Maximal Correlation Secrecy

Cheuk Ting Li, *Student Member, IEEE*, and Abbas El Gamal, *Fellow, IEEE*

***Abstract*—This paper shows that the Hirschfeld-Gebelein-Rényi maximal correlation between the message and the ciphertext provides good secrecy guarantees for cryptosystems that use short keys. We first establish a bound on the eavesdropper's advantage in guessing functions of the message in terms of maximal correlation and the Rényi entropy of the message. This result implies that maximal correlation is stronger than the notion of entropic security introduced by Russell and Wang. We then show that a small maximal correlation ρ can be achieved via a randomly generated cipher with key length $\approx 2\log(1/\rho)$, independent of the message length, and by a stream cipher with key length $2\log(1/\rho)+\log n+2$ for a message of length n. We establish a converse showing that these ciphers are close to optimal. This is in contrast to entropic security for which there is a gap between the lower and upper bounds. Finally, we show that a small maximal correlation implies secrecy with respect to several mutual information based criteria but is not necessarily implied by them. Hence, maximal correlation is a stronger and more practically relevant measure of secrecy than mutual information.**

***Index Terms*—Information-theoretic secrecy, Hirschfeld-Gebelein-Rényi maximal correlation, entropic security, stream cipher, expander graph.**

I. Introduction

Consider the symmetric-key cryptosystem setting in which Alice encrypts a message (plaintext) M of length n bits using a shared secret key K of length s bits into a ciphertext C and sends it to Bob who recovers the message using the ciphertext and the key. The system is said to provide perfect secrecy if the eavesdropper Eve cannot gain *any* information about the message from the ciphertext C alone, that is, if M and C are independent. Shannon [1] showed that achieving perfect secrecy requires the key to be as long as the message. This result is considered impractical for most cryptographic applications and much shorter keys than the message are commonly used.

To analyze the secrecy of cryptosystems that use short keys, less stringent criteria than perfect secrecy have been proposed. One such criterion is to limit Eve's ability to guess functions of M, by requiring that the difference between Eve's probability of correctly guessing a function $f(M)$ of the message by a function of the ciphertext $\tilde{f}(C)$ and the maximum probability of correctly guessing $f(M)$ without knowledge of C, referred to as the *advantage* of Eve

$$\mathrm{Adv}(f,\tilde{f}) = \mathsf{P}\left\{f(M)=\tilde{f}(C)\right\} - \max_i \mathsf{P}\left\{f(M)=i\right\} \quad (1)$$

to be small. While perfect secrecy is equivalent to requiring the advantage to be less than or equal to zero for all functions of M, we show that requiring the advantage to be less than a small positive value for all functions can be satisfied by keys that are much shorter than the message. In semantic security [2], a small advantage is required with the additional restriction that Eve uses only probabilistic polynomial-time algorithms. Although satisfied by short keys, proofs of semantic security rely on unproven computational hardness assumptions. The closest work to this paper is entropic security introduced by Russell and Wang [3] and studied by Dodis and Smith [4], which requires a small advantage assuming the min-entropy of M is large. They proposed several ciphers with short keys that achieve entropic security and established lower bounds on the key length needed to achieve entropic security. Their lower bounds and achievability results do not match, however (refer to Remark 2 in Section IV for details).

In this paper, we show that the Hirschfeld-Gebelein-Rényi maximal correlation [5], [6], [7] between the message and the ciphertext, defined as

$$\rho_{\mathrm{m}}(M;C) = \max_{\substack{f(m),g(c):\,\mathsf{E}(f(M))=\mathsf{E}(g(C))=0,\\ \mathsf{E}(f^2(M))=\mathsf{E}(g^2(C))=1}} \mathsf{E}\big(f(M)g(C)\big), \quad (2)$$

is a natural measure of secrecy for ciphers with short keys. We say that a cipher achieves *ρ-maximal correlation secrecy* if $\rho_{\mathrm{m}}(M;C)\le\rho$ when M is uniformly distributed.

Ciphers achieving maximal correlation secrecy can guarantee a small advantage. If M is uniformly distributed and f is a one-bit function, e.g., one of the bits of the message, then the relationship between maximal correlation and the advantage follows readily by the work of Witsenhausen [8]. Applying the result by Calmon *et al.* [9], the advantage for uniformly distributed M and general f is upper-bounded by ρ.

Maximal correlation has numerous applications in information theory and statistics; see [10] for an overview. The work of Zhao and Chia [11] relates maximal correlation and secret key generation. Asoodeh, Alajaji and Linder [12] studied the privacy-utility tradeoff using maximal correlation as a measure of privacy. The role of maximal correlation and principal inertia components in security and privacy was investigated by Calmon *et al.* [9], [13], [14].

Several other information theoretic secrecy criteria that do not require long keys have also been proposed. In [15], Wyner proposed the wiretap channel where Eve observes a noisy version of the ciphertext. In [16], Ozarow and Wyner studied the wiretap channel setting in which Eve can choose a subset

The work of C. T. Li was supported in part by a Hong Kong Alumni Stanford Graduate Fellowship. This paper was presented in part at the IEEE International Symposium on Information Theory, Hong Kong, June 2015.

C. T. Li was with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA. He is now with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA 94720, USA (e-mail: ctli@berkeley.edu).

A. El Gamal is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA (e-mail: abbas@ee.stanford.edu).

of the ciphertext bits to observe. In such settings, secrecy criteria based on the mutual information between the message and the eavesdropper's observation (e.g., weak secrecy and strong secrecy [17], [18], [19]) are typically used. Semantic security has also been applied to wiretap channels [20] without limitations on Eve's computational power. In [21], Massey and Ingemarsson proposed a cipher with a long transmission delay which is secure assuming a memory constraint on Eve. In [22], Cachin and Maurer proposed a cipher assuming a memory constraint on Eve, but without long delay. In [23], Maurer considered the scenario where Alice, Bob and Eve observe a random source over different noisy channels. In [24], Maurer studied the case where there is a large public random source, and the number of bits in the random source that Eve can examine is limited. In [25], [26], Calmon *et al.* proposed a secrecy criterion called ϵ-symbol secrecy which limits Eve's knowledge on subsets of bits of the message. Note that the security criteria in the above works either depend on the bit structure of the message or ciphertext (e.g., noise is applied to the bits in wiretap channel, and ϵ-symbol secrecy aims at protecting subsets of bits of the message), or impose a memory constraint. In contrast, ρ-maximal correlation secrecy (like entropic security) does not depend on the bit structure of the message, guaranteeing that the cipher hides every function of the messages (not only bits) equally well from an eavesdropper with unlimited memory and computational power.

The main contributions of this paper, which is an extended and more complete version of [27], are as follows (also see Figure 1).

Rényi entropy constrained security. We bound the advantage of Eve using the maximal correlation in scenarios in which the distribution of M is not fixed and Eve may have access to some side information about the message (compared to [9] where the distribution of M is fixed to the one used to evaluate ρ). This allows us to use the same ρ to provide secrecy guarantees for a range of distributions, which is more practical since the distribution of M is often not known in practice. In Section III, we define the notion of Rényi entropy constrained security and show that a ρ-maximal correlation secure cipher also achieves a variant of semantic security with computationally unbounded adversary. In Theorem 1, we show that for non-uniform M, the advantage is upper-bounded by $2^{(n-H_2(M))/2}\rho$, where $H_2(M)$ is the Rényi entropy of M. Therefore we are able to provide secrecy guarantees for a cipher used on data with different pmfs, and even if some partial information about M is provided to Eve. We further show that a small maximal correlation secrecy implies entropic security (refer to Remark 1 in Section III for details). The proof of our result is given in Section VI-B.

Maximal correlation secrecy key length. In Section IV, we show the surprising result that ρ-maximal correlation secrecy can be achieved by short keys of length independent of the message length. We first establish a converse result showing that every ρ-maximal correlation secure cipher must have a key length $s \geq 2\log(1/\rho) - \log\left(1 + 2^{-n}\rho^{-2}\right)$ bits (Theorem 2). We then show that a cipher constructed using expander graphs can achieve ρ-maximal correlation secrecy with a key length $s = (2 + O(1/\log n))\log(1/\rho) + O(1)$ as $n \to \infty$ (Theorem 4). We further show that ρ-maximal correlation secrecy can be achieved with high probability via a randomly generated binary additive stream cipher with a key length $s = 2\log(1/\rho) + \log n + 2$ (Theorem 5). These results show that the tradeoff $s \approx 2\log(1/\rho)$ is optimal for large n. In contrast, the lower bounds on the key length for entropic security is not close to the achievable key length (refer to Remark 2 in Section IV for details). The proofs of these results are given in Section VI. For example, for a 1GB message and a 512-bits key, an advantage can be bounded by $\approx 10^{-70}$.

Relationship to other secrecy criteria. In Section V we show that ρ-maximal correlation secrecy is a stronger measure of secrecy than the notions of strong secrecy [17], [19], weak secrecy [15], and leakage rate [28], which use the mutual information between the message and the ciphertext. We show that these measures are implied by ρ-maximal correlation secrecy with suitable choices of ρ, but do not imply ρ-maximal correlation secrecy for any $\rho < 1$.

$$\begin{tikzcd}
\mathrm{Adv} & & \text{Key length } s \arrow[ll, "s \approx n - H_2(M) + 2\log\frac{1}{\mathrm{Adv}}"'] \arrow[dl, leftrightarrow, "{\begin{array}{l} s \geq \log\left(\frac{1}{\rho^2+2^{-n}}\right) \text{ (Theorem 2)} \\ s = \left(2\log\frac{1}{\rho}\right)\left(1+O(\frac{1}{\log n})\right)+O(1) \\ \text{(Theorem 4)} \end{array}}"] \\
& \rho \arrow[ul, "{\begin{array}{c} \mathrm{Adv} \leq 2^{\frac{n-H_2(M)}{2}}\rho \\ \text{(Theorem 1)} \end{array}}"] &
\end{tikzcd}$$

Figure 1. Summary of the relationship between Adv, s and ρ, given in Theorem 1, 2 and 4.

II. Definitions and Notation

Throughout this paper, we denote the joint probability matrix of X and Y by $P_{X,Y} \in \mathbb{R}^{|\mathcal{X}|\times|\mathcal{Y}|}$. We denote the spectral norm of the matrix $A \in \mathbb{R}^{m\times n}$ as

$$\|A\| = \max_{v\in\mathbb{R}^n,\, \|v\|=1} \|Av\|$$

and its Frobenius norm as $\|A\|_F$. We denote the $m \times n$ matrix consisting of all ones by $\mathbf{1}_{m\times n}$. The log function is base 2 and the entropy is measured in bits. We use the notation $[1 : n] = \{1, 2, \ldots, n\}$ and $\mathrm{Unif}(\mathcal{A})$ to be the uniform probability mass function (pmf) over a finite set $\mathcal{A}$.

We consider a cryptosystem that consists of

- a message $M \in \mathcal{M}$, where $\mathcal{M} = [1 : 2^n]$, i.e., M is an n-bit message, unless specified otherwise,
- a random secret key $K \sim \mathrm{Unif}(\mathcal{K})$, where $\mathcal{K} = [1 : 2^s]$ unless specified otherwise,
- an encryption function $E(k, m)$ that maps every pair $(k, m) \in \mathcal{K} \times \mathcal{M}$ into a ciphertext $c \in \mathcal{C}$, where $\mathcal{C} = [1 : 2^n]$ unless specified otherwise, and
- a decryption function $D(k, c)$ that maps every pair $(k, c) \in \mathcal{K} \times \mathcal{C}$ into a message $m \in \mathcal{M}$ such that $D(k, E(k, m)) = m$ for any m, k.

The pair of encryption and decryption functions (E, D) is called a (block) cipher. We assume throughout the paper that the eavesdropper knows the ciphertext C but not the message M or the key K. A cipher is said to be ρ-maximal correlation secure if $\rho_{\mathrm{m}}(M, E(K, M)) \leq \rho$ assuming that $M \sim \mathrm{Unif}(\mathcal{M})$, where ρ_{m} is as defined in (2). The encryption function can also be probabilistic. In this case, the ciphertext $C = E(K, M, W)$ is also a function of a random variable W, which is generated using the local randomness at the sender, and is unknown to the receiver and the eavesdropper. The cipher is assumed to be deterministic unless specified otherwise.

III. Rényi Entropy Constrained Security

In this section we show that every ρ-maximal correlation secure cipher satisfies a variant of semantic security. Recall that a cipher is said to be semantically secure [29] if for any pmf $p(m)$ on M, any function $f(m)$, and any partial information function $h(m)$ of the message, if M is generated according to $p(m)$, the eavesdropper who observes the ciphertext C and $h(M)$ (and also knows the choices of n, p, f and h) cannot correctly guess $f(M)$ using a probabilistic, polynomial-time algorithm with probability non-negligibly higher than the best probabilistic, polynomial-time algorithm for guessing $f(M)$ using only $h(M)$ (and also the choices of n, p, f and h). In other words, the eavesdropper cannot improve the probability of guessing $f(M)$ correctly by observing C. Note that the definition in [29] allows $p(m)$ to be a pmf on messages with different lengths. For simplicity, we consider $p(m)$ to be a pmf on messages with the same length n.

We assume $p(m)$ is a pmf on message with the same length n, and leave out the computational complexity assumptions on p, f and h here since they are not the main concern of this paper.

Rényi entropy constrained security is a variant of semantic security in which we remove the limitation on computational power but restrict the choice of the pmf $p(m)$ to have a large Rényi entropy $H_2(M)$ (or equivalently a small χ^2-divergence [30] from the uniform pmf), that is,

$$H_2(M) = -\log \sum_m (p(m))^2 \geq t$$

for some $t \geq 0$. It is formally defined below.

Definition 1. A cipher is said to be (t, ϵ)-Rényi entropy constrained secure if for any pmf $p(m)$ with $H_2(M) \geq t$, any function $f(m)$ of the message, and any eavesdropper's guess $\tilde{f}(c)$ of $f(m)$, when the message M is generated according to $p(m)$, the advantage as defined in (1) is bounded as $\mathrm{Adv}(f, \tilde{f}) \leq \epsilon$.

The min-entropy $H_\infty(M) = -\log(\max_m p(m))$ is often used in cryptography. In particular, entropic security [3] requires that the message has a high min-entropy. Rényi entropy constrained security imposes a less stringent requirement on the message since $H_2(M) \geq H_\infty(M)$, hence any message with high $H_\infty(M)$ also has a high $H_2(M)$. The case of partial information $h(m)$ will be addressed later.

We now show that maximal correlation secrecy implies Rényi entropy constrained security.

Theorem 1. *A ρ-maximal correlation secure cipher is (t, ϵ)-Rényi entropy constrained secure for any $t \geq 0$ and*

$$\epsilon = 2^{(n-t)/2}\rho.$$

Moreover, if the choice of $f(m)$ is restricted to one-bit functions $f(m) \in \{0, 1\}$, then the advantage is bounded by

$$\epsilon = 2^{(n-t)/2-1}\rho.$$

The proof of this theorem is given in Section VI-B. Note that ρ is always measured assuming the pmf of the message is uniform, though the theorem shows that ρ can also be used to bound the advantage when the actual $p(m)$ is non-uniform. Also note that the value of $n - t$ corresponds to the deviation of $p(m)$ from the uniform distribution, and can be very small when $p(m)$ is close to uniform.

The work in [9], which implies $\mathrm{Adv} \leq \rho_{\mathrm{m}}(M; C)$, also relates the advantage in guessing functions to maximal correlation (measured using the actual distribution $p(m)$). The main difference between Theorem 1 and the result in [9] is that in Theorem 1, the actual distribution $p(m)$ (where the advantage is measured) does not need to be the same as the distribution of the message used in measuring ρ (which is always set to uniform). This allows us to use the same ρ to give guarantees for a range of distributions $p(m)$.

The Rényi entropy constrained security can be extended to scenarios in which partial information $h(m)$ is available to Eve. We restrict the choices of $p(m)$ and $h(m)$ to satisfy the condition

$$\sum_a \mathsf{P}\left\{h(M) = a\right\} 2^{-H_2(M \,|\, h(M)=a)/2} \leq 2^{-\tau/2},$$

where $\tau \geq 0$ is a constant and $H_2(M \,|\, h(M) = a)$ is the Rényi entropy of the conditional pmf of M given $h(M) = a$. The eavesdropper's guess $\tilde{f}(c, h(m))$ can depend on $h(M)$, and the advantage is now defined as

$$\mathsf{P}\{f(M) = \tilde{f}(C, h(M))\} - \mathsf{E}\left(\max_i \mathsf{P}\{f(M) = i \,|\, h(M)\}\right).$$

where the second term is the probability of guessing $f(M)$ correctly using the maximum a posteriori estimation of $f(M)$ given $h(M)$. As a consequence of Theorem 1, for a ρ-maximal correlation secure cipher, the advantage is upper bounded by $2^{(n-\tau)/2}\rho$.

The value of ρ directly corresponds to the eavesdropper advantage and the correct probability of the eavesdropper's guess. For example, if the message M is uniformly distributed, (i.e., $t = n$), then the eavesdropper cannot correctly guess any one-bit function such that $\mathsf{P}\{f(M) = 1\} = 1/2$ with probability larger than $(1 + \rho)/2$. As another example, if M is uniformly distributed and l bits of M (at fixed positions) are provided to the eavesdropper via the partial information $h(m)$, then the advantage of the eavesdropper is upper bounded by $2^{l/2}\rho$.

To illustrate our results, suppose we wish to protect a message of length $n = 8 \times 10^9$ (i.e., 1GB) with a key of length $s = 512$. By Theorem 5, we can achieve ρ-maximal correlation

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2816066, IEEE Transactions on Information Theory

secrecy for $\rho = 1.54 \times 10^{-72}$ using a binary additive stream cipher. As a result, if M is uniformly distributed, then the advantage of the eavesdropper is upper bounded by 1.54×10^{-72}. If $l = 100$ bits of M are provided to the eavesdropper, then the advantage is bounded by 1.74×10^{-57}. We can see that a cipher with key length much shorter than the message length can provide good security guarantees.

Remark 1. We can show that ρ-maximal correlation secrecy implies (t, ϵ)-entropic security (as defined in [4]) for $\epsilon = 2^{(n-t)/2}\rho$. In fact the implication holds even when the min-entropy $H_\infty(M)$ in entropic security is replaced by Rényi entropy $H_2(M)$, i.e., for any pmf $p(m)$ of M with $H_2(M) \ge t$, and any function $\tilde{f}(c)$, there exists a random variable $\tilde{\tilde{F}}$ independent of M such that for any function $f(m)$,

$$\left| \mathsf{P}\{f(M) = \tilde{f}(C)\} - \mathsf{P}\{f(M) = \tilde{F}\} \right| \le \epsilon.$$

Since $H_2(M) \ge H_\infty(M)$ and the difference can be quite large if one of the messages has a high probability, maximal correlation secrecy can be much stronger than entropic security. The proof is given in Section VI-B.

IV. Maximal Correlation Secrecy Key Length

We provide bounds on the key length of a ρ-maximal correlation secure cipher in terms of ρ and the message length n. We first establish the following lower bound on the key length.

Theorem 2. *If a cipher is ρ-maximal correlation secure, then its key length is lower bounded as*

$$s \ge \log\left(\frac{1}{\rho^2 + 2^{-n}}\right).$$

The proof of this theorem is given in Section VI-C. Note that when $\rho > 0$, this lower bound can be written as

$$s \ge 2\log\frac{1}{\rho} - \log\left(1 + \frac{1}{2^n\rho^2}\right),$$

which approaches $2\log(1/\rho)$ as n tends to infinity. Also note that this bound applies to any ciphertext length (not necessarily equal to message length) and to probabilistic encryption functions.

We now consider a construction of a cipher with key length close to the lower bound using expander graphs similar to [4], [31]. Let G be a d-regular graph with vertices $[1:2^n]$ and edges $(m,\, E(k,m))$ for $m \in [1:2^n]$, $k \in [1:d]$, where E is a labeling of the edges of G. The graph may be a multigraph with multiple instances of the same edge, and we assume the graph is undirected, i.e., the number of edges (m,c) is the same as the number of edges (c,m). Further assume the labeling is invertible, i.e., there exists function $D(k,c)$ such that $D(k, E(k,m)) = m$ for all m, k. The adjacency matrix of G is given by

$$A \in \mathbb{R}^{2^n \times 2^n},\ A_{m,c} = |\{k :\, E(k,m) = c\}|.$$

Such a graph is referred to as an *expander graph* if the magnitude of the second largest eigenvalue (in absolute value) of A

$$|\lambda_2(A)| = \left\| A - \frac{d}{2^n}\mathbf{1}_{2^n\times 2^n} \right\|$$

is small. An expander graph can be constructed explicitly. For example, a (non-bipartite) Ramanujan graph [32] has a second eigenvalue

$$|\lambda_2(A)| \le 2\sqrt{d-1}.$$

Given an expander graph, we can define a corresponding *expander graph cipher* with $\mathcal{M} = \mathcal{C} = [1:2^n]$, $\mathcal{K} = [1:d]$, encryption function $E(k,m)$, and decryption function $D(k,c)$. We now find the maximal correlation for such an expander graph cipher.

Theorem 3. *The cipher defined by an expander graph with adjacency matrix A has maximal correlation*

$$\rho_{\mathrm{m}}(M;C) = \frac{1}{d}|\lambda_2(A)|.$$

As a result, the cipher corresponding to a non-bipartite Ramanujan graph is ρ-maximal correlation secure if

$$\log d \ge 2\log\frac{1}{\rho} + 2.$$

where $s = \log d$ corresponds to the key length if it is an integer.

The proof of this theorem is given in Section VI-D. It is a consequence of the characterization of maximal correlation in [8]. The relationship between maximal correlation and the second eigenvalue of a graph is also studied in [33]. A limitation of this construction is that there may not be constructions of Ramanujan graphs for a desired n and s. Using the result in [34] on the second eigenvalue of random regular graphs, we can show the existence of maximal correlation secure ciphers with key lengths close to the lower bound for any large enough n and s.

Theorem 4. *There exists a ρ-maximal correlation secure cipher with message length $n \ge 2$ and key length $s \ge 2$ if*

$$s \ge \left(2\log\frac{1}{\rho}\right)\left(1 + \frac{\alpha}{\log n}\right) + \alpha,$$

where $\alpha > 0$ is a constant.

The following corollary provides a bound on s which is independent of n.

Corollary 1. *There exists a ρ-maximal correlation secure cipher with message length $n \ge 2$ and key length $s \ge 2$ if*

$$s \ge \left(2\log\frac{1}{\rho}\right)\left(1 + \frac{3\alpha/2}{\log(\log(1/\rho)+1)}\right),$$

where $\alpha > 0$ is a constant.

The proofs of Theorem 4 and Corollary 1 are in Section VI-E. This corollary shows that for any ρ, a key length which depends only on ρ is sufficient to achieve ρ-maximal correlation secrecy for any message length. This is in a strong contrast to perfect secrecy, which requires the key length to be at least the message length.

Maximal correlation secrecy can also be achieved by a simpler cipher with a sightly longer key length. Consider a *binary additive stream cipher* $\mathcal{M} = \mathcal{C} = \{0,1\}^n$, $\mathcal{K} =$

$\{0,1\}^s$, $E(k,m) = m \oplus g(k)$, $D(k,c) = c \oplus g(k)$, where $g(k) = (g_1(k), g_2(k), \ldots, g_n(k)) \in \{0,1\}^n$ is the *keystream generator* and $\oplus$ is component-wise $\bmod 2$ addition. The following theorem shows that most binary additive stream ciphers with slightly longer key than the lower bound in Theorem 2 are ρ-maximal correlation secure.

Theorem 5. *Let $G_i(k)$, $i \in [1:n]$, $k \in [1:2^s]$ be i.i.d.* $\mathrm{Bern}(1/2)$ *random keystream components. Let $\rho > 0$, $\epsilon > 0$, then*

$$\mathsf{P}\{\rho_{\mathrm{m}}(M;\, M \oplus G(K)) \le \rho\} > 1 - \epsilon,$$

where the randomness of $\rho_{\mathrm{m}}(M; M \oplus G(K))$ is induced by the random keystream generator, if the key length

$$s \ge 2\log\frac{1}{\rho} + \log n + \log\left(1 + \frac{1}{n}\log\frac{1}{\epsilon}\right) + 2.$$

The proof of this theorem is given in Section VI-F. Substituting $\epsilon = 1$ in the theorem shows that there exists a binary additive stream cipher that is ρ-maximal correlation secure with a key length

$$s \ge 2\log\frac{1}{\rho} + \log n + 2.$$

Hence for a constant $\rho > 0$, a key size of around $\log n$ is sufficient.

Figure IV plots the lower bound on the key length in Theorem 2, the key length achievable by the expander graph cipher using the Ramanujan graphs in Theorem 3, and the key length achievable by the random stream cipher in Theorem 5 versus ρ for $n = 10000$.

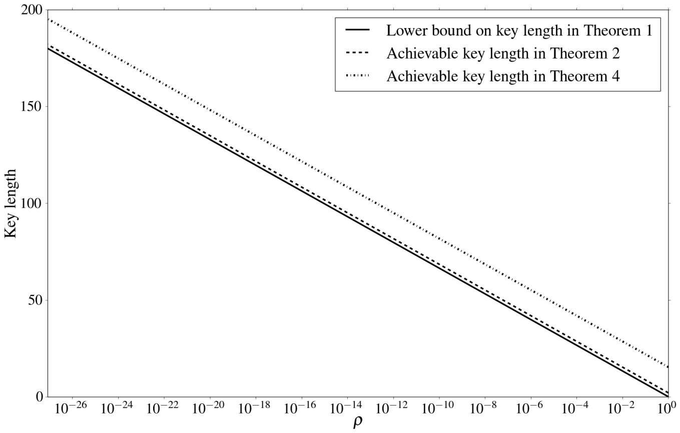

Figure 2. Comparison of the lower bound and the achievable key lengths for $n = 10000$.

Remark 2. By the achievability result in Theorem 3 where $s = 2\log(1/\rho) + 2$, there exist ciphers which achieve (t,ϵ)-Rényi entropy constrained security with key length

$$s = n - t + 2\log\frac{1}{\epsilon} + 2.$$

This is the same key length required by entropic security in [4] (Corollary 3.3). Since (t,ϵ)-Rényi entropy constrained security implies (t,ϵ)-entropic security, it is possible to show the achievability result of entropic security via maximal correlation.

Also note that unlike maximal correlation secrecy where there are tight upper and lower bounds on the optimal key length $s \approx 2\log(1/\rho)$ for large n, the lower bounds on the key length given in [4]:

$$s \ge n - t$$

for any entropically secure cipher, and

$$s \ge n - t + \log\frac{1}{\epsilon} - O(1)$$

for public-coin schemes, are much smaller than the achievable key length. It is conjectured in [4] that the achievable key length is tight.

V. Relationship to other Secrecy Criteria

In this section, we compare maximal correlation secrecy to other secrecy criteria that use the mutual information $I(M;C)$ between the message M and the ciphertext C. We show that ρ-maximal correlation secrecy is stronger than these other secrecy criteria as they are implied by ρ-maximal correlation secrecy with suitable choices of ρ, but they do not imply ρ-maximal correlation secrecy for any $\rho < 1$.

As pointed out by Maurer [17], mutual information is too loose a criterion if it is not required to approach zero because it does not guarantee hiding any bit (or more generally any function) of the message. For an n bit message and an s-bit key the mutual information is lower bounded as $I(M;C) \ge n - s$ and this bound can be achieved simply by encrypting the most significant s bits of M perfectly and sending the rest of the bits in the clear. Hence unless $s \ge n$, mutual information cannot guarantee that any bit of M is not leaked. There are other ciphers that can achieve the same minimum mutual information but ensure better secrecy. For example, suppose we wish to encrypt a 2-bit message $M \sim \mathrm{Unif}(\{0,1,2,3\})$ using a 1-bit key $K \sim \mathrm{Bern}(1/2)$. Let $C_1 = M + 2K \mod 4$ be the cipher which encrypts only the most significant bit of the message and $C_2 = M + K \mod 4$ be another cipher that achieves $I(M;C) = 1$. It is easy to see that using C_2, Eve cannot correctly guess any bit of the message with probability greater than $3/4$ signifying that C_2 provides better secrecy than C_1. Moreover the second cipher achieves lower maximal correlation $\rho_{\mathrm{m}}(M;C_2) = \sqrt{2}/2$ than the first $\rho_{\mathrm{m}}(M;C_1) = 1$, hence maximal correlation is a better measure of secrecy than mutual information.

We now formalize the relationship between maximal correlation secrecy and the mutual information secrecy measures of strong secrecy [17], weak secrecy [15], and leakage rate [28]. By the following proposition which follows from Theorem 5 in [35], a ρ-maximal correlation secrecy guarantees a small mutual information.

Proposition 1. *Let X and Y be two discrete random variables, then*

$$I(X;Y) \le \log\left((\min\{|\mathcal{X}|, |\mathcal{Y}|\} - 1) \cdot \rho_{\mathrm{m}}^2(X;Y) + 1\right). \quad (3)$$

In the following we assume that $M \sim \mathrm{Unif}([1:2^n])$, which reduces (3) to

$$I(M;C) \le \log\left((2^n - 1)\rho_{\mathrm{m}}^2(M;C) + 1\right). \quad (4)$$

We now use the above proposition to compare ρ-maximal correlation secrecy to secrecy criteria that use the mutual information.

Strong secrecy. This criterion requires that $\lim_{n\to\infty} I(M;C) = 0$. From (4) this is implied by ρ-maximal correlation secrecy for

$$\rho = o(2^{-n/2}).$$

Weak secrecy. This criterion requires that $\lim_{n\to\infty} I(M;C)/n = 0$. From (4), this is implied by ρ-maximal correlation secrecy for

$$\rho = 2^{-n/2+o(n)}.$$

Leakage rate. Note that both weak and strong secrecy require the key rate $\lim_n s/n = 1$. By requiring that $\lim_n I(M;C)/n \le R_L$ for some leakage rate R_L, a key rate of $1-R_L$ can be achieved. From (4), this is implied by ρ-maximal correlation secrecy for

$$\rho = 2^{-(1-R_L)n/2+o(n)}.$$

Note that Theorem 5 implies that such ρ can be achieved also by a key rate of $1-R_L$. Hence maximal correlation secrecy provides a better security guarantee than leakage rate with no penalty on the key rate.

The above results show that ρ-maximal correlation secret implies secrecy criteria involving mutual information. We now show that a small $I(M;C)$ does not necessarily imply ρ-maximal correlation secrecy. Consider the following cipher: Let $\mathcal{M} = \mathcal{C} = \mathcal{K} = [0 : 2^n - 1]$, and the encryption and decryption functions be

$$E(k,m) = \begin{cases} m+k \mod 2^n - 1 & \text{if } m < 2^n - 1 \\ 2^n - 1 & \text{if } m = 2^n - 1, \end{cases}$$

and $D(k,c) = E(-k,c)$. Direct computation yields $I(M;C) = 2^{-n}(n+2-2^{-(n-1)})$, which goes to zero as n tends to infinity, and thus the cipher satisfies strong secrecy. However, since one can determine if $M = 2^n - 1$ or not by observing C, $\rho_{\mathrm{m}}(M;C) = 1$. Hence ρ-maximal correlation secrecy is a strictly stronger secrecy criterion than criteria that use mutual information.

VI. Proof of the Results

A. Properties of Maximal Correlation

We use a characterization of maximal correlation in terms of the spectral norm that follows directly from the singular value characterization of maximal correlation in [8].

Lemma 1. *Let $(X,Y) \sim p(x,y)$ be discrete random variables with marginals $p(x)$ and $p(y)$. Define the matrix $B \in \mathbb{R}^{|\mathcal{X}|\times|\mathcal{Y}|}$ with entries*

$$B_{xy} = \frac{p(x,y)}{\sqrt{p(x)p(y)}} - \sqrt{p(x)p(y)}.$$

Then,

$$\rho_{\mathrm{m}}(X;Y) = \|B\|.$$

Next we state a result relating maximal correlation and the χ^2-divergence between the joint pmf and the product of the marginal pmfs, also known as χ^2 measure of correlation which follows directly from [9].

Lemma 2. *Let $(X,Y) \sim p(x,y)$ be discrete random variables with marginals $p(x)$ and $p(y)$. Then,*

$$\frac{1}{\min\{|\mathcal{X}|,|\mathcal{Y}|\}-1} \le \frac{\rho_{\mathrm{m}}^2(X;Y)}{\chi^2\left(p(x,y)\,\|\,p(x)p(y)\right)} \le 1,$$

where

$$\chi^2(p(x)\|q(x)) = \sum_{\mathrm{m}} \frac{(p(x))^2}{q(x)} - 1$$

is the χ^2-divergence between $p(x)$ and $q(x)$.

B. Proof of Theorem 1

We prove Theorem 1, which shows that maximal correlation secrecy implies Rényi entropy constrained security. First prove the case for general functions $f(m)$. Note that it is implied by the following more general result.

Proposition 2. *Consider any two pmfs $p_M(m)$ and $\hat{p}_M(m)$ on $\mathcal{M}$, and a Markov kernel $p_{C|M}(c|m)$. The two pmfs induce the joint probability measures P and $\hat{\mathsf{P}}$ on (M,C), respectively. Let $\rho_{\mathrm{m}}(M;C)$ be the maximal correlation in P. For any functions $f:\mathcal{M}\to\mathbb{N}$ and $\tilde{f}:\mathcal{C}\to\mathbb{N}$, we have*

$$\begin{aligned}&\left|\hat{\mathsf{P}}\left\{f(M)=\tilde{f}(C)\right\} - \sum_i \hat{\mathsf{P}}\left\{f(M)=i\right\}\cdot\mathsf{P}\{\tilde{f}(C)=i\}\right| \\ &\le \rho_{\mathrm{m}}(M;C)\sqrt{\chi^2\left(\hat{p}_M\|p_M\right)+1}.\end{aligned}$$

Proof: All expectations, variances and covariances in this proof are in P. Assume the range of $f(m)$ is $\{1,...,l\}$. Let $Z_1,...,Z_l$ be i.i.d. Rademacher random variables. Let $g(m) = Z_{f(m)}\hat{p}_M(m)/p_M(m)$ and $\tilde{g}(c) = Z_{\tilde{f}(c)}$. Write $\chi^2 = \chi^2\left(\hat{p}_M\|p_M\right)$, $\hat{p}_f(i) = \hat{\mathsf{P}}\left\{f(M)=i\right\}$, $p_{\tilde{f}}(i) = \mathsf{P}\{\tilde{f}(C)=i\}$ and $\hat{p}_{eq} = \hat{\mathsf{P}}\{f(M)=\tilde{f}(C)\}$. Observe that

$$\mathsf{E}(g(M)\mid Z_1^l) = \sum_i \hat{p}_f(i)Z_i,$$

$$\mathsf{E}(\tilde{g}(C)\mid Z_1^l) = \sum_i p_{\tilde{f}}(i)Z_i,$$

$$\begin{aligned}\mathsf{E}\left((g(M))^2\mid Z_1^l\right) &= \sum_m p_M(m)\left(\frac{Z_{f(m)}\hat{p}_M(m)}{p_M(m)}\right)^2 \\ &= \chi^2+1,\end{aligned}$$

$$
\begin{aligned}
&\mathsf{E}\left(\mathrm{Cov}\big(g(M),\, \tilde{g}(C) \mid Z_1^l\big)\right) \\
&= \mathsf{E}\left(\mathsf{E}\left(g(M)\tilde{g}(C) \mid Z_1^l\right) - \mathsf{E}(g(M) \mid Z_1^l)\,\mathsf{E}(\tilde{g}(C) \mid Z_1^l)\right) \\
&= \mathsf{E}\Bigg(\sum_m p(m)g(m) \sum_c p(c|m)\tilde{g}(c) \\
&\qquad\qquad - \Big(\sum_i \hat{p}_f(i) Z_i\Big)\Big(\sum_i p_{\tilde{f}}(i) Z_i\Big)\Bigg) \\
&= \mathsf{E}\Bigg(\sum_m \hat{p}_M(m) \sum_c p(c|m) Z_{\tilde{f}(c)} Z_{f(m)} \\
&\qquad\qquad - \Big(\sum_i \hat{p}_f(i) Z_i\Big)\Big(\sum_i p_{\tilde{f}}(i) Z_i\Big)\Bigg) \\
&= \hat{p}_{eq} - \sum_i \hat{p}_f(i) p_{\tilde{f}}(i).
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{E}\left(\left|\mathrm{Cov}\big(g(M),\, \tilde{g}(C) \mid Z_1^l\big)\right|\right) &\geq \left|\mathsf{E}\left(\mathrm{Cov}\big(g(M),\, \tilde{g}(C) \mid Z_1^l\big)\right)\right| \\
&= \Big|\hat{p}_{eq} - \sum_i \hat{p}_f(i) p_{\tilde{f}}(i)\Big|.
\end{aligned}
$$

Hence there exists constant z_1^l such that

$$
\left|\mathrm{Cov}\big(g(M),\, \tilde{g}(C) \mid Z_1^l = z_1^l\big)\right| \geq \Big|\hat{p}_{eq} - \sum_i \hat{p}_f(i) p_{\tilde{f}}(i)\Big|.
$$

We have

$$
\begin{aligned}
\rho_{\mathrm{m}}(M;C) &\geq \frac{\left|\mathrm{Cov}\big(g(M),\, \tilde{g}(C) \mid Z_1^l = z_1^l\big)\right|}{\sqrt{\mathrm{Var}(g(M) \mid Z_1^l = z_1^l)\mathrm{Var}(\tilde{g}(C) \mid Z_1^l = z_1^l)}} \\
&\geq \frac{\left|\mathrm{Cov}\big(g(M),\, \tilde{g}(C) \mid Z_1^l = z_1^l\big)\right|}{\sqrt{\mathsf{E}((g(M))^2 \mid Z_1^l = z_1^l)\,\mathsf{E}((\tilde{g}(C))^2 \mid Z_1^l = z_1^l)}} \\
&\geq \frac{\left|\hat{p}_{eq} - \sum_i \hat{p}_f(i) p_{\tilde{f}}(i)\right|}{\sqrt{\chi^2+1}}.
\end{aligned}
$$

The result follows. ■

To obtain Theorem 1, take P to be the uniform distribution on M and $\hat{\mathsf{P}}$ to be the actual distribution. The result follows from $\sum_i \hat{\mathsf{P}}\left\{f(M) = i\right\} \cdot \mathsf{P}\{\tilde{f}(C) = i\} \leq \max_i \hat{\mathsf{P}}\left\{f(M) = i\right\}$ and

$$
\chi^2\left(p_M \| \mathrm{Unif}[1:2^n]\right) = 2^{n - H_2(M)} - 1.
$$

To show the result in Remark 1, take $\tilde{F}$ to be a random variable in the probability space $\hat{\mathsf{P}}$ independent of M with $\hat{\mathsf{P}}\{\tilde{F} = i\} = \mathsf{P}\{\tilde{f}(C) = i\}$, then the result follows from

$$
\hat{\mathsf{P}}\{f(M) = \tilde{F}\} = \sum_i \hat{\mathsf{P}}\left\{f(M) = i\right\} \cdot \mathsf{P}\{\tilde{f}(C) = i\}.
$$

Then we prove a slightly better result for one-bit functions $f(m)$. Note that it is implied by the following more general result.

Proposition 3. *Consider any two pmfs $p_M(m)$ and $\hat{p}_M(m)$ on $\mathcal{M}$, and a Markov kernel $p_{C|M}(c|m)$. The two pmfs induce the joint probability measures P and $\hat{\mathsf{P}}$ on (M, C), respectively. Let $\rho_{\mathrm{m}}(M;C)$ be the maximal correlation in P. For any one-bit functions $f : \mathcal{M} \to \{0,1\}$ and $\tilde{f} : \mathcal{C} \to \{0,1\}$, we have*

$$
\begin{aligned}
&\hat{\mathsf{P}}\left\{f(M) = \tilde{f}(C)\right\} - \frac{1}{2} \\
&\leq \Bigg(\frac{1}{4}\rho_{\mathrm{m}}^2(M;C)\left(\chi^2\left(\hat{p}_M \| p_M\right) + 1\right) \\
&\qquad + \left(1 - \rho_{\mathrm{m}}^2(M;C)\right)\left(\hat{\mathsf{P}}\left\{f(M) = 0\right\} - \frac{1}{2}\right)^2\Bigg)^{1/2}.
\end{aligned}
$$

Proof: All expectations, variances and covariances in this proof are in P. Let $g(m) = (-1)^{f(m)}\hat{p}_M(m)/p_M(m)$ and $\tilde{g}(c) = (-1)^{\tilde{f}(c)}$. Write $\chi^2 = \chi^2\left(\hat{p}_M \| p_M\right)$, $\hat{p}_f(i) = \hat{\mathsf{P}}\left\{f(M) = i\right\}$, $p_{\tilde{f}}(i) = \mathsf{P}\{\tilde{f}(C) = i\}$ and $\hat{p}_e = \hat{\mathsf{P}}\{f(M) \neq \tilde{f}(C)\}$. It is straightforward to check that Proposition 3 is true if $\hat{p}_e \geq \min\{\hat{p}_f(0), \hat{p}_f(1)\}$. Hence we assume $\hat{p}_e < \min\{\hat{p}_f(0), \hat{p}_f(1)\}$. Observe that

$$
\begin{aligned}
\mathsf{E}(g(M)) &= \hat{p}_f(0) - \hat{p}_f(1), \\
\mathsf{E}(\tilde{g}(C)) &= p_{\tilde{f}}(0) - p_{\tilde{f}}(1),
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}(g(M)) &= \sum_m \frac{\left(\hat{p}_M(m)\right)^2}{p_M(m)} - \left(\hat{p}_f(0) - \hat{p}_f(1)\right)^2 \\
&= (\chi^2 + 1) - \left(\hat{p}_f(0) - \hat{p}_f(1)\right)^2, \\
\mathrm{Var}(\tilde{g}(C)) &= 1 - \left(p_{\tilde{f}}(0) - p_{\tilde{f}}(1)\right)^2,
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Cov}\big(g(M),\, \tilde{g}(C)\big) \\
&= \mathsf{E}\left(g(M)\tilde{g}(C)\right) - \mathsf{E}(g(M))\,\mathsf{E}(\tilde{g}(C)) \\
&= \sum_m p(m)g(m) \sum_c p(c|m)\tilde{g}(c) \\
&\quad - \big(\hat{p}_f(0) - \hat{p}_f(1)\big)\big(p_{\tilde{f}}(0) - p_{\tilde{f}}(1)\big) \\
&= 1 - 2\hat{p}_e - \left(\hat{p}_f(0) - \hat{p}_f(1)\right)\left(p_{\tilde{f}}(0) - p_{\tilde{f}}(1)\right).
\end{aligned}
$$

We have

$$
\begin{aligned}
&\rho_{\mathrm{m}}^2(M;C) \\
&\geq \frac{\left(\mathrm{Cov}\big(g(M), \tilde{g}(C)\big)\right)^2}{\mathrm{Var}\big(g(M)\big)\mathrm{Var}\big(\tilde{g}(C)\big)} \\
&= \frac{\left(1 - 2\hat{p}_e - \big(\hat{p}_f(0) - \hat{p}_f(1)\big)\big(p_{\tilde{f}}(0) - p_{\tilde{f}}(1)\big)\right)^2}{\left((\chi^2 + 1) - \left(\hat{p}_f(0) - \hat{p}_f(1)\right)^2\right)\left(1 - \left(p_{\tilde{f}}(0) - p_{\tilde{f}}(1)\right)^2\right)} \\
&\geq \frac{\left(1 - 2\hat{p}_e\right)^2 - \left(1 - 2\hat{p}_f(0)\right)^2}{(\chi^2 + 1) - \left(1 - 2\hat{p}_f(0)\right)^2},
\end{aligned}
$$

where the last inequality is due to $\hat{p}_e < \min\{\hat{p}_f(0), \hat{p}_f(1)\}$ and

$$
\frac{(a - bx)^2}{1 - x^2} = \left(\frac{a}{\sqrt{1 - x^2}} - \frac{bx}{\sqrt{1 - x^2}}\right)^2 \geq a^2 - b^2
$$

for any a, b such that $|a| > |b|$ and $-1 < x < 1$. Hence,

$$
\begin{aligned}
&\left(1 - 2\hat{p}_e\right)^2 \\
&\leq \rho_{\mathrm{m}}^2(M;C)\left((\chi^2 + 1) - \left(1 - 2\hat{p}_f(0)\right)^2\right) + \left(1 - 2\hat{p}_f(0)\right)^2 \\
&= \rho_{\mathrm{m}}^2(M;C)(\chi^2 + 1) + \left(1 - \rho_{\mathrm{m}}^2(M;C)\right)\left(1 - 2\hat{p}_f(0)\right)^2.
\end{aligned}
$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2816066, IEEE Transactions on Information Theory

The result follows. ■

To prove Theorem 1 for one-bit functions $f(m)$, note that by Proposition 3,

$$\begin{aligned}
&\hat{\mathsf{P}}\Big\{f(M)=\tilde{f}(C)\Big\}-\frac{1}{2}\\
&\le \bigg(\frac{1}{4}\rho_{\mathrm{m}}^2(M;C)\left(\chi^2\left(\hat{p}_M\|p_M\right)+1\right)\\
&\qquad + \left(1-\rho_{\mathrm{m}}^2(M;C)\right)\left(\hat{\mathsf{P}}\left\{f(M)=0\right\}-\frac{1}{2}\right)^2\bigg)^{1/2}\\
&\le \bigg(\frac{1}{4}\rho_{\mathrm{m}}^2(M;C)\left(\chi^2\left(\hat{p}_M\|p_M\right)+1\right)\\
&\qquad + \left(\hat{\mathsf{P}}\left\{f(M)=0\right\}-\frac{1}{2}\right)^2\bigg)^{1/2}\\
&\le \sqrt{\frac{1}{4}\rho_{\mathrm{m}}^2(M;C)\left(\chi^2\left(\hat{p}_M\|p_M\right)+1\right)}\\
&\qquad + \sqrt{\left(\hat{\mathsf{P}}\left\{f(M)=0\right\}-\frac{1}{2}\right)^2}\\
&= \frac{1}{2}\rho_{\mathrm{m}}(M;C)\sqrt{\chi^2\left(\hat{p}_M\|p_M\right)+1}\\
&\qquad + \max\Big\{\hat{\mathsf{P}}\left\{f(M)=0\right\},\,\hat{\mathsf{P}}\left\{f(M)=1\right\}\Big\}-\frac{1}{2}.
\end{aligned}$$

This completes the proof.

C. Proof of Theorem 2

Here we assume the ciphertext length is arbitrary, and the encryption function can be randomized, i.e., the ciphertext is $C=E(K,M,W)$ where $W\sim p(w)$ is the local randomness at the sender. We require that $D(k,E(k,m,w))=m$ for any k,m,w. From Lemma 2,

$$\begin{aligned}
\rho^2 &\ge \rho_{\mathrm{m}}^2(M;C)\\
&\ge 2^{-n}\chi^2\left(p(m,c)\,\|\,p(m)p(c)\right)\\
&= 2^{-n}\left(\sum_{m,c}\frac{(p(m,c))^2}{p(m)p(c)}-1\right)\\
&= \sum_{m,c}\frac{(p(m,c))^2}{p(c)}-2^{-n}\\
&= \sum_{c}\frac{\sum_m (p(m,c))^2}{p(c)}-2^{-n}\\
&\ge \sum_{c}\frac{|\{m:\,p(m,c)>0\}|^{-1}\left(\sum_m p(m,c)\right)^2}{p(c)}-2^{-n}\\
&= \mathsf{E}\left(|\{m:\,p(m,C)>0\}|^{-1}\right)-2^{-n}\\
&\ge \mathsf{E}\left(|\{D(k,C):\,k\in[1:2^s]\}|^{-1}\right)-2^{-n}\\
&\ge 2^{-s}-2^{-n}.
\end{aligned}$$

Hence,

$$\begin{aligned}
s &\ge \log\left(\frac{1}{\rho^2+2^{-n}}\right)\\
&= 2\log\frac{1}{\rho}-\log\left(1+\frac{1}{2^n\rho^2}\right).
\end{aligned}$$

This completes the proof.

D. Proof of Theorem 3

Theorem 3 is a direct consequence of Lemma 1. Since $M,C\sim\mathrm{Unif}[1:2^n]$, we have

$$\begin{aligned}
\rho_{\mathrm{m}}(M;C) &= 2^n\left\|P_{MC}-2^{-2n}\mathbf{1}_{2^n\times 2^n}\right\|\\
&= 2^n\left\|\frac{1}{d\cdot 2^n}A-2^{-2n}\mathbf{1}_{2^n\times 2^n}\right\|\\
&= \frac{1}{d}\left\|A-\frac{d}{2^n}\mathbf{1}_{2^n\times 2^n}\right\|\\
&= \frac{1}{d}|\lambda_2(A)|.
\end{aligned}$$

Ramanujan graphs have second eigenvalue $|\lambda_2(A)|\le 2\sqrt{d-1}$, hence their maximal correlation is

$$\begin{aligned}
\rho_{\mathrm{m}}(M;C) &\le \frac{2\sqrt{d-1}}{d}\\
&\le \frac{2}{\sqrt{d}}.
\end{aligned}$$

As a result, if $\log d\ge 2\log(1/\rho)+2$, we have $\rho_{\mathrm{m}}(M;C)\le\rho$.

E. Proofs of Theorem 4 and Corollary 1

We first prove a lemma on the cascade of two ciphers with the same message length n but with possibly different key lengths s_1 and s_2, which yields a cipher with message length n and key length s_1+s_2.

Lemma 3. *Let (E_1,D_1) and (E_2,D_2) be two ciphers with key lengths s_1 and s_2, respectively, and the same message length n. Define the cascade of these two ciphers to be the cipher $\mathcal{K}=[1:2^{s_1}]\times[1:2^{s_2}]$, $\mathcal{M}=\mathcal{C}=[1:2^n]$,*

$$\begin{aligned}
E(k_1,k_2,m) &= E_2\left(k_2,E_1(k_1,m)\right),\\
D(k_1,k_2,m) &= D_1\left(k_1,D_2(k_2,m)\right).
\end{aligned}$$

Then we have

$$\begin{aligned}
&\rho_{\mathrm{m}}\left(M;\,E(K_1,K_2,M)\right)\\
&\quad\le \rho_{\mathrm{m}}\left(M;\,E_1(K_1,M)\right)\cdot\rho_{\mathrm{m}}\left(M;\,E_2(K_2,M)\right).
\end{aligned}$$

Proof: Consider the following alternate characterization of maximal correlation in [7]

$$\rho_{\mathrm{m}}(X;Y)=\max_{f(x):\,\mathsf{E}(f(X))=0,\,\mathsf{E}(f^2(X))=1}\sqrt{\mathsf{E}\Big(\left(\mathsf{E}\left(f(X)\,|\,Y\right)\right)^2\Big)}.$$

Let $M_1\sim\mathrm{Unif}[1:2^n]$, $M_2=E_1(K_1,M_1)$, $C=E_2\left(K_2,M_2\right)$. Note that for any $f,g:[1:2^n]\to\mathbb{R}$ with $\mathsf{E}(f(M_1))=\mathsf{E}(g(C))=0$, $\mathsf{E}(f^2(M_1))=\mathsf{E}(g^2(C))=1$, by the alternate characterization,

$$\begin{aligned}
&\mathsf{E}\left(f(M_1)g(C)\right)\\
&= \mathsf{E}\left(\mathsf{E}\left(f(M_1)\,|\,M_2\right)\cdot\mathsf{E}\left(g(C)\,|\,M_2\right)\right)\\
&\le \sqrt{\mathsf{E}\left(\left(\mathsf{E}\left(f(M_1)\,|\,M_2\right)\right)^2\right)\cdot\mathsf{E}\left(\left(\mathsf{E}\left(g(C)\,|\,M_2\right)\right)^2\right)}\\
&\le \rho_{\mathrm{m}}\left(M_1;M_2\right)\cdot\rho_{\mathrm{m}}\left(C;M_2\right).
\end{aligned}$$

The result follows. ■

Now consider the following result from [34]. Let $A_1,\ldots,A_d\in\mathbb{R}^{N\times N}$ be i.i.d. random permutation matrices

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2816066, IEEE Transactions on Information Theory

uniformly distributed in the set of permutations of $\{1, \dots, N\}$. Then we have

$$\begin{aligned}&\mathsf{E}\left(\left|\lambda_2\left(\sum_{i=1}^{d}\left(A_i+A_i^T\right)\right)\right|\right)\\&\le 2\sqrt{2d-1}\left(1+\frac{\ln d}{\sqrt{2d}}+O\left(d^{-1/2}\right)\right)+O\left(\frac{d^{3/2}\ln\ln N}{\ln N}\right).\end{aligned} \tag{5}$$

We use the above result to construct a cipher as follows. Generate $d = 2^{s-1}$ permutations on $[1 : 2^n]$, namely $\sigma_1, \dots, \sigma_d$, i.i.d. uniformly at random. Let $\sigma_{i+d} = \sigma_i^{-1}$ for $i = 1, \dots, d$. The cipher is defined as $\mathcal{K} = [1 : 2^s]$, $\mathcal{M} = \mathcal{C} = [1 : 2^n]$, $E(k, m) = \sigma_k(m)$, $D(k, c) = \sigma_k^{-1}(c)$. By Lemma 1,

$$\begin{aligned}\rho_{\mathrm{m}}(M;C) &= \left\|2^n P_{MC} - \frac{1}{2^n}\mathbf{1}_{2^n\times 2^n}\right\|\\&= \left\|\frac{1}{2d}\sum_{i=1}^{d}\left(A_i+A_i^T\right) - \frac{1}{2^n}\mathbf{1}_{2^n\times 2^n}\right\|\\&= \frac{1}{2d}\left|\lambda_2\left(\sum_{i=1}^{d}\left(A_i+A_i^T\right)\right)\right|.\end{aligned}$$

Hence by (5), there exist fixed $\sigma_1, \dots, \sigma_d$ and a constant $\eta > 0$ (that does not depend on s or n) such that

$$\begin{aligned}&\rho_{\mathrm{m}}(M;C)\\&\le \frac{1}{2d}\left(2\sqrt{2d-1}\left(1+\frac{\ln d}{\sqrt{2d}}+\eta d^{-1/2}\right)+\eta\cdot\frac{d^{3/2}\log n}{n}\right)\\&\le \frac{2}{\sqrt{2d}}\left(1+\frac{\ln d}{\sqrt{2d}}+\eta\left(d^{-1/2}+\frac{d\log n}{n}\right)\right)\\&\le \frac{4}{\sqrt{2d}}\\&= 2^{-s/2+2}\end{aligned} \tag{6}$$

if $d \ge 16\eta^2$ and $n/\log n \ge 4\eta d$, or equivalently,

$$2\log\eta + 5 \le s \le \log n - \log\log n - \log\eta - 1. \tag{7}$$

Note that this construction only works for very short key lengths. We now provide a construction for general key length s by the cascade of several ciphers with short key lengths. Let

$$t = \left\lceil\frac{s}{\log n - \log\log n - \log\eta - 2}\right\rceil,\ \tilde{s} = \left\lfloor\frac{s}{t}\right\rfloor,$$

$$a = t\left(\left\lfloor\frac{s}{t}\right\rfloor + 1\right) - s,\ b = s - t\left\lfloor\frac{s}{t}\right\rfloor,$$

then we have $a + b = t$ and $s = a\tilde{s} + b(\tilde{s} + 1)$. Consider the cascade of a ciphers with key length $\tilde{s}$ and b ciphers of key length $\tilde{s} + 1$, which gives a cipher with key length s. Let s_0 be an integer satisfying

$$s_0 \ge \max\{4\log\eta + 12\,,\ 2^{20}\}$$

and

$$\log s_0 - \log\log s_0 \ge 5\log\eta + 14.$$

Consider any $s \ge s_0$. If $n \le s$, then perfect secrecy can be achieved. Hence we assume $n > s \ge s_0$. To check the conditions in (7) for $\tilde{s}$ and $\tilde{s} + 1$,

$$\begin{aligned}\tilde{s} &= \left\lfloor\frac{s}{\left\lceil s\left(\log n - \log\log n - \log\eta - 2\right)^{-1}\right\rceil}\right\rfloor\\&\ge \frac{s}{s\left(\log n - \log\log n - \log\eta - 2\right)^{-1} + 1} - 1\\&= \frac{1}{\left(\log n - \log\log n - \log\eta - 2\right)^{-1} + s^{-1}} - 1\\&\ge \frac{1}{\left(4\log\eta + 12\right)^{-1} + \left(4\log\eta + 12\right)^{-1}} - 1\\&= 2\log\eta + 5.\end{aligned}$$

And also

$$\begin{aligned}\tilde{s} + 1 &= \left\lfloor\frac{s}{\left\lceil s\left(\log n - \log\log n - \log\eta - 2\right)^{-1}\right\rceil}\right\rfloor + 1\\&\le \log n - \log\log n - \log\eta - 1.\end{aligned}$$

Hence by (6) and Lemma 3, the maximal correlation of the resultant cipher is

$$\begin{aligned}\rho_{\mathrm{m}}(M;C) &\le \left(2^{-\tilde{s}/2+2}\right)^a\left(2^{-(\tilde{s}+1)/2+2}\right)^b\\&= 2^{-s/2+2t},\end{aligned}$$

where

$$\begin{aligned}2t &= 2\left\lceil\frac{s}{\log n - \log\log n - \log\eta - 2}\right\rceil\\&\le \frac{2s}{\log n - \log\log n - \log\eta - 2} + 2\\&\le \frac{2s}{\log n - \log\log n - \left(\log n - \log\log n\right)/5} + 2\\&= \frac{5s/2}{\log n - \log\log n} + 2\\&\le \frac{4s}{\log n} + 2,\end{aligned}$$

where the last inequality is due to $\log n \ge \log s_0 \ge 14$. Therefore,

$$2\log\frac{1}{\rho_{\mathrm{m}}(M;C)} \ge s\left(1 - \frac{8}{\log n}\right) - 4.$$

Rearranging, we have

$$\begin{aligned}s &\le \left(2\log\frac{1}{\rho_{\mathrm{m}}(M;C)} + 4\right)\left(1 - \frac{8}{\log n}\right)^{-1}\\&\le \left(2\log\frac{1}{\rho_{\mathrm{m}}(M;C)}\right)\left(1 + \frac{16}{\log n}\right) + 16.\end{aligned}$$

Hence if

$$s \ge \left(2\log\frac{1}{\rho}\right)\left(1 + \frac{\alpha}{\log n}\right) + \alpha.$$

where $\alpha = \max\{16, s_0\}$, then $s \ge s_0$, and $\rho_{\mathrm{m}}(M;C) \le \rho$. This completes the proof of Theorem 4.

Now we prove Corollary 1. If

$$s \geq \left(2\log\frac{1}{\rho}\right)\left(1+\frac{3\alpha/2}{\log(\log(1/\rho)+1)}\right),$$

then

$$\begin{aligned}\log\frac{1}{\rho}+1 &\leq \frac{s}{2}\left(1+\frac{3\alpha/2}{\log(\log(1/\rho)+1)}\right)^{-1}+1\\ &\leq \frac{s}{2}+1\\ &\leq s\end{aligned}$$

due to the assumption that $s \geq 2$. Hence,

$$\begin{aligned}s &\geq \left(2\log\frac{1}{\rho}\right)\left(1+\frac{3\alpha/2}{\log(\log(1/\rho)+1)}\right)\\ &= \left(2\log\frac{1}{\rho}\right)\left(1+\frac{\alpha}{\log(\log(1/\rho)+1)}\right)+\frac{\alpha\log(1/\rho)}{\log(\log(1/\rho)+1)}\\ &\geq \left(2\log\frac{1}{\rho}\right)\left(1+\frac{\alpha}{\log(\log(1/\rho)+1)}\right)+\alpha\\ &\geq \left(2\log\frac{1}{\rho}\right)\left(1+\frac{\alpha}{\log s}\right)+\alpha\\ &\geq \left(2\log\frac{1}{\rho}\right)\left(1+\frac{\alpha}{\log n}\right)+\alpha,\end{aligned}$$

where the last step is due to the assumption that $n > s$. This complete the proof of Corollary 1.

F. Proof of Theorem 5

We first compute the maximal correlation of a binary additive stream cipher. The following proposition follows by Fourier analysis of Boolean functions [36]; see [13].

Proposition 4. *A binary additive stream cipher has a maximal correlation*

$$\rho_{\mathrm{m}}(M;C) = \max_{v\in\{0,1\}^n\backslash 0^n}\left|\frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}}(-1)^{\sum_{l=0}^{n-1} v_l G_l(k)}\right|.$$

We now proceed to prove Theorem 5. Assume we generate $G_i(k)$ i.i.d. $\mathrm{Bern}(1/2)$ across k and i. For each fixed $v \neq 0^n$, consider

$$\frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}}(-1)^{\sum_{l=0}^{n-1} v_l G_l(k)}.$$

The terms $(-1)^{\sum_{l=0}^{n-1} v_l G_l(k)}$ are i.i.d. Rademacher. By the Chernoff bound,

$$\mathsf{P}\left\{\left|\frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}}(-1)^{\sum_{l=0}^{n-1} v_l G_l(k)}\right| \geq \rho\right\} \leq 2^{1-|\mathcal{K}|\cdot D_{\mathrm{KL}}\left(\frac{1+\rho}{2}\,\middle\|\,\frac{1}{2}\right)}.$$

By the union bound on all possible $v \in \{0,1\}^n\backslash 0^n$ and observing that $(\ln 2)\, D_{\mathrm{KL}}\left(\frac{1+\rho}{2}\,\middle\|\,\frac{1}{2}\right) > \rho^2/2$ for $\rho > 0$,

$$\begin{aligned}\mathbb{P}\left\{\rho_{\mathrm{m}}(M;\, C) \geq \rho\right\} &\leq 2^{(n+1)-|\mathcal{K}|\cdot D_{\mathrm{KL}}\left(\frac{1+\rho}{2}\,\middle\|\,\frac{1}{2}\right)}\\ &< 2^{(n+1)-|\mathcal{K}|\rho^2/(2\ln 2)}.\end{aligned}$$

Hence if

$$s \geq 2\log\frac{1}{\rho}+\log n+\log\left(1+\frac{1}{n}\log\frac{1}{\epsilon}\right)+2,$$

then

$$\begin{aligned}2^s &\geq 4\rho^{-2}n\left(1+\frac{1}{n}\log\frac{1}{\epsilon}\right)\\ &\geq 4\,(\ln 2)\,\rho^{-2}\left(n+\log\frac{1}{\epsilon}\right).\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{P}\left\{\rho_{\mathrm{m}}(M;\, C) \geq \rho\right\} &< 2^{(n+1)-2^s\rho^2/(2\ln 2)}\\ &\leq 2^{(n+1)-2(n-\log\epsilon)}\\ &\leq 2^{1-n+\log\epsilon}\\ &\leq \epsilon.\end{aligned}$$

This completes the proof of Theorem 5.

VII. Acknowledgments

The authors would like to acknowledge the invaluable input by the anonymous reviewers and the associate editor, which has greatly improved the presentation of our results and their relationships with earlier works.

References

[1] C. E. Shannon, "Communication theory of secrecy systems," *Bell Syst. Tech. J.*, vol. 28, no. 4, pp. 656–715, 1949.

[2] S. Goldwasser and S. Micali, "Probabilistic encryption," *Journal of computer and system sciences*, vol. 28, no. 2, pp. 270–299, 1984.

[3] A. Russell and H. Wang, "How to fool an unbounded adversary with a short key," in *Advances in Cryptology-EUROCRYPT 2002*. Springer, 2002, pp. 133–148.

[4] Y. Dodis and A. Smith, "Entropic security and the encryption of high entropy messages," in *Theory of Cryptography*. Springer, 2005, pp. 556–577.

[5] H. O. Hirschfeld, "A connection between correlation and contingency," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 31, no. 04. Cambridge Univ Press, 1935, pp. 520–524.

[6] H. Gebelein, "Das statistische problem der korrelation als variations-und eigenwertproblem und sein zusammenhang mit der ausgleichsrechnung," *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 21, no. 6, pp. 364–379, 1941.

[7] A. Rényi, "On measures of dependence," *Acta mathematica hungarica*, vol. 10, no. 3, pp. 441–451, 1959.

[8] H. S. Witsenhausen, "On sequences of pairs of dependent random variables," *SIAM J. Appl. Math.*, vol. 28, no. 1, pp. 100–113, 1975.

[9] F. P. Calmon, M. Varia, M. Médard, M. M. Christiansen, K. R. Duffy, and S. Tessaro, "Bounds on inference," in *Proc. 51st Ann. Allerton Conf. Commun., Contr., and Comput.*, Oct. 2013, pp. 567–574.

[10] V. Anantharam, A. A. Gohari, S. Kamath, and C. Nair, "On maximal correlation, hypercontractivity, and the data processing inequality studied by erkip and cover," *CoRR*, vol. abs/1304.6133, 2013. [Online]. Available: http://arxiv.org/abs/1304.6133

[11] L. Zhao and Y.-K. Chia, "The efficiency of common randomness generation," in *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*, Sept 2011, pp. 944–950.

[12] S. Asoodeh, F. Alajaji, and T. Linder, "On maximal correlation, mutual information and data privacy," in *Information Theory (CWIT), 2015 IEEE 14th Canadian Workshop on*. IEEE, 2015, pp. 27–31.

[13] F. P. Calmon, M. Varia, and M. Médard, "An exploration of the role of principal inertia components in information theory," in *Proc. IEEE Inf. Theory Workshop*, Nov. 2014.

[14] F. P. Calmon, A. Makhdoumi, and M. Médard, "Fundamental limits of perfect privacy," in *2015 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2015, pp. 1796–1800.

[15] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, 1975.

[16] L. H. Ozarow and A. D. Wyner, "Wire-tap channel—II," in *Advances in cryptology (Paris, 1984)*, ser. Lecture Notes in Comput. Sci. Berlin: Springer, 1985, vol. 209, pp. 33–50.

[17] U. M. Maurer, "The strong secret key rate of discrete random triples," in *Communications and Cryptography*. Springer, 1994, pp. 271–285.

[18] I. Csiszár, "Almost independence and secrecy capacity," *Problemy Peredachi Informatsii*, vol. 32, no. 1, pp. 48–57, 1996.

[19] U. Maurer and S. Wolf, "Information-theoretic key agreement: From weak to strong secrecy for free," in *Advances in Cryptology-EUROCRYPT 2000*. Springer, 2000, pp. 351–368.

[20] M. Bellare, S. Tessaro, and A. Vardy, "Semantic security for the wiretap channel," in *Advances in Cryptology–CRYPTO 2012*. Springer, 2012, pp. 294–311.

[21] J. L. Massey and I. Ingemarsson, "The Rip van Winkle cipher–a simple and provably computationally secure cipher with a finite key," in *IEEE International Symposium on Information Theory (Abstracts)*, 1985, p. 146.

[22] C. Cachin and U. Maurer, "Unconditional security against memory-bounded adversaries," in *Advances in Cryptology-CRYPTO'97*. Springer, 1997, pp. 292–306.

[23] U. M. Maurer, "Perfect cryptographic security from partially independent channels," in *Proc. 23rd Annual ACM Symp. Theory of Computing*. ACM, 1991, pp. 561–571.

[24] ——, "Conditionally-perfect secrecy and a provably-secure randomized cipher," *Journal of Cryptology*, vol. 5, no. 1, pp. 53–66, 1992.

[25] F. P. Calmon, M. Médard, L. M. Zeger, J. Barros, M. M. Christiansen, and K. R. Duffy, "Lists that are smaller than their parts: A coding approach to tunable secrecy," in *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*. IEEE, 2012, pp. 1387–1394.

[26] F. P. Calmon, M. Médard, M. Varia, K. R. Duffy, M. M. Christiansen, and L. M. Zeger, "Hiding symbols and functions: New metrics and constructions for information-theoretic security," *arXiv preprint arXiv:1503.08513*, 2015.

[27] C. T. Li and A. El Gamal, "Maximal correlation secrecy," *Proc. IEEE Symp. Info. Theory*, 2015.

[28] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 339–348, 1978.

[29] O. Goldreich, *Foundations of Cryptography: Volume 2, Basic Applications*. Cambridge University Press, 2009, vol. 2.

[30] K. Pearson, "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *Philosophical Magazine Series 5*, vol. 50, no. 302, pp. 157–175, 1900.

[31] M. Cheraghchi, F. Didier, and A. Shokrollahi, "Invertible extractors and wiretap protocols," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1254–1274, 2012.

[32] A. Lubotzky, R. Phillips, and P. Sarnak, "Ramanujan graphs," *Combinatorica*, vol. 8, no. 3, pp. 261–277, 1988.

[33] M. Bolla and G. Molnar-Saska, "Optimization problems for weighted graphs and related correlation estimates," *Discrete Mathematics*, vol. 282, no. 1-3, pp. 23 – 33, 2004.

[34] J. Friedman, "On the second eigenvalue and random walks in randomd-regular graphs," *Combinatorica*, vol. 11, no. 4, pp. 331–362, 1991.

[35] A. L. Gibbs and F. E. Su, "On choosing and bounding probability metrics," *International statistical review*, vol. 70, no. 3, pp. 419–435, 2002.

[36] R. O'Donnell, *Analysis of boolean functions*. Cambridge University Press, 2014.

Cheuk Ting Li (S'12) received the B.Sc. degree in mathematics and B.Eng. degree in information engineering from The Chinese University of Hong Kong in 2012, and the M.S. and Ph.D. degree in electrical engineering from Stanford University in 2014 and 2018 respectively. He is currently a postdoctoral scholar at the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley. His research interests include generation of random variables, one-shot schemes in information theory, wireless communications and information-theoretic secrecy.

Abbas El Gamal (S'71-M'73-SM'83-F'00) is the Hitachi America Professor in the School of Engineering at Stanford University. He received his B.Sc. Honors degree from Cairo University in 1972, and his M.S. in Statistics and Ph.D. in Electrical Engineering both from Stanford University in 1977 and 1978, respectively. From 1978 to 1980, he was an Assistant Professor of Electrical Engineering at USC. From 2003 to 2012, he was the Director of the Information Systems Laboratory at Stanford University. From 2012-2017, he was the Fortinet Founders Chair of the Department of Electrical Engineering. His research contributions have been in network information theory, FPGAs, and digital imaging devices and systems. He has authored or coauthored over 230 papers and holds 35 patents in these areas. He is coauthor of the book Network Information Theory (Cambridge Press 2011). He is a member of the US National Academy of Engineering and a Fellow of the IEEE. He received several honors and awards for his research contributions, including the 2016 IEEE Richard Hamming Medal, the 2014 Viterbi Lecture, the 2013 Shannon Memorial Lecture, the 2012 Claude E. Shannon Award, the inaugural Padovani Lecture, and the 2004 INFOCOM Paper Award. He served on the Board of Governors of the Information Theory Society from 2009 to 2016 and was President in 2014.