Secure Multi-Source Multicast

Abstract—The principal mission of Multi-Source Multicast (MSM) is to disseminate all messages from all sources in a network to all destinations. MSM is utilized in numerous applications. In many of them, securing the messages disseminated is critical.

A common secure model is to consider a network where there is an eavesdropper which is able to observe a subset of the network links, and seek a code which keeps the eavesdropper ignorant regarding all the messages. While this is solved when all messages are located at a single source, Secure MSM (SMSM) is an open problem, and the rates required are hard to characterize in general.

In this paper, we consider Individual Security, which promises that the eavesdropper has zero mutual information with each message individually. We completely characterize the rate region for SMSM under individual security, and show that such a security level is achievable at the full capacity of the network, that is, the cut-set bound is the matching converse, similar to non-secure MSM. Moreover, we show that the field size is similar to non-secure MSM and does not have to be larger due to the security constraint.

I. INTRODUCTION

Linear Network Coding (LNC) [1] and Random Linear Network Coding (RLNC) [2] are essential for efficient utilization of network resources. With network coding, multiple sources can multicast information to all destinations simultaneously, at rates up to the min-cut between the sources and the destinations. Figure 1 depicts a simple example: the min-cut from any source to any destination is 2, and from both sources to any destination is 4, hence one can disseminate 2 messages from each source to all destinations. However, in many practical multicast applications, it is important to ensure privacy is not compromised if an eavesdropper (Eve) is present in the network. Indeed, the theory of secure network coding is vast. We include here only the most relevant works.

When the sources are co-located at a single node, several secure network coding solutions were suggested [3]-[8]. Such solutions guarantee the mutual information between Eve’s data, \( Z \), and all the messages is 0. For example, returning to Figure 1, if only source \( s_1 \) had messages to send, and Eve would be able to wiretap one link in the network, then secure network coding would guarantee secure dissemination of one message from the source to all destinations. This is a reduction in rate compared to the full capacity, as the min-cut from \( s_1 \) to any destination is 2. However, when requiring zero mutual information with all messages from the source, this rate reduction is essential, and matches the converse result.

Figure 1: Secure multi-source multicast with LNC, for two sources \( s_1 \), with two messages each and four legitimate destination nodes \( d_i \). The eavesdropper min-cut is at most 1. The edges in the graph point downward.

When the network includes multiple sources which are not co-located, the problem is more involved. Clearly, applying a single-source, secure network coding solution at each source would give an achievable scheme. In the example, if Eve wiretaps one link, one can clearly multicast one message from each source, to all destinations. This solution, however, may be wasteful, as it is half of the full capacity of the network, “wasting” one message per source, although Eve may capture only a single link regardless of the number of sources. Indeed, there is no matching converse result for the above solution.

In [9], [10], the authors gave a necessary and sufficient condition for Secure Multi-Source Multicast (SMSM). However, it is a condition on ranks of matrices having the global encoding vectors as columns, and, unlike non-secure MSM or secure single-source multicast, it does not translate directly to rate or min-cut constraints. Thus, the problem of determining the rate region in SMSM is an open problem in general [11], and as mentioned in [12, Section VI], seeking models for which it is solvable is important. In [13], the authors characterized the network coding capacity of several models, including SMSM, via the entropic region \( \Gamma^* \). Yet, to date, this region is not fully characterized.

Main Contribution

In this paper, we consider SMSM under an Individual Security constraint. In this model, the eavesdropper is kept ignorant, in the sense of having zero mutual information, regarding each message separately, yet may potentially obtain insignificant information about mixtures of packets transmitted. Such a security model was recently used in various
canonical problems, e.g., wiretap channels \cite{14}, more general broadcast channels \cite{15,18} and multiple-access channels \cite{19,20}, and, although not specifically mentioned as such, is also related to weakly secure network coding \cite{21} and the notion of algebraic security \cite{22,23}, which consider the information in linear combinations of messages.

We completely characterize the rate region for individually secure MSM. Specifically, we show that secure communication is achievable up to the min-cut, that is, without any decrease in the rate or any message “blow-up” by extra randomness. In fact, due to the individual security constraint, messages protect one another, and in the context of Figure 1 one is able to send two messages from each source securely, although Eve may observe any single link.

We then turn to a few applications where the suggested coding scheme can be useful. Specifically, we consider data centers, wireless networks and live broadcasting of video using multipath streaming, and show how the individual security coding schemes suggested in this paper is applicable, achieving the full capacity of those systems. Finally, we show that the coding scheme is applicable to algebraic gossip as well \cite{24}, resulting in secure gossip without extra rounds. For example, consider the “Random Phone Call” model. This model was introduced in \cite{25} as special case of uniform gossip. In each round of communication, every participant may “call” a random participant, and send one unit of information. The goal is, of coarse, to disseminate messages from the source to all participants. Rigorously, the underlying graph is complete and unweighted. A detailed analysis of this model is given in \cite{26,27}. It was shown that in a random phone call model with \( v \) nodes, the flooding time is \( \Theta(\log v) \), with constant throughput. Of course this is without any secrecy constraint. Any phone call which Eve listens to contains relevant information, and results in leakage. Using the code suggested in this paper, we will show that one can design a secure gossip scheme, which make sure that as long as Eve dose not listen to too many calls, she remains completely ignorant regarding any specific message, and all this without any loss in throughput or number of rounds.

The structure of this paper is as follows. In Section II a SMSSM model is formally described. Section III includes our main results, with the individually-SMSSM direct proved in Section IV and converse proved in Section V. In Section VI, we show a few important examples, for which the individual security coding is applicable. Section VII includes a linear code construction for the individually-SMSSM model. Section VIII describes a Strongly-SMSSM algorithm and proves a direct result for it. Section IX concludes the paper.

II. MODEL AND PROBLEM FORMULATION

SMSSM is specified by a graph \( G = (\mathcal{V}, \mathcal{E}) \), where \( \mathcal{V} \) and \( \mathcal{E} \) are the node set and the edge set, respectively. We assume noise-free links of unit capacity. This capacity can be thought of as one “packet” of \( c \) bits, plus some negligible overhead.\footnote{As in most LNC solutions, a header is required for each message. Thus, we assume messages of length \( c \), large enough to make the overhead in the header negligible.}

The node set \( \mathcal{V} \) contains a subset of source nodes \( S = \{S_1, \ldots, S_{|S|}\} \) and a subset of legitimate destination nodes \( D = \{D_1, \ldots, D_{|D|}\} \). Each of the sources has its own set of \( k \) independent messages of length \( \epsilon \) each, over the binary field. We denote them by a messages matrix

\[
\mathbf{M}_s = [\tilde{M}_{s,1}; \tilde{M}_{s,2}; \ldots; \tilde{M}_{s,k}] \in \{0, 1\}^{k \times c},
\]

where each row corresponds to a separate message \( \tilde{M}_{s,j}, j \in \{1, \ldots, k\} \).

We assume an eavesdropper which can obtain a subset of \( w \) packets traversing the network. Specifically, we define the eavesdropper matrix as

\[
\mathbf{Z}_w = [Z^c_1; Z^c_2; \ldots; Z^c_w] \in \{0, 1\}^{w \times c}.
\]

We denote the values of min-cuts in the network by \( \rho(\cdot; \cdot) \). For example, for \( s_1 \in S \) and \( d_1 \in D \), \( \rho(s_1; d_1) \) represents the value of the min-cut from source node \( s_1 \) to legitimate node \( d_1 \). \( \rho(s_1; z) \) represents the value of the min-cut from source node \( s_1 \) to the eavesdropper (assuming \( z \) is a virtual node with infinite capacity from the \( w \) edges observed by Eve) and \( \rho(S; d_1) \) represents the value of the min-cut from all the source nodes to legitimate node \( d_1 \).

The goal is to design secure multi-source multicast coding scheme where legitimate nodes send their available messages in order to disseminate all the messages to all the legitimate destination nodes, yet, observing \( w \) packets from the communication between legitimate nodes, the eavesdropper is ignorant regarding the messages.

Definition 1. An MSM algorithm with parameters \( k \) and \( w \) is Reliable and Individually or Strongly secure if:

1. Reliable: At the legitimate destination node \( d \in D \), letting \( Y_d \) denote the message matrix obtained, for any set of messages \( \mathbf{M}_s, s \in S \), we have

\[
P(\mathbf{M}_s(Y_d) \neq \mathbf{M}_s) \leq \epsilon,
\]

where \( \mathbf{M}_s(Y_d) \) is the estimation of messages \( \mathbf{M}_s \) at \( d \).

2. Individually secure: At the eavesdropper, observing \( w \) packets, we have

\[
H(\mathbf{M}_{s,j}|\mathbf{Z}_w) = H(\mathbf{M}_{s,j}),
\]

for all \( j \in \{1, \ldots, k\} \) and for all \( s \in S \).

3. Strongly secure: At the eavesdropper, observing \( w \) packets, for all \( s \in S \) we have

\[
H(\mathbf{M}_s|\mathbf{Z}_w) = H(\mathbf{M}_s).
\]

Remark 1. The individual-secrecy constraint given in Definition \cite{12} does not promise perfect, strong-secrecy \cite{5,8,9}, which is, having the mutual information with all messages negligible. Individual-secrecy ensures secrecy only on each message \( M_{s,j} \) separately. The eavesdropper, observing \( \mathbf{Z}_w \), may obtain some information on the combination of \( k \) messages since the messages are not independent given \( \mathbf{Z}_w \). However, since the \( k \) original messages are mutually independent, the leaked information has no meaning \cite{14,19,28}. In other words, since
\[
I(M_s; Z_w) = \sum_k I(M_{s,k}; Z_w | M_s \gamma - 1) \geq \sum_k I(M_{s,k}; Z_w),
\]

we require that the r.h.s will be small, however, this does not guarantee that the l.h.s is small. If the eavesdropper receives message \( M_{s,j} \) by any other manner than the Individual-SMSM transmissions, Eve may obtain some information on other messages \( M_{s,i}, i \neq j \), from \( M_{s,j} \) and \( Z_w \). If it is required to prevent the possibility of such an attack, one can get perfect secrecy using Definition [13], yet at the price of a lower rate, as given in Section [VIII].

**Remark 2.** For multicast problems and LNC, the condition in (1) can be used with \( \epsilon = 0 \) [1, 2]. Yet, we allow a small error to cope with protocols such as randomized gossip [24], [29], which we discuss later in this paper.

**Remark 3.** The first code construction we consider, given in Section [IV] is based on random coding. Therefore, in that case, the individual secrecy constraint will hold only asymptotically, that is, \( H(M_{s,j}; Z_w)/H(M_{s,j}) \to 1 \) as \( k \) grows. Then, in Section [VII] we suggest a structured linear code, which results in zero mutual information, such that there is no requirement for \( k \) to grow.

### A. Source and Network Coding

We assume a source \( s \in \mathcal{S} \) may use an encoder,

\[
f : M_s \to \mathcal{X}_s \in \{0, 1\}^{n \times c},
\]

which maps each message matrix \( M_s \) to a matrix \( \mathcal{X}_s \) of codewords. When using a strong security constraints, e.g., [5], [8], \( n > k \) and this represents a message “blow-up” using a random key, used to confuse Eve. However, the main contribution herein, is that **under individual-secrecy, \( n = k \) suffices, and there will be no rate loss due to the secrecy constraint**.

Then, the source packets \( \hat{Y} \) transmitted are linear combinations of \( \{X_r\}_{r=1}^{n} \) with coefficients in the usual LNC sense, i.e.,

\[
\hat{Y} = \sum_{r=1}^{n} \mu_r X_r.
\]

Each node maintains a subspace \( Y_v \) that is the span of all packets known to it. In RLNC, when node \( v \) sends a packet, \( Out(\hat{Y}) \), it chooses uniformly a packet from \( Y_v \), by taking a random linear combination. If a deterministic algorithm is used, e.g., [30], the coefficients are calculated based on the network topology. The code we suggest herein is only at the sources, and then utilizes any capacity-achieving, non-secure network code.

### B. Gossip in Oblivious Networks

While the results in this paper are tailored to LNC in the sense of [1], [2], they easily apply to **algebraic gossip** [24] as well. Such algebraic gossip protocol have been considered in the literature for many tasks, such as ensuring database consistency, computing aggregate information and other functions of the data [25], [31–33]. We briefly describe this model. The network operates in rounds. In each round \( t \), the sources, as well as any legitimate node which has messages it previously received, pick a random node to exchange information with. The information exchange is done by either sending (PUSH) or receiving (PULL) a message. In algebraic gossip, the message sent by a node \( v \) is simply a random linear combination of the vectors which form a basis for \( Y_v \). The process stops when all the legitimate nodes have all the messages, i.e.,, have a full rank matrix. We briefly review the definitions and results from [29] for non-secure gossip networks, which we will use to formulate our result in this context.

**Definition 2.** A network is **oblivious** if the topology of the network, \( G_t \) at time \( t \), only depends on \( t \), \( G_{t'} \) for any \( t' < t \) and some randomness. We call an oblivious network model furthermore i.i.d., if the topology \( G_t \) is independent of \( t \) and prior topologies.

The importance of Definition [2] lies in the fact that the topology of an oblivious network may change in time, but only based on the past topology and some external randomness. Topology dose not change based on the data traversing the network. Consider a single (uncoded) message, and the set of nodes \( S_t \) which received that message after \( l \) rounds. \( S_t \) advances like a flooding process \( F \). That is, \( S_t \subseteq S_{t'} \subseteq V \) for \( l \leq l' \), with an absorbing sate \( V \). We say that \( F \) stops at time \( t \) if the message is received at all nodes after \( t \) rounds. Let \( S_F \) be the random variable denoting the stopping time of \( F \).

**Definition 3.** We say an oblivious network with a vertex set \( V \) floods in time \( T \) with throughput \( \alpha \) if there exists a prime power \( q \) such that for every vertex \( v \in V \) and every \( k > 0 \) we have \( P[S_F \geq T + k] < q^{-\alpha k} \).

### III. Main Results

The three main results in this paper completely characterize the rate region for individually secure multi-source multicast. Specifically, we give tight achievability and converse, and a tight characterization of the number of rounds required under a gossip model. Thus, the first main result is the following achievability theorem, which states that individually-secure multi-source multicast is achievable at rates up to the network min-cuts, using LNC.

**A. Individually Secure MSM**

**Theorem 1.** Assume an SMSM network \((V, E, S, D, w)\). There exists a coding scheme which disseminates \( k \) messages from each source in \( S \) to all destinations in \( D \), while keeping an eavesdropper which observes \( w \) links ignorant with respect to each message individually if:

1. For all \( s \in S \) and all \( d \in D \), \( \rho(s, d) \geq k \).
2. For all \( d \in D \), \( \rho(S, d) \geq k|S| \).

Under strong-secrecy, i.e., requiring Eve’s mutual information with all messages simultaneously to be zero, the problem of MSM is still open [11], [12, Section VI]. Clearly, if Eve observes \( w \) links, a naive implementation, which increases the message rates from each source by \( w \), can send \( k \) messages from each source and achieve strong secrecy if:

1. For all \( s \in S \) and all \( d \in D \), \( \rho(s, d) \geq k + w \).
2) For all $d \in \mathcal{D}$, $\rho(s, d) \geq (k + w)|S|$. However, such an implementation is clearly wasteful, and, to date, the optimal strategy is unknown. Obviously, the required rates under strong secrecy are higher than the min-cut bound, as even for single-source multicast one needs $\rho(s_1, d_1) \geq k + w$ \cite{5}. The importance of Theorem 1 is that under individual secrecy, not only the cut set region can be characterized, and is achievable using linear network coding, individually secure MSM is possible up to the min-cuts in the network.

In Section VIII we provide a code for Strong-SMSM. It is important to note that in the code suggested, the alphabet size does not increase with the network parameters due to the strong-security constraint.

The tightness of Theorem 1 in terms of rates, is trivially achieved using the cut-set bound. That is, the conditions $\rho(s, d) \geq k$ and $\rho(S, d) \geq k|S|$ are required solely to achieve reliability, nevertheless when security is an additional constraint. However, a stronger notion of a converse can be given. To this end, we first note that while Theorem 1 guarantees individual secrecy according Definition 1 it still gives a slightly weaker level of security than is possible at the same rates. Specifically, Theorem 1 guarantees $I(M_s; Z) = 0$ for any single message $j$. However, ensuring the mapping from $M_s$ to $X_s$ mixes the messages appropriately, i.e., satisfies rank constraints similar to \cite[Lemma 3.1]{12}, can, in fact, ensure Eve is kept ignorant of any set of $k - w$ messages. That is, guarantee $k_w$ individual secrecy with respect to any set of $k_w \leq k - w$ messages. Let $M_s^{k-w}$ denote a set of $k - w$ messages from $s$, and $M_s^w$ denote the remaining $w$. Thus, we also have the following corollary.

**Corollary 1.** Assume an SMSM network $(\mathcal{V}, \mathcal{E}, S, \mathcal{D}, w)$. There exists a coding scheme which disseminates $k$ messages from each source in $S$, to all destinations in $\mathcal{D}$, while keeping an eavesdropper which observes $w$ links ignorant with respect to any set of $k_w \leq k - w$ messages individually, such that $I(M_s^{k-w}, Z_w) = 0$, if:

1) For all $s \in S$ and all $d \in \mathcal{D}$, $\rho(s, d) \geq k$.
2) For all $d \in \mathcal{D}$, $\rho(S, d) \geq k$.

In Section IV-B we prove the $k_w$-individual secrecy constraint is indeed met.

Under such an individual secrecy constraint, the converse below gives a stronger result than the min-cut bound.

**Theorem 2.** Assume an SMSM network $(\mathcal{V}, \mathcal{E}, S, \mathcal{D}, w)$. Under individual security for $k - w$ messages, that is, requiring $I(M_s^{k-w}, Z_w) = 0$ for any set of $k - w$ messages, one must have

$$H(M_s) \leq \rho(s, d_s) - \rho(s, z) + w.$$ 

This result should be interpreted as follows. If Eve observes $w$ independent links, and $\rho(s; z) = w$, then one must have $H(M_s) \leq \rho(s, d_s)$, which is the cut set bound. Of coarse, as mentioned before, the surprising part is that this bound is tight, hence such a level of security is available without any loss in rate. However, if Eve observes more than $w$ links, and one wishes to maintain the individual-secrecy constraint, then $H(M_s)$ should be strictly smaller than $\rho(s, d_s)$ and at the same amount. E.g., if Eve observes $w + e$ links, we have $H(M_s) \leq \rho(s, d_i) - e$. This means a linear increase in Eve’s power results in a linear decrease in rate.

Finally, it is important to note that the constraint on how many messages Eve catches is set on the entire network, thus, Eve may catch $w$ messages of a single source, or $w$ messages from several sources all together. Secrecy is maintained in any case, as under individual secrecy, messages from other sources can only increase secrecy, and any network code cannot create linear combinations with other messages which reduce the secrecy level. This is another benefit of the model, and hence the network code can be any LNC, without an increase in alphabet size.

### B. Algebraic Gossip

As mentioned earlier, the suggested code easily applies to algebraic gossip as well, since this can be viewed as linear network coding over a time-extended graph. The following results capture the number of rounds required to (individually) securely disseminate $k$ messages from each of the $|S|$ sources to all nodes in the network.

**Theorem 3.** Assume an oblivious network that floods in time $T$ with throughput $\alpha$. Then, for $|S|$ nodes in the network with $k$ messages each, algebraic gossip spreads the $k|S|$ messages to all nodes with probability $1 - \epsilon$ after

$$T' = T + \frac{1}{\alpha}(k|S| + \log \epsilon^{-1})$$

rounds, while keeping any eavesdropper which observes at most $w \leq k - k_w$ packets, ignorant with respect to any set of $k_w$ messages individually.

Thus, compared to only a reliability constraint, the number of rounds required for both reliability and individual-secrecy is exactly the same as in the original non-secure gossip protocol. Note that the result above is constant-optimal, as $T'$ is the number of rounds required for a single message, hence one cannot expect less that $T'$ above for $k|S|$ messages. This is a perfect pipelining property \cite{29}, thus, surprisingly, one can gossip securely messages to all parties in the network, without any loss in rate and without any centralized mechanism for routing, key exchange or any other encryption mechanism, as long as the eavesdropper is interested in single messages.

### C. Alphabet Size

Without secrecy constraints, Jaggi et al. proved that a field with size greater than or equal to the number of destinations is sufficient for multicast under LNC \cite{30}. However, this may not hold if it is required to keep an eavesdropper ignorant. Cai et al. \cite{5} devised a code which requires a field of exponential size to obtain secrecy. There, the field size must be larger than $\left(\frac{|E|}{w}\right)^4$. Feldman et al. \cite{34} showed that there exist networks that require a field of size at least $\Theta(|E|^{2k})$. In \cite{6}, the authors demonstrate that secure network coding can be considered as a network generalization of the wiretap channel of type II. When $d$ is the number of destinations in the multicast connection, a field of size $\left(\frac{w^2}{w-1+d}\right)^{2k}$ is sufficient, which is independent of $|V|$ and $|E|$ but is still exponential in other network parameters.
identically distributed codewords, one codeword, \( x^k(e(i)) \), from the bin indexed by \( M_{s,1}(i); \ldots; M_{s,k'}(i) \), where \( e(i) = M_{s,k'+1}(i); \ldots; M_{s,k}(i) \). That is, \( k' = k - w \) bits of the column choose the bin, and the remaining \( w \) bits choose the codeword within the bin.

Then, similar to many RLNC protocols, the sources transmit linear combinations of the rows, with random coefficients. Nodes transmit random linear combinations of the vectors in \( \mathcal{S}_v \), which is maintained by each node according to the messages received at the node.

### A. Reliability

The reliability proof using RLNC is almost a direct consequence of \([2]\). Clearly, the min-cut is given by Theorem \([1]\). Hence, the legitimate nodes can easily reconstruct \( X_s \) for each \( s \) (simple, non-secure, multi-source multicast). Then, each destination maps \( X_s \) back to \( M_s \), as this is a 1 : 1 mapping.

In the same way, using a gossip protocol, the reliability proof is almost a direct consequence of \([29]\ Theorem 1]. Hence, the number of rounds required is given by Theorem \([3]\).

An example, obtaining both reliability and individual secrecy for two sources, with two messages each and four legitimate destination nodes, where the eavesdropper min-cut is at most 1, is given in Figure \([1]\). Note that secure communication with respect to one message is possible while sending two messages from each source to all destinations.

### B. Information Leakage at the Eavesdropper

We now prove the \( k_s \)-individual security constraint is met, that is, \( I(M_{s,k';w}; Z_w) \rightarrow 0 \) for any set of \( k_s \leq k - w \) messages. In particular, for the individual constraint, we wish to show that \( I(M_{s,j}; Z_w) \) is small for all \( s \in S \) and all \( j \). We will do that by showing that given \( Z_w \), Eve’s information, all possibilities for any set of \( k_s \leq k - w \) messages \( M_{s,k';w} \) are equally likely. Hence Eve has no intelligent estimation for \( M_{s,k';w} \) and \( M_{s,j} \).

Denote by \( C_k \) the random codebook and by \( X_s \) the set of codewords corresponding to \( M_{s,1}, \ldots, M_{s,k} \). To analyze the information leakage at the eavesdropper, note that Eve has access to at most \( w \) linear combinations on the rows of \( X_s \). We will assume these linear combinations are in fact independent, and since Eve has access to the coefficients, we will assume Eve can even use Gaussian elimination and have access to \( w \) rows from original matrix \( X_s \).

Next, note that the columns of \( X_s \) are independent (by the construction of the codebook, creating \( X_s \) is done independently per-column; \( c \) columns are used only to reduce the NC overhead). Hence, it suffices to consider the information leakage for each column \( i \in \{1, \ldots, c\} \) from \( X_s \) separately.

For each column \( i \) of \( M_s \), the encoder has \( \Delta \) independent and identically distributed codewords, out of which one is selected. Hence, there is an exponential number of codewords,
from the eavesdropper’s perspective, that can generate a column in \( X_s \), and we require that Eve is still confused even given the \( w \) bit from each column. Let \( Z_w(i) \) be the \( w \) bits Eve has from column \( i \). Denote \( l = k - w \). Similar to the technique used in \[35\] to prove that myopic adversaries are blind, we define \( Sh(Z_w(i), l) \) the set of all \( k \)-tuples consistent with \( Z_w(i) \), i.e.,

\[ Sh(Z_w(i), l) = \{ b^k : b^k(S_Z) = Z_w(i) \}, \]

where \( S_Z \) denotes indices of the rows Eve has. Clearly, there are \( 2^l \) tuples in \( Sh(Z_w(i), l) \). See Figure 3 for a graphical illustration. We assume Eve has the codebook, yet does not know which column from each bin is selected to be the codeword. Hence, we wish to show that given \( Z_w(i) \), Eve will have at least one candidate per bin. The probability for a codeword to fall in a given shell is

\[ Pr(X^k_s(i) \in C_k \cap X^k_s(i) \in Sh(Z_w(i), l)) = \frac{Vol(Sh(Z_w(i), l))}{2^k} = \frac{2^{k-w}}{2^k}. \]

In each bin of \( C_k \), we have \( \Delta = 2^{w+k\varepsilon} \) codewords. Thus, the number of codewords Eve sees on a shell, per bin is

\[ \#(m(i) : X^k(i) \in Sh(Z(i), l)) = \frac{2^{w+k\varepsilon} + 2^{k-w}}{2^k} = 2^{k\varepsilon}. \]

Hence, we can conclude that on average, and if \( k\varepsilon \) is not too small, for every column in \( M_s \) Eve has a few possibilities in each bin, hence cannot locate the right bin. However, it is still important to show that all bins have (asymptotically) equally likely number of candidate codewords, hence Eve cannot locate a preferred bin.

To this end, we proved that the average number of codewords per column is \( 2^{k\varepsilon} \). We wish to show that now the probability that the actual number of options deviates from the average by more than \( \varepsilon \) is small. Define

\[ \mathcal{E}_{C_i}(Z(i), l) := Pr\{(1-\varepsilon)2^{k\varepsilon} \leq \#(m(i) : X^k(i) \in Sh(Z(i), l)) \leq (1+\varepsilon)2^{k\varepsilon}\}. \]

By the Chernoff bound, we have

\[ Pr(\mathcal{E}_{C_i}(Z(i), l)) \geq 1 - 2^{-\varepsilon^22^{k\varepsilon}}. \]

Due to the super exponential decay in \( k \), when taking a union bound over all columns, the probability that Eve decodes correctly some column is small. Hence, by the chain rule for entropies, since all the codewords in the codebook are independent \( H(X_s) = \sum_c H(X_s(i)) \), for Eve, all codewords are almost equiprobable and \( I(M_{s-w}^k; Z_w) \rightarrow 0 \). In particular, \( I(M_{s-w}^k; Z_w) \rightarrow 0 \).

V. CONVERSE (THEOREM 2)

In this section, we derive a converse result, which shows that under individual secrecy on a group of \( k - w \) messages, not only the rate is bounded by the min-cut, but, more importantly, any independent link that Eve observes above \( w \) will require to reduce the rate at the same amount in order to achieve both reliability and secrecy.

Let \( Z \) denote the random variable corresponding to the links which are not available to Eve. Hence, \( Y_d = (Z, Z) \). Let \( M_{s-w}^k \) denote a set of \( k - w \) messages, and \( M_{s-w}^k \) denote the remaining \( w \). We will show that reliability, that is \( H(M_s|Y_d) = 0 \), and individual secrecy, that is, \( I(M_{s-w}^k; Z) = 0 \), imply that \( H(M_s) \) is upper bounded by the term in Theorem 2.

\[ H(M_s) = H(M_{s-w}^k|M_w^k) + H(M_{s-w}^k) \]

\[ \leq I(M_{s-w}^k; Y_d|M_w^k) + H(M_{s-w}^k|Y_d) + w \]

\[ \leq I(M_{s-w}^k; Z|M_w^k) + I(M_{s-w}^k; Z|M_w^k) + w \]

\[ = I(M_{s-w}^k; Z) + I(M_{s-w}^k; Z|M_{s-w}^k) - I(Z; M_{s-w}^k) + I(M_{s-w}^k; Z) \]

\[ + I(M_{s-w}^k; Z|M_{s-w}^k) + w \]

\[ = I(M_{s-w}^k; Z) - I(Z; M_{s-w}^k) + I(M_{s-w}^k; Z|M_{s-w}^k) + w \]

\[ = I(M_{s-w}^k; Z|M_{s-w}^k) - I(Z; M_{s-w}^k) + I(M_{s-w}^k; Z|M_{s-w}^k) + w \]

\[ = I(M_{s-w}^k; Z|M_{s-w}^k) + H(Z; M_{s-w}^k) + H(Z) - w \]

\[ = I(M_{s-w}^k; Z|M_{s-w}^k) + H(Z; M_{s-w}^k) + H(Z) - H(Z) + w \]

\[ = I(M_{s-w}^k; Z|M_{s-w}^k) - I(Z; M_{s-w}^k) + H(Z) + w \]

\[ \leq \rho(s_i; d_i) - \rho(s_i; z) + w, \]

where (a) is since conditioning reduces entropy, (b) is due to the reliability constraint, (c) follows since we assume that Eve is kept ignorant regarding any group of \( w - k \) messages, hence \( I(M_{s-w}^k; Z) = 0 \), and (d) follows since \( \rho(s_i; d_i) - \rho(s_i; z) \) is the maximum amount that may not be available to Eve, if she has a min-cut \( \rho(s_i; z) \). Again, we assume unit capacity links and normalize the information in a message to "1" accordingly.
VI. Applications

In previous sections we suggested an SMSM code and proved that under the suggested code an eavesdropper which can capture a subset of the packet’s traversing the network (up to \( w \) packets) is kept ignorant regarding each packets content, under the Individual Security constraint, without compromising the rate (i.e., achieving full network capacity). In this section, we show several common applications which exemplify the applicability of the suggested code to a diverse range of protocols and applications. The first two examples include only a single source, merely to show the applicability of the individual secrecy setup. The third example is multi-source in nature, and includes all aspects of our solution.

A. Data Centers

One of the most prominent facilities characterizing our new information explosion era are distributed Data Centers. Such facilities, which aim to cope with the rapidly increasing volumes of data generated, archived and expected to be accessible, are vital to many services such as video sharing, social networks, peer-to-peer cloud storage and many more. Google’s GFS [36], Amazon’s Dynamo [37], Google’s BigTable [38], Facebook’s Apache Hadoop [39], Microsoft’s WAS [40] and LinkedIn’s Voldemort [41] are just a few examples of such ubiquitous applications. Obviously, the security and reliability of such Data Centers are critical for such applications to be adopted by users and organizations.

In the basic non-secure model [42], [43], a source \( s \) needs to store a file \( M \) which is decomposed into \( k \) messages, in \( v \) servers (nodes), such that any legitimate user \( d \) (destination) can reconstruct the file by collecting the stored information from any \( l \) servers \((l = \rho(s, d_i) \geq k)\). The secured version constraints the stored chunks such that an eavesdropper which can observe the information stored at any \( w \) servers will be kept ignorant regarding the actual file stored (see Figure 4).

For the secured version, we can leverage the individual-SMSSM coding scheme suggested herein to enhance the non-secure solution suggested in [42], [44]–[47], which consider each node in the network as a server which maintains pieces of data using RLNC. We will be able to guarantee that any eavesdropper that can access any \( w \) servers will have no information regarding any stored message individually (zero mutual information regarding each message separately). Specifically, each source \( s \) encodes the original data file \( M \) using the individual security coding scheme suggested herein (Sections IV and VII) and then uploads the encoded packets to the \( v \) servers. The number and the size of packets uploaded to the servers in the secure solution suggested are as in the non-secure model; thus, we obtain the full capacity of the system.

It is important to note that utilizing the individual security coding scheme suggested in this paper, one not only ensures individual secrecy from potential eavesdroppers, but also can guarantee privacy from the hosting servers themselves, such that, each server not only will not be able to decode the original data but will have zero information regarding any of the stored message individually. For example, assume that in the example depicted in Figure 4, the source \( s \) (private user) wants to store a file \( M \) in the cloud. To do that, the source can utilize 3 different cloud storage providers, such as Google Drive, Microsoft OneDrive, Dropbox, etc. However, the source wants to keep the original information private. Hence, by encoding the original data using the individual security coding scheme suggested at the source, and then uploading at most 3 encoded packets \( Y_{i1}, \ldots, Y_{i3} \) to any provider, the provider will store the packets in their servers \( v \), but these will be kept ignorant of the original file.

B. Wireless Networks

The inherent broadcast nature of the wireless medium makes network coding techniques pertinent for wireless networks. Specifically, relying on network coding, instead of sending packets (unicast, multicast or broadcast packets) to each intended addressee individually, a source (or an intermediate node which needs to relay packets toward the destination) can transmit a manipulation (usually a linear combination) of the packets destined to the various receivers. A receiver collecting sufficient number of such combinations (coded packets) can reconstruct (decode) the original packets. Relying on NC when the channel is lossy, i.e., there is a probability that a sent packet will not be received (decoded) by its intended receiver (receivers), has great advantages as instead of resending each unencoded packet until received correctly by its intended receiver, a sender keeps sending combinations of the original packets until each receiver collects a sufficient number of combinations (e.g., [48]–[54]). Accordingly, a sender can \( \text{a priori} \) estimate the number of coded packets needed according to the most
lossy channel and send coded packets accordingly, without relying on any feedbacks mechanism.

The secured version of this data dissemination problem requires that an eavesdropper with a degraded channel which can obtain only a subset of the transmitted packet will not be able to obtain any information regarding any of the original packets. Utilizing the individual security coding scheme suggested in this paper, in which the source estimates the number of packets needed to be sent according to the estimated packet loss to each receiver, encodes the messages before the wireless transmission according to the procedure presented in Section IV and the anticipated packet loss to the eavesdropper and broadcast the coded packets ensuring individual security as proved in this paper.

C. Live Broadcast of Video with Multi-Path Streaming

Multi-Path routing techniques which enable the use of multiple alternative paths between a source and a destination through the network, has been widely exploited over the years to provide a variety of benefits such as load balancing, fault tolerance, bandwidth enhancement, etc. One such ubiquitous example is LiveU innovative solution for distributing live video streams via wireless networks [55], [56]. In these systems, the real-time recorded video is encoded in packets by the source. These encoded packets include pieces of the data to be transmitted through different distributed media. For example, the pieces of the data transmitted over various technologies such as cellular networks, WiFi, satellite, fiber internet, etc. or various providers, e.g., Sprint, T-Mobile, AT&T Verizon, etc. A local server at the legitimate client decodes the data received from the different distributed media. This distributed streaming system maintains a high-quality viewer experience and cost-efficiency since the source can adapt the number of pieces dynamically to be transmitted by the different media. For example, if the connection using cellular or WiFi is lost during the real-time transmission, the source can route the pieces of the data dynamically by other connections or medias, taking into account the cost of each transmission by the optional connections.

In context to individual security suggested herein, we consider the case where there is an eavesdropper which has access to only a subset of the connections during the real-time distributed streaming (we assume that the eavesdropper can access any set of the streams unknown to the source, yet only a subset thereof). Utilizing the individual security coding scheme suggested in this paper, i.e., encoding the packets prior to the transmission, according to the coding scheme suggested in Section IV guarantees Individual Secure Live Broadcast of Video with Multi-Path Streaming, such that an eavesdropper which can capture at most \( w \) streams transmitted over the different distributed media is kept ignorant in the sense of having zero mutual information, regarding any set of \( k \) messages individually, yet may potentially obtain insignificant information about mixtures of packets transmitted. Figure 6 depicts a graphical representation of this system.
VII. LINEAR CODE CONSTRUCTION FOR INDIVIDUAL-SMSM

The suggested code given in Section IV is based on random coding. Therefore, in that case, the individual secrecy constraint holds only asymptotically, that is, \( H(M_{s,j}|Z_{w})/H(M_{s,j}) \to 1 \) as \( k \) grows. In this section, we design a structured linear code, which results in zero mutual information, such that there is no requirement for \( k \) to grow, yet, without any decrease in the rate or any message “blow-up” by extra randomness. Hence, we prove the following corollary.

Corollary 2. With binary linear code, \( k_s \)-individual security in SMSM networks holds, keeping an eavesdropper which observes \( \rho(s;z) \) links ignorant with respect to any set of \( k_s \leq k - w \) messages, if for each source \( s \in S \) to each destination \( d \in D \), \( \rho(s,d) \geq k_s \), for all \( d \in D \), \( \rho(S,d) \geq (k)|S| \), \( \rho(s;z) \leq w \) and \( k \) satisfies

\[
    k \geq \left\lceil \frac{\rho(s,d)}{\rho(s,d) - \rho(s;z)} \right\rceil \geq 2.
\]

We may now turn to the detailed construction and proof of the Individual-SMSM structured linear code.

1) Codebook Generation: Let \( C \) be a binary linear code of length \( k \) and dimension \( w \), and set \( k' = k - w \). Then, let

\[
    \mathbf{H} = \begin{bmatrix} \mathbf{H}_1; \mathbf{H}_2; \ldots; \mathbf{H}_{k'} \end{bmatrix} \in \{0,1\}^{k' \times k}
\]

be a parity check matrix, which we assume has rank \( k' \). This linear code defines \( 2^{k-w} \) cosets, one of them is the code itself. We denote the cosets by \( \{A_m\} \), \( 1 \leq m \leq 2^{k-w} \). Note that each coset is of size \( 2^w \). Hence the cosets of this code correspond to the bins we used in Section IV.

Let \( \mathbf{G} \) be a generator matrix for \( C \). We thus denote

\[
    \mathbf{G} = \begin{bmatrix} \mathbf{G}_1; \mathbf{G}_2; \ldots; \mathbf{G}_{k'} \end{bmatrix} \in \{0,1\}^{w \times k},
\]

and we select a matrix

\[
    \mathbf{G}^* = \begin{bmatrix} \mathbf{G}_1^*; \mathbf{G}_2^*; \ldots; \mathbf{G}_{k'}^* \end{bmatrix} \in \{0,1\}^{k' \times k}
\]

with \( k' \) linearly independent rows from \( \{0,1\}^k \setminus C \). That is, \( \mathbf{G}^* \) spans the null space of \( C \).

2) Source and legitimate Node encodings: At each source node, \( s \), the encoder selects, for each column \( i \) a codeword \( \mathbf{x}^k(e(i)) \) out of the \( 2^w \) members of the coset \( A_m \), where \( m \) is given by the index \( M_{s,1}(i); \ldots; M_{s,k}(i) \) and \( e(i) = M_{s,k+1}(i); \ldots; M_{s,k}(i) \). That is similar to Section IV \( k' = k - w \) bits of the column choose the coset, and the remaining \( w \) bits choose the codeword within the coset. This is equivalent to letting \( \mathbf{X}_s(i) \) be a choice from the \( 2^w \) solutions of

\[
    (M_{s,1}(i); \ldots; M_{s,k}(i)) = \mathbf{HX}_s(i).
\]

Again note that \( \mathbf{X}_s \) is of the same size as \( \mathbf{M} \).

Proposition 1 below shows that, in fact, \( \mathbf{X}_s \) can be easily computed using matrix multiplication.

\[
    \mathbf{X}_s(i)^T = M_{s,1}(i)\mathbf{G}_1^* + \ldots + M_{s,k'}(i)\mathbf{G}_{k'}^* \\
    + M_{s,k'+1}(i)\mathbf{G}_1 + \ldots + M_{s,k}(i)\mathbf{G}_w \\
    = \mathbf{M}_s(i)^T \begin{bmatrix} \mathbf{G}^* \end{bmatrix}.
\]

Proof. Define \( \mathbf{X}_s(i)^T \) according (2). We wish to show that this definition is indeed consistent with (1), that is, using the definition in (2) the bits \( M_{s,1}(i); \ldots; M_{s,k'}(i) \) define the coset in which \( \mathbf{X}_s(i) \) resides, and, furthermore, the remaining \( w \) bits, \( M_{s,k'+1}(i); \ldots; M_{s,k}(i) \), uniquely define the word within the coset.

To this end, take the transposed of equation (2), and multiply it by \( \mathbf{H} \). We have:

\[
    \mathbf{HX}_s(i) = \mathbf{HM}_{s,1}(i)(\mathbf{G}_1^*)^T + \ldots + \mathbf{HM}_{s,k'}(i)(\mathbf{G}_{k'}^*)^T \\
    + \mathbf{HM}_{s,k'+1}(i)(\mathbf{G}_1)^T + \ldots + \mathbf{HM}_{s,k}(i)(\mathbf{G}_w)^T \\
    = \begin{bmatrix} M_{s,1}(i) & \ldots & M_{s,k'}(i) & 0 & \ldots & 0 \\
                     \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
                     0 & \ldots & 0 & 1 & \ldots & 1 \\
    \end{bmatrix} \\
    = \begin{bmatrix} M_{s,1}(i) \\
                     \vdots \\
                     M_{s,k'}(i) \\
    \end{bmatrix},
\]

where the second inequality is since \( \mathbf{G}^* \) is our choice of a basis to the null space of the code, hence, we can take \( \mathbf{G}^* \) such that \( \mathbf{H}(\mathbf{G}^*)^T = \mathbf{I} \). Moreover, since \( \mathbf{H} \) is a parity check matrix for the code, it is orthogonal to all codewords. Thus, the first \( k' \) bits define the coset.

Now, since the bits \( M_{s,k'+1}(i); \ldots; M_{s,k}(i) \) create a linear combination of codewords, the addition of such a linear combination does not change the coset. \( \mathbf{X}_s(i)^T \) remains in the same coset regardless of these bits. Yet, as \( (\mathbf{G}_1; \ldots; \mathbf{G}_w) \) is of rank \( w \), all \( 2^w \) possibilities for the linear combination are distinct, creating distinct vectors \( \mathbf{X}_s(i)^T \) within the coset.

Then, the network code is similar to Section IV. That is, the sources transmit linear combinations of the rows, with random coefficients. Nodes transmit random linear combinations of the vectors in \( \mathbf{S}_w \), which is maintained by each node according to the messages received at the node.

A. Reliability

As for the reconstruction of \( \mathbf{X}_s \), the reliability part using RLNC is almost a direct consequence of (2). Again, the min-cut is given by Theorem 1 and, the legitimate nodes can easily reconstruct \( \mathbf{X}_s \) for each \( s \).

Now, since the code \( C \) is chosen to be \( (k,w) \) code, each partial column \( M_{s,1}(i); \ldots; M_{s,k}(i) \) can be seen as a syndrome of \( C \) with respect to a \( k' \times k \) parity-check matrix \( \mathbf{H} \) (as given
in [57] Section 4]). Thus, each destination can map $X_s$ back to $M_s$. First, compute the bin index: where

$$(M_{s,1}(i); \ldots ; M_{s,k}(i)) = HX_s(i).$$

Then, $M_{s,k'}(i); \ldots ; M_{s,k}(i)$ are simply the index of $X_s$ within that bin.

In the same way, using a gossip protocol, the reliability proof is almost a direct consequence of [29, Theorem 1]. Hence, the number of rounds required is given by Theorem 3.

### B. Information Leakage at the Eavesdropper

Denoted by $C$ the code and by $X_s$ the set of codewords corresponding to $M_{s,1} \ldots M_{s,k}$. We assume that the eavesdropper has full knowledge of the code $C$ as well.

As given in Section IV-B to analyze the information leakage at the eavesdropper, note that Eve has access to at most $w$ linear combinations on the rows of $X_s$. We will assume these linear combinations are in fact independent, and since Eve has access to the coefficients, we will assume Eve can even use Gaussian elimination and have access to $w$ rows from the original matrix $X_s$.

Next, we note that the columns of $X_s$ are independent (by the construction of the codebook, as creation $X_s$ is done independently per-column). Hence, it suffices to consider the information leakage for each column $i \in \{1, \ldots , c\}$ from $X_s$ separately. Thus, using techniques given in [57], we calculate the eavesdroppers uncertainty $H(M_{s,k-w}(Z_w))$ by first evaluating $H(M_{s,k-w}(Z_w)|Z_w)$ for each equally likely column.

If a coset of $C$ contains at least one word that agrees with $Z_w$, we say that the coset is consistent with $Z_w$. Let $N(C,Z_w(i))$ denote the total number of cosets consistent with $Z_w$. If the number of words consistent with $Z_w(i)$ in each coset is the same, we will have

$$H(M_{s,k-w}(Z_w)|Z_w) = \log_2 N(C,Z_w(i)).$$

For an $(k,w)$ code $C$, the maximum possible value for $N(C,Z_w(i))$ is the total number of cosets, $2^k$. If $N(C,Z_w(i)) = 2^k$, we say that $Z_w(i)$ is individually secured by $C$, since the total number of words which can be consistent with $Z_w(i)$ is $2^k$ (as $Z_w(i)$ is of length $w = k - k'$), hence they are all in different cosets, hence $H(M_{s,k-w}(Z_w(i)) = \log_2 N(C,Z_w(i))$ and $Pr(M_{s,k-w}(Z_w(i)) = 1/2^k$ for every possible column $M_{s,k-w}(Z_w(i))$.

The following lemma states a condition for a column $Z_w(i)$ at the eavesdropper to be individually secured by a code $C$.

**Lemma 1.** Let $(k,w)$ binary linear code $C$ have a generator matrix $G$, where $G_j$ denote the $j$-th column of $G$. Consider an eavesdroppers observation $Z_w(i)$, with access to $w$ positions $\{j : X_{s,j}(i) \neq ?\} = \{j_1,j_2, \ldots, j_w\}$ from a certain column $i \in \{1, \ldots , c\}$ of the original matrix $X_s$. $Z_w(i)$ is individually secured by $C$ iff the matrix $G_w = [G_{j_1}; G_{j_2}, \ldots ; G_{j_w}]$ has rank $w$ for any set of indices $\{j_1,j_2, \ldots, j_w\}$.

**Proof.** Remember that the encoding is defined by $M_s(i)^T [G^*; G]$. The eavesdropper has access to at most $w$ bits from each codeword. Eve is interested in the linear combination which created the codeword $X_s(i)^T, M_s(i)^T$. If $G_w$ has rank $w$, the code $C$ has codewords with all possible sequences of $2^w$ in the $w$ positions $\{j : X_{s,j}(i) \neq ?\} = \{j_1,j_2, \ldots, j_w\}$ obtained by the eavesdropper. Since Eve sees only $w$ positions, there are $2^k - w$ additional dimensions. Again, since cosets are obtained by translating the code $C$, each coset of $C$ has one of these possibilities.

If, however, $G_w$ has rank less than $w$, the code $C$ does not have all $2^w$ possible sequences in the $w$ revealed positions. Hence, there exists at least one coset that does not contain a given sequence which matches Eve’s data in the $w$ revealed positions, and therefore $N(C,Z_w(i)) < 2^k$.

To conclude, if one uses a good linear $(k,w)$ code to create the cosets, for Eve, all matching codewords are equiprobable and reside in different cosets, hence, $I(M_{s,k-w};Z_w) = 0$.

### VIII. Code Construction and a Proof for Strong-SMMSM

In this section, we design a random code, which results with strong-secrecy, i.e., requiring Eve’s mutual information with all messages simultaneously to be zero, yet, at price of rate as given in [5]. However, using the suggest random code herein, the field size is determined only by the network coding scheme, that is, only by the requirement for reliability, and is not increased by the strong-security constraints.

At each source node $s \in \{1, \ldots , 5\}$, we randomly map each column of the message matrix $M_s$. As depicted in Figure 7 in the code construction phase, for each possible column of the $s$-th message matrix we generate a bin, containing several columns. The number of such columns corresponds to $w$, the number of packets that the eavesdropper can wiretap, in a relation that will be made formal in the sequel. Then, to encode, for each column of the message matrix, we randomly select a column from its corresponding bin. This way, a new, $n \times c$ message matrix $X_s$ is created. Specifically, a Strong-SMMSM code at the $s$-th source node consists of a messages matrix $M_s$ of $M_{s,1} \ldots M_{s,k}$ messages of length $c$ bits over the
binary field, we denote the set of matrices by $M_s$. A discrete memoryless source of randomness over the alphabet $\mathcal{R}$ and some known statistics $p_R$; An encoder,

$$f : M_s \times \mathcal{R} \to X_s \in \{0, 1\}^{n \times c}$$

which maps each message matrix $M_s$ to a matrix $X_s$ of codewords. This message matrix contains $n \geq k + w$ new messages of size $c$.

The need for a stochastic encoder is similar to most encoders ensuring information theoretic security, as randomness is required to confuse the eavesdropper about the actual information [58]. Hence, we define by $R_k$ the random variable encompassing the randomness required for the $k$ messages at the source node, and by $\Delta$ the number of columns in each bin. We may now turn to the detailed construction and analysis.

1) Codebook Generation: Set $\Delta = 2^{w+n e}$. Where $P(x) \sim$ Bernoulli $(1/2)$, using a distribution $P(X^n) = \prod_{j=1}^{n} P(x_j)$, for each possible column in the message matrix generate $\Delta$ independent and identically distributed codewords $x^n(e)$, $1 \leq e \leq \Delta$, where $\epsilon \geq 1/n$.

2) Source and legitimate Node encodings: For each column $i$ of the $s$-th message matrix $M_s$, the $s$-th source node selects uniformly at random one codeword $x^n(e)$ from the $i$-th bin. Therefore, the $s$-th source Strong-SMSM matrix $X_s$ contains $c$ randomly selected codewords of length $n$, one for each column of the $s$-th message matrix. Then, the sources transmit linear combinations of the rows, with random coefficients. Nodes transmit random linear combinations of the vectors in $S_s$, which is maintained by each node according to the messages received at the node.

The reliability in the Strong-SMSM algorithm is inherited from the reliability in RLNC. That is if min-cuts are $p(s, d) \geq k + w$ and $p(S, d) \geq (k + w)|S|$ for each $s \in S$ and $d \in D$ then $k + w = n$ messages can be transmitted reliably from each source to all destinations. Since the transformation $M_s$ to $X_s$ can be inverted, the destinations can decode the original messages.

A. Information Leakage at the Eavesdropper

We now prove the strong-security constraint is met. In particular, for the strong constraint, we wish to show that $I(M_i; Z_w)$ is small for all $s \in S$. We will do that by showing that given $Z_w$, Eve’s information, all possibilities for $M_s$ are equally likely, hence Eve has no intelligent estimation for $M_s$.

Denote by $C_n$ the random codebook and by $X_s$ the set of codewords corresponding to $M_{s1} \ldots M_{sk}$. To analyze the information leakage at the eavesdropper, note that Eve has access to at most $w$ linear combinations on the rows of $X_s$. We will assume these linear combinations are in fact independent, and since Eve has access to the coefficients, we will assume Eve can even use Gaussian elimination and have access to $w$ rows from original matrix $X_s$.

Next, note that the columns of $X_s$ are independent (by the construction of the codebook, creating $X_s$ is done independently per-column; $c$ columns are used only to reduce the NC overhead). Hence, it suffices to consider the information leakage for each column $i \in \{1, \ldots, c\}$ from $X_s$ separately.

For each column $i$ of $M_s$, the encoder has $\Delta$ independent and identically distributed codewords, out of which one is selected. Hence, there is an exponential number of codewords, from the eavesdropper’s perspective, that can generate a column in $X_s$, and we require that Eve is still confused even given the $w$ bit from each column.

Let $Z_w(i)$ be the $w$ bits Eve has from column $i$. Denote $l = n - w$. Define by $\text{Sh}(Z_w(i), l)$ the set of all $n$-tuples consistent with $Z_w(i)$, i.e.,

$$\text{Sh}(Z_w(i), l) = \{b^n : b^n(S_Z) = Z_w(i)\},$$

where $S_Z$ denotes indices of the rows Eve has. Clearly, there are $2^l$ tuples in $\text{Sh}(Z_w(i), l)$.

We assume Eve has the codebook, yet does not know which column from each bin is selected to be the codeword. Hence, we wish to show that given $Z_w(i)$, Eve will have at least one candidate per bin. The probability for a codeword to fall in a given shell is

$$Pr(X^n_s(i) \in C_n \cap X^n_s(i) \in \text{Sh}(Z_w(i), l)) = \frac{\text{Vol}(\text{Sh}(Z_w(i), l))}{2^n} = \frac{2^{n-w}}{2^n}.$$ 

In each bin of $C_n$, we have $\Delta = 2^{w+n e}$ codewords. Thus, the average number of codewords Eve sees in her shell, per bin is

$$|\{m(i) : X^n(i) \in \text{Sh}(Z(i), l)\}| = \frac{2^{w+n e} + 2^{n-w}}{2^n} = 2^{n e}.$$ 

Hence, we can conclude that on average, and if $ne$ is not too small, for every column in $M_s$, Eve has a few possibilities in each bin, hence cannot locate the right bin. However, it is still important to show that all bins have (asymptotically) equally likely number of candidate codewords, hence Eve cannot locate a preferred bin.

To this end, we proved that the average number of codewords per column is very close to $2^{n e}$ with high probability. We wish to show that the probability that the actual number of options deviates from the average by more than $\epsilon$ is small.

Define

$$E_{C_1}(Z(i), l) := Pr\{(1-\epsilon)2^{n e} \leq |m(i) : X^n_s(i) \in \text{Sh}(Z_w(i), l)| \leq (1+\epsilon)2^{n e}\}.$$ 

By the Chernoff bound, we have

$$Pr(E_{C_1}(Z(i), l)) \geq 1 - 2^{-\epsilon^2 2^{n e}}.$$ 

Due to the super exponential decay in $n$, when taking a union bound over all columns, the probability that Eve decodes correctly some column is small. Hence, by the chain rule for entropies, since all the codewords in the codebook are independent $H(X_s) = \sum_{i=1}^{n_c} H(X_s(i))$, for Eve, all codewords are almost equiprobable and $I(M_i; Z_w) \to 0$.

IX. Conclusions

In this paper, we proposed SMSM codes under an Individual Security constraint. In this model, the eavesdropper is kept ignorant, in the sense of having zero mutual information regarding each message separately, yet may potentially obtain
information about mixtures of packets transmitted. In fact, it ensures Eve is kept ignorant of any set of $k - w$ messages. That is, guarantee zero mutual information, with respect to any set of $k - w$ messages.

We completely characterized the rate region for individually secure MSM. Specifically, we showed that secure communication is achievable up to the min-cut, that is, without any decrease in the rate or any message “blow-up” by extra randomness. Moreover, we provided a code for Strong-SMSSM by extra randomness, i.e., requiring Eve’s mutual information with all messages simultaneously to be zero. While this included a rate loss, it is important to note that in the code suggested the alphabet size did not increase with the network parameters due to the strong-security constraint.

Finally, we showed a few examples out of many important applications, like data centers, wireless networks, gossip and live broadcasting of video, for which the individual security coding schemes suggested is applicable, and achieves the full capacity of these systems.

REFERENCES

[1] S.-Y. Li, R. W. Yeung, and N. Cai, “Linear network coding,” IEEE transactions on information theory, vol. 49, no. 2, pp. 371–381, 2003.
[2] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong. “A random linear network coding approach to multicast,” IEEE Transactions on Information Theory, vol. 52, no. 10, pp. 4413–4430, 2006.
[3] N. Cai and R. W. Yeung, “Secure network coding,” in Information Theory, 2002. Proceedings. 2002 IEEE International Symposium on. IEEE, 2002.
[4] T. Chan and A. Grant. “Capacity bounds for secure network coding,” in Communications Theory Workshop, 2008. AusCTW 2008. Australian. IEEE, 2008, pp. 95–100.
[5] N. Cai and R. W. Yeung, “Secure network coding on a wiretap network,” Information Theory, IEEE Transactions on, vol. 57, no. 1, pp. 424–435, 2011.
[6] S. Y. El Rouayheb and E. Soljanin, “On wiretap networks II,” in Information Theory, 2007. ISIT 2007. IEEE International Symposium on. IEEE, 2007, pp. 551–555.
[7] D. Silva and F. R. Kschischang, “Security for wiretap networks via rank-metric codes,” in Information Theory, 2008. ISIT 2008. IEEE International Symposium on. IEEE, 2008, pp. 176–180.
[8] S. El Rouayheb, E. Soljanin, and A. Sprintson, “Secure network coding for wiretap networks of type II,” Information Theory, IEEE Transactions on, vol. 58, no. 3, pp. 1361–1371, 2012.
[9] N. Cai and R. W. Yeung, “A security condition for multi-source linear network coding,” in Information Theory, 2007. ISIT 2007. IEEE International Symposium on. IEEE, 2007, pp. 561–565.
[10] Z. Zhang and R. W. Yeung, “A general security condition for multi-source linear network coding,” in Information Theory, 2009. ISIT 2009. IEEE International Symposium on. IEEE, 2009, pp. 1155–1158.
[11] N. Cai, “Valuable messages and random outputs of channels in linear network coding,” in Information Theory, 2009. ISIT 2009. IEEE International Symposium on. IEEE, 2009, pp. 413–417.
[12] N. Cai and T. Chan, “Theory of secure network coding,” Proceedings of the IEEE, vol. 99, no. 3, pp. 421–437, 2011.
[13] T. H. Chan and A. Grant, “Network coding capacity regions via entropy functions,” IEEE Transactions on Information Theory, vol. 60, no. 9, pp. 5347–5374, 2014.
[14] D. Kobayashi, H. Yamamoto, and T. Ogawa, “Secure multiplex coding attaining channel capacity in wiretap channels,” IEEE Transactions on Information Theory, vol. 59, no. 12, pp. 8131–8143, 2013.
[15] A. S. Mansour, R. F. Schaefer, and H. Boche, “Secrecy measures for broadcast channels with receiver side information: Joint vs individual,” in Information Theory Workshop (ITW), 2014 IEEE. IEEE, 2014, pp. 426–430.
[16] Y. Chen, O. O. Koyluoglu, and A. Sezgin, “On the individual secrecy rate region for the broadcast channel with an external eavesdropper,” in 2015 IEEE International Symposium on Information Theory (ISIT). IEEE, 2015, pp. 1347–1351.
[17] A. S. Mansour, R. F. Schaefer, and H. Boche, “The individual secrecy capacity of degraded multi-receiver wiretap broadcast channels,” in 2015 IEEE International Conference on Communications (ICC). IEEE, 2015, pp. 4181–4186.
[18] ——, “On the individual secrecy capacity regions of the general, degraded and gaussian multi-receiver wiretap broadcast channel,” IEEE Transactions on Information and Security, 2016, vol. 11, no. 9, pp. 2107–2122, 2016.
[19] M. Goldenbaum, R. F. Schaefer, and H. V. Poor, “The multiple-access channel with an external eavesdropper: Trusted vs. untrusted users,” in 2015 49th Asilomar Conference on Signals, Systems and Computers. IEEE, 2015, pp. 564–568.
[20] Y. Chen, O. O. Koyluoglu, and A. H. Vinck, “On secure communication over the multiple access channel,” International Symposium on Information Theory and Its Applications (ISITA), 2016 IEEE, 2016.
[21] K. Bhattacharjee and K. R. Narayanan, “Weakly secure network coding,” NetCod, Apr, vol. 104, 2005.
[22] L. Lima, M. Médard, and J. Barros, “Random linear network coding: A free cipher?” in Information Theory, 2007. ISIT 2007. IEEE International Symposium on. IEEE, 2007, pp. 546–550.
[23] J. Claidige and I. Chatzigeorgiou, “Probability of partially solving random linear systems in network coding,” arXiv preprint arXiv:1607.04725, 2016.
[24] S. Deb, M. Médard, and C. Choute, “Algebraic gossip: A network coding approach to optimal multiple rumor mongering,” Information Theory, IEEE Transactions on, vol. 52, no. 6, pp. 2486–2507, 2006.
[25] A. Demers, D. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. Sturgis, D. Swinehart, and D. Terry, “Epidemic algorithms for replicated database maintenance,” in Proceedings of the sixth annual ACM Symposium on Principles of distributed computing. ACM, 1987, pp. 1–12.
[26] S. Deb, M. Medard, and C. Choute, “On random network coding based information dissemination,” in Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on. IEEE, 2005, pp. 278–282.
[27] B. Haeupler, “Analyzing network coding gossip made easy,” in Proceedings of the 43rd annual ACM symposium on Theory of computing. ACM, 2011, pp. 293–302.
[28] A. S. Mansour, R. F. Schaefer, and H. Boche, “The individual secrecy capacity of the gaussian SISO and degraded gaussian MIMO multi-receiver wiretap channel,” in 2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE, 2015, pp. 365–369.
[29] A. Cohen, B. Haeupler, C. Amin, and M. Médard, “Network coding based information spreading in dynamic networks with correlated data,” IEEE Journal on Selected Areas in Communications, vol. 33, no. 2, pp. 213–224, 2015.
[30] S. Jaggi, P. Sanders, P. A. Chou, M. Effros, S. Egner, K. Jain, and L. M. Tolhuizen, “Polynomial time algorithms for multicast network code construction,” IEEE Transactions on Information Theory, vol. 51, no. 6, pp. 1973–1982, 2005.
[31] R. Karp, C. Schindelhauer, S. Shenker, and B. Vocking, “Randomized rumor spreading,” in Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on. IEEE, 2000, pp. 565–574.
[32] D. Kempe, A. Dobra, and J. Gehrke, “Gossip-based computation of aggregate information,” in Foundations of Computer Science, 2003. Proceedings. 44th Annual IEEE Symposium on. IEEE, 2003, pp. 482–491.
[33] S. Boyd, A. Ghosh, P. Prabhakar, and D. Shah, “Randomized gossip algorithms,” Information Theory, IEEE Transactions on, vol. 52, no. 6, pp. 2508–2530, 2006.
[34] J. Feldman, T. Malkin, R. A. Servedio, and C. Stein, “Secure network coding via filtered secret sharing.” Citeseer.
[35] B. K. Dey, S. Jaggi, and M. Langberg, “Sufficiently myopic adversaries are blind,” in Information Theory (ISIT), 2015 IEEE International Symposium on. IEEE, 2015, pp. 1164–1168.
[36] S. Ghemawat, H. Gobioff, and S.-T. Leung, “The google file system,” in ACM SIGOPS operating systems review, vol. 37, no. 5. ACM, 2003, pp. 29–43.
[37] G. DeCandia, D. Hastorun, M. Jampani, G. Kakulapati, A. Lakshman, A. Pilchin, S. Sivasubramanian, P. Voshall, and W. Vogels, “Dynamo: amazon’s highly available key-value store,” ACM SIGOPS operating systems review, vol. 41, no. 6, pp. 205–220, 2007.
[38] F. Chang, J. Dean, S. Ghemawat, W. C. Hsieh, D. A. Wallach, M. Burrows, T. Chandra, A. Fikes, and R. E. Gruber, “Bigtable: A distributed storage system for structured data,” ACM Transactions on Computer Systems (TOCS), vol. 26, no. 2, p. 4, 2008.
