Wireless Network Secrecy with Public Feedback

Etienne Perron, Suhas Diggavi, Emre Telatar
EPFL, Lausanne, Switzerland
Email: {etienne.perron,suhas.diggavi,emre.telatar}@epfl.ch

Abstract— **In this paper we consider secret communication between two special nodes ("source" and "destination") in a wireless network with authenticated relays: the message communicated to the destination is to be kept information-theoretically (unconditionally) secret from any eavesdropper within a class. A public feedback channel from the destination to all nodes may or may not be available. We focus on a one-relay network with a bottleneck link between the source and the relay to illustrate the ideas. For this particular network, we derive the rate-equivocation capacity region when the public feedback channel is absent, and an achievable secret key rate when public feedback is present.**

I. Introduction

A fundamental aspect of wireless communication is its broadcast nature, *i.e.*, transmissions from a node can be overheard (albeit through different channels) at several locations. This property makes wireless communication inherently vulnerable to eavesdropping by an adversary. As the use of wireless networks grows, this is an important concern to be addressed.

Clearly the secrecy of any system can be enhanced if one could have even a small amount of shared key between the source–destination pair, which can be kept unconditionally secret from an eavesdropper. In particular one of the motivations for the formulation in this paper is that we can potentially generate such a secret key using physical wireless channel properties.

In wireless communication, even though the signal from the source is broadcast, it is received at the destination and the potential eavesdropper through *different* (fading) channels. It is this distinction that is exploited in information-theoretic secrecy for broadcast channels in the seminal work of wiretap channels [1], [2]. However, not much is known about the cooperative secrecy setup, where there are relay nodes facilitating secure communication between a source and a destination[1]. The main difficulties in dealing with arbitrary relay networks are (i) the broadcast nature of wireless communications, (ii) the fact that signals from simultaneously transmitting nodes interfere with one another at other nodes. These give rise to complex signal interactions making the understanding of wireless networks difficult.

In [5], using the recent understanding of wireless network information flow in [6], [7], we studied the secrecy problem for a wireless network, where a source node is communicating with a destination node, with the help of (authenticated) relay nodes, when an unknown (passive) eavesdropper is also present in the network. We called this setup *cooperative secrecy*, since the (authenticated) nodes in the network cooperate to provide secrecy against potential eavesdroppers. The main result in [5] was an achievable trade-off between the reliable transmission rate from the source to the legitimate destination and the amount of information leaked to a class of eavesdroppers over an *arbitrary* wireless relay network, when the channels are either deterministic or Gaussian. Roughly speaking, the trade-off is related to the information-theoretic min-cuts[2] between the source-destination and source-eavesdropper pairs. In particular, one extreme point of the trade-off is that we can ensure perfect secrecy (zero rate of information leakage to the eavesdropper) for an information rate of (approximately) the "difference" between these two min-cut values. The other extreme point is to transmit information to the legitimate destination at its min-cut while leaking an information rate related to the "difference" between these two information theoretic min-cut values to the eavesdropper. The main results of [5] are summarized in Section III of this paper.

If for any possible transmission strategy (any given set of input distributions) the min-cut value between at least one source-eavesdropper pair is larger than the min-cut value between the source-destination pair, then the results of [5] do not guarantee any secrecy against the class of eavesdroppers. In particular, establishing a secret key shared by the source and the destination might not be possible. The rate of such a key is called the secret key rate. The result by Csiszár and Körner [2] for broadcast channels has a very similar limitation. In seminal work [9], Maurer discovered that for broadcast channels, an infinite-capacity and *public* feedback channel from the destination to the source and all the eavesdroppers can improve the secret key capacity. In particular, the secret key capacity can be positive for broadcast channels that have zero secret key capacity in the absence of the public feedback link. A more complete treatment of this class of problems can be found in [10].

The existence of feedback from the destination is a reasonable assumption for wireless networks, since wireless channels are generally bi-directional, and hence, a backward transmission from the destination is easy to implement. This fact, together with the encouraging results by Maurer and

E. Perron was supported in part by SNSF Grant # 200021-105640/1.

[1] Notable exceptions are some recent studies of techniques when there is a single relay node present, as an extension of the classical relay channel to the secrecy problem [3], [4].

[2] The information-theoretic min-cut is related to the cooperative information transfer rate [8], [7].

others, motivates the study of secrecy protocols with feedback for wireless networks. Towards this aim, we simplify the problem by assuming that the backward transmission is over a perfectly reliable public channel, as in [9] and [10]. One would expect that feedback which is not public, *i.e.*, which yields different received signals at the source and at the eavesdroppers, can only improve the situation compared to public feedback. On the other hand, the assumption that the public feedback channel is of infinite capacity is quite strong, and needs to be refined in future work. However, we believe that assuming a public, infinite-capacity feedback channel is an interesting first step, and as seen in this paper, yields valuable intuitions about achievable secret key rates in wireless networks.

The main new difficulty in proving the existence of feedback-utilizing protocols for wireless networks stems from communication via relays. The protocols in [9] and [10] use the fact that when an i.i.d. sequence of symbols is transmitted over a memoryless broadcast channel, then this sequence, together with the received sequences, has the same statistics as the output of a memoryless multi-source. When the transmission happens over a relay network, then the relays need to map from their receive signals to their transmit signals. Depending on the signal rate, the whole typical transmit set may not be used. This leads to a thinning of the transmit signal spaces at the relays. As a consequence, after a forward transmission over the network, the information available at the source, the destination and the eavesdropper does *not* have the statistics of a memoryless multisource. Therefore the techniques introduced in [9], [10] are not applicable. This difficulty is not related to the interference of received signals, but stems only from the thinning at the relays.

For this reason, we restrict our attention to a simple one-relay network with feedback without any phenomena of signal interference. We show the existence of protocols that use the public backward channel to achieve a certain secret key rate for this network. Interaction between the nodes is over general discrete memoryless channels. For comparison, we derive the secret key capacity for the case when no backward channel is available. We give a simple sufficient condition under which the presence of a public backward channel improves the achievable secret key rate. These are the main contributions of this paper.

The paper is organized as follows. We give the precise problem statement in Section II. The results from [5] are stated in Section III. We give the main results in Section IV and the corresponding proof outlines in Section V. A list-size computation crucial to the paper is presented in the appendix. We conclude in Section VI with a discussion about possible extensions and open questions raised by this work.

II. Problem Statement

We consider transmission over a wireless relay network $\mathcal{G} = (\mathcal{V}, \mathcal{L})$, where $\mathcal{V}$ is the set of vertices representing the communication nodes in the relay network and $\mathcal{L}$ is the set of pairs of nodes which describes the signal interactions. Note that the channels are not point-to-point links, rather, they model how the transmitted signals are superimposed and received at the receiving nodes (*i.e.*, there can be broadcasting and interference). We consider a special node $S \in \mathcal{V}$ (the source) which wants to secretly communicate with another special node $D \in \mathcal{V}$ (the destination) with the help of a set of (authenticated) relay nodes $\mathcal{A} \subset \mathcal{V}$ in the network. The secrecy is with respect to a set of possible (passive) eavesdropper nodes $\mathcal{E} \subset \mathcal{V}$ where $\mathcal{E}$ is disjoint from $\mathcal{A} \cup \{S, D\}$. The aim is that S and D agree on a message (or a key), which is to be kept secret from any one of the possible eavesdropper nodes $E \in \mathcal{E}$, that listen to the wireless transmissions in the relay network[3]. When no feedback channel exists in the network, then the message needs to be generated at the source S and is destined for D. When a feedback channel is available, the names "source" and "destination" are less meaningful, but for consistency, we also use them in that case.

The interaction between nodes is defined by a collection of alphabets $\{\mathcal{X}_j, \mathcal{Y}_j\}_{j\in\mathcal{V}}$ and a collection of channels

$$p_{Y_j|\{X_i\}_{i\in\mathcal{N}_j}}(y_j[t]|\{x_i[t]\}_{i\in\mathcal{N}_j}), \tag{1}$$

where $y_j[t] \in \mathcal{Y}_j$ is the symbol received at node j during time t, $x_i[t] \in \mathcal{X}_i$ is the symbol transmitted at node i during time t, and $\mathcal{N}_j$ is the set of nodes that interact with node j. For a fixed set $\{x_i[t]\}_{i\in\mathcal{N}_j}$, the received symbol $y_j[t]$ is drawn from the distribution (1) conditionally independently of all other received or transmitted symbols.

When a public feedback channel is present, we assume that at any time, D can designate an arbitrary integer I to be transmitted over the public feedback channel. I is then immediately and reliably known to S and to all eavesdropping nodes in $\mathcal{E}$.

Given a relay network with the signal interaction model given in (1), we consider two different problems:

Secret Communication: In secret communication, we want to ensure that we can communicate reliably *and* secretly between S and D. The notion of reliability is the standard information-theoretic notion that a message of rate R_S chosen at S can be decoded with arbitrarily small probability of error at D. More formally, we have the following definitions.

Definition 1: A (T, ϵ)-code for a wireless network without feedback is given by a (possibly) probabilistic source encoding function f_S mapping W to $\mathbf{X}_S$, a set of (possibly) probabilistic relay encoding functions f_i, mapping $\mathbf{Y}_i$ to $\mathbf{X}_i$ at each relay $i \in \mathcal{A}$, and a deterministic decoding function f_D, mapping $\mathbf{Y}_D$ into an estimate $\hat{W}$ of the message. Here, $\mathbf{X}_i = (x_i[1], \dots, x_i[T])$ and $\mathbf{Y}_i$ are blocks of T symbols (each symbol being a member of $\mathcal{X}_i$ or $\mathcal{Y}_i$, respectively), and W is the message, which is uniformly distributed in the set $\{1, \dots, 2^{TR}\}$. The quantity R is the rate of the code.

[3] Our results apply also when *all* eavesdroppers in $\mathcal{E}$ are present, but cannot collude since they are not allowed to transmit anything. In this paper we restrict ourselves to the passive eavesdropper model, motivated by secure authentication protocols [11] which could potentially discover an active eavesdropper.

The probability of error of a (T,ϵ)-code is required to be bounded by ϵ: $\mathbf{P}(W \neq \hat{W}) < \epsilon$.

The notion of information-theoretic secrecy is defined through the *equivocation* rate R_e, which is the residual uncertainty about the message when the observation of the strongest eavesdropper is given. More formally, [1], [2]:

Definition 2: Given a (T,ϵ)-code, the equivocation rate is

$$\frac{1}{T}\min_{E\in\mathcal{E}} H(W|\mathbf{Y}_E), \tag{2}$$

where W is the uniformly distributed source message, $\mathbf{Y}_E$ is the sequence of observations at eavesdropper E and $H(\cdot|\cdot)$ denotes the (conditional) entropy [8].

In this paper, we also refer to the equivocation rate as simply the "equivocation".

Definition 3: A rate-equivocation pair (R, R_e) is called "achievable" if for any $\epsilon > 0$, there exists a blocklength T and a (T,ϵ)-code of rate R and equivocation R_e.

By "perfect secrecy", we mean a situation where a rate-equivocation pair (R, R_e) such that $R = R_e$ is achievable. We then say that a perfect secrecy rate R is achievable.

Secret Key Agreement: In secret key agreement, we want to ensure that S and D agree with high probability on the same key (message) K, while keeping K secret from the eavesdroppers. For a protocol that achieves this, the rate of the key K is called the secret key rate and denoted by R_K. This can be formalized as follows.

Definition 4: For a network without feedback, we say that a secret key rate R_K is achievable if the rate-equivocation pair (R_K, R_K) is achievable for secret communication.

In other words, the secret key rate and the perfect secrecy rate are equivalent for a wireless network without feedback. In the presence of a public feedback channel, we use a different definition:

Definition 5: A (T,ϵ)-protocol for key generation over a wireless network with feedback is given by a distribution $p_{\mathbf{X}_S}$ for the transmitted sequence $\mathbf{X}_S$ of length T, a set of (possibly) probabilistic relay encoding functions f_i, mapping $\mathbf{Y}_i$ to $\mathbf{X}_i$ at each relay $i \in \mathcal{A}$, a (possibly) probabilistic feedback function f_D, mapping $\mathbf{Y}_D$ to a public message I, and two key generation functions g_S (mapping $(\mathbf{X}_S, I)$ to $\hat{K}$) and g_D (mapping $(\mathbf{Y}_D, I)$ to K). The key K should be uniformly distributed in $\{1,\ldots,2^{TR_K}\}$, where R_K is the secret key rate. The probability of disagreement of a (T,ϵ)-protocol is required to be bounded by ϵ: $\mathbf{P}(K \neq \hat{K}) < \epsilon$.

Definition 6: Given a (T,ϵ)-protocol for key generation, the key equivocation rate is

$$\frac{1}{T}\min_{E\in\mathcal{E}} H(K|\mathbf{Y}_E, I), \tag{3}$$

where K is the key, $\mathbf{Y}_E$ is the sequence of observations at eavesdropper E, and I is the public feedback.

Definition 7: For a network with feedback, we say that a secret key rate R_K is achievable if for any $\epsilon > 0$, there exists an integer T and a secret key generation (T,ϵ)-protocol that has key rate *and* equivocation rate R_K.

III. Secrecy Rates for General Networks without Feedback

The results in this section have already been presented in [5] and we summarize them here for completeness. We consider the following two interaction models.

Wireless interaction model: In this well-accepted model [12], transmitted signals get attenuated by (complex) gains to which independent (Gaussian) receiver noise is added. More formally, the received signal y_j at node $j \in \mathcal{V}$ at time t is given by,

$$y_j[t] = \sum_{i\in\mathcal{N}_j} h_{ij}x_i[t] + z_j[t], \tag{4}$$

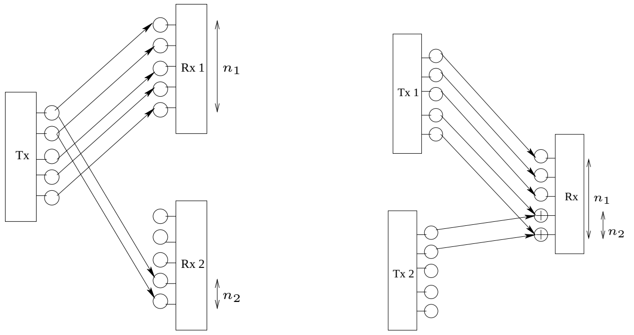
where h_{ij} is the complex channel gain between node i and j, x_i is the signal transmitted by node i, and $\mathcal{N}_j$ are the sets of nodes that have non-zero channel gains to j. We assume that the average transmit power constraint for all nodes is 1 and the additive receiver Gaussian noise is of unit variance. We use the terminology *Gaussian wireless network* when the signal interaction model is governed by (4).

Deterministic interaction model: In [13], a simpler deterministic model which captures the essence of wireless interaction was developed. The advantage of this model is its simplicity, which gives insight to strategies for the noisy wireless network model in (4). The results in this section are developed for both this deterministic model as well as the model in (4). The deterministic model of [13] simplifies the wireless model (4) by eliminating the noise and discretizing the channel gains through a binary expansion of q bits. Therefore, the received signal $\mathbf{y}_j^{(d)}$ which is a binary vector of size q is modeled as

$$\mathbf{y}_j^{(d)}[t] = \sum_{i\in\mathcal{N}_j} \mathbf{G}_{ij}\mathbf{x}_i^{(d)}[t], \tag{5}$$

where $\mathbf{G}_{ij}$ is a $q \times q$ binary matrix representing the (discretized) channel transformation between nodes i and j and $\mathbf{x}_i^{(d)}$ is the (discretized) transmitted signal. All operations in (5) are done over the binary field. We use the terminology *deterministic wireless network* when the signal interaction model is governed by (5).

An illustration of this deterministic model is given in Figure 1 for the broadcast and multiple access networks. Figure 1(a) shows a deterministic model of the broadcast channel, where the channel from the transmitter to Receiver 1 is stronger than that to Receiver 2. This is represented by the deterministic model developed in [13] with the 5 most significant bits (MSB) of the transmitted signal captured by Rx 1 and only the 2 MSB of the transmitted signal captured by Rx 2. The deterministic model of the multiple access channel shown in Figure 1(b) adds one more ingredient, which is how the bits from two transmitting nodes interact at a receiver. In Figure 1(b) the channel from Tx 1 to Rx is stronger than that of Tx 2. Therefore, the interaction is between the 2 MSB of Tx 2 with the lower significant bits of Tx 1, and the interaction is modeled with an addition over the binary field (*i.e.*, `xor`). This interaction captures the dynamic range of the signal interactions. It was shown in [13], that this model *approximately*[4] captures the wireless interaction model of (4) for the broadcast and multiple access channels. For general networks the deterministic model yields insights which, when translated to the noisy wireless network, lead one to develop cooperative strategies for the model in (4), which are (provably) approximately[5] optimal [7].



(a) Broadcast channel

(b) Multiple access channel

Fig. 1. The linear deterministic model for a Gaussian broadcast channel (BC) is shown in (a) and for a Gaussian multiple access channel (MAC) is shown in (b).

Before continuing, we introduce the notion of an information-theoretic min-cut in a network.

Definition 8: We denote by Λ_i the set of all cutsets in the network that separate the source S from node i:

$$\Lambda_i = \{\Omega \subset \mathcal{V} : S \in \Omega, i \in \Omega^c\}.$$

For a subset $\Omega \subset \mathcal{V}$, we denote by X_Ω the tuple of all X_i for $i \in \Omega$.

Definition 9: The quantity $\min_{\Omega \in \Lambda_i} I(X_\Omega; Y_{\Omega^c} | X_{\Omega^c})$ is called the min-cut between the pair (S, i), where $I(\cdot;\cdot|\cdot)$ is the (conditional) mutual information.

Now, we state the two main results of [5].

Definition 10: Let a distribution $p(\{x_i\}_{i \in \mathcal{V}})$ over all transmit alphabets be given. We define $\mathcal{R}(p)$ to be the set of all pairs (R, R_e) satisfying

$$\begin{aligned} R &< \min_{\Omega \in \Lambda_D} H(Y_{\Omega^c}|X_{\Omega^c}), \\ R_e &< \min\big\{R, \min_{\Omega \in \Lambda_D} H(Y_{\Omega^c}|X_{\Omega^c}) \\ &\quad - \max_{E \in \mathcal{E}} \min_{\Omega \in \Lambda_E} H(Y_{\Omega^c}|X_{\Omega^c})\big\}. \end{aligned}$$

Theorem 1: In an arbitrary network with deterministic signal interaction as given in (5), all rate-equivocation pairs in the region

$$\mathcal{R} = \cup_{\prod_{i \in \mathcal{V}} p(x_i)} \mathcal{R}(p)$$

are achievable, where the union is taken over all possible product-distributions on the transmit alphabets.

Theorem 1 guarantees the existence of codes for the region of rate-equivocation pairs $\mathcal{R}$. However, this is not a complete characterization in that, we do not have a matching converse stating that no strategy can achieve pairs (R, R_e) that lie outside this region. Also note that the result in Theorem 1 can be generalized to include arbitrary deterministic interaction functions, rather than the linear model given in (5), by using techniques similar to those in [6].

Corollary 1: In a network with deterministic signal interaction as given in (5), any secret key rate R_K satisfying

$$R_K \leq \max_{\prod_{i \in \mathcal{V}} p(x_i)} \left[\min_{\Omega \in \Lambda_D} H(Y_{\Omega^c}|X_{\Omega^c}) - \max_{E \in \mathcal{E}} \min_{\Omega \in \Lambda_E} H(Y_{\Omega^c}|X_{\Omega^c})\right]$$

is achievable.

Proof: From Theorem 1, whenever $(R, R_e) \in \mathcal{R}$, so is (R_e, R_e), and perfect secrecy at rate R_e is possible. Corollary 1 then directly follows from Definition 4. ■

Definition 11: Let a distribution $p(\{x_i\}_{i \in \mathcal{V}})$ over all transmit alphabets be given. We define $\tilde{\mathcal{R}}(p)$ to be the set of all pairs (R, R_e) satisfying

$$\begin{aligned} R_S &< \min_{\Omega \in \Lambda_D} I(X_\Omega; Y_{\Omega^c}|X_{\Omega^c}) - \gamma, \\ R_e &< \min\big\{R_S - \gamma, \min_{\Omega \in \Lambda_D} I(X_\Omega; Y_{\Omega^c}|X_{\Omega^c}) \\ &\quad - \max_{E \in \mathcal{E}} \min_{\Omega \in \Lambda_E} I(X_\Omega; Y_{\Omega^c}|X_{\Omega^c}) - \gamma\big\} \end{aligned}$$

where γ is a constant which depends on the topology of the network but *not* on the channel gains h_{ij} or the signal-to-noise ratio (SNR) of operation.

Theorem 2: In a Gaussian wireless network, all rate-equivocation pairs in the region

$$\tilde{\mathcal{R}} = \cup_{\prod_{i \in \mathcal{V}} p(x_i)} \tilde{\mathcal{R}}(p)$$

are achievable, where the union is taken over all possible product-distributions on the transmit alphabets.

Remark 1: The proofs of Theorems 1 and 2 can be found in [5]. The codes that achieve the rates given in the theorems use a random encoding function at S, but deterministic encoding functions at the relays. The writeup [5] contains examples that suggest that random encoding functions (noise insertion) at the relays would improve the achievable rate-equivocation region.

Remark 2: Note that since the gap γ is a constant and the rates given in Definition 11 can grow with the SNR, this gap can be made small with respect to the rates of operation. Therefore, even though there is a gap of γ bits in the bound on the equivocation in Theorem 2, it can be small with respect to the total rate, *i.e.*, we might leak only a small number of bits for Gaussian wireless networks. However, we also believe that this gap is a purely technical issue in the proof, and not a fundamental problem arising in Gaussian wireless networks.

IV. One-Relay Network

In this section, we present the main contribution of this paper, namely an achievable secret key rate for a specific network, shown in Figure 2. The one-relay network consists of a source, a destination, only one relay A and only one eavesdropping node E. The network is defined by alphabets

[4] The approximation is in the sense that the capacity region of the deterministic model is within 1 bit of the capacity region of the Gaussian counterparts.

[5] It has been shown that for single unicast there is an *approximate* max-flow, min-cut result where the difference is within a constant number of bits, which depends on the topology of the network, but not the values of the channel gains [7].

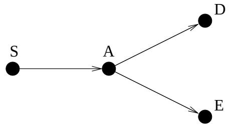


Fig. 2. The one-relay network.

$\mathcal{X}_S$, $\mathcal{Y}_A$, $\mathcal{X}_A$, $\mathcal{Y}_D$ and $\mathcal{Y}_E$, and channels $p_{Y_A|X_S}$, $p_{Y_D|X_A}$ and $p_{Y_E|X_A}$.

First, we consider the one-relay network without feedback.

Definition 12: Let $\mathcal{V}_A$ be the alphabet of an auxiliary random variable V_A, and let a distribution p_{X_S,V_A,X_A} over $\mathcal{X}_S \times \mathcal{V}_A \times \mathcal{X}_A$ be given. We define $\breve{\mathcal{R}}(\mathcal{V}_A, p)$ to be the set of all pairs (R, R_e) satisfying

$$\begin{aligned} R &< \min\{I(X_S;Y_A), I(V_A;Y_D)\}, \\ R_e &< [I(V_A;Y_D) - I(V_A;Y_E)]^+, \\ R_e &\leq R. \end{aligned}$$

Theorem 3: In the one-relay network *without feedback*, a rate-equivocation pair (R, R_e) is achievable if and only if it lies in the region

$$\breve{\mathcal{R}} = \cup_{\mathcal{V}_A, p_{X_S} p_{V_A} p_{X_A|V_A}} \breve{\mathcal{R}}(\mathcal{V}_A, p), \tag{6}$$

where the union is over all choices of the alphabet $\mathcal{V}_A$ and over distributions on $\mathcal{X}_S \times \mathcal{V}_A \times \mathcal{X}_A$ for which X_S and (V_A, X_A) are independent.

Using Caratheodory's theorem [14], one can show that it is sufficient to consider alphabets $\mathcal{V}_A$ no larger than $|\mathcal{Y}_D| + |\mathcal{Y}_E| + 1$ in the optimization (6). The proof of Theorem 3 can be found in Section V.

The following corollary follows directly from Theorem 3:

Corollary 2: The secret key capacity of the one-relay network without feedback is

$$\begin{aligned} C_K = \max_{\mathcal{V}_A, p_{X_S} p_{V_A} p_{X_A|V_A}} &\min\{I(X_S;Y_A), \\ &[I(V_A;Y_D) - I(V_A;Y_E)]^+\}. \end{aligned} \tag{7}$$

Note that Theorem 3 and Corollary 2 give necessary and sufficient conditions for secret communication and secret key agreement over the one-relay network without feedback. The following theorem gives sufficient conditions for secret key agreement over the one-relay network *with* feedback.

Definition 13: For a given distribution p_{X_A} and a number u, let the function $F(u, p_{X_A})$ be defined as

$$F(u, p_{X_A}) = \begin{cases} 0 & \text{if } u < I(X_A;Y_E) \\ u - I(X_A;Y_E) & \text{if } I(X_A;Y_E) < u < I(X_A;Y_D,Y_E) \\ I(X_A;Y_D|Y_E) & \text{if } I(X_A;Y_D,Y_E) < u. \end{cases}$$

It can be readily checked that for any fixed p_{X_A}, $F(u, p_{X_A})$ is a continuous function of u.

Theorem 4: In the one-relay network *with feedback*, a secret key rate R_K is achievable if it satisfies

$$R_K < \max_{p_{X_S} p_{X_A}} F(I(X_S;Y_A), p_{X_A}). \tag{8}$$

The proof of Theorem 4 can be found in Section V.

Remark: Fix a distribution $p_{X_S} p_{X_A}$. Note that Theorem 4 says that when $I(X_S;Y_A) > I(X_A;Y_D,Y_E)$, we can achieve a secret key rate of $I(X_A;Y_D|Y_E)$. Since the channel from X_A to (Y_D, Y_E) is such that $Y_D -\!\!\circ\!\!- X_A -\!\!\circ\!\!- Y_E$ is a Markov chain, we have $I(X_A;Y_D|Y_E) = I(X_A,Y_E;Y_D) - I(Y_D;Y_E) = I(X_A;Y_D) - I(Y_D;Y_E)$, which is the secret key rate given in [9] and [10] for a conditionally independent broadcast channel. Hence, we see that when the channel between S and A is strong enough compared to the channel between A and (D, E), then the secret key rate is limited by the latter channel.

The following proposition compares the secret key rates (8) and (7).

Proposition 1: Let X_S^* and X_A^* be such that $p_{X_S^*} p_{X_A^*}$ maximizes (8), let Y_A^*, Y_D^* and Y_E^* be the corresponding channel outputs, and let $R_K^* = F(I(X_S^*;Y_A^*), p_{X_A^*})$ be the largest secret key rate guaranteed by (8). Then, the secret key rate

$$R' \triangleq \min\{I(X_S^*;Y_A^*), [I(X_A^*;Y_D^*) - I(X_A^*;Y_E^*)]^+\} \tag{9}$$

is achievable without the use of the public feedback channel. In addition, if $I(X_S^*;Y_A^*) \leq I(X_A^*;Y_D^*)$, then

$$R' \geq R_K^*. \tag{10}$$

Proof: First, note that $(X_S, V_A, X_A) = (X_S^*, X_A^*, X_A^*)$ is a valid member of the maximization set in (7). Hence, when ignoring the feedback channel,

$$R' = \min\{I(X_S^*;Y_A^*), [I(X_A^*;Y_D^*) - I(X_A^*;Y_E^*)]^+\}$$

is achievable, because it is smaller than the secret key capacity given in (7). Assume that $I(X_S^*;Y_A^*) \leq I(X_A^*;Y_D^*)$. Then, it follows from (8) that

$$R_K^* = \max\{0, I(X_S^*;Y_A^*) - I(X_A^*;Y_E^*)\}.$$

Since $I(X_S^*;Y_A^*) \leq I(X_A^*;Y_D^*)$, we have $R' \geq I(X_S^*;Y_A^*) - I(X_A^*;Y_E^*)$. From the non-negativity of the mutual information, we also have $R' \geq 0$. Hence, $R' \geq R_K^*$. ■

The example in Figure 3 shows that the inequality in (10) can be strict.

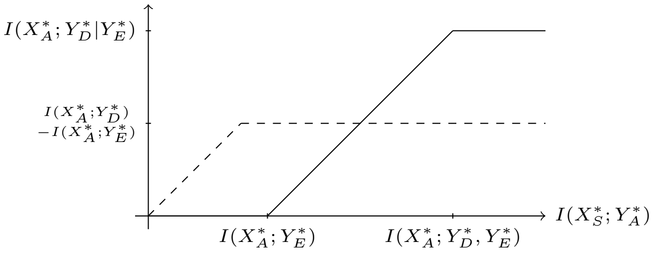


Fig. 3. Illustration of Proposition 1. R_K^* (solid line) and R' (dashed line) are plotted as functions of the mutual information between S and A.

V. Proof Outlines

A. Proof of Theorem 3

Achievability: Assume that $(R, R_e) \in \breve{\mathcal{R}}$. It follows that for a certain $\mathcal{V}_A$ and a distribution $p_{X_S} p_{V_A} p_{X_A|V_A}$, we have

$$R < I(X_S;Y_A), \tag{11}$$

$$R < I(V_A;Y_D), \tag{12}$$

$$R_e \leq R, \tag{13}$$

$$R_e < [I(V_A;Y_D) - I(V_A;Y_E)]^+. \tag{14}$$

From the channel coding theorem [14], we know that if (11) is true, then there exists a reliable code of rate R for the channel $p_{Y_A|X_S}$. Since $R_e \leq R$, and since E does not receive any information about this transmission, the equivocation rate at E about this message is at least R_e. Now, assume that A knows the transmitted message. From the achievability in [2], we know that as long as (12) and (14) are true, there exists a code that allows reliable transmission of a message of rate R from A to D while achieving an equivocation rate R_e.

Converse: Assume that (R, R_e) are achievable. Then, from the channel coding theorem [14], we know that $R < \max_{p_{X_S}} I(X_S; Y_A)$. Moreover, since the equivocation at E cannot be more than the number of messages, $R_e \leq R$. Hence,

$$(R, R_e) \in \cup_{p_{X_S}} \mathcal{A}(p_{X_S}), \tag{15}$$

where $\mathcal{A}(p_{X_S}) = \{(R, R_e) : R < I(X_S; Y_A), R_e \leq R\}$. From the converse proof in [2], we know that

$$(R, R_e) \in \cup_{\mathcal{V}_A, p_{V_A} p_{X_A|V_A}} \mathcal{B}(\mathcal{V}_A, p_{V_A} p_{X_A|V_A}), \tag{16}$$

where $\mathcal{B}(\mathcal{V}_A, p_{V_A} p_{X_A|V_A}) = \{(R, R_e) : R < I(V_A; Y_D), R_e < [I(V_A; Y_D) - I(V_A; Y_E)]^+\}$. From (15) and (16), it follows that

$$\begin{aligned}(R, R_e) \in &\left(\cup_{p_{X_S}} \mathcal{A}(p_{X_S})\right) \\ &\cap \left(\cup_{\mathcal{V}_A, p_{V_A} p_{X_A|V_A}} \mathcal{B}(\mathcal{V}_A, p_{V_A} p_{X_A|V_A})\right) \\ = &\cup_{p_{X_S}} \cup_{\mathcal{V}_A, p_{V_A} p_{X_A|V_A}} \Big(\mathcal{A}(p_{X_S}) \\ &\cap \mathcal{B}(\mathcal{V}_A, p_{V_A} p_{X_A|V_A})\Big) \\ = &\cup_{\mathcal{V}_A, p_{X_S} p_{V_A} p_{X_A|V_A}} \breve{\mathcal{R}}(\mathcal{V}_A, p) \\ = &\breve{\mathcal{R}}.\end{aligned}$$

B. Proof of Theorem 4

Let $p_{X_S} p_{X_A}$ be fixed. Consider the following protocol: S generates a random message of rate just below $I(X_S; Y_A)$ and sends it over the channel $p_{Y_A|X_S}$ using a good channel code. Then, it follows from the channel coding theorem [8] that A can decode that message with small error probability. Before the next transmission block, A re-encodes the message into a transmit sequence $\mathbf{x}_A$ using a codebook $\mathcal{C} \subseteq T(X_A)$ of rate $\min\{I(X_S; Y_A), H(X_A)\}$, where $T(X_A)$ denotes the set of all typical sequences $\mathbf{x}_A$ [14]. If $I(X_S; Y_A) < H(X_A)$, then the codebook $\mathcal{C}$ is generated uniformly at random and fixed for all time. If $H(X_A) \leq I(X_S; Y_A)$, then $\mathcal{C} = T(X_A)$.

Definition 14: Define $\mathcal{L}(Y_j)$, $j \in \{D, E\}$ as the list of all sequences $\mathbf{y}_j$ such that there is at least one codeword $\mathbf{x}_A \in \mathcal{C}$ for which $(\mathbf{x}_A, \mathbf{y}_j) \in T(X_A, Y_j)$.
Similarly, define $\mathcal{L}(Y_D|\mathbf{y}_E)$ as the list of all sequences $\mathbf{y}_D$ such that there is at least one $\mathbf{x}_A \in \mathcal{C}$ for which $(\mathbf{x}_A, \mathbf{y}_D, \mathbf{y}_E) \in T(X_A, Y_D, Y_E)$.

D and E receive sequences $\mathbf{y}_D$ and $\mathbf{y}_E$, respectively, that are the outputs of the broadcast channel $p_{Y_D|X_A} p_{Y_E|X_A}$ when $\mathbf{x}_A$ is the input.

For the backward transmission, we are facing a source coding problem similar to the one in [9] and [10]. The destination D wants to describe $\mathbf{y}_D$ to S, while keeping part of it secret from E. S and E observe the sequences $\mathbf{x}_S$ and $\mathbf{y}_E$, respectively, which serve as side-information in the source coding problem.

Definition 15: Define α_S and α_E to be non-negative real numbers such that: If $\mathbf{x}_S$ is known, then with high probability, $\mathbf{y}_D$ is one of $2^{T\alpha_S}$ possible sequences. If $\mathbf{y}_E$ is known, then with high probability, $\mathbf{y}_D$ is one of $2^{T\alpha_E}$ possible sequences.
The exponents α_S and α_E play an important role in deriving and analyzing a source code for the backward transmission. In [9] and [10], the computation of these exponents is straightforward, because the receive sequence at D and the two side-information sequences at S and at E together have the same statistics as the output of a discrete memoryless multisource. For the one-relay network, the derivation of α_S and α_E is more involved.

The decoder at A has a small error probability. Hence, with high probability, S knows which message was decoded and re-encoded at A, and we obtain

$$2^{T\alpha_S} \doteq |T(Y_D|\mathbf{x}_A)| = 2^{TH(Y_D|X_A)}. \tag{17}$$

Note that the sequences $\mathbf{x}_A$, $\mathbf{y}_D$ and $\mathbf{y}_E$ are highly likely to be jointly typical and thus, with high probability, $\mathbf{y}_E$ lies in $\mathcal{L}(Y_E)$ and

$$2^{T\alpha_E} = |\mathcal{L}(Y_D|\mathbf{y}_E)|.$$

Using Lemma 2 in the appendix and taking $\tilde{R} = I(X_S; Y_A)$, we obtain

$$\alpha_E = \begin{cases} H(Y_D|X_A) & \text{if } 0 < I(X_S; Y_A) < I(X_A; Y_E) \\ \left.\begin{array}{c} I(X_S;Y_A) \\ -I(X_A;Y_E) \\ +H(Y_D|X_A) \end{array}\right\} & \text{if } I(X_A; Y_E) < I(X_S; Y_A) < I(X_A; Y_D, Y_E) \\ H(Y_D|Y_E) & \text{otherwise.} \end{cases} \tag{18}$$

Lemma 1: Through public transmission by D, any secret key rate R_K that satisfies

$$R_K < \alpha_E - \alpha_S \tag{19}$$

is achievable between D and S, when E is the eavesdropper.

Evaluating $\alpha_E - \alpha_S$ for α_S and α_E as given by (17) and (18) concludes the proof of the theorem. When verifying this, keep in mind that because of the Markov chain $Y_D \mathrel{-\!\!\circ\!\!-} X_A \mathrel{-\!\!\circ\!\!-} Y_E$, we have $H(Y_D|X_A) = H(Y_D|X_A, Y_E)$. It remains to prove Lemma 1:

Proof: We show the existence of a source code that achieves R_K as given in the lemma.

Code construction: We construct a random binning scheme to be used at D in the following way:

- Every sequence $\mathbf{y}_D \in \mathcal{L}(Y_D)$ is thrown into a randomly (uniformly) chosen bin, indexed by $i \in \{1, \ldots, 2^{T(\alpha_S + \epsilon_1)}\}$.

- Then, for every bin i, every sequence $\mathbf{y}_D$ in it is thrown into a randomly (uniformly) chosen sub-bin indexed by $k \in \{1, \ldots, 2^{T(\alpha_E - \alpha_S)}\}$.

Let $\mathbf{Y}_D$ be the received sequence from the forward transmission over the network. The bin index $I = i(\mathbf{Y}_D)$ will be communicated over the public channel, while the sub-bin index $K = k(\mathbf{Y}_D)$ is the secret message.

Encoding: D sends I over the public channel.

Decoding: S identifies a unique sequence $\hat{\mathbf{y}}_D$ in bin I such that $\hat{\mathbf{y}}_D \in \mathcal{L}(Y_D|\mathbf{x}_S)$. The dominant error event is that $\hat{\mathbf{y}}_D$ is not unique in bin I.

Virtual Decoding at the Eavesdropper: If a genie gave K to E, then E would know (I, K). It could then identify a unique sequence $\hat{\hat{\mathbf{y}}}_D$ in sub-bin K of bin I such that $\hat{\hat{\mathbf{y}}}_D \in \mathcal{L}(Y_D|\mathbf{y}_E)$. The dominant error event is that $\hat{\hat{\mathbf{y}}}_D$ is not unique in sub-bin K of bin I.

Expected Error Probability: Through analysis of the expected error probability, one can conclude that there exists a binning scheme for which

$$\mathbf{P}(\text{error at } S) \leq \epsilon_0 \tag{20}$$

and

$$\frac{1}{2^{T(\alpha_E - \alpha_S)}} \sum_{k=1}^{2^{T(\alpha_E - \alpha_S)}} \mathbf{P}(A_k) \leq \epsilon_0, \tag{21}$$

where A_k is the error event of the virtual (genie-aided) decoder at the eavesdropper for a given $K = k$. We then use an adaption of Fano's inequality (similar to Lemma 3 in [5]) to conclude that for that binning scheme,

$$H(\mathbf{Y}_D|K, I, \mathbf{Y}_E) \leq T\epsilon_2. \tag{22}$$

Equivocation Analysis: From Lemma 2 in [5], we have

$$\begin{aligned}
H(K|I,\mathbf{Y}_E) \geq{} & H(\mathbf{Y}_D) - I(\mathbf{Y}_D; \mathbf{Y}_E, I) \\
& - H(\mathbf{Y}_D|K, I, \mathbf{Y}_E) \\
={} & H(\mathbf{Y}_D) - I(\mathbf{Y}_D; \mathbf{Y}_E) \\
& - I(\mathbf{Y}_D; I|\mathbf{Y}_E) - H(\mathbf{Y}_D|K, I, \mathbf{Y}_E) \\
\geq{} & H(\mathbf{Y}_D) - I(\mathbf{Y}_D; \mathbf{Y}_E) \\
& - H(I) - H(\mathbf{Y}_D|K, I, \mathbf{Y}_E) \\
={} & H(\mathbf{Y}_D|\mathbf{Y}_E) - H(I) - H(\mathbf{Y}_D|K, I, \mathbf{Y}_E) \qquad (23) \\
\geq{} & T\big(\alpha_E - (\alpha_S + \epsilon_1) - \epsilon_2\big),
\end{aligned}$$

where we have used (22) and the fact that $\mathbf{Y}_D$ is uniformly distributed in $\mathcal{L}(Y_D)$ to bound the last two terms. For the first term in (23) we have used

$$\begin{aligned}
H(\mathbf{Y}_D|\mathbf{Y}_E) &= \sum_{\mathbf{y}_E \in \mathcal{L}(Y_E)} \mathbf{P}(\mathbf{Y}_E = \mathbf{y}_E) H(\mathbf{Y}_D|\mathbf{Y}_E = \mathbf{y}_E) \\
&\doteq \sum_{\mathbf{y}_E \in \mathcal{L}(Y_E)} \mathbf{P}(\mathbf{Y}_E = \mathbf{y}_E) \log L(Y_D|\mathbf{y}_E) \\
&= T\alpha_E \cdot \underbrace{\sum_{\mathbf{y}_E \in \mathcal{L}(Y_E)} \mathbf{P}(\mathbf{Y}_E = \mathbf{y}_E)}_{=1} \\
&= T\alpha_E,
\end{aligned}$$

where the asymptotic limit of $H(\mathbf{Y}_D|\mathbf{Y}_E = \mathbf{y}_E)$ follows from the uniformity stated in Lemma 2 in the appendix.

Rate Analysis: The rate of the sub-bin index K is equal to $\alpha_E - \alpha_S$.

Hence, our scheme achieves perfect secrecy of the key K, at a rate R_K as stated in Lemma 1. ■

VI. Discussion

In this paper, we introduce the one-relay network with feedback with a secrecy perspective, which, to the best of our knowledge, is a new problem setup. The motivation for this setting is that it captures a critical aspect of secret key generation via public feedback in networks, namely the thinning effect created by relays. This appears to be a new phenomenon, creating a further difficulty in the application of the techniques of Maurer and others for generating secrecy via feedback. The model introduced here isolates this phenomenon by considering it in the simplest possible setting. We derive a list size calculation to deal with it. This technical result might be of independent interest.

However, new challenges arise when more general networks are considered. Even without a secrecy constraint, no description of the capacity of general wireless networks is known. One possible way to proceed may be to combine the insight from the one-relay network discussed here with the more tractable signal interaction models presented in [6], [7].

References

[1] A. D. Wyner, “The wire-tap channel,” *Bell System Tech. J.*, vol. 54, pp. 1355–1387, October 1975.

[2] I. Csiszár and J. Körner, “Broadcast channels with confidential messages,” *IEEE Trans. Inform. Theory*, vol. 24, no. 3, May 1978.

[3] Y. Oohama, “Relay channels with confidential messages,” *IEEE Trans. Inform. Theory*, Nov. 2006, submitted.

[4] L. Lai and H. E. Gamal, “The relay-eavesdropper channel: Cooperation for secrecy,” *IEEE Trans. Inform. Theory*, vol. 54, no. 9, pp. 4005–4019, Sept. 2008.

[5] E. Perron, S. Diggavi, and E. Telatar, “On cooperative wireless network secrecy,” École Polytechnique Fédérale de Lausanne, Switzerland, Tech. Rep. LICOS-REPORT-2008-009, Aug. 2008, http://infoscience.epfl.ch/record/126166, submitted to IEEE Infocom 2009.

[6] A. Avestimehr, S. Diggavi, and D. Tse, “Wireless network information flow,” in *Proceedings of Allerton Conference on Communication, Control, and Computing*, Illinois, USA, Sept. 2007, see: http://licos.epfl.ch/index.php?p=research_projWNC.

[7] ——, “Approximate capacity of gaussian relay networks,” in *Proc. of the IEEE Int. Symposium on Inform. Theory*, Toronto, Canada, July 2008, see: http://licos.epfl.ch/index.php?p=research_projWNC.

[8] T.M.Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[9] U. M. Maurer, “Secret key agreement by public discussion from common information,” *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 733–742, May 1993.

[10] R. Ahlswede and I. Csiszár, “Common randomness in information theory and cryptography – part I: Secret sharing,” *IEEE Trans. Inform. Theory*, vol. 39, no. 4, pp. 1121–1132, July 1993.

[11] M. Wegman and J. Carter, “New hash functions and their use in authentication and set equality,” *Journal of Computer and System Sciences*, vol. 22, pp. 265–279, 1981.

[12] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, May 2005.

[13] A. Avestimehr, S. Diggavi, and D. Tse, “A deterministic approach to wireless relay networks,” in *Proceedings of Allerton Conference on Communication, Control, and Computing*, Illinois, USA, Sept. 2007, see: http://licos.epfl.ch/index.php?p=research_projWNC.

[14] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press, 1981.

VII. Appendix

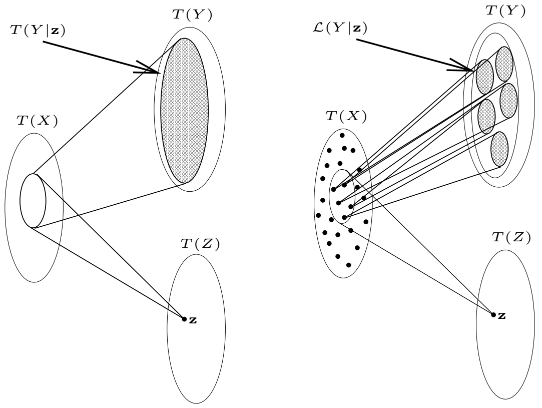


Fig. 4. Illustration of the list $\mathcal{L}(Y|\mathbf{z})$ for two cases: when $I(X;Y,Z) < \tilde{R}$ (on the left) and when $\tilde{R} < I(X;Y,Z)$ (on the right).

Lemma 2: Consider discrete alphabets $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ and a probability distribution p_{XYZ} for which $Y \multimap X \multimap Z$ form a Markov chain. Let $\mathcal{C}$ be a randomly generated code of blocklength T and $2^{T\tilde{R}}$ codewords for the $\mathcal{X}$ alphabet where each codeword is chosen independently and uniformly from the p_X-typical set $T(X)$. Let $\mathcal{L}(Z)$ and $\mathcal{L}(Y|\mathbf{z})$ be as in Definition 14. Then for large T, for almost every $\mathcal{C}$, and all $\mathbf{z} \in \mathcal{L}(Z)$ the list $\mathcal{L}(Y|\mathbf{z})$ is of size $L(Y|\mathbf{z}) \doteq 2^{TA}$, where

$$A = \begin{cases} H(Y|X) & \text{if } \tilde{R} < I(X;Z) \\ \tilde{R} - I(X;Z) + H(Y|X) & \text{if } I(X;Z) < \tilde{R} < I(X;Y,Z) \\ H(Y|Z) & \text{if } I(X;Y,Z) < \tilde{R}. \end{cases} \tag{24}$$

In addition, if the codeword $\mathbf{x} \in \mathcal{C}$ is chosen uniformly at random, then the members of $\mathcal{L}(Y|\mathbf{z})$ are also uniformly distributed.

Remark: The lemma tells us that when $\tilde{R} < I(X;Y,Z)$, the list $\mathcal{L}(Y|\mathbf{z})$ consists of disjoint fans, each of size $2^{TH(Y|X)}$ and associated with a codeword in $\mathcal{C}$. On the other hand, as soon as $I(X;Y,Z) < \tilde{R}$, $\mathcal{L}(Y|\mathbf{z})$ and $T(Y|\mathbf{z})$ are essentially the same. Figure 4 shows an illustration of this phenomenon.

Proof: When $\tilde{R} < I(X;Z)$, then it is well known that $L(X|\mathbf{z}) = 1$ for $\mathbf{z} \in \mathcal{L}(Z)$ and for almost all $\mathcal{C}$. Hence, for almost all $\mathcal{C}$, there is exactly one possible transmitted $\mathbf{x}$. Hence, the list $\mathcal{L}(Y|\mathbf{z})$ is the fan going from that $\mathbf{x}$ to $T(Y)$, which is, with high probability, of size $2^{TH(Y|X)}$.

Now, assume that $\tilde{R} > I(X;Z)$. Note that in this case, almost any typical $\mathbf{z}$ is jointly typical with a codeword and thus, $\mathcal{L}(Z)$ and $T(Z)$ are essentially the same. Fix any $\mathbf{z} \in T(Z)$. Consider an arbitrary sequence $\mathbf{y} \in T(Y|\mathbf{z})$. Define $\tilde{p}$ to be the probability that this particular sequence $\mathbf{y}$ is possible for $\mathbf{z}$, *i.e.*, $\tilde{p}$ is the probability that there exists a codeword $\mathbf{x}$ such that $(\mathbf{x}, \mathbf{y}, \mathbf{z}) \in T(X,Y,Z)$.

Define $\mathcal{E}_{\mathbf{x}}$ to be the event that the sequence $\mathbf{x} \in T(X)$ was selected as a codeword. Then, we have

$$\begin{aligned} \tilde{p} =& \mathbf{P}\Big(\cup_{\mathbf{x} \in T(X|\mathbf{y},\mathbf{z})} \mathcal{E}_{\mathbf{x}} \Big) && (25) \\ \leq& \min\Big\{1, \sum_{\mathbf{x} \in T(X|\mathbf{y},\mathbf{z})} \mathbf{P}(\mathcal{E}_{\mathbf{x}})\Big\} && (26) \\ =& \min\Big\{1, 2^{TH(X|Y,Z)} \frac{2^{T\tilde{R}}}{2^{TH(X)}}\Big\} && (27) \\ =& \min\Big\{1, 2^{T(\tilde{R} - I(X;Y,Z))}\Big\}, && (28) \end{aligned}$$

where we used the union bound for the inequality. Note that the events $\mathcal{E}_{\mathbf{x}}$ are independent.

Case 1: If $\tilde{R} > I(X;Y,Z)$, then $2^{T(\tilde{R} - I(X;Y,Z))}$ gets very large for large T and hence the union bound is tight and equal to 1.

Case 2: If $\tilde{R} < I(X;Y,Z)$, then $2^{T(\tilde{R} - I(X;Y,Z))}$ goes to zero for large T and again the union bound is tight and equal to $2^{T(\tilde{R} - I(X;Y,Z))}$.

The expected list-size at Z is found by multiplying this probability with the total number of sequences $\mathbf{y}$ that are jointly typical with $\mathbf{z}$, which yields

$$\tilde{p} \cdot 2^{TH(Y|Z)} = \begin{cases} 2^{TH(Y|Z)} & \text{when } \tilde{R} > I(X;Y,Z) \\ 2^{T(\tilde{R} - I(X;Z) + H(Y|X))} & \text{when } \tilde{R} < I(X;Y,Z), \end{cases} \tag{29}$$

where the second exponent is obtained by noting that $H(Y|Z) - I(X;Y|Z) = H(Y|X,Z) = H(Y|X)$, using the Markov chain $Y \multimap X \multimap Z$ in the last step.

Let $L(Y|\mathbf{z})$ be the random list-size. It remains to be shown that for $\tilde{R} > I(X;Z)$, $L(Y|\mathbf{z})$ hardens to the expected list-size $\mathrm{E}\left[L(Y|\mathbf{z})\right]$ for large T. We can write

$$\begin{aligned} L(Y|\mathbf{z}) &= \sum_{\mathbf{y} \in T(Y|\mathbf{z})} \mathbf{1}_{\{\bigcup_{\mathbf{x} \in T(X|\mathbf{z})} \{(\mathbf{x},\mathbf{y}) \in T(X,Y|\mathbf{z}), \mathcal{E}_{\mathbf{x}}\}\}} \\ &= \sum_{\mathbf{y} \in T(Y|\mathbf{z})} \mathbf{1}_{\{\bigcup_{\mathbf{x} \in T(X|\mathbf{z}) \cap \mathcal{C}} \{(\mathbf{x},\mathbf{y}) \in T(X,Y|\mathbf{z})\}\}} \\ &\leq \sum_{\mathbf{y} \in T(Y|\mathbf{z})} \sum_{\mathbf{x} \in T(X|\mathbf{z}) \cap \mathcal{C}} \mathbf{1}_{\{(\mathbf{x},\mathbf{y}) \in T(X,Y|\mathbf{z})\}}. \qquad (30) \end{aligned}$$

By computing the second moment of (30), one can show that $\mathrm{Var}\left(L(Y|\mathbf{z})\right)$ is of the same order as $\mathrm{E}\left[L(Y|\mathbf{z})\right]$. Then, using Chebyshev's inequality, we conclude that the probability that $L(Y|\mathbf{z})$ deviates considerably from $\mathrm{E}\left[L(Y|\mathbf{z})\right]$ is small.

Assume now that $\mathbf{X} \in \mathcal{C}$ is a uniformly chosen codeword from a given code $\mathcal{C}$. The uniform distribution of $\mathbf{Y} \in \mathcal{L}(Y|\mathbf{z})$ can be shown in three steps. First, one can use the law of total probabilities to show that the random sequence $\mathbf{Z}$ is uniformly distributed in $\mathcal{L}(Z)$. Then, from Baye's rule, one can conclude that for a given $\mathbf{z}$, the conditional distribution of the random codeword $\mathbf{X}$ is uniform over the set $\mathcal{C} \cap T(X|\mathbf{z})$. Finally, one can again use the law of total probabilities to show that conditioned on $\mathbf{z}$, the sequence $\mathbf{Y}$ is uniformly distributed in $\mathcal{L}(Y|\mathbf{z})$. ■