Secure Groupcast With Shared Keys

Hua Sun, *Member, IEEE*

***Abstract*— We consider a transmitter and K receivers, each of which shares a key variable with the transmitter. Through a noiseless broadcast channel, the transmitter wishes to send a common message W securely to N out of the K receivers while the remaining $K - N$ receivers learn no information about W. We are interested in the maximum message rate, i.e., the maximum number of bits of W that can be securely groupcast to the legitimate receivers per key block and the minimum broadcast bandwidth, i.e., the minimum number of bits of the broadcast information required to securely groupcast the message bits. We focus on the setting of combinatorial keys, where every subset of the K receivers share an independent key of arbitrary size. Under this combinatorial key setting, the maximum message rate is characterized for the following scenarios - 1) $N = 1$ or $N = K - 1$, i.e., secure unicast to 1 receiver with $K - 1$ eavesdroppers or secure groupcast to $K - 1$ receivers with 1 eavesdropper, 2) $N = 2, K = 4$, i.e., secure groupcast to 2 out of 4 receivers, and 3) the symmetric setting where the key size for any subset of the same cardinality is equal for any N, K. Further, for the latter two cases, the minimum broadcast bandwidth for the maximum message rate is characterized.**

***Index Terms*— Capacity, broadcast, groupcast, information theoretic security.**

I. Introduction

THE first theoretical analysis of cryptography and secrecy system was carried out by Shannon in the groundbreaking 1949 work [1], where the mathematical framework of information theoretic security was introduced to establish the fundamental limits of secure point-to-point communication. Shannon studied the one-time pad system (see Fig. 1.1), where Alice shares a key Z with Bob and wishes to send an independent message W to Bob such that even if the transmit signal X is fully eavesdropped by Eve, Eve cannot learn anything about W as long as Eve has no knowledge of the key Z. The simple one-time pad scheme $X = W + Z$, where '$+$' represents bit-wise binary addition is proved information theoretically secure and communication-wise optimal in the following sense.

- To send one bit of the message W securely, one bit of the key Z must be shared. That is, the maximum *message rate* is 1 bit per key bit under perfect secrecy.
- To send one bit of the message W securely, one bit of the transmit signal X must be broadcast (seen by everyone). That is, the minimum *broadcast bandwidth* is 1 bit per message bit under zero error decoding.

In this work, motivated by the need of secure *group* (beyond point-to-point) communication under *complex* adversarial scenario (beyond a single eavesdropper knowing nothing about the key), we consider the following secure groupcast communication scenario. A transmitter shares a key variable $Z_k, k \in \{1, 2, \cdots, K\}$ with Receiver k and Z_k may be arbitrarily correlated (see Fig. 1.2). Aided by the shared keys, the transmitter wishes to send a common message W securely to N out of the K receivers through broadcasting the signal X to all receivers, in a way that any one of the remaining $K - N$ receivers learns no information about W in the information theoretic sense.

This secure groupcast problem naturally generalizes Shannon's one-time pad system, which is a special case of secure unicast $(N = 1)$ over a $K = 2$ receiver broadcast channel and the eavesdropping receiver knows nothing about the key of the legitimate receiver. Following the communication metrics considered by Shannon, we focus on the following two questions regarding the fundamental limits of secure groupcast.

- What is the maximum message rate, defined as the maximum number of bits of the message W that can be securely groupcast per key block (a classic Shannon theoretic formulation where we may code over a long key block and the block size is allowed to approach infinity)?
- What is the minimum broadcast bandwidth, defined as the minimum number of bits of the broadcast information X required to securely groupcast a message of certain rate?

Beyond being an elemental model for information theoretic security, the above shared key secure groupcast problem arises naturally in many applications, where we interpret the keys either as digital tokens or information from memory devices (e.g., in premiere streaming or game distribution), or more generally as side-information variables that could be sensed from the environment or obtained from prior communication (e.g., in wireless networking). In addition, the model can be easily extended from groupcasting a single message for a single group to multiple messages, each exclusively for an arbitrary group under various security constraints, i.e., the secure groupcast model is introduced to enable broadcasting to a selected set of qualified receivers while unqualified receivers obtain no useful information.

Combinatorial Key Setting: As an initial step, we mainly focus on the combinatorial key setting, where every subset $\mathcal{U}$

Manuscript received November 13, 2020; revised September 20, 2021; accepted March 9, 2022. Date of publication March 17, 2022; date of current version June 15, 2022. This work was supported in part by NSF under Grant CCF-2007108 and Grant CCF-2045656.

The author is with the Department of Electrical Engineering, University of North Texas, Denton, TX 76203 USA (e-mail: hua.sun@unt.edu).

Communicated by A. Beimel, Associate Editor for Sequences and Cryptography.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2022.3160507.

Digital Object Identifier 10.1109/TIT.2022.3160507

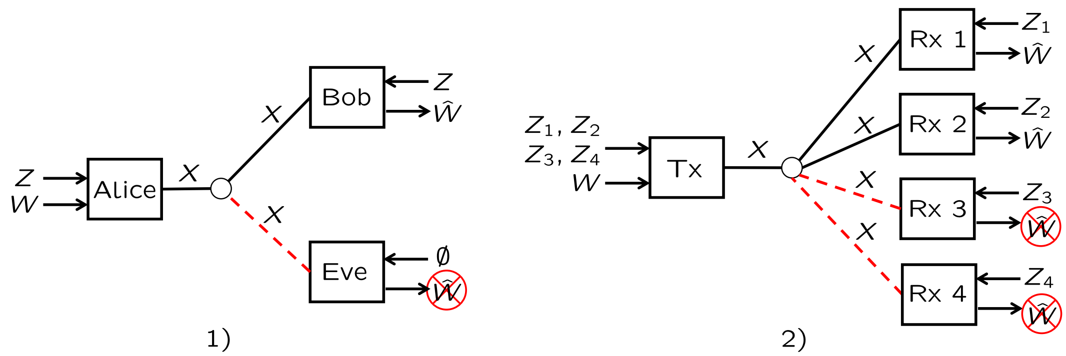

Fig. 1. 1) The one-time pad system. 2) The secure groupcast problem (to 2 out of 4 receivers).

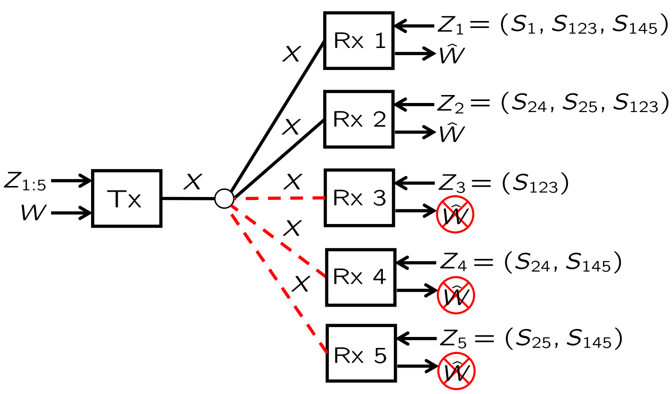

Fig. 2. A secure groupcast problem to 2 out of 5 receivers with combinatorial keys (i.e., the S variables are independent). $Z_{1:5}$ denotes $(Z_1, Z_2, Z_3, Z_4, Z_5)$.

of the K receivers share an independent key $S_{\mathcal{U}}$ of arbitrary size. An example is shown in Fig. 2, where S_1 denotes the key that is known only to Receiver 1 (and the transmitter), S_{145} (abbreviation of $S_{\{1,4,5\}}$ for simplicity) is known to Receiver 1, Receiver 4, and Receiver 5 etc. Further, the S variables with different subscripts are independent of each other.

The combinatorial key setting turns out to be technically challenging due to the necessity of highly structured coding of the message symbols and the key symbols (for which the setting in Fig. 2 is a representative example even when the S variables all have the same size), and the abundance of parameters (as the key size for different subsets may be different so that overall the order of parameters is exponential in K). The essence is to accommodate for and utilize the complex correlation among the keys so that legitimate receivers can decode the common message while eavesdropping receivers cannot obtain anything from the correlated keys (i.e., need to avoid leakage under multiple correlated views). Our results are summarized next.

Main Results and Techniques: The main results of this work include the exact characterization of the maximum message rate and the minimum broadcast bandwidth for settings listed below.

- $N = 1$, any K: This is the secure unicast setting, with only 1 desired receiver. Both the maximum message rate and the minimum broadcast bandwidth are characterized. The achievable scheme is based on random linear coding over the key symbols. Refer to Theorem 3.
- $N = K - 1$, any K: This can be viewed as the secure multicast setting, with only 1 eavesdropper. The maximum message rate is characterized (and the minimum broadcast bandwidth when $K \leq 4$ or when the total secure key size is the same for all receivers). The achievable scheme is based on random linear coding over the message symbols. Refer to Theorem 4.
- $N = 2, K = 4$: Both the maximum message rate and the minimum broadcast bandwidth are characterized. The achievable scheme requires a delicate structured decomposition to basic components according to the key sizes. Refer to Theorem 5.
- The symmetric setting for any N, K, where the size of the key $S_{\mathcal{U}}$ only depends on $|\mathcal{U}|$ (i.e., the cardinality of the subset): Both the maximum message rate and the minimum broadcast bandwidth are characterized. The achievable scheme requires an intricate coding over keys from various subsets that handles correctness and security jointly. Refer to Theorem 6.
- The converse bounds on the message rate for all results above are given by a simple conditional entropy term (refer to Theorem 1); the converse bounds on the broadcast bandwidth for all results above have an interesting unified form that can be interpreted through common information (refer to Theorem 2).
- The simple conditional entropy converse bound in Theorem 1 is not tight in general. Specifically, a stronger bound is derived for the setting in Fig. 2 when the S variables have the same size and the maximum message rate is characterized with a matching vector linear coding scheme (refer to Theorem 7).

We have also explored the generalization to the following scenarios.

- The rate region of secure groupcasting multiple messages. Specifically, we consider 2 legitimate receivers with 3 desired messages (1 for each individual receiver so that the other receiver learns nothing and a common message for both receivers) and all these 3 messages must be kept fully secure to an eavesdropping receiver. Refer to Theorem 8.

- The discrete memoryless key[1] setting. Interestingly, the scenarios where random linear codes suffice for the combinatorial key setting (i.e., $N = 1$ and $N = K - 1$) generalize fully to discrete memoryless keys by random binning. Refer to Theorem 9.

Before proceeding to the problem statement, we end the introduction section with discussions on the connection between secure groupcast and related prior work.

A. Related Work

The elemental problem of secure groupcast has interesting connections to several problems that have been studied in prior work and this section is devoted to the discussion of these connections. Due to space limits, we will focus on the connections to secure groupcast and leave further details to the references cited.

1) Secret Key Agreement (Generation): In the problem of key agreement [2]–[6], multiple terminals observing correlated sources wish to agree on a common key through public communication and it is required that an eavesdropper learns nothing about the key from public communication.

Secret key agreement provides a natural achievable scheme for secure groupcast, where the legitimate receivers first agree on a secret key that is not known to the eavesdropping receivers (with the help of the transmitter and the noiseless broadcast channel). Then the secret key can be used to encrypt the desired message. This idea will be used in Section IV-B. Unfortunately, secret key agreement is only understood when there is a *single* eavesdropper [4] but in secure groupcast, we have *multiple* eavesdroppers, each with a different view of the source. Also, for key agreement, only the maximum key rate (corresponding to the groupcast rate in secure groupcast) is known and the communication cost (corresponding to the broadcast bandwidth in secure groupcast) remains open in general [4]. Lastly, key agreement does not appear necessary for secure groupcast.

2) Broadcast Encryption: As a related problem, broadcast encryption [7]–[10] studies how to *design* the key variables at a number of receivers so that a transmitter may securely send a message through noiseless broadcasting in a communication efficient manner to some legitimate receivers (from one out of a number of selected subsets of the receivers) while the remaining eavesdropping receivers learn nothing. There are two main differences to secure groupcast - first, in secure groupcast the key variables are given (e.g., with fixed joint distribution) while in broadcast encryption the keys are subject to design; second, in secure groupcast, the identities of legitimate and eavesdropping receivers are fixed while in broadcast encryption, there are multiple choices of legitimate receiver sets and the crux is to design the keys (instead of using given keys) to enable multiple secure groupcast tasks (instead of one). Therefore, the broadcast encryption problem can be regarded as a 'compound' version of secure groupcast and has been studied in a follow-up work using Shannon theoretic formulations [11].

[1]That is, the keys are drawn from a discrete memoryless source, such that the key variables are generated independently symbol by symbol from a fixed joint distribution.

3) Latent Capacity Region: The latent capacity region of broadcast channels [12]–[14] studies the implication of a rate tuple of common messages for various subsets of receivers being achievable, i.e., how can the achievable rates of certain group of receivers be exchanged for those of other groups of receivers? This interesting open problem is conceptually related to combinatorial secure groupcast, where from a rate exchange perspective, we are asking how to exchange various key variables shared by subsets of receivers to a common message for the group of desired receivers. However, latent capacity region has no security constraint and the required techniques in achievability and converse appear different.

4) Secure Broadcasting: How to send messages securely over a broadcast channel has been studied along the line of Wyner's wiretap channel [15], and its generalizations to confidential messages (see e.g., [16], [17]), and secure broadcasting over wireless channels (see e.g., [18], [19]). The enabler of secure communication in this line of work is that different receivers experience different channels, i.e., the channel itself has relative advantage to be exploited. In contrast, in secure groupcast every receiver sees the same noiseless broadcast channel and relative advantage comes from the shared keys. Notably, a recent work has studied a model (with a few users and a simple key structure) where both shared keys and discrete memoryless broadcast channels are simultaneously present [20].

5) Secure (Private) Index Coding: Index coding [21] is a canonical problem that studies how to efficiently broadcast under various side information at the receiver side with a noiseless broadcast channel. There are several variants of index coding that include security constraints (see e.g., [22]–[24]) with and without shared keys and with and without external eavesdroppers. The main focus of index coding works is on the side information structure and its interplay with multiple desired messages. Secure groupcast highlights the shared key structure and its influence on message rate and broadcast bandwidth.

Notation: For positive integers $K_1, K_2, K_1 \leq K_2$, we use the notation $[K_1 : K_2] = \{K_1, K_1+1, \cdots, K_2\}$. The notation $|\mathcal{U}|$ is used to denote the cardinality of a set $\mathcal{U}$ and the notation $|X|$ is used to denote the number of elements of a vector X. For a matrix $\mathbf{V}$, $\mathbf{V}(i, j)$ represents the element in the i-th row and j-th column. For two matrices $\mathbf{V}_1, \mathbf{V}_2$ (with the same number of columns), $[\mathbf{V}_1; \mathbf{V}_2]$ denotes the row stack of $\mathbf{V}_1, \mathbf{V}_2$. A binomial coefficient $\binom{K}{U}$ is defined as 0 if $K < U$.

II. Problem Statement

Define K discrete random variables $z_1, z_2, \cdots, z_K$ of finite cardinality, drawn from an arbitrary joint distribution $P_{z_1,z_2,\cdots,z_K}$. Following the convention, $Z_1, Z_2, \cdots, Z_K$ denote L length extensions of $z_1, z_2, \cdots, z_K$, i.e., $Z_1, Z_2, \cdots, Z_K$ are sequences of length L, such that the sequence of tuples $[Z_1(l), Z_2(l), \cdots, Z_K(l)]_{l=1}^{L}$ is produced i.i.d. according to $P_{z_1,z_2,\cdots,z_K}$.

Consider a transmitter that knows the keys $Z_1, Z_2, \cdots, Z_K$, and K receivers such that Receiver k knows $Z_k, k \in [1 : K]$. The transmitter wishes to send a common message W securely to the first N receivers, where $1 \leq N \leq K - 1$.

The message W consists of L_W i.i.d. uniform symbols from a finite field[2] $\mathbb{F}_p$ for a prime power p, so $H(W) = L_W \log_2 p$ bits. We assume that the message W is independent of the key variables $Z_1, Z_2, \cdots, Z_K$,

$$I(W; Z_1, Z_2, \cdots, Z_K) = 0. \tag{1}$$

The communication channel is a noiseless broadcast channel, i.e., the transmit signal X is sent by the transmitter and seen by every receiver. To securely groupcast the message W, the transmit signal X consists of L_X symbols from $\mathbb{F}_p$.

From the transmit signal X and the key Z_k, each legitimate receiver must be able to decode the message W, with probability of error P_e. The probability of error must approach zero as the key block length L approaches infinity.[3] From Fano's inequality, we have

$$\text{[Correctness]} \quad H(W|X, Z_k) = o(L), \ \forall k \in [1:N] \tag{2}$$

where any function of L, say $f(L)$, is said to be $o(L)$ if $\lim_{L\to\infty} f(L)/L = 0$. From the transmit signal X and the key Z_k, each eavesdropping receiver obtains a negligible amount of information about the message W,

$$\text{[Security]} \quad I(W; X, Z_k) = o(L), \ \forall k \in [N+1:K]. \tag{3}$$

The groupcast *rate* characterizes how many bits of the message are securely groupcast per key block, and is defined as follows,

$$R = \frac{H(W)}{L} = \frac{L_W \log_2 p}{L}. \tag{4}$$

A rate R is said to be achievable if there exists a sequence of secure groupcast schemes (indexed by L), each of rate greater or equal to R, for which $P_e \to 0$ as $L \to \infty$ (i.e., the correctness constraint (2) and the security constraint (3) are satisfied). The supremum of achievable rates is called the capacity C.

The broadcast bandwidth $\beta(R)$ characterizes how many bits of the transmit signal are broadcast per key block to securely groupcast a message of rate R, and is defined as follows,

$$\beta(R) = \frac{L_X \log_2 p}{L}. \tag{5}$$

The achievable broadcast bandwidth is defined similarly, i.e., broadcast bandwidth $\beta(R)$ is said to be achievable if there exists a sequence of secure groupcast schemes, each of rate greater than or equal to R and each of broadcast bandwidth smaller than or equal to $\beta(R)$, for which $P_e \to 0$ as $L \to \infty$. The infimum of achievable broadcast bandwidth is called the minimum broadcast bandwidth $\beta^*(R)$.

We will be mainly interested in the capacity, C and the minimum broadcast bandwidth when the rate value is the capacity, $\beta^*(C)$.

[2]As usual for an information theoretic formulation, the actual size of the message is allowed to approach infinity. We allow the optimization of both parameters of the key block length L and the field size p, to match the code dimensions and simplify the presentation of the coding scheme.

[3]If P_e is required to be exactly zero, then the $o(L)$ term can be replaced with 0. The situation is similar if zero leakage instead of vanishing leakage is required in the security constraint (3).

A. Combinatorial Keys

The combinatorial key setting refers to a specific type of joint distribution of the keys and is defined as follows. Consider $2^K - 1$ independent random variables $s_{\mathcal{U}}$, where $\mathcal{U}$ may be any non-empty subset of $[1:K]$. For example, when $K = 3$, we have[4] $s_1, s_2, s_3, s_{12}, s_{13}, s_{23}, s_{123}$ and

$$H(s_1, s_2, \cdots, s_{\mathcal{U}}, \cdots, s_{1:K}) = H(s_1) + H(s_2) + \cdots + H(s_{\mathcal{U}}) + \cdots + H(s_{1:K}). \tag{6}$$

We assume that $s_{\mathcal{U}}$ consists of an integer number, say $L_{\mathcal{U}}$, of i.i.d. uniform symbols from $\mathbb{F}_p$,

$$H(s_{\mathcal{U}}) = L_{\mathcal{U}} \log_2 p \text{ bits}. \tag{7}$$

The variable z_k is the collection of all $s_{\mathcal{U}}$ variables such that $k \in \mathcal{U}$,

$$z_k = (s_{\mathcal{U}} : k \in \mathcal{U}). \tag{8}$$

For example, when $K = 3$, $z_2 = (s_2, s_{12}, s_{23}, s_{123})$. The symmetric setting is defined as follows,

$$\text{(symmetric setting)} \quad H(s_{\mathcal{U}_1}) = H(s_{\mathcal{U}_2}), \forall \mathcal{U}_1, \mathcal{U}_2 \text{ such that } |\mathcal{U}_1| = |\mathcal{U}_2|. \tag{9}$$

The extension of the above system model to include multiple groupcast messages is immediate and will be presented when we consider this generalization in Section IV-A.

III. Main Results

In this section, we summarize our main results along with illustrative examples and observations.

A. Converse on Rate R and Broadcast Bandwidth $\beta(R)$

We present a simple converse (upper) bound on the groupcast rate R in the following theorem.

Theorem 1 (Rate Converse): For any secure groupcast problem (to the first N of K receivers),

$$R \leq H(z_q|z_e), \ \forall q \in [1:N], \forall e \in [N+1:K]. \tag{10}$$

The proof of Theorem 1 is presented in Section V-A. The conditional entropy bound (10) is very intuitive, because W must be decoded by any qualified Receiver $q \in [1:N]$ and cannot be learned by any eavesdropping Receiver $e \in [N+1:K]$. Surprisingly, this simple conditional entropy bound turns out to be tight for many settings of interest (see below). However, it is not sufficient in general (refer to Remark 2 after Theorem 7).

Next, we present an interesting converse (lower) bound on the broadcast bandwidth $\beta(R)$ in the following theorem.

Theorem 2 (Broadcast Bandwidth Converse): For any secure groupcast problem (to the first N of K receivers), consider any set of qualified receivers $\mathcal{Q} \triangleq \{q_1, \cdots, q_{|\mathcal{Q}|}\} \subset [1:N]$ and consider any random

[4]For $s_{\mathcal{U}}$, we may simplify the subscript when the elements of $\mathcal{U}$ are easy to list, e.g., we may write $s_{\{1,2\}}$ as s_{12}.

variable u_e that is a function of the key of an eavesdropping Receiver $e \in [N+1:K]$, i.e., $H(u_e|z_e) = 0$. We have

$$\begin{aligned}
&\beta(R)\\
\geq\ & |\mathcal{Q}|R - \left(\sum_{i=1}^{|\mathcal{Q}|} H(z_{q_i}|u_e) - H(z_{q_1}, z_{q_2}, \cdots, z_{q_{|\mathcal{Q}|}}|u_e)\right)\\
=\ & |\mathcal{Q}|R - \sum_{i=1}^{|\mathcal{Q}|-1} I(z_{q_1}, \cdots, z_{q_i}; z_{q_{i+1}}|u_e). \quad (11)
\end{aligned}$$

The proof of Theorem 2 is presented in Section V-B. The negative term on the RHS of (11) captures the benefits of correlated keys in reducing the broadcast bandwidth. On one extreme when the keys are fully independent, this negative term is zero and we have to send the message W to all $|\mathcal{Q}|$ qualified receivers one by one, so that the broadcast bandwidth is $|\mathcal{Q}|$ times of the rate, R of the message, i.e., $\beta(R) \geq |\mathcal{Q}|R$. On the other extreme when the keys are identical and independent of the key at the eavesdropping receiver (i.e., $z_{q_1} = \cdots = z_{q_Q}$), this negative term becomes $(|\mathcal{Q}|-1)H(z_{q_1})$. Now suppose $R = H(z_{q_1})$, then the broadcast bandwidth bound becomes $\beta(R) \geq R$ and it might suffice to simply send out the one-time pad signal $W + Z_{q_1}$. In general, between the two extremes, the saving is given by the difference between the sum of individual entropy of each key and the joint entropy of all keys, which can be interpreted as a form of common information. Interestingly, this common information type of term can also be written as the sum of a chain of mutual information terms.

Equipped with the above converse results, we are now ready to proceed to consider the combinatorial key setting, which is referred to as the combinatorial secure groupcast problem for short. Note that for combinatorial secure groupcast, all achievable schemes satisfy zero error and zero leakage (i.e., $o(L)$ is replaced with 0 in (2), (3)).

B. Secure Unicast $N = 1$ and Secure Multicast $N = K - 1$ Settings

When there is only $N = 1$ desired receiver, secure groupcast reduces to secure unicast, and this combinatorial secure unicast problem can be solved by random linear coding over the key symbols for the achievability side and the bounds given above for the converse side. This result is presented in the following theorem.

Theorem 3 (Secure Unicast): For the combinatorial secure unicast problem (to the first of K receivers), the capacity and the minimum broadcast bandwidth for capacity achieving schemes are

$$\begin{aligned}
C &= \min_{e \in [2:K]} H(z_1|z_e)\\
&= \min_{e \in [2:K]} \sum_{\substack{\mathcal{U} \subset [1:K]:\\ 1 \in \mathcal{U}, e \notin \mathcal{U}}} H(s_{\mathcal{U}}), \quad (12)\\
\beta^*(C) &= C. \quad (13)
\end{aligned}$$

The proof of Theorem 3 is presented in Section V-C. The achievability is based on creating a key from what the legitimate Receiver 1 knows so that any one of the $K-1$ eavesdropping receivers cannot learn anything about the created key. This key can be created by random linear coding and after the key is created, one-time pad coding suffices to achieve the capacity and the minimum broadcast bandwidth. An example is presented below to explain this idea.

Example 1: Consider a combinatorial secure unicast instance with $K = 4$ receivers, where the key configurations are given as follows,

$$\begin{aligned}
z_1 = (s_{12}, s_{13}, s_{14}, s_{134}),\ z_2 = (s_{12}),\\
z_3 = (s_{13}, s_{134}),\ z_4 = (s_{14}, s_{134}); \quad (14)\\
H(s_{12}) = 4\log_2\ p,\ H(s_{13}) = 2\log_2\ p,\\
H(s_{14}) = \log_2\ p,\ H(s_{134}) = 3\log_2\ p. \quad (15)
\end{aligned}$$

For example, $s_{12} \in \mathbb{F}_p^{4\times 1}$ contains 4 symbols from $\mathbb{F}_p$. From Theorem 3, we have

$$\begin{aligned}
C = \min\Big(&H(s_{13}) + H(s_{14}) + H(s_{134}),\\
&H(s_{12}) + H(s_{14}), H(s_{12}) + H(s_{13})\Big)\\
= 5\log_2\ p&, \quad (16)\\
\beta^*(C) = C &= 5\log_2\ p. \quad (17)
\end{aligned}$$

The converse follows immediately from Theorem 1 (taking the minimum converse bound over all eavesdropping receivers) and Theorem 2 (taking $\mathcal{Q} = \{1\}$ so that $\beta(R) \geq R$). The achievable scheme is presented next. Consider $L = 1$ block of the keys (then $Z_k = z_k$) and we wish to send $L_W = 5$ message symbols from $\mathbb{F}_p$ by broadcasting $L_X = 5$ symbols, i.e., W, X are both 5×1 vectors over $\mathbb{F}_p$. The combinatorial key variables are each precoded by a beamforming matrix to produce a mixed key, which is then added with the message W to produce the transmit signal X,

$$X = W + \mathbf{V}_{12}s_{12} + \mathbf{V}_{13}s_{13} + \mathbf{V}_{14}s_{14} + \mathbf{V}_{134}s_{134} \quad (18)$$

where the precoding matrices have 5 rows each and the number of columns matches the dimension of the key variables, e.g., $\mathbf{V}_{12} \in \mathbb{F}_p^{5\times 4}$. The correctness constraint (2) is trivially satisfied. For eavesdropping Receiver 2, after canceling the known key, he can recover

$$\begin{aligned}
&W + \mathbf{V}_{13}s_{13} + \mathbf{V}_{14}s_{14} + \mathbf{V}_{134}s_{134}\\
&= W + \Big[\mathbf{V}_{13}\ \mathbf{V}_{14}\ \mathbf{V}_{134}\Big]\begin{bmatrix} s_{13}\\ s_{14}\\ s_{134}\end{bmatrix} \quad (19)
\end{aligned}$$

so that in order to make sure nothing is revealed, we need

$$[\mathbf{V}_{13}\ \mathbf{V}_{14}\ \mathbf{V}_{134}]_{5\times 6} \text{ to have full rank (Receiver 2).} \quad (20)$$

Similarly, we need

$$\begin{aligned}
&[\mathbf{V}_{12}\ \mathbf{V}_{14}]_{5\times 5} \text{ to have full rank (Receiver 3)} \quad (21)\\
\text{and}\quad &[\mathbf{V}_{12}\ \mathbf{V}_{13}]_{5\times 6} \text{ to have full rank (Receiver 4).} \quad (22)
\end{aligned}$$

That is, we simply need the matrices to have full rank, which is easily satisfied by generic MDS matrices, e.g., Cauchy matrices over a properly large field. The details are deferred to the proof presented in Section V-C. Finally, the rate and broadcast bandwidth achieved match the converse.

We next consider the (somewhat) dual of secure unicast (a single legitimate receiver and any number of eavesdroppers) - secure multicast (a single eavesdropper and any number of legitimate receivers), whose capacity is solved by a similar random linear coding idea (but over the message symbols instead of over the key symbols). This result is presented in the following theorem.

Theorem 4 (Secure Multicast): For the combinatorial secure multicast problem (to the first $K-1$ of K receivers), the capacity is

$$\begin{aligned} C &= \min_{q\in[1:K-1]} H(z_q|z_K) \\ &= \min_{q\in[1:K-1]} \sum_{\mathcal{U}\subset[1:K-1]:q\in\mathcal{U}} H(s_{\mathcal{U}}). \end{aligned} \tag{23}$$

Further, the minimum broadcast bandwidth for capacity achieving schemes is characterized in the following two cases,

1. $H(z_1|z_K) = \cdots = H(z_{K-1}|z_K)$:
$$\beta^*(C) = \sum_{\mathcal{U}\subset[1:K-1]} H(s_{\mathcal{U}}); \tag{24}$$
2. $K=4$ (assume $H(z_1|z_4) \leq \min\Big(H(z_2|z_4), H(z_3|z_4)\Big)$, $H(s_{12}) \leq H(s_{13})$ with no loss[5]) :
$$\begin{aligned} \beta^*(C) = H(s_{123}) + \max\Big(&2H(s_1) + H(s_{12}) + 2H(s_{13}), \\ &3H(s_1) + 2H(s_{12}) + 2H(s_{13}) - H(s_{23})\Big). \end{aligned} \tag{25}$$

The proof of Theorem 4 is presented in Section V-D. To achieve the capacity, we simply generate random linear combinations of the message symbols and mix them with each of the combinatorial keys. Each legitimate receiver can decode the message after collecting a sufficient number of coded message symbols. This idea is explained in the following example.

Example 2: Consider a combinatorial secure multicast instance with $K=4$ receivers, where the key configurations are given as follows,

$$z_1 = (s_1, s_{13}),\ z_2 = (s_{23}),\ z_3 = (s_{13}, s_{23}),\ z_4 = (); \tag{26}$$
$$H(s_1) = \log_2\ p,\ H(s_{13}) = 2\log_2\ p,\ H(s_{23}) = 3\log_2\ p. \tag{27}$$

From Theorem 4, we have

$$\begin{aligned} C &= \min\Big(H(s_1) + H(s_{13}), H(s_{23}), H(s_{13}) + H(s_{23})\Big) \\ &= 3\log_2\ p, \end{aligned} \tag{28}$$
$$\begin{aligned} \beta^*(C) = \max\Big(&2H(s_1) + 2H(s_{13}), \\ &3H(s_1) + 2H(s_{13}) - H(s_{23})\Big) = 6\log_2\ p. \end{aligned} \tag{29}$$

The rate converse is simply given by the minimum entropy of the legitimate key variables (and follows from Theorem 1). The broadcast bandwidth converse is given by Theorem 2, where $\mathcal{Q} = \{1,2\}$ and $u_e = ()$ so that $\beta(C) \geq 2C - I(z_1; z_2) = 2C$. The achievable scheme is presented next. Consider $L=1$ and $W \in \mathbb{F}_p^{3\times 1}$. The transmit signal $X \in \mathbb{F}_p^{6\times 1}$ is produced as follows,

$$X = \begin{bmatrix} \mathbf{V}_1 W + s_1 \\ \mathbf{V}_{13} W + s_{13} \\ \mathbf{V}_{23} W + s_{23} \end{bmatrix} \tag{30}$$

where the dimensions of the precoding matrices are specified as $\mathbf{V}_1 \in \mathbb{F}_p^{1\times 3}, \mathbf{V}_{13} \in \mathbb{F}_p^{2\times 3}, \mathbf{V}_{23} \in \mathbb{F}_p^{3\times 3}$. For the secure multicast problem, security is trivial as the keys known by the eavesdropping receiver are never used and correctness requires that

$$\begin{aligned} &[\mathbf{V}_1; \mathbf{V}_{13}] \text{ has full rank (Receiver 1)}, \\ &[\mathbf{V}_{23}] \text{ has full rank (Receiver 2)}, \qquad (31)\\ &\text{and } [\mathbf{V}_{13}; \mathbf{V}_{23}] \text{ has full rank (Receiver 3)}. \qquad (32) \end{aligned}$$

The above constraints can be satisfied by generic (e.g., Cauchy or any MDS) matrices and details are deferred to the full proof presented in Section V-D.

Remark 1: While the capacity of secure multicast is solved simply by random linear codes, the minimum broadcast bandwidth for capacity achieving schemes is generally an open problem (e.g., when $K \geq 5$). When $K \leq 4$, we need to analyze carefully which combinatorial key has redundancy and quantify the amount so as to use only the minimum required (see Section V-D.1).

Interestingly, the above random coding idea for both combinatorial secure unicast and secure multicast generalizes to the discrete memoryless key setting, where corresponding results are obtained using standard existing random binning arguments (see Theorem 9).

C. Secure Groupcast to $N=2$ of $K=4$ Receivers

The capacity and minimum broadcast bandwidth for the combinatorial secure groupcast problem to 2 out of 4 receivers are characterized in the following theorem.

Theorem 5 ($N=2, K=4$): For the combinatorial secure groupcast problem to the first $N=2$ of $K=4$ receivers, the capacity is

$$\begin{aligned} C = &\min_{q\in\{1,2\}, e\in\{3,4\}} H(z_q|z_e) \\ &= H(s_{12}) + \min\Big(H(s_1) + H(s_{14}) + H(s_{124}), \\ &\quad H(s_1) + H(s_{13}) + H(s_{123}), H(s_2) + H(s_{24}) + H(s_{124}), \\ &\quad H(s_2) + H(s_{23}) + H(s_{123})\Big) \end{aligned} \tag{33}$$

and the minimum broadcast bandwidth for capacity achieving schemes is

$$\beta^*(C) = 2C - H(s_{12}) - \min\Big(H(s_{123}), H(s_{124})\Big). \tag{34}$$

The proof of Theorem 5 is presented in Section V-E. The complexity mainly lies in the abundance of the parameters so that we need to decompose the problem instance into multiple basic components (also how to identify basic components) and depending on the key configurations, there are many case studies. To this end, we need a decomposition result of two

[5]This assumption has no loss of generality because we can switch the indices of the receivers. Specifically, $H(z_1|z_4) \leq \min(H(z_2|z_4), H(z_3|z_4))$ means that we label the receiver with minimum key entropy conditioned on Receiver 4's key as Receiver 1 (among Receivers 1, 2, 3). After fixing Receiver 1, $H(s_{12}) \leq H(s_{13})$ means that we label the receiver that shares a smaller key with Receiver 1 as Receiver 2 (among Receivers 2, 3).

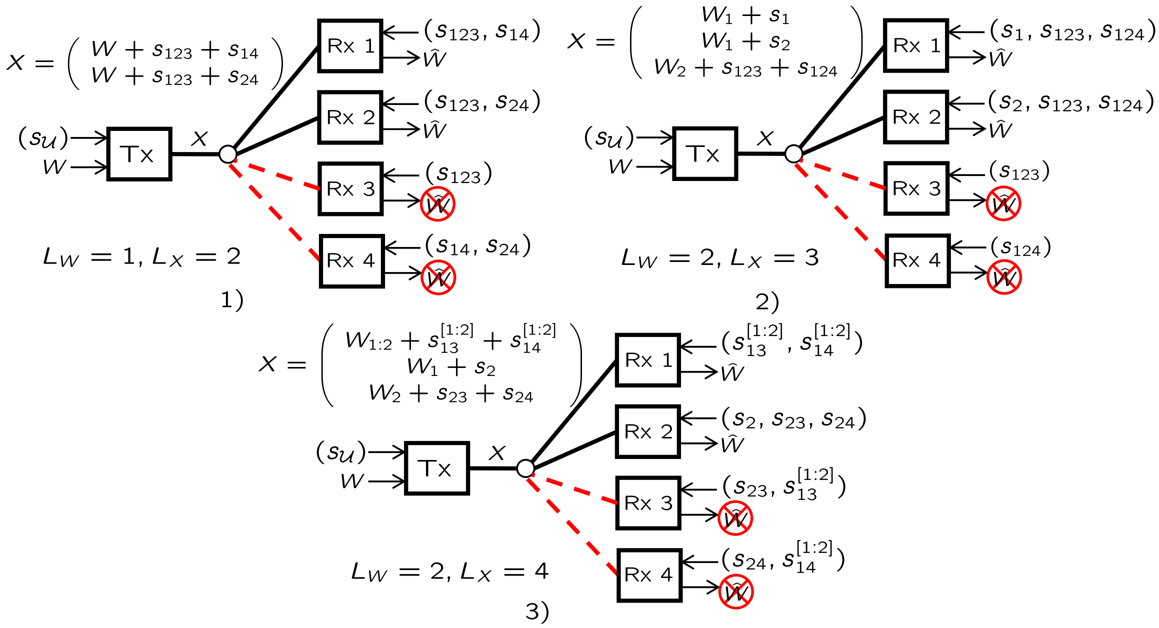

Fig. 3. The 3 sub-networks, where the variables are independent for different sub-networks. When some variable has 2 symbols, it is denoted as $W = (W_1; W_2) \triangleq W_{1:2}, s_{13} = (s_{13}^{[1]}; s_{13}^{[2]}) \triangleq s_{13}^{[1:2]}$ etc.

achievable schemes with two independent sets of keys, stated in the following lemma.

Lemma 1: Consider two sets of independent keys $(Z_1^{[1]}, \cdots, Z_K^{[1]})$ and $(Z_1^{[2]}, \cdots, Z_K^{[2]})$ such that $L_W^{[1]}$ and $L_W^{[2]}$ symbols of the messages $W^{[1]}, W^{[2]}$ can be securely groupcast with $L_X^{[1]}$ and $L_X^{[2]}$ symbols of the transmit signals $X^{[1]}, X^{[2]}$, respectively. Then we can concatenate the two schemes to one such that the keys are $Z_k = (Z_k^{[1]}, Z_k^{[2]}), k \in [1:K]$, $L_W = L_W^{[1]} + L_W^{[2]}$ symbols of $W = (W^{[1]}, W^{[2]})$ are securely groupcast with $L_X = L_X^{[1]} + L_X^{[2]}$ symbols of $X = (X^{[1]}, X^{[2]})$.

Proof: The proof is almost immediate. As long as each component scheme is correct and secure, the concatenated scheme will be correct and secure as the keys are independent and the message and transmit signal symbols are also independent. Further, this concatenation generalizes trivially to any number of independent key sets. ∎

We are now ready to give an example of the combinatorial secure groupcast problem to 2 of 4 receivers, to illustrate the main idea.

Example 3: Consider a combinatorial secure groupcast instance to 2 of 4 receivers, where the key configurations are given as follows,

$$\begin{aligned}&(H(s_1), H(s_2), H(s_{13}), H(s_{14}), H(s_{23}), H(s_{24}),\\ &H(s_{123}), H(s_{124})) = (1, 2, 2, 3, 1, 2, 2, 1).\end{aligned} \tag{35}$$

Remember that $z_k = (s_{\mathcal{U}} : k \in \mathcal{U}), k \in \{1, 2, 3, 4\}$. From Theorem 5, we have

$$\begin{aligned}C &= \min_{q \in \{1,2\}, e \in \{3,4\}} H(z_q | z_e)\\ &= \min(5, 5, 5, 5) = 5,\end{aligned} \tag{36}$$

$$\beta^*(C) = 2C - 0 - \min(2, 1) = 9. \tag{37}$$

The rate converse follows from the conditional entropy bound in Theorem 1 and the broadcast bandwidth converse follows from Theorem 2 by setting $\mathcal{Q} = \{1, 2\}$ and $u_e = z_3$ or z_4. The achievable scheme is shown in Fig. 3, where we decompose the instance into 3 sub-networks. We operate over the binary field $\mathbb{F}_2$, i.e., $p = 2$ and key block size is $L = 1$. The total number of bits in the message and the transmit signal match the converse above (5 and 9, respectively). All key bits are used, e.g., $H(s_{14}) = 3$ so that we have 3 bits of s_{14}, and sub-network 1 uses 1 bit and sub-network 3 uses 2 bits (see Fig. 3). Correctness and security are easy to verify.

D. Secure Groupcast: Symmetric Setting

We consider now the symmetric setting, where the key size only depends on the cardinality of the set of the receivers that have the same key. For any set $\mathcal{U} \subset [1:K]$ with cardinality $|\mathcal{U}| = u, u \in [1:K]$, we denote the key size as $H(s_{\mathcal{U}}) = L^{[u]} \log_2 p$. The capacity and minimum broadcast bandwidth for the symmetric setting are characterized in the following theorem.

Theorem 6 (Symmetric Setting): For the symmetric combinatorial secure groupcast problem (to the first N of K receivers), the capacity and the minimum broadcast bandwidth for capacity achieving schemes are

$$\begin{aligned}C &= \sum_{u=1}^{K} \binom{K-2}{u-1} L^{[u]} \log_2 p,\\ \beta^*(C) &= \sum_{u=1}^{K} \left(\binom{K-1}{u} - \binom{K-N-1}{u} \right) L^{[u]} \log_2 p.\end{aligned} \tag{38}$$

We refer to a key that is known to u receivers as a u-key. From the capacity and broadcast bandwidth formula, we see that it suffices to consider u-keys separately for distinct u values, i.e., joint coding across different u-keys is not necessary. This is generally not true (e.g., see Fig. 3.1) and greatly simplifies the problem. After we notice this simplification (we may limit to only u-keys of one u value), the problem still requires an intricate decomposition of the keys, depending on how many qualified receivers and how many eavesdropping receivers know the key. An example is presented below to illustrate the main idea and the full proof is presented in Section V-F.

Example 4: Consider a symmetric combinatorial secure groupcast instance to $N = 3$ of $K = 6$ receivers, where we only have 3-keys, i.e., $\binom{6}{3}$ keys of the same length $L^{[3]} = 1$.

From Theorem 6, we have

$$C = \binom{4}{2} \log_2 p = 6 \log_2 \ p,$$
$$\beta^*(C) = \binom{5}{3} \log_2 p = 10 \log_2 \ p. \tag{39}$$

The rate converse follows from the conditional entropy bound in Theorem 1, where we may pick any qualified receiver and any eavesdropping receiver such that the qualified receiver knows $\binom{4}{2}$ keys that are not known to the eavesdropping receiver. The broadcast bandwidth converse follows from Theorem 2 by setting $\mathcal{Q} = [1:3]$ and $u_e = z_6$ so that $\beta(C) \geq H(z_1, z_2, z_3|z_6) = \binom{5}{3} \log_2 p$.

The achievability is designed based on dividing the 3-keys into 3 groups.

1) The first group involves the key that is known only to 3 qualified receivers, i.e., s_{123}. As s_{123} is not known to the eavesdropping receivers, we simply send 1 message symbol with 1 symbol of one-time pad transmit signal, i.e., we have achieved $R^1 = \beta^{[1]}(R^1) = \log_2 p$.
2) The second group involves the keys that are known to 2 qualified receivers and 1 eavesdropping receiver. We need to further divide these keys depending on the set of 2 qualified receivers. Suppose the set of qualified receivers is $\{1, 2\}$, i.e., we are considering the keys $(s_{124}, s_{125}, s_{126})$ that are common to qualified Receiver 1 and qualified Receiver 2. Further, any eavesdropping receiver only knows 1 key from $(s_{124}, s_{125}, s_{126})$. In other words, we have the secure unicast situation (note that here Receiver 1 and Receiver 2 both require the same message and hold the same key) where the desired receiver has 4 equal-size combinatorial key variables while the eavesdropping receivers have 1 combinatorial key each. Therefore, combining generic linear coding ideas for key symbols in Theorem 3 and for message symbols in Theorem 4, we can send $3 - 1 = 2$ symbols securely to Receiver 1 and Receiver 2 by transmitting
$$X^{2,12} = \mathbf{V}_{12}^w W^2 + \mathbf{V}_{124}^s s_{124} + \mathbf{V}_{125}^s s_{125} + \mathbf{V}_{126}^s s_{126} \tag{40}$$
where $\mathbf{V}_{124}^s, \mathbf{V}_{125}^s, \mathbf{V}_{126}^s \in \mathbb{F}_p^{2\times 1}$ and $\mathbf{V}_{12}^w \in \mathbb{F}_p^{2\times 4}$ (the reason of setting this size will be clear soon). We repeat the same coding procedure for the other $\binom{3}{2} - 1 = 2$ sets of keys, i.e., $(s_{134}, s_{135}, s_{136})$ (common keys to qualified Receiver 1 and qualified Receiver 3) and $(s_{234}, s_{235}, s_{236})$ (common to Receiver 2 and Receiver 3). So we set
$$X^{2,13} = \mathbf{V}_{13}^w W^2 + \mathbf{V}_{134}^s s_{134} + \mathbf{V}_{135}^s s_{135} + \mathbf{V}_{136}^s s_{136} \tag{41}$$
$$X^{2,23} = \mathbf{V}_{23}^w W^2 + \mathbf{V}_{234}^s s_{234} + \mathbf{V}_{235}^s s_{235} + \mathbf{V}_{236}^s s_{236} \tag{42}$$
where $\mathbf{V}^s \in \mathbb{F}_p^{2\times 1}, \mathbf{V}^w \in \mathbb{F}_p^{2\times 4}$. From the transmit signal $X^2 = (X^{2,12}, X^{2,13}, X^{2,23})$, each qualified receiver can obtain 4 generic desired message combinations (so the size of $\mathbf{V}^w$ is chosen to match this total number of combinations), from which 4 symbols of W^2 can be recovered as long as the $\mathbf{V}^w$ matrices are chosen in a generic manner. For example, qualified Receiver 1 can obtain $(\mathbf{V}_{12}^w W^2, \mathbf{V}_{13}^w W^2)$. Security is guaranteed as long as the $\mathbf{V}^s$ matrices are chosen generically so that eavesdropping receivers see a sufficiently number of generic key combinations. To sum up, the overall rate and broadcast bandwidth achieved for all keys in the second group are
$$R^2 = \binom{3-1}{2-1} 2 \log_2 p = 4 \log_2 \ p,$$
$$\beta^2(R^2) = \binom{3}{2} 2 \log_2 p = 6 \log_2 p. \tag{43}$$
3) The third group involves the keys that are known to 1 qualified receiver and 2 eavesdropping receivers. We need to further divide these keys depending on the identity of the qualified receiver. Suppose the qualified receiver is Receiver 1, i.e., we are considering the keys $(s_{145}, s_{146}, s_{156})$ such that any eavesdropping receiver knows 2 of these 3 keys (e.g., eavesdropping Receiver 4 knows s_{145}, s_{146}). From the result of secure unicast (refer to Theorem 3), we can achieve $R = \beta(R) = (2-1) \log_2 p = \log_2 p$ for these 3 keys. Repeat the same procedure for $(s_{245}, s_{246}, s_{256})$ and $(s_{345}, s_{346}, s_{356})$. The overall rate and broadcast bandwidth achieved for all keys in the third group are
$$R^3 = \log_2 \ p, \quad \beta^3(R^3) = 3 \log_2 \ p. \tag{44}$$

Finally, we combine the performance of all 3 groups (using Lemma 1 for independent keys), so the total rate and broadcast bandwidth achieved are

$$R = R^1 + R^2 + R^3 = 6 \log_2 \ p,$$
$$\beta(R) = \beta^1(R^1) + \beta^2(R^2) + \beta^3(R^3) = 10 \log_2 p \tag{45}$$

which match the converse.

E. A Secure Groupcast Instance With $N = 2, K = 5$

For all capacity results presented above, the conditional entropy converse bound in Theorem 1 turns out to be tight. We wonder if the bound is always tight. Along this line, we find that the answer is negative. We identify a simplest setting of combinatorial secure groupcast instance to $N = 2$ of

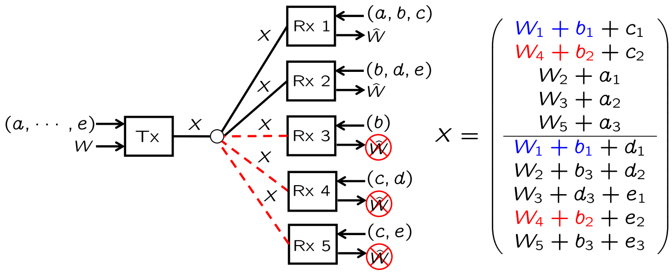

Fig. 4. A combinatorial secure groupcast instance to 2 of 5 receivers. The keys a, b, c, d, e are independent and each key has 1 bit per block. The message has 5 bits, $W = (W_1, \cdots, W_5)$, sent over 3 key blocks so that each key has 3 bits, e.g., $a = (a_1, a_2, a_3)$. The broadcast signal X has 10 bits.

$K = 5$ receivers (note that all settings with smaller N, K values are settled by the converse bound in Theorem 1) where a strictly stronger converse is required. The setting turns out to be that in Fig. 2 and is redrawn here with simplified notations (refer to Fig. 4). We have characterized its capacity and minimum broadcast bandwidth, and this result is presented in the following theorem.

Theorem 7: For the combinatorial secure groupcast instance in Fig. 4, the capacity and the minimum broadcast bandwidth for capacity achieving schemes are

$$C = 5/3, \quad \beta^*(C) = 10/3. \tag{46}$$

The achievable scheme is shown in Fig. 4. Correctness is easy to verify, e.g., qualified Receiver 1 knows a, b, c such that from the first 5 rows of the transmit signal X, he can obtain all message bits $(W_1, W_2, W_3, W_4, W_5)$. Security is more interesting. Eavesdropping Receiver 3 learns nothing because even if b is known, all message bits in X are protected by c, a, d, e (not known to Receiver 3). Eavesdropping Receiver 4 knows c, d and after canceling c, d, he can obtain $W_1 + b_1, W_4 + b_2, W_1 + b_1, W_2 + b_3$ that contains b. However, nothing is leaked because the first term is the same as the third term (highlighted in blue). In fact, this is the key of the design. Therefore,

$$\begin{aligned} I(W; X, c, d) &\overset{(1)}{=} I(W; X|c, d) && (47)\\ &= H(X|c, d) - H(X|W, c, d) \\ &= 9 - 9 = 0 && (48) \end{aligned}$$

where (47) follows from the independence of the message and the keys (refer to (1)). The situation for eavesdropping Receiver 5 is similar, where the same noise of b_2 (noise alignment) appears in the same red signal $W_4 + b_2$ (signal alignment). Interestingly, similar alignment view has been proved useful recently in several other security and privacy primitives [25]–[27].

We now discuss the rate converse. Here we give an intuitive argument for linear schemes, which guides the design of the achievable scheme, and defer the information theoretic proof to Section V-G, which is based on the sub-modularity property of entropy functions. Consider qualified Receiver 1, who knows only a, b, c and can decode W. Then W must be fully recoverable from the key variables a, b, c in X. As a is only known to Receiver 1, it can easily be used to transmit L message bits. Then to achieve rate R, the message bits carried by b, c must be $(R-1)L$ bits. As b and c are known to eavesdropping Receiver 3 and eavesdropping Receiver 4, respectively, the $(R-1)L$ message bits must be protected by both b and c. Denote these $(R-1)L$ dimensions of b as B_1. Now consider qualified Receiver 2, who knows b, d, e such that there must exist RL dimensional space of W that is covered by b, d, e. These RL dimensions must be fully covered by d, e as b is known to eavesdropping Receiver 3. As d, e have dimension L each, their overlap is $(2-R)L$ and each of them separately covers $\frac{1}{2}(R-(2-R))L = (R-1)L$ dimensions. Therefore, eavesdropping Receiver 4 can fully recover the $(R-1)L$ dimensions covered only by d (and b) as d is known. This $(R-1)L$ dimensional space of b is denoted as B_2. Symmetrically, the $(R-1)L$ space of b (mixed with e) after e is known is denoted as B_3. Finally, we connect B_1, B_2, B_3. The desired message bits in B_2, B_3 are independent, so $B_1 \cap (B_2 \cap B_3) = \emptyset$. Otherwise, in B_1, we have the same b space $(B_2 \cap B_3)$ mixed with different desired message bits (security violated). Then $L \geq \dim(B_1) + \dim(B_2 \cap B_3) \geq (R-1)L + (2(R-1)-1)L = (3R-4)L$, and $3R \leq 5$. We note that the translation of this linear argument into an information theoretic proof with entropy terms is highly non-trivial. The converse for the broadcast bandwidth is immediate, by setting $\mathcal{Q} = \{1, 2\}$ and[6] $u_e = z_3 = b$ in Theorem 2: $\beta(C) \geq 2C - I(a, b, c; b, d, e|b) = 2C - I(a, c; d, e|b) = 2C = 10/3$.

Remark 2: The conditional entropy converse bound in Theorem 1 is $R \leq 2$ for the secure groupcast instance in Fig. 4, which is strictly weaker than the capacity $5/3$. Thus the conditional entropy converse bound is not tight in general, for combinatorial secure groupcast (and for secure groupcast).

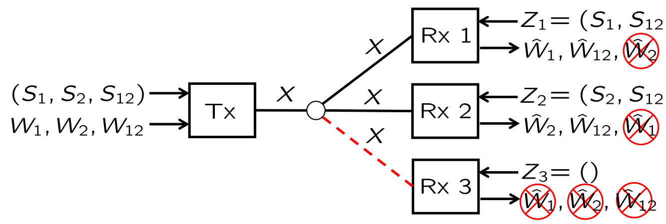

Fig. 5. A secure groupcast problem with 3 messages and 3 receivers.

IV. Generalizations

In this section, to show how insights generalize, we consider two extensions of the basic combinatorial secure groupcast model - the first one includes multiple messages and in the second one, keys are discrete memoryless sources.

A. Secure Groupcasting Multiple Messages

We consider an elementary 3 receiver broadcast network with 3 messages (see Fig. 5).

We first succinctly describe the model, which generalizes that in Section II. A transmitter wishes to deliver 3 messages

[6] Here we slightly abuse the notation. Note that in u_e, e denotes the index of an eavesdropping receiver and not the combinatorial key e in the secure groupcast instance.

W_1, W_2, W_{12} (of size $L_{W_1}, L_{W_2}, L_{W_{12}}$ i.i.d. uniform bits, respectively) through broadcasting a signal X of size L_X bits such that Receiver 1 only learns W_1, W_{12}, Receiver 2 only learns W_2, W_{12}, and Receiver 3 learns nothing. Receiver 1 is equipped with key $Z_1 = (S_1, S_{12})$, Receiver 2 is equipped with key $Z_2 = (S_2, S_{12})$ and Receiver 3's key is empty. The key variables S_1, S_2, S_{12} are L length extensions of uniform bits s_1, s_2, s_{12} (of size L_1, L_2, L_{12} bits, respectively),

$$\begin{aligned} H(W_1, W_2, W_{12}, S_1, S_2, S_{12}) &= H(W_1) + H(W_2) \\ &+ H(W_{12}) + H(S_1) + H(S_2) + H(S_{12}). \end{aligned} \tag{49}$$

The correctness and security constraints are as follows,

$$\begin{aligned} &\text{(Receiver 1)} \quad H(W_1, W_{12}|X, S_1, S_{12}) = o(L), \\ &\qquad\qquad\qquad I(W_2; X, S_1, S_{12}) = o(L). &(50) \\ &\text{(Receiver 2)} \quad H(W_2, W_{12}|X, S_2, S_{12}) = o(L), \\ &\qquad\qquad\qquad I(W_1; X, S_2, S_{12}) = o(L). &(51) \\ &\text{(Receiver 3)} \quad I(W_1, W_2, W_{12}; X) = o(L). &(52) \end{aligned}$$

The rate of the messages and the broadcast bandwidth are defined as follows,

$$\begin{aligned} &R_1 = \frac{L_{W_1}}{L},\ R_2 = \frac{L_{W_2}}{L},\ R_{12} = \frac{L_{W_{12}}}{L}, \\ &\beta(R_1, R_2, R_{12}) = \frac{L_X}{L}. \end{aligned} \tag{53}$$

The closure of the set of achievable rate tuples (R_1, R_2, R_{12}) is called the capacity region $\mathcal{C}$ and the minimum broadcast bandwidth for a rate tuple (R_1, R_2, R_{12}) is denoted as $\beta^*(R_1, R_2, R_{12})$.

We present the capacity region and the minimum broadcast bandwidth for the 3 message secure groupcast problem in Fig. 5 in the following theorem.

Theorem 8 (Rate Region): For the 3 message combinatorial secure groupcast problem in Fig. 5, the capacity region and the minimum broadcast bandwidth are

$$0 \le R_1 + R_{12} \le H(s_1) + H(s_{12}) \tag{54}$$
$$0 \le R_2 + R_{12} \le H(s_2) + H(s_{12}) \tag{55}$$
$$0 \le R_1 \le H(s_1) \tag{56}$$
$$0 \le R_2 \le H(s_2) \tag{57}$$
$$\begin{aligned} \beta^*(R_1, R_2, R_{12}) &= R_1 + R_2 \\ &+ \max(R_{12}, 2R_{12} - H(s_{12})). \end{aligned} \tag{58}$$

The proof of Theorem 8 is presented in Section V-H.

B. Discrete Memoryless Keys

The results for secure unicast and multicast with combinatorial keys generalize to discrete memoryless keys and are presented in the following theorem.

Theorem 9 (Secure Unicast and Multicast under Discrete Memoryless Keys): For the secure unicast problem (to the first of K receivers), the capacity and the minimum broadcast bandwidth for capacity achieving schemes are

$$C = \beta^*(C) = \min_{e \in [2:K]} H(z_1|z_e). \tag{59}$$

For the secure multicast problem (to the first $K-1$ of K receivers), the capacity is

$$C = \min_{q \in [1:K-1]} H(z_q|z_K). \tag{60}$$

The converse proof of Theorem 9 is identical to that under the combinatorial key setting. The achievability proof of Theorem 9 is presented in Section V-I. We give an intuitive overview here. First, consider secure unicast. Based on z_1, we wish to generate a key that is secure to any eavesdropping Receiver $e \in [2 : K]$. With a discrete memoryless source, we can use random binning (whose mapping does not depend on $z_2, \cdots, z_K$) to obtain $H(z_1|z_e)L + o(L)$ secure bits over L key blocks. This step is well known and is typically referred to as privacy amplification [28] (refer to Lemma 5 in Section V-I for a technical description). Given these secure bits, the rate value of the capacity is easily achieved by one-time pad coding. Second, consider secure multicast, which is similar, but with an additional step of communication for omniscience [4] (well known as well). This is implemented as follows. We assume the key Z_K known by the eavesdropping Receiver K is globally known (e.g., the transmitter may broadcast Z_K to everyone). Next we wish to make the qualified receivers 1 to $K-1$ all know $Z_1, \cdots, Z_{K-1}$ (i.e., common randomness). To this end, by Slepian Wolf coding [29] (random binning), the transmitter needs to broadcast $\max_{q \in [1:K-1]} H(z_1, \cdots, z_{K-1}|z_q, z_K)L + o(L)$ bits over L key blocks and note that these bits are available to the eavesdropping Receiver K as well. After this communication for omniscience step, the qualified receivers all know $Z_1, \cdots, Z_{K-1}$ so that from privacy amplification (under eavesdropped public communication), they can agree on a key of size $(H(z_1, \cdots, z_{K-1}|z_K) - \max_{q \in [1:K-1]} H(z_1, \cdots, z_{K-1}|z_q, z_K))L + o(L) = \min_{q \in [1:K-1]} H(z_q|z_K)L + o(L)$ bits that are almost unknown to the eavesdropping Receiver K (i.e., the conditional entropy subtracts the amount of leaked communication). Equipped with these secure key bits, the desired rate can be easily achieved with one-time pad coding.

Remark 3: Similar to combinatorial secure multicast (see Remark 1), the minimum broadcast bandwidth of secure multicast under the discrete memoryless key setting is an open problem. In particular, the step of communication for omniscience is not necessary and might cause additional broadcast bandwidth (this statement is also true for the key agreement problem [4]).

V. Proofs

Note that in the proofs, the relevant equations needed to justify each step are specified by the equation numbers set on top of the (in)equality symbols.

A. Proof of Theorem 1: Converse on R

Consider any qualified Receiver $q \in [1 : N]$ and any eavesdropping Receiver $e \in [N+1 : K]$. We have

$$RL \overset{(4)}{=} H(W) \tag{61}$$
$$\overset{(1)}{=} H(W|Z_e) \tag{62}$$

$$\begin{aligned}
&\overset{(2)}{=} I(W;X,Z_q|Z_e)+o(L) &(63)\\
&\overset{(3)}{=} I(W;Z_q|X,Z_e)+o(L) &(64)\\
&\leq H(Z_q|Z_e)+o(L) &(65)\\
&= LH(z_q|z_e)+o(L). &(66)
\end{aligned}$$

Note that the notation $\overset{(4)}{=}$ in (61) means that (61) follows from equation (4), introduced earlier in the paper. Normalizing (66) by L and letting $L\to\infty$, we have the desired converse bound $R\leq H(z_q|z_e)$.

B. Proof of Theorem 2: Converse on $\beta(R)$

To simplify the notations, we set $\mathcal{Q}=\{1,2,\cdots,Q\}\subset[1:N]$, which has no loss of generality. Let us start with a useful lemma.

Lemma 2: For any $q\in[1:Q-1]$, we have

$$\begin{aligned}
&I(X;Z_{q+1}|Z_1,\cdots,Z_q,U_e,W)\\
\geq\ &H(W)-I(Z_1,\cdots,Z_q;Z_{q+1}|U_e)+o(L). &(67)
\end{aligned}$$

Proof:

$$\begin{aligned}
&I(X;Z_{q+1}|Z_1,\cdots,Z_q,U_e,W)\\
\overset{(1)}{=}\ &I(X,W;Z_{q+1}|Z_1,\cdots,Z_q,U_e) &(68)\\
=\ &I(X,W,Z_1,\cdots,Z_q;Z_{q+1}|U_e)\\
&-I(Z_1,\cdots,Z_q;Z_{q+1}|U_e) &(69)\\
\geq\ &I(W;Z_{q+1}|U_e,X)-I(Z_1,\cdots,Z_q;Z_{q+1}|U_e) &(70)\\
\overset{(2)}{=}\ &H(W|U_e,X)-I(Z_1,\cdots,Z_q;Z_{q+1}|U_e)+o(L) &(71)\\
\overset{(3)}{=}\ &H(W|U_e)-I(Z_1,\cdots,Z_q;Z_{q+1}|U_e)+o(L) &(72)\\
\overset{(1)}{=}\ &H(W)-I(Z_1,\cdots,Z_q;Z_{q+1}|U_e)+o(L). &(73)
\end{aligned}$$

■

Next, we apply Lemma 2 to decompose the term $I(X;W,Z_1,\cdots,Z_Q|U_e)$. We have

$$\begin{aligned}
&I(X;W,Z_1,\cdots,Z_Q|U_e)\\
=\ &I(X;W,Z_1|U_e)+\sum_{q=1}^{Q-1}I(X;Z_{q+1}|Z_1,\cdots,Z_q,U_e,W) &(74)\\
\overset{(67)}{\geq}\ &I(X;W|Z_1,U_e)\\
&+\sum_{q=1}^{Q-1}\Big(H(W)-I(Z_1,\cdots,Z_q;Z_{q+1}|U_e)+o(L)\Big) &(75)\\
\overset{(1)}{=}\ &I(X,Z_1,U_e;W)\\
&+\sum_{q=1}^{Q-1}\Big(H(W)-I(Z_1,\cdots,Z_q;Z_{q+1}|U_e)+o(L)\Big) &(76)\\
\overset{(2)}{=}\ &H(W)+o(L)\\
&+\sum_{q=1}^{Q-1}\Big(H(W)-I(Z_1,\cdots,Z_q;Z_{q+1}|U_e)+o(L)\Big) &(77)\\
=\ &QH(W)-\sum_{q=1}^{Q-1}I(Z_1,\cdots,Z_q;Z_{q+1}|U_e)+o(L). &(78)
\end{aligned}$$

Finally, note that

$$\begin{aligned}
\beta(R)L\ &\overset{(5)}{=}\ L_X\log_2 p &(79)\\
&\geq\ H(X) &(80)\\
&\geq\ H(X|U_e) &(81)\\
&\geq\ I(X;W,Z_1,\cdots,Z_Q|U_e). &(82)
\end{aligned}$$

Combining (78), (82), we have

$$\beta(R)L\geq QH(W)-\sum_{q=1}^{Q-1}I(Z_1,\cdots,Z_q;Z_{q+1}|U_e)+o(L). \tag{83}$$

Normalizing (83) by L and letting $L\to\infty$, we have the desired converse bound on $\beta(R)$. Note that as the keys are discrete memoryless, $I(Z_1,\cdots,Z_q;Z_{q+1}|U_e)=LI(z_1,\cdots,z_q;z_{q+1}|u_e)$.

C. Proof of Theorem 3: Secure Unicast

The converse bounds for the capacity and the broadcast bandwidth follow immediately from Theorem 1 and Theorem 2, respectively. The achievable scheme is presented as follows.

We show that rate $R=\min_{e\in[2:K]}\sum_{\mathcal{U}\subset[1:K]:1\in\mathcal{U},e\notin\mathcal{U}}H(s_{\mathcal{U}})$ and broadcast bandwidth $\beta(R)=R$ are achievable. Set $L_W=R/\log_2 p$ and $L=1$. Set the field size p to be the least prime power such that $p\geq L_W+\sum_{\mathcal{U}\subset[1:K]:1\in\mathcal{U}}L_{\mathcal{U}}$ and

$$\begin{aligned}
X\ &=\ W+\sum_{\mathcal{U}\subset[1:K]:1\in\mathcal{U}}\mathbf{V}_{\mathcal{U}}s_{\mathcal{U}}\\
&=\ W+\underbrace{\Big[\mathbf{V}_1\ \mathbf{V}_{12}\ \cdots\ \mathbf{V}_{1:K}\Big]}_{\triangleq\mathbf{V}}\begin{bmatrix}s_1\\ s_{12}\\ \vdots\\ s_{1:K}\end{bmatrix} &(84)
\end{aligned}$$

where $X,W\in\mathbb{F}_p^{L_W\times 1}$, $\mathbf{V}_{\mathcal{U}}\in\mathbb{F}_p^{L_W\times L_{\mathcal{U}}}$, $s_{\mathcal{U}}\in\mathbb{F}_p^{L_{\mathcal{U}}\times 1}$ and $\mathbf{V}$ is chosen as a full-rank Cauchy matrix of dimension $L_W\times\sum_{\mathcal{U}\subset[1:K]:1\in\mathcal{U}}L_{\mathcal{U}}$ such that the element in i-th row and j-th column is given by

$$\mathbf{V}(i,j)=\frac{1}{a_i-b_j},\quad a_i,b_j \text{ are distinct elements over } \mathbb{F}_p. \tag{85}$$

The correctness constraint (2) is trivially satisfied. We verify the security constraint (3). Consider any eavesdropping

Receiver $e \in [2:K]$,

$$\begin{aligned} & I(W;X,Z_e) \\ \stackrel{(1)(84)}{=} & I(W;W+\sum_{\mathcal{U}\subset[1:K]:1\in\mathcal{U}} \mathbf{V}_{\mathcal{U}}s_{\mathcal{U}}|z_e) && (86) \\ = & I\Big(W;W+\sum_{\mathcal{U}\subset[1:K]:1\in\mathcal{U},e\notin\mathcal{U}} \mathbf{V}_{\mathcal{U}}s_{\mathcal{U}}|(s_{\mathcal{U}}:e\in\mathcal{U})\Big) && (87) \\ \stackrel{(1)(6)}{=} & I\Big(W;W+\sum_{\mathcal{U}\subset[1:K]:1\in\mathcal{U},e\notin\mathcal{U}} \mathbf{V}_{\mathcal{U}}s_{\mathcal{U}}\Big) && (88) \\ \stackrel{(1)}{=} & H\Big(W+\sum_{\mathcal{U}\subset[1:K]:1\in\mathcal{U},e\notin\mathcal{U}} \mathbf{V}_{\mathcal{U}}s_{\mathcal{U}}\Big) \\ & -H\Big(\sum_{\mathcal{U}\subset[1:K]:1\in\mathcal{U},e\notin\mathcal{U}} \mathbf{V}_{\mathcal{U}}s_{\mathcal{U}}\Big) && (89) \\ \leq & L_W\log_2 p - L_W\log_2 p = 0 && (90) \end{aligned}$$

where in the last step, the first term of $L_W\log_2 p$ follows from the fact that the vector has L_W symbols from $\mathbb{F}_p$ and the second term of $-L_W\log_2 p$ follows from the fact that the sub-matrix $[\mathbf{V}_{\mathcal{U}}:\mathcal{U}\subset[1:K], 1\in\mathcal{U}, e\notin\mathcal{U}]$ of the Cauchy matrix $\mathbf{V}$ has rank L_W (as it has at least L_W columns and exactly L_W rows) and $s_{\mathcal{U}}$ are i.i.d. uniform symbols.

Finally, the rate and broadcast bandwidth achieved match the converse such that the proof of Theorem 3 is complete.

D. Proof of Theorem 4: Secure Multicast

We first consider the capacity of combinatorial secure multicast. The converse follows directly from Theorem 1 and we consider the achievability.

We show that rate $R = \min_{q\in[1:K-1]} \sum_{\mathcal{U}\subset[1:K-1]:q\in\mathcal{U}} H(s_{\mathcal{U}})$ is achievable. Set $L_W = R/\log_2 p$ and $L=1$. Set the field size p to be the least prime power such that $p \geq L_W + \sum_{\mathcal{U}\subset[1:K-1]} L_{\mathcal{U}}$. The transmit signal has $2^{K-1}-1$ row blocks, each corresponding to a subset of $[1:K-1]$,

$$\begin{aligned} X &= [X_1;X_2;\cdots;X_{\mathcal{U}};\cdots;X_{1:K-1}] \\ X_{\mathcal{U}} &= \mathbf{V}_{\mathcal{U}}W + s_{\mathcal{U}}, \forall \mathcal{U}\subset[1:K-1] \quad (91) \end{aligned}$$

where $X_{\mathcal{U}} \in \mathbb{F}_p^{L_{\mathcal{U}}\times 1}, \mathbf{V}_{\mathcal{U}} \in \mathbb{F}_p^{L_{\mathcal{U}}\times L_W}, W \in \mathbb{F}_p^{L_W\times 1}$, $s_{\mathcal{U}} \in \mathbb{F}_p^{L_{\mathcal{U}}\times 1}$ and the precoding matrices $\mathbf{V}_{\mathcal{U}}$ are sub-matrices of a Cauchy matrix, set as follows,

$$\mathbf{V} = [\mathbf{V}_1;\mathbf{V}_2;\cdots;\mathbf{V}_{\mathcal{U}};\cdots;\mathbf{V}_{1:K-1}]_{\sum_{\mathcal{U}\subset[1:K-1]} L_{\mathcal{U}}\times L_W}, \quad (92)$$

$$\mathbf{V}(i,j) = \frac{1}{a_i - b_j}, \quad a_i, b_j \text{ are distinct elements over } \mathbb{F}_p. \quad (93)$$

Security is guaranteed because the keys known to the eavesdropping Receiver K do not appear in the transmit signal and the $s_{\mathcal{U}}$ variables are independent. Correctness constraint is satisfied because each qualified Receiver $q \in [1:K-1]$ can recover at least L_W linear combinations of the message symbols W, i.e., $(\mathbf{V}_{\mathcal{U}}W : q\in\mathcal{U}\subset[1:K-1])$, from which W can be decoded as any sub-matrix of a full-rank Cauchy matrix has full rank.

Next we proceed to the minimum broadcast bandwidth of combinatorial secure multicast. The broadcast bandwidth achieved by the scheme above is $\beta(C) = \sum_{\mathcal{U}\subset[1:K-1]} H(s_{\mathcal{U}})$, which is optimal when $H(z_1|z_K) = \cdots = H(z_{K-1}|z_K)$. This follows from Theorem 2, where we set $\mathcal{Q} = [1:K-1]$ and $u_e = z_K$. We have

$$\begin{aligned} \beta(C) &\geq (K-1)C \\ &\quad - \Big(\sum_{q=1}^{K-1} H(z_q|z_K) - H(z_1,\cdots,z_{K-1}|z_K)\Big) && (94) \\ &= H(z_1,\cdots,z_{K-1}|z_K) = \sum_{\mathcal{U}\subset[1:K-1]} H(s_{\mathcal{U}}) && (95) \end{aligned}$$

where (95) follows from the fact that $C = H(z_1|z_K) = \cdots = H(z_{K-1}|z_K)$. The proof when $K=4$ is more involved as we need to improve the above achievable scheme, and is presented next.

1) $\beta^*(C)$ When $K=4$: First, we consider the converse. Note that in this case, $C = H(z_1|z_4)$. From (25), we need two converse bounds. The first one is obtained by setting $\mathcal{Q} = \{1,2\}, u_e = z_4$ in Theorem 2,

$$\begin{aligned} \beta(C) &\geq 2C - I(z_1;z_2|z_4) = 2H(z_1|z_4) - I(z_1;z_2|z_4) && (96) \\ &= 2H(s_1) + 2H(s_{12}) + 2H(s_{13}) + 2H(s_{123}) \\ &\quad - (H(s_{12}) + H(s_{123})) && (97) \\ &= 2H(s_1) + H(s_{12}) + 2H(s_{13}) + H(s_{123}). && (98) \end{aligned}$$

The second one is obtained by setting $\mathcal{Q} = \{1,2,3\}$, $u_e = z_4$ in Theorem 2,

$$\begin{aligned} \beta(C) &\geq 3C - (I(z_1;z_2|z_4) + I(z_1,z_2;z_3|z_4)) && (99) \\ &= 3H(s_1) + 3H(s_{12}) + 3H(s_{13}) + 3H(s_{123}) \\ &\quad - (H(s_{12}) + H(s_{13}) + H(s_{23}) + 2H(s_{123})) \\ &= 3H(s_1) + 2H(s_{12}) + 2H(s_{13}) - H(s_{23}) + H(s_{123}). && (100) \end{aligned}$$

Second, we consider the achievability, where we need to adjust the size of the keys used in (91) depending on the key configuration. We present the scheme that achieves rate $R = H(z_1|z_4) = H(s_1) + H(s_{12}) + H(s_{13}) + H(s_{123})$. Set $L_W = R/\log_2 p$ and $L=1$.

We have 3 cases depending on the relationship between $H(s_{23}), H(s_1) + H(s_{13}), H(s_1) + H(s_{12})$. For each case, set the field size p to be the least prime power such that $p \geq L_X + L_W$. The transmit signal has 7 row blocks and the 4 blocks $X_1, X_{12}, X_{13}, X_{123}$ (corresponding to the keys known to Receiver 1) are the same for all 3 cases, where all key symbols are used,

$$\begin{aligned} X &= [X_1;X_2;X_3;X_{12};X_{13};X_{23};X_{123}] \\ X_1 &= \mathbf{V}_1W + s_1, X_{12} = \mathbf{V}_{12}W + s_{12}, \\ X_{13} &= \mathbf{V}_{13}W + s_{13}, X_{123} = \mathbf{V}_{123}W + s_{123} \quad (101) \end{aligned}$$

where the sizes of the matrices and vectors above are the same as before (see (91)). Note that now Receiver 1 can achieve rate R and in the remaining proof, we only need to consider

Receiver 2 and Receiver 3. The remaining blocks X_2, X_3, X_{23} are designed for each case separately, where not all the key symbols may be used. Note that $H(s_{12}) \leq H(s_{13})$, i.e., $L_{12} \leq L_{13}$.

Case 1. $H(s_{23}) \geq H(s_1) + H(s_{13})$.

$$X_2 = (), X_3 = (), X_{23} = \mathbf{V}_{23}^{[1]} W + s_{23}^{[1]} \tag{102}$$

where $s_{23}^{[1]} \in \mathbb{F}_p^{(L_1+L_{13})\times 1}$ is comprised of the first $L_1 + L_{13}$ symbols from s_{23} (which has more symbols, i.e., $L_{23} \geq L_1 + L_{13}$) and $\mathbf{V}_{23}^{[1]} \in \mathbb{F}_p^{(L_1+L_{13})\times L_W}$. We verify that the number of linear combinations of the message decodable by Receiver 2 and Receiver 3 is no less than L_W.

$$\begin{aligned}&\text{For Receiver 2: } |X_{12}| + |X_{23}| + |X_{123}| \\ &\quad = L_{12} + (L_1 + L_{13}) + L_{123} = L_W, \end{aligned}\tag{103}$$

$$\begin{aligned}&\text{For Receiver 3: } |X_{13}| + |X_{23}| + |X_{123}| \\ &\quad = L_{13} + (L_1 + L_{13}) + L_{123} \geq L_W. \end{aligned}\tag{104}$$

The transmit signal size is

$$\begin{aligned}L_X &= |X_1| + |X_{12}| + |X_{13}| + |X_{123}| + |X_{23}| \\ &= L_1 + L_{12} + L_{13} + L_{123} + (L_1 + L_{13}) \end{aligned}\tag{105}$$

which matches the converse bound (98) for broadcast bandwidth. The other cases are similar.

Case 2. $H(s_1) + H(s_{12}) \leq H(s_{23}) \leq H(s_1) + H(s_{13})$.

$$X_2 = \mathbf{V}_2^{[2]} W + s_2^{[2]}, X_3 = (), X_{23} = \mathbf{V}_{23} W + s_{23} \tag{106}$$

where $s_2^{[2]} \in \mathbb{F}_p^{(L_1+L_{13}-L_{23})\times 1}$ is comprised of the first $L_1 + L_{13} - L_{23}$ symbols from s_2 (which has more symbols, i.e., $L_2 \geq L_1 + L_{13} - L_{23}$ because $H(z_1|z_4) \leq H(z_2|z_4)$) and $\mathbf{V}_2^{[2]} \in \mathbb{F}_p^{(L_1+L_{13}-L_{23})\times L_W}$. We verify that the number of linear combinations of the message decodable by Receiver 2 and Receiver 3 is no less than L_W.

$$\begin{aligned}&\text{For Receiver 2: } |X_2| + |X_{12}| + |X_{23}| + |X_{123}| \\ &\quad = (L_1 + L_{13} - L_{23}) + L_{12} + L_{23} + L_{123} \\ &\quad = L_1 + L_{12} + L_{13} + L_{123} = L_W, \end{aligned}\tag{107}$$

$$\begin{aligned}&\text{For Receiver 3: } |X_{13}| + |X_{23}| + |X_{123}| \\ &\quad = L_{13} + L_{23} + L_{123} \geq L_W. \end{aligned}\tag{108}$$

The transmit signal size is

$$\begin{aligned}L_X &= |X_1| + |X_{12}| + |X_{13}| + |X_{123}| \\ &\quad + |X_2| + |X_{23}| \end{aligned}\tag{109}$$

$$\begin{aligned} &= L_1 + L_{12} + L_{13} + L_{123} \\ &\quad + (L_1 + L_{13} - L_{23}) + L_{23} \end{aligned}\tag{110}$$

which matches the converse bound (98) for broadcast bandwidth.

Case 3. $H(s_{23}) \leq H(s_1) + H(s_{12})$.

$$\begin{aligned}&X_2 = \mathbf{V}_2^{[3]} W + s_2^{[3]}, X_3 = \mathbf{V}_3^{[3]} W + s_3^{[3]}, \\ &X_{23} = \mathbf{V}_{23} W + s_{23} \end{aligned}\tag{111}$$

$$s_2^{[3]} \in \mathbb{F}_p^{(L_1+L_{13}-L_{23})\times 1}, \mathbf{V}_2^{[3]} \in \mathbb{F}_p^{(L_1+L_{13}-L_{23})\times L_W} \tag{112}$$

$$s_3^{[3]} \in \mathbb{F}_p^{(L_1+L_{12}-L_{23})\times 1}, \mathbf{V}_3^{[3]} \in \mathbb{F}_p^{(L_1+L_{12}-L_{23})\times L_W} \tag{113}$$

where $L_2 \geq L_1 + L_{13} - L_{23}$ and $L_3 \geq L_1 + L_{12} - L_{23}$ because $H(z_1|z_4) \leq \min(H(z_2|z_4), H(z_3|z_4))$. We verify that the number of linear combinations of the message decodable by Receiver 2 and Receiver 3 is no less than L_W.

$$\begin{aligned}&\text{For Receiver 2: } |X_2| + |X_{12}| + |X_{23}| + |X_{123}| \\ &\quad = (L_1 + L_{13} - L_{23}) + L_{12} + L_{23} + L_{123} \\ &\quad = L_1 + L_{12} + L_{13} + L_{123} = L_W, \end{aligned}\tag{114}$$

$$\begin{aligned}&\text{For Receiver 3: } |X_3| + |X_{13}| + |X_{23}| + |X_{123}| \\ &\quad = (L_1 + L_{12} - L_{23}) + L_{13} + L_{23} + L_{123} \\ &\quad = L_1 + L_{12} + L_{13} + L_{123} = L_W. \end{aligned}\tag{115}$$

The transmit signal size is

$$\begin{aligned}L_X &= |X_1| + |X_{12}| + |X_{13}| + |X_{123}| \\ &\quad + |X_2| + |X_3| + |X_{23}| \end{aligned}\tag{116}$$

$$\begin{aligned} &= L_1 + L_{12} + L_{13} + L_{123} + (L_1 + L_{13} - L_{23}) \\ &\quad + (L_1 + L_{12} - L_{23}) + L_{23} \end{aligned}\tag{117}$$

which matches the converse bound (100) for broadcast bandwidth.

After the sizes are specified, the remaining proof is the same as that presented above, where we set $\mathbf{V}$ as a full-rank Cauchy matrix and the correctness and security constraints are satisfied. The proof of $\beta^*(C)$ when $K = 4$ is thus complete.

E. Proof of Theorem 5: the $N = 2, K = 4$ Case

The converse of rate and broadcast bandwidth follows from Theorem 1 and Theorem 2, and we present the achievable scheme now. Set $p = 2$ (i.e., binary field) and $L = 1$ (i.e., one key block). The idea is to decompose every instance into multiple component sub-networks, where the basic components are listed in Fig. 6.

Next, we divide the problem instance into multiple cases, where each case requires a different decomposition. As s_{12} is only known to qualified Receiver 1 and Receiver 2, we can easily use s_{12} to achieve rate $H(s_{12})$ with broadcast bandwidth $H(s_{12})$, by one-time pad. Without loss of generality, we assume $H(s_1) \leq H(s_2), H(s_{124}) \leq H(s_{123})$. As such, for all cases we invoke Component 1 $H(s_{124})$ times (i.e., use $H(s_{124})$ bits of s_{123}, s_{124}) and Component 2 $H(s_1)$ times (use $H(s_1)$ bits of s_1, s_2). We write this succinctly as

$$H(s_{124}) \times \text{ Cmp 1 } + H(s_1) \times \text{ Cmp 2.} \tag{118}$$

We proceed next depending on the key configurations. Specifically, all cases are divided as follows.

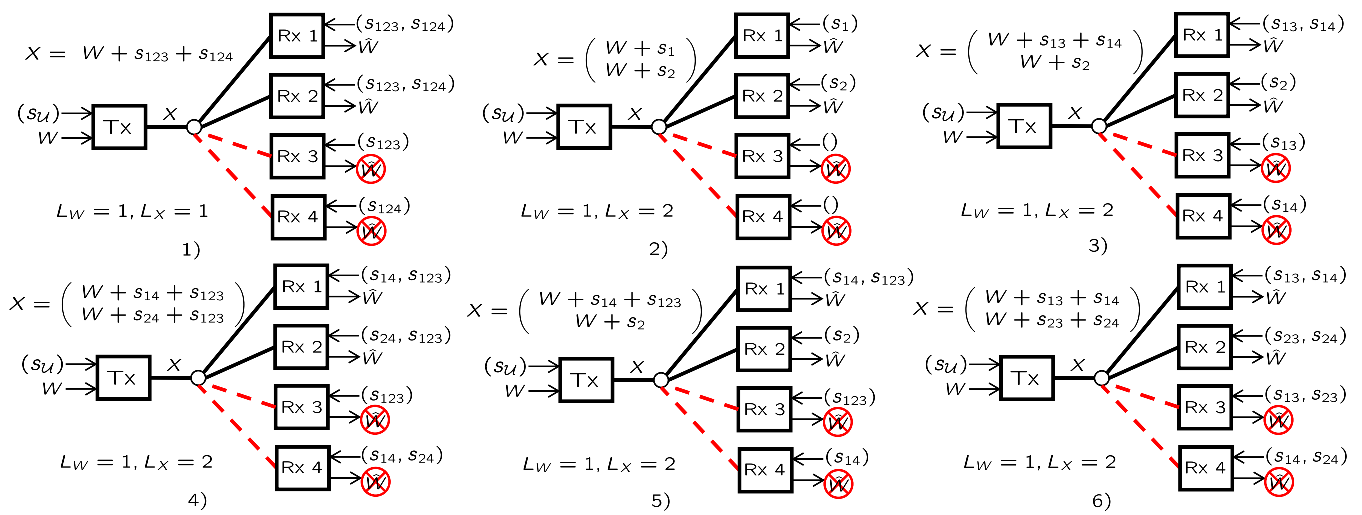

Fig. 6. The 6 basic components with (correct and secure) achievable schemes.

Case 1. $H(s_2) - H(s_1) \geq \min(H(s_{13}), H(s_{14}))$. We further invoke

$$\min(H(s_{13}), H(s_{14})) \times \text{ Cmp } 3 \tag{119}$$

where we can employ the scheme in Component 3 a number of $\min(H(s_{13}), H(s_{14}))$ times because we have $H(s_2) - H(s_1)$ bits left of s_2 and $H(s_2) - H(s_1) \geq \min(H(s_{13}), H(s_{14}))$. The remaining steps need further division.

1) Case 1.1. $H(s_{14}) \leq H(s_{13})$. No further action is needed. Tracing back, we have invoked one-time pad of s_{12}, (118), and (119). Therefore we have achieved

$$R = H(s_{12}) + H(s_{124}) + H(s_1) + H(s_{14}),$$
$$\beta(R) = H(s_{12}) + H(s_{124}) + 2H(s_1) + 2H(s_{14})$$

which match the converse (33) and (34). Note that the converse bounds are minimum or maximum of several terms and it suffices to show the achievability of one term.

2) Case 1.2. $H(s_{14}) \geq H(s_{13})$. We need to further consider the following cases.

3) Case 1.2.1. $\min(H(s_{14}) - H(s_{13}), H(s_{123}) - H(s_{124}), H(s_{24})) = H(s_{14}) - H(s_{13})$. We further invoke

$$(H(s_{14}) - H(s_{13})) \times \text{ Cmp } 4 \tag{120}$$

and the description of the scheme is complete for this case. We trace back and find that

$$R = H(s_{12}) + H(s_{124}) + H(s_1) + H(s_{14}),$$
$$\beta(R) = H(s_{12}) + H(s_{124}) + 2H(s_1) + 2H(s_{14})$$

are achieved and they are optimal as the formulas match the converse.

4) Case 1.2.2. $\min(H(s_{14}) - H(s_{13}), H(s_{123}) - H(s_{124}), H(s_{24})) = H(s_{123}) - H(s_{124})$. We further invoke

$$(H(s_{123}) - H(s_{124})) \times \text{ Cmp } 4 \tag{121}$$

and the description of the scheme is complete for this case. We trace back and find that

$$R = H(s_{12}) + H(s_1) + H(s_{13}) + H(s_{123}), \tag{122}$$
$$\beta(R) = H(s_{12}) - H(s_{124}) + 2H(s_1) + 2H(s_{13}) + 2\,H(s_{123}) \tag{123}$$

are achieved and they are optimal as the formulas match the converse.

5) Case 1.2.3. $\min(H(s_{14}) - H(s_{13}), H(s_{123}) - H(s_{124}), H(s_{24})) = H(s_{24})$. We further invoke

$$H(s_{24}) \times \text{ Cmp } 4 \tag{124}$$

and need to consider the following cases.

Case 1.2.3.1. $\min(H(s_2) - H(s_1) - H(s_{13}), H(s_{14}) - H(s_{13}) - H(s_{24}), H(s_{123}) - H(s_{124}) - H(s_{24})) = H(s_2) - H(s_1) - H(s_{13})$. We further invoke

$$\big(H(s_2) - H(s_1) - H(s_{13})\big) \times \text{ Cmp } 5$$
such that overall (125)
$$R = H(s_{12}) + H(s_{124}) + H(s_{24}) + H(s_2), \tag{126}$$
$$\beta(R) = H(s_{12}) + H(s_{124}) + 2H(s_{24}) + 2H(s_2). \tag{127}$$

Case 1.2.3.2. $\min(H(s_2) - H(s_1) - H(s_{13}), H(s_{14}) - H(s_{13}) - H(s_{24}), H(s_{123}) - H(s_{124}) - H(s_{24})) = H(s_{14}) - H(s_{13}) - H(s_{24})$. We further invoke

$$\big(H(s_{14}) - H(s_{13}) - H(s_{24})\big) \times \text{ Cmp } 5$$
such that overall (128)
$$R = H(s_{12}) + H(s_{124}) + H(s_1) + H(s_{14}), \tag{129}$$
$$\beta(R) = H(s_{12}) + H(s_{124}) + 2H(s_1) + 2H(s_{14}). \tag{130}$$

Case 1.2.3.3. $\min(H(s_2) - H(s_1) - H(s_{13}), H(s_{14}) - H(s_{13}) - H(s_{24}), H(s_{123}) -$

$H(s_{124})-H(s_{24})) = H(s_{123})-H(s_{124})-H(s_{24})$. We further invoke

$$\big(H(s_{123}) - H(s_{124}) - H(s_{24})\big) \times \text{ Cmp 5 such that overall} \quad (131)$$
$$R = H(s_{12}) + H(s_1) + H(s_{13}) + H(s_{123}), \quad (132)$$
$$\beta(R) = H(s_{12}) - H(s_{124}) + 2H(s_1) + 2H(s_{13}) + 2H(s_{123}). \quad (133)$$

Case 2. $H(s_2) - H(s_1) \leq \min(H(s_{13}), H(s_{14}))$. We further invoke

$$\big(H(s_2) - H(s_1)\big) \times \text{ Cmp 3} \quad (134)$$

and consider the following cases.

6) Case 2.1. $\min(H(s_{24}), H(s_{14}) - H(s_2) + H(s_1), H(s_{123}) - H(s_{124})) = H(s_{24})$. We further invoke

$$H(s_{24}) \times \text{ Cmp 4} \quad (135)$$

and the description of the scheme is complete for this case. We trace back and find that

$$R = H(s_{12}) + H(s_{124}) + H(s_2) + H(s_{24}), \quad (136)$$
$$\beta(R) = H(s_{12}) + H(s_{124}) + 2H(s_2) + 2H(s_{24}) \quad (137)$$

are achieved and they are optimal as the formulas match the converse.

7) Case 2.2. $\min(H(s_{24}), H(s_{14}) - H(s_2) + H(s_1), H(s_{123}) - H(s_{124})) = H(s_{14}) - H(s_2) + H(s_1)$. We further invoke

$$\big(H(s_{14}) - H(s_2) + H(s_1)\big) \times \text{ Cmp 4} \quad (138)$$

and the description of the scheme is complete for this case. We trace back and find that

$$R = H(s_{12}) + H(s_{124}) + H(s_{14}) + H(s_1), \quad (139)$$
$$\beta(R) = H(s_{12}) + H(s_{124}) + 2H(s_{14}) + 2H(s_1) \quad (140)$$

are achieved and they are optimal as the formulas match the converse.

8) Case 2.3. $\min(H(s_{24}), H(s_{14}) - H(s_2) + H(s_1), H(s_{123}) - H(s_{124})) = H(s_{123}) - H(s_{124})$. We further invoke

$$\big(H(s_{123}) - H(s_{124})\big) \times \text{ Cmp 4} \quad (141)$$

and need to consider the following cases.

9) Case 2.3.1. $\min(H(s_{14}) - H(s_2) + H(s_1) - H(s_{123}) + H(s_{124}), H(s_{13}) - H(s_2) + H(s_1), H(s_{24}) - H(s_{123}) + H(s_{124}), H(s_{23})) = H(s_{14}) - H(s_2) + H(s_1) - H(s_{123}) + H(s_{124})$. We further invoke

$$\big(H(s_{14}) - H(s_2) + H(s_1) - H(s_{123}) + H(s_{124})\big) \times \text{ Cmp 6 such that overall} \quad (142)$$
$$R = H(s_{12}) + H(s_{124}) + H(s_{14}) + H(s_1), \quad (143)$$
$$\beta(R) = H(s_{12}) + H(s_{124}) + 2H(s_{14}) + 2H(s_1). \quad (144)$$

10) Case 2.3.2. $\min(H(s_{14}) - H(s_2) + H(s_1) - H(s_{123}) + H(s_{124}), H(s_{13}) - H(s_2) + H(s_1), H(s_{24}) - H(s_{123}) + H(s_{124}), H(s_{23})) = H(s_{13}) - H(s_2) + H(s_1)$. We further invoke

$$\big(H(s_{13}) - H(s_2) + H(s_1)\big) \times \text{ Cmp 6 such that overall} \quad (145)$$
$$R = H(s_{12}) + H(s_{123}) + H(s_{13}) + H(s_1), \quad (146)$$
$$\beta(R) = H(s_{12}) - H(s_{124}) + 2\ H(s_{123}) + 2H(s_{13}) + 2H(s_1). \quad (147)$$

11) Case 2.3.3. $\min(H(s_{14}) - H(s_2) + H(s_1) - H(s_{123}) + H(s_{124}), H(s_{13}) - H(s_2) + H(s_1), H(s_{24}) - H(s_{123}) + H(s_{124}), H(s_{23})) = H(s_{24}) - H(s_{123}) + H(s_{124})$. We further invoke

$$\big(H(s_{24}) - H(s_{123}) + H(s_{124})\big) \times \text{ Cmp 6 such that overall} \quad (148)$$
$$R = H(s_{12}) + H(s_{124}) + H(s_2) + H(s_{24}) \quad (149)$$
$$\beta(R) = H(s_{12}) + H(s_{124}) + 2\ H(s_2) + 2H(s_{24}) \quad (150)$$

12) Case 2.3.4. $\min(H(s_{14}) - H(s_2) + H(s_1) - H(s_{123}) + H(s_{124}), H(s_{13}) - H(s_2) + H(s_1), H(s_{24}) - H(s_{123}) + H(s_{124}), H(s_{23})) = H(s_{23})$. We further invoke

$$H(s_{23}) \times \text{ Cmp 6 such that overall} \quad (151)$$
$$R = H(s_{12}) + H(s_2) + H(s_{123}) + H(s_{23}), \quad (152)$$
$$\beta(R) = H(s_{12}) - H(s_{124}) + 2\ H(s_2) + 2H(s_{123}) + 2H(s_{23}). \quad (153)$$

F. Proof of Theorem 6: the Symmetric Setting

The rate converse follows from Theorem 1, where among u-keys, any qualified Receiver $q \in [1 : N]$ knows $\binom{K-2}{u-1}$ keys that are not known to any eavesdropping Receiver $e \in [N+1 : K]$ because we may pick any $u-1$ receivers from any $K-2$ receivers other than Receiver q and Receiver e to form a u-key (note that Receiver q must be included). The broadcast bandwidth converse follows from Theorem 2, where we set $\mathcal{Q} = [1 : N]$, $u_e = z_K$ and obtain $\beta(C) \geq H(z_1, \cdots, z_N | z_K)$. Among u-keys, we have $\binom{K-1}{u} - \binom{K-N-1}{u}$ keys in the term $H(z_1, \cdots, z_N | z_K)$ because we pick u-keys from receivers 1 to $K-1$ and need to remove the ones that are only known to receivers $N+1$ to $K-1$.

To sum up for the converse part, we have proved that

$$C \leq \sum_{u=1}^{K} \binom{K-2}{u-1} L^{[u]} \log_2 \ p,$$
$$\beta^*(C) \geq \sum_{u=1}^{K} \left(\binom{K-1}{u} - \binom{K-N-1}{u} \right) L^{[u]} \log_2 \ p. \tag{154}$$

We next show that the above rate and broadcast bandwidth are achievable. Similar to Example 4, we consider u-keys separately for different u values and then combine the decomposed schemes using Lemma 1. Consider a fixed value of $u \in [1:K]$ and further consider the u-keys that are known to i qualified receivers and $u-i$ eavesdropping receivers, where $i \in [1:u]$.

We focus on one specific set of i qualified receivers, say receivers from the set $\mathcal{I}$ where $\mathcal{I} \subset [1:N], |\mathcal{I}| = i$. That is, we consider the keys $(s_{\mathcal{U}} : [1:N] \cap \mathcal{U} = \mathcal{I}, |\mathcal{U}| = u)$ and there are $\binom{K-N}{u-i}$ such u-keys. Further these $\binom{K-N}{u-i}$ keys are known to all qualified receivers from $\mathcal{I}$, and each eavesdropping receiver knows $\binom{K-N-1}{u-i-1}$ keys from these keys. Invoking generic linear codes similar to Example 4, we can securely send $(\binom{K-N}{u-i} - \binom{K-N-1}{u-i-1})L^{[u]} = \binom{K-N-1}{u-i}L^{[u]}$ generic message symbols to receivers from $\mathcal{I}$ by transmitting $\binom{K-N-1}{u-i}L^{[u]}$ symbols.

$$X^{[u],\mathcal{I}} = \mathbf{V}_{\mathcal{I}}^{w} W^{[u]} + \sum_{\mathcal{U}:[1:N]\cap\mathcal{U}=\mathcal{I}, |\mathcal{U}|=u} \mathbf{V}_{\mathcal{U}}^{s} s_{\mathcal{U}} \tag{155}$$

where $\mathbf{V}_{\mathcal{I}}^{w}$ is a $\binom{K-N-1}{u-i}L^{[u]} \times \binom{N-1}{i-1}\binom{K-N-1}{u-i}L^{[u]}$ matrix over $\mathbb{F}_p$ and $\mathbf{V}_{\mathcal{U}}^{s}$ is a $\binom{K-N-1}{u-i}L^{[u]} \times L^{[u]}$ matrix over $\mathbb{F}_p$. We now explain the dimension of $\mathbf{V}_{\mathcal{I}}^{w}$. The number of rows of $\mathbf{V}_{\mathcal{I}}^{w}$ corresponds to the number of message symbols in $X^{[u],\mathcal{I}}$, which is set to be equal to the number of key symbols that are known to all qualified receivers from $\mathcal{I}$ minus the number of key symbols among these keys that are known to each eavesdropper, i.e., $(\binom{K-N}{u-i} - \binom{K-N-1}{u-i-1})L^{[u]} = \binom{K-N-1}{u-i}L^{[u]}$. The number of columns of $\mathbf{V}_{\mathcal{I}}^{w}$ corresponds to the total number of message symbols that are sent under the protection of u-keys (i.e., $H(W^{[u]})$). Consider all sets of i qualified receivers (i.e., all $\mathcal{I}$ sets), there are $\binom{N}{i}$ choices and any particular qualified receiver is picked $\binom{N-1}{i-1}$ times (i.e., each qualified receiver appears in $\binom{N-1}{i-1}$ different $X^{[u],\mathcal{I}}$ terms), so the size of $W^{[u]}$ is set as $\binom{N-1}{i-1}\binom{K-N-1}{u-i}L^{[u]}$.

Repeat the same coding procedure for all sets $\mathcal{I}$ such that $\mathcal{I} \subset [1:N]$ and $|\mathcal{I}| = i$. Consider the row stack of all the $\mathbf{V}_{\mathcal{I}}^{w}$ matrices appeared (denoted as $\mathbf{V}^{[u],w}$) and the column stack of all the $\mathbf{V}_{\mathcal{U}}^{s}$ matrices appeared (denoted as $\mathbf{V}^{[u],s}$). As a result, $\mathbf{V}^{[u],w}$ has dimension $\binom{N}{i}\binom{K-N-1}{u-i}L^{[u]} \times \binom{N-1}{i-1}\binom{K-N-1}{u-i}L^{[u]}$ and $\mathbf{V}^{[u],s}$ has dimension $\binom{K-N-1}{u-i}L^{[u]} \times \binom{N}{i}\sum_{\mathcal{U}:[1:N]\cap\mathcal{U}=\mathcal{I},|\mathcal{U}|=u} L^{[u]}$. The field size p is chosen to be no smaller than the sum of the number of rows and the number of columns of $\mathbf{V}^{[u],w}$ and $\mathbf{V}^{[u],s}$ for all u, i.e., $p = \max_{u\in[1:K]} \max\Big((\binom{N}{i} + \binom{N-1}{i-1})\binom{K-N-1}{u-i}L^{[u]}, \binom{K-N-1}{u-i}L^{[u]} + \binom{N}{i}\sum_{\mathcal{U}:[1:N]\cap\mathcal{U}=\mathcal{I},|\mathcal{U}|=u} L^{[u]}\Big)$. Set $\mathbf{V}^{[u],w}$ and $\mathbf{V}^{[u],s}$ as two Cauchy matrices over $\mathbb{F}_p$ generated by distinct field elements (same as (85), (93)).

Security and correctness on the message symbols $W^{[u]}$ hold by the generic property of Cauchy matrices, i.e., any square sub-matrix of a full rank Cauchy matrix also has full rank. Specifically, any qualified receiver can obtain $H(W^{[u]})$ independent linear combinations of $W^{[u]}$ symbols (from $H(W^{[u]})$ rows of the Cauchy matrix $\mathbf{V}^{[u],w}$ so the rows are linearly independent). Hence all $W^{[u]}$ symbols can be decoded with no error. For any eavesdropping receiver, consider $X^{[u],\mathcal{I}}$. The number of unknown key symbols in $X^{[u],\mathcal{I}}$ is equal to the number of rows of $X^{[u],\mathcal{I}}$ and the linear combinations of the unknown key symbols have full rank (as the columns of $\mathbf{V}_{\mathcal{U}}^{s}$ are from the Cauchy matrix $\mathbf{V}^{[u],s}$). Further, the key symbols are independent for different $\mathcal{U}$, so the messages symbols are perfectly hidden from any eavesdropping receiver.

Counting all sets of i qualified receivers and all u-keys, where $i \in [1:u], u \in [1:K]$, we calculate the overall performance as follows,

$$\begin{aligned} R &= \sum_{u=1}^{K}\sum_{i=1}^{u} H(W^{[u]}) \\ &= \sum_{u=1}^{K}\sum_{i=1}^{u} \binom{N-1}{i-1}\binom{K-N-1}{u-i} L^{[u]} \log_2 p \qquad (156) \\ &= \sum_{u=1}^{K}\sum_{j=0}^{u-1} \binom{N-1}{j}\binom{K-N-1}{u-1-j} L^{[u]} \log_2 p \\ &= \sum_{u=1}^{K} \binom{K-2}{u-1} L^{[u]} \log_2 p \qquad (157) \end{aligned}$$

and

$$\begin{aligned} \beta(R) &= \sum_{u=1}^{K}\sum_{i=1}^{u} \binom{N}{i}\binom{K-N-1}{u-i} L^{[u]} \log_2 p \qquad (158) \\ &= \sum_{u=1}^{K} \left(\sum_{i=0}^{u} \binom{N}{i}\binom{K-N-1}{u-i} \right. \\ &\qquad \left. - \binom{N}{0}\binom{K-N-1}{u} \right) L^{[u]} \log_2 p \qquad (159) \\ &= \sum_{u=1}^{K} \left(\binom{K-1}{u} - \binom{K-N-1}{u} \right) L^{[u]} \log_2 p \end{aligned} \tag{160}$$

where both rate and broadcast bandwidth match the converse bounds. Note that we have used decompositions of schemes with independent u-keys (refer to Lemma 1) and for each u, we invoke the generic coding scheme (155). To ensure the overall scheme operates over the same field, we have used the maximum field size p required for all component schemes (i.e., p is picked as the maximum over all $u \in [1:K]$). The proof of Theorem 6 is thus complete.

G. Proof of Theorem 7: Rate Converse for the $N = 2, K = 5$ Instance

The rate converse is split into two lemmas. Before presenting the lemmas, we first summarize the entropy identities from the problem description. We have

$$\text{(Combinatorial Keys)} \quad H(a,b,c,d,e) = H(a)+H(b) + H(c)+H(d)+H(e) \tag{161}$$
$$\text{(Same Key Sizes)} \quad H(a)=H(b)=H(c) = H(d)=H(e)=L \tag{162}$$
$$\text{(Correctness)} \quad H(W|X,a,b,c) = H(W|X,b,d,e) = o(L) \tag{163}$$
$$\text{(Security)} \quad I(W;X,b) = I(W;X,c,d) = I(W;X,c,e) = o(L). \tag{164}$$

Lemma 3: For the secure groupcast instance in Fig. 4, we have

$$H(d,e|W,X,b) \leq 2L - H(W) + o(L). \tag{165}$$

Proof:

$$\begin{aligned}
& H(d,e|W,X,b) \\
= \;& H(d,e|X,b) - I(d,e;W|X,b) && (166)\\
\overset{(162)}{\leq} \;& 2L - H(W|X,b) + H(W|X,b,d,e) && (167)\\
\overset{(164)(163)}{=} \;& 2L - H(W) + o(L). && (168)
\end{aligned}$$

■

Lemma 4: For the secure groupcast instance in Fig. 4, we have

$$H(d,e|W,X,b) \geq 2H(W) - 3L + o(L). \tag{169}$$

Proof: Consider eavesdropping Receiver 4 such that X,c,d shall not reveal anything about W. We have

$$\begin{aligned}
& H(W,X,c,d) \\
\overset{(164)}{=} \;& H(W) + H(X,c,d) + o(L) && (170)\\
\geq \;& H(W) + H(X,b,c,d) - H(b) + o(L). && (171)
\end{aligned}$$

Symmetrically, consider eavesdropping Receiver 5 such that X,c,e shall not reveal anything about W. We have

$$\begin{aligned}
& H(W,X,c,e) \\
\overset{(164)}{=} \;& H(W) + H(X,c,e) + o(L) && (172)\\
\geq \;& H(W) + H(X,b,c,e) - H(b) + o(L). && (173)
\end{aligned}$$

Adding (171) and (173) and applying sub-modularity to $H(X,b,c,d) + H(X,b,c,e)$, we have

$$\begin{aligned}
& (171)+(173) \\
\Rightarrow \;& H(W,X,c,d) + H(W,X,c,e) \\
\geq \;& 2H(W) - 2H(b) + H(X,b,c,d,e) \\
& + H(X,b,c) + o(L) && (174)\\
\overset{(163)}{\geq} \;& 2\,H(W) - 2H(b) + H(W,X,b,c,d,e) \\
& + H(X,a,b,c) - H(a) + o(L) && (175)\\
\overset{(163)}{\geq} \;& 2H(W) - 2H(b) + H(W,X,c,d) \\
& + H(W,X,a,b,c) - H(a) + o(L). && (176)
\end{aligned}$$

Rearranging terms above and applying (162), we have

$$\begin{aligned}
& 2H(W) - 3L + o(L) \\
\leq \;& H(W,X,c,e) - H(W,X,a,b,c) && (177)\\
\leq \;& H(W,X,a,b,c,d,e) - H(W,X,a,b,c) && (178)\\
= \;& H(d,e|W,X,a,b,c) && (179)\\
\leq \;& H(d,e|W,X,b). && (180)
\end{aligned}$$

■

Finally, combining Lemma 3 and Lemma 4, we have

$$\begin{aligned}
& 2H(W) - 3L + o(L) \leq 2L - H(W) + o(L) && (181)\\
\Rightarrow \;& R = \frac{H(W)}{L} \leq \frac{5}{3} + \frac{o(L)}{L} && (182)
\end{aligned}$$

and letting $L \to \infty$ produces the desired bound $R \leq 5/3$.

H. Proof of Theorem 8: Multiple Messages

Let us start with the converse proof, which is a generalization of that in Theorem 1 and Theorem 2. Consider (54) and (55) follows from symmetry. We have

$$\begin{aligned}
& (R_1 + R_{12})L \\
= \;& H(W_1) + H(W_{12}) && (183)\\
\overset{(50)}{=} \;& I(W_1,W_{12};X,S_1,S_{12}) + o(L) && (184)\\
\overset{(52)}{=} \;& I(W_1,W_{12};S_1,S_{12}|X) + o(L) && (185)\\
\leq \;& H(S_1,S_{12}) + o(L) = (H(s_1) + H(s_{12}))L + o(L). && (186)
\end{aligned}$$

Consider (56) and (57) follows from symmetry. We have

$$\begin{aligned}
R_1 L \;=\;& H(W_1) \overset{(50)}{=} I(W_1;X,S_1,S_{12}) + o(L) && (187)\\
\overset{(51)}{=}\;& I(W_1;S_1|X,S_{12}) + o(L) && (188)\\
\leq\;& H(S_1) + o(L) = H(s_1)L + o(L). && (189)
\end{aligned}$$

Consider (58). We have

$$\begin{aligned}
& \beta(R_1,R_2,R_{12})L \\
\geq \;& H(X) \geq H(X|S_1,S_2,S_{12}) && (190)\\
\geq \;& I(X;W_1,W_2,W_{12}|S_1,S_2,S_{12}) && (191)\\
\overset{(50)(51)}{=} \;& H(W_1,W_2,W_{12}|S_1,S_2,S_{12}) + o(L) && (192)\\
\overset{(49)}{=} \;& H(W_1) + H(W_2) + H(W_{12}) + o(L) && (193)\\
= \;& (R_1 + R_2 + R_{12})L + o(L), && (194)\\
& \beta(R_1,R_2,R_{12})L \\
\geq \;& I(X;W_1,W_2,W_{12},Z_1,Z_2) && (195)\\
= \;& I(X;W_1,W_{12},Z_1) + I(X;W_2,Z_2|W_1,W_{12},Z_1) && (196)\\
\geq \;& I(X;W_1,W_{12}|Z_1) + I(X;Z_2|W_1,W_2,W_{12},Z_1) && (197)\\
\overset{(49)}{=} \;& I(X,Z_1;W_1,W_{12}) + I(X,W_2,W_{12};Z_2|W_1,Z_1) && (198)\\
\overset{(50)}{=} \;& H(W_1,W_{12}) + I(X,W_2,W_{12},Z_1;Z_2|W_1) \\
& -I(Z_1;Z_2|W_1) \; + o(L) && (199)
\end{aligned}$$

$$\overset{(49)}{\geq} H(W_1, W_{12}) + I(W_2, W_{12}; Z_2|X, W_1) - I(Z_1; Z_2) + o(L) \tag{200}$$

$$\overset{(51)}{=} H(W_1, W_{12}) + H(W_2, W_{12}|X, W_1) - I(Z_1; Z_2) + o(L) \tag{201}$$

$$\overset{(52)}{=} H(W_1, W_{12}) + H(W_2, W_{12}|W_1) - I(Z_1; Z_2) + o(L) \tag{202}$$

$$\overset{(49)}{=} (R_1 + R_{12} + R_2 + R_{12} - H(s_{12}))L + o(L). \tag{203}$$

Next, we consider the achievability. Consider any rational rate tuple $(R_1, R_2, R_{12}) \in \mathcal{C}$, i.e., (R_1, R_2, R_{12}) satisfies (54) - (57). Without loss of generality, assume R_1, R_2, R_{12} are integers (for rationals, we may consider blocks over the least common multiple of the denominators so that the number of bits becomes integers). We operate over the binary field and consider $L = 1$ block. The transmit signal is designed as follows. Denote by $W^{[a_1:a_2]}$ the a_1-th to a_2-th bits in the vector W. We have two cases.

Case 1. $R_{12} \leq H(s_{12})$.

$$X = (W_1 + s_1^{[1:R_1]};\ W_2 + s_2^{[1:R_2]};\ W_{12} + s_{12}^{[1:R_{12}]}). \tag{204}$$

The broadcast bandwidth achieved is $\beta(R_1, R_2, R_{12}) = R_1 + R_2 + R_{12}$.

Case 2. $R_{12} > H(s_{12})$.

$$X = \begin{pmatrix} W_1 + s_1^{[1:R_1]};\ W_2 + s_2^{[1:R_2]};\ W_{12}^{[1:H(s_{12})]} + s_{12} \\ W_{12}^{[H(s_{12})+1:R_{12}]} + s_1^{[R_1+1:R_1+R_{12}-H(s_{12})]} \\ W_{12}^{[H(s_{12})+1:R_{12}]} + s_2^{[R_2+1:R_2+R_{12}-H(s_{12})]} \end{pmatrix}. \tag{205}$$

Note that as (R_1, R_2, R_{12}) satisfies $(54) - (57)$, the key bits in the above scheme exist. The broadcast bandwidth achieved is $\beta(R_1, R_2, R_{12}) = R_1 + R_2 + 2R_{12} - H(s_{12})$.

Thus any rational rate tuple in the capacity region is achievable and as rational tuples are dense over the reals, the proof of Theorem 8 is complete.

I. Proof of Theorem 9: Achievability Under Discrete Memoryless Keys

Before presenting the achievability proof for Theorem 9, we cite a lemma on privacy amplification,[7] which encapsulates most technicalities of the achievability proof.

Lemma 5 (Lemma 5.18 in [32]): Consider random variables Z_c, Z_e, X_e (with finite cardinality) such that Z_c, Z_e are L length extensions of z_c, z_e and L_{X_e} denotes the number of bits in X_e. Then there exists a random mapping (independent of Z_e, X_e) from Z_c to a uniform random variable Z with L_Z bits such that

$$L_Z = H(z_c|z_e)L - L_{X_e} + o(L), \tag{206}$$

$$I(Z; Z_e, X_e) = o(L). \tag{207}$$

[7]Lemma 5 on secret key extraction suffices for our purposes over long key block lengths. Stronger non-asymptotic results and more efficient constructions of the random mappings are available in the literature (see e.g., [30], [31]).

In Lemma 5, we may interpret Z as the secret key to be generated from a known variable Z_c such that Z is almost independent of an eavesdropped variable Z_e (that has certain joint distribution with Z_c) and a prior knowledge variable X_e (that is arbitrarily correlated with Z_c, Z_e). The secret key size turns out to be given by the conditional entropy value minus the leaked prior knowledge.

Consider secure unicast first, whose achievability proof follows immediately from Lemma 5. Set $Z_c = Z_1$, i.e., the key for qualified Receiver 1, Z_e as the key for eavesdropping Receiver $e \in [2 : K]$, and $X_e = ()$. From Lemma 5, we know that Receiver 1 can generate a key Z that is almost independent from any eavesdropping receiver. Note that the random mapping used in Lemma 5 does not depend on the eavesdropped variable Z_e so that the secret key Z generated is simultaneously independent of any eavesdropped variable as long as we pick the key length to be $L_Z = \min_{e \in [2:K]} H(z_1|z_e)L + o(L)$. We use the key to send the desired message through one-time pad, i.e., $X = W + Z$ where the length of W is the same as the length of Z. Correctness and security are easy to verify (as Z is almost independent of Z_e, see (207)). The rate and broadcast bandwidth achieved are $R = \beta(R) = \min_{e \in [2:K]} H(z_1|z_e)$ as $L \to \infty$.

Then consider secure multicast. In Lemma 5, we set $Z_c = (Z_1, Z_2, \cdots, Z_{K-1})$, $Z_e = Z_K$, and X_e as the random bin index of $(Z_1, Z_2, \cdots, Z_{K-1})$ of length $\max_{q \in [1:K-1]} H(z_1, \cdots, z_{K-1}|z_q, z_K)L + o(L)$. The key Z is generated from Z_c and has length as specified in Lemma 5. The transmit signal sent by the transmitter is $X = (Z_K, X_e, W + Z)$. From Z_K, X_e, every qualified Receiver $q, q \in [1 : K-1]$ can recover Z_c. This result (in fact, in a more general form) was first proved in Theorem 2 of [33], where through Slepian Wolf coding (random binning, see Section 10.3 of [34]), the overall information seen by Receiver q, i.e., Z_K, X_e, Z_q has entropy whose value is at least the joint entropy, $H(Z_1, \cdots, Z_K)$), then $Z_1, \cdots, Z_K$ can be successfully decoded with vanishing error. Next every qualified receiver can generate Z with the same random mapping used by the transmitter. After extracting the common key Z, W can be decoded with vanishing error by every qualified receiver. Security is guaranteed by Lemma 5 as Z is almost independent of the information available to the eavesdropping Receiver K, i.e., Z_K, X_e. The rate achieved is $R = H(z_1, \cdots, z_{K-1}|z_K) - \max_{q \in [1:K-1]} H(z_1, \cdots, z_{K-1}|z_q, z_K) = \min_{q \in [1:K-1]} H(z_q|z_K)$ as $L \to \infty$.

VI. Conclusion

We introduce the problem of secure groupcast, where a transmitter wishes to securely communicate with a group of selected receivers while ensuring the other illegitimate receivers are fully ignorant of the desired communication, through noiseless broadcasting and correlated keys.

The communication efficiency of secure groupcast is measured by the message rate (number of message bits securely groupcast) and the broadcast bandwidth resource used (number of bits in the transmit signal). The main emphasis is placed on the most elementary setting of combinatorial keys and one common message, and limited extensions are also explored. Complete answers are obtained for certain preliminary cases, e.g., one legitimate receiver or one eavesdropping receiver, symmetric cases, while other cases remain unsolved. Interesting insights emerge out of this study, e.g., the necessity of decomposition and both structured and generic coding, the quest for tighter general converse bounds, and the potential of alignment view of the correlated key, message and transmit signal spaces. We find secure groupcast to be an interesting and challenging information theoretic security primitive with many open questions, and this work is a first step towards understanding coding opportunities for group communications under multiple correlated eavesdroppers.

References

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

[2] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 733–742, May 1993.

[3] R. Ahlswede and I. Csiszar, "Common randomness in information theory and cryptography. I. Secret sharing," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1121–1132, Jul. 1993.

[4] I. Csiszár and P. Narayan, "Secrecy capacities for multiple terminals," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3047–3061, Dec. 2004.

[5] A. Aminzadeh Gohari and V. Anantharam, "Information-theoretic key agreement of multiple terminals—Part I," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3973–3996, Aug. 2010.

[6] C. Chan and L. Zheng, "Mutual dependence for secret key agreement," in *Proc. 44th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2010, pp. 1–6.

[7] A. Fiat and M. Naor, "Broadcast encryption," in *Proc. Annu. Int. Cryptol. Conf.* Berlin, Germany: Springer, 1993, pp. 480–491.

[8] C. Blundo and A. Cresti, "Space requirements for broadcast encryption," in *Proc. Workshop Theory Appl. Cryptogr. Techn.* Berlin, Germany: Springer, 1994, pp. 287–298.

[9] D. R. Stinson, "On some methods for unconditionally secure key distribution and broadcast encryption," in *Selected Areas in Cryptography*. Boston, MA, USA: Springer, 1997, pp. 3–31.

[10] C. Blundo, L. A. F. Mattos, and D. R. Stinson, "Trade-offs between communication and storage in unconditionally secure schemes for broadcast encryption and interactive key distribution," in *Proc. Annu. Int. Cryptol. Conf.* Berlin, Germany: Springer, 1996, pp. 387–400.

[11] H. Sun, "Compound secure groupcast: Key assignment for selected broadcasting," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021, pp. 694–699.

[12] L. Grokop and D. Tse, "Fundamental constraints on multicast capacity regions," in *Proc. 45th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, 2007, pp. 1–25.

[13] C. Tian, "Latent capacity region: A case study on symmetric broadcast with common messages," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3273–3285, Jun. 2011.

[14] A. Salimi, T. Liu, and S. Cui, "Polyhedral description of the symmetrical latency capacity region of broadcast channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2014, pp. 2122–2126.

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

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

[17] R. Liu, I. Maric, P. Spasojević, and R. D. Yates, "Discrete memoryless interference and broadcast channels with confidential messages: Secrecy rate regions," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2493–2507, Jun. 2008.

[18] A. Khisti, A. Tchamkerten, and G. W. Wornell, "Secure broadcasting over fading channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2453–2469, Jun. 2008.

[19] E. Ekrem and S. Ulukus, "Secure broadcasting using multiple antennas," *J. Commun. Netw.*, vol. 12, no. 5, pp. 411–432, Oct. 2010.

[20] R. F. Schaefer, A. Khisti, and H. V. Poor, "Secure broadcasting using independent secret keys," *IEEE Trans. Commun.*, vol. 66, no. 2, pp. 644–661, Feb. 2018.

[21] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol, "Index coding with side information," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1479–1494, Mar. 2011.

[22] S. H. Dau, V. Skachek, and Y. M. Chee, "On the security of index coding with side information," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3975–3988, Jun. 2012.

[23] M. M. Mojahedian, M. R. Aref, and A. Gohari, "Perfectly secure index coding," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7382–7395, Nov. 2017.

[24] V. Narayanan, J. Ravi, V. K. Mishra, B. K. Dey, N. Karamchandani, and V. M. Prabhakaran, "Private index coding," *IEEE Trans. Inf. Theory*, vol. 68, no. 3, pp. 2020–2049, Mar. 2022.

[25] Z. Li and H. Sun, "Conditional disclosure of secrets: A noise and signal alignment approach," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2020, pp. 944–948.

[26] Y. Zhou, H. Sun, and S. Fu, "On the randomness cost of linear secure computation," in *Proc. 53rd Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2019, pp. 1–6.

[27] H. Sun and S. A. Jafar, "The capacity of private information retrieval," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4075–4088, Jul. 2017.

[28] M. Sudan, H. Tyagi, and S. Watanabe, "Communication for generating correlation: A unifying survey," *IEEE Trans. Inf. Theory*, vol. 66, no. 1, pp. 5–37, Jan. 2020.

[29] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.

[30] C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer, "Generalized privacy amplification," *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pp. 1915–1923, Nov. 1995.

[31] M. Hayashi, "Security analysis of ε-almost dual universal$_2$ hash functions: Smoothing of min entropy versus smoothing of Rényi entropy of order 2," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3451–3476, Jun. 2016.

[32] P. Narayan *et al.*, "Multiterminal secrecy by public discussion," *Found. Trends Commun. Inf. Theory*, vol. 13, nos. 2–3, pp. 129–275, 2016.

[33] A. Sgarro, "Source coding with side information at several decoders," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 2, pp. 179–182, Mar. 1977.

[34] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

Hua Sun (Member, IEEE) received the B.E. degree in communications engineering from the Beijing University of Posts and Telecommunications, China, in 2011, and the M.S. degree in electrical and computer engineering and the Ph.D. degree in electrical engineering from the University of California Irvine, USA, in 2013 and 2017, respectively.

He is currently an Assistant Professor with the Department of Electrical Engineering, University of North Texas, USA. His research interests include information theory and its applications to communications, privacy, security, and storage. He was a recipient of the NSF CAREER Award in 2021 and the UNT College of Engineering Distinguished Faculty Fellowship in 2021. His coauthored papers received the IEEE Jack Keil Wolf ISIT Student Paper Award in 2016 and an IEEE GLOBECOM Best Paper Award in 2016.