be an encryption function and ChA an adversary channel. Let $\epsilon = \mathbf{Adv}^{\text{ds}}(\mathcal{E}; \mathsf{ChA})$. Then $\mathbf{Adv}^{\text{mis}}(\mathcal{E}; \mathsf{ChA}) \leq 2\epsilon \cdot \lg(2^c/\epsilon)$.* ∎

The somewhat strange-looking form of the bound of Theorem 8 naturally raises the question of whether Lemma 6 is tight. The following says that it is up to a constant factor of 4. The proof is in [5].

Proposition 9. *Let $n > k \geq 1$ be integers. Let $\epsilon = 2^{-k}$ and $N = 1 + \epsilon 2^n$. Then there are distributions P, Q with $|\text{SUPP}(P) \cup \text{SUPP}(Q)| = N$ and $\mathbf{SD}(P; Q) = \epsilon$ and $\mathbf{H}(P) - \mathbf{H}(Q) \geq 0.5 \cdot \epsilon \cdot \lg(N/\epsilon)$.* ∎

OTHER RELATIONS. We have now justified all the numbered implication arrows in Fig. 2. The un-numbered implication **MIS** → **MIS-R** is trivial. The intuition for the separation **MIS-R** ↛ **MIS** is simple. Let $\mathcal{E}$ be the identity function. Let ChA faithfully transmit inputs 0^m and 1^m and be very noisy on other inputs. Then **MIS** fails because the adversary has high advantage when the message takes on only values $0^m, 1^m$ but **MIS-R**-security holds since these messages are unlikely. This example may seem artificial. In [5] we give a more complex example where ChA is a BSC and the encryption function is no longer trivial.

4 DS-Secure Encryption Achieving Secrecy Capacity

This section presents our main technical result, an encryption scheme achieving DS-security. Its rate, for a large set of adversary channels, is optimal.

HIGH-LEVEL APPROACH. We start by considering an extension of the usual setting where sender and receiver share a *public* random value S, i.e., known to the adversary, and which we call the *seed*. We will call an encryption function in this setting a *seeded encryption function*. For simplicity, this discussion will focus on the case where ChR and ChA are BSCs with respective crossover probabilities $p_R < p_A \leq 1/2$, and we assume that sender and receiver only want to agree on a joint secret key. If we let S be the seed of an extractor Ext: $\text{SDS} \times \{0,1\}^k \to \{0,1\}^m$ and given an error-correcting code E: $\{0,1\}^k \to \{0,1\}^n$ for reliable communication over BSC_{p_R}, a natural approach consists of the sender sending $\mathsf{E}(R)$, for a random k-bit R, to the receiver, and both parties now derive the key as $K = \mathsf{Ext}(S, R)$, since the receiver can recover R with very high probability.

The achievable key length is at most $\mathbf{H}_\infty(R|Z)$, where $Z = \mathsf{BSC}_{p_A}(\mathsf{E}(R))$. Yet, it is not hard to see that the most likely outcome, when $Z = z$, is that R equals the unique r such that $\mathsf{E}(r) = z$, and that hence $\mathbf{H}_\infty(R|Z) = n \cdot \lg(1/(1-p_A))$, falling short of achieving capacity $h(p_A) - h(p_R)$. To overcome this, we will observe the following: We can think of BSC_{p_A} as adding an n-bit vector E to its input $\mathsf{E}(R)$, where each bit $E[i]$ of the noise vector takes value one with probability p_A. With overwhelming probability, E is (roughly) uniformly distributed on the set of n-bit vectors with hamming weight (approximately) $p_A \cdot n$ and there are (approximately) $2^{n \cdot h_2(p_A)}$ such vectors. Therefore, choosing the noise uniformly from such vectors does not change the experiment much, and moreover, in this new experiment, one can show that roughly $\mathbf{H}_\infty(R|Z) \geq k - n \cdot (1 - h_2(p_A))$, which yields optimal rate using an optimal code with $k \approx (1-h(p_R)) \cdot n$. We will make this precise for a general class of symmetric channels via the notion of *smooth min-entropy* [36].

But recall that our goal is way more ambitious: Alice wants to send an *arbitrary message of her choice*. The obvious approach is obtain a key K as above and then send an error-corrected version of $K \oplus M$. But this at least halves the rate, which becomes far from optimal. Our approach instead is to use an extractor Ext that is *invertible*, in the sense that given M and S, we can sample a random R such that $\mathsf{Ext}(S, R) = M$. We then encrypt a message M as $\mathsf{E}(R)$, where R is a random preimage of M under $\mathsf{Ext}(S, \cdot)$. However, the above argument only yields, at best, security for randomly chosen messages. In contrast, showing DS-security accounts to proving, for any two messages M_0 and M_1, that $\mathsf{BSC}_{p_A}(\mathsf{E}(R_0))$ and $\mathsf{BSC}_{p_A}(\mathsf{E}(R_1))$ are statistically close, where R_i is uniform such that $\mathsf{Ext}(S, R_i) = M_i$. To make things even worse, we allow the messages M_0 and M_1 are allowed to depend on the seed. The main challenge is that such proof appears to require detailed knowledge of the combinatorial structure of E and Ext, as the actual ciphertext distribution depends on them.

Instead, we will take a completely different approach: We show that any seeded encryption function with appropriate linearity properties is DS-secure whenever it is secure for random messages. This result is surprising, as random-message security does *not*, in general, imply chosen-message security. A careful choice of the extractor to satisfy these requirements, combined with the above idea, yields a DS-secure seeded encryption function. The final step is to remove the seed, which is done by transmitting it (error-corrected) and amortizing out its impact on the rate to essentially zero by re-using it with the above seeded encryption function across blocks of the message. A hybrid argument is used to bound the decoding error and loss in security.

Seeded encryption. A *seeded encryption function* $\mathcal{SE}$: $\textsc{Sds} \times \{0,1\}^b \to \{0,1\}^n$ takes a seed $S \in \textsc{Sds}$ and message $M \in \{0,1\}^b$ to return a sender ciphertext denoted $\mathcal{SE}(S, M)$ or $\mathcal{SE}_S(M)$; each seed S defines an encryption function $\mathcal{SE}_S$: $\{0,1\}^b \to \{0,1\}^n$. There must be a corresponding seeded decryption function $\mathcal{SD}$: $\textsc{Sds} \times \{0,1\}^n \to \{0,1\}^b$ such that $\mathcal{SD}(S, \mathcal{SE}(S, M)) = M$ for all S, M. We consider an extension of the standard wiretap setting where a seed $S \xleftarrow{\$} \textsc{Sds}$ is a public parameter, available to sender, receiver and adversary. We extend DS-

transform $\mathcal{SE}(S, M)$:	**transform** $\mathcal{D}(C)$: // $C \in \text{OUTR}^{(c+t)n}$
// $S \in \text{SDS}$, $M \in \{0,1\}^b$	$C[1], \ldots, C[c+t] \xleftarrow{n} C$
$R \leftarrow\!\!\$\ \{0,1\}^r$; Ret $\mathsf{E}(\mathsf{Inv}(S, R, M))$.	$S \leftarrow \mathsf{D}(C[1]) \,\|\, \ldots \,\|\, \mathsf{D}(C[c])$
	For $i = 1$ to t do
transform $\mathcal{E}(M)$: // $M \in \{0,1\}^m$	$\quad X[i] \leftarrow \mathsf{D}(C[c+i])$
$S \leftarrow\!\!\$\ \text{SDS}$; $S[1], \ldots, S[c] \xleftarrow{k} S$	$\quad M[i] \leftarrow \mathsf{Ext}(S, X[i])$
$M[1], \ldots, M[t] \xleftarrow{b} M$	Ret $M[1] \,\|\, \cdots \,\|\, M[t]$.
For $i = 1$ to t do $C[i] \leftarrow\!\!\$\ \mathcal{SE}(S, M[i])$	
Ret $\mathsf{E}(S[1]) \,\|\, \cdots \,\|\, \mathsf{E}(S[c]) \,\|\, C[1] \,\|\, \cdots \,\|\, C[t]$.	

Fig. 3. Encryption function $\mathcal{E} = \mathbf{RItE}_t[\mathsf{Inv}, \mathsf{E}]$ using $\mathcal{SE} = \mathbf{ItE}[\mathsf{Inv}, \mathsf{E}]$ and decryption function $\mathcal{D}$. By $X[1], \ldots, X[c] \xleftarrow{b} X$ we mean that bc-bit string X is split into b-bit blocks.

security to this setting by letting $\mathbf{Adv}^{\text{ds}}(\mathcal{SE}; \mathsf{ChA})$ be the expectation, over S drawn at random from SDS, of $\mathbf{Adv}^{\text{ds}}(\mathcal{SE}_S; \mathsf{ChA})$. The rate of $\mathcal{SE}$ is defined as b/n, meaning the seed is ignored.

EXTRACTORS. A function Ext: $\text{SDS} \times \{0,1\}^k \to \{0,1\}^b$ is an (h, α)-*extractor* if $\mathbf{SD}((\mathsf{Ext}(\mathsf{S}, \mathsf{X}), \mathsf{Z}, \mathsf{S}); (\mathsf{U}, \mathsf{Z}, \mathsf{S})) \leq \alpha$ for all pairs of (correlated) random variables (X, Z) over $\{0,1\}^k \times \{0,1\}^*$ with $\mathbf{H}_\infty(\mathsf{X}|\mathsf{Z}) \geq h$, where additionally S and U are uniform on SDS and $\{0,1\}^b$, respectively. (This is a strong, average case extractor in the terminology of [16].) We will say that Ext is *regular* if for all $S \in \text{SDS}$, the function $\mathsf{Ext}(S, \cdot)$ is regular, meaning every point in the range has the same number of preimages.

INVERTING EXTRACTORS. We say that a function $\mathsf{Inv} : \text{SDS} \times \{0,1\}^r \times \{0,1\}^b \to \{0,1\}^k$ is an *inverter* for an extractor $\mathsf{Ext} : \text{SDS} \times \{0,1\}^k \to \{0,1\}^b$ if for all $S \in \text{SDS}$ and $Y \in \{0,1\}^b$, and for R uniform over $\{0,1\}^k$, the random variable $\mathsf{Inv}(S, \mathsf{R}, Y)$ is uniformly distributed on $\{\ X \in \{0,1\}^k\ :\ \mathsf{Ext}(S, X) = Y\ \}$, the set of preimages of Y under $\mathsf{Ext}(S, \cdot)$. To make this concrete we give an example of an extractor with an efficiently computable inverter. Recall that k-bit strings can be interpreted as elements of the finite field $\text{GF}(2^k)$, allowing us to define a multiplication operator $\odot$ on k-bit strings. Then, for $\text{SDS} = \{0,1\}^k \setminus 0^k$, we consider the function $\mathsf{Ext} : \text{SDS} \times \{0,1\}^k \to \{0,1\}^b$ which, on inputs $S \in \text{SDS}$ and $X \in \{0,1\}^k$, outputs the first b bits of $X \odot S$. It is easy to see that Ext is regular, as 0^k is not in the set of seeds. In [4] we prove the following using the average-case version of the Leftover Hash Lemma of [20], due to [16].

Lemma 10. *For all $\alpha \in (0, 1]$ and all $b \leq k - 2\lg(1/\alpha) + 2$ the function Ext is a $(b + 2\lg(1/\alpha) - 2, \alpha)$-extractor.*

An efficient inverter $\mathsf{Inv} : \text{SDS} \times \{0,1\}^{k-b} \times \{0,1\}^b \to \{0,1\}^k$ is obtained via $\mathsf{Inv}(S, R, M) = S^{-1} \odot (M \,\|\, R)$ where S^{-1} is the inverse of S with respect to multiplication in $\text{GF}(2^k)$.

THE RITE CONSTRUCTION. Let $\mathsf{Ext} : \text{SDS} \times \{0,1\}^k \to \{0,1\}^b$ be a regular extractor with inverter $\mathsf{Inv} : \text{SDS} \times \{0,1\}^r \times \{0,1\}^b \to \{0,1\}^k$. Also let $\mathsf{E} : \{0,1\}^k \to \{0,1\}^n$ be an injective function with $k \leq n$, later to be instantiated via an ECC. Assume without loss of generality that for some $c \geq 1$, we have $|S| = c \cdot k$ for all $S \in \text{SDS}$. The encryption function $\mathcal{E}$ is described in Fig. 3 and is obtained via the construction $\mathbf{RItE}_t$ (Repeat Invert-then-Encode), where $t \geq 1$ is a parameter: As its main component, it relies on the construction **ItE** (Invert-then-Encode) of a seeded encryption function $\mathbf{ItE}[\mathsf{Inv}, \mathsf{E}] : \text{SDS} \times \{0,1\}^b \to \{0,1\}^n$ which applies the inverter Inv to the message, and then applies E to the result. The final, seed-free, encryption function $\mathbf{RItE}_t[\mathsf{Inv}, \mathsf{E}]$ then takes an input $M \in \{0,1\}^m$, where $m = t \cdot b$, splits it into t b-bit blocks $M[1], \ldots M[t]$, chooses a random seed S, and combines an encoding of S with the encryptions of the blocks using $\mathcal{SE}_S$ for $\mathcal{SE} = \mathbf{ItE}[\mathsf{Inv}, \mathsf{E}]$.

DECRYPTABILITY. Given a channel $\mathsf{ChR} : \{0,1\} \to \text{OUTR}$, a decoder for E over ChR is a function $\mathsf{D} : \text{OUTR}^n \to \{0,1\}^k$. Its decoding error is defined as $\mathbf{DE}(\mathsf{E}; \mathsf{D}; \mathsf{ChR}) = \max_{M \in \{0,1\}^k} \Pr[\, \mathsf{D}(\mathsf{ChR}(\mathsf{E}(M))) \neq M \,]$. Therefore, for any output alphabet OUTR and function $\mathsf{D} : \text{OUTR}^n \to \{0,1\}^b$, we define the corresponding decryption function for $\mathcal{E}$ over ChR as in Fig. 3. The following lemma summarizes the relation between its decryption error and the one of D.

Lemma 11. **[Correct decryption]** *Let* $\mathsf{ChR} : \{0,1\} \to \text{OUTR}$ *be a channel, and let* $\mathcal{E}$, $\mathcal{D}$, E, *and* D *be as above. Then,* $\mathbf{DE}(\mathcal{E}; \mathcal{D}; \mathsf{ChR}) \leq (c + t) \cdot \mathbf{DE}(\mathsf{E}; \mathsf{D}; \mathsf{ChR})$. ∎

STEP I: FROM RITE TO ITE. We reduce security of **RItE** to that of **ItE**. The proof of the following [4] uses a hybrid argument.

Lemma 12. *Let* $t \geq 1$, $\mathcal{E} = \mathbf{RItE}_t[\mathsf{Inv}, \mathsf{E}]$ *and* $\mathcal{SE} = \mathbf{ItE}[\mathsf{Inv}, \mathsf{E}]$. *For all* $\mathsf{ChA} : \{0,1\}^n \to \text{OUTA}$ *we have* $\mathbf{Adv}^{\text{ds}}(\mathcal{E}; \mathsf{ChA}) \leq t \cdot \mathbf{Adv}^{\text{ds}}(\mathcal{SE}; \mathsf{ChA})$. ∎

STEP II: RDS-SECURITY OF ITE. Towards determining the DS-security of **ItE** we first address the seemingly simpler question of proving security under *random* messages. Specifically, for a seeded encryption function $\mathcal{SE} : \text{SDS} \times \{0,1\}^b \to \{0,1\}^n$, we define the rds advantage $\mathbf{Adv}^{\text{rds}}(\mathcal{SE}; \mathsf{ChA})$ as the expectation of $\mathbf{SD}((\mathsf{ChA}(\mathcal{SE}(S, \mathsf{U})), \mathsf{U}); (\mathsf{ChA}(\mathcal{SE}(S, \mathsf{U}')), \mathsf{U}))$ where U and U' are uniformly chosen and independent b-bit messages, and the expectation is taken over the choice of the seed S. Exploiting the notion of ϵ-*smooth min-entropy* [36], the following, proven in [4], establishes RDS-security of **ItE**:

Lemma 13. **[RDS-security of ItE]** *Let* $\delta > 0$, *let* $\mathsf{ChA} : \{0,1\} \to \text{OUTA}$ *be a symmetric channel, let* $\mathsf{Inv} : \text{SDS} \times \{0,1\}^r \times \{0,1\}^b \to \{0,1\}^k$ *be the inverter of a regular* $(k - n \cdot (\lg(|\text{OUTA}|) - \mathbf{H}(\mathsf{ChA}) + \delta), \alpha)$-*extractor, and let* $\mathsf{E} : \{0,1\}^k \to \{0,1\}^n$ *be injective. Then, for* $\mathcal{SE} = \mathbf{ItE}[\mathsf{Inv}, \mathsf{E}]$, *we have*

$$\mathbf{Adv}^{\text{rds}}(\mathcal{SE}; \mathsf{ChA}) \leq 2 \cdot 2^{-\frac{n\delta^2}{2 \lg^2(|\text{OUTA}|+3)}} + \alpha \,. \quad ∎$$

STEP III: FROM RDS- TO DS-SECURITY. In contrast to RDS-security, proving DS-security of **ItE** seems to require a better grasp of the combinatorial structure

of E and Inv. More concretely, think of any randomized (seeded) encryption function $\mathcal{SE} : \textsc{Sds} \times \{0,1\}^b \to \{0,1\}^n$ as a deterministic map $\mathcal{SE} : \textsc{Sds} \times \{0,1\}^r \times \{0,1\}^b \to \{0,1\}^n$ (for some r), where the second argument takes the role of the random coins. We call $\mathcal{SE}$ *separable* if $\mathcal{SE}(S,R,M) = \mathcal{SE}(S,R,0^b) \oplus \mathcal{SE}(S,0^r,M)$ for all $S \in \textsc{Sds}$, $R \in \{0,1\}^r$, and $M \in \{0,1\}^b$. Also, it is *message linear* if $\mathcal{SE}(S,0^r,\cdot)$ is linear for all $S \in \textsc{Sds}$. The following is true for encryption functions with both these properties, and is proven in [4].

Lemma 14. [RDS ⇒ DS] *Let* $\mathsf{ChA} : \{0,1\} \to \textsc{OutA}$ *be symmetric. If* $\mathcal{SE}$ *is separable and message linear, then* $\mathbf{Adv}^{\mathrm{ds}}(\mathcal{SE};\mathsf{ChA}) \le 2 \cdot \mathbf{Adv}^{\mathrm{rds}}(\mathcal{SE};\mathsf{ChA})$. ∎

Coming back to **ItE**, we say that $\mathsf{Inv} : \textsc{Sds} \times \{0,1\}^r \times \{0,1\}^b \to \{0,1\}^k$ is *output linear* if $\mathsf{Inv}(S,0^r,\cdot)$ is linear for all $S \in \textsc{Sds}$. Moreover, it is *separable* if $\mathsf{Inv}(S,R,M) = \mathsf{Inv}(S,R,0^b) \oplus \mathsf{Inv}(S,0^r,M)$ for all $S \in \textsc{Sds}$, $R \in \{0,1\}^r$, and $M \in \{0,1\}^b$. For example, the inverter for the above extractor based on finite-field multiplication is easily seen to be output linear and separable, by the linearity of the map $M \mapsto S^{-1} \odot M$.

Security. If we instantiate $\mathbf{ItE}[\mathsf{Inv},\mathsf{E}]$ so that Inv is both output linear and separable, and we let E be linear, the encryption function $\mathcal{SE}$ is easily seen to be message linear and separable. The following theorem now follows immediately by combining Lemma 12, Lemma 14, and Lemma 13.

Theorem 15. [DS-security of RItE] *Let* $\delta > 0$ *and* $t \ge 1$. *Also, let* $\mathsf{ChA} : \{0,1\} \to \textsc{OutA}$ *be a symmetric channel, let* $\mathsf{Inv} : \textsc{Sds} \times \{0,1\}^r \times \{0,1\}^b \to \{0,1\}^k$ *be the output-linear and separable inverter of a regular* $(k - n \cdot (\lg(|\textsc{OutA}|) - \mathbf{H}(\mathsf{ChA}) + \delta), \alpha)$*-extractor, and let* $\mathsf{E} : \{0,1\}^k \to \{0,1\}^n$ *be linear and injective. Then, for* $\mathcal{E} = \mathbf{RItE}_t[\mathsf{Inv},\mathsf{E}]$, *we have*

$$\mathbf{Adv}^{\mathrm{ds}}(\mathcal{E};\mathsf{ChA}) \le 2t \cdot \left(2 \cdot 2^{-\frac{n\delta^2}{2\lg^2(|\textsc{OutA}|+3)}} + \alpha\right) . \blacksquare$$

Instantiating the scheme. Recall that if $\mathsf{ChA} : \{0,1\}^l \to \textsc{OutA}$ and $\mathsf{ChR} : \{0,1\}^l \to \textsc{OutR}$ are symmetric channels, their secrecy capacity equals [27] $(\mathbf{H}(\mathsf{U}|\mathsf{ChA}(\mathsf{U})) - \mathbf{H}(\mathsf{U}|\mathsf{ChR}(\mathsf{U})))/l$, for a uniform l-bit U. Also, for a channel ChR, we denote its (Shannon) capacity as $\mathbf{C}(\mathsf{ChR}) = \max_{\mathsf{X}} \mathbf{I}(\mathsf{X};\mathsf{ChR}(\mathsf{X}))/l$. We will need the following result (cf. e.g. [18] for a proof).

Lemma 16. [18] *For any* $l \in \mathbb{N}$ *and any channel* $\mathsf{ChR} : \{0,1\}^l \to \textsc{OutR}$, *there is a family* $\mathsf{E} = \{\mathsf{E}_s\}_{s\in\mathbb{N}}$ *of linear encoding functions* $\mathsf{E}_s : \{0,1\}^{k(s)} \to \{0,1\}^{n(s)}$ *(where* $n(s)$ *is a multiple of* l*), with corresponding decoding functions* $\mathsf{D}_s : \textsc{OutR}^{n(s)/l} \to \{0,1\}^{k(s)}$, *such that (i)* $\mathbf{DE}(\mathsf{E}_s;\mathsf{D}_s;\mathsf{ChR}) = 2^{-\Theta(k(s))}$, *(ii)* $\lim_{s\to\infty} k(s)/n(s) = \mathbf{C}(\mathsf{ChR})$, *and (iii)* E *and* D *are PT computable.* ∎

We now derive a scheme $\overline{\mathcal{E}} = \{\mathcal{E}_s\}_{s\in\mathbb{N}}$ achieving secrecy capacity for the most common case $\mathsf{ChR} = \mathsf{BSC}_{p_R}$ and $\mathsf{ChA} = \mathsf{BSC}_{p_A}$, where $0 \le p_R < p_A \le \frac{1}{2}$. We start with a family of codes $\{\mathsf{E}_s\}_{s\in\mathbb{N}}$ for BSC_{p_R} guaranteed to exist by Lemma 16, where $E_s : \{0,1\}^{k(s)} \to \{0,1\}^{n(s)}$ and $\lim_{s\to\infty} k(s)/n(s) = 1 - h_2(p_R)$, or,

equivalently, there exists ν such that $\nu(s) = o(1)$ and $k(s) = (1 - h_2(p_R) - \nu(s)) \cdot n(s)$. Then, we let $\delta(s) = (2\lg^2(5))^{1/2} \cdot n(s)^{-1/4}$ and $\alpha(s) = 2^{-n(s)^{1/2}}$, and use the finite-field based extractor $\mathsf{Ext}_s : \{0,1\}^{k(s)} \times \{0,1\}^{k(s)} \to \{0,1\}^{b(s)}$, where $b(s) = k(s) - n(s) \cdot (1 - h_2(p_A) + \delta(s)) + 2\lg(\alpha) = (h_2(p_A) - h_2(p_R) - \nu(s) - \delta(s) - 2 \cdot n(s)^{-1/2}) \cdot n(s)$. We note that the resulting scheme is equivalent to the one described in the introduction (with $A = S^{-1}$). With these parameters,

$$\mathbf{Adv}^{\mathrm{ds}}(\mathcal{E}_s; \mathsf{BSC}_{p_A}) \leq 6 \cdot t(s) \cdot 2^{-\sqrt{n}} \ , \quad \mathbf{DE}(\mathcal{E}_s; \mathcal{D}_s; \mathsf{BSC}_{p_R}) \leq (t(s)+1) \cdot 2^{-\Theta(k(s))}$$

by Theorem 15 and Lemma 11, respectively. The rate of $\mathcal{E}_s$ is

$$\mathbf{Rate}(\mathcal{E}_s) = \frac{t(s)}{t(s)+1} \cdot \left(h_2(p_A) - h_2(p_R) - \nu(s) - \delta(s) - \frac{2}{\sqrt{n(s)}} \right) \ .$$

Setting $t(s) = \lg(k(s))$ yields $\lim_{s\to\infty} \mathbf{Rate}(\mathcal{E}_s) = h_2(p_A) - h_2(p_R)$.

Extensions. The proof applies also for any pair of symmetric channels ChR and ChA, and the resulting rate is the secrecy capacity if the capacity of $\mathsf{ChA} : \{0,1\} \to \textsc{OutA}$ is $\lg(|\textsc{OutA}|) - \mathbf{H}(\mathsf{ChA})$, which is the case if and only if a uniform input to ChA produces a uniform output. For other channels, such as *erasure channels* (where each bit is left unchanged with probability δ and mapped to an erasure symbol with probability $1 - \delta$) our technique still yields good schemes which, however, may fall short of achieving capacity. We also remark that the above presentation is constrained to single input-bit base channels for simplicity only. Our results can be extended to discrete memoryless channels with l-bit inputs for $l > 1$. For example, Lemma 13 extends to arbitrary symmetric channels $\mathsf{ChA} : \{0,1\}^l \to \textsc{OutA}$, at the price of replacing n by n/l in the security bound and in the extractor's entropy requirement. In contrast, we do not know whether Lemma 14 applies to arbitrary symmetric channels with l-bit inputs, but it does, for instance, extend to any channel of the form $\mathsf{ChA}(X) = X \oplus \mathsf{E}$, where E is an l-bit string sampled according to an input-independent noise distribution.

Acknowledgments

Bellare was supported in part by NSF grants CCF-0915675, CNS-0904380 and CNS-1116800. Tessaro was supported in part by NSF grants CCF-0915675, CCF-1018064.

This material is based on research sponsored by DARPA under agreement numbers FA8750-11-C-0096 and FA8750-11-2-0225. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA or the U.S. Government.

References

1. M. Andersson, V. Rathi, R. Thobaben, J. Kliewer, and M. Skoglund. Nested polar codes for wiretap and relay channels. Available at arxiv.org/abs/1006.3573, 2010.
2. E. Arıkan. Channel polarization: A method for constructing capacity achieving codes for symmetric binary-input memoryless channels. *IEEE Transactions on Information Theory*, 55(7):3051–3073, 2009.
3. M. Bellare, A. Desai, E. Jokipii, and P. Rogaway. A concrete security treatment of symmetric encryption. In *38th FOCS*, pages 394–403. IEEE Computer Society Press, Oct. 1997.
4. M. Bellare and S. Tessaro. Polynomial-time, semantically-secure encryption achieving the secrecy capacity, Jan. 2012. Available as arxiv.org/abs/1201.3160 and Cryptology Eprint Archive Report 2012/022.
5. M. Bellare, S. Tessaro, and A. Vardy. A cryptographic treatment of the wiretap channel, Jan. 2012. Available as arxiv.org/abs/1201.2205 and Cryptology Eprint Archive Report 2012/15.
6. M. Bloch and J. Barros. *Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge Academic Press, 2011.
7. M. Bloch and J. N. Laneman. On the secrecy capacity of arbitrary wiretap channels. In *Proceedings of the 46th Allerton Conference on Communications, Control, and Computing*, pages 818–825, Sep 2008.
8. M. Cheraghchi, F. Didier, and A. Shokrollahi. Invertible extractors and wiretap protocols. *IEEE Transactions on Information Theory*, 58(2):1254–1274, 2012.
9. G. Cohen and G. Zémor. The wiretap channel applied to biometrics. In *Proc. of the International Symposium on Information Theory and Applications*, 2004.
10. G. Cohen and G. Zémor. Syndrome coding for the wire-tap channel revisited. In *Proc. of the IEEE Information Theory Workshop (ITW '06)*, pages 33–36. IEEE, 2006.
11. T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley and Sons, 1991.
12. I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*, 2:299–318, 1967.
13. I. Csiszár. Almost independence and secrecy capacity. *Problems of Information Transmission*, 32(1):40–47, 1996.
14. I. Csiszár and J. Körner. Broadcast channels with confidential messages. *IEEE Transactions on Information Theory*, 24(3):339–348, 1978.
15. I. Damgard, T. Pedersen, and B. Pfitzmann. Statistical secrecy and multibit commitments. *IEEE Transactions on Information Theory*, 44(3):1143–1151, 1998.
16. Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *SIAM Journal on Computing*, 38(1):97–139, 2008.
17. Y. Dodis, L. Reyzin, and A. Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. In C. Cachin and J. Camenisch, editors, *EUROCRYPT 2004*, volume 3027 of *LNCS*, pages 523–540. Springer, May 2004.
18. I. Dumer. Concatenated codes and their multilevel generalizations. In *The Handbook of Coding Theory*, pages 1191–1988. Elsevier, 1998.
19. S. Goldwasser and S. Micali. Probabilistic encryption. *Journal of Computer and System Sciences*, 28(2):270–299, 1984.

20. J. Håstad, R. Impagliazzo, L. A. Levin, and M. Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999.
21. M. Hayashi and R. Matsumoto. Construction of wiretap codes from ordinary channel codes. In *Proceedings of the 2010 IEEE International Symposium on Information Theory (ISIT 2010)*, pages 2538–2542. IEEE, 2010.
22. S. Ho and R. Yeung. The interplay between entropy and variational distance. *IEEE Transactions on Information Theory*, 56(12):5906–5929, 2010.
23. E. Hof and S. Shamai. Secrecy-achieving polar-coding. In *Proceedings of the IEEE Information Theory Workshop (ITW 2010)*. IEEE, 2010.
24. ICC 2011 workshop on physical-layer security, June 2011. Kyoto, Japan.
25. M. Iwamoto and K. Ohta. Security notions for information theoretically secure encryptions. In *Proceedings of the 2011 IEEE International Symposium on Information Theory (ISIT 2011)*, pages 1777–1781. IEEE, 2011.
26. O. Koyluoglu and H. ElGamal. Polar coding for secure transmission. In *Proceedings of the IEEE International Symposium on Personal Indoor and Mobile Radio Communication*, pages 2698–2703, 2010.
27. S. Leung-Yan-Cheong. On a special class of wire-tap channels. *IEEE Transactions on Information Theory*, 23(5):625–627, 1977.
28. Y. Liang, H. Poor, and S. Shamai. Information theoretic security. *Foundations and Trends in Communications and Information Theory*, 5(4):355–580, 2008.
29. H. Mahdavifar and A. Vardy. Achieving the secrecy capacity of wiretap channels using polar codes. In *Proceedings of the 2010 IEEE International Symposium on Information Theory (ISIT 2010)*, pages 913 – 917. IEEE, 2010.
30. H. Mahdavifar and A. Vardy. Achieving the secrecy capacity of wiretap channels using polar codes. *IEEE Transactions on Information Theory*, 57(10):6428–6443, 2011.
31. U. Maurer. The strong secret key rate of discrete random triples. In R. E. Blahut, editor, *Communication and Cryptography – Two Sides of One Tapestry*, pages 271–285. Kluwer, 1994.
32. U. M. Maurer and S. Wolf. Information-theoretic key agreement: From weak to strong secrecy for free. In B. Preneel, editor, *EUROCRYPT 2000*, volume 1807 of *LNCS*, pages 351–368. Springer, May 2000.
33. J. Muramatsu and S. Miyake. Construction of wiretap channel codes by using sparse matrices. In *Proc. of the IEEE Information Theory Workshop (ITW 2009)*, pages 105–109. IEEE, 2009.
34. L. H. Ozarow and A. D. Wyner. Wire-tap channel II. In T. Beth, N. Cot, and I. Ingemarsson, editors, *EUROCRYPT'84*, volume 209 of *LNCS*, pages 33–50. Springer, Apr. 1985.
35. M. S. Pinsker. *Information and information stability of random variables and processes*. Holden Day, San Francisco, 1964.
36. R. Renner and S. Wolf. Simple and tight bounds for information reconciliation and privacy amplification. In B. K. Roy, editor, *ASIACRYPT 2005*, volume 3788 of *LNCS*, pages 199–216. Springer, Dec. 2005.
37. C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423,623–656, July, October 1948.
38. A. Suresh, A. Subramanian, A. Thangaraj, M. Bloch, and S.W.McLaughlin. Strong secrecy for erasure wiretap channels. In *Proc. of the IEEE Information Theory Workshop (ITW 2010)*. IEEE, 2010.
39. I. Tal and A. Vardy. How to construct polar codes. In *Proc. of the IEEE Information Theory Workshop (ITW 2010)*. IEEE, 2010.

40. A. Thangaraj, S. Dihidar, A. Calderbank, S. McLaughlin, and J. Merolla. Applications of LDPC codes to the wiretap channel. *IEEE Transactions on Information Theory*, 53(8):2933–2945, 2007.
41. A. D. Wyner. The wire-tap channel. *Bell Systems Tech. Journal*, 54(8):1355–1387, 1975.
42. Ziva corporation. `http://www.ziva-corp.com/`.