Group secret key agreement over state-dependent wireless broadcast channels

M. Jafari Siavoshani[†] S. Mishra[‡] S. N. Diggavi[‡] C. Fragouli[†]
[†]Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
Email: {mahdi.jafarisiavoshani,christina.fragouli}@epfl.ch
[‡]University of California (UCLA), Los Angeles, USA
Email: {shaunakmishra,suhasdiggavi}@ucla.edu

***Abstract*—We consider a group of m trusted nodes that aim to create a shared secret key $\mathcal{K}$, using a state-dependent wireless broadcast channel that exists from one of the honest nodes to the rest of the nodes including a passive eavesdropper Eve. All of the trusted nodes can also discuss over a cost-free and unlimited rate public channel which is also observed by Eve. For this setup, we develop an information-theoretically secure secret key agreement protocol. We show the optimality of this protocol for linear deterministic wireless broadcast channels as well as in the high-SNR regime for wireless channels with large dynamic range over channel states.**

I. Introduction

Secret key agreement protocols between a pair of nodes, allowing for unlimited public discussion, was initiated in seminal work of Maurer [1]. Group key agreement with public discussion was developed in [4], where the key agreement rate was established when the trusted users had access to a private broadcast channel. The case when the eavesdropper also had access to the broadcast channel was the main focus of recent work in [7] which developed lower and outer bounds for secrecy rates.

In this work we initiate the study of group secret key agreement over a state dependent Gaussian broadcast channel. This can be motivated by fading wireless channels, where the channel states vary over time. The use of state-dependent channels for secrecy has been of interest recently (see for example [8] and references therein). To gain insight into our problem, we first investigate a deterministic approximation of the wireless channel as defined in [5]. For the deterministic channel we will show that using a superposition based secrecy scheme [9], we can develop a group key agreement protocol that can be shown to be information-theoretically optimal. In particular, we show that we can get the same key agreement rate for the entire group as we would get for a single pair of nodes. Therefore this result demonstrates that with unlimited public discussion, we get secret key-agreement rates for linear deterministic channels, that is invariant to network size.

To the best of our understanding, the scheme proposed in [7] simulates source model and in order to create a secret key requires the "communication for omniscience by a neutral observer", *i.e.,* a neutral observer can reconstruct all the sources if given the eavesdropper output and the shared common randomness. However, our scheme seems to be different since we do not need such a condition.

We use the deterministic achievability scheme to get an insight about the wireless broadcast channel with state. To this end, we use a nested message set, degraded channel wiretap code based on the broadcast approach of [9] to develop a key-agreement protocol for the noisy broadcast problem. This enables a scheme that converts the wireless channel with state to behave similar to the deterministic case. Though this is not optimal, we can demonstrate that when there is a large dynamic range between the channel states, this scheme is optimal in the degrees of freedom sense.

The paper is organized as follows. In §II we introduce our notation and the problem formulation. In §III we state a general upper bound for the key generation capacity and give an achievability scheme for the Gaussian channel discussed in §IV.

II. Notation and Setup

A. Notation

We use uppercase letters (e.g., X) to represent random variables (or more generally random objects). Given random variables $X_1, \ldots, X_m$, we write $X_{1:m}$ to denote $(X_1, \ldots, X_m)$. We use also $X^{t_0:t}$ to denote $(X[t_0], \ldots, X[t])$ where t is the discrete time index. When $t_0 = 1$ we simply write X^n to denote $(X[1], \ldots, X[t])$.

All vectors are column vectors unless otherwise stated. Bold capital letters (e.g., $\boldsymbol{A}$) are reserved for deterministic matrices.

For convenience, we use $[i : j]$ to denote $\{i, i+1, \ldots, j\}$ where $i, j \in \mathbb{Z}$. Let $\mathsf{Uni}(\mathcal{M})$ denote the uniform distribution over the set $\mathcal{M}$. For example, we use $\mathsf{Uni}(\mathbb{F}_q^\ell)$ to denote the uniform distribution over vectors of length ℓ that are defined over finite field $\mathbb{F}_q$.

We abuse the notation $H(\cdot)$ to denote both entropy and differential entropy depending on the context. During the paper, all the logarithms are in base two.

The work of M. Jafari Siavoshani and C. Fragouli was supported by Swiss National Science Foundation Award No. PP00P2-128639.

B. Problem Statement

In our problem setup, we consider $m \geq 2$ terminals $\mathsf{T}_0, \ldots, \mathsf{T}_{m-1}$, which we will for convenience call "Alice," "Bob," "Calvin," etc. Alice can broadcast information to the remaining nodes using a state dependent broadcast additive white Gaussian channel. The goal of her broadcast transmissions is to establish a secret key $\mathcal{K}$ among all the m terminals, in the presence of a passive eavesdropper, Eve. All terminals can also utilize a cost-free public channel to send information to each other.

In out state dependent channel model, we assume that for each channel use the state remains the same for a block of symbols of length L and changes independently from one block to another block. We also assume that L is large enough that enables us to apply information theoretical arguments within each block, for noisy channels. The transmitted vector sent by Alice is denoted by $X \in \mathbb{R}^L$. The received vectors for every terminal and Eve depend on their channel states for the particular time instant. We define a random variable $S_{\mathsf{T}_i} \in [0 : s]$ corresponding to the state of the channel for the ith terminal and similarly define the random variable $S_{\mathsf{E}} \in [0 : s]$ for Eve. For the channel state of a receiver $r \in \{\mathsf{T}_0, \ldots, \mathsf{T}_{m-1}, \mathsf{E}\}$ we assume that[1]

$$\mathbb{P}[S_r = k] = \delta_k, \qquad k \in [0 : s].$$

Then we model the received vector by the receiver r by a state dependent white Gaussian channel as follows[2]

$$\widetilde{X}_r[t] = h_{S_r[t]} X_0[t] + Z_r[t], \tag{1}$$

where $\widetilde{X}_r[t] \in \mathbb{R}^L$ and $Z_r[t] \in \mathbb{R}^L$. For the additive noise of each receiver we have $Z_r[t] \sim \mathcal{N}(0, I_L)$. The channel gains h_i are some real constants such that

$$h_0 \leq \cdots \leq h_s.$$

We also assume that the channel input is subject to an average power constraint P, i.e.,

$$\frac{1}{L}\mathbb{E}\left[\|X_0\|^2\right] \leq P.$$

Moreover, we assume that the channel state information (CSI) is completely known by each receiver. So we define a composite received vector for the receiver r as follows

$$X_r[t] = (\widetilde{X}_r[t], S_r[t]).$$

III. Preliminary Results

In this section we discuss some preliminary results which are the foundation of our achievability scheme for the secret key sharing over the Gaussian broadcast channel introduced in (1).

[1]For simplicity, we consider a symmetric problem where the probability distribution over the states is the same for all of the receivers (including Eve). Moreover, we focus on a finite number of states. Both these restrictions can be relaxed.

[2]During the paper, we use T_i and i interchangeably when they are used as subscript. So instead of X_{T_i} we sometimes write X_i. At some points, we also use X_{A} and X_{B} to denote for X_0 and X_1.

A. The General Upper Bound

As explained in [6, Section IV] combining the results of [3] and [4] we can find an upper bound for the key generation capacity for the cases where the channels from Alice to the other terminals are independent, as follows

Theorem 1 ([6, Theorem 3])**.** *The key generation capacity of a multiterminal secret key sharing problem*[3] *can be upper bounded as follows*[4]

$$C_s \leq \sup_{p(x_0)} \min_{j \in [1:m-1]} I(X_0, X_j | X_{\mathsf{E}}).$$

Interpretation: Theorem 1 states that the key generation capacity of a multi-terminal problem is upper bounded by the best pairwise key generation upper bound between Alice and the other honest terminals.

B. Deterministic Broadcast Channel

In this section we consider a channel model which is the deterministic model for Gaussian channels as defined in [5]. The transmitted vector sent by Alice is denoted by $X_0 \in \mathbb{F}_q^L$.

The received vectors for every terminal and Eve depend on their channel states for the particular time instant. We define a random variable $S_{\mathsf{T}_i} \in [0 : s]$ corresponding to the state of the channel for the ith terminal and similarly define the random variable $S_{\mathsf{E}} \in [0 : s]$ for Eve. For the channel state of a receiver $r \in \{\mathsf{T}_0, \ldots, \mathsf{T}_{m-1}, \mathsf{E}\}$ we write

$$\mathbb{P}[S_r = k] = \delta_k, \qquad k \in [0 : s].$$

We then model the received vector by the receiver r by a state dependent deterministic channel as follows

$$\widetilde{X}_r[t] = \boldsymbol{F}_{S_r[t]} X_0[t], \tag{2}$$

where $\boldsymbol{F}_i \in \mathbb{F}_q^{L \times L}$ for $i \in [0 : s]$. Moreover, we assume that the state of a particular channel is available at the corresponding receiver. We define a composite received vector for the receiver r as follows

$$X_r = (\widetilde{X}_r, S_r).$$

In order to capture and model the different SNR level for the Gaussian channel we use the deterministic matrix model developed in [5]. This implies that the matrices $\boldsymbol{F}_i$ have the following nested structure, capturing the degraded broadcast structure:

$$\mathbf{0} = \ker \boldsymbol{F}_s \subset \ker \boldsymbol{F}_{s-1} \subset \cdots \subset \ker \boldsymbol{F}_0 = \mathbb{F}_q^L, \quad \text{and} \tag{3}$$

$$\operatorname{rank}(\boldsymbol{F}_i - \boldsymbol{F}_{i-1}) = \operatorname{rank}(\boldsymbol{F}_i) - \operatorname{rank}(\boldsymbol{F}_{i-1}). \tag{4}$$

For convenience we assume that $\boldsymbol{F}_s = \boldsymbol{I}_L$.

We can then state the following result, Theorem 2.

[3]For a precise definition of the "secret key generating protocol" and its corresponding "key generation capacity" we refer to [1], [2], [4], [7].

[4]For the Gaussian version of our problem, we can adapt the result to include an additional power constraint.

Theorem 2. *The key generation capacity of the deterministic broadcast channel introduced in §III-B is given by*

$$C_s = \sum_{j=1}^{s} [\operatorname{rank} \boldsymbol{F}_j - \operatorname{rank} \boldsymbol{F}_{j-1}] \, (\sum_{i=0}^{j-1} \rho_i) \log q,$$

where $\rho_i \triangleq \delta_i - 2\delta_i(\delta_0 + \cdots + \delta_{i-1}) - \delta_i^2$.

Proof: 1) Upper Bound: Using Theorem 1 we can upper bound the secrecy capacity. For the proof see the Appendix.

2) Lower Bound (Achievability Scheme): Because of (3) we can find subspaces $\Pi_1, \ldots, \Pi_s$, such that $\Pi_i \cap \Pi_j = \mathbf{0}$ and

$$\begin{aligned} \Pi_1 \oplus \ker \boldsymbol{F}_1 &= \mathbb{F}_q^L, \\ \Pi_2 \oplus \Pi_1 \oplus \ker \boldsymbol{F}_2 &= \mathbb{F}_q^L, \\ &\vdots \\ \Pi_s \oplus \cdots \oplus \Pi_1 \oplus \ker \boldsymbol{F}_s &= \mathbb{F}_q^L. \end{aligned} \tag{5}$$

Then for $i \in [1:s]$ we have $\dim \Pi_i = \operatorname{rank} \boldsymbol{F}_i - \operatorname{rank} \boldsymbol{F}_{i-1}$.

In our proposed achievability scheme, Alice uses superposition coding to create the vector

$$X_{\mathsf{A}}[t] = X_{\mathsf{A}1}[t] + \cdots + X_{\mathsf{A}s}[t], \tag{6}$$

such that $X_{\mathsf{A}i}[t] \in \Pi_i$. Because of (5), $\{\Pi_i\}$ form a basis for $\mathbb{F}_q^L$ so every vector $X_{\mathsf{A}}[t] \in \mathbb{F}_q^L$ can be uniquely decomposed as (6). Now each $X_{\mathsf{A}i}[t] \in \Pi_i$ can be considered as a vector that is transmitted by Alice and will be received independently by each trusted terminal and Eve with erasure probability $\theta_i \triangleq \sum_{j=0}^{i-1} \delta_i$ (the vector $X_{\mathsf{A}i}[t]$ is correctly received by the rth receiver only if $S_r \geq i$).

So we may view the broadcast channel from Alice to the rest of terminals as s independent *packet erasure channels*; where Π_i is the set of messages transmitted over the ith channel (layer) and the erasure probability of the ith channel is θ_i.

We then proceed as follows. On each layer k we run the group secrecy scheme proposed in [10] independently from the other layers. The group secrecy rate for the erasure channel secrecy with unlimited public discussion is given in [6] and is summarized in Lemma 1.

Lemma 1 ([6, Theorem 1])**.** *The achievable group secret key generation rate for an packet erasure broadcast channel with unlimited public discussion is given by*

$$R = (1-\theta)\theta\ell \log q,$$

where θ is the symmetric erasure probability for all nodes and $\ell \log q$ is the packet size.

So by applying the above scheme for each layer and using Lemma 1, for the total achievable secrecy rate we have

$$R_s = \sum_{k=1}^{s} R_k = \sum_{k=1}^{s} [\operatorname{rank} \boldsymbol{F}_k - \operatorname{rank} \boldsymbol{F}_{k-1}] \, (\sum_{i=0}^{k-1} \rho_i) \log q,$$

because $(1-\theta_k)\theta_k = \sum_{i=0}^{k-1} \rho_i$ and $\ell_k = \dim(\Pi_k) = \operatorname{rank} \boldsymbol{F}_k - \operatorname{rank} \boldsymbol{F}_{k-1}$. ∎

This result can be easily extended to the asymmetric case where the channels to the legitimate users are not statistically identical. Moreover, notice that in the symmetric case, the key-generation rate is the same for any $m \geq 2$, and therefore this protocol scales ideally with the network size. Finally, the critical difference between $m = 2$ and $m > 2$ is that the key-reconciliation used ideas from network coding [10].

IV. Gaussian Broadcast Channel

A. Upper Bound

Here, we apply the result of Theorem 1 to obtain an upper bound for the key generation capacity as follows in Theorem 3.

Theorem 3. *The key generation capacity of the Gaussian broadcast channel given in* (1) *using public discussions is upper bounded as follows*

$$C_s \leq \sum_{i=0}^{s} \sum_{j=0}^{s} \delta_i \delta_j \log \left(1 + \frac{h_i^2 P}{1 + h_j^2 P}\right).$$

Proof: See Appendix. ∎

B. Lower Bound (Achievability Scheme)

Before giving the achievability scheme, let us define a nested message set, degraded channel wiretap scenario as follows

Definition 1. *Assume a wiretap channel scenario where there is a transmitter called Alice that broadcasts X and there are $s+1$ receivers Y_i where the ith receiver receives Y_i according to the broadcast channel $(\mathcal{X}, p(y_0, \ldots, y_s|x), \mathcal{Y}_0 \times \cdots \times \mathcal{Y}_s))$ such that*

$$p(y_0, \ldots, y_s|x) = p(y_s|x) \cdot p(y_{s-1}|y_s) \cdots p(y_0|y_1).$$

Suppose that Alice has s messages $W_1, \ldots, W_s$ where $W_i \in \{1, \ldots, 2^{LR_i}\}$ and $W_i \sim \mathsf{Uni}(1 : 2^{LR_i})$. The goal is that she wants to broadcast these messages such that $\forall i$:
(i) *each message W_i should be decodable by the receivers $Y_i, \ldots, Y_s$ with a negligible error probability, and*
(ii) *all the receivers $Y_0, \ldots, Y_{i-1}$ should be ignorant about the message W_i, namely for the leakage rate we have*

$$R_{l,i}^{(L)} \triangleq \frac{1}{L} I(W_{i+1}, \ldots, W_s; Y_i^{1:L}) \leq \epsilon_L, \forall i \in [0:s].$$

Then we can state the following result.

Theorem 4. *Using a properly designed layered wiretap code similar to [9] we can achieve the following set of rates for the nested message set, degraded Gaussian wiretap channel.*

$$R_i = \log \left(1 + \frac{h_i^2 P_i}{1 + h_i^2 I_i}\right) - \log \left(1 + \frac{h_{i-1}^2 P_i}{1 + h_{i-1}^2 I_i}\right), \tag{7}$$

$\forall i \in [1:s]$, where $I_i \triangleq \sum_{j=i+1}^{s} P_j$.

By using a layered coding scheme for the nested message set, degraded channel wiretap channel defined in Definition 1, we can convert the Gaussian channel given in (1) to a set of s independent erasure channels where the erasure of the messages for each channel (layer) depends on the receiver

channel state. In fact using the layered coding scheme for the wiretap channel, we mimic the orthogonality behavior that we have for the deterministic channel as described by (3) and (5).

More precisely, we assume that Alice broadcasts the L-length vector $X_{\mathsf{A}}[t] = \sum_{i=1}^{s} X_{\mathsf{A}i}[t]$, where she maps W_i (the messages corresponding to the ith layer) to $X_{\mathsf{A}i}[t]$ according to the codebook described in the following. We construct s codebooks $\hat{\mathcal{C}}_i(2^{L\hat{R}_i}, L)$ each contains $2^{L\hat{R}_i}$ codewords $X_{\mathsf{A}i}^L$ by choosing L symbols independently from the Gaussian distribution $\mathcal{N}(0, P_i)$ where

$$\hat{R}_i = \log\left(1 + \frac{h_i^2 P_i}{1 + h_i^2 I_i}\right).$$

Each codebook $\hat{\mathcal{C}}_i$, $i \in [1:s]$, is divided into 2^{LR_i} bins where R_i is given by (7). At each layer i, the message W_i is coded so as to be secure from all receivers in states $j < i$. This is done by a standard wiretap code (see also [9]), where the message W_i is the bin-index and the transmit sequence $X_{\mathsf{A}i}$ is a (random) sequence from the the bin. So, the ith layer can transmits 2^{LR_i} messages securely from the "weaker" receivers. Following a similar argument as stated in [9], we can show that the receiver r which observes the channel state $S_r = i$ can decode messages up to layer i and is ignorant about messages of layers above i. So, equivalently, we can say that the message W_i experiences erasure probability $\theta_i = \sum_{j=0}^{i-1} \delta_j$ when it passes through the channel (1).

Now for each layer i, we run the interactive scheme of §III-B where Alice broadcasts a sequence of random messages W_i^n. Then, by discussing over the public channel, the trusted terminals reconcile their secret messages to build a common key. The key generation rate for each layer is $\Delta_i R_i$ so for a fixed power allocation we achieve the following secrecy rate

$$R_s \le \sum_{i=1}^{s} \Delta_i R_i,$$

where R_i is defined in (7) and $\Delta_i \triangleq (1-\theta_i)\theta_i$.

The maximum secrecy rate is obtained by optimizing the above rate over the power allocations $\{P_i\}$. So we can write

$$R_s = \left\{ \begin{array}{ll} \max & \sum_{i=1}^{s} \Delta_i R_i \\ \text{subject to} & \sum_{i=1}^{s} P_i \le P \\ & P_i \ge 0, \quad \forall i \in [1:s]. \end{array} \right. \tag{8}$$

Because R_1 is an increasing function of P_1 when other P_i are kept fixed and R_i does not depend on P_1 for $i > 1$ we can write the power constant inequality as an equality.

The important special case of this optimization problem is when there is a large dynamic range between the channel states, and we focus on this case applied to the high SNR regime. This enables us to demonstrate an optimal power allocation for this regime in Section IV-C.

C. High SNR Regime

By large dynamic range in the states, we mean that $h_i \gg h_{i-1}$, $\forall i \in [1:s]$, where this comparison is done with respect to SNR. In particular, we denote $h_i^2 = \mathrm{SNR}^{-\alpha_i}$, for $i \in [0:s]$, where $h_i^2 > h_{i-1}^2$ implying that $\alpha_i < \alpha_{i-1}$. Suppose we apply a power allocation which is given as follows, $P_i = \mathrm{SNR}^{\beta_i}$, $\forall i \in [1:s]$, with the assumption that $\beta_i < \beta_{i-1}$. Since $\beta_i < \beta_{i-1}$, in high SNR regime I_i is dominated by $\mathrm{SNR}^{\beta_{i+1}}$. Using this approximation, we can rewrite the expression for R_i from (7) as follows

$$R_i \doteq \log\left(1 + \frac{\mathrm{SNR}^{\beta_i - \alpha_i}}{1 + \mathrm{SNR}^{\beta_{i+1} - \alpha_i}}\right) - \log\left(1 + \frac{\mathrm{SNR}^{\beta_i - \alpha_{i-1}}}{1 + \mathrm{SNR}^{\beta_{i+1} - \alpha_{i-1}}}\right),$$

that simplifies to

$$R_i \doteq \Big[\big((\beta_i - \alpha_i) - (\beta_{i+1} - \alpha_i)^+\big)^+\Big] \log \mathrm{SNR} - \Big[\big((\beta_i - \alpha_{i-1}) - (\beta_{i+1} - \alpha_{i-1})^+\big)^+\Big] \log \mathrm{SNR},$$

where we use the notation "$\doteq$" and "$\dot{\le}$" for exponential equality and inequality with respect to SNR. Using the power allocation $\beta_1 = 1 - \epsilon$ and $\beta_i = \alpha_{i-1} - \epsilon$, $\forall i \in [2:s]$, with $\epsilon > 0$ and $\epsilon \ll 1$ we can write

$$R_i = (\alpha_{i-1} - \epsilon - \alpha_i) \log \mathrm{SNR} \doteq (\alpha_{i-1} - \alpha_i) \log \mathrm{SNR}.$$

Intuitively $\beta_i = \alpha_{i-1}$ can be interpreted as a power allocation *matching* the channel gains. So, for the total achievable secrecy rate we have

$$R_s \doteq \sum_{i=1}^{s} \Delta_i (\alpha_{i-1} - \alpha_i) \log \mathrm{SNR} = \sum_{i=1}^{s} \Delta_i \log \frac{h_i^2}{h_{i-1}^2}. \tag{9}$$

Now, let us state the upper bound to C_s in the high SNR regime in Theorem 5.

Theorem 5. *Assuming high-SNR regime and large dynamic range over channel states we can upper bound C_s as follows*

$$C_s \dot{\le} \sum_{i=1}^{s} \Delta_i (\alpha_{i-1} - \alpha_i) \log \mathrm{SNR}.$$

This upper bound is matched with the achievable rate derived in (9)*, so the above equation characterizes C_s in this regime.*

REFERENCES

[1] U. M. Maurer, "Secret Key Agreement by Public Discussion From Common Information," *IEEE Trans. Inf. Theory*, vol. 39, 1993.

[2] R. Ahlswede and I. Csiszar, "Common Randomness in Information Theory and Cryptography, Part I: Secret Sharing," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1121–1132, Jul. 1993.

[3] I. Csiszar and P. Narayan, "Secrecy capacities for multiple terminals," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, Dec. 2004.

[4] I. Csiszar and P. Narayan, "Secrecy capacities for multiterminal channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, Jun. 2008.

[5] A S. Avestimehr, S. N. Diggavi and D. N. C. Tse, "Wireless Network Information Flow," *45th Allerton Conf. On Comm., Control, and Computing*, Monticello, Illinois, USA, Sep. 2007.

[6] M. Jafari Siavoshani, C. Fragouli, S N. Diggavi, U. Pulleti, and K. Argyraki, "Group secret key generation over broadcast erasure channels," *Asilomar*, Nov. 2010.

[7] A. A. Gohari and V. Anantharam, "Information-Theoretic key agreement of multiple terminals - Part II: channel model," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3997–4010, Aug. 2010.

[8] A. Khisti, S. N. Diggavi, and G. W. Wornell, "Secret-key agreement over wiretap channels with random state parameters," to appear *IEEE Transactions on Information Forensics and Security,* 2011.
[9] Y. Liang, L. Lai, H. V. Poor, and S. Shamai (Shitz), "The Broadcast Approach over Fading Gaussian Wiretap Channels," *IEEE Information Theory Workshop*, 2009.
[10] M. Jafari Siavoshani, U. Pulleti, E. Atsan, I. Safaka, C. Fragouli, K. Argyraki, S. Diggavi, "Exchanging Secrets without Using Cryptography," Arxiv preprint http://arxiv.org/abs/1105.4991.

V. Appendix

Proof of Theorem 2 (Upper bound): Using Theorem 1 we can write $C_s \le I(X_\mathsf{A}, X_\mathsf{B}|X_\mathsf{E})$, where we have $H(X_\mathsf{A}|X_\mathsf{E}) = \sum_{i=0}^{s-1} \delta_i \left[H(X_\mathsf{A}) - H(\boldsymbol{F}_i X_\mathsf{A})\right]$, and $H(X_\mathsf{A}|X_\mathsf{B}, X_\mathsf{E}) = \sum_{i=0}^{s-1} \left[2\delta_i(\delta_0 + \cdots + \delta_{i-1}) + \delta_i^2\right] \left[H(X_\mathsf{A}) - H(\boldsymbol{F}_i X_\mathsf{A})\right]$. So we have $I(X_\mathsf{A}, X_\mathsf{B}|X_\mathsf{E}) = \sum_{i=0}^{s-1} \rho_i \left[H(X_\mathsf{A}) - H(\boldsymbol{F}_i X_\mathsf{A})\right]$, where $\rho_i \triangleq \delta_i - 2\delta_i(\delta_0 + \cdots + \delta_{i-1}) - \delta_i^2$. Knowing that $H(\boldsymbol{F}_i X_\mathsf{A}) = H(\boldsymbol{F}_i X_\mathsf{A}, \boldsymbol{F}_{i-1} X_\mathsf{A})$, applying the chain rule, and doing some mathematical manipulation, we can write

$$
\begin{aligned}
C_s \le I(X_\mathsf{A}, X_\mathsf{B}|X_\mathsf{E}) &= \sum_{j=1}^{s} H(\boldsymbol{F}_j X_\mathsf{A}|\boldsymbol{F}_{j-1} X_\mathsf{A}) \sum_{i=0}^{j-1} \rho_i \\
&= \sum_{j=1}^{s} H\left([\boldsymbol{F}_j - \boldsymbol{F}_{j-1}] X_\mathsf{A}|\boldsymbol{F}_{j-1} X_\mathsf{A}\right) \sum_{i=0}^{j-1} \rho_i \\
&\le \sum_{j=1}^{s} H\left([\boldsymbol{F}_j - \boldsymbol{F}_{j-1}] X_\mathsf{A}\right) \sum_{i=0}^{j-1} \rho_i \\
&\overset{(a)}{\le} \sum_{j=1}^{s} \operatorname{rank}\left(\boldsymbol{F}_j - \boldsymbol{F}_{j-1}\right) \left(\sum_{i=0}^{j-1} \rho_i\right) \log q \\
&\overset{(b)}{=} \sum_{j=1}^{s} \left[\operatorname{rank} \boldsymbol{F}_j - \operatorname{rank} \boldsymbol{F}_{j-1}\right] \left(\sum_{i=0}^{j-1} \rho_i\right) \log q, \qquad (10)
\end{aligned}
$$

where (a) is true because uniform distribution on X_A achieves the maximum values for all the entropies in the summation and (b) is true because of (4). Also, note that $\sum_{i=0}^{j-1} \rho_i = \theta_j(1-\theta_j) \ge 0$, where $\theta_j = \sum_{i=0}^{j-1} \delta_i$. ∎

Proof of Theorem 3: Using Theorem 1 we can write $C_s \le I(X_\mathsf{A}; X_\mathsf{B}|X_\mathsf{E}) = I(X_\mathsf{A}; \widetilde{X}_\mathsf{B}, S_\mathsf{B}|\widetilde{X}_\mathsf{E}, S_\mathsf{B})$, Hence, we have

$$
\begin{aligned}
C_s &\le I(X_\mathsf{A}; X_\mathsf{B}|X_\mathsf{E}) \\
&= H(\widetilde{X}_\mathsf{B}, S_\mathsf{B}|\widetilde{X}_\mathsf{E}, S_\mathsf{E}) - H(\widetilde{X}_\mathsf{B}, S_\mathsf{B}|\widetilde{X}_\mathsf{E}, S_\mathsf{E}, X_\mathsf{A}) \\
&\overset{(a)}{=} H(\widetilde{X}_\mathsf{B}, S_\mathsf{B}|\widetilde{X}_\mathsf{E}, S_\mathsf{E}) - H(\widetilde{X}_\mathsf{B}, S_\mathsf{B}|X_\mathsf{A}) \\
&= H(\widetilde{X}_\mathsf{B}, S_\mathsf{B}|\widetilde{X}_\mathsf{E}, S_\mathsf{E}) - H(S_\mathsf{B}|X_\mathsf{A}) - H(\widetilde{X}_\mathsf{B}|S_\mathsf{B}, X_\mathsf{A}) \\
&= H(\widetilde{X}_\mathsf{B}, S_\mathsf{B}|\widetilde{X}_\mathsf{E}, S_\mathsf{E}) - H(S_\mathsf{B}) - H(Z_\mathsf{B}) \\
&= H(\widetilde{X}_\mathsf{B}, \widetilde{X}_\mathsf{E}|S_\mathsf{E}, S_\mathsf{B}) + H(S_\mathsf{E}, S_\mathsf{B}) \\
&\quad - H(\widetilde{X}_\mathsf{E}, S_\mathsf{E}) - H(S_\mathsf{B}) - H(Z_\mathsf{B}) \\
&= H(\widetilde{X}_\mathsf{B}, \widetilde{X}_\mathsf{E}|S_\mathsf{E}, S_\mathsf{B}) - H(\widetilde{X}_\mathsf{E}|S_\mathsf{E}) - H(Z_\mathsf{B}) \\
&= \sum_{i=0}^{s}\sum_{j=0}^{s} \delta_i\delta_j H(\widetilde{X}_\mathsf{B}, \widetilde{X}_\mathsf{E}|S_\mathsf{E} = j, S_\mathsf{B} = i) \\
&\quad - \sum_{k=0}^{s} \delta_k H(\widetilde{X}_\mathsf{E}|S_\mathsf{E} = k) - H(Z_\mathsf{B}) \\
&= \sum_{i=0}^{s}\sum_{j=0}^{s} \delta_i\delta_j H(h_i X_\mathsf{A} + Z_\mathsf{B}, h_j X_\mathsf{A} + Z_\mathsf{E}) \\
&\quad - \sum_{k=0}^{s} \delta_k H(h_k X_\mathsf{A} + Z_\mathsf{E}) - H(Z_\mathsf{B}) \\
&= \sum_{i=0}^{s}\sum_{j=0}^{s} \delta_i\delta_j H(h_i X_\mathsf{A} + Z_\mathsf{B}|h_j X_\mathsf{A} + Z_\mathsf{E}) - H(Z_\mathsf{B}) \\
&\overset{(b)}{\le} \sum_{i=0}^{s}\sum_{j=0}^{s} \delta_i\delta_j \log(2\pi e(\operatorname{var}(h_i X_\mathsf{A} + Z_\mathsf{B}|h_j X_\mathsf{A} + Z_\mathsf{E}))) \\
&\quad - H(Z_\mathsf{B}), \qquad (11)
\end{aligned}
$$

where (a) is true since we have $X_\mathsf{B} \leftrightarrow X_\mathsf{A} \leftrightarrow X_\mathsf{E}$ and (b) follows from the fact that for a fixed variance, Gaussian distribution maximizes the entropy.

Equality in (b), (11), is achieved when $(h_i X_\mathsf{A} + Z_\mathsf{B}|h_j X_\mathsf{A} + Z_\mathsf{E})$ has a Gaussian distribution. A sufficient condition for this to be satisfied is when X_A, Z_B, and Z_E are Gaussian and independent. This observation makes the calculation of $\log(2\pi e \operatorname{var}(h_i X_\mathsf{A} + Z_\mathsf{B}|h_j X_\mathsf{A} + Z_\mathsf{E}))$ much easier as it is equivalent to the evaluation of $H(h_i X_\mathsf{A} + Z_\mathsf{B}, h_j X_\mathsf{A} + Z_\mathsf{E}) - H(h_j X_\mathsf{A} + Z_\mathsf{E})$ when X_A, Z_B, and Z_E are Gaussian and independent as shown below,

$$
\begin{aligned}
&H(h_i X_\mathsf{A} + Z_\mathsf{B}, h_j X_\mathsf{A} + Z_\mathsf{E}) - H(h_j X_\mathsf{A} + Z_\mathsf{E}) = \\
&\quad \log((2\pi e)^2(1 + h_i^2 P + h_j^2 P)) - \log(2\pi e(1 + h_j^2 P)),
\end{aligned}
$$

where $\mathbb{E}\left[X_\mathsf{A}^2\right] = P$. Hence, the secrecy upper bound becomes

$$
\begin{aligned}
C_s &\le \sum_{i=0}^{s}\sum_{j=0}^{s} \delta_i\delta_j \log((2\pi e)^2(1 + h_i^2 P + h_j^2 P)) \\
&\quad - \sum_{i=0}^{s}\sum_{j=0}^{s} \delta_i\delta_j \log(2\pi e(1 + h_j^2 P)) - \log(2\pi e) \\
&= \sum_{i=0}^{s}\sum_{j=0}^{s} \delta_i\delta_j \log\left(1 + \frac{h_i^2 P}{1 + h_j^2 P}\right),
\end{aligned}
$$

and we are done. ∎