Active Adversaries from an
Information-Theoretic Perspective:
Data Modification Attacks

Mahtab Mirmohseni and Panagiotis Papadimitratos
KTH Royal Institute of Technology, Stockholm, Sweden
Email: {mahtabmi,papadim}@kth.se

Abstract

We investigate the problem of reliable communication in the presence of active adversaries that can tamper with the transmitted data. We consider a legitimate transmitter-receiver pair connected over multiple communication paths (routes). We propose two new models of adversary, a “memoryless” and a “foreseer” adversary. For both models, the adversaries are placing themselves arbitrarily on the routes, keeping their placement fixed throughout the transmission block. This placement may or may not be known to the transmitter. The adversaries can choose their best modification strategy to increase the error at the legitimate receiver, subject to a distortion limit. We investigate the communication rates that can be achieved in the presence of the two types of adversaries and the channel (benign) stochastic behavior. For memoryless adversaries, the capacity is derived. Our method is to use the typical set of the anticipated received signal in all possible adversarial strategies (including their best one) in a compound channel that also captures adversarial placement. For foreseer adversaries, with enhanced observation capabilities, we propose a new coding scheme to guarantee resilience, i.e., recovery of the code word independently of the adversarial (best) choice. For this case, we derive the achievable rate. Moreover, we propose an upper bound on the capacity. We evaluate our general results for specific cases (e.g., binary symbol replacement or erasing attacks), to gain insights.

Index Terms

Physical-layer active adversaries; Modification attacks; Replacement attacks; Erasing attacks; Multi-route transmission.
I. INTRODUCTION

Operation in adverse networks requires secure and reliable communication: data modifications should not be merely detected but data should be delivered (decoded correctly) at their destination. Cryptographic primitives can ensure detection but not correction and thus data delivery. Consider a general network connecting a Transmitter (Tx) - Receiver (Rx) pair over multiple disjoint communication paths (e.g., multiple frequency bands or antennas in wireless networks, or multiple routes in multi-hop networks); adversaries can be present in a number of those paths. The challenge is how to leverage the alternative paths and achieve reliable communication in the presence of the adversary. What is the best one can do against a powerful adversary? More generally, what is the best communication rate one can achieve in the face of malicious faults (adversarial modifications) and benign faults, due to the communication channel stochastic behavior?

Facets of this problem were addressed in the literature. One approach leverages cryptographic primitives to detect modifications and attempt retransmissions over alternative communication paths (while introducing redundancy to tolerate faults) [1]. This, however, does not address the fundamental limits of the system performance. Without cryptographic assumptions, with Tx-Rx communication over \( n \) disjoint paths, termed wires, and disrupted by active adversaries that compromise a subset of these wires, the minimum needed connectivity is derived [2]; the scenario is referred to as the Dolev model. This work does not consider communication rates and thus does not even attempt to achieve the best performance. It does not model channel noise and does not consider adversarial limitations or fine-grained actions.

In contrast, confidentiality has received significant attention, notably after Wyner’s seminal paper [3], with the majority of works concerned with passive eavesdroppers [4, Chapter 22]. Less attention, in an information-theoretic sense, was paid to active adversaries that modify the channel input of the legitimate transmitter. An early characteristic model is the Arbitrarily Varying Channel (AVC) [5], which assumes worst-case interference: the adversary controls the channel state to maximize the error probability at the receiver. Based on what the adversary knows and the common randomness of the legitimate nodes, the capacity can differ considerably [6], [7]. However, it is not easy to translate erasing and replacement attacks to the AVC worst-case interference notations. Therefore, AVC cannot capture data modification attacks or network structure, e.g., as per the Dolev model [2]. Given that confidentiality (passive adversaries) is broadly researched in the information-theoretic sense (also in [2]), the challenge is how achieve reliable communication in the presence of active adversaries, in addition to channel noise, and derive fundamental limits of the capacity?
In this paper, we address this challenge. We propose a novel information-theoretic setup that captures network structure, fine-grained and strong yet realistic active adversarial behavior, along with channel stochastic behavior. We consider a Tx-Rx pair communicating across a number of disjoint paths (routes). The adversaries compromise a fixed number of these routes, thus they get access to the (noiseless) transmitted signal. The adversaries can choose their best strategy (knowing the transmitted signal) to modify and increase the error at the Rx, subject to a distortion limit. This limit, given a distortion measure (depending on the specific attack), determines the distance between the transmitted codeword and its modified version; e.g., for an erasing attack on binary transmissions, the percentage of bits the adversary can erase. The adversaries’ placement (on the routes) is arbitrary but fixed throughout one transmission block and it may be known to the Tx. The adversaries’ observation (of the transmitted signal) can be either instantaneous or cover the entire codeword. We propose accordingly two adversary types, memoryless and foreseer. Our goal is to find the reliable communication rate the Tx-Rx can achieve in each case.

Our average distortion limit and the consideration of channel stochastic behavior (noise) on top of adversarial faults lead to a generalized model over the Dolev one for active adversaries. Our distortion limit can be omitted by setting it to 1 (given it is defined as an average), reducing our channel to the Dolev model [2]. The distortion limit allows practical assumptions, e.g., adversaries with noisy observation, with tactics to remain undetected, limited resources or time or attempts to mount an attack, or even cryptographic integrity protection for parts of the messages (e.g., immutable fields). Moreover, the channel noise is not taken into account in the Dolev model.

We derive the capacity for the memoryless adversaries. For the achievability part, we use a compound channel to model the adversaries’ placement. In each compromised route, we consider the typical set of the anticipated received signal in all possible adversarial scenarios (including the one for the best adversarial strategy), subject to the distortion limit. For the foreseer adversaries, we propose a coding scheme using two techniques: (i) the Hamming approach [8], to cope with the worst-case errors inflicted by the adversaries with access to the entire codeword and (ii) a random coding argument to recover from the channel stochastic noise. For the former, we use the Varshamov construction [9], to guarantee the required minimum distance needed to mitigate adversary-inflicted errors. Moreover, we obtain an upper bound to the capacity, taking an approach similar to Hamming bound (i.e., limiting the volume of the Hamming balls). Finally, we gain insights through three special cases: replacement and erasing attacks on binary transmission and Gaussian jamming. We determine the proper distortion measures and channel distributions to model attacks that correspond to realistic situations, e.g., bit or packet
replacement and dropping (selective forwarding) and evaluate our derived rates for those. For these cases, we consider explicitly the best strategy of the adversaries: we show they can achieve the lower bounds on the capacity we derived without specific assumptions on the adversary strategy. This reveals that (i) knowing the adversaries’ placement is not useful in terms of the achievable reliable rate, (ii) memory helps the adversaries significantly, and (iii) differentiates the foreseer effect from channel noise; while the memoryless effect is equivalent to channel noise.

II. CHANNEL MODEL

Notation: Upper-case letters (e.g., $X$) denote Random Variables (RVs) and lower-case letters (e.g., $x$) their realizations. $X_i^j$ indicates a sequence of RVs $(X_i, X_{i+1}, ..., X_j)$; we use $X^j$ instead of $X_i^j$ for brevity. The probability mass function (p.m.f) of a RV $X$ with alphabet set $\mathcal{X}$ is denoted by $p_X(x)$; occasionally subscript $X$ is omitted. All possible distributions on $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$. $\pi(x, y | x^n, y^n) \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ shows the joint type (i.e., empirical p.m.f) of two sequences of length $n$, which can be extended to several $n$-length sequences. $\mathcal{P}_n(\mathcal{X}) \subset \mathcal{P}(\mathcal{X})$ consists all possible types of sequences $x^n \in \mathcal{X}^n$. For $q \in \mathcal{P}_n(\mathcal{X})$, the type class is defined as $\mathcal{T}_n(q) = \{x^n, p_X(x) = q\}$. $A^n(X, Y)$ is the set of $\epsilon$-strongly, jointly typical sequences of length $n$. $\mathcal{N}(0, \sigma^2)$ denotes a zero-mean Gaussian distribution with variance $\sigma^2$. $\mathcal{B}(\alpha)$ is a Bernoulli distribution with parameter $\alpha \in [0, 1]$. $\mathbb{F}_q$ is the finite field with $q$ elements. We define $[x]^+ = \max\{x, 0\}$. Unless specified, Logarithms are in base 2. Throughout the paper, $i$ and $j$ indices are used for time and route number, respectively. $H_q : [0, 1] \rightarrow \mathbb{R}$ is the Hilbert $q$-ary entropy function $H_q(x) = x \log_q(q-1) - x \log_q x - (1-x) \log_q(1-x)$. Bold letters are used to show the column vectors of length $n_r$, e.g., $x^n = \begin{bmatrix} x_1; & \cdots & x_n \end{bmatrix}$ and $x^n(j)$ shows its $j$th row.

Channel model: Consider a single unicast scenario: Tx sends a message $M$ to Rx, with $n_r$ available disjoint routes. $n_a$ out of the $n_r$ routes are attacked by the adversaries, with their placement being arbitrary but fixed throughout one transmission block. The placement can be chosen by the adversaries...
to maximize the error at Rx and it may be known to the Tx. One can implicitly assume there are \( n_a \) adversaries: more than one adversary in a route can be modeled with an stronger adversary (i.e., with a higher distortion limit). We model this scenario with a (compound) state-dependent multi-route Point-to-Point channel with Modifying Adversaries (PP-MA) in Fig. 1: its transition probability is not entirely specified unless the Channel State Information (CSI) (i.e., adversary presence information) is known [10]. Consider finite alphabets \( \mathcal{X}, \mathcal{X}_a, \mathcal{Y} \). Channel inputs at Tx and the adversaries are defined by \( \mathbf{X} \in \mathcal{X}^{n_r} \) and \( \mathbf{X}_a \in \mathcal{X}_a^{n_r} \) respectively. \( \mathbf{Y} \in \mathcal{Y}^{n_r} \) is the output of the channel at Rx. The \( j \)-th element of state vector \( \mathbf{S} \in \{0, 1\}^{n_r} \), i.e., \( \mathbf{S}(j) \), determines the presence of an adversary in \( j \)-th route. The received signal at Rx only depends on the adversary input, if present. Each adversary chooses its channel input subject to a distortion limit (knowing the Tx input). Hence, we define the \( D \) class of adversaries for \( j \in \{1, \ldots, n_r\} \) by all probability distributions:

\[
P_j(D) = \{ p_j(x^n_a|x^n) : E_{p_j}[d(X^n_a, X^n)] \leq D \} \tag{1}\]

where \( d \) is a distortion measure defined by the mapping \( d : \mathcal{X} \times \mathcal{X}_a \mapsto [0, \infty) \) and the average distortion for two sequences is \( d(x^n_a, x^n) = \frac{1}{n} \sum_{i=1}^{n} d(x_{a,i}, x_i) \). We assume the \( \mathcal{X}_a \mapsto \mathcal{Y} \) channel is memoryless, thus the transition probability can be shown by the conditional p.m.f on \( \mathcal{Y} \times \mathcal{X}_a \) as:

\[
p(y^n, x^n_a|x^n, s^n) = p(y^n|x^n_a)p(x^n_a|x^n, s^n) \tag{2}
\]

\[
= \prod_{j=1}^{n_a} p_j(y^n(j)|x^n(j))p_j(x^n(j)|x^n, s^n(j))
\]

\[
= \prod_{j=1}^{n_r} p_j(x^n_a(j)|x^n(j), s^n(j))\prod_{i=1}^{n} p_j(y_i(j)|x_{a,i}(j))
\]

The state vector, assumed fixed in one transmission block, models the channel statistics transmission block: \( s_i(j) = s(j) \) for \( i \in \{1, \ldots, n\} \) with at most \( n_a \leq n_r \) adversaries, i.e., \( w_H(s) \leq n_a \). Hence for \( j \in \{1, \ldots, n_r\} \):

\[
p_j x_a|x, s(x^n_a(j)|x^n(j), s^n(j)) = p_j x_{a,s}|x(x^n_a(j)|x^n(j)) = q_j, s(x^n_a(j)|x^n(j)) \tag{3}\]

where

\[
q_j, s(x^n_a|x^n) \in P_j^n(D_j, s = s \cdot D_j) \tag{3}\]

which is due to the \( D_j \) distortion limit at each adversary. In \( n \) channel uses, Tx sends \( M \) to Rx using the following code:

**Definition 1:** A \( (2^nR, n, P_e^{(n)}) \) code for multi-route PP-MA consists of:
1) A message set $\mathcal{M} = [1 : 2^{nR}]$, where $M$ is uniformly distributed over $\mathcal{M}$.

2) An encoding function $f^n$ at Tx that maps the message $M$ to a codeword $x^n \in \mathcal{X}^{n_r \times n}$.

3) A set of adversaries’ mapping $h^n$ where $h^n(j) : \mathcal{X}^n \times \{0, 1\} \mapsto \mathcal{X}_a^n$ for $j \in \{1, \ldots, n_r\}$ satisfying (3).

4) A decoding function at Rx, $g : Y^{n_r \times n} \mapsto \mathcal{M}$.

5) Probability of error for this code is defined as:

$$P_e^n = \frac{1}{2^{nR}} \sum_{m \in \mathcal{M}} Pr(g(y^n) \neq m | m \text{ sent}).$$

In case of available CSI at Tx, we have: $f^n : \mathcal{M} \times \{0, 1\}^{n_r} \mapsto \mathcal{X}^{n_r \times n}$. All codewords are revealed to all nodes (including adversaries). However, the adversaries’ mapping is not known to the legitimate user.

**Definition 2:** A rate $R$ is achievable if there exists a sequence of $(2^{nR}, n, P_e^n)$ codes such that for $\forall s \in \{0, 1\}^{n_r}$ : $w_H(s) \leq n_a$ and $\forall h^n$, we have $P_e^n \to 0$ as $n \to \infty$. The capacity $C$ is the supremum of all achievable rates $R$.

**Memoryless active adversary:** The mapping at each adversary satisfies:

$$p_j(x^n_a(j) | x^n(j), s^n(j)) = \prod_{i=1}^{n} p_j(x_{a,i}(j) | x_i(j), s_i(j))$$

i.e., the adversary uses the same probability distribution to modify the transmitted symbols in each channel use. For each route the distribution in (5) is independent and identically distributed (i.i.d) and fixed over time; but the distributions can differ across routes.

**Foreseeer active adversary:** It observes the transmitted codeword in the entire block (i.e., $x^n(j)$) upon which it bases its strategy. That is, satisfying (3), the adversary can choose the position and value of the symbols in the codeword to be modified. In this case, we concentrate on two types of attacks:

**Replacement attacks:** $\mathcal{X} = \mathcal{X}_a = \mathcal{Y}$ with hamming distortion measure:

$$d(x, \hat{x}) = \begin{cases} 1, & \text{if } x \neq \hat{x} \\ 0, & \text{if } x = \hat{x} \end{cases}$$

**Erasing (dropping) attacks** (also known as selective forwarding): $\mathcal{X}_a, \mathcal{Y} = \{\mathcal{X}, e\}$ where for all $x, x' \in \mathcal{X}, x \neq x', d(x, x) = 0, d(x, x') = \infty$ and $d(e, x) = d(x, e) = 1$. By this definition, we limit the adversaries only to erase the data and they cannot replace as long as their distortion limits are finite, i.e., $D_j < \infty$ for $j \in \{1, \ldots, n_r\}$.

These two types cover all possible modification attacks. It is reasonable to assume that anything outside the alphabet is rejected by Rx; thus, this can be modeled as an erased symbol. Therefore, the adversary does not gain by modifying to a non-existent symbol.
III. MAIN RESULTS

For the multi-route PP-MA, for memoryless and foreseer adversaries, we consider either no CSI or CSI at Tx. The adversaries have perfect CSI.

A. Memoryless active adversaries

We state the capacity for the channel in (5), first assuming no CSI available at the Tx and Rx.

**Theorem 1:** The capacity of multi-route PP-MA satisfying (5), with no CSI available at either the Tx or the Rx is:

$$C_{i.i.d}^{n} = \sup_{p(x)} \min_{w_H(s) \leq n_a} \inf_{\sum_{j=1}^{n_r} I(X(j); Y_s(j)) \geq 1} p(x) \prod_{j=1}^{n_r} p_j(x_a(j)|x(j), s(j))$$

where \( \forall s \in \{0, 1\}^{n_r} \), we have \( Y_s \in \mathcal{Y}^{n_r} \) and

$$p(y^n|x^n, x^n_0, s^n) = \prod_{j=1}^{n_r} p_j y|x_a(j), x, s(y^n(j)|x^n(j), x^n_0(j), s^n(j)) = \prod_{j=1}^{n_r} p_j y|x_a(j), x(y^n(j)|x^n(j), x^n_0(j)).$$

Hence, the mutual information term is evaluated with respect to the joint p.m.f (2).

**Proof outline:** For the achievability part, we use a random coding argument in a compound channel (to model the adversaries’ placement). To take into account all possible i.i.d adversaries’ strategies, we consider all possible joint types of \((x^n(j), y^n(j))\) for the \(j\)-th route, subject to the distortion limit on \(x^n_0(j)\). The converse follows from Fano’s inequality, by noting that for every adversaries’ placement and mapping \((s, h^n)\) we must have \(P_e(n) \rightarrow 0\). Detailed proof in Appendix.

**Remark 1:** On the \(j\)-th route, the conditional distribution of the adversary’s channel input, i.e., \(p_j(x_a(j)|x(j), s(j))\), can model all possible memoryless active attacks (e.g., replacement or dropping). To specify a certain attack, it is enough to properly define the input alphabets \(X_a, X\) and the distortion measure \(d(\cdot, \cdot)\). Thus, \(\inf\) is calculated over a feasible set of \(X \times X_a\) distributions \((p_j)\) where the feasibility constraint is determined by \(E_{p_j}[d(x_a(j), x(j))] \leq D_{j,s} = s \cdot D_j\).

Next, we obtain the capacity when CSI is available at Tx (proof in Appendix).

**Theorem 2:** The capacity of multi-route PP-MA satisfying (5), with CSI available to Tx is:

$$C_{i.i.d}^{TC} = \min_{s \in \{0, 1\}^{n_r}} \sup_{p(x)} \inf_{\sum_{j=1}^{n_r} I(X(j); Y_s(j)) \geq 1} p(x) \prod_{j=1}^{n_r} p_j(x_a(j)|x(j), s(j))$$

where \( \forall j \in \{1, \ldots, n_r\} : E_{p_j}[d(x_a(j), x(j))] \leq D_{j,s} = s \cdot D_j\), the notation \(Y_s\) is defined in Theorem 1 and the mutual information term is evaluated with respect to the joint p.m.f (2).
B. Foresee active adversaries

Now, we derive lower and upper bounds on the capacity for all possible foresee adversaries strategies. The bounds are based on the possible minimum distances the legitimate user codewords can tolerate under each attack.

Theorem 3: A lower bound to the capacity of the multi-route PP-MA with foresee adversaries (no CSI available at Tx or Rx) is:

\[
\mathcal{R}_I^{\text{nc}} = \sup \inf_{\mathbf{h}^n} \sum_{j=1}^{n_r} \left[ H(\mathbf{V}) - H(\mathbf{X}_a(j)|\mathbf{Y}(j)) - \frac{H[\mathbf{X}](d_j)}{\log|\mathbf{X}|} \right]^+
\]

where the supremum and the minimum are taken over \( p(\mathbf{x})p(\mathbf{v}|\mathbf{x}); \forall j \in \{1, \ldots, n_r\} : E_{\mathbf{p}_j}[d(\mathbf{v}(j), \mathbf{x}(j))] \leq d_j \) and \( s \in \{0,1\}^{n_r} : w_H(s) \leq n_a, \) respectively; \( d_j = f(D_{j,a} = s(j) \cdot D_j) \) is determined based on the attack type and the distortion measure (i.e., \( d_j = s(j) \cdot 2D_j \) for replacement attacks and \( d_j = s(j) \cdot D_j \) for erasing attacks); the second entropy is evaluated with respect to \( \mathbf{p}_j(\mathbf{y}^n(j)|\mathbf{x}_a^n(j)) = \prod_{i=1}^{n} \mathbf{p}_j(\mathbf{y}_i(j)|\mathbf{x}_{a,i}(j)) \) (Note: this channel is memoryless).

Proof outline: We apply a random coding technique on top of a random linear code (Varshamov construction [9]), by introducing proper auxiliary codewords. Random coding is used to combat the stochastic behavior of the channel and the Varshamov construction guarantees recovery from the worst-case errors, by making the minimum distance of the code greater than the number of errors. First, we generate auxiliary codewords \( u \) and then we apply a random linear coding \( n_r \) times to these codewords, to generate the transmitted codewords \( x^n \). To decode from the \( j \)-th route, if Rx can decode the adversary’s channel input \( x_a^n(j) \), the transmitted codeword is the only \( x^n(j) \) in a Hamming ball with radius \( d_j \). To apply this scheme, we choose \( v^n(j) \) as the possible \( x_a^n(j) \) and try to decode it after receiving \( y^n(j) \), by decreasing its rate to satisfy the stochastic limitation forced by \( \mathcal{X}_a \mapsto \mathcal{Y} \) channel. Proof details in Appendix.

Theorem 4: A lower bound on the capacity of the multi-route PP-MA with foresee adversaries (CSI available at Tx) is:

\[
\mathcal{R}_I^{\text{nc}} = \min \sup \inf_{\mathbf{h}^n} \sum_{j=1}^{n_r} \left[ H(\mathbf{V}) - H(\mathbf{X}_a(j)|\mathbf{Y}(j)) - \frac{H[\mathbf{X}](d_j)}{\log|\mathbf{X}|} \right]^+
\]

where the minimum and the supremum are taken over \( s \in \{0,1\}^{n_r} : w_H(s) \leq n_a \) and \( p(\mathbf{x})p(\mathbf{v}|\mathbf{x}); \forall j \in \{1, \ldots, n_r\} : E_{\mathbf{p}_j}[d(\mathbf{v}(j), \mathbf{x}(j))] \leq d_j, \) respectively; \( d_j \) and \( H_q(x) \) are defined in Theorem 3; the second entropy is evaluated with respect to \( \mathbf{p}_j(\mathbf{y}^n(j)|\mathbf{x}_a^n(j)) = \prod_{i=1}^{n} \mathbf{p}_j(\mathbf{y}_i(j)|\mathbf{x}_{a,i}(j)) \).
\textbf{Remark 2:} In both $R^\text{nc}_I$ and $R^\text{TC}_I$, the first term is independent of $s$ and the second term is independent of $p(x)$. Therefore, we have $R^\text{nc}_I = R^\text{TC}_I$. That is, CSI does not help the achieving strategy for these rates.

\textbf{Theorem 5:} These are upper bounds to the capacity of the multi-route PP-MA with foreseer adversaries:

$$R^\text{nc}_u = \sup p(x) \min_{h^n} \sum_{j=1}^{n_r} [I(X_a(j); Y(j)) - \frac{H(X)}{\log|X|} \frac{d_j}{2}]^+$$

$$R^\text{TC}_u = \min \sup p(x) \min_{h^n} \sum_{j=1}^{n_r} [I(X_a(j); Y(j)) - \frac{H(X)}{\log|X|} \frac{d_j}{2}]^+$$

where the minimum is taken over $s \in \{0, 1\}^{n_r} : w_H(s) \leq n_a$ and $d_j$ and $H_q(x)$ are defined in Theorem 3.

\textbf{Proof outline:} We follow an approach similar to the one used to derive the Hamming bound, which means we limit the volume of the coding balls. Detailed proof in Appendix.

\section*{IV. Examples}

\textbf{Replacement attacks to binary transmission:} The channel inputs and output have binary alphabets (i.e., $\mathcal{X}, \mathcal{X}_a, \mathcal{Y} = \{0, 1\}$) and $d$ is the Hamming distortion measure (see Section II). The stochastic channel from the adversary to the Rx is assumed as a Binary Symmetric Channel (BSC). Thus, the channel output at Rx at time $i \in \{1, \ldots, n\}$ is:

$$Y_i(j) = X_{a,i}(j) \oplus Z_i(j)$$

where $Z_i(j) \sim B(N_j)$ for $j \in \{1, \ldots, n_r\}$. First, consider memoryless active adversaries. We obtain the results of Theorems 1 and 2 as:

\textbf{Corollary 1:} The capacity of the multi-route PP-MA with binary alphabets satisfying (5), (11), for both no CSI and CSI at Tx, is:

$$C^\text{nc}_{i.t.d} = C^\text{TC}_{i.t.d} = n_r - \max_{s \in \{0, 1\}^{n_r}} \sum_{j=1}^{n_r} \max_{w_H(s) \leq n_a} H(N_j * N_j')$$

where $\hat{D}_j = \min\{D_j, 1 - D_j\}$ and $\alpha * \beta = \alpha(1 - \beta) + \beta(1 - \alpha)$. If we assume equal route conditions, $D_j = D \leq \frac{1}{2}$ and $N_j = N$ for $j \in \{1, \ldots, n_r\}$, the capacity is: $n_r - (n_r - n_a)H(N) - n_aH(N_D) - D_j$.

\textbf{Proof:} Let $P_j = Pr(X(j) = 1)$ and without loss of generality assume $P_j \leq \frac{1}{2}$. To find the inf $\sum_{j=1}^{n_r} I(X(j); Y_s(j))$ in (7), first we find a lower bound to it and then we show it is achievable by the adversaries.

$$I(X(j); Y_s(j)) = H(X(j)) - H(X(j) | Y_s(j))$$

$$\geq H(P_j) - H(X(j) \oplus X_{a}(j) \oplus Z(j)) \geq H(P_j) - H(N_j * N_j')$$
in (a) we define $N_j' \leq s(j) \cdot D_j$ and use $Pr(X(j) \neq X_a(j)) \leq s(j) \cdot D_j$. This lower bound is achievable by the $j$-th adversary if it chooses a joint distribution given by two backward BSCs $\mathcal{V} \rightarrow X_a$ and $X_a \rightarrow X$ with cross-over probabilities $N_j$ and $N_j'$, respectively. This results in $Pr(X_a(j) = 1) = \frac{P_j - N_j'}{1 - 2N_j'}$. Hence, we need $P_j \geq N_j'$ to hold. Therefore, (7) for this channel is:

$$C_{i,i,d}^{nC} = \sup_{0 \leq D_j \leq P_j \leq \frac{1}{2}} \min_{s \in \{0,1\}^n} \sum_{j=1}^{n_r} \left[ H(P_j) - \max_{N_j' \leq s(j) \cdot D_j} H(N_j * N_j') \right]$$

The rest of the proof is straightforward.

Now, consider foreseer active adversaries. We obtain the results of Theorems 3 and 4 as:

**Corollary 2:** The lower bound to the capacity of the multi-route PP-MA with foreseer adversaries, binary alphabets satisfying (11), for both no CSI and CSI at Tx is:

$$R_l^{nC} = R_l^{TC} = n_r - \sum_{j=1}^{n_r} H(N_j) - \max_{s \in \{0,1\}^n} \sum_{j=1}^{n_r} H(s(j) \cdot 2D_j)$$  \hspace{1cm} (12)

For equal route conditions, $D_j = D \leq \frac{1}{2}$ and $N_j = N$ for $j \in \{1, \ldots, n_r\}$, the rate is $n_r(1 - H(N)) - n_aH(2D)$.

**Proof:** Let $V_i(j) = X_i(j)$ and $P_j = Pr(X(j) = 1)$. Recall that for all $h^n$ (Definition 1 satisfying (3)), we have $Pr(X(j) \neq X_a(j)) \leq s(j) \cdot D_j$. After some calculations, we can compute $H(V(j)) = H(P_j)$ and

$$H(X_a(j)|Y(j)) \leq H(X_a(j) \oplus Y(j)) = H(Z_i(j)) = H(N_j)$$

and obtain (3) as:

$$R_l^{nC} \geq \sup_{0 \leq P_j \leq \frac{1}{2}} \min_{s \in \{0,1\}^n} \sum_{j=1}^{n_r} \left[ H(P_j) - H(N_j) - H(s(j) \cdot 2D_j) \right]$$  \hspace{1cm} (13)

which will be maximized for $P_j = \frac{1}{2}$ independent of $s(j)$ for $j \in \{1, \ldots, n_r\}$. This results in (12). It is easy to see that computing $R_l^{TC}$ in Theorem 4 results in the same rate.

We adapt Theorem 5 for binary alphabets and BSC of (12):

**Corollary 3:** The upper bound to the capacity of the multi-route PP-MA with foreseer adversaries, binary alphabets satisfying (11), for both no CSI and CSI at Tx, is:

$$R_u^{nC} = R_u^{TC} = n_r - \sum_{j=1}^{n_r} H(N_j) - \max_{s \in \{0,1\}^n} \sum_{j=1}^{n_r} H(s(j) \cdot D_j)$$  \hspace{1cm} (14)

For equal route conditions, $D_j = D$ and $N_j = N$ for $j \in \{1, \ldots, n_r\}$, the rate is $n_r(1 - H(N)) - n_aH(D)$.

**Proof:** We can combine the methods of Corollaries 1 and 2. We can show that the sum of the first and the second terms in the right side of (13) makes an upper bound on the first term of (9) and (10). To
do this, it is enough to choose the proper joint distribution for adversaries’ input that achieves this bound. This distribution consists of two backward BSCs $\mathcal{Y} \rightarrow \mathcal{X}_a$ and $\mathcal{X}_a \rightarrow \mathcal{X}$ with cross-over probabilities $N_j$ and $D_j$, respectively. The rest of the proof is similar to Corollary 2.

**Erasing attacks on binary transmission:** To reduce the erasing attacks to binary alphabets, we set: $\mathcal{X} = \{0,1\}$, $\mathcal{X}_a, \mathcal{Y} = \{0,1,e\}$, $d(0,0) = d(1,1) = 0$, $d(0,1) = d(1,0) = \infty$, and $d(0,e) = d(1,e) = 1$. In $\mathcal{X}_a \rightarrow \mathcal{Y}$ channel, additional erasing is introduced for the received signal at Rx (not distinguishable from the adversarial erasing at Rx). Thus, the channel output at Rx at time $i \in \{1,\ldots,n\}$ is:

$$Y_i(j) = \begin{cases} \text{BEC}(X_{a,i}(j), N_j), & \text{if } X_{a,i}(j) \neq e \\ X_{a,i}(j), & \text{if } X_{a,i}(j) = e \end{cases}$$

(15)

where $\text{BEC}(x, \beta)$ shows a Binary Erasure Channel (BEC) with input $x$ and probability of erasure $\beta$. Here, we state our results for both memoryless and foreseer adversaries. Proofs in Appendix.

**Corollary 4:** The capacity of the multi-route PP-MA with $\mathcal{X} = \{0,1\}$ and $\mathcal{X}_a, \mathcal{Y} = \{0,1,e\}$, satisfying (5) and (15), for both no CSI and CSI at Tx, is:

$$C_{i.i.d}^{ac} = C_{i.i.d}^{TC} = \min_{s \in \{0,1\}^n} \sum_{w_H(s) \leq n_a} n_r (1-s(j) \cdot D_j)(1-N_j).$$

For equal route conditions, $D_j = D$ and $N_j = N$ for $j \in \{1,\ldots,n_r\}$, the capacity is $(1-N)(n_r-n_a D)$.

**Corollary 5:** The lower bound to the capacity of the multi-route PP-MA with foreseer adversaries, $\mathcal{X} = \{0,1\}$ and $\mathcal{X}_a, \mathcal{Y} = \{0,1,e\}$, satisfying (15), for both no CSI and CSI at Tx, is:

$$R_i^{ac} = R_i^{TC} = \min_{s \in \{0,1\}^n} \sum_{w_H(s) \leq n_a} n_r \left(1-N_j(1-N'_j) - \tilde{N}_j H(N'_j) - H(s(j) \cdot D_j) \right)$$

where $\tilde{N}_j = N_j(1-s(j) \cdot D_j) + s(j) \cdot D_j$. For equal route conditions, $D_j = D$ and $N_j = N$ for $j \in \{1,\ldots,n_r\}$, the rate is $n_r(1-N) - n_a((N(1-D))H(D(1-N)/N)) + H(D(1-D)).$

**Corollary 6:** The upper bound to the capacity of the multi-route PP-MA with foreseer adversaries, $\mathcal{X} = \{0,1\}$ and $\mathcal{X}_a, \mathcal{Y} = \{0,1,e\}$, satisfying (15), for both no CSI and CSI at Tx, is:

$$R_u^{ac} = R_u^{TC} = \min_{s \in \{0,1\}^n} \sum_{w_H(s) \leq n_a} n_r \left[H((1-N'_j)(1-N_j)) + (1-N'_j)(1-N_j - H(N_j)) - H(s(j) \cdot D_j/2) \right]$$

where $N'_j = s(j) \cdot D_j$. For equal route conditions, $D_j = D$ and $N_j = N$ for $j \in \{1,\ldots,n_r\}$, the rate is $n_a(H((1-N)(1-D)) + (1-D)(1-N - H(N))) - H(D/2)) + (n_r-n_a)(1-N)$.

**Gaussian replacement attacks:** We assume Gaussian distributions for channel inputs and output. The distortion measure now is the squared error distortion:

$$d(x, \hat{x}) = (x - \hat{x})^2$$
and the channel model can be shown as:

\[ Y_i(j) = X_{n,i}(j) + Z_i(j) \]  \hspace{1cm} (16)

where \( Z_i(j) \sim \mathcal{N}(0, N_j) \) for \( j \in \{1, \ldots, n_r\} \) are independent and i.i.d Gaussian noise components.

We assume the average power constraint on input signal \( X(j) \) as \( \frac{1}{n} \sum_{i=1}^{n} |x_i(j)|^2 \leq P_j \). Hence, \( X(j) \sim \mathcal{N}(0, P_j) \) for \( j \in \{1, \ldots, n_r\} \). Here, we only consider the memoryless adversaries and obtain the results of Theorems 1 and 2 (proof in Appendix).

**Corollary 7:** The capacity of the multi-route PP-MA with Gaussian distributions for channel inputs and output, satisfying (5) and (16), for both no CSI and CSI at Tx, is:

\[
C_{n,i.d}^\text{nc} = C_{n,i.d}^\text{TC} = \max_{s \in \{0,1\}^{n_r}} \sum_{j=1}^{n_r} \theta(\frac{P_j - s(j) \cdot D_j + N_j}{s(j) \cdot D_j + N_j})
\]

where \( \theta(x) = \frac{1}{2} \log(x) \). For equal route conditions, \( D_j = D \) and \( N_j = N \) with equal power constraints \( P_j = P \) for \( j \in \{1, \ldots, n_r\} \), the capacity is: \( n_r \theta(1 + \frac{P}{N}) - n_a \theta(1 + \frac{D(P+2N)}{N(P-D+N)}) \).

**Comparison:** Along with equal route conditions, to simplify, let \( n_r = n_a = 1 \). Table I shows the results for the replacement and erasing attacks on binary transmission. Obviously, for zero distortion for the adversary \( (D = 0) \), we have BSC and BEC with parameter \( N \). The rate reduction caused by a foreseer adversary is considerable. Consider only the adversary’s effect by setting \( N = 0 \): the foreseer is twice more powerful than the memoryless one (in terms of the lower bound) for the replacement attack. For the erasing attack, the foreseer reduces (compared to the memoryless) the rate from a BEC rate \( (i.e., 1 - H(D)) \) to a BSC rate \( (i.e., 1 - H(D)) \). For Gaussian replacement attacks (under these simplified assumptions), the capacity is \( \frac{1}{2} \log(1 + \frac{P}{D+N}) \); while for Gaussian independent jamming with power \( D \), we achieve \( \frac{1}{2} \log(1 + \frac{P}{D+N}) \). Thus, knowing the transmitted codeword (even in a memoryless case) worsens the scenario compared to an independent jammer.

---

**TABLE I**

<table>
<thead>
<tr>
<th></th>
<th>Replacement</th>
<th>Erasing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memoryless Capacity</td>
<td>( 1 - H(N \cdot D) )</td>
<td>( (1 - N)(1 - D) )</td>
</tr>
<tr>
<td>Foreseer lower</td>
<td>( 1 - H(N) - H(2D) )</td>
<td>( 1 - N(1 - D) - (N(1 - D) + D)H(\frac{D}{N(1-D)+D}) - H(D) )</td>
</tr>
<tr>
<td>Foreseer upper</td>
<td>( 1 - H(N) - H(D) )</td>
<td>( H((1 - N)(1 - D)) + (1 - D)(1 - N - H(N)) - H(D) )</td>
</tr>
</tbody>
</table>

\[
<table>
<thead>
<tr>
<th>\text{Replacement}</th>
<th>1 - H(N \cdot D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erasing</td>
<td>( (1 - N)(1 - D) )</td>
</tr>
</tbody>
</table>

| Foreseer lower | \( 1 - H(N) - H(2D) \) | \( 1 - N(1 - D) - (N(1 - D) + D)H(\frac{D}{N(1-D)+D}) - H(D) \) |
| Foreseer upper  | \( 1 - H(N) - H(D) \)  | \( H((1 - N)(1 - D)) + (1 - D)(1 - N - H(N)) - H(D) \) |
REFERENCES


APPENDIX

Proof of Theorem 1:

Achievability: We use random encoding and joint typicality decoding. Considering the problem setup in Section II, we denote the set of possible joint types of triple of sequences \((x^n, x^a_n, y^n) \in \mathcal{X}^n \times \mathcal{X}^a_n \times \mathcal{Y}^n\) as:

\[
P^n_{j,s}(\mathcal{X} \times \mathcal{X}^a \times \mathcal{Y}) = \left\{ \pi(x, x_a, y|x^n, x^a_n, y^n) : E_{p^n_{j,s}}[d(X^n_a, X^n)] \leq D_{j,s} = s \cdot D_j \right\}
\]

and the possible pairs of \((x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n\) for some \(X^a_n \in \mathcal{X}^a_n\):

\[
P^n_{j,s}(\mathcal{X} \times \mathcal{Y}) = \left\{ \pi(x, y|x^n, y^n) : \exists X^a_n \in \mathcal{X}^a_n \text{ such that } (x^n, X^a_n, y^n) \in P^n_{j,s}(\mathcal{X} \times \mathcal{X}^a \times \mathcal{Y}) \right\}
\]

For \(q \in P^n_{j,s}(\mathcal{X} \times \mathcal{Y})\), the type class is defined as \(T^n_{j,s}(q) = \{(x^n, y^n), p_{XY}(x, y) = q\}\). Note that, since
the adversaries are memoryless, we have:

\[ p(y^n_s|x^n) = \prod_{i=1}^{n} p_{Y_s|X}(y_{s,i}|x_i) = \prod_{i=1}^{n} p_Y(x_s,y|x_{\cdot,i}, s) \]

\[ = \sum \prod_{x_a} \prod_{i=1}^{n} q_{j,s}(x_{a,i}(j)|x_{i}(j)) p_{j}(y_i(j)|x_{a,i}(j)) \]

(19)

where \( q_{j,s} \) is defined in (3).

Fix \( p_X(x) \) and generate \( 2^{nR} \) i.i.d sequences \( x^n[m] \) each with probability \( \prod_{i=1}^{n} p_X(x_i) \) where \( m \in [1: 2^{nR}] \). To transmit \( m \), Tx sends \( x^n[m] \). Rx after receiving \( y^n \), looks for a unique index \( \hat{m} \) such that satisfies:

\[ (x^n[\hat{m}], y^n) \in A^n(\mathbf{X}, \mathbf{Y}_s). \]

Due to the symmetry of the random codebook generation, the probability of error is independent of the specific messages. Hence, to analyze the probability of error, without loss of generality, we assume that \( m = 1 \) is encoded and transmitted. The error events at Rx are:

\[ E_1 = \{ \forall s \in \{0,1\}^n, w_H(s) \leq n_a : (x^n[1], y^n) \notin A^n(\mathbf{X}, \mathbf{Y}_s) \} \]

\[ E_2 = \{ \exists s \in \{0,1\}^n, w_H(s) \leq n_a \text{ such that } (x^n[m], y^n) \in A^n(\mathbf{X}, \mathbf{Y}_s) \text{ for some } m \neq 1 \} \]

Due to the Asymptotic Equipartition Property (AEP) [11], \( Pr(E_1) \to 0 \) as \( n \to \infty \). Now, to consider the probability of \( E_2 \), let \( \forall s \in \{0,1\}^{n_r} \):

\[ E_2^{(2)} = \{ \exists m \neq 1 \text{ such that } (x^n[m], y^n) \in A^n(\mathbf{X}, \mathbf{Y}_s) \} \]

(20)

with probability:

\[ Pr(E_2^{(2)}) \leq \sum_{m \neq 1} \prod_{j=1}^{n_r} \sum_{q \in \mathcal{P}^{n_r}_{j,s}(X \times Y)} Pr((x^n(j)[m], y^n(j)) \in T^n_{j,s}(q)) \]

\[ \leq \sum_{m \neq 1} \prod_{j=1}^{n_r} |\mathcal{P}^{n_r}_{j,s}(X \times Y)| \sup_{q \in \mathcal{P}^{n_r}_{j,s}(X \times Y)} Pr((x^n(j)[m], y^n(j)) \in T^n_{j,s}(q)) \]

\[ \leq \sum_{m \neq 1} \prod_{j=1}^{n_r} |\mathcal{P}^{n_r}_{j,s}(X \times Y)| \sup_{p_{j}(x_{s,i}(j)|x_{i}(j), s(j)) \in E_{p_j} \: d(x_{s,i}(j),x_{i}(j)) \leq D_{j,s}=s-D_{j}} 2^{-n(I(X;j);Y_s(j))-\delta(\epsilon)} \]

\[ \leq \sum_{m \neq 1} \prod_{j=1}^{n_r} \left( \begin{array}{c} |X| \\ n \end{array} \right) \left( \begin{array}{c} |Y| \\ n \end{array} \right) \left( \begin{array}{c} |X_a| \\ n \end{array} \right) \sup_{p_{j}(x_{s,i}(j)|x_{i}(j), s(j)) \in E_{p_j} \: d(x_{s,i}(j),x_{i}(j)) \leq D_{j,s}=s-D_{j}} 2^{-n(I(X;j);Y_s(j))-\delta(\epsilon)} \]

\[ \leq \sum_{m \neq 1} \prod_{j=1}^{n_r} \frac{n^{|X|+|Y|+|X_a|}}{|X| - 1|(|Y| - 1)|(|X_a| - 1)!} \sup_{p_{j}(x_{s,i}(j)|x_{i}(j), s(j)) \in E_{p_j} \: d(x_{s,i}(j),x_{i}(j)) \leq D_{j,s}=s-D_{j}} 2^{-n(I(X;j);Y_s(j))-\delta(\epsilon)} \]
(d) \leq n^{|X|+|Y|+|X_0|}2^{nR}2^{-n(\Theta_s-\delta(\epsilon))} \tag{21}

where (a) follows from (19) and \(x^n(j)[m]\) shows the \(j\)th element of vector \(x^n[m]\), (b) follows from joint typicality lemma and the memoryless property of the channel \(X \rightarrow Y\) according to (19), (c) follows from [12, Lemma II.1], (17), (18), where \(\binom{n}{m} = \binom{n+k-1}{k-1}\) is the multiset number, (d) follows from the independence of disjoint paths, where we define

\[
\Theta_s = \inf \frac{n_r}{n_a} \max_{w_H(s) \leq n_a} \sum_{j=1}^{n_r} I(X(j); Y_s(j))
\]

\[
\forall j \in \{1, \ldots, n_r\}, : E_{s_j} |d(x_s(j), x_s(j))| \leq D_j = s \cdot D_j
\]

Therefore:

\[
P_r(E^c_2) \leq \left( \frac{n_r}{n_a} \right) \max_{w_H(s) \leq n_a} P_r(E^c_{2,s})
\]

\[
\leq \left( \frac{n_r}{n_a} \right) \max_{w_H(s) \leq n_a} n^{|X|+|Y|+|X_0|}2^{n(R-\Theta_s+\delta(\epsilon))}
\]

\[
= \left( \frac{n_r}{n_a} \right) n^{|X|+|Y|+|X_0|}2^{n(R-\min_{w_H(s) \leq n_a} \Theta_s+\delta(\epsilon))}
\]

Hence, considering the finite alphabets, if \(R \leq \min_{s \in \{0,1\}^n} \Theta_s - \delta(\epsilon), P_r(E_2)\) goes to zero as \(n \to \infty\). This completes the achievability proof.

Converse: The converse easily follows from the Fano’s inequality by noting that for every \(s \in \{0,1\}^{n_r} : w_H(s) \leq n_a\) and every \(h^n\) defined in Definition 1, we must have \(H(M|Y^n_s) \leq n\epsilon_n \) for some \(\epsilon_n \xrightarrow{n \to \infty} 0\).

Proof of Theorem 2: The proof is similar to the proof of Theorem 1. Hence, we only describe the differences in the achievability part. Here, Tx generates \(|S| = \binom{n_r}{n_a}\) codebooks \(C_s\) similar to the one in Theorem 1 (with fixed \(p_X(x)\) for each codebook). To transmit \(m\), knowing the current state of the channel \(s_i = s\) for \(i \in \{1, \ldots, n\}\), selects \(C_s\) and transmits \(x^n[m]\) from that codebook. The rest of the proof is similar to Theorem 1.

Proof of Theorem 3: We apply a random coding technique on top of a random linear code (Varshamov construction [9]). To make this combination possible, we propose a new coding scheme by using proper auxiliary codewords.

First consider the replacement attacks as defined in Section III-B and let \(d_j = s \cdot 2D_j, j \in [1 : n_r]\). Now, fix \(p_X(x)\) and generate \(2^{nR}\) i.i.d sequences \(u^k[m]\) each with probability \(\prod_{i=1}^{n} p_X(u_i)\) for some \(k \geq nR \log |X| 2\) where \(m \in [1 : 2^{nR}]\). Repeat the following codebook generation process \(n_r\) times (for \(j \in \{1, \ldots, n_r\}\)) to produce \(x^n:\)
Choose a random $|\mathcal{X}|$-ary matrix $G_j \in \mathbb{F}_{|\mathcal{X}|}^{k \times n}$ where its elements are uniformly and independently chosen from $\mathbb{F}_{|\mathcal{X}|}$. Also, let $\mathcal{V}$ be the set of all $n$-length sequences in $\mathcal{X}^n$. Now, use the matrix $G$ as a generator matrix to generate $2^{nR}$ sequences $x^n(j)[m] = u^k(j)[m]G_j$ (Varshamov construction) with minimum distance $d_j$. Therefore, this code satisfies the Gilbert-Varshamov bound: for every $|\mathcal{X}| \geq 2$ and real $0 \leq d_j \leq 1 - \frac{1}{|\mathcal{X}|}$, the volume of the Hamming ball centered at $x^n(j)[m]$ (for $\forall m \in [1 : 2^{nR}]$, $j \in [1 : n_r]$) is bounded as:

$$|\mathcal{X}|^{nH_{\mathcal{X}}(d_j)} \leq Vol_{\mathcal{X}}(x^n(j)[m], n \times d_j) \leq |\mathcal{X}|^{nH_{\mathcal{X}}(d_j)}$$  \hspace{1cm} (22)

Then, pick the sequences in these hamming balls and call them codewords $v^n(j)[m, l]$: $v^n(j)[m, l] \in B_{\mathcal{X}}(x^n(j)[m], n \times d_j)$, where $l$ shows the codeword’s index in the ball, $l \in [1 : Vol_{\mathcal{X}}(x^n(j)[m], n \times d_j)]$. Let $L_j = \frac{1}{n \log |\mathcal{X}|} \max_m \log_{|\mathcal{X}|} Vol_{\mathcal{X}}(x^n(j)[m], n \times d_j)$. This means that the $v^n(j)$ is selected according to $p(v|x) : E_{p_j}(d(v(j), x(j)))] \leq d_j$ for $j \in \{1, \ldots, n_r\}$.

To transmit $m$, Tx sends $x^n[m]$. Rx after receiving $y^n$, looks for a unique index $\tilde{m}$ and some $\tilde{l}$ such that:

$$(v^n[\tilde{m}, \tilde{l}], y^n) \in A^n_{v}(X_a, Y).$$

Due to the symmetry of the random codebook generation, the probability of error is independent of the specific messages. Hence, to analyze the probability of error, without loss of generality, we assume that $m = 1$ is encoded and transmitted. Note that although the foreseeer adversaries’ channel inputs are chosen with memory, the channel from adversaries to Rx is i.i.d, i.e.,

$$p(y^n|x^n) = \prod_{j=1}^{n_r} \prod_{i=1}^{n} p_{y_j}(y_{i(j)}|x_{a_i}(j))$$  \hspace{1cm} (23)

Due to (3), we have $x^n(j) \in B_{\mathcal{X}}(x^n(j)[1], n \times d_j)$ for $j \in [1 : n_r]$. Thus, the error events at Rx are:

$${\mathcal{E}} = \{s \in \{0, 1\}^{n_r}, w_H(s) \leq n_a, (v^n[1, l'], y^n) \in A^n_{v}(X_a, Y)\}$$

$${\mathcal{E}}_2 = \{s \in \{0, 1\}^{n_r}, w_H(s) \leq n_a, \text{ such that } (v^n[m, l'], y^n) \in A^n_{v}(X_a, Y) \text{ for some } m \neq 1 \text{ and some } l'\}$$

Due to the problem definition, we are sure that $x^n(j) \in B_{\mathcal{X}}(x^n(j)[m], n \times d_j)$. Since $v^n(j)[m, l]$ covers all the codewords in this ball, $x^n(j) = v^n(j)[m, l']$ for some $l'$. Therefore, due to the AEP [11], $Pr(\mathcal{E}_1) \rightarrow 0$ as $n \rightarrow \infty$. Now, to consider the probability of $\mathcal{E}_2$, define:

$${\mathcal{E}}_2' = \{\exists m \neq 1 \text{ such that } (v^n[m, l'], y^n) \in A^n_{v}(X_a, Y) \text{ for some } l'\}$$

for $\forall s \in \{0, 1\}^{n_r}$. Considering (23), the joint AEP [11] implies:

$$Pr({\mathcal{E}}_2') \leq 2^{n(R + \sum_{j=1}^{n_r} L_j)} - 2^{-n(H(Y) + H(V) - H(X_a, Y) - \epsilon)}$$
Therefore, if \( R + \sum_{j=1}^{n_r} L_j \leq H(V) - H(X_a(Y)) - \delta(\epsilon) \), \( Pr(\mathcal{E}_{2,a}^\prime) \) goes to zero as \( n \to \infty \). Using (22) and the disjoint path property, we have:

\[
R \leq \sum_{j=1}^{n_r} \left[ H(V(j)) - H(X_a(j)|Y(j)) - \frac{H[X](d_j)}{\log|X|^2} \right] - \delta(\epsilon)
\]  

(24)

for all \( h^n \). Thus, \( Pr(\mathcal{E}_2) \leq \binom{n_r}{n_a} \max_{w_H(s) \leq n_a} Pr(\mathcal{E}_{2,s}^\prime) \) goes to zero if (24) holds for all \( s \in \{0, 1\}^{n_r} \): \( w_H(s) \leq n_a \) which results in \( R \leq \min_{s \in \{0, 1\}^{n_r}} \inf_{h^n} \sum_{j=1}^{n_r} [H(V(j)) - H(X_a(j)|Y(j)) - \frac{H[X](d_j)}{\log|X|^2}] - \delta(\epsilon) \). The proof for erasing attacks is similar by defining \( d_j = s \cdot D_j \) and noting that a code with minimum distance \( d_j \) can recover from \( d_j \) erasures. This completes the proof.

**Proof of Theorem 4:** The proof is straightforward considering the proofs of Theorems 2 and 3.

**Proof of Theorem 5:**

*No CSI:* We use the asymptotic Hamming bound (i.e., sphere packing bound) to limit the rate of a code that wishes to correct \( D_j \) errors. Similar to the proof of Theorem 3, first consider the replacement attacks with \( d_j = s \cdot 2D_j, j \in [1 : n_r] \). Since \( M \to X(j) \to X_a(j) \to Y(j) \) forms a Markov chain for \( j \in \{1, \ldots, n_r\} \), using Fano’s inequality, for every \( h^n, h^n(j) : \mathcal{X}^n \times \{0, 1\} \to \mathcal{X}^n_a \) satisfying (3), we have:

\[
H(X_a^n|Y^n) \leq H(M|Y^n) \leq n\epsilon_n
\]

where \( \epsilon_n \to 0 \) as \( n \to \infty \). Hence

\[
\sum_{j=1}^{n_r} H(X_a^n(j)) \stackrel{(a)}{=} H(X_a^n) \leq I(X_a^n; Y^n) + n\epsilon_n
\]

\[
\leq n \sum_{j=1}^{n_r} I(X_a(j); Y(j)) + n\epsilon_n
\]

(25)

for all \( h^n \), where (a) and (b) follow from (2).

To consider all possible \( h^n \), we must consider the type class \( \mathcal{T}^n(\mathcal{P}^n_{j,s}(\mathcal{X} \times \mathcal{Y})) \). Recall that one must be able to correct any possible \( \frac{d_j}{2} = s \cdot D_j \) errors made by the adversary on \( j \)th route. This means that the minimum distance of the codewords must be greater than \( d_j \). Otherwise, the adversary intentionally always chooses the closer codeword which cannot be distinguished at the Rx. Thus, the rate of the code for every \( s \in \{0, 1\}^{n_r} : w_H(s) \leq n_a \) and every \( h^n \) defined in Definition 1 must satisfy the Hamming
Proof of Theorem 2. This completes the proof. Therefore, (9) is proved. The proof for the case of CSI at Tx follows a similar lines by considering the limits (i.e., $D_j$ where we defined $D_j$).

Proof of Corollary 4: Let $P_j = Pr(X(j) = 1)$ and without loss of generality assume that $P_j \leq \frac{1}{2}$. Also, let $Pr(X_a(j) = e) = N'_j$, considering the distortion measure defined above with finite distortion limits (i.e., $D_j$s) and adversaries’ model in Definition 1, we have $N'_j = Pr(X_a(j) \neq X(j)) \leq s(j) \cdot D_j$. Thus,

$$H(Y_s(j)) = H((1 - P_j)(1 - N'_j)(1 - N_j), P_j(1 - N'_j)(1 - N_j), 1 - (1 - N'_j)(1 - N_j))$$

$$= H((1 - N'_j)(1 - N_j)) + (1 - N'_j)(1 - N_j)H(P_j)$$

$$H(Y_s(j)|X(j)) = H((1 - N'_j)(1 - N_j))$$

Therefore, we have: $I(X(j); Y_s(j)) = (1 - N'_j)(1 - N_j)H(P_j)$. Now, we can obtain (7), as:

$$C_{n,c,i,d}^{\text{nc}} = \sup_{0 \leq D_j \leq P_j \leq \frac{1}{2}} \min_{s \in \{0,1\}^{nr}} \min_{w_H(s) \leq n_a} \sum_{j=1}^{n_r} \min_{N'_j \leq s(j) \cdot D_j} (1 - N'_j)(1 - N_j)H(P_j)$$

which will be maximized for $P_j = \frac{1}{2}$ independent of $s(j)$ for $j \in \{1, \ldots, n_r\}$. Hence, the rest of the proof is straightforward.

Proof of Corollary 5: Let $V_i(j) = X_i(j)$ for $j \in \{1, \ldots, n_r\}$, $P_j = Pr(X(j) = 1)$ and $N'_j = s(j) \cdot D_j$. Recall that for all $h^n$ (defined in Definition 1 and satisfies (3)), we have $Pr(X(j) \neq X_a(j)) = Pr(X_a(j) = e) \leq N'_j$. After some calculations, one can compute:

$$H(V(j)) = H(P_j)$$

$$H(X_a(j)|Y(j)) \leq \tilde{N}_j H(N'_j/N'_j) + N_j(1 - N'_j)H(P_j)$$

where we defined $\tilde{N}_j = N_j(1 - N'_j) + N'_j$. The rest of the proof is straightforward.
Proof of Corollary 6: Similar to the proof of Corollary 3, it is enough to choose the proper joint distribution for adversaries’ input. Hence, let $X_{a,i}(j) = \text{BEC}(X_i(j), S(j) \cdot D_j)$. Computing the mutual information term in (9) completes the proof.

Proof of Corollary 7: The achievability follows by the standard arguments that extend the achievable rate to the Gaussian case with continuous alphabets [11]. Since we assume that the channel inputs are Gaussian, let $X_a(j) \sim \mathcal{N}(0, P_{a,j})$ where $E[d(X^n_a(j), X^n(j))] \leq s(j) \cdot D_j$. To find the infimum of $\sum_{j=1}^n I(X(j); Y_s(j))$ in (7), first we find a lower bound and then we show it is achievable by the adversaries.

$$I(X(j); Y_s(j)) = H(Y_s(j)) - H(Y_s(j)|X(j)) \geq H(X_a(j) + Z(j)) - H(X_a(j) - X(j) + Z(j))$$

$$= \frac{1}{2} \log(2\pi e (P_{a,j} + N_j)) - H(X(j) - X_a(j) - Z(j))$$

$$\geq \frac{1}{2} \log(2\pi e (P_{a,j} + N_j)) - \frac{1}{2} \log(2\pi e (s(j) \cdot D_j + N_j))$$

where in (a) we use $E[d(X^n_a(j), X^n(j))] \leq s(j) \cdot D_j$. (26) should be minimized over all $P_{a,j}$ satisfying the distortion limit. Thus, $P_{a,j} = P_j - s(j) \cdot D_j$. This lower bound is achievable by adversaries if they choose a joint distribution with a backward Gaussian test channel $X_{i,j} = X_{a,i}(j) + Z'_i(j)$ where $Z'_i(j) \sim \mathcal{N}(0, s(j) \cdot D_j)$ and $X_{a,i}(j) \sim \mathcal{N}(0, P_j - s(j) \cdot D_j)$. The rest of the proof is straightforward.