A Coding Approach to Guarantee Information Integrity Against a Byzantine Relay

Eric Graves and Tan F. Wong
Department of Electrical & Computer Engineering
University of Florida, FL 32611
{ericsgra,twong}@ufl.edu

***Abstract*—This paper presents a random coding scheme with which two nodes can exchange information with guaranteed integrity over a two-way Byzantine relay. This coding scheme is employed to obtain an inner bound on the capacity region with information integrity. No pre-shared secret or secret transmission is needed for the proposed scheme. Hence the inner bound obtained is generally larger than those achieved based on secret transmission schemes. This approach advocates the separation of supporting information integrity and secrecy.**

I. Introduction

In order for two parties to communicate through an intermediary node, it is required that the intermediary (hereto referred to as a relay) must faithfully forward the information. In a communication network, such cooperative behavior is not guaranteed as relays may have reasons for forwarding false information in order to fool the intended participants. Such attacks, oft referred to as *Byzantine attacks*, have major ramifications on protocols that operate within the network. There has been much research into secrecy, or the ability to obfuscate the information from unintended destinations, which is but a single aspect of information theoretic security. Clearly this need for secrecy is important, but only an aspect of the larger goal of safe reliable communications, and to this point more recently there has been research on ways information theoretic techniques can be used to secure more than just the privacy of the information. For instance, Maurer [1] addressed the need for authentication in order to support secret key agreement in a simple two-node system with an active eavesdropper. In much the same way, integrity of information must be guaranteed for larger systems to ensure secrecy. The genesis of this problem has roots in cryptography [2] where codes like [3], [4] have been studied as a means of determining Byzantine attacks, while more recent work has focused on the integrity of a network using linear network coding [5], [6], as well as the study of using coding to determine manipulation in basic channels which are supported by a relay [7], [8].

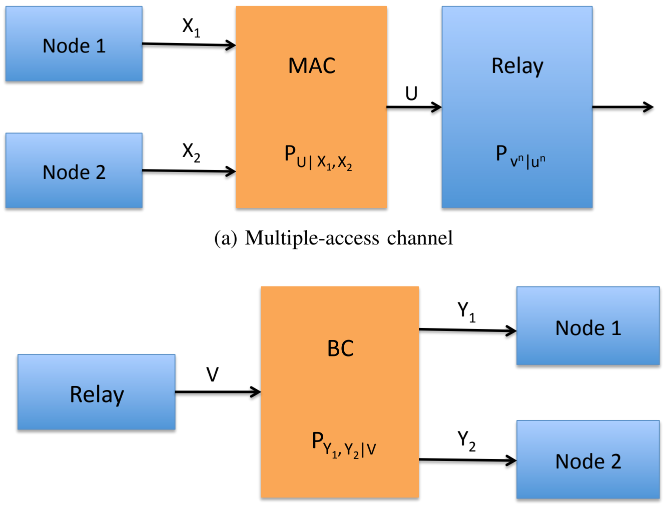
Historically this problem of guaranteeing information integrity is treated by the use of a non-observable key to add redundancy to information that allows attack detection. In [3], the problem, originally motivated by a dishonest gambling pit boss, was studied with use of planar codes and random codes. While in [4], Cramer et al. show that any random linear transformation into a space of greater dimension is with high probability invertible. Thus if one were to modify the transformation or the symbols, it is with high probability that the modified sequence would not be one corresponding to any valid input sequence. The resulting codes are known as algebraic manipulation detection (AMD) codes.

In order to extend these ideas to the physical layer, additional redundancy is required to cope with the possibility of channel errors. Mao and Wu [8] posed the problem of trying to determine which relay in a multiple relay two-hop network was manipulating the data. A cross-layer method is set forth in which a cryptographic key is inserted into the signal, by which the intended destination determines from the physical layer error rate if manipulation has occurred. In slight contrast, the more recent work of He and Yener [7], mainly focused on the problem in the two-way two-hop channel studied in this paper, does not require use of a shared secret key. Instead, an LDPC code is employed to support secret transmission which in turns allows the use of an AMD code to detect attacks by the relay. The major drawback of this solution is that it does not provide a deeper understanding of what is actually needed to support the two different requirements of secrecy and integrity, making extensions to beyond the addition channel considered in [7] difficult.

In contrast we propose a different strategy by separating the concepts of secrecy and integrity to ensure that communication can be verified without use of any key or any secret transmission. Using random coding techniques, we obtain an inner bound on the capacity region with information integrity in the general scenario of two nodes which must communicate through a Byzantine relay node. Section II describes the channel model in detail. The inner bound on the capacity region with information integrity is provided in Section III. An outline of the achievability proof that leads to the inner bound is given in Section IV. The notation employed in the paper is summarized in Table I.

II. Two-way Relay Model

Consider the two-way, half-duplex relay channel model shown in Fig. 1, in which two nodes (1 and 2) simultaneously send symbols to a relay node through a discrete, memoryless multiple-access channel (MAC). The half-duplex relay node is then supposed to broadcast its received symbols back to the two nodes. Furthermore we restrict the input and output alphabets of the relay to be the same. Also, for simplicity, we assume that the broadcast channel (BC) from the relay



(a) Multiple-access channel

(b) Broadcast channel

Fig. 1: Two-way, half-duplex relay model.

TABLE I: Notation

X	random variable
$\mathsf{x} \in \mathcal{X}$	the element x from the alphabet of X
x^n	n instances of random variable X over $\mathcal{X}$
$N(\mathsf{x}\vert x^n)$	number of times x occurs in x^n
$P_{x^n}(\mathsf{x})$	$\frac{1}{n}N(\mathsf{x}\vert x^n)$
$P_{x^n\vert y^n}(\mathsf{x}\vert\mathsf{y})$	$\frac{P_{x^n,y^n}(\mathsf{x},\mathsf{y})}{P_{y^n}(\mathsf{y})}$
$[X]^n_\delta$	$\{P_{x^n} : \vert P_{x^n}(\mathsf{x}) - P(\mathsf{x})\vert \leq \delta\}$
$T^n_{[X]_\delta}$	$\{x^n : P_{x^n} \in [X]^n_\delta\}$
$P_{X\vert Y}(\mathsf{x}\vert\mathsf{y})$	cond. pmf of X given Y; also treated as a matrix
$\mathbb{1}(\cdot)$	indicator function

back to the nodes is perfect. That is, both nodes perfectly observe the symbols sent out by the relay. There is some possibility that the relay may modify its received symbols in an attempt to conduct a manipulation attack. The design goal is for each node to at least detect any malicious act of the relay in the event that it can not decode the information sent by the other node; in other word, to guarantee the integrity of the information forwarded by the relay.

More specifically, let $\mathcal{X}_1$, $\mathcal{X}_2$, and $\mathcal{U}$ denote the discrete alphabets of node 1's input, node 2's input, and the output of the MAC. Over time instants $1, 2, \ldots, n$, suppose that nodes 1 and 2 transmit the symbol sequences x_1^n and x_2^n, respectively, through the memoryless MAC. The output sequence U^n is conditionally distributed according to

$$p(u^n|x_1^n, x_2^n) = \prod_{i=1}^{n} P_{U|X_1,X_2}(u_i|x_{1,i}, x_{2,i}) \qquad (1)$$

where the conditional pmf $P_{U|X_1,X_2}(\mathsf{u}|\mathsf{x}_1,\mathsf{x}_2)$ specifies the MAC. The relay node, during time instants $1, 2, \ldots, n$, observes the output symbol sequence U^n of the MAC, processes (or manipulates) it, and then broadcasts the processed symbol sequence to nodes 1 and 2 at time instants $n+1, n+2, \ldots, 2n$ via the perfect BC. Let V^n denote the relay's output sequence. The assumption of perfect BC from the relay to the nodes implies that $\mathcal{Y}_1 = \mathcal{Y}_2 = \mathcal{U}$, $P_{Y_1|V} = P_{Y_2|V} = I$, and

$$p(y_1^n, y_2^n|v^n) = \mathbb{1}(y_1^n = y_2^n = v^n). \qquad (2)$$

For convenience hereafter, we simply make V^n the symbol sequence observed by both nodes.

Fix n. Let R_1 and R_2 be two positive rates. Consider the encoder-decoder quadruple $(\mathbf{C}_1^n, \mathbf{C}_2^n, g_1^n, g_2^n)$:

$$\begin{aligned}
\mathbf{C}_1^n &: \{1, 2, \ldots, 2^{nR_1}\} \to \mathcal{X}_1^n \\
\mathbf{C}_2^n &: \{1, 2, \ldots, 2^{nR_2}\} \to \mathcal{X}_2^n \\
g_1^n &: \mathcal{U}^n \times \{1, 2, \ldots, 2^{nR_1}\} \to \{1, 2, \ldots, 2^{nR_2}\} \cup \{\mathbf{!}\} \\
g_2^n &: \mathcal{U}^n \times \{1, 2, \ldots, 2^{nR_2}\} \to \{1, 2, \ldots, 2^{nR_1}\} \cup \{\mathbf{!}\}
\end{aligned}$$

where $\mathbf{C}_1^n$ and g_1^n are the encoder and decoder used by node 1, and $\mathbf{C}_2^n$ and g_2^n are the encoder and decoder used by node 2. Note that we allow the encoders $\mathbf{C}_1^n$ and $\mathbf{C}_2^n$ to be random. The symbol $\mathbf{!}$ in the decoder output alphabets denotes the decision that the received sequence is deemed untrustworthy, i.e., the relay has possibly been malicious. Let W_1 and W_2 be independent messages of nodes 1 and 2 that are uniformly distributed over $\{1, 2, \ldots, 2^{nR_1}\}$ and $\{1, 2, \ldots, 2^{nR_2}\}$, respectively. Then $X_1^n = \mathbf{C}_1^n(W_1)$ and $X_2^n = \mathbf{C}_2^n(W_2)$ are the codewords sent by nodes 1 and 2 to the relay through the MAC.

The potential manipulation by the relay is specified by the conditional distribution of V^n given the other random quantities mentioned above. We impose the Markovity restriction on the conditional distribution that

$$p(v^n|u^n, x_1^n, x_2^n, w_1, w_2, \mathbf{c}_1^n, \mathbf{c}_2^n) = p(v^n|u^n, \mathbf{c}_1^n, \mathbf{c}_2^n) \qquad (3)$$

which means that the relay may potentially manipulate the transmission based only on the output symbols of the MAC that it observes as well as its knowledge about the codebooks used by the nodes. If $p(v^n|u^n, \mathbf{c}_1^n, \mathbf{c}_2^n) = \mathbb{1}(v^n = u^n)$, then we regard the relay as *non-malicious*. Otherwise the relay is malicious. For later presentation clarity, we will employ H_1 and H_0 to denote the conditions that the relay is and is not malicious, respectively.

With the scheduling and coding scheme described above, we say that the rate pair (R_1, R_2) is *achievable with information integrity* if there exists a sequence of encoder-decoder quadruples $\{(\mathbf{C}_1^n, \mathbf{C}_2^n, g_1^n, g_2^n)\}$ such that:

$$\begin{aligned}
&\text{Under } H_0 : \\
&\quad \Pr\{g_1^n(V^n, W_1) \neq W_2 \cup g_2^n(V^n, W_2) \neq W_1\} \to 0 \\
&\text{Under } H_1 : \\
&\quad \Pr\{g_1^n(V^n, W_1) \notin \{W_2, \mathbf{!}\} \cup g_2^n(V^n, W_2) \notin \{W_1, \mathbf{!}\}\} \to 0
\end{aligned}$$

as $n \to \infty$. Note that the requirement under H_1 forces the decoders to either detect the substitution attack by the relay or correct the symbols modified. The *capacity region with information integrity* in this case can then be defined as the closure of the set of achievable rate pairs with information integrity. Note that we have not counted the use of the perfect BC from the relay back to the nodes in the rate definition above.

Note that from (1), (2), and (3), the joint distribution of $(V^n, U^n, X_1^n, X_2^n, W_1, W_2, \mathbf{C}_1^n, \mathbf{C}_2^n)$ is given by

$$\begin{aligned} &p(v^n, u^n, x_1^n, x_2^n, w_1, w_2, \mathbf{c}_1^n, \mathbf{c}_2^n) \\ &\quad = p(v^n|u^n, \mathbf{c}_1^n, \mathbf{c}_2^n)\, p(u^n|x_1^n, x_2^n)\, \mathbb{1}\,(\mathbf{c}_1^n(w_1) = x_1^n) \\ &\qquad \cdot \mathbb{1}\,(\mathbf{c}_2^n(w_2) = x_2^n)\; p(w_1)\, p(w_2)\, p(\mathbf{c}_1^n, \mathbf{c}_2^n). \end{aligned} \tag{4}$$

III. Inner bound on capacity region

Under the operational definition of information transfer with guaranteed integrity given in Section II, it turns out that the matrix-algebraic structure of *manipulability* given in [9] is critical to our inner bound on the capacity region. For easy reference, we repeat the definition of manipulability for node 1's *observation channel* specified by the stochastic matrix pair $(P_{U|X_1}, P_{Y_1|V})$ in the notation of this paper:

Definition 1. *The observation channel $(P_{U|X_1}, P_{Y_1|V})$ is manipulable if there exists a stochastic $|\mathcal{U}| \times |\mathcal{U}|$ non-zero matrix Φ, such that*

$$P_{Y_1|X} = P_{Y_1|V} \Phi P_{U|X_1}$$

Otherwise, $(P_{U|X_1}, P_{Y_1|V})$ is said to be non-manipulable.

Manipulability of node 2's observation channel $(P_{U|X_2}, P_{Y_2|V})$ is the same. A more detailed discussion on manipulability can be found in [9].

Theorem 1. *An inner bound on the capacity region with information integrity is the closure of the convex hull of all (R_1, R_2) satisfying*

$$\begin{aligned} R_1 &< I(X_1; U|X_2) \\ R_2 &< I(X_2; U|X_1) \end{aligned}$$

for some

$$P_{U,X_1,X_2}(\mathsf{u}, \mathsf{x}_1, \mathsf{x}_2) = P_{U|X_1,X_2}(\mathsf{u}|\mathsf{x}_1, \mathsf{x}_2) P_{X_1}(\mathsf{x}_1) P_{X_2}(\mathsf{x}_2)$$

having the property that $(P_{U|X_1}, I)$ and $(P_{U|X_2}, I)$ are both non-manipulable.

Note that the difference between the above region and the standard capacity region (without information integrity) is that the former does not contain the rate pairs generated by input distributions that do not give non-manipulable observation channels while the latter does.

Let us apply Theorem 1 to the simple example two-way relay channel made up of a binary erasure MAC [10] and a perfect BC. That is, X_1 and X_2 have the same binary alphabet $\{0, 1\}$. The MAC is described by $U = X_1 + X_2$. Hence the alphabet of U and V are both $\{0, 1, 2\}$. The BC is defined by $Y_1 = V$ and $Y_2 = V$, and $P_{Y_1|V} = P_{Y_2|V} = I$. Let $P_{X_1}(1) = p$ and $P_{X_2}(1) = q$, where $0 < p, q < 1$. Then

$$P_{U|X_1} = \begin{pmatrix} 1-q & 0 \\ q & 1-q \\ 0 & q \end{pmatrix} \quad \text{and} \quad P_{U|X_2} = \begin{pmatrix} 1-p & 0 \\ p & 1-p \\ 0 & p \end{pmatrix}.$$

It is not hard to check [9, Thm. 2] that both $(P_{U|X_1}, I)$ and $(P_{U|X_2}, I)$ are non-manipulable for all choices of p and q. Thus the inner bound in Theorem 1 implies that the capacity regions with and without information integrity are identical as shown by the thick-lined square region in Fig. 2, and can be achieved by choosing both X_1 and X_2 to be equally likely binary random variables. With this choice of input distributions, the shaded area shown in Fig. 2 is the capacity region of the binary erasure MAC channel, treating the relay as the MAC receiver. From the achievability argument in Section IV, to guarantee integrity while operating in the shaded region, we need randomization in the encoders to move the operating point up to the unshaded triangular portion of the capacity region. Physically, this means that we need to confuse the relay from decoding both nodes' messages. Note that for this two-way relay channel, uncoded transmission can achieve capacity without information integrity, while coded transmission is needed in order to achieve capacity with information integrity.

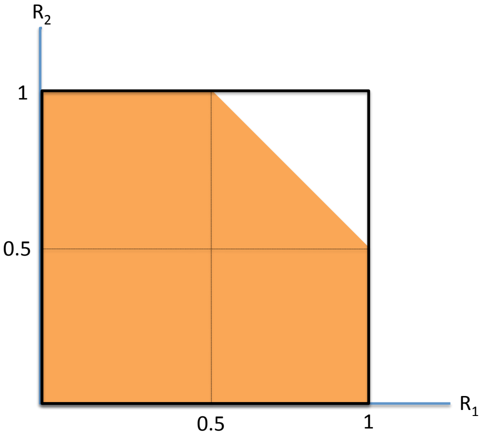


Fig. 2: Capacity region with information integrity of the two-way relay channel with a binary erasure MAC and a perfect BC.

IV. Outline of Achievability

We employ the standard random coding argument to show that the probabilities of both error events under H_0 and H_1 employed in the definition of achievable rate pairs in Section II, averaged over all random codebooks, converge to zero as n increases under the conditions stated in Theorem 1. Then the existence of a codebook pair (and the corresponding decoding functions) having the same property is guaranteed.

A. Code Construction

Fix $P_{X_1}(\mathsf{x}_1)$ and $P_{X_2}(\mathsf{x}_2)$ that satisfy the condition of non-manipulable $(P_{U|X_1}, I)$ and $(P_{U|X_2}, I)$. It can then be shown that $I(X_1; U) < I(X_1; U|X_2)$ and $I(X_2; U) < I(X_2; U|X_1)$. If the rate pair (R_1, R_2) satisfies the following constraints that $I(X_1; U) < R_1 < I(X_1; U|X_2)$, $I(X_2; U) < R_2 < I(X_2; U|X_1)$, and $R_1 + R_2 > I(X_1, X_2; U)$, independently and uniformly pick

2^{nR_1} codewords $\mathbf{C}_1^n(1), \mathbf{C}_1^n(2), \ldots, \mathbf{C}_1^n(2^{nR_1})$ from the typical set $T^n_{[X_1]_{\delta_n}}$, where $\{\delta_n\}$ satisfies the delta convention set forth in [11]. Similarly, pick 2^{nR_2} codewords $\mathbf{C}_2^n(1), \mathbf{C}_2^n(2), \ldots, \mathbf{C}_2^n(2^{nR_2})$ from the typical set $T^n_{[X_2]_{\delta_n}}$. Instantiations of $\mathbf{C}_1^n$ and $\mathbf{C}_2^n$ define the deterministic encoding functions for nodes 1 and 2, respectively. On the other hand, if the rate pair does not satisfy the above constraints, we still generate the same number of codewords as above. The codebook of each user will be partitioned into equal size subsets, each of which is associated to a unique message. The actual codeword that is sent will be uniformly chosen from the subset associated with the message to be sent. Note that the encoding functions are random in this case. It is easy to see that we may assume that the rate pair satisfies the above constraints (i.e., deterministic encoding functions are used) without loss of generality.

For a fixed codebook pair $(\mathbf{c}_1^n, \mathbf{c}_2^n)$, nodes 1 employs (and symmetrically for node 2) the following typicality decoder:

$$g_1^n(v^n, w_1) = \begin{cases} \hat{w}_2 & \text{if } (v^n, \mathbf{c}_1^n(w_1), \mathbf{c}_2^n(\hat{w}_2)) \in T^n_{[U,X_1,X_2]_{2\mu_n}} \\ & \text{and there is no } w_2' \neq \hat{w}_2 \text{ such that} \\ & (v^n, \mathbf{c}_1^n(w_1), \mathbf{c}_2^n(w_2')) \in T^n_{[U,X_1,X_2]_{2\nu_n}}, \\ \mathbf{!} & \text{otherwise,} \end{cases}$$

where μ_n is a function of δ_n and the alphabets $\mathcal{U}$, $\mathcal{X}_1$, and $\mathcal{X}_2$, and $\nu_n \geq \mu_n$ will be specified later. Both μ_n and ν_n satisfy the delta convention. Below we will also make use of μ_n', μ_n'', $\tilde{\mu}_n$, and $\hat{\mu}_n$, which are all constant multiples of μ_n.

B. Error analysis under H_0

Under this case, we have $V^n = U^n$. Because of symmetry and the union bound, it suffices to consider $\Pr\{g_1^n(V^n, W_1) \neq W_2\}$. To that end, define

$$\mathcal{U}(x_1^n, x_2^n) \triangleq \left\{u^n : (u^n, x_1^n, x_2^n) \in T^n_{[U,X_1,X_2]_{2\mu_n}}\right\},$$

$$\mathcal{V}_\beta(x_1^n; w_2, \mathbf{c}_2^n) \triangleq \left\{ u^n : \bigcup_{\hat{w}_2 \neq w_2} (u^n, x_1^n, \mathbf{c}_2^n(\hat{w}_2)) \cap T^n_{[U,X_1,X_2]_\beta} \neq \emptyset \right\},$$

respectively. Then

$$\begin{aligned} \Pr\{g_1^n(U^n, W_1) \neq W_2\} &\leq \Pr\{U^n \notin \mathcal{U}(\mathbf{C}_1^n(W_1), \mathbf{C}_2^n(W_2))\} \\ &+ \Pr\{U^n \in \mathcal{U}(\mathbf{C}_1^n(W_1), \mathbf{C}_2^n(W_2)) \\ &\qquad \cap \mathcal{V}_{2\nu_n}(\mathbf{C}_1^n(W_1); W_2, \mathbf{C}_2^n)\}. \end{aligned} \tag{5}$$

It remains to show that both probabilities on the right hand side of (5) converge to zero as $n \to \infty$.

By the combination of [11, Problem 2.9], (1), and [11, Lemma 2.12], it is shown that

$$\begin{aligned} &\Pr\{U^n \notin \mathcal{U}(\mathbf{C}_1^n(W_1), \mathbf{C}_2^n(W_2))\} \\ &\quad \leq (n+1)^{2|\mathcal{X}_1||\mathcal{X}_2|} 2^{-n\delta_n} + 2|\mathcal{U}||\mathcal{X}_1||\mathcal{X}_2| e^{-2n\mu_n^2}. \end{aligned} \tag{6}$$

Additionally, using the standard argument (cf. [10, Ch. 7]), based on the fact that the codewords in the codebooks are chosen independently, it is easy to establish

$$\begin{aligned} &\Pr\{U^n \in \mathcal{U}(\mathbf{C}_1^n(W_1), \mathbf{C}_2^n(W_2)) \cap \mathcal{V}_{2\nu_n}(\mathbf{C}_1^n(W_1); W_2, \mathbf{C}_2^n)\} \\ &\quad \leq 2^{-n[I(X_2;U|X_1) - R_2 - \epsilon_n]}, \end{aligned}$$

for some $\epsilon_n \to 0$.

C. Error analysis under H_1

Again by symmetry and the union bound, it suffices to show that $\Pr\{g_1^n(V^n, W_1) \notin \{W_2, \mathbf{!}\}\}$ vanishes as n increases. Allowing λ_n to be set later, define

$$E_1 = \{(u^n, v^n) : \min I\left(\tilde{X}_1; \tilde{V}|\tilde{U}\right) > \lambda_n\} \tag{7}$$

$$E_2 = \{(u^n, v^n) : \min I\left(\tilde{X}_1; \tilde{V}|\tilde{U}\right) \leq \lambda_n\} \tag{8}$$

where for each pair of (u^n, v^n), the minimization of mutual information above is over all triples $(\tilde{X}_1, \tilde{U}, \tilde{V})$ of random variables, which respectively take values over $\mathcal{X}_1$, $\mathcal{U}$, and $\mathcal{U}$, and have distributions satisfying the following constraints:

$$\begin{aligned} &P_{\tilde{U}\tilde{V}}(\mathsf{u}, \mathsf{v}) = P_{u^n v^n}(\mathsf{u}, \mathsf{v}) \\ &\left|P_{\tilde{X}_1|\tilde{U}}(\mathsf{x}_1|\mathsf{u}) - P_{X_1|U}(\mathsf{x}_1|\mathsf{u})\right| \leq \tilde{\mu}_n \\ &\left|P_{\tilde{X}_1|\tilde{V}}(\mathsf{x}_1|\mathsf{v}) - P_{X_1|U}(\mathsf{x}_1|\mathsf{v})\right| \leq \tilde{\mu}_n. \end{aligned} \tag{9}$$

Note that for $(u^n, v^n) \in E_2$, $\min I(\tilde{X}; \tilde{V}|\tilde{U}) \leq \lambda_n$. To simplify notation, let $(\tilde{X}, \tilde{U}, \tilde{V})$ be the choice that achieves the minimum. Then by the Pinsker inequality [10, Lemma 11.6.1], there exists a constant k such that

$$\left|P_{\tilde{V}|\tilde{U}}(\mathsf{v}|\mathsf{u}) - P_{\tilde{V}|\tilde{U},\tilde{X}_1}(\mathsf{v}|\mathsf{u}, \mathsf{x}_1)\right| \leq k\sqrt{\lambda_n}.$$

This inequality itself also implies that there exists another constant k' such that

$$\left|\sum_u P_{\tilde{V}|\tilde{U}}(\mathsf{v}|u) P_{U|X_1}(u|\mathsf{x}_1) - P_{U|X_1}(\mathsf{v}|\mathsf{x}_1)\right| < k'\sqrt{\lambda_n}.$$

This situation though is analyzed in [9], and can be shown to imply that there exists another scalar k'' giving $\left|P_{\tilde{V}|\tilde{U}}(\mathsf{v}|\mathsf{u}) - \mathbb{1}(\mathsf{v} = \mathsf{u})\right| \leq k''\sqrt{\varepsilon}$, provided that $(P_{U|X_1}, I)$ is non-manipulable. Consequently, it can be shown that $(v^n, \mathbf{c}_1^n(w_1), \mathbf{c}_2^n(w_2)) \in T^n_{[U,X_1,X_2]_{2\tilde{k}\sqrt{\lambda_n}}}$ if $u^n \in \mathcal{U}(\mathbf{c}_1^n(w_1), \mathbf{c}_2^n(w_2))$, for some constant $\tilde{k}$. By setting, $\nu_n = \tilde{k}\sqrt{\lambda_n}$, we can conclude that $\{U^n \in \mathcal{U}(\mathbf{C}_1^n(W_1), \mathbf{C}_2^n(W_2)), (U^n, V^n) \in E_2\}$ is not an error event under H_1. Therefore we can bound $\Pr\{g_1^n(V^n, W_1) \notin \{W_2, \mathbf{!}\}\}$ as below:

$$\begin{aligned} &\Pr\{g_1^n(V^n, W_1) \notin \{W_2, \mathbf{!}\}\} \\ &\quad \leq \Pr\{U^n \notin \mathcal{U}\left(\mathbf{C}_1^n(W_1), \mathbf{C}_2^n(W_2)\right)\} \\ &\qquad + \Pr\{U^n \in \mathcal{U}\left(\mathbf{C}_1^n(W_1), \mathbf{C}_2^n(W_2)\right), \\ &\qquad\qquad V^n \in \mathcal{V}_{2\mu_n}(\mathbf{C}_1^n(W_1); W_2, \mathbf{C}_2^n),\ (U^n, V^n) \in E_1\} \end{aligned} \tag{10}$$

The probability $\Pr\{U^n \notin \mathcal{U}\left(\mathbf{C}_1^n(W_1), \mathbf{C}_2^n(W_2)\right)\}$ has been shown to vanish above (cf. (6)). It thus remains to show that the second probability on the right hand side of (10) decreases to zero as $n \to \infty$.

To that end, define the following sets for convenience:

$$\mathcal{Q}_1(u^n, v^n; \mathbf{c}_1^n) \triangleq \left\{ w_1 : \mathbf{c}_1^n(w_1) \in T^n_{[X_1|U]_{\tilde{\mu}_n}}(u^n) \bigcap T^n_{[X_1|U]_{\tilde{\mu}_n}}(v^n) \right\},$$

$$\mathcal{Q}_2(u^n, v^n; \mathbf{c}_1^n, \mathbf{c}_2^n) \triangleq \left\{ w_2 : \mathbf{c}_2^n(w_2) \in \bigcup_{w_1 \in \mathcal{Q}_1(u^n, v^n; \mathbf{c}_1^n)} T^n_{[X_2|U,X_1]_{\mu''_n}}(u^n, \mathbf{c}_1^n(w_1)) \right\},$$

$$\mathcal{Q}_3(w_2; u^n; \mathbf{c}_1^n, \mathbf{c}_2^n) \triangleq \left\{ w_1 : \mathbf{c}_1^n(w_1) \in T^n_{[X_1|U,X_2]_{\mu'_n}}(u^n, \mathbf{c}_2^n(w_2)) \right\}.$$

Because our goal is to count the number of codewords, $\mathbf{c}_1^n(w_1)$ such that $\mathbf{c}_1^n(w_1) \in T^n_{[X|UX_2]}(u^n, \mathbf{c}_2^n(w_2)) \cap T^n_{[X|V]}(v^n)$ for some $\mathbf{c}_2^n(w_2)$, these sets are of particular importance. First, the set $\mathcal{Q}_1$ can be viewed as the total number of possible $\mathbf{c}_1^n(w_1)$ that will be typical with both u^n and v^n. Secondly, the set $\mathcal{Q}_2$ is the number of $\mathbf{c}_2^n(w_2)$ that will be typical with u^n and a $\mathbf{c}_1^n(w_1)$ in set $\mathcal{Q}_1$. Finally, $\mathcal{Q}_3$ can be viewed as a bookkeeping necessity, which bounds the maximum number of values in $\mathcal{Q}_1$ per value in $\mathcal{Q}_2$. By [11, Problem 2.10], there exists ζ_n satisfying the delta convention that

$$\left| T^n_{[X_1|U]_{\tilde{\mu}_n}}(u^n) \cap T^n_{[X_1|U]_{\tilde{\mu}_n}}(v^n) \right| \leq 2^{n\left[\max H(\tilde{X}_1|\tilde{U},\tilde{V}) + \zeta_n\right]},$$

where $\tilde{X}_1, \tilde{U}, \tilde{V}$ are restricted to distributions that satisfy (9). Through judicious use of this and [11, p. 409, Lemma 17.9], the following bounds can be shown:

$$\Pr\left\{ |\mathcal{Q}_1(u^n, v^n; \mathbf{C}_1^n)| > \gamma 2^{n\left[\left|R_1 - \min I(\tilde{X}_1; \tilde{U}, \tilde{V})\right|^+ + \zeta_n + \epsilon_n\right]} \right\} \leq e^{-\sigma(\gamma) 2^{n\left[\left|R_1 - \min I(\tilde{X}_1; \tilde{V}|\tilde{U})\right|^+ + \zeta_n + \epsilon_n\right]}},$$

$$\Pr\left\{ |\mathcal{Q}_2(u^n, v^n; \mathbf{C}_1^n, \mathbf{C}_2^n)| > \gamma 2^{n\left[\left|R_2 - I(X_2;U|X_1) + \left|R_1 - \min I(\tilde{X}_1;\tilde{U},\tilde{V})\right|^+\right|^+ + \zeta_n + 3\epsilon_n\right]}, \ |\mathcal{Q}_1(u^n, v^n; \mathbf{C}_1^n)| \leq \gamma 2^{n\left[\left|R_1 - \min I(\tilde{X}_1;\tilde{U},\tilde{V})\right|^+ + \zeta_n + \epsilon_n\right]} \right\} \leq e^{-\sigma(\gamma) 2^{n\left[\left|R_2 - I(X_2;U|X_1) + \left|R_1 - \min I(\tilde{X}_1;\tilde{U},\tilde{V})\right|^+\right|^+ + \zeta_n + 3\epsilon_n\right]}},$$

$$\Pr\left\{ \max_{w_2} |\mathcal{Q}_3(w_2; u^n; \mathbf{C}_1^n, \mathbf{C}_2^n)| > \gamma 2^{2n\epsilon_n} \right\} \leq e^{-\sigma(\gamma) 2^{2n\epsilon_n} + nR_2 \ln 2},$$

for any $\gamma > 1$ and $\sigma(\gamma) \triangleq \gamma \ln \gamma - \gamma + 1$. Hence all the three probabilities above can be bounded by the double exponential term $e^{-\sigma(\gamma)2^{n\epsilon'_n}}$ for some ϵ'_n with the property that $n\epsilon'_n \to \infty$. Using this and (4), we can bound

$$\Pr\left\{ U^n \in \mathcal{U}\left(\mathbf{C}_1^n(W_1), \mathbf{C}_2^n(W_2)\right), V^n \in \mathcal{V}_{2\mu_n}(\mathbf{C}_1^n(W_1); W_2, \mathbf{C}_2^n), (U^n, V^n) \in E_1 \right\}$$

$$\leq 2^{-n[H(U|X_1X_2) + R_1 + R_2 - \epsilon_n]} \sum_{u^n, v^n} \mathbb{1}\left((u^n, v^n) \in E_1\right) \cdot \mathbb{1}\left(u^n \in T^n_{[U]_{\hat{\mu}_n}}\right) \mathbb{1}\left(v^n \in T^n_{[U]_{\hat{\mu}_n}}\right) \sum_{\mathbf{c}_1^n, \mathbf{c}_2^n} p(v^n|u^n, \mathbf{c}_1^n, \mathbf{c}_2^n) \cdot |\mathcal{Q}_2(u^n, v^n; \mathbf{c}_1^n, \mathbf{c}_2^n)| \cdot \max_{w_2} |\mathcal{Q}_3(w_2; u^n; \mathbf{c}_1^n, \mathbf{c}_2^n)| \, p(\mathbf{c}_1^n, \mathbf{c}_2^n)$$

$$\leq 3e^{-\sigma(\gamma)2^{n\epsilon'_n} + n[H(U) + I(X_1, X_2; U) - R_1 - R_2 + 3\epsilon_n] \ln 2} + \gamma^2 2^{-n[R_1 + R_2 - I(X_1, X_2; U) - \zeta_n - 7\epsilon_n]} + \gamma^2 2^{-n[\lambda_n - \zeta_n - 8\epsilon_n]}.$$

Hence by setting $\lambda_n = 2\zeta_n + 8\epsilon_n$, the upper bound vanishes. Thus particular choice of λ_n is of consequence to the set defined by (8), which can now be seen to be shrinking. One may conclude from this that total percentage of values of changed by the relay must grow asymptotically small to avoid detection, and ad infinitum must proportionally vanish. Thus to avoid detection, the relay must not alter the sequence.

V. Conclusion

We have presented an inner bound on the capacity region of which two nodes can exchange information with guaranteed integrity through a two-way Byzantine relay. The inner bound is specified by the non-manipulability property of the channel. The coding scheme that achieves the bound requires neither any pre-shared secret nor secret transmission. As a result, the inner bound is generally larger than any achievable region that is obtained based on secret transmission. Finally we point out that the coding scheme described can be easily modified to support the additional requirement of secret transmission.

References

[1] U. Maurer, "Information-theoretically secure secret-key agreement by not authenticated public discussion," in *Advances in Cryptology — EUROCRYPT '97* (W. Fumy, ed.), vol. 1233 of *Lecture Notes in Computer Science*, pp. 209–225, Springer-Verlag, May 1997.

[2] M. Bloch and J. Barros, *Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge University Press, 2011.

[3] E. Gilbert, F. J. MacWilliam, and N. J. A. Sloane, "Codes which detect deception," *Bell Syst. Tech. J.*, vol. 53, no. 3, pp. 405–424, 1974.

[4] R. Cramer, Y. Dodis, S. Fehr, C. Padró, and D. Wichs, "Detection of algebraic manipulation with applications to robust secret sharing and fuzzy extractors." Cryptology ePrint Archive, Report 2008/030, 2008.

[5] O. Kosut, L. Tong, and D. Tse, "Nonlinear network coding is necessary to combat general byzantine attacks," in *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, pp. 593 –599, Oct. 2009.

[6] D. Wang, D. Silva, and F. Kschischang, "Robust network coding in the presence of untrusted nodes," *IEEE Transactions on Information Theory*, vol. 56, pp. 4532–4538, Sep. 2010.

[7] X. He and A. Yener, "Secure communication with a byzantine relay," in *Proc. IEEE International Symposium on Information Theory*, pp. 2096–2100, July 2009.

[8] Y. Mao and M. Wu, "Tracing malicious relays in cooperative wireless communications," *IEEE Trans. Inform. Forensics and Security*, vol. 2, pp. 198–212, June 2007.

[9] E. Graves and T. F. Wong, "Detectability of symbol manipulation by an amplify-and-forward relay," *CoRR*, vol. abs/1205.2681, 2012.

[10] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York, NY, USA: Wiley-Interscience, 2nd ed., 2006.

[11] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2nd ed., 2011.