Strong Secrecy and Reliable Byzantine Detection in the Presence of an Untrusted Relay

Xiang He, *Member, IEEE*, and Aylin Yener, *Member, IEEE*

***Abstract*—We consider a Gaussian two-hop network where the source and the destination can communicate only via a relay node who is both an eavesdropper and a Byzantine adversary. Both the source and the destination nodes are allowed to transmit, and the relay receives a superposition of their transmitted signals. We propose a new coding scheme that satisfies two requirements simultaneously: the transmitted message must be kept secret from the relay node, and the destination must be able to detect any Byzantine attack that the relay node might launch reliably and fast. The three main components of the proposed scheme are the nested lattice code, the privacy amplification scheme, and the algebraic manipulation detection (AMD) code. Specifically, for the Gaussian two-hop network, we show that lattice coding can successfully pair with AMD codes enabling its first application to a noisy channel model. We prove, using this new coding scheme, that the probability that the Byzantine attack goes undetected decreases exponentially fast with respect to the number of channel uses, while the loss in the secrecy rate, compared to the rate achievable when the relay is honest, can be made arbitrarily small. In addition, in contrast with prior work in Gaussian channels, the notion of secrecy provided here is strong secrecy.**

***Index Terms*—Algebraic manipulation detection (AMD) code, Byzantine detection, information-theoretic secrecy, lattice code, relay channel, strong secrecy.**

I. Introduction

INFORMATION-theoretic secrecy, first proposed by Shannon [1], provides confidentiality of transmitted information against an adversary regardless of its computational power. Shannon proved that if the adversary has access to the signals transmitted by the sender of the secret message through a noiseless channel, then, to achieve complete independence between the confidential message and the adversary's observation, the sender and the receiver have to share a secret key of the same rate as the message. Although Shannon's result implied that secret communication was impractical in this setting, it was later shown by Wyner [2] that this pessimistic result was a consequence of the noiseless channel assumption. Specifically, it was shown that when the adversary has noisy observations of the signals transmitted by the sender, a nonzero transmission rate for the secret message is achievable without requiring the transmitter to preshare a key with the receiver [2]–[4]. More recently, the fundamental rate limits at which the secret communication can take place in the presence of an eavesdropper were studied for a number of multiterminal models, e.g., the broadcast channel [5], [6], the multiple access channel [7]–[9], and the interference channel [10]–[12].

Secure communication for channel models with a relay node has been studied from a variety of perspectives, including the relay node as a helper to the legitimate communication link [13], or to an eavesdropper [14]. In [15]–[18], the authors consider the case where the relay node itself is the eavesdropper from whom the information transmitted from the source to the destination must be kept secret. This setting, which provides theoretical foundations toward the utilization of untrusted relay nodes in network design, is relevant in practice: The potentially untrusted routers of today's Internet routinely relay sensitive information for its users. The current approach is that the authenticity and secrecy of the information is protected by security protocols assuming these routers are *limited in computational power* [19]. It is interesting to address the role of these routers if they are adversaries with unlimited computation power.

To answer this question, in [16], [17], and [20], as a first step, we considered the case where the relay node was "honest but curious." This means that the curious relay node is not trusted with confidential messages. On the other hand, it is honest in the sense that it conforms to the system rules and performs the designated relaying scheme in every channel use. He and Yener [16] considered the three-node relay network with such a relay. In [17] and [20], the authors considered the two-way relay channel where two nodes could only communicate through such a relay node. In these works, we showed that if the relay was not trusted but honest, recruiting it to help relay information can provide a *higher* secrecy rate than simply treating the relay node as an eavesdropper. This effect is most pronounced in the two-hop model studied in [17], in which the achievable rate is 0 if the relay node is excluded from communication, and increases to being within 1bit of the rate of having trusted relay if the untrusted relay node is properly utilized. Similar observations can be made in networks with multiple confidential messages [18].

It is the next natural step to consider the problem where the relay node is curious and is potentially *dishonest*, i.e., when the relay can deviate from its designated behavior. This can be as benign as the relay node experiencing a failure and stopping

Manuscript received April 01, 2010; revised June 09, 2012; accepted June 13, 2012. Date of publication August 31, 2012; date of current version December 19, 2012. This work was supported in part by the National Science Foundation under Grants CNS 07-16325 and CCF 09-64362 and in part by the DARPA ITMANET Program under Grant W911NF-07-1-0028. This paper was presented in part at the 2009 IEEE International Symposium on Information Theory and in part at the IEEE Information Theory Workshop, Cairo, Egypt, January 2010.

X. He was with the Department of Electrical Engineering, Pennsylvania State University, University Park, PA 16802 USA. He is now with Microsoft, Redmond, WA 98052-6399 USA (e-mail: xianghe@microsoft.com).

A. Yener is with the Department of Electrical Engineering, Pennsylvania State University, University Park, PA 16802 USA (e-mail: yener@ee.psu.edu).

Communicated by S. Ulukus, Associate Editor for Communication Networks.

Digital Object Identifier 10.1109/TIT.2012.2216952

transmission, which is obviously easy to detect. However, if the relay is a malicious entity, or is captured by one, a more detrimental scenario can materialize. Specifically, the relay can attempt to deceive the destination into accepting a counterfeit message by actively manipulating the signals it relays. Such behavior is a "Byzantine attack" [21]. When the adversary is limited in computational power, this type of attack can be detected via message authentication code or digital signatures [19]. The security guarantee promised by these schemes is essentially based on the absence of known effective attack strategies and the fact that their reliability can be proved if a very small set of assumptions is made (see [22, Th. 12.14, Th. 13.11], for example).

In this paper, we tackle the case where the Byzantine adversary has unlimited computational power. In an effort to demonstrate the simplest network which relies on an untrusted node to communicate, we consider a two-hop network [17]. In contrast to [17], which considered an honest but curious relay, we allow the relay node to actively modify the transmitted signal in any way it desires. The goal of the destination thus becomes detecting reliably that the message has been altered, whenever the relay node chooses to do so.

Toward accomplishing this goal, there are several known results that can be leveraged, each with their own limitations. For example, Byzantine attack detection can be viewed as an authentication problem, by treating the counterfeit message as a message from a "wrong" source node. An information-theoretic secrecy scheme with an authentication capability was proposed in [23]. However, like other message authentication codes [22], the source has to share an authentication key with the destination beforehand.

It is known, on the other hand, that to detect the Byzantine attack, which is a milder requirement than authentication, it is not essential to share keys. In [24], the so-called algebraic manipulation detection (AMD) code was used for encoding the data from the source node which ensures the probability that the Byzantine attack succeeds can be made arbitrarily small with an arbitrarily small loss in rate. A limitation of this scheme is that it has to be used along with a secrecy sharing scheme that has a certain *linearity* property [24], a property that is easily fulfilled in a noiseless network as shown in [21] and [25]. Indeed, in [25], we considered a deterministic two-hop network and it was shown that by using an AMD code, the probability that the Byzantine adversary wins decreases exponentially fast with respect to the total number of channel uses n_{total} while the loss in rate can be made arbitrarily small. On the other hand, for *noisy* channels, secret sharing schemes generally fail to have the required linearity property. As a result, to date, the strongest result that could have been obtained is that, for a noisy two-hop network, the probability that a Byzantine attack goes undetected decreases exponentially only with respect to $\sqrt{n_{\text{total}}}$[25].

The main contribution of this study is to demonstrate that for the Gaussian two-hop network, the probability that a Byzantine attack goes undetected, i.e., the adversary wins, can be made to decrease exponentially fast with respect to n_{total}, while the loss in secrecy rate can be made arbitrarily small. Hence, the same result achievable for the deterministic two-hop network is attainable for the *noisy* two-hop network. This represents a departure from traditional security approaches that assume a noiseless bit pipe for communication and brings the physical characteristics of the channel into the picture while providing a guarantee thought to be possible only in the noiseless setting. The key to prove this result is the introduction of a new strong secrecy scheme. Its existence is proved using the representation theorem presented in [11] and [26] and the privacy amplification technique presented in [27] and [28]. Compared to the previously known strong secrecy schemes, the main differences are as follows.

1) Unlike the randomly generated codes in [29], the decoder of the new scheme is linear for certain rate configurations.
2) Unlike [11] and [26], the codeword consists of a single lattice point rather than multiple lattice points. This allows the mutual information between the confidential message and eavesdropper's observation to decrease exponentially with respect to n_{total} as in [30] while still preserving the linear property of the decoder.

The first item provides the linear property required by AMD codes. The exponential decay property in the second item is essential in preserving the Byzantine detection performance offered by AMD codes. As will be explained in Section VI-B, existing strong secrecy methods are not sufficient for these purposes.

The remainder of this paper is organized as follows: In Section II, we describe the system model and formulate the Byzantine detection and secrecy problem. In Section III, we review known Byzantine detection schemes, in particular, the AMD code and describe the technical obstacles to be overcome in this study. Section IV describes the overall architecture of the Byzantine detection scheme proposed in this paper. Sections V–VII examine each component in the scheme in detail and analyze the performance of the scheme. Section VIII concludes this paper.

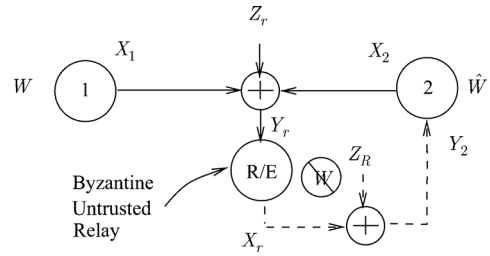

Fig. 1. Gaussian two-hop network. Phase 1 is indicated by solid line, and phase 2 by dashed line. R/E: Relay/Eavesdropper. Y_1 is not shown.

II. System Model and Problem Formulation

The Gaussian two-hop network with a Byzantine relay node is shown in Fig. 1. In this model, node 1 wants to send a confidential message W to node 2. Since it cannot communicate with node 2 directly, it recruits the help of a relay node, who is not trusted with the message W. The signal received by the relay node consists of the signals transmitted by both nodes 1 and 2, and the signal broadcasted by the relay node is heard by both nodes as well. These are fitting assumptions for the wireless medium.

Let $X_i, i = 1, 2, X_r$ denote the signal transmitted by node 1, 2 and the relay. Let $Y_i, i = 1, 2$ and Y_r denote their received signals, respectively. After normalizing the channel gains, we have

$$Y_r = X_1 + X_2 + Z_r \tag{1}$$

$$Y_2 = X_r + Z_R, \quad Y_1 = hX_r + Z_R\prime \tag{2}$$

where Z_r, Z_R, and $Z_R\prime$ are independent Gaussian random variables with zero mean and unit variance, and h is the normalized channel gain. Since Y_1 is not used in the scheme described in this paper, it is omitted in Fig. 1 for clarity. Each node is assumed to be half-duplex. For simplicity, the relay node transmits in half of all channel uses. Without loss of generality, it is assumed that nodes 1 and 2 do not transmit when the relay node transmits since the relay node cannot receive and relay their transmitted signals simultaneously. It is also assumed that during the n channel uses that the relay node transmits, its transmission power averaged over these channel uses should not exceed $\bar{P}$. During the remaining n channel uses that nodes 1 and 2 may transmit, the transmission power of each of these two nodes averaged over these channel uses should not exceed $\bar{P}$.

We assume the Byzantine adversary at the relay node can employ any stochastic function to compute its current transmitted signal. Let $X_{r,i}$ be its transmitted signal at the ith channel use. Let M_r be the local randomness available to the relay node. Let Y_r^{i-1} be the signals it received in the past. Let W be the confidential message it is currently relaying. Let f_i be the relaying function. Then, the attacker (relay) can compute

$$X_{r,i} = f_i(M_r, Y_r^{i-1}, W). \tag{3}$$

It might seem inconsistent at first glance to assume the Byzantine adversary knows the message, which should be kept secret from the relay node in the first place. However, in reality, the secrecy of the message can be broken due to a nontechnical reason, for example, by human error. In that case, the definition (3) will guarantee that the performance of the Byzantine detection scheme is not affected. Through this pessimistic assumption, we are able to claim that the scheme can deal with worst case attacks.

Let the total number of channel uses be $n_{\text{total}} = 2n$, during which each node transmits during n channel uses. Let $\hat{W}$ be the estimate of W computed by the destination, i.e., node 2, based on its observation. Note that because the relay can be a Byzantine adversary, node 2 may or may not accept $\hat{W}$ as a genuine message from node 1 based on certain criteria.

The Byzantine detection problem for secure communication using an untrusted relay can be stated as follows:

Find the secrecy rate R_s of W, defined as

$$R_s = \lim_{n\to\infty} \frac{1}{n_{\text{total}}} H(W) \tag{4}$$

such that the following conditions hold.

1) When the relay node is honest, and W is uniformly distributed over the message set, then both $\Pr\left(W \neq \hat{W}\right)$ and

$$\Pr\left(\hat{W} \text{ is not accepted by Node 2} | W = \hat{W}\right) \tag{5}$$

should decrease exponentially fast with respect to the total number of channel uses. Hence, the transmission of W is *reliable*.

2) For all w_0 in the message set, the probability that the adversary wins, $\Pr(A \text{ wins})$, given by

$$\Pr(A \text{ wins}) = \Pr\left(\hat{W} \text{ is accepted by Node 2} | W = w_0, W \neq \hat{W}\right) \tag{6}$$

should decrease exponentially fast with respect to the total number of channel uses. Hence, any modification on W is detected reliably.[1]

3) $I(W; Y_r^n, X_r^n)$ should decrease exponentially fast with respect to the total number of channel uses. This means the information that the adversary has regarding the value of W is negligible.

Remark 1: Observe that the condition of reliable Byzantine detection in 2) is independent of the distribution of W. □

Remark 2: For the achievable scheme, we develop in the sequel, nodes 1 and 2 do not use the signals they receive in the past to compute the signals they transmit in the future. Consequently, we have $I(W; X_r^n | Y_r^n) = 0$ and $I(W; Y_r^n, X_r^n) = I(W; Y_r^n)$. Hence, in reality, we only need to prove that $I(W; Y_r^n)$ is negligible for 3) to hold. □

Remark 3: It should be noted that the Byzantine detection problem described here is different from that in references [31]–[34]. In these works, the adversaries also actively manipulate the signals received by the destination. The goal in these references is to find a way for reliable communication in the presence of such adversaries carrying out the worst case attack. In the two-hop network considered in this paper, this is not possible since there is no direct link between the two legitimate communicating nodes. Hence, when Byzantine behavior is detected, we need to forgo the relay. □

Remark 4: For simplicity, we use power constraint $\bar{P}$ for all nodes. The Byzantine detection scheme proposed in this paper can be easily adapted to the case where nodes have unequal power constraints, since the coding scheme does not depend on the rate at which the relay node can transmit to the second node. Let $\bar{P}_r$ denote the power constraint of the relay, and let $\bar{P}_i, i = 1, 2$ denote the power constraints of the two source nodes. $\bar{P}_1 \neq \bar{P}_2 \neq \bar{P}_r$. The relay would need more channel uses to transmit if $\bar{P}_r$ is small. When $\bar{P}_1 \neq \bar{P}_2$, the scheme can be applied by treating the power constraints of the two sources to be $\min\{\bar{P}_1, \bar{P}_2\}$. □

III. Known Byzantine Detection Schemes

In this section, we review some known Byzantine detection schemes and explain why they are insufficient for the scenario considered in this study.

[1]When conditioned on a fixed message value w_0, (3) becomes $X_{r,i} = f_i(M_r, Y_r^{i-1}, W = w_0)$. In this case, the set of strategy available to the attacker is identical regardless of whether the Byzantine adversary knows w_0 or not and hence the proof which shows $\Pr(A\ wins)$ is negligible is identical in both cases.

A. Algebraic Watchdog

Kim *et al.* [35] proposed to use the sender of the confidential message to monitor the behavior of the relay node. When the message is confidential as in our setting, using this so-called watchdog is not possible. This is because there is no direct link between the two legitimate communicating nodes which means the sender has no information regarding the signals transmitted by the destination. As will be explained in Section V, these signals are necessary in order to deploy cooperative jamming [7] to keep the message secret from the relay node (see also [17]). Since the received signals at the relay is corrupted by signals transmitted by the destination, so are the signals transmitted from it. This prevents the source from detecting whether the relay misbehaves by just looking at its transmitted signals without the knowledge of the signals transmitted from the destination.

B. AMD Codes

Cramer *et al.* [24] proposed AMD codes, with which the receiver can achieve Byzantine detection. They are defined as follows [24]. Let $\mathcal{GF}(q^r)$ denote a Galois field that has q^r elements, where q is a prime number and r is a positive integer. An AMD codeword is composed of three parts: $\{s, x, h\}$, where s is the $d \times 1$ vector on $\mathcal{GF}(q^r)$ representing the message. The component x is called the random seed and is generated from $\mathcal{GF}(q^r)$ by the encoder itself. h is the hash tag and is computed according to the *hash rule*:

$$h = x^{d+2} + \sum_{i=1}^{d} s_i x^i \tag{7}$$

where s_i is the ith component of s and the addition and multiplication is defined over $\mathcal{GF}(q^r)$. Suppose that node 2 receives $s\prime$, $x\prime$, $h\prime$, where $s\prime \neq s$. Let $\Delta_x = x\prime - x$. $\Delta_h = h\prime - h$. Then, [24] has the following result:

Theorem 1 [24, Th. 2]: Assume at least one of $s\prime - s$, Δ_x, Δ_h is not zero. If the distribution of x conditioned on $\{\Delta_x, \Delta_h, s\prime, s\}$ is uniform over the field $\mathcal{GF}(q^r)$, q being a prime, and $d+2$ is not divisible by q, then the probability that the hash rule (7) holds for $\{s\prime, x\prime, h\prime\}$ is bounded by $\frac{d+1}{q^r}$.

Remark 5: The rate of the AMD code is $\frac{d}{d+2}$. The rate can be made arbitrarily close to 1 by choosing a large enough value for d.

On the other hand, an AMD codeword can be represented by less than $(d+2)r\log_2 q + 1$ bits. Hence, if we fix d and q, the codeword length is a linear function of r. Consequently, for a given code rate, the probability that $\{s\prime, x\prime, h\prime\}$ can pass the hash rule check (7) decreases exponentially fast with respect to the codeword length. □

Despite the excellent performance of the AMD code, applying it in a noisy channel is by no means straight-forward. This is exemplified by the condition in Theorem 1: The distribution of x conditioned on $\{\Delta_x, \Delta_h, s\prime, s\}$ must be uniform over the field $\mathcal{GF}(q^r)$. In a noisy channel, in general, Δ_x and x are not independent. In the two-hop network considered in this paper, this can be seen from the expression of Δ_x. Let g be the decoding function used by node 2. Let Y_2^n be the signal received by node 2 if the relay is honest, and $\tilde{Y}_2^n$ if the relay is dishonest. Assuming the decoding result is correct at all nodes if the relay is honest, Δ_x is given by

$$\Delta_x = x\prime - x \tag{8}$$

$$= g\left(\tilde{Y}_2^n, X_2^n\right) - g\left(Y_2^n, X_2^n\right). \tag{9}$$

By observing (9), we notice that the condition in Theorem 1 can be fulfilled if g is linear in its first parameter and $\tilde{Y}_2^n - Y_2^n$ is independent from x. In general, it is difficult for a decoding function g to be linear without a specifically designed signaling scheme.

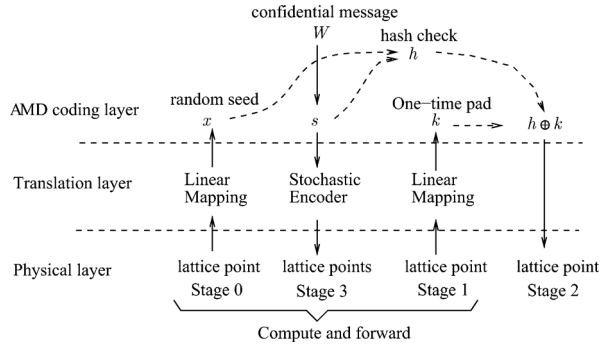

Fig. 2. Architecture of the Byzantine detection scheme.

IV. Architecture

This section provides an architectural overview of the Byzantine detection scheme proposed in this paper. As shown in Fig. 2, conceptually the scheme can be divided into three layers.

1) The bottom layer is the physical layer responsible for signals from node 1 to node 2 to be transmitted reliably. For this purpose, the relay is asked to perform compute-and-forward [36], which we shall review in Section V. In this scheme, signals are transmitted in the unit of so-called lattice points [37]. For now, for conceptual ease of understanding, these lattice points can be viewed as elements from a Galois field $\mathcal{GF}(q^N)$.
2) Since the compute-and-forward scheme by itself is generally insufficient to ensure the secrecy of the data, a middle layer, called "translation layer," is introduced to achieve secrecy. In the translation layer, data from upper layers are mapped to lattice points so that when these lattice points are transmitted by the physical layer through the relay, the eavesdropper at the relay can obtain little information about the data. Furthermore, the mapping is designed such that the linearity property required by AMD codes, as described in Section III, can be satisfied.
3) The top layer, the AMD coding layer, takes the confidential message W in the format of d elements from $\mathcal{GF}(q^r)$ for some positive integer $r < N$. This layer then computes the AMD codeword from W, which becomes the s component in the AMD codeword, as described in Section III. The

resulting AMD codeword is then sent to the translation layer for transmission.

On the time axis, the transmission of an AMD codeword can be divided into the following four stages (see Fig. 2).

1) The zeroth stage generates a random seed, which is the component x in an AMD codeword (s, x, h). $x \in \mathcal{GF}(q^r)$.
2) The first stage generates a one-time pad k, which will be used to secure the hash tag h. $k \in \mathcal{GF}(q^r)$.
3) The second stage transmits $u = h \oplus k$, where $\oplus$ is the addition operation defined over $\mathcal{GF}(q^r)$.
4) The third stage transmits s.

During stages 0, 1, and 3, we use compute-and-forward relaying at the physical layer. During stage 2, since h is already secured by the one-time pad k, a conventional two-hop decode-and-forward relay scheme is sufficient to transmit u. Hence, in this stage, node 2 remains silent.

The zeroth stage, during which the random seed x is generated, is composed of the following steps.

1) A lattice point t^N is chosen from the nested lattice codebook, which we shall define in Section V, according to a uniform distribution and transmitted to node 2 through the physical layer using compute-and-forward.
2) x is computed by the translation layer through $x = \mathbf{g}(t^N)$, where $\mathbf{g}$ is a linear mapping that maps a transmitted lattice point t^N from $\mathcal{GF}(q^N)$ to $\mathcal{GF}(q^r)$. We shall prove in Theorem 2 in Section VI that there exists such a linear mapping that preserves the confidentiality of x against the eavesdropper.

The same steps are used to generate the one-time pad k in stage 1.

The third stage, during which s is transmitted, is composed of the following steps.

1) First s is mapped to an element in a finite field, which is represented by $\mathbf{S}_{r_0 \times 1}$ in (39).
2) The translation layer then maps the field element to a lattice point t_1^N using a stochastic encoder, which will be described explicitly in (39).
3) t_1^N is then transmitted by the physical layer using compute-and-forward relaying.

Node 2 computes its estimate for x, k, and u, denoted by $\hat{x}$, $\hat{k}$, and $\hat{u}$ respectively, using the decoder offered by the physical layer. The estimate for h, $\hat{h}$, is computed from $\hat{u} \oplus (-\hat{k})$. It then accepts $\hat{s}$ as genuine if $\hat{s}$, $\hat{x}$, $\hat{h}$ satisfies (7) by substituting s, x, h with $\hat{s}$, $\hat{x}$, $\hat{h}$.

Remark 6: Here, the transmission of h uses the idea of message authentication codes with key manipulation security in [24, Sec. 4]. Note that for a given s, the distribution of hash tag h is in general not uniform. Hence, the distribution of h depends on the distribution of s. However, as we shall see in Section VI-A, if we want to use the strongly secure scheme in Section VI-A to transmit h and desire to fix the hash function $\mathbf{G} = \mathbf{g}$, we need to know the distribution of h beforehand, which is difficult since the distribution of s is hard to determine beforehand. To solve this problem, we introduce another random seed k from $\mathcal{GF}(q^r)$, which can be generated via the linear coding scheme in Section VI-A. From Lemma, k is uniformly distributed over $\mathcal{GF}(q^r)$. Hence, h can be transmitted by using k as a one-time pad. □

V. Compute-and-Forward

In this section, we review the compute-and-forward scheme that uses nested lattice codes. This scheme was used in [36] for a Gaussian two-way relay channel without eavesdroppers. Later, it was used in [26] as a building block to transmit confidential messages when the relay is honest but curious, i.e., is an eavesdropper but not a Byzantine adversary. In this study, we use this relay method in the physical layer as part of the Byzantine detection scheme.

We begin by introducing some notations for the nested lattice structure: For a lattice Λ_c, the modulus operation $x \bmod \Lambda_c$ is defined as $x \bmod \Lambda_c = x - \arg\min_{t \in \Lambda_c} d(x,t)$, where $d(x,t)$ is the Euclidean distance between x and t. The fundamental region of a lattice $\mathcal{V}(\Lambda_c)$ is defined as the set $\{x : x \bmod \Lambda_c = x\}$. A pair of N-dimensional lattices $\{\Lambda, \Lambda_c\}$ is said to have a nested structure if $\Lambda_c \subset \Lambda$[37].

Now consider a pair of N-dimensional nested lattice pair $\{\Lambda, \Lambda_c\}$ which is designed as in [37]. The signal transmitted by each node is given by

$$X_i^N = \left(t_i^N + d_i^N\right) \bmod \Lambda_c, \quad i = 1, 2 \tag{10}$$

where $t_i^N \in \Lambda \cap \mathcal{V}(\Lambda_c)$, and $d_i^N, i = 1, 2$ are two fixed vectors in $\mathcal{V}(\Lambda_c)$ and are known by the relay node. For our purpose, t_1^N will be computed from the confidential message. t_2^N is independent from t_1^N and is chosen from $\Lambda \cap \mathcal{V}(\Lambda_c)$ according to a uniform distribution. As a result, $X_2^N = (t_2^N + d_2^N) \bmod \Lambda_c$ serves as the cooperative jamming signal to confuse the untrusted relay node.

An honest relay node will then decode $(t_1^N + t_2^N) \bmod \Lambda_c$ and transmit $(t_1^N + t_2^N + d_3^N) \bmod \Lambda_c$ during phase two, where d_3^N is a fixed vector in $\mathcal{V}(\Lambda_c)$ and is known by node 2. Node 2 then decodes $\hat{t}^N = (t_1^N + t_2^N) \bmod \Lambda_c$ from the signal it received during phase two. An estimate of t_1^N, denoted by $\hat{t}_1^N$, is then computed from $(\hat{t}^N - t_2^N) \bmod \Lambda_c$.

Define $|\mathcal{S}|$ be the cardinality of a set $\mathcal{S}$. Define R_0 as

$$R_0 = \frac{1}{N} \log_2 |\Lambda \cap \mathcal{V}(\Lambda_c)|. \tag{11}$$

Define P as the average transmission power per dimension of the nested lattice $\Lambda \cap \mathcal{V}(\Lambda_c)$:

$$P = \frac{1}{N \, \mathbf{vol}(\mathcal{V}(\Lambda_c))} \int_{x \in \mathcal{V}(\Lambda_c)} \|x\|_2^2 \, dx \tag{12}$$

where $\|x\|_2$ denote the Euclidean distance between x and 0 in N dimensional real space $\mathbf{R}^N$. Then, it was shown in [38] that if

$$R_0 < \frac{1}{2} \log_2 \left(\frac{1}{2} + P\right) \tag{13}$$

the probability $\Pr(\hat{t}_1^N \neq t_1^N)$ decreases exponentially with respect to N.

Remark 7: It is clear that if the relay chooses to transmit $(t_3^N + d_3^N) \bmod \Lambda_c$ for some arbitrary $t_3^N \in \Lambda \cap \mathcal{V}(\Lambda_c)$, then

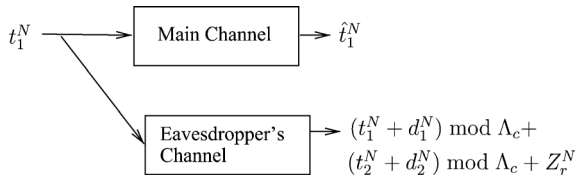

Fig. 3. Lattice input wiretap channel.

node 2 will be forced to accept a message that does not originate from node 1. This shows that unless some proper measures are taken, a Byzantine adversary can easily succeed in this scenario. □

Remark 8: $d_i^N, i = 1,2,3$, are conventionally defined as random variables uniformly distributed over $\mathcal{V}(\Lambda_c)$[37]. The reason of defining them to be random is that it is easier to analyze the average error performance of an ensemble of lattice code books parameterized by the dithering vectors than to analyze the error performance of a specific lattice code book [39]. However, from the result on the average performance, we can also claim that there must exist some fixed $d_i^N, i = 1,2,3$, which corresponds to fixed lattice codebooks in the ensemble, and these $d_i^N, i = 1,2,3$ also provide vanishing error probability and meet the average power constraints [11]. Hence, in the sequel, we assume $d_i^N, i = 1,2,3$, are fixed. □

VI. Using Nested Lattice Codes to Provide Strong Secrecy

The channel input–output relationship implied by the compute-and-forward scheme described in Section V is summarized by the *lattice input wiretap channel* shown in Fig. 3.[2] The main channel takes input $t_1^N \in \Lambda \cap \mathcal{V}(\Lambda_c)$, and produces output $\hat{t}_1^N$. The eavesdropper channel also takes input t_1^N, and has the same observation as the signals received by the relay node in the two-hop network. It is clear that the eavesdropper observation is not independent from t_1^N. Hence, an additional measure is necessary to secure the data if this scheme is used to transmit confidential message. This operation, as mentioned in Section IV, is carried out in the translation layer and is described in this section.

A. Strongly Secure Scheme

1) When $\Lambda_c = q\Lambda$ for a Prime q: The self-similar nested lattice code with a prime nesting ratio, i.e., $\Lambda_c = q\Lambda$, is a special case of the good nested lattice ensemble proposed in [37, Sec. 7]. We first consider this case since when q is a prime, the set $(\Lambda + d^N) \cap \mathcal{V}(\Lambda_c)$ is isomorphic to a finite field, as shown by the following lemma.

Lemma 1: When $\Lambda_c = q\Lambda$ for a prime q and the generation matrix of Λ has full rank, $((\Lambda + d^N) \cap \mathcal{V}(\Lambda_c); +)$ is isomorphic[3] to $(\mathcal{GF}(q^N); +)$.

[2]The only difference from the original two-hop network is that in the two-hop network, it takes another N channel uses for the relay to relay the lattice point to node 2 during which node 1 and 2 do not transmit. Here, to simplify the argument, we omit this detail and will take these additional channel uses into account when we revisit the two-hop network in Section VII.

[3]This is a group isomorphism over the addition operation. It is not a field isomorphism.

Proof: The proof is provided in Appendix A. ■

Remark 9: The isomorphism in Lemma 1 is not affected by the choice of d. The fixed dithering vector d is simply used to constrain the average power of the lattice code book. □

As we will show later in the proof of Theorem 2, the isomorphism property proved by Lemma 1 allows the resulting decoder to be linear and proves to be of critical importance in the Byzantine detection scheme in Section VII.

The next theorem declares the existence of the strong secrecy scheme.

Theorem 2: For a given constant $\varepsilon > 0$ that can be arbitrarily small, assume q is a prime large enough such that

$$1 - \frac{1+\varepsilon}{\log_2 q} > 0. \tag{14}$$

Then for an integer r, such that

$$0 \le r \le N\left(1 - \frac{1+\varepsilon}{\log_2 q}\right) \tag{15}$$

there exists a linear mapping $\mathbf{g}$ from $\mathcal{GF}(q)^N$ to $\mathcal{GF}(q)^r$ such that

1) $\mathbf{g}$ has full row rank r;
2) when $t_i^N, i = 1,2$ are uniformly distributed over $(\Lambda + d_i^N) \cap \mathcal{V}(\Lambda_c)$ and are independent of each other, there exists a positive constant β such that

$$I\left(\mathbf{g}\left(t_1^N\right); \bar{Y}_r^N\right) \le 2e^{-\beta N}. \tag{16}$$

Before proving the theorem, we need several supporting results.

First, the following representation theorem from [26] is useful.

Theorem 3 [26]: For any u_1, u_2, such that $u_i \in \mathcal{V}(\Lambda_c), i = 1,2$, there exists a function mapping from $\sum_{k=1}^2 u_k$ to an integer T in $[1, 2^N]$ such that $\sum_{k=1}^2 u_k$ is uniquely determined given the value of the function, T, and $(\sum_{k=1}^2 u_k) \bmod \Lambda_c$. Theorem 3 can be used to prove the following result. Define $\bar{Y}_r^N$ as

$$\bar{Y}_r^N = (t_1^N + d_1^N) \bmod \Lambda_c + (t_2^N + d_2^N) \bmod \Lambda_c \tag{17}$$

which is obtained by subtracting the channel noise Z_r^N from Y_r^N. Then, based on Theorem 3, $\bar{Y}_r^N$ in (17) can be represented by $\{(\sum_{i=1}^2 (t_i^N + d_i^N)) \bmod \Lambda_c, T\}$. Since $d_i^N, i = 1,2$, are known by each node, this means $\bar{Y}_r^N$ in (17) can be represented by $\{(t_1^N + t_2^N) \bmod \Lambda_c, T\}$.

We also need the following lemma which says most matrices have full rank.

Lemma 2: Let $\mathbf{G}$ be taken from the set of linear mappings from $\mathcal{GF}(q)^N$ to $\mathcal{GF}(q)^r$ according to a uniform distribution. Hence, $\mathbf{G}$ can be represented as a matrix over $\mathcal{GF}(q)$ with r rows and N columns. The probability that $\mathbf{G}$ has full row rank is greater than $1 - q^{r-N}$.

Proof: The lemma can be derived from [40, Lemma 6]. A self-contained proof is provided in Appendix B for the reader's convenience. ■

Finally, we need the following results on privacy amplification [27], which we state here for completeness.

Definition 1 [27, Definition 1]: A set of functions $\mathcal{A} \to \mathcal{B}$ is a class of *universal hash function* if for a function g taken from the set according to a uniform distribution, and $x_1, x_2 \in \mathcal{A}, x_1 \neq x_2$, the probability that $g(x_1) = g(x_2)$ holds is at most $1/|\mathcal{B}|$.

We next state the results based on these definitions.

Lemma 3 [27]: The set of linear mappings as defined in Lemma 2 is a class of universal hash function.

Theorem 4[27, Corollary 4]: Let $\mathbf{G}$ be selected according to a uniform distribution from a class of universal hash function from $\mathcal{A}$ to $\mathcal{GF}(q)^r$. For two random variables A, B, A being defined over $\mathcal{A}$, if for a constant c, $H_2(A|B=b) > c$, then

$$H(\mathbf{G}(A)|\mathbf{G}, B=b) > r\log_2 q - \frac{2^{r\log_2 q - c}}{\ln 2} \tag{18}$$

where $H_2(X) = -\log_2 \sum_x \Pr(X=x)^2$ is the Rényi entropy and $H(X)$ denotes the Shannon entropy.

The following theorem provides a bound on the decrease of Rényi entropy given that a finite number of bits are revealed.

Theorem 5 [41, Th. 5.2] [28, Lemma 3]: Let X and Q be random variables. Let $\mathcal{Q}$ be the alphabet set of Q. Let $s > 0$. Then with probability at least $1 - 2^{-(s/2-1)}$, we have $H_2(X) - H_2(X|Q=q) \leq \log_2 |\mathcal{Q}| + s$.

Denote $(a+b) \bmod \Lambda_c$ with $a \oplus b$ for notational simplicity. With these preparations, we are now ready to prove Theorem 2:

Proof of Theorem 2: For the distribution for $t_i^N, i = 1, 2$ stated in Theorem 2, $t_1^N \oplus t_2^N$ is independent from t_1^N. Therefore, we have

$$H_2\left(t_1^N | t_1^N \oplus t_2^N = t^N\right) = H_2\left(t_1^N\right) = N\log_2 q. \tag{19}$$

Let T be the integer defined in Theorem 3. Then, according to Theorem 5, for a given integer a, $1 \leq a \leq 2^N$ and $t^N \in \Lambda \cap \mathcal{V}(\Lambda_c)$, with probability $1 - 2^{-(s/2-1)}$:

$$\begin{aligned} & H_2\left(t_1^N | t_1^N \oplus t_2^N = t^N, T=a\right) && (20)\\ & \geq H_2\left(t_1^N | t_1^N \oplus t_2^N = t^N\right) - \log_2 |T| - s && (21)\\ & = N(\log_2 q - 1) - s. && (22)\end{aligned}$$

Thanks to Lemma 1, an element t_1^N can be identified to an element of $\mathcal{GF}(q^N)$, which is isomorphic to $\mathcal{GF}(q)^N$; therefore, with a slight abuse of notation we write $\mathbf{G}(t_1^N)$ in place of $\mathbf{G}(\mathbf{I}(t_1^n))$, where $\mathbf{I}$ is the group isomorphism between $\Lambda \cap \mathcal{V}(\Lambda_c)$ and $\mathcal{GF}(q)^N$.

According to Lemma 3, $\mathbf{G}$ is a universal hash function. Hence, according to Theorem 4, we have

$$\begin{aligned} & H\left(\mathbf{G}\left(t_1^N\right)|\mathbf{G}, t_1^N \oplus t_2^N = t^N, T=a\right) \\ & \geq r\log_2 q - \frac{2^{r\log_2 q - c}}{\ln 2} \end{aligned} \tag{23}$$

where c is given by (22)

$$c = N(\log_2 q - 1) - s. \tag{24}$$

Since depending on the value of t^N and a (22) holds with probability $1 - 2^{-(s/2-1)}$, from (23), we have

$$\begin{aligned} & H\left(\mathbf{G}\left(t_1^N\right)|\mathbf{G}, t_1^N \oplus t_2^N, T\right) \\ & \geq \left(1 - 2^{-(s/2-1)}\right)\left(r\log_2 q - \frac{2^{r\log_2 q - c}}{\ln 2}\right). \end{aligned} \tag{25}$$

Note that

$$H(\mathbf{G}(t_1^N)|\mathbf{G}) \leq r\log_2 q. \tag{26}$$

Hence, in order for $I(\mathbf{G}(t_1^N); t_1^N \oplus t_2^N, T|\mathbf{G})$ to be negligible, we expect $2^{-(s/2-1)}$ and $2^{r\log_q - c}$ to decrease exponentially with respect to N. To achieve this, we choose $s = \varepsilon' N$, where $0 < \varepsilon' < \log_2 q - 1$ so that c in (24) is positive. We choose r such that for $\delta > 0$

$$\begin{aligned} r\log_2 q &< c - N\delta && (27)\\ &= N(\log_2 q - 1) - s - N\delta && (28)\\ &= N(\log_2 q - 1 - \varepsilon' - \delta). && (29)\end{aligned}$$

We observe that if (27)–(29) are satisfied, $2^{r\log_q - c}$ to decrease exponentially with respect to N. We also observe that if we let $\varepsilon = \varepsilon' + \delta$, then (27)–(29) lead to (15).

For these choices of r and s, from (25) and (26), we observe that there exists $\beta > 0$, such that

$$I\left(\mathbf{G}\left(t_1^N\right); t_1^N \oplus t_2^N, T|\mathbf{G}\right) \leq e^{-\beta N}. \tag{30}$$

We next use the fact that for sufficiently large N, most $\mathbf{G}$s have full row rank as shown in Lemma 2. Therefore, for a uniform distribution for $t_i^N, i = 1, 2, t_1^N$ and t_2^N being independent, there must exists a $\mathbf{G} = \mathbf{g}$, such that

1) $\mathbf{g}$ has full rank;
2) from the Markov inequality

$$I\left(\mathbf{G}\left(t_1^N\right); t_1^N \oplus t_2^N, T|\mathbf{G} = \mathbf{g}\right) \leq 2e^{-\beta N}. \tag{31}$$

Finally, we use Theorem 3 which says $t_1^N \oplus t_2^N, T$ in (31) can be replaced by $\bar{Y}_r^N$. Hence, we have proved Theorem 2. ■

The secrecy generation scheme described above is useful only if the generated random variable, $\mathbf{g}(t_1^N)$, can serve as the random seed, x, in the AMD tuple as described in Section III. Hence, we need the following lemma on the distribution of $\mathbf{g}(t_1^N)$.

Lemma 4: If t_1^N is uniformly distributed over $\mathcal{GF}(q^N)$, and $\mathbf{g}$ has full row rank, Then, $\mathbf{g}(t_1^N)$ is uniformly distributed over $\mathcal{GF}(q^r)$.

Proof: Since $\mathbf{g}$ has full row rank, and its elements are taken from the field $\mathcal{GF}(q)$, it can always be represented as

$$\mathbf{g} = [\mathbf{I}, \mathbf{P}]\mathbf{O} \tag{32}$$

where $\mathbf{O}$ is an $N \times N$ invertible matrix, $\mathbf{P}$ is an $r \times (N-r)$ matrix. Hence, $\mathbf{O}(t_1^N)$ is uniformly distributed over $\mathcal{GF}(q^N)$. $\mathbf{I}$ is an $r \times r$ identity matrix. Since the sum of any two independent field elements will be uniformly distributed if one of the field element is uniformly distributed [39, Lemma 2], it can be verified that $\mathbf{g}(t_1^N)$ is uniformly distributed over $\mathcal{GF}(q^r)$. ■

2) General Case: When (Λ, Λ_c) does not have the self-similar relationship as described in Section VI-A1, we can still extract a strongly secure random variable from a lattice point using the same method as shown in Section VI-A1. The only difference is that the map between the extracted random variable and the lattice point will not be linear.

Consider a general N-dimensional nested lattice codebook $\Lambda \cap \mathcal{V}(\Lambda_c)$. Recall that R_0, as defined in (11), is the rate of the codebook. Assume $R_0 > 1$. Let $\lfloor x \rfloor$ be the operation that rounds x to the nearest integer less than or equal to x. Define N_0 as

$$N_0 = \lfloor \log_2 |\Lambda \cap \mathcal{V}(\Lambda_c)| \rfloor. \tag{33}$$

Then

$$N_0 \geq NR_0 - 1. \tag{34}$$

Choose any subset K of the codebook $(\Lambda + d_1^N) \cap \mathcal{V}(\Lambda_c)$ that yields the minimal average decoding error probability with the lattice decoder and has size $|K| = 2^{N_0}$. Define v as a bijective mapping from K to $\mathcal{GF}(2^{N_0})$. Then, we have the following theorem.

Theorem 6: Let $\varepsilon > 0$ be a constant such that

$$R_0 - 1 - \varepsilon > 0. \tag{35}$$

Then, for an integer r_0, such that

$$0 \leq r_0 \leq N(R_0 - 1 - \varepsilon) \tag{36}$$

there exists a linear mapping $\mathbf{g}$ from $\mathcal{GF}(2)^{N_0}$ to $\mathcal{GF}(2)^{r_0}$ such that

1) $\mathbf{g}$ has full row rank r_0;
2) when t_1^N is uniformly distributed over K, t_2^N is uniformly distributed over $(\Lambda + d_2^N) \cap \mathcal{V}(\Lambda_c)$, t_1^N, t_2^N are independent of each other, we have

$$I\left(\mathbf{g}\left(v(t_1^N)\right); \bar{Y}_r^N\right) \leq 2e^{-\beta N} \tag{37}$$

for a certain $\beta > 0$.

Proof: The proof is similar to that of Theorem 2, and is given in Appendix C. ■

3) Encoder Construction: Although both Theorem 2 and Theorem 6 can be used to prove the existence of an encoder with rate arbitrarily close to $\max\{R_0 - 1, 0\}$, with R_0 defined in (11), only Theorem 6 is used in the sequel to transmit confidential messages. Theorem 2 is only used to generate strongly secure random seeds, for which Theorem 2 is sufficient by itself. Hence, in this section, we discuss Theorem 6 only. The argument we use is as follows.

For a given $\mathbf{g}$ that has full row rank, let $\mathbf{g}'$ be an $(N_0 - r_0) \times N_0$ matrix such that $\begin{bmatrix} \mathbf{g}' \\ \mathbf{g} \end{bmatrix}$ is a square matrix that is invertible. Define $\mathbf{S}$ and $\mathbf{S}'$ such that

$$\begin{bmatrix} \mathbf{g}'_{(N_0 - r_0) \times N_0} \\ \mathbf{g}_{r_0 \times N_0} \end{bmatrix} v(t_1^N) = \begin{bmatrix} \mathbf{S}'_{(N_0 - r_0) \times 1} \\ \mathbf{S}_{r_0 \times 1} \end{bmatrix}. \tag{38}$$

Then, $\mathbf{S} = \mathbf{g}(v(t_1^N))$. Define $\mathbf{A}$ as the inverse of $\begin{bmatrix} \mathbf{g}' \\ \mathbf{g} \end{bmatrix}$, then the encoder is given by

$$t_1^N = v^{-1}\left(\mathbf{A}\begin{bmatrix} \mathbf{S}'_{(N_0 - r_0) \times 1} \\ \mathbf{S}_{r_0 \times 1} \end{bmatrix}\right) \tag{39}$$

where $\mathbf{S} \in \mathcal{GF}(2^{r_0})$ be the input to the encoder. We assume $\mathbf{S}$ is uniformly distributed over $\mathcal{GF}(2^{r_0})$. $t_1^N \in \Lambda \cap \mathcal{V}(\Lambda_c)$ is the output of the encoder. $\mathbf{S}'$ represents the randomness in the encoding scheme. We observe that, if $\{\mathbf{S}'_{(N_0 - r_0) \times 1}, \mathbf{S}_{r_0 \times 1}\}$ is uniformly distributed over $\mathcal{GF}(2)^{N_0}$ and (39) is used as the encoder, t_1^N is also uniformly distributed over the set K. Since $\mathbf{G} = \mathbf{g}$ is chosen when t_1^N has a uniform distribution over K, this means that when (39) is used as an encoder, the secrecy constraint in Theorem 6, (37), still holds.

Since the encoder (39) uses N channel uses to transmit a $r_0 \times 1$ binary vector, the rate achieved by the encoder is

$$R_e = [R_0 - 1 - \varepsilon]^+ \tag{40}$$

where $[x]^+$ equals x if $x \geq 0$ or 0 otherwise. According to (13), this means R_e can be arbitrarily close to[4]

$$\left[\frac{1}{2}\log_2\left(\frac{1}{2} + P\right) - 1\right]^+. \tag{41}$$

B. Comparison With Other Wiretap Coding Schemes

Although this work leverages the same technique, namely, privacy amplification as [28], it is distinct from [28] in the following aspects:

Maurer and Wolf [28] proposed that one can invoke any weakly secure scheme multiple times and extract a strongly secure key using privacy amplification. Let $\Theta(x)$ denote the set of functions $ax + b, a > 0, b \neq 0$, and a, b are constants. In our model, each invocation of the weakly secure scheme involves $\Theta(N)$ channel uses, where N is the dimension of the lattice code. Suppose this scheme is invoked for M times. Then, the total number of channel uses is MN. Let K denote the generated key and Y_r^{MN} be the signals observed by the eavesdropper; then it can be shown by following [28] that[5]

$$\lim_{M \to \infty} -\frac{1}{M}\log_2 I\left(K; Y_r^{MN}\right) > 0. \tag{42}$$

In this paper, $\mathbf{g}(t_1^N)$ in Theorem 2 can be viewed as the strongly secure key. Based on Theorem 2, we have

$$\lim_{N \to \infty} -\frac{1}{N}\log_2 I\left(K; Y_r^N\right) > 0. \tag{43}$$

By comparing (43) with (42), we observe their relationship depends on whether the lattice dimension N can be kept as a constant. If N is to be kept as a constant, then another layer of error correction code must be used as an outer code to correct the errors from the inner code which is the nested lattice code. The

[4]Due to the structural limitation, (41) is not known to be achievable via self-similar nested lattices.

[5]To obtain (42), it is necessary to replace the weak typicality notion in [28] with strong typicality.

redundancy introduced by the outer code is called *error reconciliation bits* in [28]. Doing so leads to a negligible loss in secrecy rates and does not affect (42). However, it should be noted that these redundancy bits can also be modified by the Byzantine adversary, which must be detected reliably. Yet, introducing the outer code renders the decoding operation to be nonlinear, since the decoder of most error correction code entails nonlinear operations, which, as we have mentioned earlier in Section III-B, makes it difficult to prove the effectiveness of AMD codes. This is the major obstacle that prevents a direct application of [28] to our problem and motivates devising a different strong secrecy scheme given in this section.

VII. Byzantine Detection

In this section, we analyze the performance of the Byzantine detection scheme.

A. Notation

Recall that, as described in Section IV, the transmission is divided into four stages.

The notations for the average transmission power for each stage are defined as follows. Recall that stage 0 and 1 are used to generate x and k using an N-dimensional lattice code as shown in Section VI-A1. We shall use P_1 to denote the average power per channel use for these two stages. Stage 2 transmits $u = h \oplus k$ via the conventional decode-and-forward two-hop protocol. For such a protocol, we can use r-dimensional lattice codes whose rate equals $\log_2 q$ bits per channel use. Let the average transmission power for this stage be P_2. Stage 3 transmits s via the encoder described in Section VI-A2, whose average transmission power per channel use is P. We choose P as $P = \bar{P}(1 - \varepsilon_P)$, where ε_P is a positive constant that can be made arbitrarily small and $\bar{P}$ is the overall transmission power limit defined in Section II.

We next introduce the notations for the signals associated with each stage: $X_i(j), i = 1, 2, X_r(j)$ denote the signals transmitted by node $1, 2$ and the relay during the jth stage, $j = 0, \ldots, 3$. Similarly, $Y_i(j), i = 1, 2,\ Y_r(j), Z_r(j), Z_R(j)$ denote the signals and channel noise observed during the jth stage. $\hat{X}_r(i), i = 0, \ldots, 3$ denotes the estimate for $X_r(i)$ computed by node 2. To simplify the notation, we omit the superscript for these signals which were used to indicate their dimensions.

Remark 10: Note that both P_1 and P_2 are only functions of the rate of their respective lattice code, which is $\log_2 q$. Hence, P_1 and P_2 are only functions of q. Therefore, we can increase r, while leaving P_1, P_2 unchanged. □

B. Performance Analysis

We next derive the following important lemma which implies the condition of AMD code stated in Theorem 1 can be fulfilled using the transmission scheme described in Section IV.

Lemma 5: Let s_0 be any $d \times 1$ vector on $\mathcal{GF}(q^r)$. Then

$$I\left(x; \Delta_x, \Delta_h, \hat{s} | s = s_0\right) < 4 \exp(-\beta N) \tag{44}$$

where β is a positive number defined in Theorem 2.

Proof: The proof of Lemma 5 is based on the strong secrecy offered by Theorem 2 and Theorem 6, and is provided in Appendix D. ■

Remark 11: Lemma 5 implies that

$$I\left(x; \Delta_x, \Delta_h, \hat{s} | s\right) < 4 \exp(-\beta N). \tag{45}$$

Since $I(x; s) = 0$, this means

$$I\left(x; \Delta_x, \Delta_h, \hat{s}, s\right) < 4 \exp(-\beta N). \tag{46}$$

□

Remark 12: Note that $I\left(x; \Delta_x, \Delta_h, \hat{s} | s = s_0\right)$ does not depend on the error exponents of the lattice decoder. Also, it does not depend on whether s_0 is known by the attacker beforehand. □

We next link Lemma 5 and Theorem 1 with Pinsker's inequality which leads to the following *main result* of this paper.

Theorem 7: For the Gaussian two-hop network, for a rate smaller but arbitrarily close to

$$0.5 \left[\frac{1}{2} \log_2 (\frac{1}{2} + \bar{P}) - 1\right]^+ \tag{47}$$

and a total number of channel uses $2n = \Theta(N)$.

1) When the relay is honest, the confidential message W can be transmitted at this rate such that all the three terms $\Pr(W \neq \hat{W})$, $I(W; Y_r^n)$ and
$$\Pr\left(\hat{W} \text{ is not accepted by Node } 2 | W = \hat{W}\right) \tag{48}$$
decrease exponentially fast with N.
2) When the relay is not honest, the probability that the Byzantine attack goes undetected, i.e., the probability that the adversary wins, denoted as $\Pr(A \text{ wins})$ in (6), decreases exponentially fast with N.

Proof: We use "HRH" for "hash rule holds" for the AMD code tuple s', x', h'

$$h' = x'^{d+2} + \sum_{i=1}^{d} s'_i x'^i. \tag{49}$$

This means the message s', x', h' will be accepted by node 2. Hence, the probability that the adversary wins is given by

$$\Pr\left(A \text{ wins}\right) = \sum_{\substack{x, \Delta_x \\ \Delta_h, s' \neq s_0}} \Pr\left(\text{HRH} | x, \Delta_h, \Delta_x, s = s_0, s'\right) \Pr\left(x | \Delta_h, \Delta_x, s = s_0, s'\right) \Pr\left(\Delta_h, \Delta_x, s' | s = s_0\right). \tag{50}$$

Define $Q(A \text{ wins})$ as the term (50) with $\Pr\left(x | \Delta_h, \Delta_x, s = s_0, s'\right)$ replaced by $\Pr(x)$.

$$Q\left(A \text{ wins}\right) = \sum_{\substack{x, \Delta_x \\ \Delta_h, s' \neq s_0}} \Pr\left(\text{HRH} | x, \Delta_h, \Delta_x, s = s_0, s'\right) \Pr\left(x\right) \Pr\left(\Delta_h, \Delta_x, s' | s = s_0\right). \tag{51}$$

Note that $Q\left(A \text{ wins}\right)$ would be the probability that the Byzantine adversary wins if x and $\Delta_h, \Delta_x, s, s'$ are truly independent. To evaluate the effect of being otherwise, we next bound the difference between $\Pr\left(A \text{ wins}\right)$ and $Q\left(A \text{ wins}\right)$

$$\left|\Pr\left(A \text{ wins}\right) - Q\left(A \text{ wins}\right)\right| \tag{52}$$

$$\leq \sum_{\substack{x,\Delta_x \\ \Delta_h, s' \neq s_0}} \begin{array}{l} \Pr\left(\text{HRH}|x, \Delta_h, \Delta_x, s = s_0, s'\right) \times \\ \Pr\left(\Delta_h, \Delta_x, s'|s = s_0\right) \times \\ \left|\Pr\left(x|\Delta_h, \Delta_x, s = s_0, s'\right) - \Pr\left(x\right)\right| \end{array}. \tag{53}$$

Equation (53) is upper bounded by

$$\sum_{\substack{x,\Delta_x \\ \Delta_h, s' \neq s_0}} \begin{array}{l} \left|\Pr\left(x|\Delta_h, \Delta_x, s = s_0, s'\right) - \Pr\left(x\right)\right| \times \\ \Pr\left(\Delta_h, \Delta_x, s'|s = s_0\right) \end{array} \tag{54}$$

$$= \sum_{\substack{x,\Delta_x \\ \Delta_h, s' \neq s_0}} \begin{array}{l} \left|\Pr\left(x|\Delta_h, \Delta_x, s', s = s_0\right) - \Pr\left(x|s = s_0\right)\right| \\ \times \Pr\left(\Delta_h, \Delta_x, s'|s = s_0\right) \end{array} \tag{55}$$

$$= \sum_{\substack{x,\Delta_x \\ \Delta_h, s' \neq s_0}} \begin{array}{l} \left|\Pr\left(x, \Delta_h, \Delta_x, s'|s = s_0\right) - \right. \\ \left. \Pr\left(x|s = s_0\right) \Pr\left(\Delta_h, \Delta_x, s'|s = s_0\right)\right| \end{array} \tag{56}$$

$$\leq \sum_{\substack{x,\Delta_x \\ \Delta_h, s'}} \begin{array}{l} \left|\Pr\left(x, \Delta_h, \Delta_x, s'|s = s_0\right) - \right. \\ \left. \Pr\left(x|s = s_0\right) \Pr\left(\Delta_h, \Delta_x, s'|s = s_0\right)\right| \end{array}. \tag{57}$$

Then, we use Pinsker's inequality [42, Th. 2.33]

$$I(A;B) \geq \frac{1}{2\ln 2}\left(\sum_{A,B} \left|p\left(A,B\right) - p\left(A\right)p\left(B\right)\right|\right)^2. \tag{58}$$

Let $p(A)$ be $\Pr(x|s = s_0)$. Let $p(B)$ be $\Pr(\Delta_h, \Delta_x, s'|s = s_0)$. Let $p(A,B)$ be given by

$$p(A,B) = \Pr(x, \Delta_h, \Delta_x, s'|s = s_0). \tag{59}$$

Then, from Lemma 5, (57) is bounded by $\sqrt{(8\ln 2)\exp(-\beta N)}$ because of Pinsker's inequality. Hence, we have

$$\left|\Pr\left(A \text{ wins}\right) - Q\left(A \text{ wins}\right)\right| \leq \sqrt{(8\ln 2)\exp(-\beta N)}. \tag{60}$$

From Theorem 1, $Q\left(A \text{ wins}\right)$ is bounded by $\frac{d+1}{q^r}$. Hence

$$\Pr\left(A \text{ wins}\right) \leq \sqrt{(8\ln 2)\exp(-\beta N)} + \frac{d+1}{q^r}. \tag{61}$$

Each $\{s\}$ conveys $dr\log_2 q$ bits of information, where r is defined in Theorem 2. Recall that the total number of channel uses is denoted by $2n$. The relay node transmits during n channel uses. Node 1 transmits during the other n channel uses. When node 1 transmits, node 2 may or may not transmit depending on which of the four stages described at the beginning of this section is being executed. For the four-stage transmission scheme, n is given by

$$n = 2N + r + \left\lceil \frac{\lceil dr\log_2 q\rceil}{NR_e} \right\rceil N. \tag{62}$$

This is because N channel uses are needed to transmit x or k, and r channel uses are needed to transmit $k \oplus h$. The third term in (62) is the number of channel uses needed to transmit s, where $\lceil x\rceil$ is the operation that rounds x to the nearest integer greater than or equal to x. In this stage, each lattice point can convey NR_e bits and takes N channel uses to transmit. s can be represented by $\lceil dr\log_2 q\rceil$ bits. Hence, $\left\lceil \frac{\lceil dr\log_2 q\rceil}{NR_e} \right\rceil$ is the number of lattice points that need to be transmitted during this stage, leading to the third term in (62) as being the total number of channel uses for this stage.

Since s is composed of d elements from $\mathcal{GF}(q^r)$, the overall secrecy rate R_s is given by

$$R_s = \frac{dr\log_2 q}{2n}. \tag{63}$$

By substituting (62) into (63), we observe R_s can be made arbitrarily close to $0.5R_e$ by choosing a sufficiently large d.

Let P_T denote the transmission power averaged over the channel uses during which a node transmits. Based on the four-stage transmission scheme, P_T of node 1 and the relay are the same. P_T of node 2 is smaller since it does not transmit during the third stage. Hence, we only need to make sure P_T of node 1 does not exceed the power constraint $\bar{P}$. P_T of node 1 is calculated as follows.

1) For the first two stages, each stage takes N channel uses and recall that we use P_1 to denote the average power per channel use for these two stages.
2) The third stage takes r channel uses and recall that we use P_2 to denote the average power per channel use for this stage.
3) The number of channel uses for the fourth stage is given by the last term in (62).

Recall that we use P to denote the average power per channel use for this stage. Hence, P_T is given by

$$P_T = \frac{P_1 2N + P_2 r + P\left\lceil \frac{\lceil dr\log_2 q\rceil}{NR_e} \right\rceil N}{n}. \tag{64}$$

P_T can be made arbitrarily close to but strictly smaller than $\bar{P}$ by choosing a sufficiently large d and a sufficiently small ε_P. This also implies that $0.5R_e$ can be made arbitrarily close to (47).

Once R_s and P_T are fixed, d is fixed. On the other hand, as shown by (62) and (15), for a fixed d, n increases linearly with respect to N.

Select r as in (15) such that r increases linearly with respect to N. Then, from (61), we observe that the probability that the adversary wins decreases exponentially fast with N. Hence, we have the bound on $\Pr(A \text{ wins})$ stated in the theorem.

We next check whether the secrecy constraint is satisfied:

$$I\left(s; Y_r\left(i\right), 0 \leq i \leq 3\right) \tag{65}$$

$$\leq I\left(x; Y_r\left(0\right)\right) + I\left(h; Y_r\left(1\right), Y_r\left(2\right)\right) + I\left(s; Y_r\left(3\right)\right). \tag{66}$$

In (66), the first term decreases exponentially fast with respect to N due to Theorem 2. For the second term, we have

$$\begin{aligned} & I\left(h; Y_r(1), Y_r(2)\right) \\ & \leq I\left(h; \bar{Y}_r(1), Z_r(1), h \oplus k\right) && (67) \\ & = I\left(h; \bar{Y}_r(1), h \oplus k\right) && (68) \\ & = I\left(h; h \oplus k\right) + I\left(h; \bar{Y}_r(1) | h \oplus k\right) && (69) \\ & = I\left(h; \bar{Y}_r(1) | h \oplus k\right) && (70) \\ & \leq I\left(h, k; \bar{Y}_r(1)\right) = I\left(k; \bar{Y}_r(1)\right) && (71) \end{aligned}$$

where $\bar{Y}_r$ was defined in (17).

Hence, the second term is bounded by $I(k, \bar{Y}_r(1))$, which also decreases exponentially fast with respect to N due to Theorem 2. The third term in (66) decreases exponentially fast with respect to N due to Theorem 6. Hence, (65) decreases exponentially fast with respect to N.

Finally, we check whether the confidential message W, which corresponds to s in our scheme, can be transmitted reliably. We observe that the probability $\Pr(W \neq \hat{W})$ does decrease exponentially fast with respect to N because the decoding error probability of the lattice decoder decreases at this speed, as stated in the end of Section V.

The probability

$$\Pr\left(\hat{W} \text{ is not accepted by Node 2} | W = \hat{W}\right) \tag{72}$$

depends on whether $x, k, k \oplus h$ can be transmitted reliably. Since they are also transmitted with the nested lattice code and decoded with a lattice decoder, the probability of decoding error when transmitting $x, k, k \oplus h$ also decreases exponentially with respect to the dimension of the lattice, which in turn increases linearly with N. Hence, (72) also decreases exponentially fast with respect to N.

Hence, we have proved the theorem. ■

Remark 13: It is evident from (60) that if Lemma 5 were weakened to just proving the left-hand side converging to 0, which is the case if the secrecy notion like the one in [43] is used, then it would not be possible to preserve the exponentially decreasing detection property offered by the AMD code. Hence, in this problem, the secrecy notion as stated in [43] is insufficient, and a stronger notion, as described by (16), is required. □

Remark 14: Theorem 7 is an achievability result. Finding good upper bounds for the secrecy rate for the two-hop network with an untrusted relay is a nontrivial problem for which we refer the interested reader to [20]. A simple upper bound for our specific channel model described in Section II is one without secrecy constraints which follows from [44] and is $\frac{0.5}{2}\log_2(1+\bar{P})$. By comparison, the secret rate achieved in this study is $[\frac{0.5}{2}\log_2(\frac{1}{2}+\bar{P})-1]^+$, which is within a constant gap from this upper bound. □

VIII. Conclusion

In this paper, we developed a coding scheme which provides strong secrecy by combining nested lattice codes and universal hash functions. In our previous work [26], the representation theorem for nested lattice codes is used to bound the Shannon entropy. Here, we showed that the same theorem is also useful in bounding another information-theoretic measure, i.e., the Rényi entropy, which in turn leads to the desired strong secrecy results in a Gaussian setting. We showed that this coding scheme can be used with AMD codes to perform Byzantine detection for a Gaussian two-hop network where the relay is both an eavesdropper and a Byzantine attacker. Using this code, we showed that the probability that a Byzantine adversary wins decreases exponentially fast with respect to the number of channel uses.

It should be noted that, in this paper, we have assumed that the channel gains are known by each node before the communication starts. It should be recognized that the Byzantine attacker at the relay node may attempt to manipulate the channel estimation process, for example, by broadcasting incorrect pilot signals, to gain an advantage. Detection of this type of misbehavior is closely related to the physical layer implementation of the system and is left as future work.

We have also assumed a model that is discrete in time, which implicitly assumes the signals interfere synchronously at the relay. The effect of synchronization on secrecy is certainly worth further investigation.

Appendix A
Proof of Lemma 1

When $\Lambda_c = q\Lambda$ and the generation matrix of Λ has full rank, there are q^N lattice points in $(\Lambda + d^N) \cap \mathcal{V}(\Lambda_c)$. Each point in $(\Lambda + d^N) \cap \mathcal{V}(\Lambda_c)$ can be represented by its coordinates, which is a vector composed of N integers: $\{c_1, \ldots, c_N\}$.

We next prove the following mapping is an isomorphism from $(\Lambda + d^N) \cap \mathcal{V}(\Lambda_c)$ to the group of a finite field $\mathcal{GF}(q^N)$:

$$\begin{aligned} & \mathbf{I} : \mathbf{I}(c_1, \ldots c_N) \\ & = \{c_1 \bmod q + (c_2 \bmod q)v \ldots + (c_N \bmod q)v^{N-1}\}. \end{aligned} \tag{73}$$

$\mathbf{I}$ maps the coordinates $(c_1, \ldots, c_N)$ to a polynomial.

First we prove that two elements in $(\Lambda + d^N) \cap \mathcal{V}(\Lambda_c)$ cannot be mapped to the same element in $\mathcal{GF}(q^N)$. This can be proved via contradiction: Suppose they can. Then, we have two points x, and y, whose coordinates are $\{a_1, \ldots, a_N\}$ and $\{b_1, \ldots, b_N\}$, respectively, such that

$$(a_i - b_i) \bmod q = 0, \quad i = 1, \ldots, N \tag{74}$$

$$\exists j, \ \ a_j \neq b_j. \tag{75}$$

This means $x - y \in q\Lambda = \Lambda_c$. Let $z \in \Lambda_c$ be $x - y$. Then $x = y + z$ and $z \neq 0$.

Define the quantization operator $Q_{\Lambda_c}(x)$ as

$$Q_{\Lambda_c}(x) = \arg\min_{t \in \Lambda_c} \|t - x\| \tag{76}$$

where $\|t - x\|$ denotes the Euclidean distance between t and x. $Q_{\Lambda_c}(x)$ has the following property: $\forall z \in \Lambda_c$, $Q_{\Lambda_c}(x + z) = Q_{\Lambda_c}(x) + z$. This can be shown as follows:

$$\begin{aligned} Q_{\Lambda_c}(x+z) &= \arg\min_{t \in \Lambda_c} \|t - x - z\| && (77) \\ &= \arg\min_{t - z \in \Lambda_c} \|(t - z) - x\| && (78) \\ &= \arg\min_{t' \in \Lambda_c} \|t' - x\| + z && (79) \\ &= Q_{\Lambda_c}(x) + z. && (80) \end{aligned}$$

Since $x, y \in \mathcal{V}(\Lambda_c)$, we have $Q_{\Lambda_c}(x) = 0$ and $Q_{\Lambda_c}(y) = 0$. However, we can also write $Q_{\Lambda_c}(x) = Q_{\Lambda_c}(y+z) = Q_{\Lambda_c}(y) + z = z \neq 0$. This leads to a contradiction.

Since $\mathbf{I}$ cannot map two different lattice points to the same field element, and the set $(\Lambda + d^N) \cap \mathcal{V}(\Lambda_c)$ has the same cardinality as $\mathcal{GF}(q^N)$, $\mathbf{I}$ must be a bijective mapping.

Finally, it is easy to verify that $\mathbf{I}$ preserves the addition operation:

$$\mathbf{I}(x+y) = \mathbf{I}(x) + \mathbf{I}(y). \tag{81}$$

This completes the proof that $\mathbf{I}$ is an isomorphism.

Appendix B
Proof of Lemma 2

Let $g_i, i = 1, \ldots, r$ be the ith row of $\mathbf{G}$. Then $\mathbf{G}$ does not have full row rank if and only if

$$a_1 g_1 + a_2 g_2 + \ldots + a_r g_r = 0, \quad a_i \in \mathcal{GF}(q). \tag{82}$$

Since at least one a_i has to be nonzero, there are $q^r - 1$ possible choices for $\{a_i\}$.

For each choice of $\{a_i\}$, since one a_i is not zero, there are $q^{N(r-1)}$ solutions for $\{g_i\}$. Hence, there are at most $q^{N(r-1)}(q^r - 1)$ $\mathbf{G}$s that do not have full row rank. There are q^{Nr} possible $\mathbf{G}$s in all, each chosen with equal probability. Hence, the probability that $\mathbf{G}$ does not have full row rank is smaller than q^{r-N}, and we have Lemma 2.

Appendix C
Proof of Theorem 6

For the distribution for $t_i^N, i = 1, 2$ stated in Theorem 6, $t_1^N \oplus t_2^N$ is independent from t_1^N. Therefore

$$H_2\left(t_1^N | t_1^N \oplus t_2^N = t^N\right) = H_2\left(t_1^N\right) = N_0. \tag{83}$$

Then, as in (22), with probability $1 - 2^{-(s/2-1)}$:

$$\begin{aligned} &H_2\left(t_1^N | t_1^N \oplus t_2^N = t^N, T = a\right) \geq \\ &H_2\left(t_1^N | t_1^N \oplus t_2^N = t^N\right) - \log_2 |T| - s = N_0 - N - s. \end{aligned} \tag{84}$$

We next use the fact that when $\mathbf{G}$ is uniformly distributed over the set of linear functions from $\mathcal{GF}(2)^{N_0}$ to $\mathcal{GF}(2)^{r_0}$, the following equation holds according to Theorem 4:

$$H\left(\mathbf{G}\left(v(t_1^N)\right) | \mathbf{G}, t_1^N \oplus t_2^N = t^N, T = a\right) \geq r_0 - \frac{2^{r_0 - c}}{\ln 2} \tag{85}$$

where $c = N_0 - N - s$.

Hence

$$\begin{aligned} &H\left(\mathbf{G}\left(v(t_1^N)\right) | \mathbf{G}, t_1^N \oplus t_2^N, T\right) \geq \\ &\left(1 - 2^{-(s/2-1)}\right)\left(r_0 - \frac{2^{r_0-c}}{\ln 2}\right). \end{aligned} \tag{86}$$

In order for $2^{-(s/2-1)}$ to decrease exponentially fast with respect to N, we choose $s = \varepsilon N$, where $0 < \varepsilon < R_0 - 1$ so that c is positive. Choose r_0 such that for $\delta > 0$

$$r_0 < c - N\delta/2 = N_0 - N - s - N\delta/2 \tag{87}$$

so that $2^{r_0 - c}$ decreases exponentially fast with respect to N. Recall by (34), we have $N_0 \geq NR_0 - 1$. Hence a sufficient condition for (87) to hold is to require

$$r_0 < N(R_0 - 1) - s - N\delta. \tag{88}$$

This yields (36). For this r_0 and s, from (86), we observe that there exists $\beta > 0$, such that

$$I\left(\mathbf{G}\left(v(t_1^N)\right); t_1^N \oplus t_2^N, T | \mathbf{G}\right) \leq e^{-\beta N}. \tag{89}$$

We next use the fact that for sufficiently large N, most $\mathbf{G}$ has full row rank as shown in Lemma 2. Therefore, under a uniform distribution for $t_i^N, i = 1, 2, t_1^N$ and t_2^N being independent, there must exists a $\mathbf{G} = \mathbf{g}$, such that:

1) $\mathbf{g}$ has full rank;
2) $I\left(\mathbf{G}\left(v(t_1^N)\right); t_1^N \oplus t_2^N, T | \mathbf{G} = \mathbf{g}\right) \leq 2e^{-\beta N}$

Hence, we have proved Theorem 6.

Appendix D
Proof of Lemma 5

As described in Section IV, the zeroth stage is used to transmit x. The 1st stage is used to transmit k. The second stage is used to transmit $k \oplus h$. The third stage is used to transmit s.

A) Outline: Let $\oplus$ in $x \oplus y$ denote the addition operation in the field where x and y are taken from. Let $-x$ denote the element such that $(-x) \oplus x = 0$. The proof can be divided into two steps.

1) In the first step, we prove that

$$\begin{aligned} &I\left(x; \Delta_x, \Delta_h, \hat{s} | s = s_0\right) \qquad (90) \\ \leq &I(x; Y_r(0), X_1(0) \oplus X_2(0), Y_r(2), k \oplus h, \\ &Y_r(1), X_1(1) \oplus X_2(1) | M_r, s = s_0). \qquad (91) \end{aligned}$$

   The proof uses basic relationships implied by the coding scheme and the channel model through we gradually replace Δ_x, Δ_h and $\hat{s}$ with random variables which are more amenable to analysis. For this purpose, we shall use the linear property of $\mathbf{g}$ in Theorem 2 repeatedly.
2) In the second step, we use Theorem 2 to upper bound (91) and prove Lemma 5.

B) Step 1: Recall that $\mathbf{g}$ is the linear mapping whose existence is proved in Theorem 2. Then, we can write Δ_x as

$$\begin{aligned} \Delta_x &= \mathbf{g}\left(\hat{X}_r(0) \oplus (-X_2(0))\right) \oplus (-x) && (92) \\ &= \mathbf{g}\left(\hat{X}_r(0) \oplus (-X_2(0))\right) \oplus \mathbf{g}(-X_1(0)) && (93) \\ &= \mathbf{g}\left(\hat{X}_r(0) \oplus (-(X_2(0) \oplus X_1(0)))\right). && (94) \end{aligned}$$

Since Δ_x is a function of $\hat{X}_r(0)$ and $X_2(0) \oplus X_1(0)$, (90) is upper bounded by

$$I\left(x; \hat{X}_r\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), \Delta_h, \hat{s} | s = s_0\right). \tag{95}$$

$\hat{X}_r(0)$ is computed from $Y_2(0)$ by node 2. Hence, (95) is upper bounded by

$$\begin{aligned} & I\left(x; Y_2\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), \Delta_h, \hat{s} | s = s_0\right) && (96)\\ \leq & I\left(x; X_r\left(0\right), Z_R\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), \Delta_h, \hat{s} | s = s_0\right) && (97)\\ = & I\left(x; X_r\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), \Delta_h, \hat{s} | s = s_0\right) + \\ & I\left(x; Z_R\left(0\right) | X_r\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), \Delta_h, \hat{s}, s = s_0\right). && (98) \end{aligned}$$

Recall that $Z_R(0)$ is the noise observed by Node 2 during the stage responsible for transmitting x. We observe that it is independent from all the other terms in the second term of (98). This is because Δ_h, $\hat{s}$, s are only related to signals transmitted in later stages. The relay node has no knowledge of $Z_R(0)$. Hence, $Z_R(0)$ cannot affect the relaying strategy. As a result, (98) equals

$$I\left(x; X_r\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), \Delta_h, \hat{s} | s = s_0\right). \tag{99}$$

Recall that M_r denotes the randomness available to the relay node. Then, the expression in (99) is upper bounded by

$$\begin{aligned} & I\left(x; M_r, X_r\left(0\right), Y_r\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), \Delta_h, \hat{s} | s = s_0\right) && (100)\\ = & I\left(x; M_r, Y_r\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), \Delta_h, \hat{s} | s = s_0\right) + \\ & I\left(x; X_r\left(0\right) | M_r, Y_r\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), \Delta_h, \hat{s}, s = s_0\right). && (101) \end{aligned}$$

Since $X_r(0)$ is computed from $Y_r(0)$ at the relay node, it is a deterministic function of $Y_r(0)$, M_r and potentially s_0. Hence, the second term in (101) is 0, and (101) equals

$$I\left(x; M_r, Y_r\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), \Delta_h, \hat{s} | s = s_0\right). \tag{102}$$

We next examine Δ_h in (102). Recall that u is defined as $k \oplus h$. $\hat{u}$ and $\hat{k}$ are the estimates for u and k computed by node 2, respectively. With these notations, we can express Δ_h as

$$\begin{aligned} \Delta_h &= \hat{u} \oplus (-\hat{k}) \oplus (-h) && (103)\\ &= \hat{u} \oplus ((-k) \oplus (-\Delta_k)) \oplus (-h) && (104)\\ &= \hat{u} \oplus (-(k \oplus h)) \oplus (-\Delta_k). && (105) \end{aligned}$$

As seen from (103)–(105), Δ_h is a function of $\hat{u}$, $k \oplus h$, and Δ_k. Therefore, (102) can be upper bounded by

$$I\left(x; M_r, Y_r\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), \hat{u}, k \oplus h, \Delta_k, \hat{s} | s = s_0\right). \tag{106}$$

Note that $\hat{u}$ is computed from $Y_2(2)$ by node 2. Therefore, (106) is upper bounded by

$$\begin{aligned} & I(x; M_r, Y_r\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), Y_2\left(2\right), k \oplus h, \\ & \quad \Delta_k, \hat{s} | s = s_0) && (107)\\ \leq & I(x; M_r, Y_r\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), X_r\left(2\right), Z_R\left(2\right), k \oplus h, \\ & \quad \Delta_k, \hat{s} | s = s_0) && (108)\\ = & I(x; M_r, Y_r\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), X_r\left(2\right), k \oplus h, \\ & \quad \Delta_k, \hat{s} | s = s_0) + I(x; Z_R\left(2\right) | M_r, Y_r\left(0\right), \\ & \quad X_1\left(0\right) \oplus X_2\left(0\right), X_r\left(2\right), k \oplus h, \Delta_k, \hat{s}, s = s_0). && (109) \end{aligned}$$

Again $Z_R(2)$ is independent from all the other terms in the second term of (109). Hence, (109) equals

$$\begin{aligned} & I(x; M_r, Y_r\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), X_r\left(2\right), k \oplus h, \\ & \quad \Delta_k, \hat{s} | s = s_0). && (110) \end{aligned}$$

For Δ_k, we have

$$\begin{aligned} \Delta_k &= \mathbf{g}\left(\hat{X}_r\left(1\right) \oplus (-X_2\left(1\right))\right) \oplus (-k) && (111)\\ &= \mathbf{g}\left(\hat{X}_r\left(1\right) \oplus (-X_2\left(1\right))\right) \oplus \mathbf{g}\left(-X_1\left(1\right)\right) && (112)\\ &= \mathbf{g}\left(\hat{X}_r\left(1\right) \oplus (-\left(X_2\left(1\right) \oplus X_1\left(1\right)\right))\right). && (113) \end{aligned}$$

Hence, Δ_k is a function of $\hat{X}_r(1)$, $X_2(1) \oplus X_1(1)$. Therefore, (110) can be upper bounded by

$$\begin{aligned} & I(x; M_r, Y_r\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), Y_r\left(2\right), k \oplus h, \\ & \quad \hat{X}_r\left(1\right), X_1\left(1\right) \oplus X_2\left(1\right), \hat{s} | s = s_0). && (114) \end{aligned}$$

$\hat{X}_r(1)$ is computed from $Y_2(1)$ by node 2. Hence, (114) is upper bounded by

$$\begin{aligned} & I(x; M_r, Y_r\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), Y_r\left(2\right), k \oplus h, \\ & \quad Y_2\left(1\right), X_1\left(1\right) \oplus X_2\left(1\right), \hat{s} | s = s_0) && (115)\\ \leq & I(x; M_r, Y_r\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), Y_r\left(2\right), k \oplus h, \\ & \quad X_r\left(1\right), Z_R(1), X_1\left(1\right) \oplus X_2\left(1\right), \hat{s} | s = s_0) && (116)\\ = & I(x; M_r, Y_r\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), Y_r\left(2\right), k \oplus h, \\ & \quad X_r\left(1\right), X_1\left(1\right) \oplus X_2\left(1\right), \hat{s} | s = s_0) \\ & + I(x; Z_R(1) | M_r, Y_r\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), Y_r\left(2\right), \\ & \quad k \oplus h, X_r\left(1\right), X_1\left(1\right) \oplus X_2\left(1\right), \hat{s}, s = s_0) && (117)\\ = & I(x; M_r, Y_r\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), Y_r\left(2\right), k \oplus h, \\ & \quad Y_r\left(1\right), X_1\left(1\right) \oplus X_2\left(1\right), \hat{s} | s = s_0). && (118) \end{aligned}$$

Finally, $\hat{s}$ is computed from $Y_2(3), X_2(3)$ by node 2. Hence, (118) is upper bounded by

$$\begin{aligned} & I(x; M_r, Y_r\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), Y_r\left(2\right), k \oplus h, Y_r\left(1\right), \\ & \quad X_1\left(1\right) \oplus X_2\left(1\right), Y_2\left(3\right), X_2\left(3\right) | s = s_0) && (119)\\ \leq & I(x; M_r, Y_r\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), Y_r\left(2\right), k \oplus h, Y_r\left(1\right), \\ & \quad X_1\left(1\right) \oplus X_2\left(1\right), X_r\left(3\right), X_2\left(3\right) | s = s_0). && (120) \end{aligned}$$

Since $X_r(3)$ is a deterministic function of M_r, $Y_r(3)$ and potentially s_0, we can upper bound (120) with the following term by replacing $X_r(3)$ with $Y_r(3)$:

$$\begin{aligned} & I(x; M_r, Y_r\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), Y_r\left(2\right), k \oplus h, \\ & \quad Y_r\left(1\right), X_1\left(1\right) \oplus X_2\left(1\right), Y_r\left(3\right), X_2\left(3\right) | s = s_0) \\ & && (121)\\ \leq & I(x; M_r, Y_r\left(0\right), X_1\left(0\right) \oplus X_2\left(0\right), Y_r\left(2\right), k \oplus h, \\ & \quad Y_r\left(1\right), X_1\left(1\right) \oplus X_2\left(1\right), X_1\left(3\right), Z_r(3), X_2\left(3\right) | s = s_0). && (122) \end{aligned}$$

Equation (122) follows from $Y_r(3) = X_1(3) + X_2(3) + Z_r(3)$. We then use the fact that the stochastic encoder used by node 1 to transmit s is independent from the stochastic mapping used at other stages. Hence, we have

$$I(x; X_1(3), X_2(3), Z_r(3) | M_r, Y_r(0), X_1(0) \oplus X_2(0), Y_r(2), k \oplus h, Y_r(1), X_1(1) \oplus X_2(1), s = s_0) = 0 \quad (123)$$

and (122) equals

$$I(x; M_r, Y_r(0), X_1(0) \oplus X_2(0), Y_r(2), k \oplus h, Y_r(1), X_1(1) \oplus X_2(1) | s = s_0) \quad (124)$$

$$= I(x; M_r | s = s_0) + I(x; Y_r(0), X_1(0) \oplus X_2(0), Y_r(2), k \oplus h, Y_r(1), X_1(1) \oplus X_2(1) | M_r, s = s_0). \quad (125)$$

Next we note that since $I(x; M_r | s = s_0) = 0$, (125) equals

$$I(x; Y_r(0), X_1(0) \oplus X_2(0), Y_r(2), k \oplus h, Y_r(1), X_1(1) \oplus X_2(1) | M_r, s = s_0). \quad (126)$$

C) Step 2: Equation (126) is upper bounded by

$$I(x, h; Y_r(0), X_1(0) \oplus X_2(0), Y_r(2), k \oplus h, Y_r(1), X_1(1) \oplus X_2(1) | M_r, s = s_0). \quad (127)$$

Recall that the notation $\bar{Y}_r$, as introduced in (17), denotes the quantity obtained by subtracting the channel noise Z_r from Y_r. Following this notation, we can upper bound (127) as

$$I(x, h; \bar{Y}_r(0), Z_r(0), X_1(0) \oplus X_2(0), \bar{Y}_r(2), Z_r(2), k \oplus h, \bar{Y}_r(1), Z_r(1), X_1(1) \oplus X_2(1) | M_r, s = s_0) \quad (128)$$

$$= I(x, h; \bar{Y}_r(0), X_1(0) \oplus X_2(0), \bar{Y}_r(2), k \oplus h, \bar{Y}_r(1), X_1(1) \oplus X_2(1) | M_r, Z_r(i), i = 1, 2, 3, s = s_0). \quad (129)$$

Since for $i = 0, 1$, $X_1(i) \oplus X_2(i) = \bar{Y}_r(i) \bmod \Lambda_c$ and hence is a function of $\bar{Y}_r(i)$, we can drop $X_1(i) \oplus X_2(i), i = 0, 1$ from (129) and write it as

$$I(x, h; \bar{Y}_r(0), \bar{Y}_r(2), k \oplus h, \bar{Y}_r(1) | M_r, Z_r(i), i = 1, 2, 3, s = s_0) \quad (130)$$

which is further upper bounded by[6]

$$H(\bar{Y}_r(0) | M_r, Z_r(i), i = 1, 2, 3, s = s_0) + H(\bar{Y}_r(2), k \oplus h, \bar{Y}_r(1) | M_r, Z_r(i), i = 1, 2, 3, s = s_0) - H(\bar{Y}_r(0), \bar{Y}_r(2), k \oplus h, \bar{Y}_r(1) | x, h, M_r, Z_r(i), i = 1, 2, 3, s = s_0) \quad (131)$$

[6]We use the inequality $I(A, B; C|D) \leq H(A|D) + H(B|D) - H(A, B|C, D)$ for discrete random variable A, B, and random variable C, D.

which, by applying the chain rule to the last term, can be written as

$$H(\bar{Y}_r(0) | M_r, Z_r(i), i = 1, 2, 3, s = s_0) + H(\bar{Y}_r(2), k \oplus h, \bar{Y}_r(1) | M_r, Z_r(i), i = 1, 2, 3, s = s_0) - H(\bar{Y}_r(0) | x, h, M_r, Z_r(i), i = 1, 2, 3, s = s_0) - H(\bar{Y}_r(2), k \oplus h, \bar{Y}_r(1) | \bar{Y}_r(0), x, h, M_r, Z_r(i), i = 1, 2, 3, s = s_0). \quad (132)$$

We then use the two Markov chains shown below:

$$\bar{Y}_r(0) - \{x, M_r, Z_r(i), i = 1, 2, 3, s\} - h \quad (133)$$

$$\{\bar{Y}_r(2), k \oplus h, \bar{Y}_r(1)\} - \{h, M_r, Z_r(i), i = 1, 2, 3, s\} - \{x, \bar{Y}_r(0)\}. \quad (134)$$

The Markov relation in (133) holds because given x, the distribution of $\bar{Y}_r(0)$ only depends on the randomness in the transmitter of nodes 1 and 2 during stage 0. The Markov chain in (133) follows because

$$k \oplus h - \{h, M_r, Z_r(i), i = 1, 2, 3, s\} - \{x, \bar{Y}_r(0)\} \quad (135)$$

and

$$\{\bar{Y}_r(2), \bar{Y}_r(1)\} - \{k \oplus h, h, M_r, Z_r(i), i = 1, 2, 3, s\} - \{x, \bar{Y}_r(0)\} \quad (136)$$

are Markov chains. Equation (135) is a Markov chain, because, given h, the distribution of $k \oplus h$ only depends on k, which is independent from all the remaining terms in (135). Equation (136) is a Markov chain, because, given $k \oplus h$ and h, which implies k and $k \oplus h$ are given, the distribution of $\{\bar{Y}_r(2), \bar{Y}_r(1)\}$ only depends on the randomness in the transmitter of nodes 1 and 2 during stage 1 and stage 2.

Applying the two Markov chains (133) and (134) to the last two terms in (132), we find that it equals

$$I(x; \bar{Y}_r(0) | M_r, Z_r(i), i = 1, 2, 3, s = s_0) + I(h; \bar{Y}_r(2), k \oplus h, \bar{Y}_r(1) | M_r, Z_r(i), i = 1, 2, 3, s = s_0). \quad (137)$$

The first term in (137) equals

$$I(x; \bar{Y}_r(0)). \quad (138)$$

Since x is extracted from a lattice point in $\mathcal{GF}(q^N)$ based on the strong secrecy scheme described in Section VI-A1, from Theorem 2, we have $I(x; \bar{Y}_r(0)) < 2\exp(-\beta N)$.

For the second term in (137), note that $\bar{Y}_r(2)$ is just $X_1(2)$, because node 2 remains silent at this stage. Therefore, this term can be expressed as

$$I(h; X_1(2), k \oplus h, \bar{Y}_r(1) | M_r, Z_r(i), i = 1, 2, 3, s = s_0) \quad (139)$$

$$= I(h; \bar{Y}_r(1), X_1(2) | M_r, Z_r(i), i = 1, 2, 3, s = s_0) + I(h; k \oplus h | \bar{Y}_r(1), X_1(2), M_r, Z_r(i), i = 1, 2, 3, s = s_0). \quad (140)$$

The second term in (140) is 0 since $k \oplus h$ is a deterministic function of $X_1(2)$. Therefore, (140) equals

$$I\left(h; \bar{Y}_r(1), X_1(2) | M_r, Z_r(i), i = 1, 2, 3, s = s_0\right) \tag{141}$$

$$= I\left(h; \bar{Y}_r(1), X_1(2)\right). \tag{142}$$

Since $X_1(2)$ is determined by $h \oplus k$, (142) is upper bounded by

$$I\left(h; \bar{Y}_r(1), h \oplus k\right) \tag{143}$$

$$= I(h; h \oplus k) + I\left(h; \bar{Y}_r(1) | h \oplus k\right) \tag{144}$$

$$= I\left(h; \bar{Y}_r(1) | h \oplus k\right) \tag{145}$$

$$\leq I\left(h, k; \bar{Y}_r(1)\right) = I\left(k; \bar{Y}_r(1)\right). \tag{146}$$

Since k is extracted from a lattice point in $\mathcal{GF}(q^N)$ based on the strong secrecy scheme described in Section VI-A1, hence from Theorem 2, (146) is bounded by $2\exp(-\beta N)$.

Therefore, (137) is bounded by $4\exp(-\beta N)$. Hence, we have Lemma 5.

REFERENCES

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

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

[3] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. IT-4, no. 3, pp. 339–348, May 1978.

[4] S. Leung-Yan-Cheong and M. Hellman, "The Gaussian wire-tap channel," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 4, pp. 451–456, Jul. 1978.

[5] R. Liu, T. Liu, and H. V. Poor, "Multiple-input multiple-output Gaussian broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4215–4227, Sep. 2010.

[6] E. Ekrem and S. Ulukus, "The secrecy capacity region of the Gaussian MIMO multi-receiver wiretap channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2083–2114, Apr. 2011.

[7] E. Tekin and A. Yener, "The general Gaussian multiple access and two-way wire-tap channels: Achievable rates and cooperative jamming," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2735–2751, Jun. 2008.

[8] E. Tekin and A. Yener, "The Gaussian multiple access wire-tap channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5747–5755, Dec. 2008.

[9] Y. Liang and H. V. Poor, "Multiple access channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 976–1002, Mar. 2008.

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

[11] X. He and A. Yener, "Providing secrecy with structured codes: Tools and applications to Gaussian two-user channels," submitted to *IEEE Trans. Inf. Theory,* Jul. 2009 [Online]. Available: http://arxiv.org/abs/0907.5388.

[12] X. He and A. Yener, "The Gaussian many-to-one interference channel with confidential messages," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2730–2745, May 2011.

[13] L. Lai and H. El Gamal, "Cooperation for secrecy: The relay-eavesdropper channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 4005–4019, Sep. 2008.

[14] M. Yuksel, X. Liu, and E. Erkip, "A secure communication game with a relay helping the eavesdropper," in *Proc. IEEE Inf. Theory Workshop*, Oct. 2009, pp. 110–114.

[15] Y. Oohama, "Coding for relay channels with confidential messages," in *Proc. IEEE Inf. Theory Workshop*, Sep. 2001, pp. 87–89.

[16] X. He and A. Yener, "Cooperation with an untrusted relay: A secrecy perspective," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3801–3827, Aug. 2010.

[17] X. He and A. Yener, "Two-hop secure communication using an untrusted relay," *Eurasip J. Wireless Commun. Netw. (Special issue in Wireless Physical Layer Security)*, vol. 2009, p. 13, 2009.

[18] E. Ekrem and S. Ulukus, "Secrecy in cooperative relay broadcast channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 137–155, Jan. 2011.

[19] C. Kaufman, R. Perlman, and M. Speciner, *Network Security: Private Communication in a Public World*. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[20] X. He and A. Yener, "The role of feedback in two-way secure communication," submitted to *IEEE Trans. Inf. Theory,* Nov. 2009 [Online]. Available: http://arxiv.org/abs/0911.4432.

[21] T. Ho, B. Leong, R. Koetter, M. Medard, M. Effros, and D. R. Karger, "Byzantine modification detection in multicast networks using randomized network coding," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2798–2803, Jun. 2008.

[22] J. Katz and Y. Lindell, *Introduction to Modern Cryptography*. London, U.K.: Chapman & Hall/CRC, 2008.

[23] L. Lai, H. El-Gamal, and H. V. Poor, "Authentication over noisy channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 1118–1134, Feb. 2009.

[24] R. Cramer, Y. Dodis, S. Fehr, C. Padro, and D. Wichs, "Detection of algebraic manipulation with applications to robust secret sharing and fuzzy extractors," *Adv. Cryptology*, vol. 4965, pp. 471–488, 2008.

[25] X. He and A. Yener, "Secure communication with a Byzantine relay," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2009, pp. 2096–2100.

[26] X. He and A. Yener, "Providing secrecy with lattice codes," in *Proc. 46th Allerton Conf. Commun., Control, Comput.*, Sep. 2008.

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

[28] U. Maurer and S. Wolf, "Information-theoretic key agreement: From weak to strong secrecy for free," in *Proc. 19th Int. Conf. Theory Appl. Cryptographic Tech.*, 2000, pp. 351–368.

[29] I. Csiszár, "Almost independence and secrecy capacity," *Probl. Inf. Transmiss.*, vol. 32, no. 1, pp. 48–57, 1996.

[30] I. Csiszár and P. Narayan, "Common randomness and secret key generation with a helper," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 344–366, Mar. 2000.

[31] S. Jaggi, M. Langberg, S. Katti, T. Ho, D. Katabi, and M. Medard, "Resilient network coding in the presence of Byzantine adversaries," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2596–2603, Jun. 2008.

[32] O. Kosut, L. Tong, and D. Tse, "Nonlinear network coding is necessary to combat general Byzantine attacks," presented at the Allerton Conf. Commun., Control, Comput., Sep. 2009.

[33] V. Aggarwal, L. Lai, R. Calderbank, and H. V. Poor, "Wiretap channel type II with an active eavesdropper," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2009, pp. 1944–1948.

[34] V. Guruswami and A. Smith, "Explicit capacity-achieving codes for worst-case additive errors," Dec. 2009 [Online]. Available: http://arxiv.org/abs/0910.1511.

[35] M. J. Kim, M. Medard, J. Barros, and R. Koetter, "An algebraic watchdog for wireless network coding," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2009, pp. 1159–1163.

[36] M. P. Wilson, K. Narayanan, H. Pfister, and A. Sprintson, "Joint physical layer coding and network coding for bi-directional relaying," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5641–5654, Nov. 2010.

[37] U. Erez and R. Zamir, "Achieving $1/2\log(1+\mathrm{SNR})$ on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2293–2314, Oct. 2004.

[38] K. Narayanan, M. P. Wilson, and A. Sprintson, "Joint physical layer coding and network coding for bi-directional relaying," presented at the 45th Allerton Conf. Commun., Control, Comput., Sep. 2007.

[39] G. D. Forney Jr, "On the role of MMSE estimation in approaching the information-theoretic limits of linear Gaussian channels: Shannon meets Wiener," presented at the 41st Allerton Conf. Commun., Control, Comput., Sep. 2003.

[40] L. H. Ozarow and A. D. Wyner, "Wire-tap channel. II," *AT & T Bell Lab. Tech. J.*, vol. 63, no. 10, pp. 2135–2157, 1984.

[41] C. Cachin, "Entropy measures and unconditional security in cryptography," Ph.D. dissertation, ETH Zurich, Zurich, Switzerland, 1997.

[42] R. W. Yeung, *A First Course in Information Theory*. New York: Kluwer/Plenum, 2002.

[43] J. Barros and M. Bloch, "Strong secrecy for wireless channels," presented at the Int. Conf. Inf.-Theoretic Security, Aug. 2008.

[44] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Math. Stat. Prob*, 1961, vol. 1, pp. 351–384.

Xiang He (S'08–M'10) received B.S. and M.S. degrees in Electrical Engineering from Shanghai Jiao Tong University, Shanghai, China in 2003 and 2006 respectively. His master study is about high speed FPGA implementation of channel encoder, decoder and MIMO detectors. He received his Ph.D. degree in 2010 from the Department of Electrical Engineering at the Pennsylvania State University and joined Microsoft in that year. In 2010, he received Melvin P. Bloom Memorial Outstanding Doctoral Research Award from the Department of Electrical Engineering at the Pennsylvania State University and the best paper award from the Communication Theory Symposium in IEEE International Conference on Communications (ICC). In 2011, he was named as one of the exemplary reviewers by the IEEE COMMUNICATION LETTERS. His research interests include information theoretic secrecy, coding theory, queuing theory, optimization techniques, distributed detection and estimation.

Aylin Yener (S'91–M'00) received the B.Sc. degree in electrical and electronics engineering, and the B.Sc. degree in physics, from Bogaziçi University, Istanbul, Turkey; and the M.S. and Ph.D. degrees in electrical and computer engineering from Wireless Information Network Laboratory (WINLAB), Rutgers University, New Brunswick, NJ. Commencing fall 2010, for three semesters, she was a P.C. Rossin Assistant Professor at the Electrical Engineering and Computer Science Department, Lehigh University, PA. In 2002, she joined the faculty of The Pennsylvania State University, University Park, PA, where she was an Assistant Professor, then Associate Professor, and is currently Professor of Electrical Engineering since 2010. During the academic year 2008–2009, she was a Visiting Associate Professor with the Department of Electrical Engineering, Stanford University, CA. Her research interests are in information theory, communication theory and network science, with recent emphasis on green communications and information security. She received the NSF CAREER award in 2003.

Dr. Yener served as the student committee chair for the IEEE Information Theory Society 2007–2011, and was the co-founder of the Annual School of Information Theory in North America co-organizing the school in 2008, 2009 and 2010. She currently serves on the board of governors as the treasurer of the IEEE Information Theory Society.