Network Information Theoretic Security With Omnipresent Eavesdropping

Hongchao Zhou, *Member, IEEE*, and Abbas El Gamal, *Life Fellow, IEEE*

***Abstract*— Shannon showed that to achieve perfect secrecy in point-to-point communication, the message rate cannot exceed the shared secret key rate giving rise to the simple one-time pad encryption scheme. In this paper, we extend this work from point-to-point to networks. We consider a connected network with pairwise communication between the nodes and assume that each node is provided with a certain amount of secret bits before communication commences. An eavesdropper with unlimited computing power has access to all communication and can hack a subset of the nodes not known to the rest of the nodes. We investigate the limits on information-theoretic secure communication with end-to-end encryption for this network. We establish a tradeoff between the secure channel rate (for a node pair) and the secure network rate (sum over all node pair rates) and show that information-theoretic secrecy can be achieved asymptotically if and only if the sum rate of any subset of unhacked channels does not exceed the shared unhacked-secret-bit rate of these channels. We also propose a practical scheme that achieves a good balance of network and channel rates with information-theoretic secrecy guarantee. This work has a wide range of potential applications for which strong secrecy is desired, such as cyber-physical systems, distributed-control systems, and ad-hoc networks.**

***Index Terms*— Network information theoretic security, all communication eavesdropped, network capacity.**

I. Introduction

THE information-theoretic security introduced by Shannon [1] and widely accepted as the strictest notation of security, is becoming increasingly attractive for many cyber-physical systems, distributed-control systems, wireless ad-hoc networks, among other applications. Secure network coding [2] has been well studied to guarantee the information-theoretic security when a subset of channels are wiretapped [3], [4] or in the presence of Byzantine adversaries [5], [6]. In this paper, we make a stronger assumption: all the channels are eavesdropped and some nodes are hacked without knowledge of the rest of the nodes. This assumption is realistic for example in wireless networks in which an eavesdropper can sense the transmitted signals, or as another example, some network nodes communicate via an insecure public network. Under this assumption, pure network-coding approaches cannot work without the help of common randomness shared among the network nodes. Physical layer security [7]–[11] can be used to distribute secret keys among network nodes, however, the channel advantage required by the receivers over any eavesdropper is not easy to guarantee in a wireless network. In some scenarios, it is allowed to pre-distribute a very large number of secret bits to network nodes to support future secure communication, which also becomes realistic with the development of data-storage technology, which enables each node to carry enough secret bits. For instance, in some applications of mobile ad-hoc networks one may pre-distribute secret bits to network nodes at the same location before the network starts to work. In this paper, we are interested in a fundamental problem: if every node in the network is allowed to carry a certain large number of secret bits, how much information can be securely transmitted over the network with end-to-end encryption under the information-theoretic security criterion?

We consider a connected network of n nodes, with each carrying up to $l \gg n$ secret bits and at most $t \leq n-2$ nodes being hacked without knowledge of the other nodes. We assume pairwise communication with end-to-end encryption, that is each sender node encrypts its message using a secret key generated from the common randomness shared with the intended receiver node and only the receiver can decrypt it using the same key. A secure network-communication scheme includes two phases - the key pre-distribution phase and the communication phase. Through the process of key pre-distribution, each node gets a sequence of at most l secret bits, which are correlated among the network nodes (a special case is that each secret bit is distributed to multiple network nodes). If a node is hacked, all its secret bits will be leaked to eavesdroppers. When two nodes communicate to each other, they would like to utilize their common randomness to realize secure communication, but the problem is that this common randomness may not be secure as some bits might be hacked and some bits might have been used for communication before. Information reconciliation [12], [13], allowing two nodes with correlated random sequences to agree on a common shared string, and privacy amplification [14]–[17], distilling a secret key from a common shared string that is partially known by an eavesdropper, have been studied to solve the problem. With

Manuscript received May 27, 2020; revised September 16, 2021; accepted September 22, 2021. Date of publication October 1, 2021; date of current version November 22, 2021. An earlier version of this paper was presented in part at the 2020 IEEE International Symposium on Information Theory (ISIT), titled "Network Information Theoretic Security" [DOI: 10.1109/ISIT44484.2020.9174074]. *(Corresponding author: Hongchao Zhou.)*

Hongchao Zhou is with the School of Information Science and Engineering, Shandong University, Qingdao, Shandong 266237, China (e-mail: hongchao@sdu.edu.cn).

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

Communicated by J. Kliewer, Associate Editor for Coding Techniques.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2021.3116962.

Digital Object Identifier 10.1109/TIT.2021.3116962

these techniques, the two communication nodes can generate a secure secret key from correlated random sequences and then use the one-time pad encryption to transmit a message. As only the destination node can decrypt the transmitted message, such an end-to-end encryption can tolerate not only passive eavesdropping of the transmitted ciphertext but also active attacking from any relay node. With end-to-end encryption, it is impossible for any hacked node to modify a transmitted message or pretend to be another node sending messages, unless the attacker can derive all the common randomness shared between the source and the destination. As the total length of the messages to be transmitted over a channel (i.e., between two terminal nodes) cannot exceed the secret-bit length l of each node, we define their ratio as the channel rate, and the sum rate of all the channels as the network rate.

The secure-communication limit of a network depends on the secret bits distributed and the method of utilizing them to deliver information. As an example, consider a network with $n = 4$ nodes and $t = 1$ nodes being hacked. A straightforward way to pre-distribute the secret bits is to assign each pair of nodes $l/3$ common secret bits as the secret key which they can use with the one-time pad scheme. In this case, the messages are secure if and only if the rate of each channel does not exceed $1/3$. As the network size n increases, this approach can only reach channel rate of at most $1/(n-1)$, limiting its applications for large networks. An alternative way to reallocate the secret bits in the 4-node example is to assign every three nodes $l/3$ common secret bits, and hence there are four sequences of secret bits denoted by $\mathbf{u}_{123}, \mathbf{u}_{124}, \ldots$, where $\mathbf{u}_{123}$ is the secret bits distributed to nodes $1, 2, 3$. When two nodes say nodes 1 and 2 communicate to each other, they use $\mathbf{u}_{123} + \mathbf{u}_{124}$ as the secret key. In this case, no matter whether node 3 or node 4 is hacked, the messages are secure as long as the channel rate between node 1 and 2 does not exceed $1/3$. Compared to the former method, it can be proved that for a larger network of size n, by allowing each secret bit to be distributed to multiple nodes instead of only two nodes, the maximum channel rate (channel capacity) can be improved to more than $1/4$ from $1/(n-1)$. Our scheme can be regarded as another application of linear network coding. By utilizing the secret bits shared by multiple channels, higher communication rates with information-theoretic secrecy can be achieved.

We address several basic questions about our network setting: (1) What is the limit on the network rate and the channel rate for secure communication if the future network communication load is unknown? (2) Given an arbitrary distribution on the secret bits, how can we determine the security of a network with given channel rates? And (3) how to design a practical network-communication scheme that has both high network rate and high channel rate?

The rest of this paper is organized as follows. Section II provides the formulation and definitions of the problem for network communication with information theoretic security. Section III summaries some asymptotic results. Section IV introduces and investigates a practical and efficient key-distribution method named the combinational key distribution, and Section V studies secure network communication with asymptotically optimal privacy amplification, followed by simplified security criteria discussed in Section VI. Section VII further discusses the network security beyond the information theoretic limit, the application of network coding, and some open questions. The proofs of the main results are given in Section VIII.

II. Definitions

We consider a network consisting of a set of nodes $\mathcal{N} = \{1, 2, \ldots, n\}$, where every two nodes can find a path (channel) connecting them. We refer each pair of nodes as a channel, and use $\mathcal{P} = \{(i,j) | i \in \mathcal{N}, j \in \mathcal{N}, i < j\}$ to denote the set of all the channels. Every two nodes of a channel communicate with end-to-end encryption: they generate a secret key from their common randomness, encrypt the message with the one-time pad scheme, and then transmit the ciphertext through a path from the source to the destination.

It is assumed that all the channels are wiretaped and up to $t \leq n-2$ nodes could be hacked by an eavesdropper. Namely, every ciphertext transmitted over the network is possible to be known by eavesdroppers, and if a node is hacked all its secret bits are revealed to the eavesdropper. Each node in the network is able to store up to $l \gg n$ secret bits. To guarantee the network security, the total message length m_{ij} of a channel $(i,j) \in \mathcal{P}$ cannot exceed l. We call the number of message bits transmitted through a channel (i,j) per a node's secret bit as the *channel rate* r_{ij}, and the sum of the channel rates as the *network rate* r. Mathematically,

$$r_{ij} = \frac{m_{ij}}{l}, \quad r = \sum_{(i,j)\in\mathcal{P}} r_{ij}. \tag{1}$$

We use $\mathcal{N}_{\mathrm{h}} \subset \mathcal{N}$ with $|\mathcal{N}_{\mathrm{h}}| \leq t$ to denote the set of hacked nodes, and use $\mathcal{N}_{\mathrm{s}} = \mathcal{N}/\mathcal{N}_{\mathrm{h}}$ to denote the set of unhacked secure nodes. As a channel is insecure if one of its terminals is hacked, our goal is to protect those channels between secure nodes, denoted by $\mathcal{P}_{\mathrm{s}}$. Note that if $t \geq n-1$, it has $|\mathcal{P}_{\mathrm{s}}| = 0$, and in this case no message bits can be securely transmitted between any two nodes. In this paper, we assume that $t \leq n-2$ by default. As $\mathcal{P}_{\mathrm{s}}$ is assumed unknown, the network rate is defined above as the sum of the channel rates over all the channels instead of only the secure channels.

A secure network-communication scheme consists of two phases. In the key pre-distribution phase, the scheme generates n sequences of secret bits, denoted by $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n \in \{0,1\}^{l_i}$ with $l_i \leq l$, distributed to nodes $1, 2, \ldots, n$ respectively. To support secure communication between two nodes, it requires the sequences of the two nodes having sufficient common randomness, measured by the mutual information of the two sequences. One example of key pre-distribution is to assign a pool of independent and unbiased truly random bits to the network nodes, with each bit possibly assigned to multiple nodes. In the communication phase, we let $\mathbf{x}_{ij} \in \{0,1\}^{m_{ij}}$ be the message transmitted in channel (i,j) with $i < j$, and the corresponding ciphertext is $\mathbf{y}_{ij}$ with

$$\mathbf{y}_{ij} = \mathbf{x}_{ij} \bigoplus \mathbf{s}_{ij},$$

where $\mathbf{s}_{ij}$ is the secret key generated from the common randomness between node i and j. Note that sending the message

from node i to j is equivalent to sending it from node j to i without affecting the network security, as they result in the same ciphertext, and hence the communication is considered undirected. For simplicity, we assume there is a single message transmitted in each channel. If there are multiple messages transmitted in a channel one can concatenate them as a single one, and this does not change the security of the network.

A secure communication scheme ψ defines how to pre-distribute secret bits to the network nodes and how to extract secret keys for end-to-end encryption from the common randomness of any two nodes. A secure communication scheme ψ should be able to work with any specified parameters (n, t, l) with n the network size, t the maximum number of hacked nodes and l the maximum number of secret bits distributed to each node. If the communication demand, i.e. the message lengths $\mathbf{m} = \{m_{ij}\}$, are known during the key pre-distribute phase, the problem of designing a good scheme is trivial: a best way is to assign m_{ij} secret bits to the terminals of each channel (i, j), which then communicate with the one-time pad encryption. In this paper, it is assumed that the communication demand is unknown in advance, especially during the key pre-distribution phase, and it is desired to design schemes that can support a variety of real-time communication demands.

According to the assumption of the attack model, all the transmitted ciphertexts $\mathbf{y} = \{\mathbf{y}_{ij} | (i, j) \in \mathcal{P}\}$, the hacked secret bits $\mathbf{u}_{\mathrm{h}} = \cup_{j \in \mathcal{N}_{\mathrm{h}}} \mathbf{u}_j$ and the secure communication scheme ψ are known to the eavesdropper. Let $\mathbf{x}_{\mathrm{s}} = \{\mathbf{x}_{ij} | (i, j) \in \mathcal{P}_{\mathrm{s}}\}$ be the set of all the messages transmitted among unhacked nodes. Given a secure communication scheme ψ with end-to-end encryption, the information-theoretic security of a network with parameters (n, t, l) is measured by the maximum amount of information that can be possibly leaked to an eavesdropper. This security depends on the message lengths $\mathbf{m} = \{m_{ij}\}$, not necessarily their exact values or which of the two terminals is the source. As $\mathbf{m} = \mathbf{r} \cdot l$ with the channel rates $\mathbf{r} = \{r_{ij}\}$, we write the security (leaked information normalized by l) as

$$I_{\psi}(n, t, l, \mathbf{r}) = \frac{\max_{\mathcal{N}_{\mathrm{h}}, \mathbf{x}_{\mathrm{s}}} \mathbf{I}[\mathbf{x}_{\mathrm{s}}; \mathbf{y}, \mathbf{u}_{\mathrm{h}}]}{l} \tag{2}$$

where the maximum is over all possible combinations of hacked nodes and distributions of the messages. Our interests focus on the case of $l \gg n$, as the development of data-storage technology makes it more realistic and attractive to carry a large number of secret bits in network nodes. In particular, this paper studies some asymptotic results of the network information-theoretic security with $l \to \infty$ as a theoretical approximation.

Definition 1: Given a network of n nodes with at most $t \leq n - 2$ nodes being hacked, the channel rates $\mathbf{r} = \{r_{ij}\}$ are achievable (asymptotically) if and only if when $l \to \infty$, there exits a scheme ψ with end-to-end encryption such that

$$I_{\psi}(n, t, l, \mathbf{r}) = 0,$$

i.e., when the maximum number of secret bits l of each node is sufficiently large, the amount of leaked information can be arbitrarily close to zero. We call the set of achievable rates using a scheme ψ as the security region of the scheme, denoted by $\mathcal{R}(n, t, \psi)$.

Usually, it is not easy to express $\mathcal{R}(n, t, \psi)$ explicitly. To characterize the security region $\mathcal{R}(n, t, \psi)$, we introduce two metrics - the maximum channel rate and the maximum network rate, reflecting the maximal amount of information that can be securely communicated through a single channel or over the entire network asymptotically. It is desired to design schemes with relatively high maximum network rate and high maximum channel rate.

Definition 2: Given a network of n nodes with at most $t \leq n - 2$ nodes being hacked, the *maximum channel rate* of channel (i, j) with a scheme ψ is defined as the supremum of r_{ij} over all the achievable channel rates. The *maximum channel rate* of a scheme ψ is the minimum of the maximum channel rates over all the channels, denoted by

$$R_{\text{channel}}(n, t, \psi) = \min_{(i,j) \in \mathcal{P}} \sup_{\{r_{ij}\} \in \mathcal{R}(n,t,\psi)} r_{ij}.$$

The minimum is taken over all the channels, due to the assumption that the communication load is not known in advance.

Definition 3: Given a network of n nodes with at most $t \leq n - 2$ nodes being hacked, the *maximum network rate* of a scheme ψ is the supremum of all the achievable network rates of the scheme, denoted by

$$R_{\text{net}}(n, t, \psi) = \sup_{\{r_{ij}\} \in \mathcal{R}(n,t,\psi)} \sum_{(i,j) \in \mathcal{P}} r_{ij}.$$

To investigate the communication limit, we define the *channel capacity* and the *network capacity* of a network as the maximum over all the maximum channel rates and maximum network rates of any scheme with end-to-end encryption.

Definition 4: Given a network of n nodes with at most $t \leq n-2$ nodes being hacked, the *channel capacity* of the network is the maximum over all the maximum channel rates of any scheme with end-to-end encryption

$$C_{\text{channel}}(n, t) = \max_{\psi} R_{\text{channel}}(n, t, \psi),$$

and the *network capacity* of the network is the maximum over all the maximum network rates of any scheme with end-to-end encryption

$$C_{\text{net}}(n, t) = \max_{\psi} R_{\text{net}}(n, t, \psi).$$

III. Asymptotic Results

This section summaries some main asymptotic results when $l \to \infty$ and provides an overview of network secure-communication schemes.

From its definition, the maximum channel rate of a scheme is at most 1. But this upper bound cannot be reached when the maximum number of hacked nodes $t > 0$. The following two theorems study the network capacity and the channel capacity of a general network of n nodes with at most $t \leq n - 2$ nodes being hacked. The proofs are given in Section VIII-A and Section VIII-B. In this result of Theorem 1, the network capacity is independent of t, partially because the network rate

is the sum rate of all the channels, not only those in $\mathcal{P}_s$ (as $\mathcal{P}_s$ is unknown to the network). From the result of Theorem 2, a network with $n = 4$ and $t = 1$ has channel capacity of $1/3$. Interestingly, if we further increase the size of the network to 5, the channel capacity is still $1/3$.

Theorem 1: Given a network of n nodes with at most $t \leq n-2$ nodes being hacked, its network capacity is

$$C_{\text{net}}(n,t) = \frac{n}{2}.$$

Theorem 2: Given a network of n nodes with at most $t \leq n-2$ nodes being hacked, its channel capacity is

$$C_{\text{channel}}(n,t) = \frac{\binom{n-t-2}{a-2}}{\binom{n-1}{a-1}}$$

with $a = \lceil \frac{n}{t+1} \rceil$ the minimal integer larger than or equal to $\frac{n}{t+1}$.

Among all the ways of key pre-distribution, we are particularly interested in those that assign a pool of independent and unbiased truly random bits to the network nodes, for which any two secret bits from two different nodes are either identical or independent. We call them random-bit assignments. In this case, no information reconciliation is necessary for two nodes to agree on a common random sequence (the process of information reconciliation costs extra secret-bit resources). For a network with assigned random bits, let $\mathbf{u}_i$ be the sequence of random bits assigned to node i with $i \in \mathcal{N}$, and let $\mathbf{u}_{ij}$ be the sequence of common random bits shared by both node i and j. The secret key $\mathbf{s}_{ij}$ between node i and j is generated with a privacy-amplification method h such that

$$\mathbf{s}_{ij} = h(\mathbf{u}_{ij}),$$

which is then used for one-time pad encryption.

The following result shows that there is a certain tradeoff between the maximum network rate and the maximum channel rate of a scheme that assigns random bits to the network nodes when $t = 0$, implying that one may sacrifice the maximum network rate to gain a better maximum channel rate, and vice versa. The proof is provided in Section VIII-C.

Theorem 3: Given a network of n nodes without any nodes being hacked, for any scheme ψ based on random-bit assignment, it satisfies

$$R_{\text{net}}(n,0,\psi)\frac{2}{n+1} + R_{\text{channel}}(n,0,\psi)\frac{n-1}{n+1} \leq 1. \quad (3)$$

For any $1 \leq R_{\text{net}} \leq \frac{n}{2}$, the equality is achievable by a scheme.

Given a scheme, it is crucial to determine whether a network with current channel rates $\mathbf{r} = \{r_{ij}\}$ is information-theoretically secure or not. When the channel rates reach the limit, it may need to terminate the network communication for guaranteeing information-theoretic secrecy. Mathematically, we need a method to check whether $\mathbf{r} \in \mathcal{R}(n,t,\psi)$ for a scheme ψ. The difficulty arises from the fact that different channels may share some common secret bits, hence "interfere" with each other. Furthermore, it is easy to derive the accurate security region of a scheme for finite l, but when l is sufficiently large, the security region can be well approximated by the asymptotic result with infinitely large l.

Theorem 4: Given a network of n nodes with each node assigned up to l random bits, the channel rates $\{r_{ij}\}$ are achievable *if and only if* when $l \to \infty$, $\forall \mathcal{N}_h$ and $P \subseteq \mathcal{P}_s$,

$$\sum_{(i,j)\in P} r_{ij} < \frac{|\cup_{(i,j)\in P} \mathbf{u}_{ij}/\mathbf{u}_h|}{l} \text{ or } \sum_{(i,j)\in P} r_{ij} = 0, \quad (4)$$

where $\mathbf{u}_{ij}$ is the sequence of common secret bits shared by both node i and j, and $\mathbf{u}_h$ is the set of secret bits in hacked nodes.

Given the set of hacked nodes $\mathcal{N}_h$ and a set of channels between unhacked nodes $P \subseteq \mathcal{P}_s$, we call $\frac{|\cup_{(i,j)\in P}\mathbf{u}_{ij}/\mathbf{u}_h|}{l}$ the shared unhacked-secret-bit rate of the channels P. The above result shows that information-theoretic secrecy can be achieved if and only if for any subset of unhacked channels, the sum of their channel rates does not exceed the shared unhacked-secret-bit rate of these channels asymptotically. The proof of this theorem is given in Section VIII-D.

Example 1: Let us use a network of 4 nodes as an example of demonstrating the security criteria in Theorem 4. Assume that the secret bits are equally distributed to 4 groups of network nodes, i.e., $(1,2,3)$, $(1,2,4)$, $(1,3,4)$ and $(2,3,4)$. It means every 3 nodes share $l/3$ common secret bits.

If no nodes are hacked, according to Theorem 4, the network is secure for sufficiently large l if and only if

$$\begin{aligned} r_{12} &< \frac{|\mathbf{u}_{12}|}{l} = 2/3 \\ r_{12}+r_{13} &< \frac{|\mathbf{u}_{12}\cup\mathbf{u}_{13}|}{l} = 1 \\ r_{12}+r_{34} &< \frac{|\mathbf{u}_{12}\cup\mathbf{u}_{34}|}{l} = 4/3 \\ r_{12}+r_{13}+r_{14} &< \frac{|\mathbf{u}_{12}\cup\mathbf{u}_{13}\cup\mathbf{u}_{14}|}{l} = 1 \\ &\cdots \\ \sum_{(i,j)} r_{ij} &< 4/3 \end{aligned}$$

holds for every node permutation (not all permutations are listed here for simplicity). As some conditions can be derived by the others, the above conditions can be simplified as

$$r_{ij} < \frac{2}{3}, \sum_j r_{ij} < 1, \sum_{(i,j)} r_{ij} < 4/3.$$

They are necessary and sufficient to guarantee the security of the network.

Example 2: In the network of the above example, assume that at most 1 node is hacked. Then the network is secure for sufficiently large l if and only if

$$\begin{aligned} r_{12} &< \frac{|\mathbf{u}_{12}/\mathbf{u}_h|}{l} = 1/3 \\ r_{12}+r_{13} &< \frac{|\mathbf{u}_{12}\cup\mathbf{u}_{13}/\mathbf{u}_h|}{l} = 1/3 \\ &\cdots \end{aligned}$$

holds for every node permutation. It is equivalent to have

$$r_{12}+r_{13}+r_{23} < 1/3 \quad (5)$$

holding for every node permutation.

The achievability of Theorem 4 uses a simple method for privacy amplification that generates a secret key $\mathbf{s}_{ij}$ from the common sequence $\mathbf{u}_{ij}$ with a random linear transformation such that $\mathbf{s}_{ij} = M_{ij}\mathbf{u}_{ij}$ with a binary random matrix M_{ij} where each entry is one with probability $O(\log l/l)$ independently. For simplicity, we call it a random matrix of density $O(\log l/l)$. This leads to the following observation.

Corollary 5: Given a network of n nodes with each node assigned up to l random bits, the privacy-amplification method that generates

$$\mathbf{s}_{ij} = M_{ij}\mathbf{u}_{ij}$$

with a binary random matrix M_{ij} of density $O(\log l/l)$ for all (i,j) is asymptotically optimal when $l \to \infty$. The optimality means that any achievable channel rates $\{r_{ij}\}$ can be realized by this method.

From Theorem 4, we further derive the maximum channel rate and the maximum network rate of a network that is assigned sufficient random bits and uses an asymptotically-optimal privacy-amplification method for generating secret keys from common random bits.

Corollary 6: Given a network of n nodes, with each node assigned up to l random bits with $l \to \infty$ and at most $t \leq n-2$ nodes being hacked, the maximum channel rate of the network with asymptotically-optimal privacy amplification is

$$R_{\text{channel}} = \min_{\mathcal{N}_h, i,j | i,j \notin \mathcal{N}_\text{h}} \frac{|\mathbf{u}_{ij}/\mathbf{u}_h|}{l} \tag{6}$$

and the maximum network rate of the network with asymptotically-optimal privacy amplification is

$$R_{\text{network}} = \sum_{\mathcal{N}_\text{h}:|\mathcal{N}_\text{h}|=t} \frac{|\cup_{(i,j)\in\mathcal{P}_\text{s}} \mathbf{u}_{ij}/\mathbf{u}_h|}{l} / \binom{n-2}{t}. \tag{7}$$

The maximum channel rate is straightforward following the result of Theorem 4. The maximum network rate is based on the observation that

$$\sum_{\mathcal{N}_\text{h}:|\mathcal{N}_\text{h}|=t} \Big(\sum_{(i,j)\in\mathcal{P}_\text{s}} r_{ij} \Big) = \binom{n-2}{t} \Big(\sum_{(i,j)\in\mathcal{P}} r_{ij} \Big),$$

where $\binom{n-2}{t}$ is the number of combinations for $\mathcal{N}_\text{h}$ that does not include specific nodes i and j, equal to the number of times that r_{ij} appears on the left of the equality.

Besides Theorem 4, the following result provides an alternative approach to check the security of a network. It is sufficient, and for some schemes it is easier to use to check the security of the network. In particular, given any subset of network nodes G, we use $\mathbf{u}_G$ to be the set of secret bits assigned only to all the nodes in set G. As a result, all the secret bits can be divided into distinct groups $\{\mathbf{u}_G\}$ with $G \subseteq \mathcal{N}$. One idea here is that one can further decompose each group of secret bits $\mathbf{u}_G$ into sub-groups $\{\mathbf{u}_{ij}^G\}$ such that the sub-group of secret bits $\mathbf{u}_{ij}^G$ is only used for the communication between node i and j. Then the maximum possible amount of information that can be communicated between node i and j is $\sum |\mathbf{u}_{ij}^G|$, where the summation is over all the node set G that includes i and j but not any hacked node. The following result extends this idea to general schemes. Here, we call $|\mathbf{u}_G|/l$ as the secret-bit rate of a set of nodes G. $\{x_{ij}^G | i,j \in G\}$ is a non-negative decomposition of $|\mathbf{u}_G|/l$ if and only if $x_{ij}^G \geq 0$ and $\sum_{i,j\in G} x_{ij}^G = |\mathbf{u}_G|/l$.

Theorem 7: Given a network of n nodes with each node assigned up to l random bits, the channel rates $\{r_{ij}\}$ are achievable if when $l \to \infty$, there exists non-negative decompositions $\{x_{ij}^G | i,j \in G\}$ of $\frac{|\mathbf{u}_G|}{l}$ for all $G \subseteq \mathcal{N}$ and they satisfy

$$\sum_{G|i,j\in G\subseteq\mathcal{N}_\text{s}} x_{ij}^G > r_{ij}, \quad \forall \mathcal{N}_\text{s} \text{ and } (i,j) \in \mathcal{P}_\text{s} \text{ with } r_{ij} > 0. \tag{8}$$

Example 3: We continue to apply the criteria to the network of 4 nodes in Example 1. There are 4 groups of secret bits, denoted by $\mathbf{u}^{(123)}$, $\mathbf{u}^{(124)}$, $\mathbf{u}^{(134)}$, $\mathbf{u}^{(234)}$, each of size $l/3$. Their corresponding non-negative decompositions satisfy

$$x_{12}^{(123)} + x_{13}^{(123)} + x_{23}^{(123)} = 1/3,$$
$$x_{12}^{(124)} + x_{14}^{(124)} + x_{24}^{(124)} = 1/3, \ldots$$

The network is secure if and only if for any $\mathcal{N}_\text{h}$ the network is secure. Without loss of generality, we assume that node 4 is hacked. In this case, only $\mathbf{u}^{(123)}$ is not hacked. According to Theorem 7, the network is secure for sufficiently large l if the decompositions exist such that

$$x_{12}^{(123)} > r_{12} \quad \text{if } r_{12} > 0,$$
$$x_{13}^{(123)} > r_{13} \quad \text{if } r_{13} > 0,$$
$$x_{23}^{(123)} > r_{23} \quad \text{if } r_{23} > 0.$$

It is equivalent to $r_{12} + r_{13} + r_{23} < \frac{1}{3}$. This should hold for any node permutation, as we don't know which node is actually hacked, reaching the same condition as (5).

We hope to develop schemes that can securely communicate as many message bits as possible not only over the entire network but also through a single channel. As defined earlier, a secure network-communication scheme consists of a key pre-distribution phase and a communication phase. In this key pre-distribution phase, the secret bits of the network nodes are assigned from a pool of independent and unbiased truly random bits, and we use $\mathbf{u}_{ij}$ to denote the sequence of common secret bits between node i and j. For each node i, it needs to identify $\mathbf{u}_{ij}$ for all $j \neq i$. In the communication phase, whenever node i needs to communicate with node j, it establishes a secret key $\mathbf{s}_{ij}$ from their common secret bits $\mathbf{u}_{ij}$ via privacy amplification, and then use the key $\mathbf{s}_{ij}$ for the one-time pad encryption to realize secure communication. Fig. 1 depicts a brief and complete diagram for secure network communication, with a short description for each of the steps. In the next two sections, we will study and analyze key distribution methods and privacy amplification methods, respectively, in detail.

IV. Combinational Key Distribution

This section focuses on the study of key distribution, i.e., how to assign truly random bits to the network nodes and how to let a node i find the common secret bits $\mathbf{u}_{ij}$ shared with another node j for any $j \neq i$.

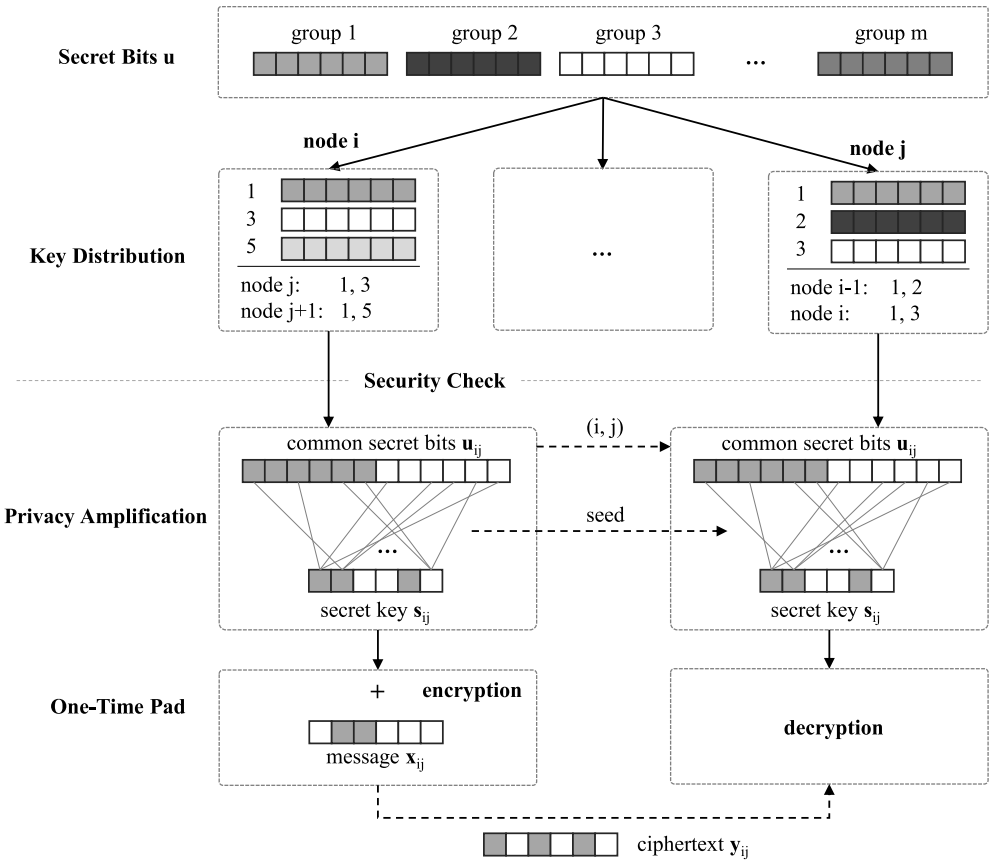


Fig. 1. A diagram for secure network communication. (a) With the key distribution, each node receives multiple groups of secret bits and the indices of the common groups shared with each other node. (b) Before sending a message or periodically, the sender checks whether the current channel can support more communication with information-theoretic security. (c) If node i needs to send a message $\mathbf{x}_{ij}$ to node j, node i determines their common secret bits $\mathbf{u}_{ij}$ and from which it generates a secret key $\mathbf{s}_{ij}$, with each bit generated by computing the XOR of d uniformly sampled common secret bits. (d) With the generated secret key $\mathbf{s}_{ij}$, node i encrypts the message $\mathbf{x}_{ij}$ using the one-time pad scheme, and sends the resulting ciphertext $\mathbf{y}_{ij}$ as well as the node indices and the pseudo-random seed for secret-bit sampling to the receiver. (e) The receiver finds the common secret bits $\mathbf{u}_{ij}$ based on the received i, and from which it further reconstructs the secret key $\mathbf{s}_{ij}$ based on the pseudo-random seed. (f) Finally, the receivers get the decrypted $\mathbf{x}_{ij}$ from the ciphertext $\mathbf{y}_{ij}$ and the secret key $\mathbf{s}_{ij}$.

A. Combinational Key Distribution

Key pre-distribution was explored in sensor networks with computational security [22], [23]. In contrast, we study key pre-distribution for the information theoretic security, where the way of distributing secret bits directly affects the communication rates. The idea of combinational key distribution is to assign the same number of distinct secret bits to each combination of a nodes with $a \geq 2$. Given a network of n nodes, there are totally $\binom{n}{a}$ combinations of a nodes. Hence, we divide all the secret bits into $\binom{n}{a}$ groups with each of size $l/\binom{n-1}{a-1}$, and assign the secret bits of each group to a unique combination of a nodes. For every two nodes, their shared common secret bits consist of the secret bits from $\binom{n-2}{a-2}$ groups. In Theorem 2, it has been shown that this method with an appropriate parameter a combined with asymptotically-optimal privacy amplification can achieve the channel capacity of a network for sufficiently large l.

Example 4: Let's consider a network of size 4 and $a = 3$. The combination key distribution assigns secret bits to 4 combinations of network nodes, corresponding to nodes $(1, 2, 3)$, $(1, 2, 4)$, $(1, 3, 4)$, $(2, 3, 4)$, respectively. As a result, each node obtains 3 groups of secret bits, with each group of size $l/3$.

The combinational key distribution described above becomes less practical when both n and a are large, as there are too many combinations of a nodes. To reduce the complexity, we would like to consider only a fraction of the combinations, saying m combinations of a nodes. Then we divide all the secret bits into m groups of the same size, and assign them to the m combinations of nodes respectively. Finding the m combinations is equivalent to constructing a binary $m \times n$ matrix, with each row including a ones that represent a combination. We call this matrix as the assignment matrix. This assignment matrix is publicly known by all the network nodes, and helps to identify the sequence of shared common secret bits between any two nodes. It is expected that each column of the assignment matrix has roughly the same number of ones, so that each node belongs to almost an equal number of combinations. So does the intersection (common ones) of any two columns, for any two nodes having enough common secret bits.

Example 5: We continue using the network of size 4 and $a = 3$ as an example. Assuming that the secret bits are equally

distributed to 3 combinations, $(1,2,3)$, $(1,2,4)$ and $(1,3,4)$. These node combinations can be represented by an assignment matrix

$$M = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}.$$

The 2nd column of the matrix has 2 ones, meaning that node 2 belongs to two of the combinations and hence receives two groups of secret bits.

B. Random Construction of Assignment Matrix

A simple construction of the assignment matrix is to construct a random matrix in which each row has exactly a ones and each column has exactly $b = \frac{am}{n}$ ones (taking m as a multiple of n). Such a construction has been studied for the parity-check matrix of regular LDPC codes [21], see Example 6. Based on this assignment matrix, the pool of secret bits are equally divided and distributed to m combinations of a nodes, with each node receiving b groups of secret bits. For each pair of nodes i and j, the expected number of their shared common groups is about

$$m_{ij} = b \cdot \frac{a-1}{n-1} = \frac{a(a-1)}{n(n-1)}m,$$

where $\frac{a-1}{n-1}$ is the probability that node j belongs to a random combination of a nodes that includes node i. Following the law of large numbers, one can choose m sufficiently large so that every pair of nodes can receive roughly the same amount of common secret bits. As m becomes larger, the performance of this method converges to that of the original combinational key distribution with $\binom{n}{a}$ groups when l is infinitely large.

Example 6: This example demonstrates how to construct a 6×10 random matrix with each row containing exactly 5 ones and each column containing exactly 3 ones. First, let the 1st row be $(1,1,1,1,1,0,0,0,0,0)$ and the 2nd row be $(0,0,0,0,0,1,1,1,1,1)$, and let π_1, π_2 be random permutations on $\{0,1\}^{10}$. Applying π_1 on the 1st and 2nd rows yields the 3rd and 4th rows, and applying π_2 on the 1st and 2nd rows yields the 5th and 6th rows. It guarantees that the resulting matrix contains exactly 5 ones in each row and 3 ones in each column, for instance,

$$M = \left(\begin{array}{cccccccccc} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ \hdashline 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ \hdashline 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \end{array}\right) \tag{9}$$

The proposed method can run on the network nodes with both low computational complexity and low space complexity. For each node, it only needs to store b rows of the assignment matrix for identifying the combinations of nodes that it belongs to, and the row indices of the b rows in the assignment matrix. Hence, the total space overhead of the method in each node is about $b(n + \log_2\ m)$, where $\log_2\ m$ is for storing the index of a row. When node i needs to establish a secret key with another node j, it can identify the groups of the common secret bits by finding the jth column of the stored b rows and mapping its ones to row indices. The running time for the two nodes to find their common groups of secret bits is $O(b)$.

Example 7: This example demonstrates the method for a node to find the groups of common secret bits shared with another node, considering the assignment matrix in (9) and node 3. Node 3 only needs to store 3 rows

$$\left(\begin{array}{cc|c|ccccccc} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{array}\right)$$

and their indices $(1,4,5)$. When node 3 needs to find the groups of common secret bits shared with node 4, it first finds the 4th column of the stored rows, which is $(1,1,0)^T$, and then based on the locations of the ones it finds the indices of the shared groups, i.e., $(1,4)$.

C. Maximum Rates

For the original combinational key distribution with $a \geq 2$, each node is assigned $\binom{n-1}{a-1}$ groups of secret bits, and every two nodes share $\binom{n-2}{a-2}$ groups of secret bits. From Corollary 6, the maximum network rate and the maximum channel rate of the combinational key distribution combined with asymptotically-optimal privacy amplification can be derived.

Corollary 8: Given a network of n nodes with at most $t \leq n-2$ nodes being hacked, let ψ be a secure communication scheme that uses the combinational key distribution with $a \geq 2$ and an asymptotically-optimal privacy amplification method. The maximum network rate of the scheme is

$$R_{\text{network}}(n,t,\psi_{\text{comb}}) = \frac{n}{a}\frac{\binom{n-t-2}{a-2}}{\binom{n-2}{a-2}}, \tag{10}$$

and the maximum channel rate of the scheme is

$$R_{\text{channel}}(n,t,\psi_{\text{comb}}) = \frac{a-1}{n-1}\frac{\binom{n-t-2}{a-2}}{\binom{n-2}{a-2}}. \tag{11}$$

The proof is given Section VIII-G. For the maximum rates of the scheme based on the combinational key distribution, they have a common term $\gamma(t,a) = \binom{n-t-2}{a-2}/\binom{n-2}{a-2}$ which is a decreasing function of t with $\gamma(0,a) = 1$. This term captures the effect of the number of hacked nodes t on the maximum rates of the scheme. From this term, we can estimate the number of hacked nodes that the scheme can tolerate. For instance, when $a = 3$, $\gamma(t,a) = \frac{n-t-2}{n-2}$, and one can tolerate relatively large t. When a is large, the scheme can only tolerate a very small number of nodes to be hacked. Besides the common term, the maximum network rate is proportional to $\frac{n}{a}$ and the maximum channel rate is proportional to $\frac{a-1}{n-1}$. The intuition behind this is that, given a random set of a nodes that includes node i, the probability for this set including another specific node j is $\frac{a-1}{n-1}$. As the number of nodes a that each secret bit distributed to increases, the number of common secret bits shared between any two nodes (corresponding to the maximum channel rate) increases by a factor of $a-1$. Meanwhile, the usage efficiency of each secret bit (corresponding to the maximum network rate) is reduced by a factor of roughly $\frac{2}{a}$, as each secret bit stored with a copies can only

be used to securely transmit at most 1-bit information all over the network.

D. Irregular Key Distribution

We call the combinational key distribution described above the *regular* combinational key distribution, where each secret bit is distributed to exactly a network nodes. The multiplication of its maximum network rate and its maximum channel rate does not exceed 1, implying that with a regular combinational key distribution, high network rate and high channel rate cannot be achieved at the same time. Generally, with a smaller a, the regular combinational key distribution can have a higher network rate and tolerate more nodes being hacked, and with a bigger a, the regular combinational key distribution can have a higher channel rate. We denote the regular combinational key distribution with $a = 2$ as the pairwise key distribution, which assigns each pair of nodes the same number of distinct secret bits.

To contrast to the regular key distributions, we introduce the *irregular* key distributions, where each node uses a fraction $\lambda_a \in [0,1]$ of its storage space to run the regular combinational key distribution with a such that $\sum_a \lambda_a = 1$. If the channel rates $\{r_{ij}^{(a)}\}$ is achievable by the regular combinational key distribution with a, then for the same n and t, the channel rates $\{r_{ij}\}$ with $r_{ij} = \sum_a \lambda_a r_{ij}^{(a)}$ is achievable by the irregular combinational key distribution with λ_a for $a \geq 2$.

Example 8: This example considers a simple irregular combinational key distribution, in which each node uses $\frac{1}{2}$ of its storage space to run the pairwise key distribution and the rest to run the combinational key distribution with $a > 2$. The maximum network rate of this irregular key distribution (combined with asymptotically-optimal privacy amplification) is the weighted sum of their respective component schemes' maximum network rates, and so is its maximum channel rate. The maximum rates of the scheme based on the irregular key distribution are

$$R_{\text{net}}(n,t,\psi_{\text{example}}) = \frac{n}{4} + \frac{n}{2a}\frac{\binom{n-t-2}{a-2}}{\binom{n-2}{a-2}},$$

$$R_{\text{channel}}(n,t,\psi_{\text{example}}) = \frac{1}{2(n-1)} + \frac{a-1}{2(n-1)}\frac{\binom{n-t-2}{a-2}}{\binom{n-2}{a-2}}.$$

The maximum network rate is strictly larger than $n/4$, and the maximum channel rate can be adjusted by selecting appropriate a.

Fig. 2 compares the maximum network rates and the maximum channel rates of the regular and irregular combinational key distributions combined with asymptotically-optimal privacy amplification in the above example, when the network size $n = 100$ and the maximum number of hacked nodes $t = 1$ or 2. For a network with $n = 100$ and $t = 1$, the maximum network and channel rates of the combinational key distribution with $a = 2$ are 50.0 and 0.0101, respectively, and those of the irregular key distribution with $a = 25$ are 26.53 and 0.0978, respectively, which improves on the maximum channel rate by sacrificing on the maximum network rate. It is worth mentioning that the maximum network rate and the maximum channel rate are only two important metrics characterizing the security region of a scheme. They are not sufficient to guarantee the security of the network. When running a scheme in a network with l sufficiently large, the network needs to keep monitoring the channel rates of all the channels and using Theorem 4 or Theorem 7 to guarantee the information-theoretic security.

V. Secure Communication With Privacy Amplification

This section discusses how secure network communication can be achieved with privacy amplification and end-to-end one-time pad encryption.

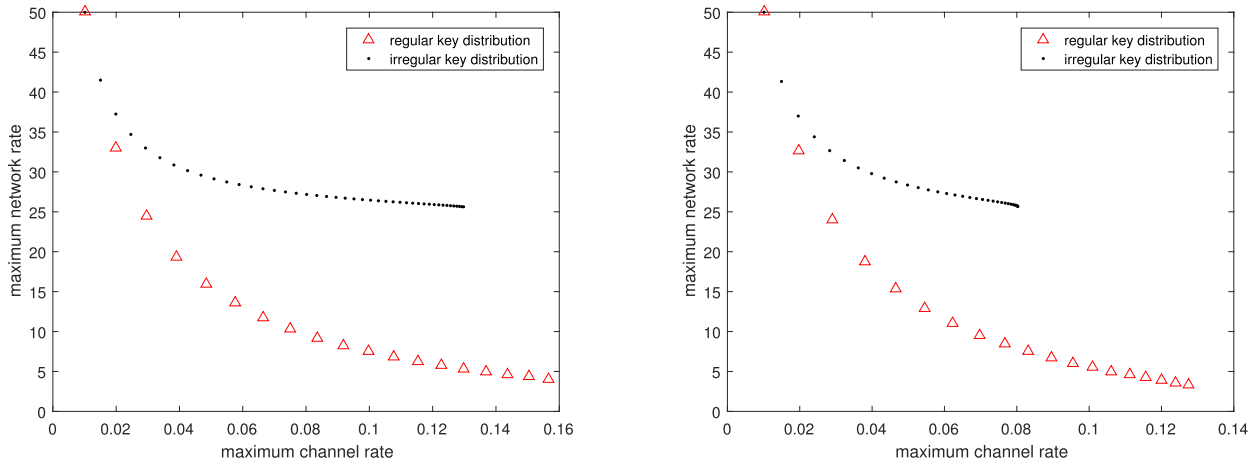
A. Secure Network Communication

Before two nodes i and j start to communicate, it is assumed that they share a sequence of common secret bits $\mathbf{u}_{ij}$, which can be quickly identified by both of the parties (this has been realized in key pre-distribution phase). However, this common sequence $\mathbf{u}_{ij}$ is not perfectly secure to be directly used as secret keys, because some of the bits might be hacked and some of the bits might have been used before to generate other secret keys. Privacy amplification is a key technique here that distils a secure secret key from a common shared random sequence that is partially known by eavesdroppers. There are a variety of methods for privacy amplification, such as universal hashing [18], random linear transformations [19] and polar codes [20]. For certain types of random sources, such as i.i.d. sources, privacy amplification can extract a secret key of length up to $H(\mathbf{x}|\mathbf{z})$, with $\mathbf{x}$ the random source and $\mathbf{z}$ the set of information possibly known by eavesdroppers.

When node i needs to send a message $\mathbf{x}_{ij}$ to node j, it first generates a secret key $\mathbf{s}_{ij}$ of length $|\mathbf{s}_{ij}| = |\mathbf{x}_{ij}|$ from $\mathbf{u}_{ij}$ with privacy amplification, and then encrypts the message $\mathbf{x}_{ij}$ with the one-time pad encryption and sends it to node j. It is desired that given all information possibly known by eavesdroppers, the generated secret key $\mathbf{s}_{ij}$ is very close to the uniform distribution. One problem arises when node i sends a series messages to node j on demand. Given a large number of common secret bits $\mathbf{u}_{ij}$, one straightforward idea is to divide the shared secret bits into blocks, and then apply privacy amplification to each block. However, this approach is not appropriate for our applications as it requires knowledge of the total message length (channel rate) before communication as well as sophisticated coordination among the nodes. More importantly, it results in unbalanced utilization of the secret bits, which may introduce some block-necks of the network security.

This motivates us to directly apply privacy amplification on the whole common sequence $\mathbf{u}_{ij}$, instead of on the blocks of $\mathbf{u}_{ij}$. We are particularly interested in the method of random linear transformations as described in Corollary 5 due to its asymptotic optimality. With this method, the generated secret key is

$$\mathbf{s}_{ij} = M_{ij}\mathbf{u}_{ij}$$



(a) n=100, t=1 (b) n=100, t=2

Fig. 2. The maximum network rate vs. the maximum channel rate for the regular and irregular combinational key distributions with different a. Here, an irregular method uses $\frac{1}{2}$ space to run the pairwise key distribution and the rest to run the regular combinational key distribution with $a > 2$.

with a random matrix M_{ij} in which each entry is one with probability $\frac{d}{|\mathbf{u}_{ij}|}$ independently. If $d = O(\log l)$ and l is sufficiently large, it is asymptotically optimal. Besides of its optimality, this method is naturally good for communication on demand. Assuming that node i needs to send a series of messages $\mathbf{x}_{ij}^1, \mathbf{x}_{ij}^2, \ldots$ to node j. Whenever a new message $\mathbf{x}_{ij}^k$ needs to be transmitted, node i creates a corresponding new random matrix M_{ij}^k for privacy amplification that yields a secret key $\mathbf{s}_{ij}^k$. Then it sends both the encrypted message $\mathbf{x}_{ij}^k + \mathbf{s}_{ij}^k$ and the random matrix M_{ij}^k to the receiver. The receiver can recover the secret key $\mathbf{s}_{ij}^k$ from both $\mathbf{u}_{ij}$ and M_{ij}^k, based on which the message can be decrypted.

For the privacy-amplification method based on random linear transformations, the random matrices are assumed known by public. These random matrices can be generated with pseudo random numbers instead of truly randomness, hence only their pseudo-random seeds need to be transmitted, not the entire matrices. This saves a lot of communication cost. In addition, because of the sparsity of these random matrices, the computational complexity of privacy amplification can be reduced by transforming matrix multiplication into secret bit sampling.

B. Privacy Amplification With Finite d

The method of random linear transforms have been proved asymptotically optimal for privacy amplification when $d = O(\log l)$ and l is sufficiently large. We further investigate how the selection of d affects its performance when d is finite. Our study is based on a simplified example only for the purpose of gaining some insight. Specifically, let $\mathbf{u} \in \{0,1\}^r$ be a sequence of independent and unbiased truly random bits, where δ random bits are known by eavesdroppers (we don't know which δ bits). A secret key $\mathbf{s} \in \{0,1\}^k$ is generated from $\mathbf{u}$ by privacy amplification, and we call the probability that $\mathbf{s}$ is not truly uniform as the failure probability.

First of all, the limit of the secret-key length k is $H(\mathbf{u}|\mathbf{z}) = r - \delta$, where $\mathbf{z}$ represents all the information known by eavesdropper. If $k > r - \delta$, the failure probability must be 1. Then, we consider that $\mathbf{s} \in \{0,1\}^k$ is generated by multiplying an $k \times r$ random matrix of density d/r on $\mathbf{u}$. Its failure probability is

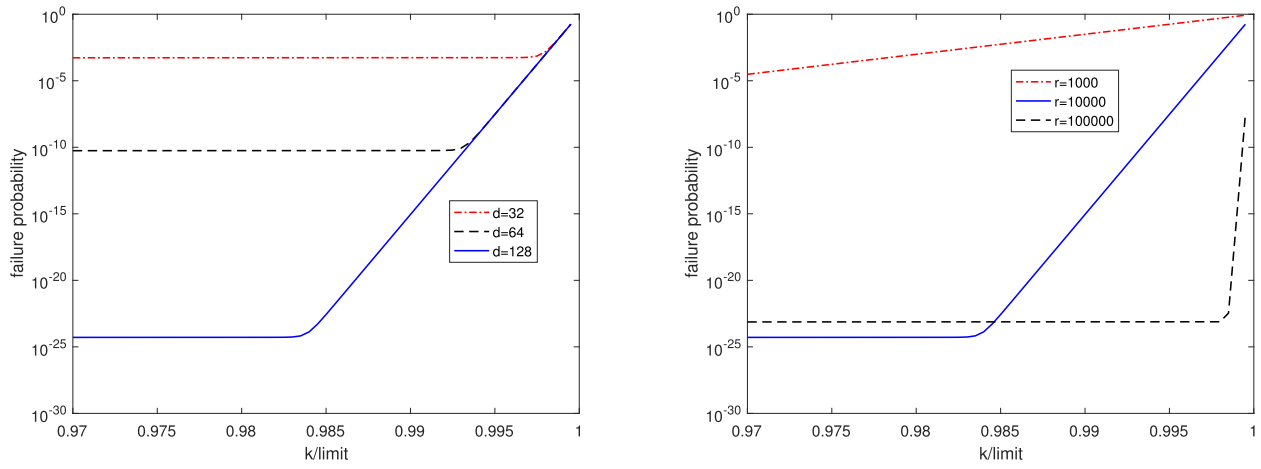
$$\mathbb{P}_1 \leq \sum_{j=1}^{k} \binom{k}{j} \left(\frac{1}{2} + \frac{1}{2}\left(1 - \frac{2d}{r}\right)^j\right)^{r-\delta}, \tag{12}$$

following (44) in the proof of Lemma 15.

Fig. 3 shows the upper bound on the probability $\mathbb{P}_1$ under different parameters n, k, d, where $\delta = r/2$ and the limit for k is $r/2$. It can be seen that given $r = 10000$, the failure probability is non-decreasing function of k. As k decreases, there is a turning point depending on d such that the probability quickly becomes saturated. On the other hand, fixing $d = 128$, the bigger r is, the smaller the probability $\mathbb{P}_1$ is. It is illustrated that when $d = 128$ and r is large, the performance of the method is very close to the limit.

C. Privacy Amplification by Sampling

We introduce a variant of the above proposed privacy-amplification method: each secret-key bit is generated by computing the XOR of d randomly sampled common secret bits from $\mathbf{u}_{ij}$ with d a large integer, for example, 128. Each common secret bit is sampled based on a uniform distribution. We can repeat this process whenever more secret-key bits are needed. The computational complexity of the privacy-amplification method to generate k secret-key bits is $O(kd)$. To reduce the times of random memory or disk accesses, one can generate $q \gg 1$ secret-key bits simultaneously by packing q secret bits together at the same location and performing the same operations on them. While the sampling process is based on pseudo random numbers, the pseudo-random seed along with the encrypted message is sent to the receiver for reconstructing the secret key.



(a) r=10000 (b) d=128

Fig. 3. The upper bound of the failure probability for privacy amplification by multiplying a random matrix of probability d/r, where the input is a sequence of r truly random bits with $r/2$ of them possibly known by eavesdroppers, and k is the output length whose limit is $r/2$.

This privacy-amplification method can also be described as a random linear transformation on $\mathbf{u}_{ij}$ such that the secret key

$$\mathbf{s}_{ij} = M_{ij}\mathbf{u}_{ij}$$

where M_{ij} is a random matrix with each row containing exactly d ones. The difference between this method and the former one is that, in the former method the entries in M_{ij} are i.i.d. and there are approximately d ones in each row when d is large (according to the law of large numbers). We call the modified method as d-sampling and the former one as i.i.d. sampling with probability d/r.

For d-sampling, the dependency of the entries in M_{ij} makes it difficult to analyze the performance. But we suspect that the d-sampling performs slightly better than the i.i.d. sampling with the same d when $|\mathbf{u}_{ij}|$ is finite. Our intuition follows that the i.i.d. sampling is actually equivalent to computing the XOR of x sampled common secret bits with x a random variable of Poisson distribution with expected value d. It seems that, when $d \ll r/2$, the smaller x is, the more chance that all the x secret bits get hacked or their XOR collides with other generated secret-key bits. Therefore, limiting the number of sampled secret bit to exact d could likely improve the performance of privacy amplification.

We continue investigate the performance of d-sampling on the above simplified example.

Lemma 9: Let $\mathbf{u} \in \{0,1\}^r$ be a sequence of independent and unbiased truly random bits, with δ random bits known by eavesdroppers. Let $\mathbf{s} \in \{0,1\}^k$ be a secret key generated from $\mathbf{u}$ such that $\mathbf{s} = \mathbf{M}\mathbf{u}$ with a binary random matrix M that contains exactly d ones in each row. The probability for $\mathbf{s}$ being not uniform (not truly random) is

$$\mathbb{P}_2 \le \sum_{j=1}^{k} \binom{k}{j} \sum_{w=0}^{r} P_j(w) \binom{\delta}{w} / \binom{r}{w}. \tag{13}$$

Here, $P_j(w)$ is the probability that the sum of given j rows in M has exactly w ones, and $\binom{\delta}{w}/\binom{r}{w}$ is the probability that all the corresponding w random bits are known by eavesdroppers.

The proof follows a similar idea as proving Lemma 15. The secret key $\mathbf{s}$ is uniform if and only if all the rows of M are linearly independent if only considering the columns of unhacked random bits. It means that the sum of any rows is a non-zero vector on those columns. Specifically, there are $\binom{k}{j}$ subsets of j rows, the probability for the sum of the j rows having w ones is $P_j(w)$, and $\binom{\delta}{w}/\binom{r}{w}$ is the probability that all the w ones are in the columns of hacked bits. This leads to the above result.

The probability $P_j(w)$ can be derived based on an iterative relation

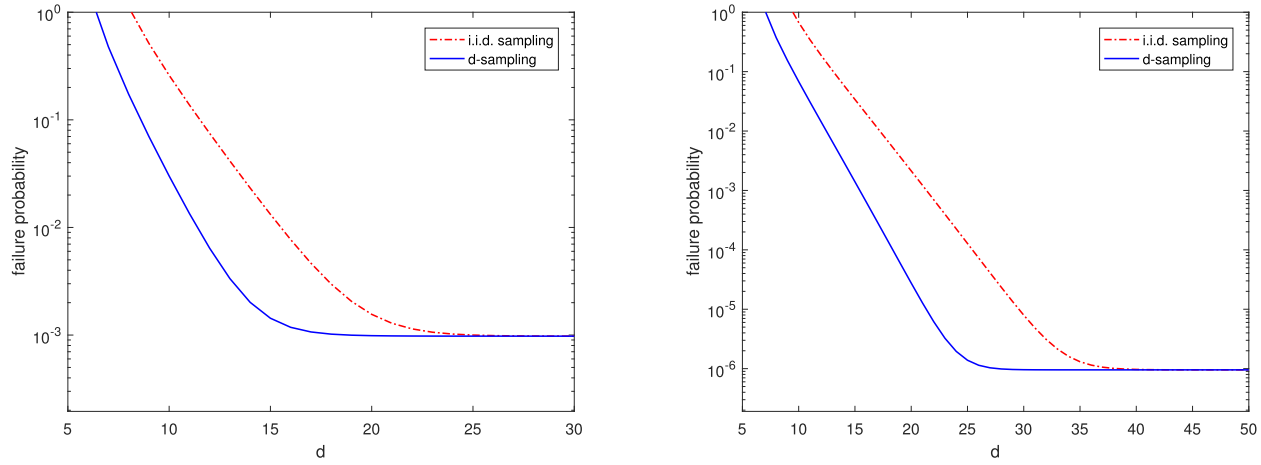
$$P_j(w) = \sum_{i=0}^{d} P_{j-1}(w+2i-d) \times \binom{w+2i-d}{i}\binom{r-w-2i+d}{d-i} / \binom{r}{d} \tag{14}$$

and it satisfies $P_1(d) = 1$, $P_0(w \neq d) = 0$, $P_j(w < 0) = 0$ and $P_j(w > r) = 0$. This relation comes from the fact that a vector of length r and weight w can be generated by adding a vector of weight d to a vector of weight $w + 2i - d$ if the two vectors has i overlapped ones. Here, $\binom{w+2i-d}{i}\binom{r-w-2i+d}{d-i}/\binom{r}{d}$ is the probability of having such a random vector of weight d, given any vector of weight $w + 2i - d$.

Based on (13), we can numerically calculate the upper bound on the failure probability of the d-sampling, and compare it with that of the i.i.d. sampling given by (12). Fig. 4 conducts two groups of experiments for $r = 100, k = 40$ and $r = 200, k = 80$ respectively and depicts the comparison of the two methods with the same d. In the experiments under the same r and k, the performance of the d-sampling surpasses that of the i.i.d. sampling when d is relatively small. As d becomes larger, the performances of the two methods converge.

VI. Simplifying Security Criteria

To guarantee the information-theoretic security of the network, it is desired that each node can monitor its neighboring channel rates and broadcasts their status to all the other



(a) r=100, k=40 (b) d=200, k=80

Fig. 4. The upper bound of the failure probability for i.i.d. sampling with probability d/r and d-sampling, where the input is a sequence of r random bits with $r/2$ of them possibly known by eavesdroppers, and k is the output length whose limit is $r/2$.

nodes periodically, e.g., through flooding [27]. So when two nodes communicate with each other, they can first check whether this communication violates the requirements of the network's information-theoretic security. However, when the network size is large, given a scheme and the channel rates, it is computationally too complex to check the security of a network directly based on Theorem 4 and Theorem 7, as the number of constraints in the criteria becomes prohibitively large even when $n = 10$. We discuss some techniques to reduce the number of constraints by relaxing or tightening the criteria, and their applications to the proposed scheme. Here, we assume that the underlying privacy-amplification method used by the schemes is asymptotically optimal.

A. Relaxing Theorem 4

As there are too many constraints in the criteria of Theorem 4, we relax the criteria by keeping only an important subset of the constraints. One way is to only consider those set of channels P that form a clique, namely, $P = \{(i,j)|i,j \in N\}$ for a set of unhacked nodes $N \subseteq \mathcal{N}_\mathrm{s}$. Then, the conditions become $\forall \mathcal{N}_\mathrm{h}$ and $N \subseteq \mathcal{N}_\mathrm{s}$,

$$\sum_{i,j\in N} r_{ij} < \frac{|\cup_{i,j\in N} \mathbf{u}_{ij}/\mathbf{u}_\mathrm{h}|}{l} \text{ or } \sum_{i,j\in N} r_{ij} = 0. \tag{15}$$

This relaxation makes the criteria not sufficient to guarantee the information-theoretic security, but it helps to approximate the boundary of the network's security region. To further reduce the number of conditions, we tighten (15) such that for every clique size w with $2 \le w \le n-t$,

$$\max_{N:|N|=w} \sum_{i,j\in N} r_{ij} < \min_{\mathcal{N}_\mathrm{h},N:|N|=w,N\subseteq\mathcal{N}_\mathrm{s}} \frac{|\cup_{i,j\in N} \mathbf{u}_{ij}/\mathbf{u}_\mathrm{h}|}{l}. \tag{16}$$

Then the number of conditions become $n-t-1$. For a network with the combinational key distribution, the right terms of the conditions can be computed explicitly based on the following result.

Lemma 10: Given a network of n nodes with at most $t \le n-2$ nodes being hacked, where each node is assigned up to l random bits by the combinational key distribution with $a \ge 2$, it has

$$\min_{\mathcal{N}_\mathrm{h},N:|N|=w,N\subseteq\mathcal{N}_\mathrm{s}} \frac{|\cup_{i,j\in N} \mathbf{u}_{ij}/\mathbf{u}_\mathrm{h}|}{l} = \frac{\binom{n-t}{a} - \binom{n-t-w}{a} - w\binom{n-t-w}{a-1}}{\binom{n-1}{a-1}}.$$

This result is due to the symmetry of the combinational key distribution (permutating the indices of the network nodes does not change the joint probability distribution of the distributed secret bits). Without loss of generality, we can assume that the node set N consists of the first w nodes, and the last t nodes are hacked. In this result, $\binom{n-1}{a-1}$ is the number of groups (of secret bits) assigned to each node, $\binom{n-t}{a}$ is the total number of unhacked groups, and $\binom{n-t-w}{a} + w\binom{n-t-w}{a-1}$ is the number of unhacked groups that are owned by at most one node in N.

Example 9: Let us continue considering the network of 4 nodes as an example. Assume that the combinational key distribution with $a = 3$ is applied, then the simplified conditions become

$$\max_{i,j} r_{ij} < \frac{1}{3}, \quad \max_{N:|N|=3} \sum_{i,j\in N} r_{ij} < \frac{1}{3},$$

i.e., within any clique of size 3 the total channel rate should be less than $1/3$. This is consistent with (5) that is derived from the criteria of Theorem 4.

B. Tightening Theorem 7

For the criteria of Theorem 7, instead of determining whether there exists feasible decompositions of $\frac{|\mathbf{u}_G|}{l}$ for all

$G \subseteq \mathcal{N}$, it is much easier to check whether given decompositions are feasible or not. Specifically, we can construct non-negative decompositions $\{x_{ij}^G\}$ of $\frac{|\mathbf{u}_G|}{l}$ such that

$$x_{ij}^G = \frac{r_{ij}}{\sum_{i',j' \in G} r_{i'j'}} \frac{|\mathbf{u}_G|}{l}. \tag{17}$$

If (8) of Theorem 7 holds with these $\{x_{ij}^G\}$, it is sufficient to guarantee the security of the network for sufficiently large l.

Example 10: For the combinational key distribution, $\frac{|\mathbf{u}_G|}{l}$ is a constant. If given a network of size 4 with at most 1 node being hacked, the combinational key distribution with $a = 3$ leads to $\frac{|\mathbf{u}_G|}{l} = \frac{1}{3}$. In this case, $\frac{|\mathbf{u}^{(123)}|}{l}$ is decomposed as

$$x_{12}^{(123)} = \frac{r_{12}}{r_{12}+r_{13}+r_{23}} \cdot 1/3,$$
$$x_{13}^{(123)} = \frac{r_{13}}{r_{12}+r_{13}+r_{23}} \cdot 1/3,$$
$$x_{23}^{(123)} = \frac{r_{23}}{r_{12}+r_{13}+r_{23}} \cdot 1/3.$$

So is $\frac{|\mathbf{u}^{(124)}|}{l}$, and so on. Without loss of generality, we assume that node 4 is hacked. To let (8) hold, it requires $x_{12}^{(123)} > r_{12}$. This leads to

$$r_{12} + r_{13} + r_{23} < 1/3.$$

This should hold for any node permutation, as we don't know which node is actually hacked, reaching the same condition as (5).

VII. Discussions

In this section we provide some further discussions, including the security strength with channel rates near their theoretical limit, the application of network coding, and several open questions.

A. Security Beyond Limit

Can a network continue to communicate when its channel rates reach or even exceed the theoretical limit? Our claim is that when l is very large, e.g. $l > 10^9$ (1GB), the network communication near the theoretical limit is more secure than widely used cryptographic approaches that are based on some unproven assumptions about computational hardness.

For the network communication model that we studied, let $\mathbf{u} \in \{0,1\}^r$ be a sequence of independent and unbiased random bits assigned to the network nodes, and let $\mathbf{s} \in \{0,1\}^k$ be the concatenation of the secret keys that generated all over the network. Then $\mathbf{s}$ can be represented as a linear transformation of $\mathbf{u}$, i.e., $\mathbf{s} = M\mathbf{u}$ with a $k \times r$ partially-random matrix M. Hence, the concatenation of the ciphertexts generated using the one-time pad encryption is

$$\mathbf{y} = \mathbf{s} + \mathbf{x} = M\mathbf{u} + \mathbf{x} \tag{18}$$

with $\mathbf{x} \in \{0,1\}^k$ the concatenation of the transmitted messages. According to our network model, the matrix M, the ciphertexts $\mathbf{y}$ and some secret bits $\mathbf{u}_\mathrm{h}$ in $\mathbf{u}$ are possibly known by eavesdroppers. Let $\mathbf{u}_\mathrm{s}$ be the set of unhacked random bits in $\mathbf{u}$, i.e., $\mathbf{u}_\mathrm{s} = \mathbf{u}/\mathbf{u}_\mathrm{h}$. Attacking the system is to derive some information about $\mathbf{u}_\mathrm{s}$ or $\mathbf{x}$ from M, $\mathbf{y}$ and $\mathbf{u}_\mathrm{h}$. By moving the part corresponding to $\mathbf{u}_\mathrm{h}$ to the left, Equation (18) can be simplified as

$$\mathbf{y}' = M'\mathbf{u}_\mathrm{s} + \mathbf{x} \tag{19}$$

with M' and $\mathbf{y}'$ known by eavesdroppers.

When $|\mathbf{y}'| > |\mathbf{u}_\mathrm{s}|$, the network is not information-theoretically secure, but it is still extremely difficult to derive some information about $\mathbf{u}_\mathbf{s}$ or $\mathbf{x}$ if $|\mathbf{u}_\mathrm{s}|$ is very large and $|\mathbf{y}'| < 2|\mathbf{u}_\mathrm{s}|$. Firstly, this attacking process is analog to the decoding of a linear random code, with $\mathbf{u}_\mathrm{s}$ being the message and $M'\mathbf{u}_\mathrm{s}$ as the codeword. It has been proven that with a general M', finding the $\mathbf{x}$ with the minimum Hamming weight is NP-complete [25].

Secondly, there are some uncertainties in the messages $\mathbf{x}$ especially when the message is compressed. Even if an eavesdropper is possible to search all the possibilities of $\mathbf{u}_\mathrm{s}$ with unlimited computing power, given M' and $\mathbf{y}'$, there are about $2^{O(H(\mathbf{x})+|\mathbf{u}_\mathrm{s}|-|\mathbf{y}'|)}$ feasible choices for $\mathbf{x}$. When $H(\mathbf{x}) + |\mathbf{u}_\mathrm{s}| - |\mathbf{y}'| \gg 1$, it is difficult for an eavesdropper to choose the right one for $\mathbf{x}$.

Thirdly, as in our proposed schemes, the secret bits are typically shared by multiple network nodes, and each secret key is generated by jointly utilizing all the common secret bits between the two terminals (not block by block). Attacking the system needs to solve the values of a very large number of secret bits together, which is very difficult in a typical application with each node storing more than gigabytes of secret bits. Even if an attacking algorithm of polynomial computational complexity exists, e.g. $O(r^4)$, it is still practically impossible to break the system.

B. Network Coding

In this paper, we mainly focus on network communications with end-to-end encryption. Another way to realize the information-theoretic security is using network coding. One idea of network coding that can be used here is called "secret sharing" [26]. In order to tolerate t nodes to be hacked, the source node encodes the message into $t+1$ packets such that no eavesdropper can obtain any information about the message unless getting all the $t+1$ packets. For example, let $\mathbf{x} \in \{0,1\}^m$ be the message to communicate, then it is encoded into

$$\mathbf{r}_1, \quad \mathbf{r}_2, \quad \ldots, \quad \mathbf{r}_t, \quad \mathbf{r}_1 + \ldots + \mathbf{r}_t + \mathbf{x}$$

with the random-bit sequence $\mathbf{r}_i \in \{0,1\}^m$ as the ith packet for $1 \le i \le t$. Then the source node sends the $t+1$ packets over node-disjoint paths (only the channels whose rates are below the limit are used) to the destination. Each path has at least one relay node that decrypts and re-encrypts the data. After receiving all the packets, the destination node can decode the original message $\mathbf{x}$. Two nodes can communicate to each other with information-theoretic secrecy if and only if there exits at least $t+1$ node-disjoint paths connecting them.

Our results on the network security with end-to-end encryption can be applied to network communications with network

coding. When sending a message from a source node to a destination node with network coding, the communication can be decomposed as multiple end-to-end encryptions, resulting in the increment of the rates of multiple channels on the paths instead of only the channel between the two terminals. For example, if node 1 sends two packets to node 4 through relay nodes 2 and 3 respectively, it increases all of $r_{12}, r_{14}, r_{13}, r_{34}$ in the same amount. In this case, we can use the criteria developed in this paper to check the security of the network.

This network-coding approach based on multiple paths is very expensive in general. It costs at least $2(t+1)$ times of secret-bit resources (more precisely proportional to the number of channels in the selected $t+1$ node-disjoint paths), and introduces much more communication latency. Furthermore, it may bring in additional adversaries, as some hacked nodes may interrupt the communication by modifying relayed packets or injecting corrupted packets, known as Byzantine adversaries [5], [6]. One possible application scenario of network coding is when there are two nodes having to communicate with information-theoretic security but their channel rate has already reached the limit. Then the two nodes can communicate via network coding by finding $t+1$ node-disjoint paths whose underlying channels have sufficient gaps to their secure communication limits.

C. Further Questions

In this paper, we work on a framework that studies the problem of network communication with the information-theoretic security when each node is allowed to carry some pre-distributed randomness. This work is an extension of the well-known one-time pad scheme from 'links' to 'networks.' There are several questions that haven't been completely answered in this paper, which deserve further studies.

1) The tradeoff between the maximum network rate and the maximum channel rate for a network without any nodes being hacked is given in Theorem 1. A natural question is how to extend it to a network with $t > 1$.
2) The criteria in Theorem 4 are both necessary and sufficient for guaranteeing the information theoretic security. It is also proved that the criteria in Theorem 7 are sufficient, but it is unclear whether they are necessary or not.
3) This paper only considers communication with end-to-end encryption to avoid active attacks. How to compute the theoretical bounds for secure communication with both end-to-end encryption and network coding is an open question.
4) The theoretical bounds and security criteria derived in this paper consider asymptotically large l. It would be useful to study the non-asymptotic bounds and security criteria.
5) Does the d-sampling for privacy amplification perform strictly better than the i.i.d. sampling for finite l and d? This paper only provides some numerical comparisons.
6) How to derive simple criteria that guarantee the network's information-theoretic security and meanwhile they are very easy to verify and very close to the theoretical limits?

This paper mainly focuses on networks with all the network nodes playing the same role, in which every node can carry the same number of secret bits. The models, methods and analysis developed can be naturally applied or extended to some other occasions, such as a clustered network or a centralized network. For example, if a network has a trustable central node with a larger storage space than the other nodes, one may distribute all the secret bits to this central node, with each secret bit also shared by some of the other nodes. This allows the central node to easily communicate with the other nodes and monitor all the messages transmitted over the network.

VIII. Proofs of Main Results

In this section we provide proofs of our main results.

A. Proof of Theorem 1

The network capacity is easy to derive: The total message length communicated with a node cannot exceed l, hence for any i,

$$\sum_{j|i<j\le n} m_{ij} + \sum_{j|1\le j<i} m_{ji} \le l.$$

Summing over all the nodes,

$$\sum_{i|1\le i\le n} \Big(\sum_{j|i<j\le n} m_{ij} + \sum_{j|1\le j<i} m_{ji} \Big) \le nl.$$

This leads to the total message length over the whole network

$$\sum_{i,j|1\le i<j\le n} m_{ij} \le \frac{nl}{2},$$

yielding the upper bound $\frac{n}{2}$ on the network capacity.

This upper bound is achievable using the simple pairwise key distribution when $t \le n-2$: for each pair of nodes, it is distributed $\frac{l}{n-1}$ secret bits, distinct from the other pairs. With this scheme, the common sequence between any two nodes can be used directly as the secret key of the one-time pad encryption, and hence the limiting channel rate is $r_{ij} = \frac{1}{n-1}$ for all (i,j). Its network capacity is

$$r = \sum_{(i,j)\in\mathcal{P}} r_{ij} \le \frac{1}{n-1}\binom{n}{2} = \frac{n}{2},$$

which reaches the upper bound $\frac{n}{2}$.

B. Proof of Theorem 2

We now prove that the channel capacity is at most $\frac{\binom{n-t-2}{a-2}}{\binom{n-1}{a-1}}$ with $a = \lceil \frac{n}{t+1} \rceil$, and it's achievable.

Given the sequence of secret bits stored in node i, $\mathbf{u}_i$, for all $i \in \mathcal{N}$, the entropy of $\mathbf{u}_i$ is at most l. Assume there are t nodes hacked and the hacked secret bits are $\mathbf{u}_{\mathrm{h}}$. The number of message bits that can be securely communicated between node i and j with $i, j \in \mathcal{N}_{\mathrm{s}}$ is upper bounded by

the mutual information between $\mathbf{u}_i$ and $\mathbf{u}_j$ conditioning on $\mathbf{u}_{\mathrm{h}}$ asymptotically. Specifically, for any $(i,j) \in \mathcal{P}_{\mathrm{s}}$, as $l \to \infty$,

$$\sup_{\{r_{ij}\} \in \mathcal{R}(n,t,\psi)} r_{ij} \leq \min_{\mathcal{N}_{\mathrm{h}} | i,j \notin \mathcal{N}_{\mathrm{h}}, |\mathcal{N}_{\mathrm{h}}|=t} \frac{\mathrm{I}[\mathbf{u}_i; \mathbf{u}_j | \mathbf{u}_{\mathrm{h}}]}{l}. \tag{20}$$

This upper bound is based on Theorem 3 in [16]. We briefly describe the result here. In order to get a common sequence from $\mathbf{u}_i$ and $\mathbf{u}_j$, node i and j may need to exchange some messages denoted by $\mathbf{c}$. Then node i computes a secret key $\mathbf{s}_{ij}$ as a function of $\mathbf{u}_i$ and $\mathbf{c}$, and node j computes a secret key $\mathbf{s}'_{ij}$ as a function of $\mathbf{u}_j$ and $\mathbf{c}$. To guarantee the security of the communication between node i and j, as $l \to \infty$, it needs

$$\mathbf{r}_{ij} \leq \frac{H(\mathbf{s}_{ij})}{l}$$

under the conditions that $\mathbf{s}_{ij}$ and $\mathbf{s}'_{ij}$ agree with very high probability and very little information about either $\mathbf{s}_{ij}$ or $\mathbf{s}'_{ij}$ is known by eavesdroppers. It was proved in [16] that

$$H(\mathbf{s}_{ij}) \leq \mathrm{I}[\mathbf{u}_i; \mathbf{u}_j | \mathbf{u}_{\mathrm{h}}] + H(\mathbf{s}_{ij} | \mathbf{s}'_{ij}) + \mathrm{I}[\mathbf{s}_{ij}; \mathbf{c}\mathbf{u}_{\mathrm{h}}],$$

where $H(\mathbf{s}_{ij}|\mathbf{s}'_{ij})$ measures how $\mathbf{s}_{ij}$ agrees with $\mathbf{s}'_{ij}$, and $\mathrm{I}[\mathbf{s}_{ij}; \mathbf{c}\mathbf{u}_{\mathrm{h}}]$ computes the amount of information about $\mathbf{s}_{ij}$ leaked to an eavesdropper. As $l \to \infty$, it requires $\frac{H(\mathbf{s}_{ij}|\mathbf{s}'_{ij})}{l} \to 0$, which ensures that node i and node j can create the same secret key $\mathbf{s}_{ij}$, and $\frac{\mathrm{I}[\mathbf{s}_{ij};\mathbf{c}\mathbf{u}_{\mathrm{h}}]}{l} \to 0$, which guarantees that little information about $\mathbf{s}_{ij}$ is leaked to eavesdroppers. As a result, we can get (20).

According to the definition of the maximum channel rate, we can get that the maximum channel rate C_{channel} satisfies

$$\begin{aligned} C_{\text{channel}} &\leq \min_{i,j} \min_{\mathcal{N}_{\mathrm{h}} | i,j \notin \mathcal{N}_{\mathrm{h}}, |\mathcal{N}_{\mathrm{h}}|=t} \frac{\mathrm{I}[\mathbf{u}_i; \mathbf{u}_j | \mathbf{u}_{\mathrm{h}}]}{l} \\ &\leq \frac{\sum_{i,j,\mathcal{N}_{\mathrm{h}} | i,j \notin \mathcal{N}_{\mathrm{h}}, |\mathcal{N}_{\mathrm{h}}|=t} \mathrm{I}[\mathbf{u}_i; \mathbf{u}_j | \mathbf{u}_{\mathrm{h}}]}{\binom{n}{t}\binom{n-t}{2} l}. \end{aligned} \tag{21}$$

On the other hand, for any node i, $H(\mathbf{u}_i) \leq l$. Hence,

$$\frac{\sum_{i \in \mathcal{N}} H(\mathbf{u}_i)}{l} \leq n. \tag{22}$$

To derive an upper bound on C_{channel} from (21) and (22), the key is to find the connection between $\sum_{i \in \mathcal{N}} H(\mathbf{u}_i)$ and $\sum_{i,j,\mathcal{N}_{\mathrm{h}} | i,j \notin \mathcal{N}_{\mathrm{h}}, |\mathcal{N}_{\mathrm{h}}|=t} \mathrm{I}[\mathbf{u}_i; \mathbf{u}_j | \mathbf{u}_{\mathrm{h}}]$.

We generalize this concept of conditional mutual information to a higher order, and let $I(a_1, a_2, \ldots, a_u | b_1, b_2, \ldots, b_v) \geq 0$ be the amount of mutual information among all the nodes $a_1, a_2, \ldots, a_u$ given the secret bits of the nodes $b_1, b_2, \ldots, b_v$. It is the maximum amount of information shared by $a_1, a_2, \ldots, a_u$ and unknown by $b_1, b_2, \ldots, b_v$. By definition,

$$\begin{aligned} I(a_1 | b_1, b_2, \ldots, b_v) &= H(\mathbf{u}_{a_1} | \mathbf{u}_{b_1} \mathbf{u}_{b_2} \ldots \mathbf{u}_{b_v}), \\ I(a_1, a_2 | b_1, b_2, \ldots, b_v) &= \mathrm{I}[\mathbf{u}_{a_1}; \mathbf{u}_{a_2} | \mathbf{u}_{b_1} \mathbf{u}_{b_2} \ldots \mathbf{u}_{b_v}]. \end{aligned}$$

It has

$$\begin{aligned} & I(a_1, a_2 | b_1, b_2, \ldots, b_v) \\ = \; & I(a_1 | b_1, b_2, \ldots, b_v) - I(a_1 | a_2, b_1, b_2, \ldots, b_v). \end{aligned}$$

Mathematically, when $u > 2$, it is defined by

$$\begin{aligned} & I(a_1, a_2, \ldots, a_u | b_1, b_2, \ldots, b_v) \\ = \; & I(a_1, a_2, \ldots, a_{u-1} | b_1, b_2, \ldots, b_v) \\ & - I(a_1, a_2, \ldots, a_{u-1} | a_u, b_1, b_2, \ldots, b_v). \end{aligned} \tag{23}$$

If $u+v=n$, $I(a_1, a_2, \ldots, a_u | b_1, b_2, \ldots, b_v)$ can be written as the form of $I(A|\mathcal{N}/A)$ with $A = \{a_1, a_2, \ldots, a_u\}$, which is the amount of mutual information among all the nodes in A given all the secret bits of the nodes not in A. If $u+v < n$, $I(a_1, a_2, \ldots, a_u | b_1, b_2, \ldots, b_v)$ can be decomposed as the sum of multiple terms in the form of $I(A|\mathcal{N}/A)$ by iteratively applying

$$\begin{aligned} & I(a_1, a_2, \ldots, a_{u-1} | b_1, b_2, \ldots, b_v) \\ = \; & I(a_1, a_2, \ldots, a_u | b_1, b_2, \ldots, b_v) \\ & + I(a_1, a_2, \ldots, a_{u-1} | a_u, b_1, b_2, \ldots, b_v). \end{aligned} \tag{24}$$

From Lemma 11 (given later in this subsection), we can get

$$\begin{aligned} \sum_{i \in \mathcal{N}} H(\mathbf{u}_i) &= \sum_{i=1}^{n} \sum_{a=1}^{n} \Big(\sum_{A: i \in A, |A|=a} I(A|\mathcal{N}/A) \Big) \\ &= \sum_{a=1}^{n} \Big(\sum_{i=1}^{n} \sum_{A: i \in A, |A|=a} I(A|\mathcal{N}/A) \Big) \\ &= \sum_{a=1}^{n} \Big(\sum_{A: |A|=a} \sum_{i \in A} I(A|\mathcal{N}/A) \Big) \\ &= \sum_{a=1}^{n} \Big(a \sum_{A: |A|=a} I(A|\mathcal{N}/A) \Big). \end{aligned} \tag{25}$$

On the other hand, from Lemma 12 (given later in this subsection), we can get

$$\begin{aligned} & \sum_{i,j,\mathcal{N}_{\mathrm{h}} | i,j \notin \mathcal{N}_{\mathrm{h}}, |\mathcal{N}_{\mathrm{h}}|=t} \mathrm{I}[\mathbf{u}_i; \mathbf{u}_j | \mathbf{u}_{\mathrm{h}}] \\ = & \sum_{i,j,\mathcal{N}_{\mathrm{h}} | i,j \notin \mathcal{N}_{\mathrm{h}}, |\mathcal{N}_{\mathrm{h}}|=t} \Big(\sum_{a=2}^{n-t} \Big(\sum_{A: i,j \in A \subseteq \mathcal{N}/\mathcal{N}_{\mathrm{h}}, |A|=a} I(A|\mathcal{N}/A) \Big) \Big) \\ = & \sum_{a=2}^{n-t} \Big(\sum_{i,j,\mathcal{N}_{\mathrm{h}} | i,j \notin \mathcal{N}_{\mathrm{h}}, |\mathcal{N}_{\mathrm{h}}|=t} \Big(\sum_{A: i,j \in A \subseteq \mathcal{N}/\mathcal{N}_{\mathrm{h}}, |A|=a} I(A|\mathcal{N}/A) \Big) \Big) \\ = & \sum_{a=2}^{n-t} \Big(\sum_{A: |A|=a} \sum_{i,j,\mathcal{N}_{\mathrm{h}} | i,j \in A \subseteq \mathcal{N}/\mathcal{N}_{\mathrm{h}}} I(A|\mathcal{N}/A) \Big) \Big) \\ = & \sum_{a=2}^{n-t} \Big(\binom{a}{2} \binom{n-a}{t} \sum_{A: |A|=a} I(A|\mathcal{N}/A) \Big). \end{aligned} \tag{26}$$

Let $\beta(a) = \sum_{A: |A|=a} I(A|\mathcal{N}/A) \geq 0$. From (25) and (26), we get

$$\begin{aligned} \frac{\sum_{i,j,\mathcal{N}_{\mathrm{h}} | i,j \notin \mathcal{N}_{\mathrm{h}}, |\mathcal{N}_{\mathrm{h}}|=t} \mathrm{I}[\mathbf{u}_i; \mathbf{u}_j | \mathbf{u}_{\mathrm{h}}]}{\sum_{i \in \mathcal{N}} H(\mathbf{u}_i)} &= \frac{\sum_{a=2}^{n-t} \left(\binom{a}{2} \binom{n-a}{t} \beta(a) \right)}{\sum_{a=1}^{n} (a \beta(a))} \\ &\leq \max_{a=2}^{n-t} \frac{\binom{a}{2} \binom{n-a}{t}}{a}, \end{aligned}$$

where the maximum is achieved with $a = \lceil \frac{n}{t+1} \rceil$. To see this, we let $\alpha(a) = \frac{\binom{a}{2}\binom{n-a}{t}}{a}$, then $\alpha(a+1) > \alpha(a)$ if and only if $a(n-a-t) > (a-1)(n-a)$, equivalent to $a < \frac{n}{t+1}$.

As a result,

$$\frac{\sum_{i,j,\mathcal{N}_\mathrm{h}|i,j\notin\mathcal{N}_\mathrm{h},|\mathcal{N}_\mathrm{h}|=t}\mathbf{I}[\mathbf{u}_i;\mathbf{u}_j|\mathbf{u}_\mathrm{h}]}{\sum_{i\in\mathcal{N}}H(\mathbf{u}_i)}\leq\frac{\binom{a}{2}\binom{n-a}{t}}{a} \tag{27}$$

with $a=\lceil\frac{n}{t+1}\rceil$.

Combining (27) with (21) and (22), we can get

$$\begin{aligned}C_{\text{channel}} &\leq \frac{\binom{a}{2}\binom{n-a}{t}n}{\binom{n}{t}\binom{n-t}{2}a}\\ &= \frac{(a-1)(n-t-2)!(n-a)!}{(n-1)!(n-a-t)!}\\ &= \frac{\binom{n-t-2}{a-2}}{\binom{n-1}{a-1}}.\end{aligned}$$

This leads to the upper bound on the channel capacity.

From Corollary 8, the maximum channel rate of the combinational key distribution that distributes each combination of a nodes the same number of distinct secret bits is

$$R_{\text{channel}}(n,t,\psi_{\text{comb}})=\frac{a-1}{n-1}\frac{\binom{n-t-2}{a-2}}{\binom{n-2}{a-2}}=\frac{\binom{n-t-2}{a-2}}{\binom{n-1}{a-1}}.$$

So the above upper bound can be achieved with the combinational key distribution with $a=\lceil\frac{n}{t+1}\rceil$ if the underlying privacy amplification is asymptotically optimal.

Lemma 11:

$$H(\mathbf{u}_i)=\sum_{a=1}^{n}\Big(\sum_{A:i\in A,|A|=a}I(A|\mathcal{N}/A)\Big). \tag{28}$$

Proof: Let us prove this by induction. Without loss of generality, we let $i=1$. Firstly, when $n=1$, the equation (28) holds as

$$\sum_{a=1}^{n}\Big(\sum_{A:1\in A,|A|=a}I(A|\mathcal{N}/A)\Big)=I(1)=H(\mathbf{u}_1).$$

We show that if (28) holds for any $n\leq k$, then it also holds for $n=k+1$.

Our idea is to concatenate the secret bits of node k and node $k+1$ as a new node k'. Therefore, we get a new groups of network nodes $\mathcal{N}'=\{1,2,\ldots,k-1,k'\}$ of size k. According to our assumption that (28) for $n=k$,

$$\begin{aligned}&H(\mathbf{u}_1)\\ =&\sum_{a=1}^{k}\Big(\sum_{A':1\in A'\subseteq\mathcal{N}',|A'|=a}I(A'|\mathcal{N}'/A')\Big)\\ =&\sum_{a=1}^{k}\Big(\sum_{A':1,k'\in A'\subseteq\mathcal{N}',|A'|=a}I(A'|\mathcal{N}'/A')\Big)\\ &+\sum_{a=1}^{k}\Big(\sum_{A':1\in A'\subseteq\mathcal{N}'\subseteq\mathcal{N}',k'\notin A',|A'|=a}I(A'|\mathcal{N}'/A')\Big).\end{aligned} \tag{29}$$

Given any $A'\subseteq\mathcal{N}'$, we define $A\subseteq\mathcal{N}$ by replacing k' in A' with k and $k+1$, we define $A_k\subseteq\mathcal{N}$ by replacing k' in A' with k, and define $A_{k+1}\subseteq\mathcal{N}$ by replacing k' in A' with $k+1$.

When $k'\in A'\subseteq\mathcal{N}'$, it has

$$\begin{aligned}&I(A'|\mathcal{N}'/A')\\ =&\ I(A'|\mathcal{N}/A)\\ =&\ I(A'k|\mathcal{N}/A)+I(A'|k(\mathcal{N}/A)) &(30)\\ =&\ I(A_k|\mathcal{N}/A)+I(A_{k+1}|\mathcal{N}/A_{k+1})\\ =&\ I(A_k(k+1)|\mathcal{N}/A)+I(A_k|(k+1)(\mathcal{N}/A))\\ &+I(A_{k+1}|\mathcal{N}/A_{k+1}) &(31)\\ =&\ I(A|\mathcal{N}/A)+I(A_k|\mathcal{N}/A_k)+I(A_{k+1}|\mathcal{N}/A_{k+1}), &(32)\end{aligned}$$

where (30) and (31) are due to the definition (24).

When $k'\notin A'\subseteq\mathcal{N}'$, it has

$$I(A'|\mathcal{N}'/A')=I(A|\mathcal{N}/A). \tag{33}$$

Substituting (32) and (33) into (29) yields

$$\begin{aligned}&H(\mathbf{u}_1)\\ =&\sum_{a=1}^{k}\Big(\sum_{A:1,k,k+1\in A\subseteq\mathcal{N},|A|=a+1}I(A|\mathcal{N}/A)\Big)\\ &+\sum_{a=1}^{k}\Big(\sum_{A_k:1,k\in A_k\subseteq\mathcal{N},k+1\notin A_k,|A_k|=a}I(A_k|\mathcal{N}/A_k)\Big)\\ &+\sum_{a=1}^{k}\Big(\sum_{A_{k+1}:1,k+1\in A_{k+1}\subseteq\mathcal{N},k\notin A_{k+1},|A_{k+1}|=a}I(A_{k+1}|\mathcal{N}/A_{k+1})\Big)\\ &+\sum_{a=1}^{k}\Big(\sum_{A:1\in A\subseteq\mathcal{N},k,k+1\notin A,|A|=a}I(A|\mathcal{N}/A)\Big)\\ =&\sum_{a=1}^{k}\Big(\sum_{A:1\in A\subseteq\mathcal{N},|A|=a}I(A|\mathcal{N}/A)\Big)\\ &+\sum_{a=k+1}^{k+1}\Big(\sum_{A:1,k,k+1\in A\subseteq\mathcal{N},|A|=a}I(A|\mathcal{N}/A)\Big)\\ &-\sum_{a=1}^{1}\Big(\sum_{A:1,k,k+1\in A\subseteq\mathcal{N},|A|=a}I(A|\mathcal{N}/A)\Big)\\ =&\sum_{a=1}^{k}\Big(\sum_{A:1\in A\subseteq\mathcal{N},|A|=a}I(A|\mathcal{N}/A)\Big)\\ &+\sum_{a=k+1}^{k+1}\Big(\sum_{A:1\in A\subseteq\mathcal{N},|A|=a}I(A|\mathcal{N}/A)\Big)\\ =&\sum_{a=1}^{k+1}\Big(\sum_{A:1\in A\subseteq\mathcal{N},|A|=a}I(A|\mathcal{N}/A)\Big).\end{aligned} \tag{34}$$

Finally, the result can be reached by induction. This completes the proof. □

Lemma 12:

$$\mathbf{I}[\mathbf{u}_i;\mathbf{u}_j|\mathbf{u}_\mathrm{h}]=\sum_{a=2}^{n-t}\Big(\sum_{A:i,j\in A\subseteq\mathcal{N}/\mathcal{N}_\mathrm{h},|A|=a}I(A|\mathcal{N}/A)\Big).$$

Proof: The proof is similar to that of Lemma 11, emitted here. □

C. Proof of Theorem 3

Given a scheme ψ, let $\mathbf{u} \in \{0,1\}^u$ be the independent sequence of random bits from which the assigned secret bits are chosen. Denote the fraction of bits that are assigned only to the set of nodes $A \subseteq \mathcal{N}$ by p_A with $0 \le p_A \le 1$ and

$$\sum_{A \subseteq \mathcal{N}} p_A = 1.$$

We define

$$\mathbb{E}(x) = \sum_{A \subseteq \mathcal{N}} (p_A x)$$

for a variable x depending on A. It is like the expectation of x.

The total amount of space in all the n nodes needed to store the assigned secret bits is

$$\sum_{A \subseteq \mathcal{N}} (p_A u)|A| = (\sum_{A \subseteq \mathcal{N}} p_A |A|)u = \mathbb{E}(|A|)u \le nl, \qquad (35)$$

where u is the total number of independent secret bits assigned to the network.

Firstly, the total number of secure message bits is upper bounded by the total number of secret bits u, hence as $l \to \infty$, the maximum network rate

$$R_{\text{net}}(n,0,\psi) \le \frac{u}{l} \le \frac{nl}{\mathbb{E}(|A|)l} = \frac{n}{\mathbb{E}(|A|)}. \qquad (36)$$

Secondly, the number of message bits that can be securely communicated over a channel is upper bounded by the number of common secret bits shared by the two terminals. Let $\mathbf{u}_{ij}$ be the sequence of common random bits shared by node i and j. As $l \to \infty$,

$$\begin{aligned} R_{\text{channel}}(n,0,\psi) &\le \min_{(i,j)\in\mathcal{P}} \frac{|\mathbf{u}_{ij}|}{l} \\ &\le \frac{\sum_{(i,j)\in\mathcal{P}} |\mathbf{u}_{ij}|}{\binom{n}{2} l} \\ &= \frac{\sum_{(i,j)\in\mathcal{P}} \sum_{A: i,j \in A} p_A u}{\binom{n}{2} l} \\ &= \frac{\sum_{A\subseteq\mathcal{N}} p_A \binom{|A|}{2} u}{\binom{n}{2} l} \\ &= \frac{u\mathbb{E}(|A|(|A|-1))}{l \cdot n(n-1)}. \end{aligned} \qquad (37)$$

Substituting (35) into the above inequality yields

$$R_{\text{channel}}(n,0,\psi) \le \frac{\mathbb{E}(|A|(|A|-1))}{\mathbb{E}(|A|)(n-1)} \qquad (38)$$

Since $|A|(|A|-1) \le (n+1)|A| - 2n$ for $2 \le |A| \le n$, it has

$$R_{\text{channel}}(n,0,\psi) \le \frac{n+1}{n-1} - \frac{2n}{\mathbb{E}(|A|)(n-1)}. \qquad (39)$$

From(36) and (39), we obtain

$$R_{\text{net}}(n,0,\psi)\frac{2}{n+1} + R_{\text{channel}}(n,0,\psi)\frac{n-1}{n+1} \le 1.$$

Let us prove the achievablity, starting from two simple schemes. In the first scheme, the same-key distribution scheme, all the nodes share the same set of secret bits, and its maximum rates are

$$R_{\text{net}}(n,0,\psi_{\text{same}}) = 1, \quad R_{\text{channel}}(n,0,\psi_{\text{same}}) = 1. \qquad (40)$$

The second scheme is the pairwise key distribution, where each pair of nodes share the same number of distinct secret bits. Its maximum rates are

$$R_{\text{net}}(n,0,\psi_{\text{pair}}) = \frac{n}{2}, \quad R_{\text{channel}}(n,0,\psi_{\text{pair}}) = \frac{1}{n-1}. \qquad (41)$$

The equality in the theorem holds both for the same-key distribution and the pairwise key distribution. Here we construct a scheme as the hybrid of the two simple schemes. For each node, it uses a fraction $0 \le \lambda \le 1$ of its storage space for the same-key distribution and the rest for the pairwise key distribution. The maximum rates for the hybrid scheme ψ_{hybrid} are

$$\begin{aligned} R_{\text{net}}(n,0,\psi_{\text{hybrid}}) &= \lambda + \frac{n}{2}(1-\lambda), \\ R_{\text{channel}}(n,0,\psi_{\text{hybrid}}) &= \lambda + \frac{1}{n-1}(1-\lambda). \end{aligned}$$

By adjusting the fraction λ, we can obtain all the maximum network rates and the maximum channel rates meeting the equality in the theorem.

D. Proof of Theorem 4

The necessity is easy to prove. For privacy amplification, the total length of the secret keys generated for a subset of channels $\mathcal{P}$ is less than the total number of unhacked secret bits, i.e.,

$$\frac{\sum_{(i,j)\in P} |\mathbf{s}_{ij}|}{|\cup_{(i,j)\in P} \mathbf{u}_{ij}/\mathbf{u}_{\text{h}}|} < 1.$$

If there exists a subset of channels $\mathcal{P}$ violating (4), i.e.,

$$\sum_{(i,j)\in P} r_{ij} \ge \frac{|\cup_{(i,j)\in P} \mathbf{u}_{ij}/\mathbf{u}_{\text{h}}|}{l},$$

then their total message length must be larger than the total secret-key length. As a result, at least one of these messages must be information-theoretically insecure.

To prove achievability, we consider a simple method for privacy amplification: for every channel (i,j), given the common secret bits $\mathbf{u}_{ij}$, the secret key $\mathbf{s}_{ij} = M_{ij}\mathbf{u}_{ij}$ with a sparse random matrix M_{ij} of density $O(\log l/l)$. The reason of using this method is not only due to its asymptotic optimality, but also to its practicality. It is the basis of our proposed network schemes.

Let $\mathbf{s}_{\text{s}} = \{\mathbf{s}_{ij} | (i,j) \in \mathcal{P}_{\text{s}}\}$ be the secret keys between unhacked nodes, and let $\mathbf{u}_{\text{h}}$ be the distinct secret bits stored in hacked nodes, which are disclosed to the eavesdropper. The network communication is information-theoretically secure if and only if for any possible set of hacked nodes $\mathcal{N}_h$, the secret keys $\mathbf{s}_{\text{s}}$ and the hacked secret bits $\mathbf{u}_{\text{h}}$ are truly random bits, and $\mathbf{s}_{\text{s}}, \mathbf{u}_{\text{h}}$ are independent. Note that both $\mathbf{s}_{\text{s}}$ and $\mathbf{u}_{\text{h}}$ can be written as linear transformations of the source sequence $\mathbf{u}$.

Let $\mathbf{z}$ be the concatenation of $\mathbf{s}_{\text{s}}$ and $\mathbf{u}_{\text{h}}$, then

$$\mathbf{z} = \mathbf{s}_{\text{s}}\mathbf{u}_{\text{h}} = M\mathbf{u} \qquad (42)$$

for some matrix M. The network is information-theoretically secure if and only if all the rows in matrix M are linearly independent.

We can write the secret key $\mathbf{s}_{ij}$ as

$$\begin{aligned}\mathbf{s}_{ij} &= A_{ij}^{'}(\mathbf{u}_{ij}/\mathbf{u}_{\rm h}) + B_{ij}^{'}(\mathbf{u}_{ij} \cap \mathbf{u}_{\rm h}) \\ &= A_{ij}(\mathbf{u}/\mathbf{u}_{\rm h}) + B_{ij}\mathbf{u}_h\end{aligned}$$

for some matrices A_{ij} and B_{ij}, where A_{ij} is an $(lr_{ij})\times|\mathbf{u}/\mathbf{u}_{\rm h}|$ matrix consisting of $|\mathbf{u}_{ij}/\mathbf{u}_{\rm h}|$ random columns of density $O(\log l/l)$ and $|\mathbf{u}/\mathbf{u}_{\rm h}| - |\mathbf{u}_{ij}/\mathbf{u}_{\rm h}|$ zero columns.

Then $\mathbf{z} = \mathbf{s}_{\rm s}\mathbf{u}_{\rm h}$ is represented by

$$\mathbf{z} = M\mathbf{u} = \begin{pmatrix} A & B \\ 0 & I \end{pmatrix} \begin{pmatrix} \mathbf{u}/\mathbf{u}_{\rm h} \\ \mathbf{u}_{\rm h} \end{pmatrix}, \tag{43}$$

where I is an identity matrix and A consists of all the matrices A_{ij} with $(i,j) \in \mathcal{P}_{\rm s}$, i.e.,

$$A = \begin{pmatrix} A_{12} \\ A_{13} \\ \vdots \\ A_{(n-1)n} \end{pmatrix}.$$

The kth column of A_{ij} is a random vector of length $|\mathbf{s}_{ij}|$ and density $O(\log l/l)$ if the kth bit in $\mathbf{u}$ is distributed to both node i and node j, i.e., $\mathbf{u}[k] \in \mathbf{u}_{ij}$, otherwise the k column of A_{ij} is the all-zero vector of length $|\mathbf{s}_{ij}|$.

The network is information-theoretically secure if the rows in M are linearly independent. This is equivalent to showing that the rows in A are linearly independent, i.e., all the rows in $\{A_{ij}|(i,j) \in \mathcal{P}_{\rm s}\}$ are linearly independent. This can be proved based on the following results.

Lemma 13: All the rows in $\{A_{ij}|(i,j) \in \mathcal{P}_{\rm s}\}$ are linearly independent if and only if for any subset of channels $P \subseteq \mathcal{P}_{\rm s}$, there does not exist any subset of rows from $\{A_{ij}|(i,j) \in P\}$ that includes at least one row from each matrix such that their sum is a zero-vector.

Lemma 14: Given any subset of channels $P \subseteq \mathcal{P}_{\rm s}$, if $\mathbf{s}_{ij} = M_{ij}\mathbf{u}_{ij}$ with a random matrix M_{ij} of density $O(\log l/l)$ and

$$\frac{\sum_{(i,j)\in P} |\mathbf{s}_{ij}|}{|\cup_{(i,j)\in P} \mathbf{u}_{ij}/\mathbf{u}_{\rm h}|} < 1$$

with $|\cup_{(i,j)\in P} \mathbf{u}_{ij}/\mathbf{u}_{\rm h}| = O(l)$, when $l \to \infty$, with probability almost 1 there does not exist any subset of rows from $\{A_{ij}|(i,j) \in P\}$ that includes at least one row from each matrix such that their sum is a zero-vector.

The proof of Lemma 14 is provided in subsection VIII-F. Finally, we can conclude that the rows of the security matrix M are linearly independent with high probability, and the criteria in Theorem 4 are sufficient.

E. Proof of Theorem 7

Using the same proof as Theorem 4, the network is information-theoretically secure if and only if the rows of the matrix A in (43) are linearly independent. In Theorem 7, for this matrix A, it has the following properties: there are $|\mathbf{u}_G/\mathbf{u}_{\rm h}|$ columns in A corresponding to the bits in $\mathbf{u}_G/\mathbf{u}_{\rm h}$, in which each column has $\sum_{(i,j)\in G} m_{ij}$ random entries with $m_{ij} = l \cdot r_{ij}$ corresponding to the bits in $\{\mathbf{s}_{ij}\}$ with $i,j \in G$.

The rank of the matrix A remains unchanged if we do elementary row or column operations on A. The rows of a matrix are linearly independent if and only if the the matrix can be reduced to the simplest form $[I, 0]$ by elementary operations such that it consists of an identity matrix and a zero matrix.

If there exists a feasible solution for $\{x_{ij}^G\}$, we can divide the columns corresponding to the bits in $\mathbf{u}_G/\mathbf{u}_{\rm h}$ into some groups of sizes $\{u_{ij}^G\}$ with $u_{ij}^G = l \cdot x_{ij}^G$ and $\sum_{i,j\in G} u_{ij}^G = |\mathbf{u}_G/\mathbf{u}_{\rm h}|$.

On the other hand, we can divide the rows corresponding to the bits in $\mathbf{s}_{ij}$ into some groups of sizes $\{m_{ij}^G\}$ with $m_{ij}^G = l \cdot y_{ij}^G$ and

$$y_{ij}^G = \frac{r_{ij}}{\sum_G x_{ij}^G} x_{ij}^G, \sum_{G|i,j\in G} m_{ij}^G = |\mathbf{s}_{ij}|.$$

According to the inequalities in the theorem, it has either $y_{ij}^G < x_{ij}^G$ or $y_{ij}^G = 0$.

Based on the row groups and the column groups, the matrix A is divided into $|\{m_{ij}^G\}|\times|\{u_{ij}^G\}|$ sub-matrices, whose dimensions are $\{m_{ij}^G\} \times \{u_{ij}^G\}$. By switching the rows and columns of the matrix A, the matrix A can be transformed into a form such that the sub-matrices of dimensions $\{m_{ij}^G \times u_{ij}^G\}$ are on the diagonal of the sub-matrices. We denote the sub-matrices on the diagonal by $[A_1, A_2, \ldots] = \{A_{ij}^G\}$, and the matrix A is transformed to

$$A \Leftrightarrow \begin{pmatrix} A_1 & \ldots & \ldots \\ \vdots & A_2 & \vdots \\ \vdots & \ldots & \ddots \end{pmatrix}.$$

The sub-matrices $[A_1, A_2, \ldots]$ are random matrices of density $O(\log l/l)$. The dimension of the sub-matrix A_i is $m_i \times u_i$ for some m_i, u_i such that $\frac{m_i}{u_i} < 1$ for $u_i = O(l)$ or $m_i = 0$.

For the sub-matrix A_1, according to Lemma 15 in subsection VIII-F, the rows of A_1 are linearly independent with high probability when l is sufficiently large. The sub-matrix A_1 can be reduced to its simplest form $[I_1, 0]$ consisting of an identity matrix and a zero matrix by elementary operations on A. Furthermore, all the other entries on the right of A_1 (in the same rows with A_1) can be reduced to 0 by elementary column operations. Right now, each sub-matrix A_i with $i > 1$ is transformed to A_i' with

$$A_i' = A_i + \overline{A_i}$$

for some $\overline{A_i}$ independent of A_i, and the matrix A is reduced to

$$A \Leftrightarrow \begin{pmatrix} I_1 0 & 0 & 0 \\ \vdots & A_2' & \vdots \\ \vdots & \ldots & \ddots \end{pmatrix}.$$

We continue repeating the above process to handle $A_2', A_3', \ldots$, iteratively. For the sub-matrix $A_i' = A_i + \overline{A_i}$, it can be proved that the conclusion of Lemma 15 still holds, and all the rows of A_i' are linearly independent with high probability when l is sufficiently large.

Finally, all the sub-matrices $[A_1, A_2, \ldots]$ are reduced to their simplest forms with high probability, and all the other entries on their right are 0s. In this case, the matrix A is reduced to

the reversed row echelon form, and it has full rank. Hence all the rows of the matrix A are linearly independent with high probability if l is sufficiently large. This leads to the achievability of the channel rates.

F. Proof of Lemma 14

We first prove the following result.

Lemma 15: Let $M \in \{0,1\}^{k\times r}$ be a random matrix such that the probability of each entry being 1 is $p = O(\log r/r)$. The rows in M are linearly independent with high probability for sufficiently large r if and only if $k/r < 1$.

Each row in M is an independent random vector. The sum of any j rows in M is still an independent random vector. Denote the probability of its entry being 1 by p_j. Let $p = \frac{d}{r}$ for a $d = O(\log r)$.

It is easy to show that

$$p_j = p_{j-1}(1-p) + (1-p_{j-1})p,$$

from which and by induction, we obtain

$$p_j = \frac{1}{2} - \frac{1}{2}(1-\frac{2d}{r})^j.$$

Furthermore, since the sum of any j rows is an independent vector, the probability for it being a zero-vector is

$$P_j(0) = (1-p_j)^r.$$

The rows of M are linearly independent if and only if for any subset of the rows, their sum is not a zero-vector. Hence, the probability of the rows of M being linearly independent

$$\mathbb{P}_{\text{indep}}(M) \geq 1 - \sum_{j=1}^{k} \binom{k}{j} P_j(0),$$

where $\binom{k}{j}$ is the number of subsets consisting of j rows.

This leads to

$$\mathbb{P}_{\text{indep}}(M) \geq 1 - \sum_{j=1}^{k} \binom{k}{j} (\frac{1}{2} + \frac{1}{2}(1-\frac{2d}{r})^j)^r. \qquad (44)$$

When $j < \frac{r}{2d}$ with r sufficiently large, it has

$$\begin{aligned}
&\sum_{j=1}^{\frac{r}{2d}} \binom{k}{j} (\frac{1}{2} + \frac{1}{2}(1-\frac{2d}{r})^j)^r \\
\leq &\sum_{j=1}^{\frac{r}{2d}} k^j (1 - \frac{dj}{r} + \frac{j(j-1)d^2}{r^2})^r \\
\leq &\sum_{j=1}^{\frac{r}{2d}} k^j (1 - \frac{dj}{2r})^r \\
\leq &\sum_{j=1}^{\frac{r}{2d}} r^j e^{-dj/2} \\
\leq &\sum_{j=1}^{\frac{r}{2d}} (e^{\log r - d/2})^j \to 0.
\end{aligned}$$

When $\frac{r}{2d} \leq j < \frac{r}{3\log r}$ with r sufficiently large, it has

$$\begin{aligned}
&\sum_{j=\frac{r}{2d}}^{\frac{r}{3\log r}} \binom{k}{j} (\frac{1}{2} + \frac{1}{2}(1-\frac{2d}{r})^j)^r \\
\leq &\sum_{j=\frac{r}{2d}}^{\frac{r}{3\log r}} \binom{k}{j} (\frac{1}{2} + \frac{e^{-1}}{2})^r \\
\leq &r^{\frac{r}{3\log r}+1} (\frac{1}{2} + \frac{e^{-1}}{2})^r \\
\leq &r(e^{\frac{1}{3}}(\frac{1}{2} + \frac{e^{-1}}{2}))^r \to 0.
\end{aligned}$$

When $j \geq \frac{r}{3\log r}$ with r sufficiently large, it has

$$\begin{aligned}
&\sum_{j=\frac{r}{3\log r}}^{k} \binom{k}{j} (\frac{1}{2} + \frac{1}{2}(1-\frac{2d}{r})^j)^r \\
\leq &\sum_{j=\frac{r}{3\log r}}^{k} \binom{k}{j} (\frac{1}{2} + \frac{e^{-\frac{2d}{3\log r}}}{2})^r \\
\leq &2^k (\frac{1}{2} + \frac{e^{-\frac{2d}{3\log r}}}{2})^r \\
= &(2^{\frac{k}{r}-1}(1 + e^{-\frac{2d}{3\log r}}))^r \\
\to &(2^{\frac{k}{r}-1})^r \to 0.
\end{aligned}$$

Summing the above results up, we obtain

$$\mathbb{P}_{\text{indep}}(M) \geq 1 - \epsilon$$

for any $\epsilon > 0$ when r is sufficiently large.

Lemma 16: Given $S_i \subseteq \{1,2,\ldots,r\}$ with $1 \leq i \leq k$, let $M_i \in \{0,1\}^{m_i \times r}$ with $1 \leq i \leq k$ be a binary matrix such that each entry in columns S_i is 1 with probability $O(\log r/r)$ and each entry not in columns S_i is 0. If $\frac{\sum_i m_i}{|\cup_i S_i|} < 1$ and $|\cup_i S_i| = O(r)$, as $r \to \infty$, with high probability there does not exist any subset of rows from $\{M_i\}$ that includes at least one row from each matrix such that their sum is a zero-vector.

We say that a set of matrices $\{M_i\}$ are linearly cross-independent if and only if there does not exist a subset of rows from $\{M_i\}$ that includes at least one row from each matrix such that their sum is a zero-vector. If the rows of $M_1, M_2, \ldots, M_k$ are linearly cross-independent, it does not necessarily imply that these rows are linearly independent. For example, consider the matrices

$$M_1 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix}, M_2 = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix}$$

The rows in M_1, M_2 are linearly cross-independent, but not linearly independent, as the rows in M_1 are not linearly independent.

One observation is that if $\{M_i\}$ are linearly cross-independent on a subset of columns, then $\{M_i\}$ are linearly cross-independent on all the columns.

We divide the columns into at most 2^k groups depending on which S_i the column belongs to. Two columns are in the same group if and only if they belong to the same subset of $\{S_1, S_2, \ldots, S_k\}$. Now, we are only interested in the groups

of size $O(r)$ (sufficiently large groups), and the union of their columns are denoted by S. Then

$$|S| > |\cup_i S_i|(1-\epsilon)$$

for sufficiently small ϵ, which leads to $\frac{\sum_i m_i}{|S|} < 1$ for sufficiently large r. We will prove that the matrices $\{M_i\}$ are linearly cross-independent on the columns in S.

Given $\ell = \{l_1, \ldots, l_k\}$ with $1 \le l_i \le m_i$, we choose l_i rows from M_i with $1 \le i \le k$, and we use $P(\ell)$ to denote the probability that the sum of all the $\sum_i l_i$ chosen rows is a zero-vector on S. Then the probability of the matrices $\{M_i\}$ being not linearly cross-independent on S is

$$\mathbb{P}_{\text{dep.}} = \sum_{l_1,l_2,\ldots} \prod_{i=1}^{k} \binom{m_i}{l_i} P(\ell).$$

There are two possibilities considering the chosen $\sum_i$ rows: (1) every column in S has more than $\frac{r}{\log r}$ random entries in the chosen rows; and (2) there exists a group (among the up to 2^k groups) of columns in S, whose size is at least $b = O(r)$ and in which each column has at most $\frac{r}{\log r}$ random entries in the chosen rows. We use $P_1(\ell)$ to denote the probability that the sum of chosen rows is a zero-vector on S in the first case, and $P_2(\ell)$ to denote that in the second case. It can be shown that

$$P_1(\ell) \le (\frac{1}{2}+\epsilon)^{|S|}$$

for sufficient small ϵ and

$$P_2(\ell) \le 2^k \sum_{j=1}^{\frac{r}{\log r}} \binom{\sum_i l_i}{j} (\frac{1}{2} + \frac{1}{2}(1-\frac{2d}{r})^j)^b.$$

Consider all possible choices of ℓ, as $r \to \infty$, the sum probability of the first case is

$$\begin{aligned}
\mathbb{P}_1 &= \sum_{l_1,l_2,\ldots} \prod_{i=1}^{k} \binom{m_i}{l_i} P_1(\ell) \\
&\le \sum_{l_1,l_2,\ldots} \prod_{i=1}^{k} \binom{m_i}{l_i} (\frac{1}{2}+\epsilon)^{|S|} \\
&\le 2^{\sum_i m_i} (\frac{1}{2}+\epsilon)^{|S|} \\
&\le \epsilon
\end{aligned}$$

for sufficiently small ϵ.

Consider all possible choices of ℓ, as $r \to \infty$, the sum probability of the second case is

$$\begin{aligned}
\mathbb{P}_2 &= \sum_{l_1,l_2,\ldots} \prod_{i=1}^{k} \binom{m_i}{l_i} P_2(\ell) \\
&\le \sum_{j=1}^{\frac{r}{\log r}} \binom{\sum_i m_i}{j} (\frac{1}{2} + \frac{1}{2}(1-\frac{2d}{r})^j)^b \\
&\le \epsilon
\end{aligned}$$

for sufficiently small ϵ. The last step follows the same proof as Lemma 15.

Finally, the matrices $\{M_i\}$ are not linearly cross-independent with probability

$$\mathbb{P}_{\text{dep.}} \le \mathbb{P}_1 + \mathbb{P}_2 \le 2\epsilon$$

as $r \to \infty$. This leads to the conclusion in Lemma 16.

It is straightforward to obtain Lemma 14 from Lemma 16.

G. Proof of Corollary 8

The proof directly follows Corollary 6. We first work on the maximum network rate of the combinational key distribution with $a \ge 2$, which is

$$R_{\text{network}}(n,t,\psi_{\text{comb}}) = \sum_{\mathcal{N}_{\text{h}}:|\mathcal{N}_{\text{h}}|=t} \frac{|\cup_{(i,j)\in\mathcal{P}_{\text{s}}} \mathbf{u}_{ij}/\mathbf{u}_h|}{l} / \binom{n-2}{t}$$

for sufficiently large l. Given any set of hacked nodes $\mathcal{N}_{\text{h}}$ with $|\mathcal{N}_{\text{h}}| = t$, the number of unhacked groups is $\binom{n-t}{a}$, with each of size $w = \frac{l}{\binom{n-1}{a-1}}$. As a result, for any $\mathcal{N}_{\text{h}}$,

$$\frac{|\cup_{(i,j)\in\mathcal{P}_{\text{s}}} \mathbf{u}_{ij}/\mathbf{u}_h|}{l} = \binom{n-t}{a}\frac{w}{l} = \frac{\binom{n-t}{a}}{\binom{n-1}{a-1}}.$$

This leads to the maximum network rate

$$R_{\text{network}}(n,t,\psi_{\text{comb}}) = \frac{\binom{n}{t}\binom{n-t}{a}}{\binom{n-2}{t}\binom{n-1}{a-1}} = \frac{n}{a}\frac{\binom{n-t-2}{a-2}}{\binom{n-2}{a-2}}.$$

It can be achieved by the equal channel rates with $r_{ij} = \frac{2}{a(n-1)}\frac{\binom{n-t-2}{a-2}}{\binom{n-2}{a-2}}$.

For the maximum channel rate,

$$R_{\text{channel}}(n,t,\psi_{\text{comb}}) = \min_{\mathcal{N}_h,i,j|i,j\notin\mathcal{N}_{\text{h}}} \frac{|\mathbf{u}_{ij}/\mathbf{u}_h|}{l}.$$

With the combinational key distribution, every pair of unhacked nodes share $\binom{n-t-2}{a-2}$ groups of secret bits that are not hacked. As a result, for any $\mathcal{N}_h$ and $i,j \notin \mathcal{N}_{\text{h}}$, $\frac{|\mathbf{u}_{ij}/\mathbf{u}_h|}{l} = \frac{\binom{n-t-2}{a-2}}{\binom{n-1}{a-1}}$. This leads to

$$R_{\text{channel}}(n,t,\psi_{\text{comb}}) = \frac{\binom{n-t-2}{a-2}}{\binom{n-1}{a-1}} = \frac{a-1}{n-1}\frac{\binom{n-t-2}{a-2}}{\binom{n-2}{a-2}}.$$

It can be achieved when there is only one channel with message transmissions.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their valuable comments.

REFERENCES

[1] C. E. Shannon, "Communication theory of secrecy systems," *Bell Labs Tech. J.*, vol. 28, no. 4, pp. 656–715, Oct. 1949.
[2] N. Cai and T. Chan, "Theory of secure network coding," *Proc. IEEE*, vol. 99, no. 3, pp. 421–437, Mar. 2011.
[3] T. Cui, T. Ho, and J. Kliewer, "On secure network coding with nonuniform or restricted wiretap sets," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 166–176, Jan. 2013.
[4] N. Cai and R. W. Yeung, "Secure network coding on a wiretap network," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 424–435, Jan. 2011.

[5] T. Ho, B. Leong, R. Koetter, M. Medard, M. Effros, and D. R. Karger, "Byzantine modification detection in multicast networks using randomized network coding," in *Proc. Int. Symp. Inf. Theory (ISIT)*, Jun. 2004, p. 144.
[6] S. Jaggi, M. Langberg, S. Katti, T. Ho, D. Katabi, and M. Médard, "Resilient network coding in the presence of Byzantine adversaries," in *Proc. 26th IEEE INFOCOM*, May 2007, pp. 616–624.
[7] Y.-S. Shiu, S. Y. Chang, H.-C. Wu, S. C.-H. Huang, and H.-H. Chen, "Physical layer security in wireless networks: A tutorial," *IEEE Wireless Commun.*, vol. 18, no. 2, pp. 66–74, Apr. 2011.
[8] M. Bloch, J. Barros, M. R. D. Rodrigues, and S. W. McLaughlin, "Wireless information-theoretic security," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2515–2534, Jun. 2008.
[9] Y. Liang, H. V. Poor, and S. Shamai (Shitz), "Information theoretic security," *Found. Trends Commun. Inf. Theory*, vol. 5, nos. 4–5, pp. 355–580, 2009.
[10] L. Dong, Z. Han, A. P. Petropulu, and H. V. Poor, "Improving wireless physical layer security via cooperating relays," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1875–1888, Mar. 2010.
[11] Y. Liu, Z. Qin, M. Elkashlan, Y. Gao, and L. Hanzo, "Enhancing the physical layer security of non-orthogonal multiple access in large-scale networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1656–1672, Mar. 2017.
[12] C. Cachin and U. M. Maurer, "Linking information reconciliation and privacy amplification," *J. Cryptol.*, vol. 10, no. 2, pp. 97–110, 1997.
[13] J. M. Renes and R. Renner, "Noisy channel coding via privacy amplification and information reconciliation," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7377–7385, Nov. 2011.
[14] U. M. Maurer and S. Wolf, "Information-theoretic key agreement: From weak to strong secrecy for free," in *Proc. Int. Conf. Theory Appl. Cryptograph. Techn.*, Berlin, Germany, May 2000., pp. 356–373
[15] C. H. Bennett, G. Brassard, and J.-M. Robert, "Privacy amplification by public discussion," *SIAM J. Comput.*, vol. 17, no. 2, pp. 210–229, Apr. 1988.
[16] 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.
[17] 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.
[18] J. L. Carter and M. N. Wegman, "Universal classes of hash functions," *J. Comput. Syst. Sci.*, vol. 18, no. 2, pp. 143–154, Apr. 1979.
[19] H. Zhou, V. Chandar, and G. Wornell, "Low-density random matrices for secret key extraction," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 2607–2611.
[20] R. A. Chou, M. R. Bloch, and E. Abbe, "Polar coding for secret-key generation," *IEEE Trans. Inf. Theory*, vol. 61, no. 11, pp. 6213–6237, Nov. 2015.
[21] R. G. Gallager, "Low-density parity-check codes," *IRE Trans. Inf. Theory*, vol. IT-8, no. 1, pp. 21–28, Jan. 1962.
[22] L. Eschenauer and V. D. Gligor, "A key-management scheme for distributed sensor networks," in *Proc. 9th ACM Conf. Comput. Commun. Secur. (CCS)*, Washington, DC, USA, Nov. 2002, pp. 41–47.
[23] H. Chan, A. Perrig, and D. Song, "Random key predistribution schemes for sensor networks," in *Proc. 19th Int. Conf. Data Eng.*, May 2003, pp. 197–213.
[24] J. Katz and Y. Lindell, *Introduction to Modern Cryptography: Principles and Protocols*. London, U.K.: Chapman & Hall, 2007.
[25] E. Berlekamp, R. McEliece, and H. C. A. van Tilborg, "On the inherent intractability of certain coding problems," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 384–386, May 1978.
[26] A. Shamir, "How to share a secret," *Commun. ACM*, vol. 22, no. 11, pp. 612–613, Nov. 1979.
[27] A. S. Tanenbaum and D. J. Wetherall, *Computer Networks*, 5th ed. London, U.K.: Pearson, 2010, pp. 368–370.

Hongchao Zhou (Member, IEEE) received the B.Sc. degree in physics and mathematics and M.Sc. degree in control science and engineering from Tsinghua University, Beijing, China, in 2006 and 2008, respectively, and the M.Sc. and Ph.D. degrees in electrical engineering from California Institute of Technology, Pasadena, CA, USA, in 2009 and 2012, respectively. From 2012 to 2015, he was a Post-Doctoral Researcher with the Signals, Information and Algorithms Laboratory, Massachusetts Institute of Technology. He is currently a Professor with the School of Information Science and Engineering, Shandong University. His current interests include information theory, data systems and learning systems. He was a recipient of the 2013 Charles Wilts Prize for the best doctoral thesis in electrical engineering at California Institute of Technology.

Abbas El Gamal (Life Fellow, IEEE) received the B.Sc. degree (Hons.) from Cairo University in 1972 and the M.S. degree in statistics and the Ph.D. degree in electrical engineering from Stanford University in 1977 and 1978, respectively. He is currently Hitachi America Professor with the School of Engineering, Stanford University. From 1978 to 1980, he was an Assistant Professor of electrical engineering with USC. He has been with the Faculty of the Department of Electrical Engineering, Stanford University, since 1981. From 2003 to 2012, he was the Director of the Information Systems Laboratory, Stanford University. From 2012 to 2017, he was the Chair of the Department of Electrical Engineering, Stanford University. His research contributions have been in network information theory, FPGAs, digital imaging devices and systems, and smart grid modeling and control. He has authored or coauthored over 230 articles and holds 35 patents in these areas. He is the coauthor of the book *Network Information Theory* (Cambridge Press, 2011). He is a member of the U.S. National Academy of Engineering. He received several honors and awards for his research contributions, including the 2016 IEEE Richard Hamming Medal, the 2012 Claude E. Shannon Award, and the 2004 INFOCOM Paper Award. He served on the Board of Governors for the Information Theory Society from 2009 to 2016 and the President in 2014.