Information-Theoretic Secrecy for Wireless Networks

Etienne Perron

Information-Theoretic Secrecy for Wireless Networks

Etienne Perron

EPFL - Ecole Polytechnique Fédérale de Lausanne

Thesis No. 4476 (September 2009)

Thesis presented to the faculty of computer and communication sciences for obtaining the degree of Docteur ès Sciences

Accepted by the jury:

S. Diggavi and **E. Telatar**
Thesis directors

C. Fragouli
Expert

A. Khisti
Expert

U. Maurer
Expert

R. Urbanke
President of the jury

Ecole Polytechnique Fédérale de Lausanne, 2009

To Muriel,
for years of love and support.

Abstract

The aim of information-theoretic secrecy is to ensure that an eavesdropper who listens to the wireless transmission of a message can only collect an arbitrarily small number of information bits about this message. In contrast to cryptography, there are no assumptions on the computational power of the eavesdropper.

Information-theoretically secret communication has been studied for many particular wireless network topologies. In the main part of this thesis, we consider such communication for *arbitrary* acyclic wireless network topologies. We provide lower and upper bounds on the strong perfect secrecy capacity for the case when the channels of the network are either Gaussian or deterministic. These results are based on the recent understanding of the capacity of wireless networks (without secrecy constraints) by Avestimehr, Diggavi and Tse.

As a side result, we give inner and outer bounds on the capacity region for the multisource problem in arbitrary wireless networks with Gaussian or deterministic signal interaction. For linear deterministic signal interaction, we find the exact capacity region. For Gaussian signal interaction, we are able to bound the gap between the two bounds on the capacity region. This gap depends only on the network topology, but not on the signal-to-noise ratio (SNR), which leads to an approximation of the capacity region for the high SNR regime.

We further consider a particular network topology, called the fan-network, in which we assume that an eavesdropper has physical access to every node in a subset of the relay nodes. We give a general upper bound on the perfect secrecy capacity, and we characterize the perfect secrecy capacity for two special cases.

In the second part of the thesis, we consider interactive secrecy, *i.e.*, secrecy in the presence of a public feedback link from the destination to the source. We focus on the problem of secret key generation rather than secret communication. The benefit of public discussion for secret key generation in a broadcast channel was first shown by Maurer. We extend his ideas to a relay network called the line network, leading to a lower bound on the strongly secret key capacity for this network topology.

Finally, we introduce a new channel coding setup called the interference-multiple access (IMA) channel. This channel is a variant of the interference channel where one of the receivers is required to decode the messages from both transmitters. We derive an inner bound on the capacity region of the IMA channel, as well as an outer bound for the so-called structured IMA channel. In a semi-deterministic version of the structured IMA channel, the bounds match, providing a characterization of the capacity region. In the Gaussian case, we obtain a 1 bit-approximation of the capacity region. We also show an inner bound on the equivocation-capacity region for the IMA channel, where we require that part of the private message for one receiver is kept information-theoretically secret from the other receiver.

Keywords: information-theoretic secrecy, unconditional secrecy, wireless relay networks, arbitrary topologies, deterministic model, Gaussian model, multisource problem, public feedback, public discussion, secret key generation, interference channel, interference-multiple access channel

Zusammenfassung

Der Zweck informationstheoretischer Geheimhaltung ist, sicherzustellen, dass während einer drahtlose Nachrichten-Übertragung ein Lauscher nicht mehr als eine beliebig kleine Anzahl von Informations-Bits über die Nachricht sammeln kann. Im Gegensatz zur Kryptographie machen wir hierbei keine Annahmen über die Rechenleistung des Lauschers.

Datenübertragung mit informationstheoretischer Geheimhaltung wurde in den letzten Jahren für viele Netzwerk-Konfigurationen untersucht. Im Hauptteil dieser Dissertation behandeln wir solch geheime Datenübertragung für *beliebige* azyklische drahtlose Netzwerk-Konfigurationen. Wir bestimmen untere und obere Schranken der streng definierten Geheimhaltungs-Kapazität unter der Annahme, dass die Kanäle im Netzwerk entweder deterministisch sind oder durch Gauss'sches Rauschen gestört werden. Dieser Teil der Dissertation baut auf den jüngsten Veröffentlichungen von Avestimehr, Diggavi und Tse auf, in welchen die Kapazität von beliebigen drahtlosen Netzwerken (ohne Geheimhaltungs-Bedarf) untersucht wurde.

Als Nebenergebnis beschreiben wir innere und äussere Schranken des Kapazitäts-Gebiets für das "Mehrfachquellen"-Problem. Hier enthält ein beliebiges drahtloses Netzwerk (mit Gauss'schen oder deterministischen Kanälen) mehrere Datenquellen, welche ihre Nachrichten alle an denselben Empfänger übermitteln (ohne Geheimhaltung). Wenn die Kanäle deterministisch *und* linear sind, erhalten wir eine genaue Beschreibung des Kapazitäts-Gebiets. Für Gauss'sche Kanäle können wir den Abstand zwischen der inneren und der äusseren Schranke begrenzen. Dieser Abstand hängt von der Netzwerk-Konfiguration ab, aber nicht von der Signalstärke der Übertragung, was bei hohen Signalstärken zu einer Annäherung des Kapazitäts-Gebiets führt.

Als nächstes betrachten wir eine bestimmte Netzwerk-Konfiguration, die wir "Fächer-Netzwerk" nennen. Hier nehmen wir an, dass ein Lauscher zu einer gewissen Anzahl der Knoten des Netzwerks physischen Zugang hat, und dass er deshalb die Signale, die von diesen Knoten empfangen werden, direkt beobachten kann. Wir berechnen eine obere Schranke der streng definierten Geheimhaltungs-Kapazität für den allgemeinen Fall. Für zwei Spezialfälle können wir diese Geheimhaltungs-Kapazität genau herleiten.

Im zweiten Teil der Dissertation betrachten wir interaktive Geheimhaltungs-

Methoden. Wir nehmen an, dass ein ungesicherter Rückmeldungs-Kanal vom Empfänger zum Sender zur Verfügung steht. Anstelle von geheimer Kommunikation interessiert uns hier eine etwas andere Fragestellung, nämlich das Erstellen einer geheimen Zufalls-Sequenz. Diese Sequenz, "Schlüssel" genannt, soll dem Sender und dem Empfänger bekannt sein, jedoch nicht dem Lauscher. Maurer hat als Erster erkannt, dass ungesicherter, bilateraler Datenaustausch beim Erstellen eines geheimen Schlüssels mit Hilfe eines Rundfunk-Kanals von Vorteil sein kann. Wir wenden diese Ideen auf ein kleines Netzwerk mit einem weiterleitenden Knoten an. Für dieses sogenannte Linien-Netzwerk finden wir eine untere Schranke der Schlüssel-Geheimhaltungs-Kapazität.

Letztendlich führen wir ein neuartiges Kanal-Kodierungs-Problem ein, genannt "Interferenz- und Mehrfachzugriffs-Kanal" (IMZ-Kanal). Dieser Kanal ist eine Variante des Interferenz-Kanals mit dem Unterschied, dass einer der Empfänger die Nachrichten von beiden Sendern wiedergeben soll. Wir ermitteln eine innere Schranke des Kapazitäts-Gebiets für den allgemeinen IMZ-Kanal sowie eine äussere Schranke für den sogenannten strukturierten IMZ-Kanal. Es stellt sich heraus, dass sich diese Schranken für eine halbdeterministische Version des strukturierten IMZ-Kanals treffen, und wir erhalten somit eine genaue Beschreibung des Kapazitäts-Gebiets für diesen Spezialfall. Für den Gauss'schen IMZ-Kanal finden wir eine Annäherung des Kapazitäts-Gebiets auf ein Bit genau. Ferner beschreiben wir eine innere Schranke des Geheimhaltungs-Kapazitäts-Gebiets für den IMZ-Kanal unter der Bedingung, dass ein Teil von der Nachricht, welche nur für einen Empfänger bestimmt ist, vor dem anderen Empfänger geheim gehalten wird.

Stichworte: informationstheoretische Geheimhaltung, bedingungslose Geheimhaltung, drahtlose Netzwerke mit weiterleitenden Knoten, beliebige Konfigurationen, deterministisches Modell, Gauss'sches Modell, Mehrfachquellen-Problem, ungesicherte Rückmeldung, ungesicherter Datenaustausch, Erstellen eines geheimen Schlüssels, Interferenz-Kanal, Interferenz- und Mehrfachzugriffs-Kanal

Acknowledgements

This thesis was supported in part by SNSF Grant # 200021-105640/1.

I am deeply thankful to many people who have all contributed to this thesis and to making my time as a student a very enriching experience.

First, I would like thank both my advisors Suhas Diggavi and Emre Telatar. I was very fortunate that Emre agreed to supervise my thesis. I am also very thankful that he introduced me to Suhas, which led to the best co-supervision one could think of. I enjoyed the countless hours that both spent with me, and I am grateful for all the occasions when they managed to be there for me even from distant locations and at unusual times to help me when I had a question or problem. I learned many things from both Emre and Suhas, both on a technical and on a personal level, and they wonderfully complemented each other in the guidance of my work. I greatly appreciated the kindness, honesty and good humour that were part of every interaction we had. These and many other values will continue to be an inspiration for me in my future professional life. I am also very grateful to Olivier Lévêque. It is partly thanks to his encouragement and support that I joined the information processing group.

I am also indebted to Sylviane Dal Mas, Rüdiger Urbanke, Bixio Rimoldi, Alex Grant, Matthias Grossglauser, and David Tse. As an undergraduate student, I received from them extraordinary support and guidance during classes and internships.

I would like to thank Christina Fragouli, Ueli Maurer, Ashish Khisti and Rüdiger Urbanke for agreeing to be on my thesis committee and for the many useful comments that they provided. In particular, the extension to strong secrecy, that turned out to be very important, was suggested by Professor Maurer. Special thanks go to Ashish for the long and fruitful discussions even before he was on my committee. Among many other sensible comments, he provided ideas for the upper bound in Theorem 3.3 and for the MISO upper bound in the numerical example of the Gaussian diamond network, as well as the reference to [26].

I am grateful to Muriel Bardet for doing all the thinking in the lab. It was very reassuring to know that she would take care of all the organizational things of my PhD studies. Also, many thanks to her, Françoise Behn and Yvonne Huskie for all the great social events of our lab, as well as to Damir

Laurenzi and Giovanni Cangiani for always being so prompt and friendly when dealing with my computer problems.

I wish to acknowledge my collaborators Dinkar Vasudevan, Chao Tian, Milan Vojnović, and Rethnakaran Pulikkoonattu. I strongly value the discussions and work sessions I had with them. Although the results that I obtained in collaboration with them did not directly contribute to this thesis, the things that I learned from them did.

My former and current colleagues at EPFL made my PhD time memorable: Dinkar, Sanket, Shrinivas, Vish, Marius, Vojislav, Christine, Stéphane B., Bertrand, Eren, Peter, Sibi, Stéphane M., John, Soheil, Dominique, Emmanuel, Ayfer, Satish, Mohammad, Cyril, Abdel, Jérémie, Shirin, Nicolae, Alberto, Olivier L., Nicolas, Chao, Iryna, Amin, Hamed, Mahdi, Ratna, Olivier R., Gianluca, Michal, Maciej, Hung, Maxim, Imad and many others. I thank them for the friendship and the good times, and for all the help I received from them. Many thanks also to my friends outside EPFL, especially to Urs and Thomas for their advice and the many discussions.

I owe a lot to my parents Doris and Jörg and to my sister Noëlle. I am very grateful for their advice, support and generosity. It is priceless for me to have a family as caring as them and to know that I can always rely on them. I would also like to thank my extended family for the continuous interest and understanding.

Words can hardly express how deeply thankful I am to my wife Muriel. I could make a long list of things that she has given me during the last years, but such a list could never be complete. Instead, I dedicate this thesis to her. She has contributed so much to this work that I truly feel that it is as much hers as mine.

Contents

1 Introduction

During the last two decades, a revolution has taken place in personal and corporate communication. Many devices like telephones, computers, keyboards or headphones, traditionally connected via cables, are now connected in a wireless manner. In many architectures, wireless technology is combined with a wired backbone. However, especially in military applications, wireless relay networks (e.g. sensor networks) that rely exclusively on the wireless medium are already being implemented. In such networks, some of the nodes play the role of relays, forwarding information from a predecessor to a successor in a flow of information through the network.

Secrecy or privacy is an important requirement of many communication applications. In some examples, communication without secrecy is literally useless, for instance when exchanging a password. A fundamental aspect of wireless communication is its broadcast nature, *i.e.,* transmission 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. In this thesis we restrict ourselves to the passive eavesdropper model, motivated by secure authentication protocols [47] which could potentially discover an active eavesdropper.

Our work aims at finding theoretical lower and upper bounds on the secret communication rate for point-to-point communication in wireless networks. To ask this question we first need to specify the secrecy notion that we seek. In 1949, Shannon introduced the notion of information theoretic secrecy [39], when he studied whether communication from a source to a destination can be kept secret from an eavesdropper who has *complete* access to the transmission. Shannon did not make any assumptions on computational capabilities of the eavesdropper. As one would expect, such a question resulted in the pessimistic answer that unless the source and destination had somehow a large amount of common randomness (key) kept secret from the eavesdropper, the task was impossible. In fact, the common randomness needed was of the same rate as the source message itself, making the resulting communication schemes (e.g. one-time pad) rather impractical.

This observation led to the *computational* approach pioneered by Diffie and Hellman [12], where instead of having such a long shared secret key, a shorter key is exploded into a larger one. The goal of the design is to guarantee that there is no efficient algorithm for the eavesdropper to discover the information transmitted. For example, the security of the well-known RSA public-key cryptosystem [38] is based on the difficulty of factoring large integers, and other cryptographic protocols are based on the difficulty of computing discrete logarithms over groups (e.g. [12]). Both these protocols are based on the (as of yet) unproven computational intractability hypothesis for these algorithmic problems. Another potential problem with this approach might be that theoretically, quantum computers could make difficult algorithmic problems tractable [41]. Clearly the secrecy of any communication system could be enhanced if one could communicate even a small number of bits in an information-theoretically secret way between a source and a destination. In particular, one of our motivations is that we can potentially generate such a secret message using physical wireless channel properties. Such a functionality can be incrementally deployed in networks by passing this secret key to the higher layers in the network protocol stack where it could then be used to enhance the secrecy of cryptographic protocols. In this respect, the two approaches to secrecy can be complementary, with the information-theoretic secrecy used to provide further secrecy opportunities.

In wireless communication, even though the signal from the source is broadcast, it is received at the destination and the potential eavesdropper through *different* channels. It is this distinction that is exploited in information-theoretic secrecy for broadcast channels in the seminal work of wiretap channels by Wyner as well as Csiszár and Körner [48, 10]. However, not much is known about the cooperative secrecy setup, where there are relay nodes facilitating secure communication between a source and a destination. In this case, interference between signals from different relay nodes can be used to confuse an eavesdropper. Relay nodes can even generate random signals in order to "jam" the channel to the eavesdropper (this idea was introduced by Tekin and Yener in [43]). However, there is a trade-off, because every jamming signal can potentially hurt the legitimate decoder as well. Particular networks like the relay channel or the multiple-access channel have been studied in [21, 43].

In the main part of this thesis, we examine this problem for an *arbitrary* acyclic wireless network and provide provable secrecy guarantees. The same question in the context of wired networks, where the nodes are connected by independent point-to-point channels, was studied by Cai and Yeung [6]. The main difficulties in dealing with wireless relay networks are the broadcast nature of wireless communications as well as the fact that signals from simultaneously transmitting nodes interfere with one another at other nodes. This gives rise to complex signal interactions making the understanding of wireless networks difficult (even without secrecy requirements). Though there has been some recent understanding in terms of scaling laws for asymptotically large wireless networks without secrecy, pioneered by Gupta and Kumar [15, 49, 36], there has not been a complete understanding of communication for the non-asymptotic regime. Our work develops on the recent study of wireless network information flow by Avestimehr, Diggavi and Tse [3, 4, 5], which approximately characterized the unicast (or multicast) capacity without secrecy constraints. Our formulation asks a natural question of additionally keeping such a cooperative communication secure against a class of potential eavesdroppers.

Although information-theoretic secrecy is sometimes called "unconditional" secrecy, the results that use this notion rely on an important assumption: the channel to the eavesdropper is assumed to be known. In practice, this would imply that the location occupied and the hardware used by the eavesdropper is approximately known, which is not very realistic. One possible workaround is to consider a class of eavesdroppers, each modeling one possible channel to the unknown eavesdropper. This was for instance done in recent work on compound wiretap channels [22] and we also adopt this strategy. If a continuous range of eavesdropper channels is possible, then the class of eavesdroppers can be thought of as being a "discretization" of this range.

The limits of information-theoretic secrecy can be described as follows. If for the optimal transmission and relaying strategy, the signal at the eavesdropper is still stronger than the signal at the destination, then the secrecy capacity is zero. This limit occurs in all known information-theoretic secrecy setups. However, Maurer and later Ahlswede and Csiszár [28, 1] showed that in such situations, a feedback link from the destination to the source can enable secret communication. The existence of feedback from the destination to the source is a reasonable assumption for wireless relay 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 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, authenticated public channel, as in [28, 1]. 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 arbitrarily large capacity is quite strong, and needs to be refined in future work. The assumption that communication over the public channel is authenticated can be motivated through the existence of secure authentication protocols. Extensions to non-authenticated public channels might be possible, similarly to [30, 31, 32]. However, we believe that assuming a public, infinite-capacity feedback channel is an interesting first step.

The thesis is organized as follows.

In Chapter 2, we give a table of notations and several important definitions, as well as a description of the previous work of [5].

In Chapter 3, we provide lower bounds on the perfect secrecy capacity for arbitrary acyclic network topologies with an arbitrary finite class of eavesdroppers and an arbitrary finite number of noise-inserting relays. The (strong) perfect secrecy capacity is the largest rate at which a source node can reliably communicate to a destination node in the network such that only an arbitrarily small number of information bits is leaked to any of the eavesdroppers. In this achievability result, we use two different models for the interaction between different nodes of the network: a widely accepted Gaussian broadcast and interference model as well as a deterministic interaction model introduced in [2]. This deterministic model does not take receiver noise into account, and hence provides good intuitions on the effect of interference. In particular, a linear deterministic version of this model allows us to find coding schemes for example networks in a straight-forward way. The intuitions gained through these examples can be valuable in the design of communication protocols. In addition, the lower bound for deterministic interaction is valid even for cyclic networks. We also provide a simple upper bound on the perfect secrecy capacity for arbitrary wireless relay networks. As an auxiliary result, we present inner and outer bounds on the capacity region of a multisource problem in wireless networks (without secrecy). For linear deterministic signal interaction, these bounds match and for Gaussian signal interaction, we show that the gap is constant with respect to the signal-to-noise ratio. We presented these results in our publications *3.* and *5.* (see CV at the end of this thesis).

In Chapter 4, we focus on a particular topology called the "fan network", where the source node is connected to all the relay nodes through one broadcast channel. The destination node collects data from all the relays through one multiple-access channel. The situation here is that we assume that each possible eavesdropper can capture a given subset of the relay nodes and observe the received signals of those nodes, without however being able to alter their operations (passive eavesdropper). By defining this particular setup, we sidestep several difficulties that can be present in a general network. For instance, interference occurs only at the destination, and hence, the cooperation between relays can not be used to confuse any eavesdroppers. We provide a general upper bound on the perfect secrecy capacity for the fan network. In two special cases, this bound is shown to be tight, hence yielding a characterization of the perfect secrecy capacity.

In Chapter 5, we consider secret communication in a wireless network with feedback. The feedback link is modeled as a public link from the destination to the source and the eavesdropper. In this chapter, the aim is to establish a secret key shared by the source and the destination, rather than secret communication of a message from the source to the destination. Performance is measured through the rate of this key. We restrict our attention to the simplest relay network, which is the line network with one relay and one eavesdropper. We find that in contrast to the broadcast channel, the common randomness established between the source and the destination after the forward transmission of an i.i.d. sequence over the network does *not* resemble the output of a discrete memoryless source. Hence, to be able to use a structured binning strategy during the public backward transmission, we first need to compute the sizes of the lists of possible received sequences at the destination from the point of view of the source and the eavesdropper. These results could be of independent interest. Using these results, we provide a lower bound on the secret key capacity for the line-network with feedback. This work was first presented in *6.* (see CV at the end of the thesis).

In Chapter 6, we present a particular small network (or channel) called the interference-multiple access (IMA) channel, which is a modification of the interference channel. It is defined by two transmitters and two receivers. As in the interference channel, each transmitter sends an independent message. However, we assume that one receiver is required to decode the message from one of the transmitters, while the other receiver is required to decode *both* messages. This setup is new and is not a special case of the interference channel. We provide an inner bound on the capacity region (without secrecy constraints) for the general IMA channel, an outer bound for a certain class of so-called structured IMA channels, as well as an expression for the gap between the inner and outer bound for structured IMA channels. We show that for a semi-deterministic channel, the bounds match, yielding a complete characterization. We also show that for the Gaussian IMA channel, the gap is at most 1 bit. These approximation results are inspired by recent results on the interference channel [44, 13]. Finally, we ask whether the private message (decoded only by one of the receivers) can be kept information-theoretically secret from the other receiver. We provide an inner bound on the equivocation-capacity region for general IMA channels, and we also show a special case where a code that was not designed for secrecy still provides a certain amount of "unintentional" secrecy. The results on the IMA channel appeared in *2.* (see CV at the end of the thesis).

Definitions and Preliminaries 2

2.1 Notation

$\triangleq$	definition
$\doteq$	relative equality, defined as $f(n) \doteq g(n) \Leftrightarrow \frac{f(n)}{g(n)} \to 1$ as $n \to \infty$
$\overset{\delta}{=}$	δ-exponential equality, defined in Definition 5.13
$\simeq$	approximate equality, used to state that the equality can be made arbitrarily precise by appropriately choosing some parameter
$\mathcal{A}$	set of all relay nodes of a network
α	"constant" that depends on the number of nodes in the network, $\mathcal{O}\left(\vert\mathcal{V}\vert\right)$
β	"constant" that depends on the number of nodes in the network and on the maximum out-degree, $\mathcal{O}\left(d\vert\mathcal{V}\vert^2\right)$
$\mathcal{B}$	set of all noise-inserting relay nodes of a network
B	a junk (or noise) rate
C	capacity of a wireless network (without secrecy)
$\bar{C}$	cut-set upper bound on the capacity of wireless networks
C_s	weak perfect secrecy capacity of a wireless network
C_s^ρ	(weak) ρ-almost perfect secrecy capacity of a wireless network
$\bar{C}_s$	strong perfect secrecy capacity of a wireless network
$\bar{C}_k$	strong secret key capacity for the line network
γ	"constant" that depends on the number of nodes in the network and on the maximum out-degree, $\mathcal{O}\left(d\vert\mathcal{V}\vert\right)$
d	maximum out-degree of any node in a network
$\mathcal{D}$	set of all destination nodes in a network
D	a particular destination node

$\mathcal{E}$	set of all eavesdroppers
E	a particular eavesdropper
f_i	block encoding function at node i
$\tilde{f}_i$	decoding function for Receiver i in the IMA channel
ϕ_i	auxiliary quantity used in the main results of Chapter 3, where $i = D, E$
Φ_E	auxiliary set used in the Gaussian diamond example of Chapter 3
g_j	channel function at node j for deterministic wireless networks
$\mathcal{G}$	wireless relay network, $\mathcal{G} = (\mathcal{V}, \mathcal{L})$
$\mathbf{G}_{i,j}$	channel transform matrix in linear deterministic signal interaction
$H(\cdot)$	entropy in bits
$h(\cdot)$	differential entropy in bits
$I(\cdot;\cdot)$	mutual information in bits
$\mathrm{In}(i)$	all "predecessors" of node i in a network
J	junk message
K	number of stages in an unfolded network
κ	constant used in some capacity theorems
$\mathcal{L}$	annotated channels of a network
L	number of layers in a layered network
$\Lambda(\mathcal{I}, j)$	set of all $\mathcal{I}$-j-cuts, see Definition 2.19
λ_i	factor used in Gaussian signal interaction, see Definition 2.21
$\log$	base-2 logarithm
M	used in two different contexts, once denoting a special message, and once denoting a large integer
N	large integer used in the relation $T_c = (N + L)T$
Ω	a subset of $\mathcal{V}$ denoting a cut (Ω, Ω^c)
p^{G}	Gaussian product distribution, see Definition 2.21
Q	time-sharing random variable
q	number of bits of one symbol in a binary expansion
R	an information rate
R_e	a (weak) equivocation rate
R_s	a (strong) secrecy rate
$R_{\mathcal{I};j}(p)$	information-theoretic min-cut between $\mathcal{I}$ and j for discrete signal interaction (see Definition 2.20)
$R^{\mathrm{G}}_{\mathcal{I};j}$	information-theoretic min-cut between $\mathcal{I}$ and j for Gaussian signal interaction (see Definition 2.21)
$\mathcal{R}$	the capacity region of the IMA channel
ρ	constant used in the definition of ρ-almost perfect secrecy
$\mathcal{S}$	set of all source nodes in a network
S	a particular source node in a network
σ_X	standard deviation of X

T	block length of a block code
T_c	the communication length of a code (for a block code run N times, $T_c = (N+L)T$)
$\mathcal{T}_\delta(X)$	set of all robustly δ-typical sequences with respect to X (see Appendix F)
$\mathcal{T}_\delta$	short form for $\mathcal{T}_\delta(X)$, where it should be clear from the context which random variable we refer to
$\mathcal{V}$	set of all nodes of a network
W	information message for a block code
$\underline{W}$	information message for an arbitrary code (for a block code run N times, $\underline{W} = (W^{(1)}, \ldots, W^{(N)})$)
X	a random variable
$\mathcal{X}$	alphabet of the random variable X
$X_{\mathcal{V}}$	tuple of random variables (X_i), $i \in \mathcal{V}$
x	a deterministic value
$\mathbf{X}$	a random sequence of length T
$\underline{\mathbf{X}}$	a random sequence of length T_c
$\mathbf{x}$	a deterministic sequence of length T
$x[t]$	the t^{th} component of $\mathbf{x}$
ξ	a random variable, independent of everything else
$\underline{\mathcal{X}}_i(w)$	set of all plausible transmit sequences at node i under message w
$\underline{\hat{\mathcal{Y}}}_i(w)$	set of all plausible quantized sequences at node i under message w

2.2 General Definitions

2.2.1 Wireless Networks and Reliable Communication

This thesis treats the problem of secret communication over wireless relay networks. A wireless relay network is given by a certain number of nodes, connected via memoryless channels.

Definition 2.1 *A* ***discrete memoryless channel*** *is defined by a finite transmit alphabet $\mathcal{X}$, a finite receive alphabet $\mathcal{Y}$, as well as a conditional probability mass function $p_{Y|X}(y|x)$ for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. For a* ***continuous memoryless channel****, the receive alphabet is continuous, the transmit alphabet is either discrete or continuous, and the statistics of the channel are given by a conditional probability measure $\mu(\cdot|x)$ on $\mathcal{Y}$ for every $x \in \mathcal{X}$. The channel has no memory, meaning that at each use, the channel acts independently of other channel uses.*

The channel defined here is not necessarily a point-to-point channel, because the alphabets $\mathcal{X}$ and $\mathcal{Y}$ can well be the Cartesian products of the alphabets of a number of users (or nodes). Note that in this thesis, the only continuous memoryless channel we consider is the Gaussian channel, where $\mathcal{X}$ and $\mathcal{Y}$ are powers of $\mathbb{C}$.

Definition 2.2 *A* ***wireless relay network*** *is defined using a pair $\mathcal{G} = (\mathcal{V}, \mathcal{L})$, where $\mathcal{V}$ is the set of vertices representing the communication nodes in the relay network and $\mathcal{L}$ is the set of annotated channels between the nodes, which describe the signal interactions. We consider a special node $S \in \mathcal{V}$ as the source of the message which wants to communicate to 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 relay nodes that are allowed to generate randomness (independently of the other nodes) are called* ***noise inserting nodes*** *and denoted by the subset $\mathcal{B} \subseteq \mathcal{A}$. For any node $i \in \mathcal{V}$, we define $In(i)$ to be the* ***set of predecessors*** *of i, i.e., all the nodes $j \in \mathcal{V}$ such that $(j, i) \in \mathcal{L}$.*

Note that the channels $\mathcal{L}$ are not point-to-point links, rather, they model how the transmitted signals are superimposed and received at the receiving nodes (*i.e.,* there is broadcast and interference). In some of our results and proofs, we make the distinction between layered and non-layered networks.

Definition 2.3 *A wireless relay network is* ***layered*** *if for every (i, j) such that $i \in \{S\} \cup \mathcal{B}$ and $j \in \mathcal{V}$, all the paths from i to j have the same length (the same number of hops in $\mathcal{L}$). A* ***non-layered*** *network is a network in which at least one node pair (i, j) does not have this property.*

Most Gaussian results in this thesis are only valid for acyclic networks.

Definition 2.4 *A wireless relay network is* ***acyclic*** *if there is no node $i \in \mathcal{V}$ such that there is a path from i to itself, where a path is defined by a sequence of hops in $\mathcal{L}$.*

To communicate a message from S to D, the channels of the network can be used several times in a row. We assume that all the nodes in $\mathcal{V}$ are synchronized and that time is divided into time slots, which are the time instances at which the channels can be used (simultaneously by all the nodes). Note that relay nodes in the network may forward information received at one time slot only during the next time slot (at the earliest), hence introducing a delay in the information flow through the network.

Definition 2.5 *Let $\underline{W} \in \underline{\mathcal{W}}$ be the message at the source node S, and assume that $\underline{W}$ is uniformly distributed in the discrete, finite message alphabet $\underline{\mathcal{W}}$. Assume that time is divided into time slots, and that the transmission of a message starts at time 1. A (T_c, ϵ)-**code** is a set of possibly random functions, each assigned to a certain node $i \in \mathcal{V}$ and a certain time t. At time t, the source node uses a function that maps from $\underline{W}$ to $X_S[t]$. For node $i \in \mathcal{V}$, its function at time t maps from $(Y_i[1], \dots, Y_i[t-1])$ to $X_i[t]$. The variable $Y_i[t]$ denotes the symbol received by node i at time t, and $X_i[t]$ denotes the symbol transmitted by node i at time t. The **communication length** T_c is defined as the fixed time slot at which the destination decodes. To decode, the destination node D applies a function to $(Y_D[1], \dots, Y_D[T_c])$ in order to obtain an estimate $\underline{\hat{W}}$ of the message. We assume that after time slot T_c, all the nodes stop processing the current message, and a new run of the code can start. We require the code to be ϵ-**reliable** in the usual sense that $\mathbf{P}(\underline{W} \neq \underline{\hat{W}}) \leq \epsilon$.*

Note that in this definition, we assumed that the source node S does not receive any information from other nodes. This assumption does not reduce the generality of our results, because a source node which receives a signal can always be modeled as two nodes, one of which observes the source and is connected via a high quality link to the other node, which processes all the wireless signals.

In all of our achievability proofs, we relate a given network to a layered network through time-expansion (or unfolding, see Appendix A.12). For layered networks, we use the notion of a *block* code as defined next.

Definition 2.6 *Consider a layered network. Let a **block length** T be given, and assume that we divide time into blocks of T time slots. Let $W^{(n)} \in \mathcal{W}$ be the message transmitted by the source node S during the n^{th} block, and assume that $W^{(n)}$ is uniformly distributed in the discrete, finite message alphabet $\mathcal{W}$ for all $n \geq 1$. A (T, ϵ)-**block code** is a set of possibly random functions f_i, each assigned to a certain node $i \in \mathcal{V}$. Before transmitting the first block, the function f_S at the source node maps from $W^{(1)}$ to $\mathbf{X}_S(W^{(1)})$, which is a sequence of length T. Then S uses the first T time slots to transmit $\mathbf{X}_S(W^{(1)})$. Each relay node i in the first layer of the network waits until the end of the first block of T time slots. Then, its function f_i maps from the length-T sequence $\mathbf{Y}_i(W^{(1)})$ to $\mathbf{X}_i(W^{(1)})$. During the second block of T time-slots, node i transmits $\mathbf{X}_i(W^{(1)})$. Hence, if the destination D is in layer L of the network, then D receives a sequence $\mathbf{Y}_D(W^{(1)})$ that contains information about $W^{(1)}$ during block L. The destination then immediately declares $\hat{W}^{(1)}$, which* *is a function of $\mathbf{Y}_D(W^{(1)})$, only. We require the code to be ϵ-**reliable** in the usual sense that $\mathbf{P}(W^{(n)} \neq \hat{W}^{(n)}) \leq \epsilon$ for all $n = 1, \ldots$. It is plain to see that the source S can start transmitting $\mathbf{X}_S(W^{(2)})$ right after $\mathbf{X}_S(W^{(1)})$, i.e., sequences that carry information about subsequent messages follow each other back to back through the network. Since the network is layered, the information about different messages never gets mixed.*

Assume that we run a block code for N consecutive messages. If we define $\underline{W} \triangleq (W^{(1)}, \ldots, W^{(N)})$, then we realize that a (T, ϵ)-block code is a $(T_c, N\epsilon)$-code, where $T_c = (N + L)T$ because D declares the last message estimate $\hat{W}^{(N)}$ after $N+L$ blocks. If N grows large, the error probability $\mathbf{P}(\underline{W} \neq \underline{\hat{W}})$ is unbounded, but by information reconciliation techniques like the ones discussed in Appendix E, this problem can be overcome.

The communication throughput is measured in bits per uses of the network channels, or equivalently in bits per time slot.

Definition 2.7 *The **communication rate** of a given (T_c, ϵ)-code is defined as*

$$\frac{1}{T_c} H(\underline{W}) = \frac{1}{T_c} \log |\underline{\mathcal{W}}|,$$

where the identity is true because we assume that $\underline{W}$ is uniformly distributed in $\underline{\mathcal{W}}$. Here and throughout the thesis, logarithms are base 2.

We also refer to the communication rate as simply the “rate”. Note that for a (T, ϵ)-block code run over $N + L$ blocks (for N messages), the rate is

$$\begin{aligned} \frac{1}{(N+L)T} \log |\underline{\mathcal{W}}| &= \frac{1}{(N+L)T} \log |\mathcal{W}|^N \\ &= \frac{N}{N+L} \frac{1}{T} \log |\mathcal{W}| \\ &\to \frac{1}{T} \log |\mathcal{W}|, \end{aligned}$$

as N grows large. Hence, we can use the following asymptotically equivalent definition for a block code.

Definition 2.8 *For a (T, ϵ)-block code, we define the **communication rate** as*

$$\frac{1}{T} \log |\mathcal{W}|.$$

2.2.2 Secrecy

We assume that eavesdroppers might be present in the network.

Definition 2.9 *An **eavesdropped node** E is a regular node, whose encoding functions are under our control. However, an eavesdropper has access to all the symbols received by this particular node. We denote the set of all eavesdropped nodes by $\mathcal{E}$.*

We also refer to an eavesdropped node E as an "eavesdropper".

The aim is to keep all or part of the message W unconditionally secret from the eavesdropper. The notion of unconditional or information-theoretic secrecy is defined as follows.

Definition 2.10 *Given a (T_c, ϵ)-code, the* ***equivocation rate*** *is defined as*

$$\frac{1}{T_c} \min_{E \in \mathcal{E}} H(\underline{W} | \underline{\mathbf{Y}}_E),$$

where $\underline{\mathbf{Y}}_E = (Y_E[1], \dots, Y_E[T_c])$ is the sequence of all received symbols at E.

In this thesis, we also refer to the equivocation rate as simply the "equivocation". Again, for block codes, we can redefine the equivocation as $\frac{1}{T} \min_{E \in \mathcal{E}} H(W|\mathbf{Y}_E)$, where W stands for any message $W^{(\mathrm{n})}$ and $\mathbf{Y}_E$ is the received sequence $\mathbf{Y}_E(W^{(\mathrm{n})})$.

Now, we are ready to define the notion of achievability.

Definition 2.11 *A pair (R, R_e) is called* ***weakly achievable*** *if $R \geq R_e$ and for any $\epsilon > 0$, there exists a communication length T_c and a (T_c, ϵ)-code (or a block length T and a (T, ϵ)-block code) of rate at least $R - \epsilon$ and equivocation rate at least $R_e - \epsilon$.*

This notion is called "weak" achievability for reasons that will soon become clear.

By "perfect secrecy", we mean a situation where a rate-equivocation pair (R, R_e) such that $R = R_e$ is achievable. In most parts of this thesis, we restrict our attention to such cases. For Gaussian signal interaction and weak achievability, we use the even weaker notion of "almost perfect secrecy", where the equivocation R_e is not more than a constant ρ away from R.

Definition 2.12 *The* ***(weak) ρ-almost perfect secrecy capacity*** *C_s^ρ of a network is defined as*

$$C_s^\rho \triangleq \max\{R : (R, R - \rho) \textit{ is weakly achievable}\}.$$

The ***weak perfect secrecy capacity*** *C_s is defined as*

$$C_s = C_s^0.$$

The weakness of this definition is the following. Assume that $C_s \geq R$ for some $R > 0$. Then, from Definition 2.12, we know that there exists a (T_c, ϵ)-code that is such that $\frac{1}{T_c} H(\underline{W}) \geq R - \epsilon$ and for every eavesdropper $E \in \mathcal{E}$,

$$\frac{1}{T_c} H(\underline{W} | \underline{\mathbf{Y}}_E) \geq \frac{1}{T_c} H(\underline{W}) - \epsilon.$$

It follows that $H(\underline{W} | \underline{\mathbf{Y}}_E) \geq H(\underline{W}) - T_c\ \epsilon$, *i.e.*, as the communication length T_c grows, the number of bits that we leak to an eavesdropper is not bounded. The following definitions avoid this problem.

Definition 2.13 *A secrecy rate R_s is called* ***strongly achievable*** *if for any $\epsilon > 0$, there exists a communication length T_c and a (T_c, ϵ)-code (or a block length T and a (T, ϵ)-block code) of rate at least $R_s - \epsilon$ and equivocation rate at least $\frac{1}{T_c}(H(\underline{W}) - \epsilon)$ (or $\frac{1}{T}(H(W) - \epsilon)$ for a block code).*

Note that in Definition 2.13, we require that for any eavesdropper E, $H(\underline{W}|\underline{\mathbf{Y}}_E) \geq H(\underline{W}) - \epsilon$. Hence, the number of bits leaked to any eavesdropper is bounded by ϵ, no matter how large T_c gets.

Definition 2.14 *The* ***strong perfect secrecy capacity*** *$\bar{C}_s$ of a network is defined as*

$$\bar{C}_s \triangleq \max\{R_s : R_s \textit{ is strongly achievable}\}.$$

Several of our results are lower bounds on $\bar{C}_s$ for different networks. If for some R_s, we can show that $\bar{C}_s \geq R_s$, then this means the following. We can guarantee that there exists a code that reliably communicates a message whose rate can be made arbitrarily close to R_s by choosing T_c large enough. In addition, by choosing T_c large enough, $H(\underline{W}|\underline{\mathbf{Y}}_E)$ can be made arbitrarily close to $H(\underline{W})$ for all eavesdroppers $E \in \mathcal{E}$, hence making sure that an arbitrarily small number of bits is leaked to each of the eavesdroppers. This is in contrast to weak secrecy, where an arbitrarily small number of bits *per channel use* is leaked.

If we want to extend the notion of strong secrecy to the case where only part of the message is kept secret, we need to explicitly consider two messages, one that is kept secret and one that is given away.

Definition 2.15 *A pair (R, R_s) is called* ***strongly achievable*** *if $R \geq R_s$ and for any $\epsilon > 0$, there exists a block length T_c and a (T_c, ϵ)-code for two messages $(\underline{W}, \underline{W}_s)$ such that*

$$\frac{1}{T_c} H(\underline{W}, \underline{W}_s) \geq R - \epsilon,$$

$$\frac{1}{T_c} H(\underline{W}_s) \geq R_s - \epsilon,$$

and for any $E \in \mathcal{E}$,

$$\frac{1}{T_c} H(\underline{W}_s|\underline{\mathbf{Y}}_E) \geq \frac{1}{T_c}(H(\underline{W}_s) - \epsilon).$$

2.2.3 Signal Interaction Models

In some of our results, we restrict our attention to one of the following two signal interaction models (channel models).

In the well-accepted Gaussian interaction model [45], transmitted signals get attenuated by (complex) gains to which independent (Gaussian) receiver noise is added. More formally, we have the following definition:

Definition 2.16 *In* ***Gaussian signal interaction****, the received signal Y_j at node $j \in \mathcal{V}$ at time t is given by,*

$$Y_j[t] = \sum_{i \in In(j)} h_{ij} X_i[t] + Z_j[t], \tag{2.1}$$

where h_{ij} is the complex channel gain between node i and j, $X_i[t]$ is the signal transmitted by node i at time t, and $In(j)$ is the set of nodes that have non-zero channel gains to j. We assume that the average transmit power constraints for all nodes is 1 and the additive receiver Gaussian noise $Z_j[t]$ is of zero mean and unit variance and independent over j and t.

Note that $X_i[t]$ is a random variable because the encoder at node i is a possibly random mapping from random information to $X_i[t]$.

In [3], the following deterministic model for wireless interaction was introduced.

Definition 2.17 *In* ***deterministic signal interaction****, the received signal $Y_j[t]$ at node $j \in \mathcal{V}$ at time t is given by*

$$Y_j[t] = g_j(\{X_i[t]\}_{i \in In(j)}), \tag{2.2}$$

where $In(j)$ is the set of input neighbours of the node j, and g_j is a deterministic function.

In [2], a simpler deterministic model which captures the essence of wireless interaction was developed, called linear deterministic signal interaction. The advantage of this model is its simplicity, which gives insight to strategies for the noisy wireless network model in Definition 2.16. The linear deterministic model simplifies the Gaussian interaction model in (2.1) by eliminating the noise and modeling the symbols and channel gains by a binary vector of dimension q which somewhat resembles a binary expansion over q bits of real numbers. More precisely, the received signal $Y_j[t]$ is modeled as follows.

Definition 2.18 *In* ***linear deterministic signal interaction****, the received signal $Y_j[t]$ at node $j \in \mathcal{V}$ at time t is given by*

$$Y_j[t] = \sum_{i \in In(j)} \mathbf{G}_{i,j} X_i[t], \tag{2.3}$$

where $X_i[t]$ and $Y_j[t]$ are binary vectors of dimension q and $\mathbf{G}_{i,j}$ is a $q \times q$ binary matrix representing the (discretized) channel transformation between nodes i and j. All operations in (2.3) are done over the binary field.

Note that this model is *not* equivalent to a discretization of the Gaussian signal interaction model, because the “carry-over” bits are ignored when computing additions.

An illustration of this deterministic model is given in Figure 2.1 for the broadcast and multiple access networks. The left illustration in Figure 2.1

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 linear deterministic model 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 on the right hand side of Figure 2.1 adds one more ingredient, which is how the bits from two transmitting nodes interact at a receiver. 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 [2], that this model *approximately*[1] captures the Gaussian interaction model of Definition 2.16 for the broadcast and multiple access channels. For general networks the linear deterministic model yields insights which, when translated to the noisy wireless network, lead one to develop cooperative communication strategies (without secrecy constraints) for the model in Definition 2.16, which are approximately optimal [4].

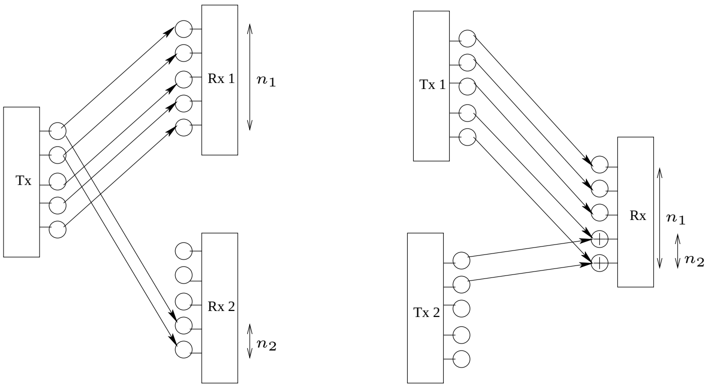

Figure 2.1: The linear deterministic model for a Gaussian broadcast channel (BC) is shown on the left and for a Gaussian multiple access channel (MAC) is shown on the right.

2.2.4 Information-Theoretic Min-Cut

Several results of this thesis are expressed in terms of information-theoretic min-cuts.

Definition 2.19 *For $\mathcal{I} \subseteq \mathcal{V}$ and $j \in \mathcal{V}$, define $\Lambda(\mathcal{I}; j)$ to be the set of all cuts (Ω, Ω^c) that separate $\mathcal{I}$ from j. More precisely, $\Lambda(\mathcal{I}; j)$ is the set of all $\Omega \subset \mathcal{V}$ such that $\mathcal{I} \subseteq \Omega$ and $j \in \Omega^c$.*

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

Definition 2.20 *Consider a wireless relay network with discrete signal alphabets. For a given transmit distribution $p(\{x_i\}_{i\in\mathcal{V}})$, for given $\mathcal{I}\subseteq\mathcal{V}$ and $j\in\mathcal{V}$, we define the* ***information-theoretic min-cut*** *between $\mathcal{I}$ and j as*

$$R_{\mathcal{I};j}(p) \triangleq \min_{\Omega\in\Lambda(\mathcal{I};j)} I(X_\Omega; Y_{\Omega^c}|X_{\Omega^c}), \tag{2.4}$$

where $X_\mathcal{V}$ is governed by p and $Y_\mathcal{V}$ is the tuple corresponding to the channel outputs.

Definition 2.21 *Consider a wireless relay network with complex signal alphabets and Gaussian signal interaction as defined in Definition 2.16. Let the transmit distribution $\prod_{i\in\mathcal{V}} p^G(x_i)$ be the i.i.d. Gaussian distribution on $\mathbb{C}^{|\mathcal{V}|}$ such that $X_i \sim \mathbb{C}\mathcal{N}(0,1)$. For given $\mathcal{I}\subseteq\mathcal{V}$ and $j\in\mathcal{V}$, we define*

$$R^G_{\mathcal{I};j} \triangleq \min_{\Omega\in\Lambda(\mathcal{I};j)} I(X_\Omega; \hat{Y}_{\Omega^c}|X_{\Omega^c}), \tag{2.5}$$

where for every $i\in\mathcal{V}$, we define

$$\hat{Y}_i \triangleq \lambda_i Y_i + \xi_i,$$

and the constant λ_i and the random variable ξ_i are defined as follows. Define $\lambda_i = \frac{\sigma^2_{Y_i}-1}{\sigma^2_{Y_i}}$, where $\sigma^2_{Y_i}$ is the variance of the Gaussian random variable Y_i. Note that $\sigma^2_{Y_i}\geq 1$ because Y_i is defined as Z_i plus other independent random variables and Z_i has unit variance. Define $\xi_i \sim \mathbb{C}\mathcal{N}(0,\sigma^2_{\xi_i})$, independent of all other quantities, with $\sigma^2_{\xi_i} = (1-\lambda_i^2)\sigma^2_{Y_i} - 1$.

2.3 Preliminaries

In this section, we state several preliminary results on the unicast capacity of wireless relay networks. Secrecy is no concern in these results.

Definition 2.22 *Let a wireless relay network as defined in Definition 2.2 be given. The* ***capacity*** *C of the network is the largest number R such that for any $\epsilon > 0$, there exists a (T_c,ϵ)-code of rate at least $R-\epsilon$.*

In [9], Cover and Thomas describe a well known upper bound on the capacity in arbitrary wireless relay networks, called the cut-set bound. We repeat the result here.

Theorem 2.1 *For an arbitrary wireless relay network as defined in Definition 2.2 the capacity is upper bounded as*

$$C < \bar{C} \triangleq \max_{p(\{x_i\},i\in\mathcal{V})} R_{S;D}(p),$$

where $R_{S;D}(p)$ is defined in Definition 2.20.

When the signal alphabets are real, the theorem still holds, with p denoting a probability density function.

In their recent work, Avestimehr, Diggavi and Tse took a new approach to the study of the capacity of wireless networks. First, they used deterministic signal interaction (Definition 2.17) to model broadcast and interference in wireless networks, without being concerned about noise. They found the following lower bound on the capacity for arbitrary deterministic network topologies [3].

Theorem 2.2 *For an arbitrary wireless relay network with deterministic signal interaction (Definition 2.17), the capacity is lower bounded as*

$$C > \max_{\prod_{i\in\mathcal{V}} p(x_i)} R_{S;D}(p),$$

where $R_{S;D}(p)$ is defined in Definition 2.20.

Note that this bound is in general different from the cut-set upper bound, because the maximization is only over product distributions on the transmit alphabets. When the signal interaction is linear deterministic (Definition 2.18), this lower bound matches the cut-set upper bound, providing the exact capacity.

Theorem 2.3 *For an arbitrary wireless relay network with linear deterministic signal interaction (Definition 2.18), the capacity is given by*

$$\begin{aligned} C &= R_{S;D}(\tilde{p}), \\ &= \min_{\Omega\in\Lambda(S;D)} \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}), \end{aligned}$$

where $\tilde{p}$ is the i.i.d. uniform distribution on $\mathcal{X}_{\mathcal{V}}$.

The matrix $\mathbf{G}_{\Omega,\Omega^c}$ is a block matrix consisting of $|\Omega^c| \times |\Omega|$ blocks, each of dimension $q \times q$. Block (i, j) of this block matrix is equal to the channel transfer matrix $\mathbf{G}_{j,i}$, where $j \in \Omega$ and $i \in \Omega^c$. Theorems 2.2 and 2.3 can be found in [3] and [5].

The intuitions gained from the deterministic models were then applied to Gaussian signal interaction, leading to the following result.

Proposition 2.1 *For an acyclic wireless relay network with Gaussian signal interaction (Definition 2.16), the capacity is lower bounded as*

$$C > R^G_{S;D} - \beta,$$

where $R^G_{S;D}$ is defined in Definition 2.21, and $\beta = \mathcal{O}\left(d|\mathcal{V}|^2\right)$ depends only on the number of nodes and on the maximum out-degree d of the nodes in the network.

This result was shown in [4] and [5], together with the following theorem, that provides an approximate characterization of the capacity for Gaussian relay networks.

Theorem 2.4 *For an acyclic wireless relay network with Gaussian signal interaction (Definition 2.16), we have*

$$\bar{C} - \kappa \leq C \leq \bar{C},$$

where $\bar{C} = \max_{\prod_{i \in \mathcal{V}} p(x_i)} R_{S;D}(p)$ *is the cut-set upper bound, and* $\kappa = \alpha + \beta$*, with* $\alpha = \mathcal{O}(|\mathcal{V}|)$ *and* $\beta = \mathcal{O}\left(d|\mathcal{V}|^2\right)$.

The power of this approximate description is that the gap κ between upper and lower bound does not depend on the signal-to-noise ratio (SNR). The cut-set bound $\bar{C}$ grows logarithmically with the SNR, and hence, the importance of the gap decreases in the high SNR regime. Note that the gap characterization could be further sharpened. This yields a *universal* approximation which is valid for arbitrary wireless networks.

Several of our results rely on the techniques introduced by Avestimehr, Diggavi and Tse. In particular, we use some of the techniques from [5] to prove multisource results given in Chapter 3.

3 Secrecy for Arbitrary Wireless Networks

3.1 Introduction

In this chapter, we consider information-theoretically secret communication for arbitrary wireless relay networks. We assume that eavesdroppers are passive and that all the relays are authenticated.

Information-theoretic secrecy for wireless networks has been introduced by Wyner in his work on the wiretap channel [48], which is essentially a discrete memoryless point-to-point channel with one eavesdropper. Wyner found the (weak) perfect secrecy capacity for this setup under the assumption that the eavesdropper's observation is a noisy version of the signal observed at the destination. Csiszár and Körner generalized this to the case where the signals at the eavesdropper and the destination are obtained from the transmitted signal through an arbitrary broadcast channel [10].

There are three main issues that limit the practical use of the wiretap channel results: First, the assumption that the channel to the eavesdropper is known is not realistic, because it would imply knowing the approximate location and hardware of the eavesdropper. Second, even if we assume that the physical (wireless) channel to the eavesdropper is known, it is unrealistic to assume that she would base her estimate of the message on a discretized and memoryless version of the channel output. Third, the secrecy guaranteed by [48, 10] is weak in the sense that the number of information bits leaked to the eavesdropper might grow with the block length T (see Section 2.2 for a precise definition). The wiretap channel results have been strengthened in the above three respects by several different contributions.

One possible improvement regarding the first issue (eavesdropper channel uncertainty) is to consider a class of possible eavesdropper channels. If the class, albeit finite, is sufficiently large, it can provide a reasonable approximation for a continuous range of the true eavesdropper channel. The compound wiretap channel studied by Liang, Kramer, Poor and Shamai in [22] shows that the perfect secrecy capacity can be lower bounded for the wiretap channel with a class of eavesdroppers. For degraded compound wiretap channels, the lower bound given in [22] is tight. This work was extended by Liu, Prabhakaran and Vishwanath in [26], where the secrecy capacity was found for a class of non-degraded parallel Gaussian compound channels.

Let us explain the second issue in more detail. In wireless communication, an electromagnetic waveform is produced by a transmit antenna and captured by a receive antenna elsewhere. A discrete memoryless channel like the binary symmetric channel models the transition from a sequence of transmit bits to a sequence of received bits, hiding all the intermediate phases of encoding into constellation points, modulating to a waveform, transmission, signal attenuation, noisy reception, and finally demodulation and decoding. While such a discrete channel might be a good model for the received symbols at the destination, there is no reason to assume that the eavesdropper follows the same demodulation and decoding steps. However, one can show that for common modulation schemes (e.g. pulse-amplitude modulation on shift-orthogonal waveforms) a projection of the received waveform using a matched filter yields a sufficient statistics. Hence, it is reasonable to assume that the eavesdropper will perform such a matched filter operation to transform the continuous-time waveform into a discrete-time sequence of real or complex numbers. It follows that the Gaussian signal interaction model (see Definition 2.16) is actually a reasonable model when dealing with an eavesdropper in wireless communication. The assumption that the channel to the eavesdropper is memoryless is a standard assumption in all previous work on information-theoretic secrecy, and we use it for the sake of tractability. Several Gaussian (multi-antenna) versions of the wiretap channel problem have recently been solved by Khisti, Wornell, Wiesel and Eldar [19, 18, 17], by Oggier and Hassibi [33] and by Liu and Shamai [27], respectively.

The third issue (weak secrecy) has been resolved by the work of Maurer and Wolf [29] for the case when the source and the destination wish to generate a secret key. Indeed, by using extractors, it is possible to transform a weak secret key into a strong secret one without sacrificing any rate. In Appendix E of this thesis, we make the additional observation that if the extractor is invertible, then the same technique provides strongly secret communication. (The same idea is also used by Cheraghchi, Didier and Shokrollahi in [7] to provide strong secrecy in linear wiretap systems.)

Recently, a considerable effort has been made to find upper and lower bounds on the perfect secrecy capacity for wireless networks with more complex topologies than the wiretap channel [21, 43, 23, 25, 34]. The results suggest that obtaining a characterization of the perfect secrecy capacity for the case when relays are present in the network is a difficult task.

In this chapter, we consider secret communication over arbitrary wireless networks. We address all three issues that we explained above. More precisely, we find bounds on the weak *and* strong perfect secrecy capacity when a *class* of eavesdroppers is present and signal interaction is *Gaussian* (but memoryless). Our results build upon the recent work by Avestimehr, Diggavi and Tse that we summarized in Section 2.3. We first provide a lower bound on the perfect secrecy capacity for arbitrary networks with deterministic signal interaction, and a similar bound for arbitrary acyclic networks with Gaussian signal interaction. As explained above, the deterministic interaction model is not directly applicable in practice. However, we show through examples that our deterministic result, especially when the signal interaction is linear, can provide valuable insights for the design of codes in general. We also provide an upper bound on the perfect secrecy capacity that is valid for both types of signal interaction.

A second contribution of this chapter is a study of the so-called multisource problem in wireless relay networks, where several source nodes observe independent data sources and wish to communicate them reliably to the same destination without secrecy requirements. We first derive a cut-set type outer bound on the multisource capacity region for arbitrary networks and any type of signal interaction. For deterministic signal interaction, we find an inner bound on the capacity region for arbitrary networks, and we show that it matches the outer bound for linear deterministic signal interaction. In the Gaussian case, we find an inner bound on the capacity region for acyclic networks, which yields an approximation of the true capacity region in the sense that the gap between inner and outer bound is bounded by a constant number of bits. Like in the results by Avestimehr, Diggavi and Tse (Section 2.3), the constant depends on the number of nodes in the network, but not on the signal-to-noise ratio. The achievability parts of these multisource results play an important role in the proof of the lower bounds on the perfect secrecy capacity.

The chapter is organized as follows. Since most of the definitions have already been presented in the previous chapter, the problem statement in Section 3.2 is short. We present the lower and upper bounds on the perfect secrecy capacity in Section 3.3. Section 3.4 presents several linear deterministic example networks, as well as a small Gaussian network called the diamond network. These examples illustrate different secret communication schemes. Section 3.5 contains our results for the multisource problem. The corresponding proofs are given in Appendix A. In Section 3.6, we give proof outlines of the main secrecy results, which rely on the multisource achievability.

3.2 Problem Statement

We consider transmission over a wireless relay network $\mathcal{G} = (\mathcal{V}, \mathcal{L})$ as described in Definition 2.2. We do not impose any constraints on the topology of the network.

Let $\mathcal{B} \subseteq \mathcal{A}$ be the set of all noise-inserting nodes, *i.e.*, the relay nodes that are allowed to insert noise into their transmission by using a random encoding function. All the relay nodes in $\mathcal{A} \backslash \mathcal{B}$ are required to use deterministic encoding functions. Without loss of generality we assume that this is done by generating a random message of arbitrary rate that is used, together with the past received symbols, as an input of a deterministic encoding function. These noise messages need not be decoded by the destination D.

The secrecy is with respect to a set of possible passive eavesdropper nodes $\mathcal{E} \subset \mathcal{V}$ where $\mathcal{E}$ is disjoint from $\{S, D\}$. We want to keep all or part of the message secret if any one of the possible eavesdropper nodes $E \in \mathcal{E}$ listens to the wireless transmissions in the relay network[1]. Note that the class of eavesdroppers that we define is finite, *i.e.*, we assume that all possible eavesdroppers and their channels can be enumerated. If there is a continuum of possible eavesdropper channels, our model can approximate this via "quantization" of this continuum. Furthermore, it might be possible to replace our proof technique by the "enhancement" of the weaker eavesdroppers, as it was for instance done in [22]. Using this technique, one would be able to show the same result for an uncountable class of eavesdroppers.

We focus on a wireless relay network with the deterministic and Gaussian signal interaction models given in Definitions 2.16 and 2.17. We want to ensure that we can communicate reliably *and* secretly between S and D. These notions have been defined in Chapter 2. The quantity of interest is the *strong* perfect secrecy capacity as defined in Definition 2.14.

Definition 3.1 *Throughout this chapter, we use the* ***"constants"*** α*,* β *and* γ *that depend only on the network topology, but not on the signal to noise ration (SNR) or the channel gains. More precisely, we have* $\alpha = \mathcal{O}\left(|\mathcal{V}|\right)$*,* $\beta = \mathcal{O}\left(d|\mathcal{V}|^2\right)$ *and* $\gamma = \mathcal{O}\left(d|\mathcal{V}|\right)$*, where* d *is the maximum out-degree of the nodes in* $\mathcal{G}$*.*

3.3 Main Results

In this section, we present the main results of this chapter. The proofs are given in Section 3.6.

3.3.1 Achievable Secrecy Rates for Deterministic Networks

We start with several definitions.

Definition 3.2 *For a given transmit distribution* $p(\{x_i\}_{i\in\mathcal{V}})$*, a subset* $\psi \subseteq \mathcal{V}$*, a node* $j \in \mathcal{E} \cup \{D\}$*, define* $\mathcal{R}_{\psi;j}(p)$ *to be the set of all tuples* $B_\psi = (B_i)_{i\in\psi}$ *such that the components of the tuple are non-negative and such that for any subset* $\mathcal{I} \subseteq \psi$*,*

$$\sum_{i\in\mathcal{I}} B_i \leq R_{\mathcal{I};j}(p).$$

The quantity $R_{\mathcal{I};j}(p)$ is the information-theoretic min-cut defined in Definition 2.20.

[1] Our results apply also when *all* eavesdroppers in $\mathcal{E}$ are present, but cannot collude.

Definition 3.3 *For a given input distribution $p(\{x_i\}_{i\in\mathcal{V}})$, a subset $\psi \subseteq \mathcal{V} \setminus \{S\}$, and a node $j \in \mathcal{E}\cup\{D\}$, define $\tilde{\mathcal{R}}_{\psi;j}(p)$ to be the set of all tuples $(B', B_\psi) = (B', (B_i)_{i\in\psi})$ such that the components of the tuple are non-negative and such that for any subset $\mathcal{I} \subseteq \psi$,*

$$B' + \sum_{i\in\mathcal{I}} B_i \leq R_{\mathcal{I}\cup\{S\};j}(p).$$

Note that for a given $\psi \subseteq \mathcal{V} \setminus \{S\}$, $\tilde{\mathcal{R}}_{\psi;j}(p)$ differs from $\mathcal{R}_{\psi\cup\{S\};j}(p)$ in the fact that $\mathcal{R}_{\psi\cup\{S\};j}(p)$ imposes constraints for subsets $\mathcal{I}$ that do not contain S. In Definition 3.3, all the constraints involve S. Recall that $\mathcal{B}$ is the set of all noise-inserting relays. We find that the nodes in $\mathcal{B}$ should be careful when inserting noise. In particular, we ensure that all noise messages are decoded by the eavesdroppers, but not necessarily by D. This is the reason for the use of the two regions $\mathcal{R}_{\psi\cup\{S\};j}(p)$ and $\tilde{\mathcal{R}}_{\psi;j}(p)$.

Definition 3.4 *For an input distribution $p(\{x_i\}_{i\in\mathcal{V}})$, we define the following function:*

$$\begin{aligned} F(p) = \max_{B_\mathcal{B}\in\cap_{E\in\mathcal{E}}\mathcal{R}_{\mathcal{B};E}(p)} \Big[& \\ & \max_x\{x : (x, B_\mathcal{B}) \in \tilde{\mathcal{R}}_{\mathcal{B};D}(p)\} \\ & - \max_x\{x : (x, B_\mathcal{B}) \in \cup_{E\in\mathcal{E}}\mathcal{R}_{\mathcal{B}\cup\{S\};E}(p)\}\Big]. \end{aligned}$$

Theorem 3.1 *The strong perfect secrecy capacity for any wireless network with deterministic signal interaction (Definition 2.17) is lower bounded as*

$$\bar{C}_s \geq \max_{\prod_{i\in\mathcal{V}} p(x_i)} F(p).$$

The proof is provided in Section 3.6. This theorem guarantees a certain strong perfect secrecy rate for arbitrary networks, expressed as a maximization problem over a finite-dimensional solution space.

We illustrate the expressions of Theorem 3.1 using an example. Assume that there is only one noise-inserting node and one eavesdropper and denote them by $\mathcal{B} = \{A\}$ and $\mathcal{E} = \{E\}$. The tuple $B_\mathcal{B}$ in Theorem 3.1 corresponds to the rates of the noise-messages inserted by the nodes in $\mathcal{B}$. In this example, there is only one such rate B_A. For a given noise rate B_A, define

$$\phi_D \triangleq \max_x\{x : (x, B_\mathcal{B}) \in \tilde{\mathcal{R}}_{\mathcal{B};D}(p)\}$$

and

$$\phi_E \triangleq \max_x\{x : (x, B_\mathcal{B}) \in \cup_{E\in\mathcal{E}}\mathcal{R}_{\mathcal{B}\cup\{S\};E}(p)\}.$$

Note that ϕ_D and ϕ_E are functions of $B_\mathcal{B} = B_A$. In this example, the regions $\tilde{\mathcal{R}}_{\mathcal{B};D}(p)$ and $\mathcal{R}_{\mathcal{B}\cup\{S\};E}(p)$ are of dimension 2. They are both depicted in Figure

3.1. The noise rate B_A can take values in $\mathcal{R}_{\mathcal{B};E}(p)$, which is equal to the interval $[0, R_{A;E}(p)]$ on the A-axis. For any fixed B_A, the strong perfect secrecy rate $\phi_D - \phi_E$ is achievable, where ϕ_D and ϕ_E are shown in Figure 3.2. Intuitively, the proof goes as follows. The source node S observes the information message W of rate roughly equal to $\phi_D - \phi_E$. In addition, it generates a random uniform junk message J of rate roughly equal to ϕ_E. The noise inserting node A generates a random uniform noise message J_A of rate B_A. Via a random code construction for this set of messages, we can show that there exists a code which has the following property: if a genie made the information message W available to the eavesdropper E, it would be able to find the correct junk (noise) messages (J, J_A) with high probability. Intuitively, this implies that before being able to decode W, the eavesdropper must decode (J, J_A). However, the pair of junk rates (ϕ_E, B_A) lies on the boundary of $\mathcal{R}_{\mathcal{B}\cup\{S\};E}(p)$. We prove a "local" converse that states that for a fixed transmit distribution p and the randomly generated code, the total information transfer from S to E cannot exceed ϕ_E. It follows that the only information that E can gather is (J, J_A), and hence, W remains secret. Finally, we maximize the quantity $\phi_D - \phi_E$ over all possible values of B_A to find the optimal amount of noise that A should insert into the network. The secrecy that we show in this manner is weak. We then use the techniques introduced by Maurer and Wolf in [29] to show the existence of strong perfect secrecy codes with the same rate. A formal proof for several noise-inserting nodes and two eavesdroppers is given in Section 3.6. Note that for the wiretap channel (a network without relays), the use of a junk message at the source node is equivalent to the binning strategy proposed by Wyner in [48]. However, it is unlikely that a binning strategy can be used to prove Theorem 3.1.

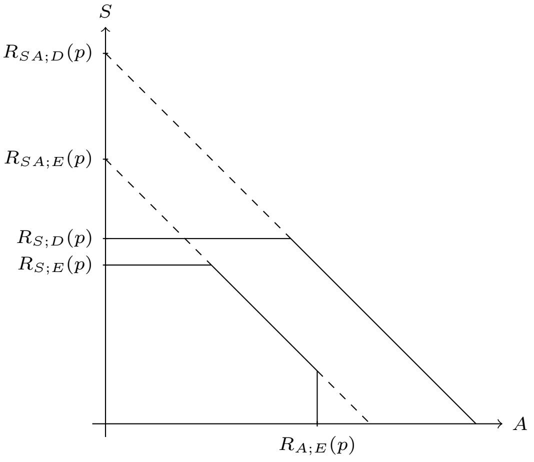

Figure 3.1: Regions for the example with one noise-inserter and one eavesdropper. The larger quadrilateral region is $\tilde{\mathcal{R}}_{\mathcal{B};D}(p)$, while the smaller pentagonal region is $\mathcal{R}_{\mathcal{B}\cup\{S\};E}(p)$. The region $\mathcal{R}_{\mathcal{B};E}(p)$ is the line interval $[0, R_{A;E}(p)]$ on the A-axis.

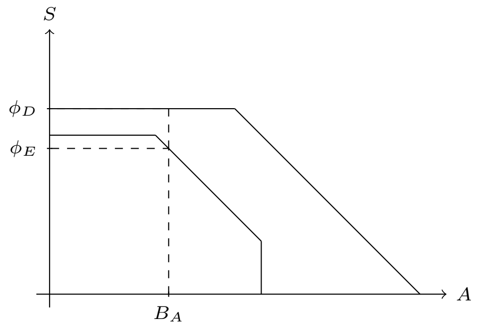

Figure 3.2: The quantities ϕ_D and ϕ_E for a particular noise rate B_A.

As one can see in Figure 3.1, $\mathcal{R}_{\mathcal{B}\cup\{S\};E}(p)$ has a similar shape as the capacity region of a multiple access channel (MAC). The multiple access region is a polymatroid (see e.g. [46]), which means that it is defined by constraints that involve a so-called rank function. The following lemma implies that $\mathcal{R}_{\mathcal{B}\cup\{S\};E}(p)$ is a polymatroid for any number of noise inserters.

Lemma 3.1 *For a network with deterministic signal interaction and for a fixed product distribution $\prod_{i\in\mathcal{V}} p(x_i)$, the function $R_{\mathcal{I},j}(p)$ is a rank function with respect to its first index. This means that for any $j \in \mathcal{E} \cup \{D\}$, we have the following properties:*
monotonicity: for any two subsets $\mathcal{A}$ and $\mathcal{B}$ of $\mathcal{V}$ such that $\mathcal{A} \subseteq \mathcal{B}$,

$$R_{\mathcal{A};j}(p) \leq R_{\mathcal{B};j}(p)$$

submodularity: for any two subsets $\mathcal{A}$ and $\mathcal{B}$ of $\mathcal{V}$,

$$R_{\mathcal{A}\cup\mathcal{B};j}(p) + R_{\mathcal{A}\cap\mathcal{B};j}(p) \leq R_{\mathcal{A};j}(p) + R_{\mathcal{B};j}(p).$$

For networks with random signal interaction, the lemma is only true for sets of cardinality at most 2. As a side-remark, we state the following lemma, which implies that $\mathcal{R}_{\mathcal{B}\cup\{S\};E}(p)$ is a so-called weak polymatroid for random signal interaction and for any number of noise inserters, meaning that it is defined by weak rank functions as stated in the lemma.

Lemma 3.2 *For a fixed product distribution $\prod_{i\in\mathcal{V}} p(x_i)$, the function $R_{\mathcal{I},j}(p)$ is a weak rank function with respect to its first index. This means that for any $j \in \mathcal{E} \cup \{D\}$, we have the following properties:*
monotonicity: for any two subsets $\mathcal{A}$ and $\mathcal{B}$ of $\mathcal{V}$ such that $\mathcal{A} \subseteq \mathcal{B}$,

$$R_{\mathcal{A};j}(p) \leq R_{\mathcal{B};j}(p)$$

weak submodularity: for any two disjoint *subsets $\mathcal{A}$ and $\mathcal{B}$ of $\mathcal{V}$,*

$$R_{\mathcal{A}\cup\mathcal{B};j}(p) \leq R_{\mathcal{A};j}(p) + R_{\mathcal{B};j}(p).$$

The notion of a weak rank function defined in this lemma differs from the usual (stronger) definition in that the submodularity is only required for disjoint sets. The proof of Lemmas 3.1 and 3.2 can be found in Appendix A. A set of tuples that is defined by rank function constraints on all sums of components is called a polymatroid. This is the case for $\mathcal{R}_{\mathcal{B}\cup\{S\};E}(p)$. The set $\tilde{\mathcal{R}}_{\mathcal{B};D}(p)$ is a polymatroid where the constraints on $B_{\mathcal{B}}$ have been dropped. Knowing this, we can inspect all possible ways the regions $\mathcal{R}_{\mathcal{B}\cup\{S\};E}(p)$ and $\tilde{\mathcal{R}}_{\mathcal{B};D}(p)$ can intersect in our two-dimensional example. By doing so, we realize that the maximum of $\phi_D - \phi_E$ over B_A occurs either at $B_A = 0$ or at $B_A = R_{A;E}(p)$. Hence, we obtain the following corollary of Theorem 3.1 for our example.

Corollary 3.1 *When $\mathcal{B} = \{A\}$ and $\mathcal{E} = \{E\}$, we have the following lower bound on the strong perfect secrecy capacity for deterministic signal interaction.*

$$\begin{aligned}\bar{C}_s \geq \max_{\prod_{i\in\mathcal{V}} p(x_i)} \max\Big\{&R_{S;D}(p) - R_{S;E}(p),\\ &\min\{R_{SA;D}(p) - R_{A;E}(p), R_{S;D}(p)\}\\ &\quad - (R_{SA;E}(p) - R_{A;E}(p))\Big\}.\end{aligned}$$

Using the polymatroidal structure of $\mathcal{R}_{\mathcal{B}\cup\{S\};E}(p)$ and $\tilde{\mathcal{R}}_{\mathcal{B};D}(p)$, a similar simplification can be found for any size of $\mathcal{B}$, provided that there is only one eavesdropper.

Note that the expressions in Theorem 3.1 are in terms of the transmit variables $X_{\mathcal{V}}$ themselves (see Definition 2.20). It would be desireable to use a set of auxiliary random variables instead, as it was done in [10] and in most of the related work. For the wiretap channel, it turns out that the use of auxiliary random variables strictly increases the achievable secrecy rate. In our relay network case, the use of such auxiliary random variables would unfortunately make our analysis impossible to carry out, since it relies on the conditional independence of the channels in the network.

3.3.2 Achievable Secrecy Rates for Gaussian Networks

We start with similar definitions as in the deterministic case.

Definition 3.5 *For a given subset $\psi \subseteq \mathcal{V}$, a node $j \in \mathcal{E} \cup \{D\}$, and any $\Delta > 0$, define $\mathcal{R}^G_{\psi;j}(\Delta)$ to be the set of all tuples $B_\psi = (B_i)_{i\in\psi}$ such that the components of the tuple are non-negative and such that for any subset $\mathcal{I} \subseteq \psi$,*

$$\sum_{i\in\mathcal{I}} B_i \leq R^G_{\mathcal{I};j} - \Delta.$$

The quantity $R^G_{\mathcal{I};j}$ is the information-theoretic min-cut defined in Definition 2.21.

Definition 3.6 *For a given subset $\psi \subseteq \mathcal{V} \setminus \{S\}$, a node $j \in \mathcal{E} \cup \{D\}$, and any $\Delta > 0$, define $\tilde{\mathcal{R}}^G_{\psi;j}(\Delta)$ to be the set of all tuples $(B', B_\psi) = (B', (B_i)_{i\in\psi})$ such* *that the components of the tuple are non-negative and such that for any subset $\mathcal{I} \subseteq \psi$,*

$$B' + \sum_{i \in \mathcal{I}} B_i \leq R^G_{\mathcal{I} \cup \{S\};j} - \Delta.$$

As in the deterministic case, $\tilde{\mathcal{R}}^{\mathrm{G}}_{\psi;j}(\Delta)$ differs from $\mathcal{R}^{\mathrm{G}}_{\psi \cup \{S\};j}(\Delta)$ in the fact that $\mathcal{R}^{\mathrm{G}}_{\psi \cup \{S\};j}(\Delta)$ as given in Definition 3.5 imposes constraints for subsets $\mathcal{I}$ that do not contain S, whereas Definition 3.6 only uses constraints that involve S.

Definition 3.7 *For any $\Delta > 0$, we define the following function:*

$$\begin{aligned} F^G(\Delta) = \max_{B_{\mathcal{B}} \in \cap_{E \in \mathcal{E}} \mathcal{R}^G_{\mathcal{B};E}(\Delta)} \Big[& \\ & \max_x \{x : (x, B_{\mathcal{B}}) \in \tilde{\mathcal{R}}^G_{\mathcal{B};D}(\Delta)\} \\ & - \max_x \{x : (x, B_{\mathcal{B}}) \in \cup_{E \in \mathcal{E}} \mathcal{R}^G_{\mathcal{B} \cup \{S\};E}(\Delta)\} \Big]. \end{aligned}$$

Theorem 3.2 *The strong perfect secrecy capacity for acyclic wireless networks with Gaussian signal interaction (Definition 2.16) is lower bounded as*

$$\bar{C}_s \geq F^G(\beta) - \alpha - \beta - \gamma,$$

where α, β and γ are constants that depend on the network topology, but not on the signal-to-noise ratio (SNR) of the channels (see Definition 3.1).

The proof is provided in Section 3.6.

It can be shown that the Gaussian min-cuts $R^{\mathrm{G}}_{\mathcal{I};j}$ in Definition 2.21 scale logarithmically with the signal-to-noise ratio (SNR). It follows that in many cases, $F^{\mathrm{G}}(\beta)$ also shows logarithmic growth in the SNR. In such cases, because of the subtractive constants α, β and γ, Theorem 3.2 provides its most useful secrecy rate guarantee for the high SNR regime. The theorem should be viewed as a "proof of concept", demonstrating that it is indeed possible to lower bound the perfect secrecy rate by a computable expression that applies to any network topology. We conjecture that the presence of non-zero α, β, γ is not due to the nature of the wireless network, but only to requirements of our analysis. It may be possible to reduce these constants with a more refined analysis.

3.3.3 Upper Bound on the Perfect Secrecy Capacity

Our third result is a simple upper bound on the perfect secrecy capacity for a wireless relay network with arbitrary memoryless signal interaction channels.

Theorem 3.3

$$C_s \leq \max_{p(\{x_i\}_{i \in \mathcal{V}})} \min_{E \in \mathcal{E}} \min_{\Omega \in \Lambda(\mathcal{B} \cup \{S\};D)} I(X_\Omega; Y_{\Omega^c} | Y_E, X_{\Omega^c}), \tag{3.1}$$

where the maximization is not only over product distributions but over all possible $p(\{x_i\}_{i \in \mathcal{V}})$.

Note that although we state Theorem 3.3 for the weak secrecy capacity, the same expression is also an upper bound on the strong secrecy capacity, because $\bar{C}_s \leq C_s$ in any case. We must also note that the upper bound provided in Theorem 3.3 can be quite loose. In particular, if all the relay nodes are allowed to insert noise ($\mathcal{B} = \mathcal{A}$), the minimization set $\Lambda(\mathcal{B} \cup \{S\}; D)$ is of size one, and it is likely that in this case, the gap to our lower bounds is rather large.

Before presenting the proof of Theorems 3.1, 3.2 and 3.3, we discuss several examples in Section 3.4 and present some auxiliary results in Section 3.5.

3.4 Examples

In this section, we provide several examples to illustrate possible coding strategies for wireless networks.

3.4.1 Deterministic Examples

Through the examples in this section we seek to convince the reader that although the linear deterministic model for signal interaction of Definition 2.18 is not too realistic, it can be very useful. For a given network with linear deterministic signal interaction, it is often easy to come up with an explicit coding scheme that achieves a certain perfect secrecy rate. Insights gained from these coding schemes can be valuable when designing codes for the same network with Gaussian signal interaction. We now provide 4 examples for linear deterministic signal interaction to illustrate the following ideas: The operation of the relays to set up cooperation in a (i) non-layered and (ii) layered network; (iii) Destructive interference and handling of a class of potential eavesdroppers; (iv) A noise insertion/jamming scheme by relays to help secrecy. In the first three examples, we exclude noise insertion by the relays although it would improve the rate-equivocation trade-off. The positive effect of noise insertion is shown in the last example. All four examples are based on the linear deterministic wireless network model of Definition 2.18.

Example 1 Consider the deterministic wireless network in Figure 3.3 with a source S, a relay R and destination D which wants secrecy from an eavesdropper E. Note that this network is not layered (Definition 2.3). This means that the information is received by D and E over multiple times, like in inter-symbol interference. It is clear that the maximum communication rate (with no concern for secrecy) between the source and the destination is 3 bits. As the channel from the source to the eavesdropper E is as strong as that to the destination, we cannot ensure any secrecy in the absence of a relay [10]. Therefore, the cooperation of the relay is crucial to ensure secrecy. Assume, however, that the relay is not allowed to insert noise. In this case, Theorem 3.1 guarantees that strong perfect secrecy is possible at a rate $R_s = \max_{p(x_S)p(x_R)}(R_{S;D}(p) - R_{S;E}(p)) = 3 - 2 = 1$ bit. This is achieved by choosing $p(x_S)$ and $p(x_R)$ to be uniform distributions over the three transmit

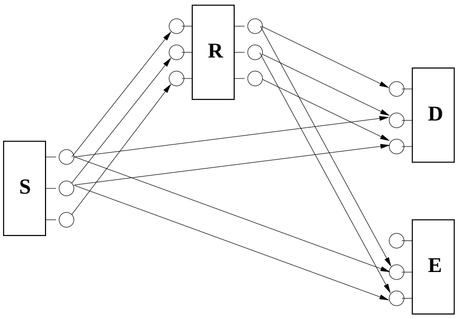

Figure 3.3: Example 1: The relay network with a single helper node providing secrecy against a single eavesdropper.

bits. Now, we show an explicit code that achieves 1 bit of perfect secrecy. Suppose the source transmits bits $(\underline{W}_1[t], \underline{W}_2[t], \underline{W}_3[t])$ at time t. Let the relay transmit bits $(\underline{W}_2[t-1], \underline{W}_1[t-1], \underline{W}_3[t-1])$ via its outputs, where the symbols are delayed because the relay needs to first hear the incoming signal before transmitting. This is a consequence of the non-layered structure of the network. The mixing at the destination (and at the eavesdropper) is therefore of bits transmitted at different times by the source. This is the reason for the time index notion in this example. Assume that the transmission takes place over $T_c + 1$ time-slots, but S only transmits during the first T_c time-slots and therefore sends $3T_c$ bits over $T_c + 1$ time-slots, resulting in a rate of $\frac{3T_c}{T_c+1} \to 3$ bits per time-slot. R transmits information during time-slots 2 through $T_c + 1$. At the end of time-slot t, we have the received signals

$$Y_D[t] = \begin{bmatrix} \underline{W}_2[t-1] \\ \underline{W}_1[t-1] \oplus \underline{W}_1[t] \\ \underline{W}_3[t-1] \oplus \underline{W}_2[t] \end{bmatrix}$$

and

$$Y_E[t] = \begin{bmatrix} 0 \\ \underline{W}_2[t-1] \oplus \underline{W}_1[t] \\ \underline{W}_1[t-1] \oplus \underline{W}_2[t] \end{bmatrix}.$$

At time 1, R does not transmit anything yet, and the destination receives $(0, \underline{W}_1[1], \underline{W}_2[1])$. At time 2, D receives $(\underline{W}_2[1], \underline{W}_1[1] \oplus \underline{W}_1[2], \underline{W}_3[1] \oplus \underline{W}_2[2])$. At time 3, D receives $(\underline{W}_2[2], \underline{W}_1[2] \oplus \underline{W}_1[3], \underline{W}_3[2] \oplus \underline{W}_2[3])$, and it has now enough linearly independent equations to decode $\underline{W}_1[1]$, $\underline{W}_1[2]$, $\underline{W}_1[3]$, $\underline{W}_2[1]$, $\underline{W}_2[2]$ and $\underline{W}_3[1]$. In general, at time-slot $t \in \{3, \ldots, T_c\}$, D can decode $(\underline{W}_1[1], \ldots, \underline{W}_1[t])$, $(\underline{W}_2[1], \ldots, \underline{W}_2[t-1])$ and $(\underline{W}_3[1], \ldots, \underline{W}_3[t-2])$. At time-slot T_c+1, the source S remains silent. Hence, D receives $(\underline{W}_2[T_c], \underline{W}_1[T_c], \underline{W}_3[T_c])$. It can use $\underline{W}_2[T_c]$ to decode $\underline{W}_3[T_c - 1]$ which was still unknown, yielding full knowledge of all transmitted bits. E receives $(0, \underline{W}_1[1], \underline{W}_2[t])$ during time-slot 1, and $(0, \underline{W}_2[1] \oplus \underline{W}_1[2], \underline{W}_1[1] \oplus \underline{W}_2[2])$ during time-slot 2, which allows

E to decode $\underline{W}_1[1], \underline{W}_1[2], \underline{W}_2[1], \underline{W}_2[2]$ at the end of time-slot 2. This argument also holds for all time-slots $t = 3, \ldots, T_c$, and thus, E can decode $(\underline{W}_1[1], \ldots, \underline{W}_1[T_c])$ and $(\underline{W}_2[t], \ldots, \underline{W}_2[T_c])$. However, it is impossible for E to decode any of the $\underline{W}_3[t]$, $t = 1, \ldots, T_c$, because it never receives any information related to these bits. Thus, the scheme described here achieves $(R, R_s) = (3, 1)$ in the sense of Definition 2.15 and hence ensures 1 bit of strong perfect secrecy. If we are not interested in sending any non-secret information, we can use a much simpler code where S and R simply route one bit of information over the "path" in the network that is invisible to E (third transmit bit at both S and R). ■

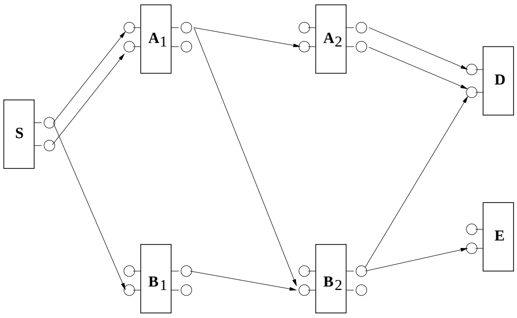

Figure 3.4: Example 2: Two-layer network with single eavesdropper, with a map-forward relaying strategy.

Example 2 We now study a larger deterministic wireless network with $\mathcal{A} = \{A_1, A_2, B_1, B_2\}$. The network is shown in Figure 3.4. Note that this network is layered, and so are all the subsequent examples. Hence, the codes provided for them are block codes as defined in Definition 2.6. If we forbid noise insertion at the relay nodes of this example network, then the perfect secrecy rate guaranteed by Theorem 3.1 is $2 - 1 = 1$ bit, again achieved by uniform transmit distributions. The following scheme illustrates that the relays do not have to decode the source message to ensure secure and reliable communication. The randomly generated code used in the proof of Theorem 3.1 also has this property. The source S sends (W_1, W_2). Each relay operates by taking a linear combination of its received signal and transmitting it. One possible operation is when A_1 sends $(0 \cdot W_1) \oplus (1 \cdot W_2) = W_2$ and B_1 forwards its received signal W_1. Therefore A_2 receives (W_2) and B_2 receives the linear combination $(W_1 \oplus W_2)$. Both A_2 and B_2 simply forward what they receive via all of their outputs. As a result, D receives $(W_2, W_2 \oplus W_1 \oplus W_2)$, and it can solve this equation system to obtain both information bits. E receives only $(W_1 \oplus W_2)$, which reveals one bit of information. This can be shown by computing the equivocation $H(W_1, W_2 | W_1 \oplus W_2) = 1$. Alternatively, we can see that only one bit is revealed by preparing the source bits (W_1, W_2) such that $W_1 = U_1, W_2 = U_1 \oplus U_2$ and therefore since E receives $W_1 \oplus W_2 = U_2$, it has no knowledge of U_1 which is secret. Hence, $R_s = 1$ is achieved by this scheme. ■

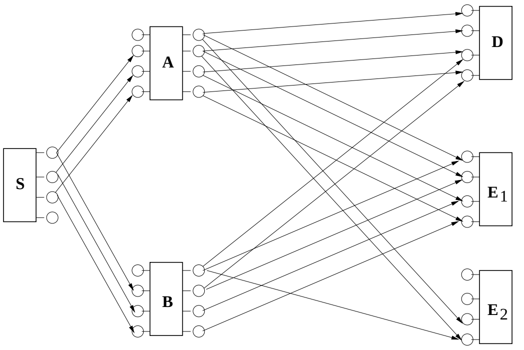

Figure 3.5: Example 3, showing the conflict created by multiple potential eavesdroppers.

Example 3 In Figure 3.5, we illustrate the conflicting needs for secrecy against a class of eavesdroppers, by having $\mathcal{E} = \{E_1, E_2\}$. In addition, this example demonstrates that there are explicit schemes for the linear interaction model that seem to achieve a perfect secrecy rate higher than the one guaranteed by Theorem 3.1. Assume for instance that only eavesdropper E_1 is present. Then, Theorem 3.1 cannot guarantee a positive perfect secrecy rate (at least if no noise-insertion is allowed). Indeed, no matter what distribution $\prod_{i\in\{S,A,B\}} p(x_i)$ we pick, the information-theoretic min-cuts $R_{S;D}(p)$ and $R_{S;E_1}(p)$ are equal. Consider now the following explicit scheme under the assumption that only E_1 is present: S sends (W_1, W_2, W_3), A and B both send $(W_1, W_2, W_3, 0)$. In this case, D receives $(W_1, W_2, W_3 \oplus W_1, W_2)$, while E_1 receives $(0, 0, 0, 0)$. Hence, if only E_1 is present, we can have a perfect secrecy situation that achieves $(R, R_s) = (3, 3)$. This is a scheme that uses deterministic encoders at A and B, *i.e.*, no noise is inserted. The particularity of this scheme is that both A and B decode all the transmitted messages. Thanks to this, they can align their transmission in order to create *destructive* interference at E_1. With a randomly generated code as used in the proof of Theorem 3.1, alignment of relay signals is not possible, and such a strategy can not be taken into account.

If there was only E_2, we could have a different scheme: S sends (W_1, W_2, W_3), A sends $(0, W_1, W_2, W_3)$, while B sends $(W_1, 0, 0, 0)$. In this case, D receives $(0, W_1, W_2 \oplus W_1, W_3)$, while E_2 receives $(0, 0, 0, W_1 \oplus W_1) = (0, 0, 0, 0)$, again leading to $(R, R_s) = (3, 3)$. Again, this strategy uses collaboration at the relays to create destructive interference at E_2.

Finally, we consider the case when $\mathcal{E} = \{E_1, E_2\}$, *i.e.,* either eavesdropper (or both eavesdroppers) could be present. In this case, we need to ensure secrecy against the set $\mathcal{E}$. A possible strategy is: S sends (W_1, W_2, W_3), A sends $(0, W_1, W_2, W_3)$, and B sends $(0, W_1, W_2, W_3)$ or (W_1, W_1, W_2, W_3). In this case, D receives $(0, W_1, W_2, W_3 \oplus W_1)$ or $(0, W_1, W_2 \oplus W_1, W_3 \oplus W_1)$, E_1 receives $(0, 0, 0, 0)$ or $(W_1, 0, 0, 0)$, and E_2 receives $(0, 0, 0, W_1)$ or $(0, 0, 0, 0)$. In either case, we only achieve $(R, R_s) = (3, 2)$ bits. Some additional thought also reveals that it is not possible to make R_s larger than 2 when E_1 and E_2 can both be present (provided that the relay nodes are not allowed to use noise insertion, *i.e.*, their encoding functions must be deterministic). We thus see the tension created by the uncertainty about the eavesdropper. ■

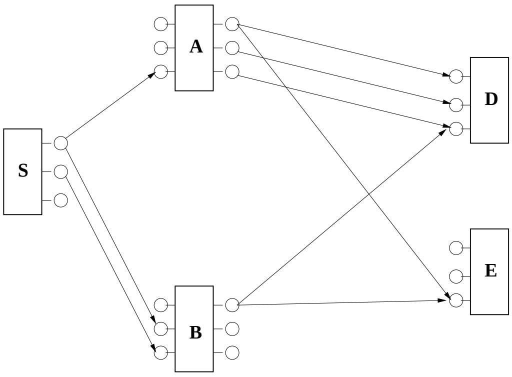

Figure 3.6: Example 4: Example showing how noise insertion at a relay can improve secrecy.

Example 4 For the network given in Figure 3.6, we allow some of the authenticated relay nodes to actively help secrecy by inserting random bits into the communication. In a sense these relays are "jamming" the eavesdropper and helping secure communication. Let the source send (W_1, W_2) at its outgoing nodes. One strategy is that A sends $(0, W_1, 0)$, and B sends (W_2). In this case, D receives $(0, W_1, W_2)$, while E receives $(0, 0, W_2)$, yielding $(R, R_s) = (2, 1)$. Any strategy where the relays do not insert randomness cannot ensure more than 1 bit of secrecy. Now we allow A to generate a random bit b before every transmission. Then, it transmits $(b, W_1, 0)$. The transmitter at B remains unchanged. Now, D receives (b, W_1, W_2). The eavesdropper E receives $(0, 0, W_2 \oplus b)$, and therefore we obtain perfect secrecy of 2 bits. This example confirms that active noise insertion by the relays can enhance secrecy. ■

3.4.2 The Gaussian Diamond Network

In this section, we focus on a particular Gaussian network with two relay nodes called the diamond network. As noted earlier, the lower bound on the perfect secrecy capacity given in Theorem 3.2 may grow logarithmically with the SNR, but this growth is "delayed" by the subtractive constants α, β, γ. For cases where the lower bound does not grow with the SNR, the presence of α, β and γ is particularly unpleasant.

In this section, we prove the existence of schemes that achieve perfect secrecy in the Gaussian diamond network even for small values of the SNR. This supports our conjecture that the constants α, β, γ are not of fundamental nature, but simply a technical artifact in our proof, needed to show the results for arbitrary networks.

Consider the diamond network shown in Figure 3.7, which is defined as follows.

Definition 3.8 *The* ***Gaussian diamond network*** *is a network with two relays A and B and one eavesdropper E, whose channels are given by the equations:*

$$Y_A[t] = h_{SA} X_S[t] + Z_A[t] \tag{3.2}$$

$$Y_B[t] = h_{SB} X_S[t] + Z_B[t] \tag{3.3}$$

$$Y_D[t] = h_{AD} X_A[t] + h_{BD} X_B[t] + Z_D[t] \tag{3.4}$$

$$Y_E[t] = h_{AE} X_A[t] + h_{BE} X_B[t] + Z_E[t], \tag{3.5}$$

where $Z_i[t] \sim \mathcal{N}(0,1)$, and the $Z_i[t]$ are independent of the $X_i[t]$ and of each other.

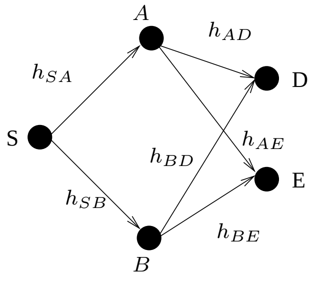

Figure 3.7: Gaussian diamond wireless network.

For simplicity assume that all channel gains are real, and that $|h_{SA}| > |h_{SB}|$. Also, we assume that the power constraint is 1 at every transmitting node. Equations (3.2) and (3.3) together describe a stochastically degraded Gaussian broadcast channel, while equations (3.4) and (3.5) each describe a Gaussian multiple-access channel (MAC).

No Noise Insertion at the Relays

Here, we do not allow the relay nodes to insert noise. For an arbitrary $\theta \in [0,1]$, define the following two functions:

$$R^{\mathrm{BC}}_{m_B}(\theta) \triangleq \frac{1}{2}\log\left(1+\frac{\theta h^2_{SB}}{1+(1-\theta)h^2_{SB}}\right) \tag{3.6}$$

$$R^{\mathrm{BC}}_{m_A}(\theta) \triangleq \frac{1}{2}\log\left(1+(1-\theta)h^2_{SA}\right). \tag{3.7}$$

It is well known [9] that for any $\theta \in [0,1]$, one can reliably transmit a message m_B of rate $R^{\mathrm{BC}}_{m_B}$ to B and simultaneously transmit (m_A, m_B) to A, where m_A is of rate $R^{\mathrm{BC}}_{m_A}$. We also define the following region of rates:

Definition 3.9 *For $k \in \{D, E\}$, we define $\mathcal{R}^{MAC}_k$ as the region of all pairs (R_A, R_B) satisfying*

$$\begin{aligned} R_A &< \frac{1}{2}\log(1+h^2_{Ak}) \\ R_B &< \frac{1}{2}\log(1+h^2_{Bk}) \\ R_A + R_B &< \frac{1}{2}\log(1+h^2_{Ak}+h^2_{Bk}). \end{aligned}$$

Note that $\mathcal{R}^{\mathrm{MAC}}_k$ is the achievable rate-region of the multiple-access channel from A and B to k, for $k \in \{D, E\}$.

Now, we describe an achievable perfect secrecy rate for the Gaussian diamond network.

Theorem 3.4 *In the Gaussian diamond network,*

$$\bar{C}_s \geq \max\left[R^{BC}_{m_A}(\theta) + R^{BC}_{m_B}(\theta) - \frac{1}{2}\log(1+h^2_{AE}+h^2_{BE})\right], \tag{3.8}$$

where the maximization is over all $\theta \in [0,1]$ for which

$$(R^{BC}_{m_A}(\theta), R^{BC}_{m_B}(\theta)) \in \mathcal{R}^{MAC}_D$$

and for which the set

$$\begin{aligned} &\left\{(R_{j_A}, R_{j_B}) \in \mathcal{R}^{MAC}_E : R_{j_A} + R_{j_B} = \frac{1}{2}\log(1+h^2_{AE}+h^2_{BE})\right\} \\ &\qquad \cap \left([0, R^{BC}_{m_A}(\theta)] \times [0, R^{BC}_{m_B}(\theta)]\right) \end{aligned}$$

is non-empty.

The proof of Theorem 3.4 is given in Appendix A.

Noise Insertion at the Relays

In this subsection, we show that the secrecy rate can be improved when noise insertion is allowed at the relay nodes. In particular, we guarantee a certain secrecy rate in the Gaussian diamond network (Definition 3.8) when the relay nodes A and B are allowed to insert noise. Both relay nodes are permitted to select a junk message of arbitrary rate uniformly at random before each transmission of a block and to use this junk message during the encoding. The following theorem states an achievable secrecy rate under this additional assumption.

Theorem 3.5 *In the Gaussian diamond network with noise insertion,*

$$\bar{C}_s \geq \max \left[R_{m_A} + R_{m_B} - \frac{1}{2}\log(1 + h_{AE}^2 + h_{BE}^2) \right], \tag{3.9}$$

where the indicated maximization is over all $(\theta, R_{m_A}, R_{m_B})$ for which

$$(R_{m_A}, R_{m_B}) \in \mathcal{R}_D^{MAC}$$

and $\exists (R_{j_A}, R_{j_B}) \in \Phi_E$ such that

$$(R_{m_A}, R_{m_B}) \in [R_{j_A}, R_{j_A} + R_{m_A}^{BC}(\theta)] \times [R_{j_B}, R_{j_B} + R_{m_B}^{BC}(\theta)],$$

where Φ_E is defined as

$$\begin{aligned} \Phi_E = \big\{ &(R_{j_A}, R_{j_B}) \in \mathcal{R}_E^{MAC} : \\ &R_{j_A} + R_{j_B} = \frac{1}{2}\log(1 + h_{AE}^2 + h_{BE}^2) \big\}. \end{aligned}$$

The proof of Theorem 3.5 is very similar to that of Theorem 3.4. The perfect secrecy rate is achievable by the same code construction, with the difference that the junk messages J_A and J_B are generated at A and B, respectively. Thanks to this, the broadcast channel between S and (A, B) can be fully used to send the information messages W_A and W_B. The secrecy rate is then limited by the capacity of this broadcast channel, as well as by the amount of secrecy the two multiple access channels (from (A, B) to D and to E) can provide. Note that this result resembles the “noise forwarding” strategy for the relay network in [20]. To illustrate the achievable perfect secrecy rates without and with noise insertion, we now provide a numerical example.

A Numerical Example

In this section, we provide a numerical example for the achievable perfect secrecy rates in Theorems 3.4 and 3.5.

Let $h_{SA} = h_{AD} = 2$ be the stronger channel gains. All other channel gains are $h_{SB} = h_{BD} = h_{AE} = h_{BE} = 1$. The multiple access channel (MAC) capacity regions are then given by

$$\begin{aligned}\mathcal{R}_D^{\text{MAC}} &= \{(R_A, R_B) :\\ R_A &< \frac{1}{2}\log 5 = 1.16 \text{ bits},\\ R_B &< \frac{1}{2}\log 2 = 0.5 \text{ bits},\\ R_A + R_B &< \frac{1}{2}\log 6 = 1.29 \text{ bits}\}\end{aligned}$$

and

$$\begin{aligned}\mathcal{R}_E^{\text{MAC}} &= \{(R_A, R_B) :\\ R_A &< \frac{1}{2}\log 2 = 0.5 \text{ bits},\\ R_B &< \frac{1}{2}\log 2 = 0.5 \text{ bits},\\ R_A + R_B &< \frac{1}{2}\log 3 = 0.79 \text{ bits}\}\end{aligned}$$

These two capacity regions are shown in Figure 3.8.

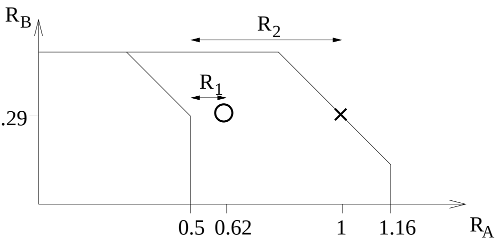

Figure 3.8: The MAC capacity regions $\mathcal{R}_E^{\text{MAC}}$ (small pentagon) and $\mathcal{R}_D^{\text{MAC}}$ (large pentagon). The overall rate pair (R_{m_A}, R_{m_B}) is marked by a cross and a circle for the strategies with and without noise insertion, respectively. The achievable secrecy rates without noise insertion (R_1) and with noise insertion (R_2) are indicated.

Case without Noise Insertion: If the relay nodes do not insert noise, we compute the perfect secrecy rate given by Theorem 3.4, *i.e.*, we maximize

$$\begin{aligned}&R_{m_A}^{\text{BC}}(\theta) + R_{m_B}^{\text{BC}}(\theta) =\\ &\quad \frac{1}{2}\log\big(1 + 4(1-\theta)\big) + \frac{1}{2}\log\Big(1 + \frac{\theta}{1 + (1-\theta)}\Big)\end{aligned}$$

under the conditions that $(R_{m_A}^{\text{BC}}(\theta), R_{m_B}^{\text{BC}}(\theta)) \in \mathcal{R}_D^{\text{MAC}}$ and that $\Phi_E \cap \big([0, R_{m_A}^{\text{BC}}(\theta)] \times [0, R_{m_B}^{\text{BC}}(\theta)]\big)$ should be non-empty. We find that the optimal value of θ is

$\theta^{\text{opt}} = 0.66$, and

$$\begin{aligned} R_{m_A} &= R_{m_A}^{\text{BC}}(\theta^{\text{opt}}) = 0.62 \text{ bits}, \\ R_{m_B} &= R_{m_B}^{\text{BC}}(\theta^{\text{opt}}) = 0.29 \text{ bits}. \end{aligned}$$

This point is marked with a circle in Figure 3.8. The achievable perfect secrecy rate is

$$\begin{aligned} R_1 &= R_{m_A} + R_{m_B} - \frac{1}{2}\log(1 + h_{AE}^2 + h_{BE}^2) \\ &= 0.62 + 0.29 - 0.79 = 0.12 \text{ bits}. \end{aligned} \tag{3.10}$$

Case with Noise Insertion: For the case when noise insertion at the relay nodes is permitted, we compute the perfect secrecy rate given by Theorem 3.5. It is easy to see that by choosing for instance $R_{j_A} = 0.5$ and $R_{j_B} = 0.29$, we can achieve an overall rate of

$$\begin{aligned} R_{m_B} &= 0.29 \text{ bits} \\ R_{m_A} &= \frac{1}{2}\log(1 + h_{AD}^2 + h_{BD}^2) - R_{m_B} \\ &= 1.29 - 0.29 = 1 \text{ bit}. \end{aligned}$$

This point is marked with a cross in Figure 3.8. In this case, the achievable perfect secrecy rate is

$$\begin{aligned} R_2 &= R_{m_A} + R_{m_B} - \frac{1}{2}\log(1 + h_{AE}^2 + h_{BE}^2) \\ &= 1 + 0.29 - 0.79 = 0.5 \text{ bits}. \end{aligned}$$

This secrecy rate is considerably larger than (3.10). The two achievable secrecy rates R_1 and R_2 are also shown in Figure 3.8. All the rates are measured in bits and hence all logarithms are base 2.

Note that the rates provided in these numerical examples are only lower bounds on the perfect secrecy capacity $\bar{C}_s$. To prove that noise insertion gives a fundamental improvement of the perfect secrecy rate, we would need an upper bound on the perfect secrecy rate without noise insertion.

Upper Bound: We can, however, derive the following general upper bound on $\bar{C}_s$. If we allow A and B to collaborate with S, then we obtain a special case of the so-called multiple input antennas, single output antenna, multiple eavesdropper antennas (MISOME) wiretap channel, which was studied by Khisti and Wornell in [18]. In our case, the eavesdropper has only one antenna. The (weak) secrecy capacity of this multi-antenna wiretap channel was found in [18] and it is clearly an upper bound on C_s, which is in turn an upper bound on $\bar{C}_s$:

$$\bar{C}_s \leq \frac{1}{2}\log \lambda_{\max}\big(\mathbf{I} + P\mathbf{h}_d\mathbf{h}_d^T, \mathbf{I} + P\mathbf{h}_e\mathbf{h}_e^T\big),$$

where $\mathbf{h}_d = (2,1)^T$ and $\mathbf{h}_e = (1,1)^T$, $P = 2$ is the sum power constraint, and $\lambda_{\max}(\cdot,\cdot)$ denotes the largest generalized eigenvalue of the argument pair. In our numerical example, the upper bound yields

$$\bar{C}_s \leq 0.87028 \text{ bits.}$$

Comparing this to our best lower bound of 0.5 bits, we see that the two bounds provide a reasonable approximation of $\bar{C}_s$.

3.5 Results for the Multisource Problem

3.5.1 Setup and Outer Bound

Here, we consider a network without eavesdropper, but with several source nodes. Let a relay network be given as defined in Section 3.2, and let the signal interaction be deterministic or Gaussian as defined in Definitions 2.17 and 2.16, respectively. However, instead of having just one source node S, assume that each node i in a subset $\mathcal{S} \subset \mathcal{V}$ observes an independent source $\underline{W}_i$ of rate R_i. The aim is that the destination node D can reliably reconstruct all the sources $\{\underline{W}_i\}_{i\in\mathcal{S}}$. All the encoders are supposed to be deterministic. We call this the multisource problem.

In the next two subsections, we provide achievability results (Theorems 3.7 and 3.9) that are used to prove the secrecy guarantees given in Section 3.3. Since the results presented in this section may be of independent interest, we give a few additional results on the multisource problem, starting with the following infeasibility result, which is valid for all types of memoryless signal interaction.

If $\mathcal{I} \subseteq \mathcal{S}$ is a subset of the source nodes, then $R_{\mathcal{I}}$ denotes a tuple of rates (one rate for each $i \in \mathcal{I}$).

Definition 3.10 *A rate tuple $R_{\mathcal{S}}$ is called* ***achievable*** *if for any $\epsilon > 0$, there exists a communication length T_c and a (T_c, ϵ)-code (for multiple sources) for which the rate of message $\underline{W}_i$ is at least $R_i - \epsilon$ for all $i \in \mathcal{S}$. In this context, a (T_c, ϵ)-code is required to have $\mathbf{P}(\underline{W}_i \neq \hat{\underline{W}}_i) \leq \epsilon$ for all $i \in \mathcal{S}$.*

Theorem 3.6 *In the multisource problem with arbitrary signal interaction, if a rate tuple $R_{\mathcal{S}}$ is achievable then it is contained in*

$$\mathcal{R}_{outer} = \cup_{p(\{x_i\}, i\in\mathcal{V})} \mathcal{R}_{\mathcal{S};D}(p),$$

where $\mathcal{R}_{\mathcal{S};D}(p)$ is defined in Definition 3.2.

Theorem 3.6 is basically the cut-set bound on communication in arbitrary networks. This theorem as well as the remaining results on the multisource problem are proved in Appendix A.

3.5.2 Deterministic Signal Interaction

We now give an inner bound on the set of all achievable rate tuples for deterministic signal interaction.

Theorem 3.7 *In the multisource problem with deterministic signal interaction, any rate tuple $R_{\mathcal{S}}$ in*

$$\mathcal{R}_{inner} \triangleq \cup_{\prod_{i\in\mathcal{V}} p(x_i)} \mathcal{R}_{\mathcal{S};D}(p)$$

is achievable.

The set $\mathcal{R}_{\mathcal{S};D}(p)$ is defined in Definition 3.2.

In particular, for the case of two source nodes S_1 and S_2 that wish to transmit messages W_1 and W_2, respectively, to D, communication rates (R_1, R_2) satisfying

$$\begin{aligned} R_1 &\leq R_{S_1;D}(p) \\ R_2 &\leq R_{S_2;D}(p) \\ R_1 + R_2 &\leq R_{S_1 S_2;D}(p) \end{aligned}$$

are achievable, where the distribution on the transmit alphabets can be any product distribution $\prod_{i\in\mathcal{V}} p(x_i)$.

For deterministic signal interaction, the only difference between inner and outer bound is that in the outer bound of Theorem 3.6, the union is over all possible distributions $p(\{x_i\}, i \in \mathcal{V})$, whereas in the inner bound, the union is only taken over product distributions. For linear deterministic signal interaction as defined in Definition 2.18, it turns out that the two bounds match. More precisely, we have

Theorem 3.8 *For an arbitrary wireless relay network with linear deterministic signal interaction as defined in Definition 2.18, the multisource capacity region is equal to*

$$\begin{aligned} \mathcal{R}_{\text{inner}} = \mathcal{R}_{\text{outer}} = \big\{ & R_{\mathcal{S}} : \forall \mathcal{I} \subseteq \mathcal{S}, \\ & \sum_{i\in\mathcal{I}} R_i \leq \min_{\Omega\in\Lambda(\mathcal{I};D)} \text{rank}\left(\mathbf{G}_{\Omega,\Omega^c}\right) \big\}. \end{aligned}$$

The capacity region is defined as the region of all achievable tuples $R_{\mathcal{S}}$. The proofs of Theorems 3.7 and 3.8 are given in Appendices A.5 and A.6.

3.5.3 Gaussian Signal Interaction

Theorem 3.9 *In the multisource problem for an acyclic network with Gaussian signal interaction, any rate tuple $R_{\mathcal{S}}$ in*

$$\mathcal{R}^{G}_{inner} \triangleq \mathcal{R}^{G}_{\mathcal{S};D}(\beta)$$

is achievable, where β is a constant as defined in Definition 3.1.

The set $\mathcal{R}^{\mathrm{G}}_{\mathcal{S};D}(\Delta)$ is defined in Definition 3.5.

For the Gaussian multisource problem, we provide an approximate characterization of the capacity region, *i.e.*, we state a bound on the gap between $\mathcal{R}_{\mathrm{outer}}$ and $\mathcal{R}^{\mathrm{G}}_{\mathrm{inner}}$. This gap is constant in the sense that it does not depend on the channel parameters (signal-to-noise ratio) but only on the number of nodes in the network and the maximum out-degree. This phenomenon is similar to the gap in Theorem 2.4.

Theorem 3.10 *For an acyclic network with Gaussian signal interaction, for a given rate-tuple $R_{\mathcal{S}} \in \mathcal{R}_{outer}$, we have that*

$$R_{\mathcal{S}} - (\alpha + \beta)\mathbf{1} \in \mathcal{R}^{G}_{inner},$$

where $\mathbf{1}$ *denotes the all-one vector of length $|\mathcal{S}|$, and α and β are defined in Definition 3.1.*

In words, this theorem states that the boundary of $\mathcal{R}^{\mathrm{G}}_{\mathrm{inner}}$ is not more than $\alpha + \beta$ away from the outer bound in every dimension. The proofs of Theorems 3.9 and 3.10 are given in Appendices A.7 and A.8.

3.6 Proof of the Main Results

We are now ready to prove the three main results of Section 3.3.

3.6.1 Proof of Theorems 3.1 and 3.2

The main parts of the proofs are the following two propositions, which give weak secrecy guarantees.

Proposition 3.1 *For an arbitrary wireless network with deterministic signal interaction, the weak perfect secrecy capacity is lower bounded as*

$$C_s \geq \max_{\prod_{i \in \mathcal{V}} p(x_i)} F(p). \tag{3.11}$$

Proposition 3.1 is proved in the next two subsections for layered and non-layered networks, respectively.

Proposition 3.2 *Let $\rho = \alpha + \beta + \gamma$. For an acyclic wireless network with Gaussian signal interaction, the (weak) ρ-almost perfect secrecy rate is lower bounded as*

$$C_s^{\rho} \geq F^{G}(\beta). \tag{3.12}$$

The proof of Proposition 3.2 is very similar to the proof of Proposition 3.1. A discussion of the differences is given in Sections 3.6.4 and 3.6.5 for layered and non-layered networks, respectively.

To strengthen these weak secrecy guarantees, we use Lemma E.5 given in Appendix E, which is a slight adaption of the results by Maurer and Wolf [29].

Define

$$R_s \triangleq \max_{\prod_{i\in\mathcal{V}} p(x_i)} F(p).$$

For deterministic signal interaction, Proposition 3.1 tells us that (R_s, R_s) is weakly achievable. Hence, by Lemma E.5, R_s is strongly achievable. This proves Theorem 3.1.

For acyclic networks with Gaussian signal interaction, Proposition 3.2 tells us that $\big(F^{\mathrm{G}}(\beta), F^{\mathrm{G}}(\beta) - \rho\big)$ is weakly achievable. Hence, by Lemma E.5, $F^{\mathrm{G}}(\beta) - \rho$ is strongly achievable, where $\rho = \alpha + \beta + \gamma$. This proves Theorem 3.2.

3.6.2 Proof of Proposition 3.1 for Layered Networks

In this section, we prove Proposition 3.1 for the case where there are two eavesdroppers $\mathcal{E} = \{E_1, E_2\}$. The proof for an arbitrary number of eavesdroppers is analogous. Assume that the network is layered (equal-path) as defined in Definition 2.3. Fix a product distribution $\prod_{i\in\mathcal{V}} p(x_i)$. Let a tuple of non-negative numbers $B_{\mathcal{B}} = \{B_i\}_{i\in\mathcal{B}}$ be given such that

$$B_{\mathcal{B}} \in \cap_{E\in\mathcal{E}} \mathcal{R}_{\mathcal{B};E}(p), \tag{3.13}$$

i.e., the tuple $B_{\mathcal{B}}$ lies inside the $|\mathcal{B}|$-dimensional multisource achievable region for each of the two eavesdroppers (see Section 3.5). Let

$$\phi_D \triangleq \max_x \{x : (x, B_{\mathcal{B}}) \in \tilde{\mathcal{R}}_{\mathcal{B};D}(p)\} \tag{3.14}$$

and

$$\phi_E \triangleq \max_x \{x : (x, B_{\mathcal{B}}) \in \mathcal{R}_{\mathcal{B}\cup\{S\};E}(p)\}, \tag{3.15}$$

for $E \in \mathcal{E}$. The quantity ϕ_E is the largest value that B' can take such that the tuple $(B', B_{\mathcal{B}})$ lies in the $(|\mathcal{B}| + 1)$-dimensional multisource achievable region for eavesdropper E, where $E \in \{E_1, E_2\}$. Without loss of generality, assume that $\phi_{E_1} \geq \phi_{E_2}$.

Define the following rates:

$$R \triangleq \phi_D - \phi_{E_1}, \tag{3.16}$$

$$B_1 \triangleq \phi_{E_1} - \phi_{E_2}, \tag{3.17}$$

and

$$B_2 \triangleq \phi_{E_2} - \epsilon_1, \tag{3.18}$$

where ϵ_1 is an arbitrary constant. This choice of rates is represented pictorially in Figure 3.9. Let W be the message (to be kept secret) that S wishes to

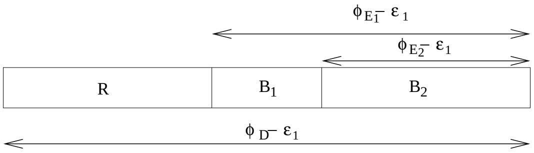

Figure 3.9: Illustration of the rate allocation in (3.16), (3.17) and (3.18).

transmit, and let it be of rate R. More precisely, W is uniformly distributed in $\{1, \ldots, 2^{\lfloor TR \rfloor}\}$. Here, T is the block length, *i.e.*, all sequences are of length T. In addition to W, S generates two independent, random junk messages J_1 and J_2 uniformly from $\{1, \ldots, 2^{\lfloor TB_1 \rfloor}\}$ and $\{1, \ldots, 2^{\lfloor TB_2 \rfloor}\}$. Likewise, each noise-inserter $i \in \mathcal{B}$ generates an independent, random junk message J_i uniformly from $\{1, \ldots, 2^{\lfloor TB_i \rfloor}\}$. We generate a random block code (Definition 2.6) in the following way. For the source node S, each (w, j_1, j_2) is mapped to a sequence $\mathbf{x}_S$, drawn uniformly at random from the set of all δ-typical sequences with respect to the distribution $p(x_S)$ (see Appendix F). Every transmit sequence $\mathbf{x}_S$ can lead to only one received sequence $\mathbf{y}_i$ for each relay node i in the first layer of the network. Hence, node i can directly re-encode this sequence (and a possibly present noise message j_i) into a transmit sequence $\mathbf{x}_i$. Since this is true for the nodes in all layers, we can construct the encoding functions at the relay nodes as follows. For each relay node $i \in \mathcal{A} \setminus \mathcal{B}$, each possible received sequence $\mathbf{y}_i$ is mapped to a transmit sequence $\mathbf{x}_i$, drawn uniformly at random from $\mathcal{T}_\delta(X_i)$, where $\mathcal{T}_\delta$ is the set of robustly typical sequences defined in Appendix F. For each noise-inserting node $i \in \mathcal{B}$, each possible pair $(\mathbf{y}_i, j_i)$ is mapped to a sequence $\mathbf{x}_i$, drawn uniformly at random from $\mathcal{T}_\delta(X_i)$. Once this random code generation process is finished, the mappings are deterministic and fixed for all time. In this proof, $(\mathbf{X}_\mathcal{V}, \mathbf{Y}_\mathcal{V})$ denotes the transmit and received sequences that depend on the same realization of the message W. Since the network is layered, none of these sequences depends on more than one message realization W. This implies that for $i, j \in \mathcal{V}$, the t^{th} component of $\mathbf{X}_i$ and the t^{th} component of $\mathbf{X}_j$ do not occur at the same point in time unless i and j lie in the same layer (at the same distance from S) in the network. The randomness of the sequences $(\mathbf{X}_\mathcal{V}, \mathbf{Y}_\mathcal{V})$ comes from the randomness of $(W, J_1, J_2, J_\mathcal{B})$.

Note that by the definition of R, B_1 and B_2, we have

$$(R + B_1 + B_2, B_\mathcal{B}) \in \tilde{\mathcal{R}}_{\mathcal{B};D}(p). \tag{3.19}$$

The region $\tilde{\mathcal{R}}_{\mathcal{B};D}(p)$ is obtained from $\mathcal{R}_{\mathcal{B}\cup\{S\};D}(p)$ by relaxing all constraints which do not involve $B' \triangleq R + B_1 + B_2$. We examine the proof of Theorem 3.7, with the first message being (W, J_1, J_2) and the subsequent messages being $\{W_i\}_{i \in \mathcal{B}}$. By doing so, we notice that when (3.19) holds, then the above code construction guarantees that for T large enough,

$$\mathrm{E}\left[\mathbf{P}(\text{decoding } (W, J_1, J_2) \text{ wrongly at } D)\right] \leq \frac{\epsilon_0}{5}, \tag{3.20}$$

where the expectation is taken over the (randomly generated) code, and ϵ_0 is an arbitrary constant. Note that nothing is known about the decodability of the relay noise messages $J_{\mathcal{B}}$ at D. At the same time, we obtain from (3.17) and (3.18) that

$$(B_1 + B_2, B_{\mathcal{B}}) \in \mathcal{R}_{\mathcal{B}\cup\{S\};E_1}(p) \tag{3.21}$$

and

$$(B_2, B_{\mathcal{B}}) \in \mathcal{R}_{\mathcal{B}\cup\{S\};E_2}(p). \tag{3.22}$$

For a fixed $W = w$, the codewords generated at S and at the nodes in $\mathcal{B}$ have the same distribution as a randomly generated multisource code of rates $(B_1 + B_2, B_{\mathcal{B}})$. Thus, from (3.21) and from Theorem 3.7 it follows that for T large enough,

$$\mathrm{E}\left[\mathbf{P}(A_w^{(1)})\right] \le \frac{\epsilon_0}{5}, \tag{3.23}$$

where $A_w^{(1)}$ is the event that E_1 makes a decoding error when trying to decode $((J_1, J_2), J_{\mathcal{B}})$, assuming that $W = w$ and assuming that W is already available at E_1. Again, the expectation is taken over the randomly generated code. Since this is true for all $w \in \{1, \ldots, 2^{\lfloor TR \rfloor}\}$, it follows that for T large enough,

$$\mathrm{E}\left[\frac{1}{2^{\lfloor TR \rfloor}} \sum_{w=1}^{2^{\lfloor TR \rfloor}} \mathbf{P}(A_w^{(1)})\right] \le \frac{\epsilon_0}{5}. \tag{3.24}$$

We apply the same argument once more for E_2, using (3.22) and keeping $(W, J_1) = (w, j_1)$ fixed. It follows that

$$\mathrm{E}\left[\frac{1}{2^{\lfloor TR \rfloor + \lfloor TB_1 \rfloor}} \sum_{w=1}^{2^{\lfloor TR \rfloor}} \sum_{j_1=1}^{2^{\lfloor TB_1 \rfloor}} \mathbf{P}(A_{w,j_1}^{(2)})\right] \le \frac{\epsilon_0}{5}, \tag{3.25}$$

where $A_{w,j_1}^{(2)}$ is the event that E_2 makes a decoding error when trying to decode $(J_2, J_{\mathcal{B}})$, assuming that $(W, J_1) = (w, j_1)$ and that (W, J_1) is already available at E_2.

Recall that we chose B_1 and B_2 such that $B_1 + B_2 = \phi_{E_1} - \epsilon_1$. Let $\mathcal{I}_1 \subseteq \mathcal{B}$ be such that when computing ϕ_{E_1}, defined in (3.15), the constraint

$$x + \sum_{i \in \mathcal{I}_1} B_i \le R_{\mathcal{I}_1 \cup \{S\};E_1}(p)$$

is dominant. Similarly, recall that we chose B_2 such that $B_2 = \phi_{E_2} - \epsilon_1$. Let $\mathcal{I}_2 \subseteq \mathcal{B}$ be the subset whose corresponding constraint in ϕ_{E_2} is dominant. We know from Lemma A.5 that if T is chosen large enough,

$$\mathrm{E}\left[\mathbf{1}_{\{I(\mathbf{X}_S, \mathbf{X}_{\mathcal{I}_1}; \mathbf{Y}_{E_1} | J_{\mathcal{B}\setminus\mathcal{I}_1}) > TR_{\mathcal{I}_1\cup\{S\};E_1}(p)\}}\right] \le \frac{\epsilon_0}{5}, \tag{3.26}$$

where the expectation is over all codes. Similarly, for T large enough,

$$\mathrm{E}\left[\mathbf{1}_{\left\{I(\mathbf{X}_S,\mathbf{X}_{\mathcal{I}_2};\mathbf{Y}_{E_2}|J_{\mathcal{B}\setminus\mathcal{I}_2})>TR_{\mathcal{I}_2\cup\{S\};E_2}(p)\right\}}\right] \leq \frac{\epsilon_0}{5}. \tag{3.27}$$

Combining (3.20), (3.24), (3.25), (3.26) and (3.27), we obtain

$$\begin{aligned}\mathrm{E}\Big[&\mathbf{P}(\text{decoding } (W, J_1, J_2) \text{ wrongly at } D)+\\ &\frac{1}{2^{\lfloor TR\rfloor}}\sum_{w=1}^{2^{\lfloor TR\rfloor}}\mathbf{P}(A_w^{(1)})\Big]+\\ &\frac{1}{2^{(\lfloor TR\rfloor+\lfloor TB_1\rfloor)}}\sum_{w=1}^{2^{\lfloor TR\rfloor}}\sum_{j_1=1}^{2^{\lfloor TB_1\rfloor}}\mathbf{P}(A_{w,j_1}^{(2)})+\\ &\mathbf{1}_{\left\{I(\mathbf{X}_S,\mathbf{X}_{\mathcal{I}_1};\mathbf{Y}_{E_1}|J_{\mathcal{B}\setminus\mathcal{I}_1})>TR_{\mathcal{I}_1\cup\{S\};E_1}(p)\right\}}+\\ &\mathbf{1}_{\left\{I(\mathbf{X}_S,\mathbf{X}_{\mathcal{I}_2};\mathbf{Y}_{E_2}|J_{\mathcal{B}\setminus\mathcal{I}_2})>TR_{\mathcal{I}_2\cup\{S\};E_2}(p)\right\}}\Big] \leq \epsilon_0.\end{aligned}$$

We conclude that for at least one code,

$$\mathbf{P}(\text{decoding } (W, J_1, J_2) \text{ wrongly at } D) \leq \epsilon_0, \tag{3.28}$$

$$\frac{1}{2^{\lfloor TR\rfloor}}\sum_{w=1}^{2^{\lfloor TR\rfloor}}\mathbf{P}(A_w^{(1)}) \leq \epsilon_0, \tag{3.29}$$

$$\frac{1}{2^{(\lfloor TR\rfloor+\lfloor TB_1\rfloor)}}\sum_{w=1}^{2^{\lfloor TR\rfloor}}\sum_{j_1=1}^{2^{\lfloor TB_1\rfloor}}\mathbf{P}(A_{w,j_1}^{(2)}) \leq \epsilon_0, \tag{3.30}$$

$$\mathbf{1}_{\left\{I(\mathbf{X}_S,\mathbf{X}_{\mathcal{I}_1};\mathbf{Y}_{E_1}|J_{\mathcal{B}\setminus\mathcal{I}_1})>TR_{\mathcal{I}_1\cup\{S\};E_1}(p)\right\}} \leq \epsilon_0 \tag{3.31}$$

and

$$\mathbf{1}_{\left\{I(\mathbf{X}_S,\mathbf{X}_{\mathcal{I}_2};\mathbf{Y}_{E_2}|J_{\mathcal{B}\setminus\mathcal{I}_2})>TR_{\mathcal{I}_2\cup\{S\};E_2}(p)\right\}} \leq \epsilon_0 \tag{3.32}$$

at the same time. From (3.28), we see that this particular code is reliable. In addition, (3.29) and (3.30) together with Lemma A.4 imply that for this particular code,

$$H(\mathbf{X}_S, \mathbf{X}_{\mathcal{B}}|W, \mathbf{Y}_{E_1}) \leq 1 + T(B_1 + B_2 + \sum_{i\in\mathcal{B}} B_i)\epsilon_0 \tag{3.33}$$

and

$$H(\mathbf{X}_S, \mathbf{X}_{\mathcal{B}}|W, J_1, \mathbf{Y}_{E_2}) \leq 1 + T(B_2 + \sum_{i\in\mathcal{B}} B_i)\epsilon_0. \tag{3.34}$$

The statements (3.31) and (3.32) are used later in the proof.

It remains to analyze the equivocations at E_1 and E_2 for the given code. Let $\mathcal{I}_1 \subseteq \mathcal{B}$ be as defined above. This definition implies that

$$B_1 + B_2 + \sum_{i \in \mathcal{I}_1} B_i = R_{\mathcal{I}_1 \cup \{S\}; E_1}(p) - \epsilon_1. \tag{3.35}$$

We can write

$$\begin{aligned} H(W|\mathbf{Y}_{E_1}) &\geq H(W|\mathbf{Y}_{E_1}, J_{\mathcal{B}\setminus\mathcal{I}_1}) \\ &\geq I(W; \mathbf{X}_S, \mathbf{X}_{\mathcal{I}_1}|\mathbf{Y}_{E_1}, J_{\mathcal{B}\setminus\mathcal{I}_1}) \\ &= I(\mathbf{X}_S, \mathbf{X}_{\mathcal{I}_1}; W, \mathbf{Y}_{E_1}|J_{\mathcal{B}\setminus\mathcal{I}_1}) \\ &\quad - I(\mathbf{X}_S, \mathbf{X}_{\mathcal{I}_1}; \mathbf{Y}_{E_1}|J_{\mathcal{B}\setminus\mathcal{I}_1}) \\ &= H(\mathbf{X}_S, \mathbf{X}_{\mathcal{I}_1}|J_{\mathcal{B}\setminus\mathcal{I}_1}) \\ &\quad - H(\mathbf{X}_S, \mathbf{X}_{\mathcal{I}_1}|W, \mathbf{Y}_{E_1}, J_{\mathcal{B}\setminus\mathcal{I}_1}) \\ &\quad - I(\mathbf{X}_S, \mathbf{X}_{\mathcal{I}_1}; \mathbf{Y}_{E_1}|J_{\mathcal{B}\setminus\mathcal{I}_1}). \end{aligned} \tag{3.36}$$

We bound the three terms above separately:

$$
\begin{aligned}
H(\mathbf{X}_S, \mathbf{X}_{\mathcal{I}_1} | J_{\mathcal{B}\setminus\mathcal{I}_1}) &\geq I(\mathbf{X}_S, \mathbf{X}_{\mathcal{I}_1}; W, J_1, J_2, J_{\mathcal{I}_1} | J_{\mathcal{B}\setminus\mathcal{I}_1}) \\
&= H(W, J_1, J_2, J_{\mathcal{I}_1} | J_{\mathcal{B}\setminus\mathcal{I}_1}) \\
&\quad - H(W, J_1, J_2, J_{\mathcal{I}_1} | J_{\mathcal{B}\setminus\mathcal{I}_1}, \mathbf{X}_S, \mathbf{X}_{\mathcal{I}_1}) \\
&\overset{(a)}{=} H(W, J_1, J_2, J_{\mathcal{I}_1}) \\
&\quad - H(W | J_{\mathcal{B}\setminus\mathcal{I}_1}, \mathbf{X}_S, \mathbf{X}_{\mathcal{I}_1}) \\
&\quad - H(J_1, J_2, J_{\mathcal{I}_1} | W, J_{\mathcal{B}\setminus\mathcal{I}_1}, \mathbf{X}_S, \mathbf{X}_{\mathcal{I}_1}) \\
&\overset{(b)}{=} H(W, J_1, J_2, J_{\mathcal{I}_1}) \\
&\quad - H(W | J_{\mathcal{B}\setminus\mathcal{I}_1}, \mathbf{X}_S, \mathbf{X}_{\mathcal{I}_1}, \mathbf{Y}_D) \\
&\quad - H(J_1, J_2, J_{\mathcal{I}_1} | W, J_{\mathcal{B}\setminus\mathcal{I}_1}, \mathbf{X}_S, \mathbf{X}_{\mathcal{I}_1}, \mathbf{Y}_{E_1}) \\
&\geq H(W, J_1, J_2, J_{\mathcal{I}_1}) \\
&\quad - H(W | \mathbf{Y}_D) \\
&\quad - \underbrace{H(J_1, J_2, J_{\mathcal{I}_1} | W, J_{\mathcal{B}\setminus\mathcal{I}_1}, \mathbf{Y}_{E_1})}_{\leq H(J_1, J_2, J_{\mathcal{I}_1}, J_{\mathcal{B}\setminus\mathcal{I}_1} | W, \mathbf{Y}_{E_1})} \\
&\overset{(c)}{\geq} H(W, J_1, J_2, J_{\mathcal{I}_1}) \\
&\quad - 1 - TR\epsilon_0 \\
&\quad - 1 - T(B_1 + B_2 + \sum_{i\in\mathcal{B}} B_i)\epsilon_0 \\
&= (\lfloor TR \rfloor + \lfloor TB_1 \rfloor + \lfloor TB_2 \rfloor + \sum_{i\in\mathcal{I}_1} \lfloor TB_i \rfloor) \\
&\quad - 2 - T(R + B_1 + B_2 + \sum_{i\in\mathcal{B}} B_i)\epsilon_0, \\
&\overset{(d)}{=} T(R + B_1 + B_2 + \sum_{i\in\mathcal{I}_1} B_i - \epsilon_2) \\
&\quad - 2 - T(R + B_1 + B_2 + \sum_{i\in\mathcal{B}} B_i)\epsilon_0, \qquad (3.37)
\end{aligned}
$$

where (a) is true because of the independence of all messages, (b) follows from the Markov chains $W -\!\!\circ\!\!- \mathbf{X}_S -\!\!\circ\!\!- (J_{\mathcal{B}\setminus\mathcal{I}_1}, \mathbf{X}_{\mathcal{I}_1}, \mathbf{Y}_D)$ and $(J_1, J_2, J_{\mathcal{I}_1}) -\!\!\circ\!\!- (W, \mathbf{X}_S, J_{\mathcal{B}\setminus\mathcal{I}_1}, \mathbf{X}_{\mathcal{I}_1}) -\!\!\circ\!\!- \mathbf{Y}_{E_1}$. In (c), we have used Fano's inequality, together with (3.28), as well as Lemma A.4, together with (3.29). Finally, the small constant ϵ_2 in (d) is introduced because we drop the rounding. As T grows large, ϵ_2 goes to zero.

The second term in (3.36) can be upper bounded as follows:

$$\begin{aligned}H(\mathbf{X}_S,\mathbf{X}_{\mathcal{I}_1}|W,\mathbf{Y}_{E_1},J_{\mathcal{B}\setminus\mathcal{I}_1}) &\leq H(\mathbf{X}_S,\mathbf{X}_{\mathcal{B}}|W,\mathbf{Y}_{E_1},J_{\mathcal{B}\setminus\mathcal{I}_1})\\ &\leq H(\mathbf{X}_S,\mathbf{X}_{\mathcal{B}}|W,\mathbf{Y}_{E_1})\\ &\leq T\big(\frac{1}{T}+(B_1+B_2+\sum_{i\in\mathcal{B}}B_i)\epsilon_0\big),\end{aligned} \tag{3.38}$$

where we used (3.33) in the last inequality. The third term in (3.36) is bounded because for $\epsilon_0 < 1$, (3.31) implies that

$$I(\mathbf{X}_S,\mathbf{X}_{\mathcal{I}_1};\mathbf{Y}_{E_1}|J_{\mathcal{B}\setminus\mathcal{I}_1}) \leq TR_{\mathcal{I}_1\cup\{S\};E_1}(p). \tag{3.39}$$

Plugging (3.37), (3.38) and (3.39) into (3.36), we obtain

$$\begin{aligned}\frac{1}{T}H(W|\mathbf{Y}_{E_1}) &\geq R+B_1+B_2+\sum_{i\in\mathcal{I}_1}B_i-\epsilon_2\\ &\quad-\frac{2}{T}-(R+B_1+B_2+\sum_{i\in\mathcal{B}}B_i)\epsilon_0\\ &\quad-\frac{1}{T}-(B_1+B_2+\sum_{i\in\mathcal{B}}B_i)\epsilon_0\\ &\quad-R_{\mathcal{I}_1\cup\{S\};E_1}(p)\\ &=R-\epsilon_1-\epsilon_2-\frac{3}{T}\\ &\quad-\big(R+2(B_1+B_2+\sum_{i\in\mathcal{B}}B_i)\big)\epsilon_0,\end{aligned} \tag{3.40}$$

where we have used (3.35) in the last step. Now, we analyze the equivocation for eavesdropper E_2. Let $\mathcal{I}_2 \subseteq \mathcal{B}$ be the subset whose corresponding constraint in ϕ_{E_2} is dominant, as defined earlier. This implies that

$$B_2+\sum_{i\in\mathcal{I}_2}B_i = R_{\mathcal{I}_2\cup\{S\};E_2}(p)-\epsilon_1. \tag{3.41}$$

We can write

$$\begin{aligned}H(W|\mathbf{Y}_{E_2}) &\geq H(W|\mathbf{Y}_{E_2},J_{\mathcal{B}\setminus\mathcal{I}_2},J_1)\\ &\geq H(\mathbf{X}_S,\mathbf{X}_{\mathcal{I}_2}|J_{\mathcal{B}\setminus\mathcal{I}_2},J_1)\\ &\quad-H(\mathbf{X}_S,\mathbf{X}_{\mathcal{I}_2}|W,\mathbf{Y}_{E_2},J_{\mathcal{B}\setminus\mathcal{I}_2},J_1)\\ &\quad-I(\mathbf{X}_S,\mathbf{X}_{\mathcal{I}_2};\mathbf{Y}_{E_2}|J_{\mathcal{B}\setminus\mathcal{I}_2},J_1),\end{aligned} \tag{3.42}$$

where we used the same chain of inequalities as in (3.36), but conditioned on

J_1. We bound the three terms above similarly as in (3.36):

$$\begin{aligned} H(\mathbf{X}_S, \mathbf{X}_{\mathcal{I}_2} | J_{\mathcal{B}\setminus\mathcal{I}_2}, J_1) &\geq H(W, J_2, J_{\mathcal{I}_2}) \\ &\quad - H(W|\mathbf{Y}_D) \\ &\quad - H(J_2, J_{\mathcal{B}}|W, J_1, \mathbf{Y}_{E_2}) \\ &\geq T(R + B_2 + \sum_{i\in\mathcal{I}_2} B_i - \epsilon_3) \\ &\quad - 1 - TR\epsilon_0 \\ &\quad - 1 - T(B_2 + \sum_{i\in\mathcal{B}} B_i)\epsilon_0 \\ &= T(R + B_2 + \sum_{i\in\mathcal{I}_2} B_i - \epsilon_3) \\ &\quad - 2 - T(R + B_2 + \sum_{i\in\mathcal{B}} B_i)\epsilon_0, \end{aligned} \tag{3.43}$$

where ϵ_3 goes to zero with increasing T,

$$\begin{aligned} H(\mathbf{X}_S, \mathbf{X}_{\mathcal{I}_2} | W, \mathbf{Y}_{E_2}, J_{\mathcal{B}\setminus\mathcal{I}_2}, J_1) &\leq H(\mathbf{X}_S, \mathbf{X}_{\mathcal{B}} | W, \mathbf{Y}_{E_2}, J_1) \\ &\leq T(\frac{1}{T} + (B_2 + \sum_{i\in\mathcal{B}} B_i)\epsilon_0), \end{aligned} \tag{3.44}$$

where we used (3.34) in the last inequality, and finally,

$$\begin{aligned} I(\mathbf{X}_S, \mathbf{X}_{\mathcal{I}_2}; \mathbf{Y}_{E_2} | J_{\mathcal{B}\setminus\mathcal{I}_2}, J_1) &\leq H(\mathbf{Y}_{E_2} | J_{\mathcal{B}\setminus\mathcal{I}_2}) \\ &\quad - H(\mathbf{Y}_{E_2} | J_{\mathcal{B}\setminus\mathcal{I}_2}, J_1, \mathbf{X}_S, \mathbf{X}_{\mathcal{I}_2}) \\ &\overset{(a)}{=} H(\mathbf{Y}_{E_2} | J_{\mathcal{B}\setminus\mathcal{I}_2}) \\ &\quad - H(\mathbf{Y}_{E_2} | J_{\mathcal{B}\setminus\mathcal{I}_2}, \mathbf{X}_S, \mathbf{X}_{\mathcal{I}_2}) \\ &= I(\mathbf{X}_S, \mathbf{X}_{\mathcal{I}_2}; \mathbf{Y}_{E_2} | J_{\mathcal{B}\setminus\mathcal{I}_2}) \\ &\leq TR_{\mathcal{I}_2\cup\{S\};E_2}(p) \end{aligned} \tag{3.45}$$

from (3.32). In (a), we used the Markov chain $J_1 \multimap \mathbf{X}_S \multimap (\mathbf{Y}_{E_2}, \mathbf{X}_{\mathcal{I}_2}, J_{\mathcal{B}\setminus\mathcal{I}_2})$.

Plugging (3.43), (3.44) and (3.45) into (3.42), we obtain

$$\begin{aligned}\frac{1}{T}H(W|\mathbf{Y}_{E_1}) \geq\; & R + B_2 + \sum_{i\in\mathcal{I}_2} B_i - \epsilon_3 \\ & - \frac{2}{T} - (R + B_2 + \sum_{i\in\mathcal{B}} B_i)\epsilon_0 \\ & - \frac{1}{T} - (B_2 + \sum_{i\in\mathcal{B}} B_i)\epsilon_0 \\ & - R_{\mathcal{I}_2\cup\{S\};E_2}(p) \\ =\; & R - \epsilon_1 - \epsilon_3 \\ & - \frac{3}{T} - \big(R + 2(B_2 + \sum_{i\in\mathcal{B}} B_i)\big)\epsilon_0, \end{aligned} \tag{3.46}$$

where we have used (3.41) in the last step. From (3.40) and (3.46), we see that by choosing ϵ_0 and ϵ_1 small enough and T large enough, we can make ϵ_2 and ϵ_3 as small as we like, and hence ensure that both equivocations are arbitrarily close to R. This concludes the proof for layered networks. The proof for non-layered networks is given in the next section.

3.6.3 Proof of Proposition 3.1 for Non-Layered Networks

When the network is not layered, we use the time-expansion technique presented in Appendix A.12, which closely follows [3, 5]. Consider a non-layered network $\mathcal{G} = (\mathcal{V}, \mathcal{L})$ with a set of noise-inserting nodes $\mathcal{B}$ and a set of eavesdroppers $\mathcal{E}$. We define an unfolded network $\mathcal{G}_{\text{unf}}^{(K)}$ as described in Appendix A.12, with $\mathcal{S} = \mathcal{B} \cup \{S\}$ and $\mathcal{D} = \mathcal{E} \cup \{D\}$. Let R be any transmission rate that satisfies

$$R \leq \max_{\prod_{i\in\mathcal{V}} p(x_i)} F(p),$$

Let p be a product distribution such that $R \leq F(p)$ for that particular p. From Lemma A.7, we know that for any $E \in \mathcal{E}$,

$$K\mathcal{R}_{\mathcal{B}\cup\{S\};E}(p) \simeq \mathcal{R}^{\text{unf}}_{\mathcal{B}\cup\{S\};E}(p),$$

where $\simeq$ means approximate equality for large unfolding length K. Define $\tilde{\mathcal{R}}^{\text{unf}}_{\mathcal{B};D}(p)$ to be the set of all tuples $(B'^{\text{unf}}, B^{\text{unf}}_{\mathcal{B}})$ such that for any subset $\mathcal{I} \subseteq \mathcal{B}$,

$$B'^{\text{unf}} + \sum_{i\in\mathcal{I}} B_i^{\text{unf}} \leq \min_{\Omega_{\text{unf}}\in\Lambda_{\text{unf}}(\mathcal{I}\cup\{S\},D)} I(X_{\Omega_{\text{unf}}}; Y_{\Omega^c_{\text{unf}}}|X_{\Omega^c_{\text{unf}}}).$$

In other words, $\tilde{\mathcal{R}}^{\text{unf}}_{\mathcal{B};D}(p)$ is the unfolded equivalent of $\tilde{\mathcal{R}}_{\mathcal{B};D}(p)$ defined in Definition 3.3. It is straightforward to apply the proof of Lemma A.7 to $\tilde{\mathcal{R}}$, yielding the result that

$$K\tilde{\mathcal{R}}_{\mathcal{B};D}(p) \simeq \tilde{\mathcal{R}}^{\text{unf}}_{\mathcal{B};D}(p).$$

It follows that

$$\begin{aligned}
F(p) = &\max_{B_{\mathcal{B}} \in \cap_{E\in\mathcal{E}} \mathcal{R}_{\mathcal{B};E}(p)} \Big[\\
&\max_x \{x : (x, B_{\mathcal{B}}) \in \tilde{\mathcal{R}}_{\mathcal{B};D}(p)\} \\
&- \max_x \{x : (x, B_{\mathcal{B}}) \in \cup_{E\in\mathcal{E}} \mathcal{R}_{\mathcal{B}\cup\{S\};E}(p)\}\Big] \\
\simeq &\max_{K B_{\mathcal{B}} \in \cap_{E\in\mathcal{E}} \mathcal{R}^{\text{unf}}_{\mathcal{B};E}(p)} \Big[\\
&\max_x \{x : (Kx, KB_{\mathcal{B}}) \in \tilde{\mathcal{R}}^{\text{unf}}_{\mathcal{B};D}(p)\} \\
&- \max_x \{x : (Kx, KB_{\mathcal{B}}) \in \cup_{E\in\mathcal{E}} \mathcal{R}^{\text{unf}}_{\mathcal{B}\cup\{S\};E}(p)\}\Big] \\
= &\frac{1}{K} \max_{B^{\text{unf}}_{\mathcal{B}} \in \cap_{E\in\mathcal{E}} \mathcal{R}^{\text{unf}}_{\mathcal{B};E}(p)} \Big[\\
&\max_x \{x : (x, B^{\text{unf}}_{\mathcal{B}}) \in \tilde{\mathcal{R}}^{\text{unf}}_{\mathcal{B};D}(p)\} \\
&- \max_x \{x : (x, B^{\text{unf}}_{\mathcal{B}}) \in \cup_{E\in\mathcal{E}} \mathcal{R}^{\text{unf}}_{\mathcal{B}\cup\{S\};E}(p)\}\Big] \\
\triangleq &\frac{1}{K} F^{\text{unf}}(p),
\end{aligned}$$

where the last step serves as a definition of $F^{\text{unf}}(p)$. Thus, $KR \leq KF(p) \simeq F^{\text{unf}}(p)$. Since the unfolded network $\mathcal{G}^{(K)}_{\text{unf}}$ is layered, the proof for layered networks lets us conclude that for deterministic signal interaction, the perfect secrecy capacity C_s is lower bounded by $F^{\text{unf}}(p)$. Hence, as K grows large, there is a block coding strategy that achieves rate KR with perfect secrecy in $\mathcal{G}^{(K)}_{\text{unf}}$. Since the implementation of such a coding strategy on the original network $\mathcal{G}$ takes K times as many transmissions, we see that rate R is achievable with perfect secrecy on $\mathcal{G}$. This concludes the proof.

3.6.4 Proof of Proposition 3.2 for Layered Networks

In this section, we prove Proposition 3.2 for layered networks. Most parts of the proof are identical to the proof of Proposition 3.1 in Section 3.6.2. Therefore, we only give an outline of the proof for Gaussian signal interaction and point out the differences. Let a tuple of noise rates $B_{\mathcal{B}}$ be given such that

$$B_{\mathcal{B}} \in \cap_{E\in\mathcal{E}} \mathcal{R}^{\text{G}}_{\mathcal{B};E}(\beta), \tag{3.47}$$

i.e., the tuple $B_{\mathcal{B}}$ lies inside the $|\mathcal{B}|$-dimensional multisource achievable region for each of the two eavesdroppers (see Section 3.5). Let

$$\phi_D \triangleq \max_x \{x : (x, B_{\mathcal{B}}) \in \tilde{\mathcal{R}}^{\text{G}}_{\mathcal{B};D}(\beta)\} \tag{3.48}$$

and

$$\phi_E \triangleq \max_x \{x : (x, B_{\mathcal{B}}) \in \mathcal{R}^{\text{G}}_{\mathcal{B}\cup\{S\};E}(\beta)\}, \tag{3.49}$$

for $E \in \mathcal{E}$. Define the rates R, B_1 and B_2 used at the source S as in (3.16) through (3.18).

In addition to the encoding functions, each node $i \in \mathcal{A}$ uses a Gaussian vector quantizer to map its received sequence $\mathbf{Y}_i$ to a representation sequence $\hat{\mathbf{Y}}_i$. The vector quantizer is constructed by picking $2^{T(I(Y_i;\hat{Y}_i)+\tilde{\epsilon})}$ sequences uniformly at random from $\mathcal{T}_\delta(\hat{Y}_i)$, where $\hat{Y}_i = \alpha_i Y_i + \xi_i$ as defined in Definition 2.21 and $\tilde{\epsilon} > 0$ can be choosen arbitrarily small. The encoding functions are constructed randomly in the same way as in Section 3.6.2, except that they act on $\hat{\mathbf{y}}_i$ instead of $\mathbf{y}_i$ directly. Then, by Theorem 3.9, all the error analysis steps of Section 3.6.2 can be done in an analogous way here, and we conclude that there exists a code and a set of quantizers for which (3.28) through (3.30) hold. Using Lemma A.4 in Appendix A.11, this implies that for this particular code,

$$H(\mathbf{X}_S, \mathbf{X}_\mathcal{B}|W, \mathbf{Y}_{E_1}) \le 1 + T(B_1 + B_2 + \sum_{i\in\mathcal{B}} B_i)\epsilon_0 + T\gamma \tag{3.50}$$

and

$$H(\mathbf{X}_S, \mathbf{X}_\mathcal{B}|W, J_1, \mathbf{Y}_{E_2}) \le 1 + T(B_2 + \sum_{i\in\mathcal{B}} B_i)\epsilon_0 \\ + T\gamma. \tag{3.51}$$

We analyze the equivocation at E_1 for the given code. Recall that we chose B_1 and B_2 such that $B_1 + B_2 = \phi_{E_1} - \epsilon_1$. Let $\mathcal{I}_1 \subseteq \mathcal{B}$ be such that when computing ϕ_{E_1}, defined in (3.49), the constraint

$$x + \sum_{i\in\mathcal{I}_1} B_i \le R^{\mathrm{G}}_{\mathcal{I}_1\cup\{S\};E_1} - \beta$$

is dominant. This implies that

$$B_1 + B_2 + \sum_{i\in\mathcal{I}_1} B_i = R^{\mathrm{G}}_{\mathcal{I}_1\cup\{S\};E_1} - \beta - \epsilon_1. \tag{3.52}$$

The bounds (3.36) and (3.37) hold without modification. Instead of (3.38), we use (3.50) to obtain

$$H(\mathbf{X}_S,\mathbf{X}_{\mathcal{I}_1}|W, \mathbf{Y}_{E_1}, J_{\mathcal{B}\setminus\mathcal{I}_1}) \le T\big(\frac{1}{T} + (B_1 + B_2 + \sum_{i\in\mathcal{B}} B_i)\epsilon_0\big) + T\gamma. \tag{3.53}$$

Also, instead of (3.39), we use

$$I(\mathbf{X}_S, \mathbf{X}_{\mathcal{I}_1}; \mathbf{Y}_{E_1}|J_{\mathcal{B}\setminus\mathcal{I}_1}) \le T\big(R^{\mathrm{G}}_{\mathcal{I}_1\cup\{S\};E_1} + \alpha\big) \tag{3.54}$$

from Lemma A.6 in Appendix A.11.

Plugging (3.37), (3.53) and (3.54) into (3.36) yields

$$\begin{aligned}\frac{1}{T}H(W|\mathbf{Y}_{E_1}) &\geq R + B_1 + B_2 + \sum_{i\in\mathcal{I}_1} B_i - \epsilon_2\\ &- \frac{2}{T} - (R + B_1 + B_2 + \sum_{i\in\mathcal{B}} B_i)\epsilon_0\\ &- \frac{1}{T} - (B_1 + B_2 + \sum_{i\in\mathcal{B}} B_i)\epsilon_0 - \gamma\\ &- R^{\mathrm{G}}_{\mathcal{I}_1\cup\{S\};E_1} - \alpha\\ &= R - \alpha - \beta - \gamma - \epsilon_1 - \epsilon_2 - \frac{3}{T}\\ &- \big(R + 2(B_1 + B_2 + \sum_{i\in\mathcal{B}} B_i)\big)\epsilon_0, \qquad (3.55)\end{aligned}$$

where we have used (3.52) in the last step. The equivocation analysis for E_2 is done analogously. We conclude that the given code is reliable and the equivocation at either of the eavesdroppers can be made arbitrarily close to $R - \alpha - \beta - \gamma$. This concludes the proof of the proposition.

3.6.5 Proof of Proposition 3.2 for Non-Layered Networks

When the network is not layered, we use the time-expansion technique presented in Appendix A.12. Consider a non-layered acyclic network $\mathcal{G} = (\mathcal{V}, \mathcal{L})$ with a set of noise-inserting nodes $\mathcal{B}$ and a set of eavesdroppers $\mathcal{E}$. We define an unfolded network $\mathcal{G}^{(K)}_{\mathrm{unf}}$ as described in Appendix A.12, with $\mathcal{S} = \mathcal{B} \cup \{S\}$ and $\mathcal{D} = \mathcal{E} \cup \{D\}$. The number of nodes in $\mathcal{G}^{(K)}_{\mathrm{unf}}$ is asymptotically upper bounded by $2K|\mathcal{V}|$. The factor 2 comes into the picture because in the worst case, every node has a virtual node added to it in the unfolded graph (see Appendix A.12). Let R be any transmission rate that satisfies

$$R \leq F^{\mathrm{G}}(\beta),$$

Lemma A.7 and the proof of Section 3.6.3 can be adapted to show that

$$\begin{aligned}F^{\mathrm{unf,G}}(\Delta) \triangleq \max_{B^{\mathrm{unf}}_{\mathcal{B}} \in \cap_{E\in\mathcal{E}} \mathcal{R}^{\mathrm{unf,G}}_{\mathcal{B};E}(\Delta)} \Big[&\max_x \{x : (x, B^{\mathrm{unf}}_{\mathcal{B}}) \in \tilde{\mathcal{R}}^{\mathrm{unf,G}}_{\mathcal{B};D}(\Delta)\}\\ &- \max_x \{x : (x, B^{\mathrm{unf}}_{\mathcal{B}}) \in \cup_{E\in\mathcal{E}} \mathcal{R}^{\mathrm{unf,G}}_{\mathcal{B}\cup\{S\};E}(\Delta)\}\Big]\end{aligned}$$

and $KF^{\mathrm{G}}(\frac{1}{K}\Delta)$ are asymptotically equal with growing K. The sets $\mathcal{R}^{\mathrm{unf,G}}_{\psi;j}(\Delta)$ and $\tilde{\mathcal{R}}^{\mathrm{unf,G}}_{\psi;j}(\Delta)$ above are the Gaussian equivalents of $\mathcal{R}^{\mathrm{unf}}_{\psi;j}(p)$ and $\tilde{\mathcal{R}}^{\mathrm{unf}}_{\psi;j}(p)$ used in Section 3.6.3. It follows that

$$\begin{aligned}KR &\leq KF^{\mathrm{G}}(\beta)\\ &\simeq F^{\mathrm{unf,G}}(K\beta).\end{aligned}$$

The proof for layered networks lets us conclude that for Gaussian signal interaction, the ρ^{unf}-almost perfect secrecy capacity $C_s^{\rho^{\text{unf}}}$ for the unfolded network is lower bounded by $F^{\text{unf,G}}(K\beta)$. Here, we define $\rho^{\text{unf}} = K(\alpha + \beta + \gamma)$, where the terms $K\alpha$ and $K\gamma$ come from the fact that the number of nodes in $\mathcal{G}_{\text{unf}}^{(K)}$ is (asymptotically) at most $2K|\mathcal{V}|$, while the term $K\beta$ follows by using Lemma A.3 instead of Lemma A.2 in the proof of Theorem 3.9. Hence, as K grows large, there is a block coding strategy that achieves the rate-equivocation pair $(KR, KR-\rho^{\text{unf}})$. If the coding scheme on $\mathcal{G}_{\text{unf}}^{(K)}$ is over one block of T transmissions, then we can transform it into a coding scheme on the original network $\mathcal{G}$ that uses TK transmissions. The actual transmission rate is therefore R, and the equivocation is $\frac{1}{K}(KR - \rho^{\text{unf}}) = R - \rho'$, where $\rho' = \alpha + \beta + \gamma$. Hence, this scheme achieves ρ'-almost perfect secrecy on the original graph at rate R. This concludes the proof.

3.6.6 Proof of Theorem 3.3

Let a network with an arbitrary number of noise-inserters be given. Consider a rate $R < C_s$, *i.e.*, such that (R, R) is weakly achievable. Fix an arbitrarily small $\epsilon_1 > 0$. Hence, there exists a communication length T_c and a (T_c, ϵ_1)-code of rate at least $R - \epsilon_1$ and equivocation at least $R - \epsilon_1$. Let $\underline{\mathbf{X}}_{\mathcal{V}}$ and $\underline{\mathbf{Y}}_{\mathcal{V}}$ be the transmit and receive sequences of that particular code. In this proof, sequences (denoted by $\underline{\mathbf{X}}_{\mathcal{V}}$ and $\underline{\mathbf{Y}}_{\mathcal{V}}$) are indexed by the *same* time-slots $1, \dots, T_c$. This is in contrast to the achievability proof in Sections 3.6.2 and 3.6.4, where the sequences $\mathbf{X}_i$ and $\mathbf{X}_j$ can occur during different communication blocks for nodes $i, j \in \mathcal{V}$. Let $\underline{W}$ be the random variable that denotes the message. Consider one particular eavesdropper $E \in \mathcal{E}$. Since the code is ϵ_1-reliable and ϵ_1-perfectly secret, Fano's inequality states that for some ϵ_2,

$$H(\underline{W}|\underline{\mathbf{Y}}_D, \underline{\mathbf{Y}}_E) \leq H(\underline{W}|\underline{\mathbf{Y}}_D) \leq T_c\epsilon_2, \tag{3.56}$$

while

$$H(\underline{W}|\underline{\mathbf{Y}}_E) \geq H(\underline{W}) - T_c\epsilon_1, \tag{3.57}$$

where ϵ_2 goes to zero as ϵ_1 decreases. Using (3.56), (3.57), and the rate constraint $H(\underline{W}) \geq T_c(R - \epsilon_1)$, we obtain

$$\begin{aligned}
T_c(R - 2\epsilon_1 - \epsilon_2) &\leq H(\underline{W}) - T_c\epsilon_1 - T_c\epsilon_2 \\
&\leq H(\underline{W}|\underline{\mathbf{Y}}_E) - T_c\epsilon_2 \\
&\leq H(\underline{W}|\underline{\mathbf{Y}}_E) - H(\underline{W}|\underline{\mathbf{Y}}_D, \underline{\mathbf{Y}}_E) \\
&= I(\underline{W}; \underline{\mathbf{Y}}_D|\underline{\mathbf{Y}}_E) \\
&\leq I(\underline{W}, \underline{\mathbf{X}}_S; \underline{\mathbf{Y}}_D|\underline{\mathbf{Y}}_E) \\
&= I(\underline{\mathbf{X}}_S; \underline{\mathbf{Y}}_D|\underline{\mathbf{Y}}_E) + \underbrace{I(\underline{W}; \underline{\mathbf{Y}}_D|\underline{\mathbf{Y}}_E, \underline{\mathbf{X}}_S)}_{=0},
\end{aligned}$$

where the indicated term is zero because $\underline{W} \multimap \underline{\mathbf{X}}_S \multimap (\underline{\mathbf{Y}}_D, \underline{\mathbf{Y}}_E)$ is a Markov chain. The remainder of the proof is very similar to the proof of Theorem

14.10.1 in [9]. Consider an arbitrary cut $\Omega \in \Lambda(S, \mathcal{B}; D)$, *i.e.*, a cut that separates $\{S\} \cup \mathcal{B}$ from D. For any such cut, we obtain

$$\begin{aligned}
I(\underline{\mathbf{X}}_S; \underline{\mathbf{Y}}_D | \underline{\mathbf{Y}}_E) &\le I(\underline{\mathbf{X}}_\Omega; \underline{\mathbf{Y}}_{\Omega^c} | \underline{\mathbf{Y}}_E) \\
&= \sum_{t=1}^{T_c} I(\underline{\mathbf{X}}_\Omega; \underline{Y}_{\Omega^c}[t] | \underline{Y}_{\Omega^c}[1], \ldots, \underline{Y}_{\Omega^c}[t-1], \underline{\mathbf{Y}}_E) \\
&= \sum_{t=1}^{T_c} \big[H(\underline{Y}_{\Omega^c}[t] | \underline{Y}_{\Omega^c}[1], \ldots, \underline{Y}_{\Omega^c}[t-1], \underline{\mathbf{Y}}_E) \\
&\quad - H(\underline{Y}_{\Omega^c}[t] | \underline{Y}_{\Omega^c}[1], \ldots, \underline{Y}_{\Omega^c}[t-1], \underline{\mathbf{Y}}_E, \underline{\mathbf{X}}_\Omega) \big] \\
&\overset{(a)}{=} \sum_{t=1}^{T_c} \big[H(\underline{Y}_{\Omega^c}[t] | \underline{Y}_{\Omega^c}[1], \ldots, \underline{Y}_{\Omega^c}[t-1], \underline{\mathbf{Y}}_E, \underline{X}_{\Omega^c}[t]) \\
&\quad - H(\underline{Y}_{\Omega^c}[t] | \underline{Y}_{\Omega^c}[1], \ldots, \underline{Y}_{\Omega^c}[t-1], \underline{\mathbf{Y}}_E, \underline{\mathbf{X}}_\Omega, \underline{X}_{\Omega^c}[t]) \big] \\
&\overset{(b)}{\le} \sum_{t=1}^{T_c} \big[H(\underline{Y}_{\Omega^c}[t] | \underline{Y}_E[t], \underline{X}_{\Omega^c}[t]) \\
&\quad - H(\underline{Y}_{\Omega^c}[t] | \underline{Y}_E[t], \underline{X}_\Omega[t], \underline{X}_{\Omega^c}[t]) \big] \\
&= \sum_{t=1}^{T_c} I(\underline{X}_\Omega[t]; \underline{Y}_{\Omega^c}[t] | \underline{Y}_E[t], \underline{X}_{\Omega^c}[t]) \\
&\overset{(c)}{=} T_c \frac{1}{T_c} \sum_{t=1}^{T_c} I(\underline{X}_\Omega[Q]; \underline{Y}_{\Omega^c}[Q] | \underline{Y}_E[Q], \underline{X}_{\Omega^c}[Q], Q = t) \\
&= T_c I(\underline{X}_\Omega[Q]; \underline{Y}_{\Omega^c}[Q] | \underline{Y}_E[Q], \underline{X}_{\Omega^c}[Q], Q) \\
&= T_c \big[H(\underline{Y}_{\Omega^c}[Q] | \underline{Y}_E[Q], \underline{X}_{\Omega^c}[Q], Q) \\
&\quad - H(\underline{Y}_{\Omega^c}[Q] | \underline{X}_\Omega[Q], \underline{Y}_E[Q], \underline{X}_{\Omega^c}[Q], Q) \big] \\
&\le T_c \big[H(\underline{Y}_{\Omega^c}[Q] | \underline{Y}_E[Q], \underline{X}_{\Omega^c}[Q]) \\
&\quad - H(\underline{Y}_{\Omega^c}[Q] | \underline{X}_\Omega[Q], \underline{Y}_E[Q], \underline{X}_{\Omega^c}[Q], Q) \big] \\
&\overset{(d)}{=} T_c \big[H(\underline{Y}_{\Omega^c}[Q] | \underline{Y}_E[Q], \underline{X}_{\Omega^c}[Q]) \\
&\quad - H(\underline{Y}_{\Omega^c}[Q] | \underline{X}_\Omega[Q], \underline{Y}_E[Q], \underline{X}_{\Omega^c}[Q]) \big] \\
&= T_c I(\underline{X}_\Omega[Q]; \underline{Y}_{\Omega^c}[Q] | \underline{Y}_E[Q], \underline{X}_{\Omega^c}[Q]),
\end{aligned}$$

where in (a), both terms remain the same, because all relay nodes in Ω^c use deterministic encoding functions, and hence, $\underline{X}_{\Omega^c}[t]$ is a function of $(\underline{Y}_{\Omega^c}[1], \ldots, \underline{Y}_{\Omega^c}[t-1])$. In (b), the second term remains the same, because $\underline{Y}_{\Omega^c}[t]$ depends only on $(\underline{X}_\Omega[t], \underline{X}_{\Omega^c}[t])$. In (c), we define Q to be a time-sharing variable, uniformly distributed in $\{1, \ldots, T_c\}$. In (d), we use the fact that when Q is given, $\underline{Y}_{\Omega^c}[Q]$ depends only on $\underline{X}_\mathcal{V}[Q]$.

Now, define $\tilde{X}_\mathcal{V} \triangleq \underline{X}_\mathcal{V}[Q]$ and $\tilde{Y}_\mathcal{V} = \underline{Y}_\mathcal{V}[Q]$. Note that the distribution of $\tilde{Y}_\mathcal{V}$ is the distribution of the received symbols of the network when $\tilde{X}_\mathcal{V}$ is transmitted. From the above derivation, we have

$$R \leq I(\tilde{X}_\Omega; \tilde{Y}_{\Omega^c} | \tilde{Y}_E, \tilde{X}_{\Omega^c}) + 2\epsilon_1 + \epsilon_2. \tag{3.58}$$

We repeat the same argument for every choice of Ω and for every $E \in \mathcal{E}$. The distribution of $\tilde{X}_\mathcal{V}$ that we find in the upper bound is the same for every choice of Ω and E. Since this bound holds for every $R < C_s$ and ϵ_1, ϵ_2 can be chosen arbitrarily small, the claim of the theorem follows.

4 Secrecy for the Fan Network

4.1 Introduction

In this section, we study a particular wireless relay network called the fan network. In such a network, the signal sent by a source node can be heard by *all* relay nodes (through different outputs of a broadcast channel). All the relay nodes are then connected to the destination via the so-called "destination channel". This second transmission can only be heard by the destination, *i.e.*, the relays do not create interference at other relays. This situation is illustrated in Figure 4.1. We assume that every eavesdropper can physically access a certain subset of the relay nodes and observe the signals received by these compromised nodes.

One possible scenario in which such a setup would occur is the following. We wish to receive data from a far away source (for instance a satellite), using a device that we plan to leave unattended and hence, the device can potentially be captured by an attacker. To solve this problem, we decide to deploy a number of independent, autonomous devices (the relays) in a certain area. There are two reasons for doing so. First, having several antennas (one per device) spread out over a large area provides a higher capacity from the source to the set of devices than if there was only one antenna. Second, collecting the received data at several independent locations provides protection against attackers that capture devices. To collect the data from all the devices, we either assume the existence of a communication channel (multiple access channel) from all the relays to the destination, or we assume that the legitimate user periodically visits all devices to physically collect the received data.

A more precise problem statement is given in Section 4.2. In Section 4.3, we state our main results, which are the following: an upper bound on the perfect secrecy capacity that is valid for a memoryless broadcast channel from the source and an arbitrary destination channel; a characterization of the perfect secrecy capacity for the case of a perfect destination channel, where the destination has access to the received signals of all the relays; a characterization of the perfect secrecy capacity for a fan network with linear deterministic signal interaction for the special case when the broadcast channel is the bottleneck of the source-destination information flow. The proofs of all the results are given in Section 4.4. For Theorem 4.2, we only give a short outline of the proof, whose details can be found in Appendix B.

4.2 Problem Statement

Consider the network shown in Figure 4.1. A source node S is connected to M relay nodes $\{A_1, \ldots, A_M\}$ through an arbitrary memoryless broadcast channel, *i.e.*, a channel as defined in Definition 2.1, with $\mathcal{X} = \mathcal{X}_S$ and $\mathcal{Y} = (\mathcal{Y}_1, \ldots, \mathcal{Y}_M)$. In general, this channel can be noisy or deterministic, discrete or continuous. Let $\mathcal{A} = \{A_1, \ldots, A_M\}$ be the set of relay nodes. The relay nodes operate independently of each other based on their respective received symbols $Y_A[t]$, $A \in \mathcal{A}$, and forward some information to the destination D via a channel of some sort. We consider two types of such destination channels. The first is the perfect channel, which allows D to directly observe the observations $Y_A[t]$, $A \in \mathcal{A}$. The other is a multiple-access channel (MAC). If the destination channel is a MAC, then the fan network is a wireless relay network as defined in Definition 2.2.

We assume that a class of eavesdroppers $\mathcal{E}$ is present in this network. Each eavesdropper $E \in \mathcal{E}$ has direct access to the received symbols $Y_A[t]$ for all relay nodes A in a certain subset $\mathcal{A}_E \subset \mathcal{A}$.

A formal definition follows.

Definition 4.1 *A **fan network** is defined by a set $\mathcal{A} = \{A_1, \ldots, A_M\}$ of relay nodes, a memoryless broadcast channel, given by alphabets $\mathcal{X}_S$ and $\mathcal{Y}_\mathcal{A}$ and a probability mass function $p_{Y_\mathcal{A}|X_S}$ (or a probability measure $\mu(\cdot|x_S)$), as well as a **destination channel** that is either a set of parallel channels of arbitrarily large rates (perfect destination channel) or a multiple access channel. Finally, the definition of a fan network includes a set $\mathcal{E}$ of eavesdroppers with a corresponding collection $\mathcal{A}_\mathcal{E}$ of subsets of $\mathcal{A}$.*

Since the fan network is layered (see Definition 2.3), we restrict our attention to block codes, defined as follows.

Definition 4.2 *Let $W \in \mathcal{W}$ be a uniformly distributed message observed by S. A (T, ϵ)-**block code** is defined as follows. When the destination channel is a MAC, we use Definition 2.6. If the destination channel is perfect, then a (T, ϵ)-block code is simply given by an encoding function $f_S : \mathcal{W} \to \mathcal{X}_S^T$, as well as a decoding function $f_D : \mathcal{Y}^{MT} \to \mathcal{W}$ that creates an estimate $\hat{W}$ of W based on $\{(Y_A[1], \ldots, Y_A[T])\}_{A \in \mathcal{A}}$. We require the code to be ϵ-**reliable** in the usual sense that $\mathbf{P}(W \neq \hat{W}) \leq \epsilon$.*

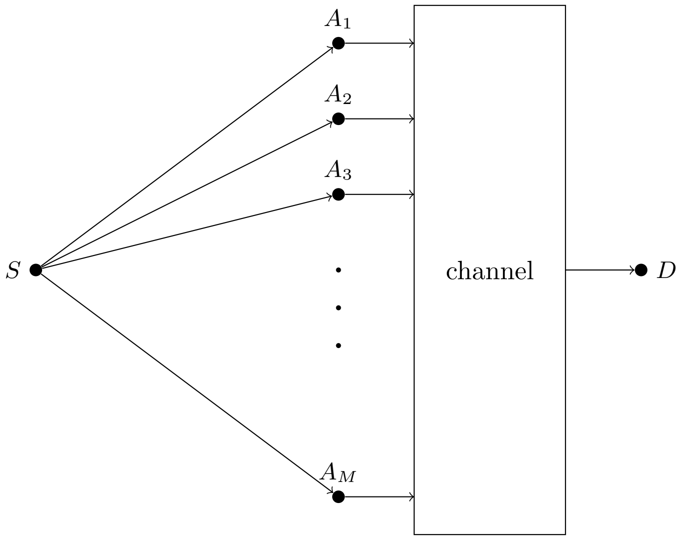

Figure 4.1: A fan network with M relay nodes.

The notions of rate, equivocation, achievability and perfect secrecy capacity are as given in Section 2.2, except that $\mathbf{Y}_E$ is replaced by $\mathbf{Y}_{\mathcal{A}_E}$ and we only consider block codes.

4.3 Main Results

4.3.1 Upper Bound on C_s

In this section, we present an upper bound on the weak perfect secrecy capacity of a fan network. This bound is valid for *any* destination channel.

Theorem 4.1 *For a fan network with an arbitrary memoryless broadcast channel and an arbitrary destination channel, we have*

$$C_s \leq \max_{p_{X_S}} \bigl[I(X_S; Y_{\mathcal{A}}) - \max_{E \in \mathcal{E}} I(X_S; Y_{\mathcal{A}_E}) \bigr].$$

This theorem is proved in Section 4.4. Note that since $\bar{C}_s \leq C_s$ for any network, Theorem 4.1 provides an upper bound on the strong perfect secrecy capacity as well. The bound given in the theorem does not depend on the nature of the destination channel. This is due to the fact that if we replace a given destination channel by the perfect destination channel for any given fan network, we help the destination without helping the eavesdroppers, and hence we increase the secrecy capacity of the network. For a network with a perfect destination channel, the bound in Theorem 4.1 is actually tight.

4.3.2 Perfect Destination Channel

Assume now that the destination channel is perfect, *i.e.*, that D can directly observe $(Y_A[1], \ldots, Y_A[T])$ for every $A \in \mathcal{A}$. As mentioned before, the upper bound in Theorem 4.1 is tight in this case. This fact is formally stated in the following theorem.

Theorem 4.2 *For a fan network with an arbitrary discrete memoryless broadcast channel and a perfect destination channel,*

$$\bar{C}_s = C_s = \max_{p_{X_S}} \big[I(X_S; Y_{\mathcal{A}}) - \max_{E \in \mathcal{E}} I(X_S; Y_{\mathcal{A}_E}) \big].$$

This result can be proved thanks to the absence of interference in this setup, avoiding the question of how relays should cooperate. Furthermore, there is only one transmit distribution p_{X_S} to choose, avoiding any requirements of independence among transmit distributions for different nodes. Note that since the observation of every eavesdropper is a degraded version of the observation at D, this setup is actually a degreded compound wiretap channel, and hence, Theorem 4.2 is a special case of the characterization in [22]. Nevertheless, we still give a detailed proof in Section 4.4.

4.3.3 Deterministic Signal Interaction

Consider a fan network with a deterministic broadcast channel and assume that the destination channel is a deterministic MAC. We have the following lower bound on the strong perfect secrecy rate for this network. We state it as a corollary to Theorem 3.1.

Corollary 4.1 *For a fan network with a deterministic broadcast channel and a deterministic MAC to the destination, we have*

$$\begin{aligned} \bar{C}_s \geq \max_{p_{X_S} \prod_{A \in \mathcal{A}} p_{X_A}} \Big[&\min_{\Omega \in \Lambda(S;D)} I(X_\Omega; Y_{\Omega^c} | X_{\Omega^c}) \\ &- \max_{E \in \mathcal{E}} I(X_S; Y_{\mathcal{A}_E}) \Big] \end{aligned}$$

To prove Corollary 4.1, we only need to show that a deterministic fan network can be viewed as a wireless relay network as defined in Definition 2.2 (without noise-inserting relays). Then, the result of the theorem follows from Theorem 3.1 in Chapter 3. The proof details can be found in Section 4.4.

Note that for deterministic signal interaction, we have $I(X_\Omega; Y_{\Omega^c} | X_{\Omega^c}) = H(Y_{\Omega^c} | X_{\Omega^c})$, because Y_{Ω^c} is a function of $(X_\Omega, X_{\Omega^c}) = X_{\mathcal{V}}$.

Assume now that signal interaction is *linear deterministic* as defined in Definition 2.18. In this case, $\mathcal{X}_S = \mathcal{X}_A = \{0,1\}^q$ for all $A \in \mathcal{A}$. Let us pick $p_{X_S} \prod_{A \in \mathcal{A}} p_{X_A}$ to be uniform over all the input alphabets. One can show that for this particular distribution p,

$$H(Y_{\Omega^c} | X_{\Omega^c}) = \text{rank}\,(\mathbf{G}_{\Omega,\Omega^c}),$$

where $\mathbf{G}_{\Omega,\Omega^c}$ is a binary matrix of dimensions $|\Omega^c|q \times |\Omega|q$. More precisely, $\mathbf{G}_{\Omega,\Omega^c}$ is a block matrix consisting of $|\Omega^c| \cdot |\Omega|$ blocks, each of dimension $q \times q$. Block (i, j) of this block matrix is equal to the channel transfer matrix $\mathbf{G}_{j,i}$, where $j \in \Omega$ and $i \in \Omega^c$ (see Definition 2.18).

Using this fact, we can show the following result.

Theorem 4.3 *Consider a fan network with a MAC destination channel and linear signal interaction. If* $\arg\min_{\Omega\in\Lambda(S;D)} \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}) = \{S\}$*, then*

$$\bar{C}_s = C_s = \operatorname{rank}(\mathbf{G}_{S,\mathcal{A}}) - \max_{E\in\mathcal{E}} \operatorname{rank}(\mathbf{G}_{S,\mathcal{A}_E}).$$

The feasibility proof of Theorem 4.3 follows by plugging a uniform transmit distribution into the lower bound of Corollary 4.1. To show the infeasibility, we evaluate Theorem 4.1 for linear signal interaction. The details are provided in Section 4.4.

The result of Theorem 4.3 is quite intuitive in the following sense. The condition $\arg\min_{\Omega\in\Lambda(S;D)} \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}) = \{S\}$ states that the bottleneck for the communication between S and D is the broadcast channel from S to $\mathcal{A}$. This means that the MAC destination channel is in some sense “stronger” than the broadcast channel. The theorem tells us that if this is the case, then the MAC is actually as strong as a perfect destination channel.

4.4 Proofs of the Main Results

4.4.1 Proof of Theorem 4.1

Consider a rate $R < C_s$, *i.e.*, such that (R, R) is weakly achievable. Fix an arbitrarily small $\epsilon_1 > 0$. It follows that there exists a communication length T and a (T, ϵ_1)-block code of rate at least $R - \epsilon_1$ and equivocation at least $R - \epsilon_1$. Let W be the message transmitted by the code. It follows that for every $E \in \mathcal{E}$,

$$H(W|\mathbf{Y}_{\mathcal{A}_E}) \geq H(W) - T\epsilon_1. \tag{4.1}$$

Let I_D be the information available at D at the end of the communication length T. From the reliability of the code and Fano's inequality, we have that

$$H(W|I_D) \leq T\epsilon_2,$$

where ϵ_2 can be made arbitrarily small by adjusting ϵ_1 accordingly. But $W \multimap \mathbf{Y}_{\mathcal{A}} \multimap I_D$ forms a Markov chain, and hence,

$$H(W|\mathbf{Y}_{\mathcal{A}}) \leq H(W|I_D) \leq T\epsilon_2. \tag{4.2}$$

Fix a certain eavesdropper $E \in \mathcal{E}$. Starting from $H(W) \geq T(R - \epsilon_1)$ and inequalities (4.1) and (4.2), we obtain the following chain of inequalities:

$$\begin{aligned}
TR - T2\epsilon_1 - T\epsilon_2 &\leq H(W) - T\epsilon_1 - T\epsilon_2 \\
&\leq H(W|\mathbf{Y}_{\mathcal{A}_E}) - H(W|\mathbf{Y}_{\mathcal{A}}) \\
&= I(W; \mathbf{Y}_{\mathcal{A}\backslash\mathcal{A}_E}|\mathbf{Y}_{\mathcal{A}_E}) \\
&\leq I(W, \mathbf{X}_S; \mathbf{Y}_{\mathcal{A}\backslash\mathcal{A}_E}|\mathbf{Y}_{\mathcal{A}_E}) \\
&\overset{(a)}{=} I(\mathbf{X}_S; \mathbf{Y}_{\mathcal{A}\backslash\mathcal{A}_E}|\mathbf{Y}_{\mathcal{A}_E}) \\
&= \sum_{t=1}^{T} I(\mathbf{X}_S; Y_{\mathcal{A}\backslash\mathcal{A}_E}[t]|Y_{\mathcal{A}\backslash\mathcal{A}_E}[1], \ldots, Y_{\mathcal{A}\backslash\mathcal{A}_E}[t-1], \mathbf{Y}_{\mathcal{A}_E}) \\
&= \sum_{t=1}^{T} \Big[H(Y_{\mathcal{A}\backslash\mathcal{A}_E}[t]|Y_{\mathcal{A}\backslash\mathcal{A}_E}[1], \ldots, Y_{\mathcal{A}\backslash\mathcal{A}_E}[t-1], \mathbf{Y}_{\mathcal{A}_E}) \\
&\quad - H(Y_{\mathcal{A}\backslash\mathcal{A}_E}[t]|Y_{\mathcal{A}\backslash\mathcal{A}_E}[1], \ldots, Y_{\mathcal{A}\backslash\mathcal{A}_E}[t-1], \mathbf{Y}_{\mathcal{A}_E}, \mathbf{X}_S) \Big] \\
&\overset{(b)}{\leq} \sum_{t=1}^{T} \Big[H(Y_{\mathcal{A}\backslash\mathcal{A}_E}[t]|Y_{\mathcal{A}_E}[t]) \\
&\quad - H(Y_{\mathcal{A}\backslash\mathcal{A}_E}[t]|Y_{\mathcal{A}_E}[t], X_S[t]) \Big] \\
&\overset{(c)}{=} T\frac{1}{T} \sum_{t=1}^{T} I(X_S[Q]; Y_{\mathcal{A}\backslash\mathcal{A}_E}[Q]|Y_{\mathcal{A}_E}[Q], Q = t) \\
&= TI(X_S[Q]; Y_{\mathcal{A}\backslash\mathcal{A}_E}[Q]|Y_{\mathcal{A}_E}[Q], Q) \\
&\leq T\Big[H(Y_{\mathcal{A}\backslash\mathcal{A}_E}[Q]|Y_{\mathcal{A}_E}[Q]) - H(Y_{\mathcal{A}\backslash\mathcal{A}_E}[Q]|Y_{\mathcal{A}_E}[Q], X_S[Q], Q) \Big] \\
&\overset{(d)}{=} T\Big[H(Y_{\mathcal{A}\backslash\mathcal{A}_E}[Q]|Y_{\mathcal{A}_E}[Q]) - H(Y_{\mathcal{A}\backslash\mathcal{A}_E}[Q]|Y_{\mathcal{A}_E}[Q], X_S[Q]) \Big] \\
&= TI(X_S[Q]; Y_{\mathcal{A}\backslash\mathcal{A}_E}[Q]|Y_{\mathcal{A}_E}[Q]),
\end{aligned}$$

where

(a) follows from the Markov chain $W -\!\!\circ\!\!- \mathbf{X}_S -\!\!\circ\!\!- \mathbf{Y}_{\mathcal{A}}$,

(b) holds because since the broadcast channel is memoryless, $Y_{\mathcal{A}\backslash\mathcal{A}_E}[t]$ depends on nothing else once $(X_S[t], Y_{\mathcal{A}_E}[t])$ is known,

(c) uses the time-sharing variable Q, that we define to be uniformly distributed in $\{1, \ldots, T\}$,

(d) is again true because the channel is memoryless and thus, $Y_{\mathcal{A}\backslash\mathcal{A}_E}[Q] -\!\!\circ\!\!- (Y_{\mathcal{A}_E}[Q], X_S[Q]) -\!\!\circ\!\!- Q$ is a Markov chain.

Hence, we have shown that R is upper bounded as

$$R \leq I(X'_S; Y'_{\mathcal{A}\backslash\mathcal{A}_E}|Y'_{\mathcal{A}_E}) + 2\epsilon_1 + \epsilon_2$$

where we have chosen the distribution $p_{X'_S}$ to be such that $X'_S \sim X_S[Q]$ and $Y'_{\mathcal{A}}$ is distributed as the output of the broadcast channel if X'_S is the input. We can repeat this argument for every $E \in \mathcal{E}$, and the random variable X'_S in the upper bound has the same distribution for all E. Hence,

$$\begin{aligned}
R &\leq \min_{E \in \mathcal{E}} I(X'_S; Y'_{\mathcal{A} \setminus \mathcal{A}_E} | Y'_{\mathcal{A}_E}) + 2\epsilon_1 + \epsilon_2 \\
&= I(X'_S; Y'_{\mathcal{A}}) - \max_{E \in \mathcal{E}} I(X'_S; Y'_{\mathcal{A}_E}) + 2\epsilon_1 + \epsilon_2 \\
&\leq \max_{p_{X_S}} \big[I(X_S; Y_{\mathcal{A}}) - \max_{E \in \mathcal{E}} I(X_S; Y_{\mathcal{A}_E}) \big] + 2\epsilon_1 + \epsilon_2,
\end{aligned}$$

where ϵ_1 and ϵ_2 can be chosen arbitrarily small. This completes the proof.

4.4.2 Proof Outline for Theorem 4.2

The proof of Theorem 4.2 follows the same lines as the proof of Theorem 3.1 in Chapter 3. For a fixed distribution p_{X_S}, we sort the eavesdroppers by decreasing $I(X_S; Y_{\mathcal{A}_E})$, *i.e.*, $I(X_S; Y_{\mathcal{A}_{E_1}}) \geq I(X_S; Y_{\mathcal{A}_{E_2}}) \geq \ldots$. Define the rates

$$\begin{aligned}
R &= I(X_S; Y_{\mathcal{A}}) - I(X_S; Y_{\mathcal{A}_{E_1}}) \\
B_1 &= I(X_S; Y_{\mathcal{A}_{E_1}}) - I(X_S; Y_{\mathcal{A}_{E_2}}) \\
B_2 &= I(X_S; Y_{\mathcal{A}_{E_2}}) - I(X_S; Y_{\mathcal{A}_{E_3}}) \\
&\cdots \\
B_K &= I(X_S; Y_{\mathcal{A}_{E_K}}) - \epsilon_1.
\end{aligned}$$

We generate a random encoding function for S using a message W of rate R and K junk messages J_i of respective rates B_i, $i = 1, \ldots, K$. There are no noise inserting nodes in this network, and no further encoding functions. We can show that there exists at least one code (out of all the randomly generated ones) such that D decodes $(W, J_1, \ldots, J_K)$ reliably, and each eavesdropper could decode some subset of the junk messages if it was given W and the remaining junk messages. All of these decoders are joint typicality decoders. We then use Fano's inequality and a local cut-set bound (simpler than the one used in Section 3.6) to lower bound the equivocation at each of the eavesdroppers. For the reader particularly interested in this chapter, we provide a detailed proof of Theorem 4.2 in Appendix B.

4.4.3 Proof of Corollary 4.1

The topology of the fan network with a MAC destination-channel is shown in Figure 4.2. The signal interaction is deterministic as defined in Definition 2.17.

The fact that eavesdropper $E \in \mathcal{E}$ can observe all the received symbols at every node $A \in \mathcal{A}_E$ can be modeled as follows. We create a node in the fan network for each eavesdropper $E \in \mathcal{E}$, and we include these new nodes in the broadcast channel used by the source node S (see Figure 4.2). The broadcast

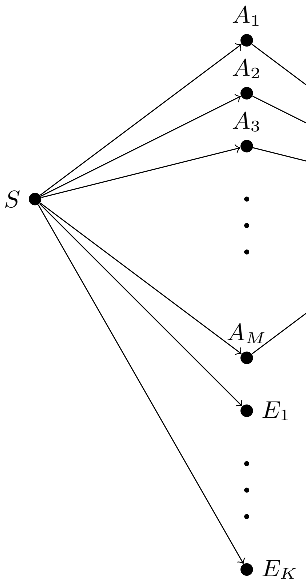

Figure 4.2: The deterministic fan network with a MAC destination-channel, where we assume that $\mathcal{E} = \{E_1, \ldots, E_K\}$ is the set of eavesdroppers.

channel from S to $(\mathcal{A}, \mathcal{E})$ is modeled such that $Y_{\mathcal{A}}$ is obtained from X_S by the given deterministic broadcast channel for this network. For eavesdropper $E \in \mathcal{E}$, we define the alphabet $\mathcal{Y}_E \triangleq \prod_{A \in \mathcal{A}_E} \mathcal{Y}_A$, *i.e.*, $\mathcal{Y}_E$ is the Cartesian product of the alphabets $\{\mathcal{Y}_A\}_{A \in \mathcal{A}_E}$. The channel output at E is then defined to be *equal* to $Y_{\mathcal{A}_E}$, *i.e.*, at every time slot, eavesdropper E observes the *same* symbols as the nodes in $\mathcal{A}_E$. It is then clear that the network shown in Figure 4.2 is an equivalent representation of the deterministic fan network. In addition, the network of Figure 4.2 is a layered wireless relay network as defined in Definitions 2.2 and 2.3 and also contains eavesdropped nodes as defined in Definition 2.9. Hence, Theorem 3.1 from Chapter 3 applies to this network. Let the set of noise-inserters $\mathcal{B}$ be empty. Then, for a given product distribution $p_{X_S} \prod_{A \in \mathcal{A}} p_{X_A}$, the function $F(p)$ defined in Definition 3.4 simplifies to

$$\begin{aligned} F(p) &= \max_x \{x \geq 0 : x \leq R_{S;D}(p)\} \\ &\quad - \max_x \{x \geq 0 : x \leq \max_{E \in \mathcal{E}} R_{E;D}(p)\} \\ &= R_{S;D}(p) - \max_{E \in \mathcal{E}} R_{S;D}(p) \end{aligned}$$

Plugging in the definition of $R_{S;j}(p)$, Theorem 3.1 says that

$$\begin{aligned}\bar{C}_s \geq \max_{p_{X_S}\prod_{A\in\mathcal{A}} p_{X_A}} \Big[&\min_{\Omega\in\Lambda(S;D)} I(X_\Omega; Y_{\Omega^c}|X_{\Omega^c}) \\ &- \max_{E\in\mathcal{E}} \min_{\Omega\in\Lambda(S;E)} I(X_\Omega; Y_{\Omega^c}|X_{\Omega^c})\Big].\end{aligned}$$

To complete the proof, it remains to note that for every $E \in \mathcal{E}$, S and E are directly connected through the broadcast channel, and hence we have

$$\begin{aligned}\min_{\Omega\in\Lambda(S;E)} I(X_\Omega; Y_{\Omega^c}|X_{\Omega^c}) &= I(X_S; Y_E) \\ &= I(X_S; Y_{\mathcal{A}_E}).\end{aligned}$$

4.4.4 Proof of Theorem 4.3

We first show that $\bar{C}_s \geq \operatorname{rank}(G_{S,\mathcal{A}}) - \operatorname{rank}(G_{S,\mathcal{A}_E})$. From Corollary 4.1, we know that for any distribution $p_{X_S}\prod_{A\in\mathcal{A}} p_{X_A}$,

$$\bar{C}_s \geq \min_{\Omega\in\Lambda(S;D)} H(Y_{\Omega^c}|X_{\Omega^c}) - \max_{E\in\mathcal{E}} H(Y_{\mathcal{A}_E}), \tag{4.3}$$

where we used the fact that the channels are deterministic to replace mutual informations by entropies. Choose p to be the uniform distribution on all alphabets $\mathcal{X}_S$ and $\mathcal{X}_A$, $A \in \mathcal{A}$. Then,

$$\begin{aligned}H(Y_{\Omega^c}|X_{\Omega^c}) &= H(G_{\mathcal{V},\Omega^c} X_{\mathcal{V}}|X_{\Omega^c}) \\ &= H(G_{\Omega,\Omega^c} X_\Omega + G_{\Omega^c,\Omega^c} X_{\Omega^c}|X_{\Omega^c}) \\ &= H(G_{\Omega,\Omega^c} X_\Omega|X_{\Omega^c}) \\ &\overset{(a)}{=} H(G_{\Omega,\Omega^c} X_\Omega) \\ &= \operatorname{rank}(G_{\Omega,\Omega^c}),\end{aligned} \tag{4.4}$$

where (a) is true because of the independence of X_i and X_j for $i \neq j$ and the last step follows because the entropy of a set of linear combinations of uniform bits is equal to the number of linearly independent such combinations in the set, which is equal to the rank of the system of equations. Plugging the uniform p into (4.3) and combining it with (4.4) yields

$$\begin{aligned}\bar{C}_s &\geq \min_{\Omega\in\Lambda(S;D)} \operatorname{rank}(G_{\Omega,\Omega^c}) - \max_{E\in\mathcal{E}} \operatorname{rank}(G_{S,\mathcal{A}_E}) \\ &= \operatorname{rank}(G_{S,\mathcal{A}}) - \max_{E\in\mathcal{E}} \operatorname{rank}(G_{S,\mathcal{A}_E}).\end{aligned} \tag{4.5}$$

In the last step, we have used the assumption that $\min_{\Omega\in\Lambda(S;D)} \operatorname{rank}(G_{\Omega,\Omega^c})$ is optimized by $\Omega = \{S\}$.

Now, we show that for linear signal interaction, the upper bound of Theorem 4.1 is equal to (4.5). For any distribution p_{X_S}, we have

$$I(X_S; Y_{\mathcal{A}}) - \max_{E\in\mathcal{E}} I(X_S; Y_{\mathcal{A}_E}) = \min_{E\in\mathcal{E}} I(X_S; Y_{\mathcal{A}\setminus\mathcal{A}_E}|Y_{\mathcal{A}_E}). \tag{4.6}$$

Further, for any $E \in \mathcal{E}$, we can write

$$\begin{aligned} I(X_S; Y_{\mathcal{A}\setminus\mathcal{A}_E}|Y_{\mathcal{A}_E}) &= H(Y_{\mathcal{A}\setminus\mathcal{A}_E}|Y_{\mathcal{A}_E}) \\ &= H(Y_{\mathcal{A}}|Y_{\mathcal{A}_E}) \\ &= H(G_{S,\mathcal{A}}X_S|G_{S,\mathcal{A}_E}X_S) \\ &\leq \operatorname{rank}(G_{S,\mathcal{A}}) - \operatorname{rank}(G_{S,\mathcal{A}_E}), \end{aligned} \tag{4.7}$$

where the last step is true for the following reason. The matrix $G_{S,\mathcal{A}}$ is of dimension $|\mathcal{A}|q \times q$ and multiplies a binary random vector X_S of length q. Hence, $V \triangleq G_{S,\mathcal{A}}X_S$ is a binary vector of length $|\mathcal{A}|q$ that contains at most q bits of information. If the information content of V is less than q bits, this means that $G_{S,\mathcal{A}}$ projects X_S onto a space of dimension $\operatorname{rank}(G_{S,\mathcal{A}}) < q$. The quantity that we are interested in is $H(V|G_{S,\mathcal{A}_E}X_S)$, where $\mathcal{A}_E \subset \mathcal{A}$. Thus, the conditioning fixes the values of $|\mathcal{A}_E|q$ bits in V, hence leaving us with $\operatorname{rank}(G_{S,\mathcal{A}}) - \operatorname{rank}(G_{S,\mathcal{A}_E})$ degrees of freedom. This proves the inequality in (4.7). The inequality is an equality if X_S is uniformly distributed in $\{0,1\}^q$.

Combining (4.7), (4.6) and Theorem 4.1, we have

$$\begin{aligned} C_s &\leq \max_{p_{X_S}} \min_{E\in\mathcal{E}} H(G_{S,\mathcal{A}}X_S|G_{S,\mathcal{A}_E}X_S) \\ &\leq \min_{E\in\mathcal{E}} \big[\operatorname{rank}(G_{S,\mathcal{A}}) - \operatorname{rank}(G_{S,\mathcal{A}_E})\big] \\ &= \operatorname{rank}(G_{S,\mathcal{A}}) - \max_{E\in\mathcal{E}} \operatorname{rank}(G_{S,\mathcal{A}_E}). \end{aligned}$$

Hence, since strong secrecy implies weak secrecy, we have

$$\begin{aligned} \operatorname{rank}(G_{S,\mathcal{A}}) - \max_{E\in\mathcal{E}} \operatorname{rank}(G_{S,\mathcal{A}_E}) \leq &\bar{C}_s \leq C_s \\ &\leq \operatorname{rank}(G_{S,\mathcal{A}}) - \max_{E\in\mathcal{E}} \operatorname{rank}(G_{S,\mathcal{A}_E}), \end{aligned}$$

which concludes the proof of the theorem.

Interactive Secrecy for the Line Network 5

5.1 Introduction

In this chapter, we assume that the destination can provide feedback to the source over a backward channel. This assumption is motivated by the fact that in most wireless networks, communication is indeed bidirectional. For simplicity, we assume that the backward link is perfectly reliable (a "bit-pipe"), of unlimited rate and that it is public. In other words, the messages sent over this channel are observed directly by the source and the eavesdropper. We assume that this public channel can only be used in the backward direction, *i.e.*, to send messages from the destination to the source. This model differs from the more general public discussion model used in previous literature.

In this chapter, we consider the problem of secret key agreement rather than secret communication. In secret key agreement, two parties (the source and the destination) wish to agree, with high probability, on a sequence of bits (the "key") which is kept secret from the eavesdropper. The key rate is the entropy in bits of the key, divided by the number of channel uses that are necessary to establish the key. This notion is weaker than the notion of secret communication introduced in Section 2.2, because of the following. Every secret communication protocol of secrecy rate R_s provides a shared "key" (the message) of key rate R_s. In contrary, if we have a secret key agreement protocol that provides a key of rate R_k, we cannot necessarily communicate a secret message of the same rate from the source to the destination. Of course, if a channel of non-zero rate from the source to the destination exists (for instance a wireless relay network), then we can achieve secret communication by one-time padding the key with the message and sending the resulting cypher-text reliably over the channel. However, this implies further uses of the channel (or the network), hence lowering the overall rate of the secret message.

In [28], Maurer demonstrated the use of public feedback for secret key agreement over the wiretap channel. He showed that if two users Alice and Bob, and an eavesdropper Eve each observe one component of a discrete memoryless multisource, then a public discussion channel can be used to establish a secret key shared by the source and the destination. Further, he observed that such a situation can be created in the wiretap channel setup [48] as follows. The source generates a random i.i.d. sequence $\mathbf{X}$ from some distribution p_X and transmits it over the wiretap channel $p_{YZ|X}$. Let $(\mathbf{Y}, \mathbf{Z})$ be the received sequences at the destination and the eavesdropper, respectively. The sequence triple $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ has the same distribution as a sample sequence from the memoryless multisource governed by p_{XYZ}. A key generation protocol can now be applied to this situation, with the source in the role of Alice and the destination in the role of Bob. A generalization of this work was provided by Ahlswede and Csiszár in [1].

The main new difficulty in proving the existence of feedback-utilizing protocols for wireless networks stems from communication via relays. In the attempt to create correlated sequences at all three ends of the setup, the source can transmit an i.i.d. sequence of symbols over the network, as it was done in the protocols of [28] and [1]. However, since the transmission happens over a relay network, the relays need to map from their received signals to their transmit signals. We analyze a decode-and-forward strategy for the relays. Depending on the signal rate, the whole typical transmit set may not be used by the forwarding relays. 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 [28, 1] are not directly 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. The backward transmission uses a structured binning scheme similar to the techniques proposed in [28, 1]. To be able to apply this scheme, we first compute the size of the list of all potential received sequences at the destination from the point of view of the eavesdropper. This computation is a result that might be of a more general use. For comparison, we derive the secret key capacity for the case when no backward channel is available.

A more detailed problem statement is given in Section 5.2. The main results are stated in Section 5.3 and proved in Section 5.4. Section 5.5 contains the list size result.

5.2 Problem Statement

In this chapter, we focus on the line network, which is essentially a concatenation of a point-to-point channel and a wiretap channel (see Figure 5.1).

Definition 5.1 *The* ***line network*** *consists of a source S, a destination D, only one relay A and only one eavesdropping node E. The network is defined by alphabets $\mathcal{X}_S$, $\mathcal{Y}_A$, $\mathcal{X}_A$, $\mathcal{Y}_D$ and $\mathcal{Y}_E$, and two discrete memoryless channels $p_{Y_A|X_S}$, $p_{Y_D,Y_E|X_A}$. The relay node A plays the role of connecting the channel $p_{Y_A|X_S}$ to the channel $p_{Y_D,Y_E|X_A}$ by mapping from its receive alphabet to its transmit alphabet. This mapping can be done block-wise and the mapping can also be random i.e., A is a noise-inserting relay.*

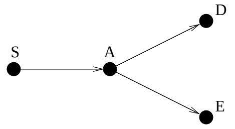

Figure 5.1: The line network.

Definition 5.2 *The* ***line network with feedback*** *is a line network, augmented with a public feedback channel over which D can transmit messages of arbitrarily large rate to S. These messages are received reliably at S* and *at E,* i.e.*, the feedback channel is public.*

As explained in the introduction of this chapter, we focus on secret key agreement rather than secret communication. In secret key agreement, we want to ensure that S and D agree with high probability on the same key 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 5.3 *For a wireless relay network without feedback, we say that a* ***secret key rate*** *R_k is* ***weakly achievable*** *if the rate-equivocation pair (R_k, R_k) is weakly achievable for secret communication. Weak achievability for secret communication has been defined in Definition 2.11. For a wireless relay network without feedback, we say that a* ***secret key rate*** *R_k is* ***strongly achievable*** *if the secrecy rate R_k is strongly achievable for secret communication. Strong achievability for secret communication has been defined in Definition 2.13.*

In other words, for a wireless relay network without feedback, the notions of secret communication rate and secret key rate are equivalent. Note that since the line network is layered (Definition 2.3), we only consider block codes.

In the presence of a public feedback channel, we use a different definition:

Definition 5.4 *A (T,ϵ)-**protocol for key generation** over a layered 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 (possibly) probabilistic 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 $\{1,\dots,2^{TR_k}\}$, where R_k is the 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 5.5 *Given a (T,ϵ)-protocol for key generation, we say that the protocol provides **weakly secret key generation** if*

$$H(K|\mathbf{Y}_E, I) \geq H(K) - T\epsilon,$$

*where K is the key, $\mathbf{Y}_E$ is the sequence of observations at eavesdropper E, and I is the public feedback. We say that the protocol provides **strongly secret key generation** if*

$$H(K|\mathbf{Y}_E, I) \geq H(K) - \epsilon.$$

Definition 5.6 *For a network with feedback, we say that a secret key rate R_k is **weakly/strongly achievable** if for any $\epsilon > 0$, there exists an integer T and a weakly/strongly secret key generation (T,ϵ)-protocol that has key rate R_k.*

Definition 5.7 *The strong **secret key capacity** $\bar{C}_k$ of a wireless network with or without feedback is defined as*

$$\bar{C}_k \triangleq \max\{R_k : R_k \textit{ is a strongly achievable secret key rate}\}.$$

5.3 Main Results

First, we consider the line network without feedback.

Definition 5.8 *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 5.1 *In the line network **without feedback**, a rate-equivocation pair (R, R_e) is weakly achievable if and only if it lies in the region*

$$\breve{\mathcal{R}} = \cup_{\mathcal{V}_A, p_{X_S} p_{V_A} p_{X_A|V_A}} \breve{\mathcal{R}}(\mathcal{V}_A, p), \tag{5.1}$$

where the union is over all choices of the alphabet $\mathcal{V}_A$ and over distributions on $\mathcal{X}_S \times \mathcal{V}_A \times \mathcal{X}_A$ such that X_S and (V_A, X_A) are independent and $V_A \multimap X_A \multimap (Y_D, Y_E)$ forms a Markov chain.

Using Caratheodory's theorem [11], 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 (5.1). The proof of Theorem 5.1 can be found in Section 5.4.

The following corollary follows from Theorem 5.1:

Corollary 5.1 *The strongly secret key capacity of the line network without feedback is*

$$\begin{aligned}\bar{C}_k = \max_{\mathcal{V}_A, p_{X_S} p_{V_A} p_{X_A|V_A}} & \min\{I(X_S;Y_A), \\ & [I(V_A;Y_D) - I(V_A;Y_E)]^+\}. \qquad (5.2)\end{aligned}$$

Proof: Theorem 5.1 tells us that every rate-equivocation pair (R, R_e) in $\breve{\mathcal{R}}$ is weakly achievable. Hence, it follows from Lemma E.5 that $R_k = R_e$ is strongly achievable over the network. Since a secretly communicated message is a secret key, $\bar{C}_k$ is lower bounded by the expression given in the corollary. To show that that expression is also an upper bound on $\bar{C}_k$, we note that for a network without feedback, $\bar{C}_k = \bar{C}_s$, which is upper bounded by the weak secrecy capacity C_s. The upper bound on C_s follows from Theorem 5.1. □

Note that Theorem 5.1 and Corollary 5.1 give necessary and sufficient conditions for secret communication and secret key agreement over the line network without feedback. The following theorem gives sufficient conditions for secret key agreement over the line network *with* feedback.

Definition 5.9 *For a given distribution p_{X_A} and a number u, let the function $\tilde{F}(u, p_{X_A})$ be defined as*

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

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

Theorem 5.2 *In the line network **with feedback**, a secret key rate R_k is strongly achievable if it satisfies*

$$R_k < \max_{p_{X_S} p_{X_A}} \tilde{F}(I(X_S;Y_A), p_{X_A}). \qquad (5.3)$$

The proof of Theorem 5.2 can be found in Section 5.4. We immediately obtain the following corollary:

Corollary 5.2 *Let*

$$C_{SA} \triangleq \max_{p_{X_S}} I(X_S;Y_A)$$

be the point-to-point capacity of the channel from S to A. A secret key rate R_k is strongly achievable if it satisfies

$$R_k < \max_{p_{X_A}} \tilde{F}(C_{SA}, p_{X_A}).$$

Proof: Consider the optimization problem (5.3). Note that the distribution p_{X_S} influences $\tilde{F}(I(X_S;Y_A), p_{X_A})$ only through its first argument. Moreover, $\tilde{F}(u, p_{X_A})$ is non-decreasing in u for any choice of p_{X_A}. It follows that to maximize $\tilde{F}(I(X_S;Y_A), p_{X_A})$ over a range of distributions $p_{X_S}p_{X_A}$, we can first maximize $I(X_S;Y_A)$ over all possible p_{X_S}, and then maximize $\tilde{F}(C_{SA}, p_{X_A})$ over all p_{X_A}. Hence, the two expressions $\max_{p_{X_S}p_{X_A}} \tilde{F}(I(X_S;Y_A), p_{X_A})$ and $\max_{p_{X_A}} \tilde{F}(C_{SA}, p_{X_A})$ are actually equal, which proves the corollary. □

Assume that the channel from X_A to (Y_D, Y_E) is such that $Y_D \multimap X_A \multimap Y_E$ is a Markov chain. Consider a modified setup for which $S = A$. This is the setup studied in [1] and it was shown there and in [29] that the strongly secret key capacity of this channel is

$$\bar{C}_k' = \max_{p_{X_A}} [I(X_A;Y_D) - I(Y_D;Y_E)]. \tag{5.4}$$

Let $\bar{C}_k$ be the secret key capacity of the line network with feedback. It is clear that $\bar{C}_k'$ is an upper bound on $\bar{C}_k$, and that $\bar{C}_k'$ is achievable in the line network with feedback if C_{SA} is infinite. Let X_A' be a random variable whose distribution maximizes (5.4) and let Y_D' and Y_E' be the corresponding outputs of the channel $p_{Y_D,Y_E|X_A}$. In the scheme described in [1], the source node (S or A in our notation) generates an i.i.d. sequence $\mathbf{X}_A$ from the distribution of X_A' and sends it over the channel. As a result, A, D and E have access to correlated memoryless sequences $\mathbf{X}_A$, $\mathbf{Y}_D$ and $\mathbf{Y}_E$, respectively. The public message over the feedback link is then used to establish a shared secret key between A and D based on these correlated sequences. Instead of generating $\mathbf{X}_A$ in an i.i.d. manner, we can also select $\mathbf{X}_A$ uniformly at random from the set of typical sequences $\mathcal{T}_\delta(X_A')$ (see Appendix F). If we do this, the statistics of the resulting $(\mathbf{X}_A, \mathbf{Y}_D, \mathbf{Y}_E)$ are with high probability the same as before. In the line network, we can select such a typical sequence $\mathbf{X}_A$ at S and describe it losslessly to A if $C_{SA} \geq H(X_A')$, hence achieving $\bar{C}_k'$. Surprisingly, Corollary 5.2 states that whenever $C_{SA} > I(X_A'; Y_D, Y_E)$, we can achieve a secret key rate of $I(X_A';Y_D') - I(Y_D';Y_E')$, which is the secret key capacity $\bar{C}_k'$ given in (5.4). Hence, $\bar{C}_k'$ is achievable in the line network even if $C_{SA} < H(X_A')$, which is unexpected. The scheme that is used to achieve this secret key rate is described in detail in Section 5.4.

The following proposition compares the achievable secret key rate with feedback given in (5.3) and the secret key capacity without feedback given in (5.2).

Proposition 5.1 *Consider a line network with a conditionally independent broadcast channel, i.e., such that $p_{Y_D,Y_E|X_A} = p_{Y_D|X_A}\, p_{Y_E|X_A}$. Let X_S^* and X_A^* be such that $p_{X_S^*}p_{X_A^*}$ maximizes (5.3), let Y_A^*, Y_D^* and Y_E^* be the corresponding channel outputs, and let*

$$R_k^* = \tilde{F}(I(X_S^*;Y_A^*), p_{X_A^*})$$

be the largest secret key rate guaranteed by (5.3). Then, the secret key rate

$$\tilde{R} \triangleq \min\{I(X_S^*;Y_A^*), [I(X_A^*;Y_D^*) - I(X_A^*;Y_E^*)]^+\} \tag{5.5}$$

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*

$$\tilde{R} \geq R_k^*. \tag{5.6}$$

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 (5.2). Hence, when ignoring the feedback channel,

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

is achievable, because it is smaller than the secret key capacity given in (5.2). Assume that $I(X_S^*;Y_A^*) \leq I(X_A^*;Y_D^*)$. Then, it follows from (5.3) 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 $\tilde{R} \geq I(X_S^*;Y_A^*) - I(X_A^*;Y_E^*)$. From the non-negativity of the mutual information, we also have $\tilde{R} \geq 0$. Hence, $\tilde{R} \geq R_k^*$. □

The example in Figure 5.2 shows that the inequality in (5.6) can be strict, implying that the achievable secret key rate given in (5.3) is not optimal. The example also shows how the two schemes can be combined via time-sharing to obtain a secret key rate that is higher than either of the individual rates.

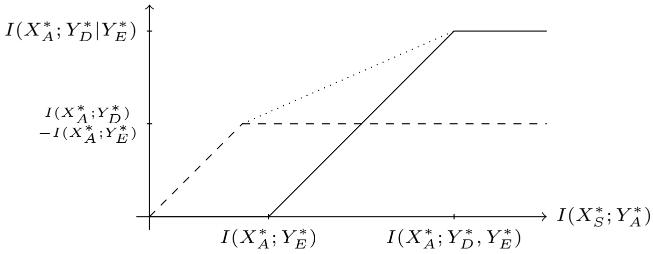

Figure 5.2: Illustration of Proposition 5.1. R_k^* (solid line) and $\tilde{R}$ (dashed line) are plotted as functions of the mutual information between S and A. The dotted line shows the convex envelope of the two functions, which is achievable via time-sharing between the two schemes.

5.4 Proof Outlines

5.4.1 Proof of Theorem 5.1

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

$$R < I(X_S;Y_A), \tag{5.7}$$

$$R < I(V_A;Y_D), \tag{5.8}$$

$$R_e \leq R, \tag{5.9}$$

$$R_e < [I(V_A;Y_D) - I(V_A;Y_E)]^+. \tag{5.10}$$

From the usual channel coding theorem [11], we know that if (5.7) 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 [10], we know that as long as (5.8) and (5.10) 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 arbitrarily close to R_e.

Converse Assume that (R, R_e) are achievable. Then, from the channel coding theorem [11], 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,

$$(R, R_e) \in \cup_{p_{X_S}} \mathcal{A}(p_{X_S}), \tag{5.11}$$

where $\mathcal{A}(p_{X_S}) = \{(R, R_e) : R < I(X_S; Y_A), R_e \leq R\}$. From the converse proof in [10], we know that

$$(R, R_e) \in \cup_{\mathcal{V}_A, p_{V_A} p_{X_A|V_A}} \mathcal{B}(\mathcal{V}_A, p_{V_A} p_{X_A|V_A}), \tag{5.12}$$

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 (5.11) and (5.12), it follows that

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

5.4.2 Proof of Theorem 5.2

Let $p_{X_S} p_{X_A}$ be fixed. Consider the following protocol: S generates a random message of rate

$$\tilde{R} = \min\{I(X_S; Y_A) - \epsilon_1, H(X_A)(1 - \delta)\}$$

and sends it over the channel $p_{Y_A|X_S}$ using a good channel code. The constants $\epsilon_1, \delta > 0$ can be chosen arbitrarily. Then, it follows from the channel coding theorem [9] 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 $\mathcal{C} \subseteq \mathcal{T}_\delta(X_A)$ of rate $\tilde{R}$, where $\mathcal{T}_\delta(X_A)$ denotes the set of all (robustly) typical sequences $\mathbf{x}_A$ (see Appendix F). The codebook $\mathcal{C}$ is generated by picking sequences from $\mathcal{T}_\delta(X_A)$ uniformly at random. Once generated, $\mathcal{C}$ is fixed for all time.

Definition 5.10 *Define $\mathcal{L}_\delta(Y_j)$, $j \in \{D, E\}$ 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 $(\mathbf{x}_A, \mathbf{y}_j) \in \mathcal{T}_\delta(X_A, Y_j)$. Similarly, define $\mathcal{L}_\delta(Y_D|\mathbf{y}_E)$ as the list of all sequences $\mathbf{y}_D$ such that there is at least one $\mathbf{x}_A \in \mathcal{C}$ for which $(\mathbf{x}_A, \mathbf{y}_D, \mathbf{y}_E) \in \mathcal{T}_\delta(X_A, Y_D, Y_E)$.*

This definition is equivalent to Definition 5.12 in the next section of this chapter.

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

For the backward transmission, we are facing a source coding problem similar to the one in [28] and [1]. 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 5.11 *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 [28] and [1], 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 line 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. Consequently, assuming $\mathbf{X}_A = \mathbf{x}_A$, we obtain

$$2^{T\alpha_S} = |\mathcal{T}_\delta(Y_D|\mathbf{x}_A)| \simeq 2^{TH(Y_D|X_A)}, \tag{5.13}$$

which is the same for all $\mathbf{x}_A$. The approximate equality can be made arbitrarily precise by choosing δ small enough (see Lemma F.2). Note that the sequences $\mathbf{X}_A$, $\mathbf{Y}_D$ and $\mathbf{Y}_E$ are highly likely to be jointly δ-typical and thus, with high probability, $\mathbf{Y}_E$ lies in $\mathcal{L}_\delta(Y_E)$ and

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

Using Lemma 5.3 and identity (5.19) in the next section, we obtain

$$\alpha_E \simeq \begin{cases} H(Y_D|X_A, Y_E) & \text{if } 0 < \tilde{R} < I(X_A; Y_E) \\ \left.\begin{smallmatrix} \tilde{R} - I(X_A; Y_E) \\ + H(Y_D|X_A, Y_E) \end{smallmatrix}\right\} & \text{if } I(X_A; Y_E) < \tilde{R} < I(X_A; Y_D, Y_E) \\ H(Y_D|Y_E) & \text{if } I(X_A; Y_D, Y_E) < \tilde{R} < H(X_A), \end{cases} \tag{5.14}$$

where the approximate equality can be made arbitrarily precise by choosing δ sufficiently small.

We conclude the proof of the theorem from the following lemma.

Lemma 5.1 *Through public transmission by D, any secret key rate R_k that satisfies*

$$R_k < \alpha_E - \alpha_S \tag{5.15}$$

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

Evaluating $\alpha_E - \alpha_S$ for α_S and α_E as given by (5.13) and (5.14) and noting that R_k should be non-negative implies that the rate range stated in the theorem is weakly achievable. It is easy to see that Lemma E.5 in Appendix E can also be applied to the problem of secret key agreement, which lets us conclude the strong achievability of that rate range.

It remains to prove Lemma 5.1:

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 $\mathbf{y}_D \in \mathcal{L}_\delta(Y_D)$ is thrown into a randomly (uniformly) chosen bin, indexed by $i \in \{1, \dots, 2^{T(\alpha_S+\epsilon_2)}\}$.

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

Let $\mathbf{Y}_D$ be the received sequence from the forward transmission over the network. The sequence $\mathbf{Y}_D$ lies in $\mathcal{L}_\delta(Y_D)$ with high probability. 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 $\hat{\mathbf{Y}}_D \in \mathcal{L}_\delta(Y_D|\mathbf{x}_S)$. The dominant error event is that $\hat{\mathbf{Y}}_D$ is not unique in bin I.

Virtual Decoding at the Eavesdropper: If a genie made K available at E, then E would know (I, K). It could then try to identify a unique sequence $\hat{\hat{\mathbf{Y}}}_D$ in sub-bin K of bin I such that $\hat{\hat{\mathbf{Y}}}_D \in \mathcal{L}_\delta(Y_D|\mathbf{y}_E)$. The dominant failure 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 since the number of bins and the number of sub-bins are chosen large enough, there exists a binning scheme for which

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

and

$$\frac{1}{2^{T(\alpha_E-\alpha_S)}} \sum_{k=1}^{2^{T(\alpha_E-\alpha_S)}} \mathbf{P}(A_k) \leq \epsilon_0, \tag{5.16}$$

where A_k is the error event of the virtual (genie-aided) decoder at the eavesdropper for a given $K = k$. Note that the left hand side of (5.16) is the error probability when estimating $\mathbf{Y}_D$ from the observable $(K, I, \mathbf{Y}_E)$. Thus, from Fano's inequality [9], we have

$$H(\mathbf{Y}_D|K, I, \mathbf{Y}_E) \leq T\epsilon_3, \tag{5.17}$$

where ϵ_3 can be made arbitrarily small by choosing ϵ_0 small enough.

Equivocation Analysis: We have

$$\begin{aligned} H(K|I, \mathbf{Y}_E) &\geq I(\mathbf{Y}_D; K|I, \mathbf{Y}_E) \\ &= I(\mathbf{Y}_D; K, I|\mathbf{Y}_E) - I(\mathbf{Y}_D; I|\mathbf{Y}_E) \\ &= H(\mathbf{Y}_D|\mathbf{Y}_E) - H(\mathbf{Y}_D|K, I, \mathbf{Y}_E) \\ &\quad - \underbrace{H(I|\mathbf{Y}_E)}_{\leq H(I)} + \underbrace{H(I|\mathbf{Y}_D, \mathbf{Y}_E)}_{\geq 0} \\ &\geq H(\mathbf{Y}_D|\mathbf{Y}_E) - H(I) - H(\mathbf{Y}_D|K, I, \mathbf{Y}_E) \qquad (5.18) \\ &\geq T\big(\alpha_E - (\alpha_S + \epsilon_2) - \epsilon_3\big), \end{aligned}$$

where we have used (5.17) to bound the last term. For the second term, we used the fact that $I \in \{1, \ldots, 2^{T(\alpha_S + \epsilon_2)}\}$, and hence $H(I) \leq T(\alpha_S + \epsilon_2)$. For the first term in (5.18) we have used

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

where $\simeq$ denotes approximate equality for large T. This approximation of $H(\mathbf{Y}_D|\mathbf{Y}_E = \mathbf{y}_E)$ in (a) follows from the uniformity stated in Lemma 5.4. Thus, the equivocation can be made arbitrarily close to $\alpha_E - \alpha_S$.

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

Hence, our scheme achieves weak secrecy of the key K, at a rate R_k as stated in the Lemma. □

5.5 List Size Lemmas

In this section, we present several lemmas that are used in Section 5.4. The proofs of all these lemmas can be found in Appendix C.

Consider discrete alphabets $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ and a joint distribution $p_{XYZ}(x, y, z)$ on $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. Let $\mathcal{C}$ be a randomly generated X-codebook of size $2^{T\tilde{R}}$, where each codeword has length T and was picked uniformly at random from $\mathcal{T}_\delta(X)$.

Definition 5.12 *We define the following sets, or "lists":*

- $\mathcal{L}_\delta(Y)$ *is the list of all sequences* $\mathbf{y}$ *such that there is at least one* $\mathbf{x} \in \mathcal{C}$ *for which* $(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_\delta$.
- $\mathcal{L}_\delta(Z)$ *is the list of all sequences* $\mathbf{z}$ *such that there is at least one* $\mathbf{x} \in \mathcal{C}$ *for which* $(\mathbf{x}, \mathbf{z}) \in \mathcal{T}_\delta$.
- *For a given* $\mathbf{z} \in \mathcal{L}_\delta(Z)$*, we define* $\mathcal{L}_\delta(X|\mathbf{z})$ *as the list of all codewords* $\mathbf{x} \in \mathcal{C}$ *such that* $(\mathbf{x}, \mathbf{z}) \in \mathcal{T}_\delta$.
- *For a given* $\mathbf{z} \in \mathcal{L}_\delta(Z)$*, we define* $\mathcal{L}_\delta(Y|\mathbf{z})$ *as the list of all sequences* $\mathbf{y} \in \mathcal{T}_\delta(Y|\mathbf{z})$ *such that there exists at least one codeword* $\mathbf{x} \in \mathcal{C}$ *for which* $(\mathbf{x}, \mathbf{y}, \mathbf{z}) \in \mathcal{T}_\delta$.

The size of a list $\mathcal{L}_\delta(\cdot)$ *is denoted as* $L_\delta(\cdot) \triangleq |\mathcal{L}_\delta(\cdot)|$.

Intuitively, the meaning of these lists is the following. If the codebook $\mathcal{C}$ is used for transmission over the channel $p_{YZ|X}(y, z|x)$, then we know from Lemma F.2 that the received sequences $\mathbf{y}$ and $\mathbf{z}$ are, with high probability, such that $(\mathbf{x}, \mathbf{y}, \mathbf{z}) \in \mathcal{T}_{\delta'}$ for some $\delta' > \delta$. It follows that in this setup, $\mathcal{L}_{\delta'}(Y)$ and $\mathcal{L}_{\delta'}(Z)$ are the lists of possible received sequences at Y and Z, respectively, given that we do not know which codeword was transmitted. For a given received sequence $\mathbf{z}$, $\mathcal{L}_{\delta'}(X|\mathbf{z})$ is the list of all codewords that seem plausible from the point of view of Z. Likewise, $\mathcal{L}_{\delta'}(Y|\mathbf{z})$ is the list of all sequences $\mathbf{y}$ that are likely to be received at Y from the point of view of Z.

Before stating the first lemma, we define the notion of exponential equality.

Definition 5.13 *We say that two expressions* A *and* B *that depend on the variable* T *are* ***δ-exponentially equal****, denoted by* $A \overset{\delta}{=} B$ *if asymptotically with large* T*, we have*

$$B2^{-\delta T} \leq A \leq B2^{\delta T}.$$

Note that δ-exponential equality is reciprocal, *i.e.*, if $A \overset{\delta}{=} B$, then $B \overset{\delta}{=} A$. For example, Lemma F.1 states that $|\mathcal{T}_\delta(X)| \overset{\delta'}{=} 2^{TH(X)}$, where $\delta' = \delta H(X)$. Instead of "δ-exponentially equal", we also use the shorter term "δ-equal".

Lemma 5.2 *For any* $\delta > 0$*, we have with high probability (over the random codebook* $\mathcal{C}$*) and for almost every* $\mathbf{z} \in \mathcal{L}_\delta(Z)$ *that the size of* $\mathcal{L}_\delta(X|\mathbf{z})$ *is*

$$L_\delta(X|\mathbf{z}) \begin{cases} = 1 & \text{if } \tilde{R} < I(X;Z) - \delta_1 \\ \overset{\delta_1}{=} 2^{T(\tilde{R} - I(X;Z))} & \text{if } I(X;Z) + \delta_1 < \tilde{R} < H(X) - \delta, \end{cases}$$

where $\delta_1 = \delta 2H(X)$.

By "almost every $\mathbf{z} \in \mathcal{L}_\delta(Z)$" we mean a subset of $\mathcal{L}_\delta(Z)$ whose size is ϵ-exponentially equal to $L_\delta(Z)$, where ϵ can be arbitrarily small.

The second lemma provides arbitrarily tight bounds on the size of the list $\mathcal{L}_\delta(Y|\mathbf{z})$:

Lemma 5.3 *For any $\delta > 0$ such that $\delta < \frac{H(Y|X)}{6H(X)+2H(Y|Z)}$, the following is true. For large T, with high probability over the random codebook $\mathcal{C}$, and for almost all $\mathbf{z} \in \mathcal{L}_\delta(Z)$, the list $\mathcal{L}_\delta(Y|\mathbf{z})$ is of size*

$$L_\delta(Y|\mathbf{z}) \begin{cases} \overset{\delta_2}{=} 2^{TH(Y|X,Z)} & \text{if } \tilde{R} < I(X;Z) - \delta_1 \\ \overset{\delta_4}{=} 2^{T\left(\tilde{R}-I(X;Y,Z)+H(Y|Z)\right)} & \text{if } I(X;Z)+2\delta_1+\delta_2+4\delta_3< \\ & \quad \tilde{R}<I(X;Y,Z)-\delta_1 \\ \overset{\delta_3}{=} 2^{TH(Y|Z)} & \text{if } I(X;Y,Z) + \delta_1 < \tilde{R} < H(X) - \delta, \end{cases}$$

where we define $\delta_1 \triangleq \delta 2H(X)$, $\delta_2 \triangleq \delta H(Y|X,Z)$, $\delta_3 \triangleq \delta H(Y|Z)$, and $\delta_4 = \delta_1 + \delta_3$.

In Lemma 5.3, the exponent for the second regime can be written as

$$\tilde{R} - I(X;Y,Z) + H(Y|Z) = \tilde{R} - I(X;Z) + H(Y|X,Z). \tag{5.19}$$

To find the list $\mathcal{L}_\delta(Y|\mathbf{z})$, we can imagine that we first find all the $\mathbf{x} \in \mathcal{L}_\delta(X|\mathbf{z})$. Then, the list $\mathcal{L}_\delta(Y|\mathbf{z})$ is the union over all such $\mathbf{x}$ of the "fans" $\mathcal{T}_\delta(Y|\mathbf{x},\mathbf{z})$. We observe the following rather surprising fact:

- If $\tilde{R} < I(X;Z) - \delta_1$, then $\mathcal{L}_\delta(X|\mathbf{z})$ is of size 1 (Lemma 5.2), and hence, $\mathcal{L}_\delta(Y|\mathbf{z})$ is simply the fan $\mathcal{T}_\delta(Y|\mathbf{x},\mathbf{z})$ for the unique $\mathbf{x} \in \mathcal{L}_\delta(X|\mathbf{z})$. The size of this fan is exponentially equal to $2^{TH(Y|X,Z)}$ as indicated in Lemma 5.3.

- If $I(X;Z)+2\delta_1+\delta_2+4\delta_3 < \tilde{R} < I(X;Y,Z)-\delta_1$, then the size of $\mathcal{L}_\delta(X|\mathbf{z})$ is exponentially equal to $2^{T(\tilde{R}-I(X;Z))}$ (see Lemma 5.2). Hence, Lemma 5.3, together with (5.19), tells us that $\mathcal{L}_\delta(Y|\mathbf{z})$ is the union of $2^{T(\tilde{R}-I(X;Z))}$ disjoint fans, each of size (exponentially equal to) $2^{TH(Y|X,Z)}$. In other words, the fans $\mathcal{T}_\delta(Y|\mathbf{x},\mathbf{z})$ are disjoint for different $\mathbf{x} \in \mathcal{L}_\delta(X|\mathbf{z})$.

- If $I(X;Y,Z) + \delta_1 < \tilde{R} < H(X) - \delta$, the size of $\mathcal{L}_\delta(X|\mathbf{z})$ is still exponentially equal to $2^{T(\tilde{R}-I(X;Z))}$. However, the fans $\mathcal{T}_\delta(Y|\mathbf{x},\mathbf{z})$ are no longer disjoint for different $\mathbf{x} \in \mathcal{L}_\delta(X|\mathbf{z})$. From Lemma 5.3, we see that $|\mathcal{L}_\delta(Y|\mathbf{z})|$ is exponentially equal to $|\mathcal{T}_\delta(Y|\mathbf{z})|$. This implies that not only are the fans $\mathcal{T}_\delta(Y|\mathbf{x},\mathbf{z})$ not disjoint, but they cover almost all of $\mathcal{T}_\delta(Y|\mathbf{z})$.

This observation suggests that a sharp transition happens when $\tilde{R}$ is close to $I(X;Y,Z)$, in the sense that $\mathcal{L}_\delta(Y|\mathbf{z})$ transits from being a disjoint union of exponentially many sets to being a covering of $\mathcal{T}_\delta(Y|\mathbf{z})$. This situation is illustrated in Figure 5.3.

In the absence of Z, the same phenomenon happens. In this case, one can show that the size of $\mathcal{L}_\delta(Y)$ is

$$L_\delta(Y) \begin{cases} \overset{\delta_3}{=} 2^{T(\tilde{R}+H(Y|X))} & \text{if } 0 < \tilde{R} < I(X;Y) - \delta_1 \\ \overset{\delta_1}{=} 2^{TH(Y)} & \text{if } I(X;Y) + \delta_1 < \tilde{R} < H(X) - \delta. \end{cases}$$

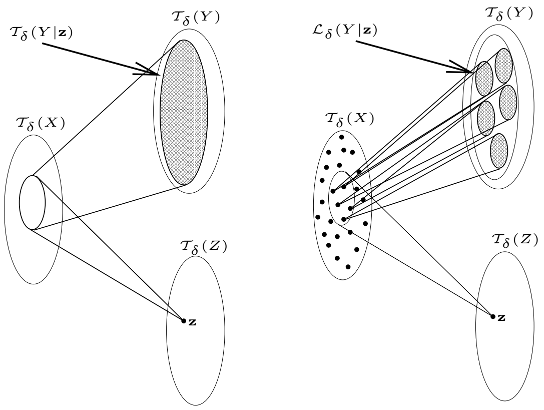

Figure 5.3: Illustration of the list $\mathcal{L}(Y|\mathbf{z})$ for two cases: when $I(X;Y,Z)+\delta_1 < \tilde{R}$ (on the left) and when $\tilde{R} < I(X;Y,Z) - \delta_1$ (on the right).

This means that the fans going from codewords in $\mathcal{C}$ to $\mathcal{T}_\delta(Y)$ are disjoint for $\tilde{R} < I(X;Y) - \delta_1$. When $\tilde{R} > I(X;Y) + \delta_1$, not only the fans are not disjoint, but they cover all of $\mathcal{T}_\delta(Y)$ without leaving any gaps. Hence, we again have a sharp transition when $\tilde{R}$ is close to $I(X;Y)$.

Another interesting observation is the following. For $I(X;Y) + \delta_1 < \tilde{R} < I(X;Y,Z) - \delta_1$, every $\mathbf{y} \in \mathcal{T}_\delta(Y)$ is covered by a fan from some codeword $\mathbf{x} \in \mathcal{C}$, but for a fixed $\mathbf{z}$, *not every* $\mathbf{y} \in \mathcal{T}_\delta(Y|\mathbf{z})$ is covered by a fan coming from a codeword $\mathbf{x} \in \mathcal{L}_\delta(X|\mathbf{z})$.

The following lemma deals with a random experiment where the (uniformly generated) codebook $\mathcal{C}$ is fixed, and we randomly pick a sequence $\mathbf{x}$ from the given $\mathcal{C}$.

Lemma 5.4 *Let $\mathcal{C}$ be a fixed code as defined in the beginning of this section. Consider the random experiment of picking a codeword $\mathbf{X}$ uniformly at random from $\mathcal{C}$ and transmitting it over the channel $p_{YZ|X}$. The received sequences $\mathbf{Y}$ and $\mathbf{Z}$ are then random variables. Fix a sequence $\mathbf{z} \in \mathcal{L}_\delta(Z)$. Then, for almost every possible code $\mathcal{C}$,*

$$\mathbf{P}(\mathbf{Y} = \mathbf{y}|\mathbf{Z} = \mathbf{z}) \overset{\tilde{\delta}}{\doteq} \frac{1}{L_\delta(Y|\mathbf{z})}$$

if $\mathbf{y} \in \mathcal{L}_\delta(Y|\mathbf{z})$ and

$$\mathbf{P}(\mathbf{Y} = \mathbf{y}|\mathbf{Z} = \mathbf{z}) \doteq 0$$

otherwise, where $\tilde{\delta}$ is a scaled version of δ, and $\doteq$ denotes relative equality as defined in Section 2.1.

This lemma states that by choosing δ sufficiently small, we can make the distribution of $\mathbf{Y}$ given $\mathbf{Z} = \mathbf{z}$ arbitrarily "close" to the uniform distribution on $\mathcal{L}_\delta(Y|\mathbf{z})$.

The Interference-Multiple Access Channel 6

6.1 Introduction

The general two transmitter - two receiver channel, shown in Figure 6.1 is a channel connecting two transmitters to two receivers. We call this small network a “channel” because none of the nodes plays the role of a relay. Assume that Transmitter 1 encodes a message W_1 and Transmitter 2 a message W_2. When Receiver 1 is only required to decode W_1 and Receiver 2 only W_2, then the communication problem is called the interference channel, because the signal sent by Transmitter 1 interferes with the communication between Transmitter 2 and Receiver 2 and vice versa. This problem was introduced in [40], and as of yet, the general capacity region for this channel is unknown. The best known inner bound was provided by Han and Kobayashi in [16]. See [8] for a survey of solved special cases. In particular, the results in [14] and [13] are related to our work. In [14], the capacity region of a class of deterministic interference channels is given, and [13] provides a 1 bit-approximation of the capacity region for the Gaussian interference channel. More recently, Telatar and Tse [44] generalized the results of [14] and [13].

A different problem arises when the receivers are required to decode both messages W_1 and W_2. In this case, for fixed product input distributions, we can achieve the intersection of the achievable multiple-access regions for Receiver 1 and Receiver 2. The capacity region for the channel is the union over all product input distributions of these intersections.

In this chapter, we introduce a new problem, where we require Receiver 1 to decode both W_1 and W_2, but Receiver 2 is only required to decode the message W_2 encoded by Transmitter 2. We call this problem formulation the *interference-multiple-access* (IMA) channel. The IMA channel is related to the cognitive interference channel studied by Liang and others in [24]. The difference is that in the cognitive interference channel, one of the transmitters observes the message that the other transmitter encodes. The authors of [24] find the exact capacity region with and without secrecy of the cognitive interference channel.

We first focus on the capacity region for the IMA channel if there is no secrecy requirement. Our main results are an achievable rate region for the general IMA channel, an outer bound for a certain class of so-called structured IMA channels, as well as an expression for the gap between the inner and outer bound for structured IMA channels. This class of structured channels is the IMA analog of the class of interference channels considered in [14] and [44]. It turns out that the inner and outer bounds match for a semi-deterministic channel, providing a complete characterization of the capacity region. For the Gaussian case, we show that the gap is at most 1 bit, yielding an approximate characterization. In the last section of this chapter, we consider the additional constraint that part of W_1 should be kept secret from Receiver 2. This requirement creates a tension for Transmitter 1, because its effort to facilitate the decoding of W_2 at Receiver 2 might harm the secrecy of its own message. Surprisingly, we can show that in a very special case, a scheme derived without concern for secrecy actually provides "unintentional" secrecy for the IMA channel. We then provide an inner bound on the equivocation-capacity region (the region of all achievable rate-equivocation triples) of the IMA channel.

In Section 6.2, we define the general and structured IMA channels in detail. Section 6.3 states the main results and Section 6.4 contains the corresponding proofs. The secrecy results are stated in Section 6.5 and proved in Appendix D.

6.2 Problem Statement

6.2.1 The General Interference-Multiple-Access Channel

Consider the discrete memoryless channel shown in Figure 6.1, connecting two transmitters to two receivers. The input and output alphabets are discrete sets $\mathcal{X}_1$, $\mathcal{X}_2$, $\mathcal{Y}_1$ and $\mathcal{Y}_2$. If input symbols (x_1, x_2) are transmitted, the outputs (Y_1, Y_2) are governed by the conditional probability distribution $p_{Y_1,Y_2|X_1,X_2}(\cdot|x_1, x_2)$. Messages W_1 and W_2 are encoded at Transmitter 1 and 2, respectively, and we require Receiver 1 to produce estimates of (W_1, W_2), while Receiver 2 is only required to estimate W_2. We refer to this problem as the general interference-multiple-access (IMA) channel.

We assume that Transmitter 1 and Transmitter 2 are operating in a synchronized way, meaning that both transmitters start transmitting a given message realization at the same time slot. This means that even if one transmitter uses less time-slots to transmit its message, it will wait with transmitting a new message realization until the other transmitter also does so. Based on this assumption, we can restrict ourselves to block codes.

Definition 6.1 *Let $W_i \in \mathcal{W}_i$ be the message observed by Transmitter i, $i =$*

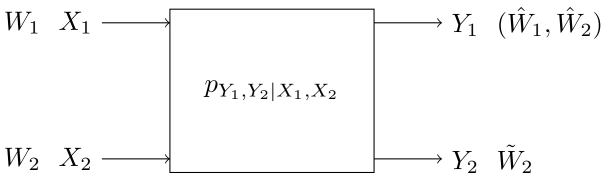

Figure 6.1: The general interference-multiple-access (IMA) channel.

$1, 2$, *and assume that* W_i *is uniformly distributed in the finite message alphabet* $\mathcal{W}_i$. *Assume that time is divided into time slots, and that the transmission of a set of messages starts at time* 1. *A* (T, ϵ)***-block code*** *is given by two possibly random encoding functions* $f_i : \mathcal{W}_i \to \mathcal{X}_i^T$, $i = 1, 2$ *and two decoding functions* $\tilde{f}_1 : \mathcal{Y}_1^T \to \mathcal{W}_1 \times \mathcal{W}_2$ *and* $\tilde{f}_2 : \mathcal{Y}_2^T \to \mathcal{W}_2$. *The* ***block length*** T *is defined as the number of time slots that go by until the transmitters start transmitting a new realization of their messages. Let* $(\hat{W}_1, \hat{W}_2) = \tilde{f}_1(\mathcal{Y}_1)$ *be the estimate of* (W_1, W_2) *at Receiver 1, and let* $\tilde{W}_2 = \tilde{f}_2(\mathcal{Y}_2)$ *be the estimate of* W_2 *at Receiver 2. We require the code to be* ϵ***-reliable*** *in the usual sense that* $\mathbf{P}((W_1, W_2) \neq (\hat{W}_1, \hat{W}_2)) \leq \epsilon$ *and* $\mathbf{P}(W_2 \neq \tilde{W}_2) \leq \epsilon$.

The capacity region is defined in the usual way as follows.

Definition 6.2 *The* ***capacity region*** $\mathcal{R}$ *of a given IMA channel is the set of all rate pairs* (R_1, R_2) *such that for any* $\epsilon > 0$ *there exists a block length* T *and a* (T, ϵ)*-block code for which we have* $\frac{1}{T} H(W_i) \geq R_i - \epsilon$, $i = 1, 2$.

Our first result, outlined in Section 6.3.1, is the description of an inner bound on $\mathcal{R}$, *i.e.*, a region of pairs (R_1, R_2) that are guaranteed to be achievable.

6.2.2 The Structured Interference-Multiple-Access Channel

In Section 6.3.2, we consider a specific class of two-user channels, depicted in Figure 6.2, and described in the following definition.

Definition 6.3 *A two-user channel is* ***structured*** *if the outputs* Y_1 *and* Y_2 *are determined from the input symbols* $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$ *in the following way. First,* x_i *is fed through a discrete memoryless channel with transition probability* $p_{V_i|X_i}$ *to obtain a random variable* V_i, *taking values in the discrete alphabet* $\mathcal{V}_i$, *for* $i = 1, 2$. *Then,* Y_1 *and* Y_2 *are computed as the outputs of the functions* $g_1(x_1, V_2)$ *and* $g_2(x_2, V_1)$, *respectively. We assume that the functions* g_i, $i = 1, 2$, *have the property that for any fixed* x_1, *the map*

$$g_1(x_1, \cdot) : \mathcal{V}_2 \to \mathcal{Y}_1, \; V_2 \mapsto Y_1$$

is invertible, and similarly for g_2. *The channel is fully specified by* $\mathcal{X}_i$, $\mathcal{V}_i$, $\mathcal{Y}_i$, $p_{V_i|X_i}$ *and* g_i, *for* $i = 1, 2$.

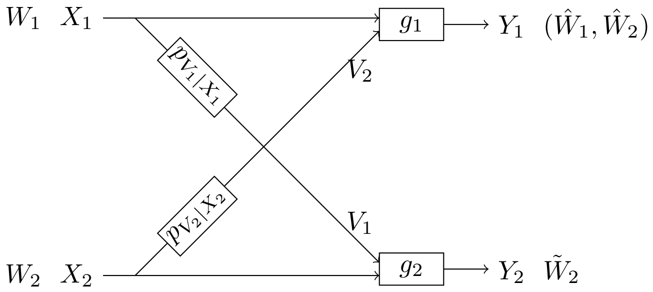

Figure 6.2: The structured IMA channel.

The structured two-user channel was introduced in this general form by Telatar and Tse in [44], and its deterministic version was earlier studied by El Gamal and Costa [14]. Both [44] and [14] considered the interference version of this channel. Telatar and Tse provided a new outer bound on the rate-region of the structured interference channel, and quantified the gap between this outer bound and the Han-Kobayashi inner bound in [16].

In Section 6.3.2, we provide an outer bound on the capacity region $\mathcal{R}$ for the structured IMA channel. We quantify the gap between inner and outer bounds on the capacity region, using the techniques introduced in [44].

6.3 Main Results

6.3.1 Inner Bound

Definition 6.4 *For a given distribution $p_{Q,X_1,X_2,U_1} = p_Q\, p_{X_1|Q}\, p_{X_2|Q}\, p_{U_1|X_1,Q}$, define $\mathcal{R}_i(Q, X_1, X_2, U_1)$ as the set of all pairs (R_1, R_2) of non-negative real numbers satisfying the 6 constraints*

$$\begin{aligned} R_1 &\le I(X_1; Y_1|X_2, Q), && (6.1)\\ R_2 &\le I(X_2; Y_1|X_1, Q), && (6.2)\\ R_1 + R_2 &\le I(X_1, X_2; Y_1|Q), && (6.3)\\ R_2 &\le I(X_2; Y_2|U_1, Q), && (6.4)\\ R_1 + R_2 &\le I(U_1, X_2; Y_2|Q) + I(X_1; Y_1|U_1, X_2, Q), && (6.5)\\ R_1 + 2R_2 &\le I(U_1, X_2; Y_2|Q) + I(X_1, X_2; Y_1|U_1, Q). && (6.6) \end{aligned}$$

In addition, define

$$\mathcal{R}_i \triangleq \cup_{Q,X_1,X_2,U_1} \mathcal{R}_i(Q, X_1, X_2, U_1),$$

where the union is over all distributions of the form $p_Q\, p_{X_1|Q}\, p_{X_2|Q}\, p_{U_1|X_1,Q}$. To simplify the presentation, we use the random variables to denote their distribution, which is an abuse of notation.

Remark: Using Caratheodory's theorem (see e.g. Lemma 3.4 in [11]), one can show that to find $\mathcal{R}_i$, one can restrict ones attention to auxiliary random variables (Q, U_1) that take values in discrete alphabets of sizes $|\mathcal{Q}| \leq 7$ and $|\mathcal{U}_1| \leq |\mathcal{Q}| \cdot |\mathcal{X}_1| + 3$.

Theorem 6.1 *For any IMA channel given by a set of alphabets and a distribution $p_{Y_1,Y_2|X_1,X_2}$, we have*

$$\mathcal{R}_i \subseteq \mathcal{R}.$$

To prove Theorem 6.1, we show the existence of a code that uses superposition at Transmitter 1. The auxiliary random variable U_1, which can be referred to as the cloud center, facilitates the decoding at Receiver 2. The proof details are given in Section 6.4.1. Note that for structured IMA channels, using the properties of the functions g_i and the fact that V_i depends only on X_i, for $i = 1, 2$, the inequalities of Definition 6.4 become

$$\begin{aligned}
R_1 \leq & H(Y_1|X_2, Q) - H(V_2|X_2, Q), && (6.7)\\
R_2 \leq & H(V_2|Q) - H(V_2|X_2, Q), && (6.8)\\
R_1 + R_2 \leq & H(Y_1|Q) - H(V_2|X_2, Q), && (6.9)\\
R_2 \leq & H(Y_2|U_1, Q) - H(V_1|U_1, Q), && (6.10)\\
R_1 + R_2 \leq & H(Y_2|Q) - H(V_1|U_1, Q)\\
& + H(Y_1|U_1, X_2, Q) - H(V_2|X_2, Q), && (6.11)\\
R_1 + 2R_2 \leq & H(Y_2|Q) - H(V_1|U_1, Q)\\
& + H(Y_1|U_1, Q) - H(V_2|X_2, Q). && (6.12)
\end{aligned}$$

The random variable Q models the fact that the users can agree on a time-sharing strategy.

6.3.2 Outer Bound for Structured IMA Channels

Definition 6.5 *For a given distribution $p_Q\, p_{X_1|Q}\, p_{X_2|Q}$, define $\mathcal{R}_o(Q, X_1, X_2)$ as the set of all pairs (R_1, R_2) of non-negative real numbers satisfying the 6 constraints*

$$\begin{aligned}
R_1 \leq & H(Y_1|X_2, Q) - H(V_2|X_2, Q),\\
R_2 \leq & H(V_2|Q) - H(V_2|X_2, Q),\\
R_1 + R_2 \leq & H(Y_1|Q) - H(V_2|X_2, Q),\\
R_2 \leq & H(Y_2|X_1, Q) - H(V_1|X_1, Q), && (6.13)\\
R_1 + R_2 \leq & H(Y_2|Q) - H(V_1|X_1, Q)\\
& + H(Y_1|U_1', X_2, Q) - H(V_2|X_2, Q),\\
R_1 + 2R_2 \leq & H(Y_2|Q) - H(V_1|X_1, Q)\\
& + H(Y_1|U_1', Q) - H(V_2|X_2, Q),
\end{aligned}$$

where the auxiliary random variable U_1' takes values in $\mathcal{V}_1$ and follows the distribution

$$p_{U_1'|Q,X_1,X_2,Y_1,Y_2}(u_1|q,x_1,x_2,y_1,y_2) = p_{V_1|X_1}(u_1|x_1),$$

i.e., *U_1' is a conditionally independent copy of V_1. In addition, define*

$$\mathcal{R}_o \triangleq \cup_{Q,X_1,X_2}\mathcal{R}_o(Q,X_1,X_2),$$

where the union is over all distributions of the form $p_Q\ p_{X_1|Q}\ p_{X_2|Q}$.

Theorem 6.2 *For any structured IMA channel given by a set of alphabets, distributions $p_{V_1|X_1}$, $p_{V_2|X_2}$, and functions g_1, g_2, we have*

$$\mathcal{R} \subseteq \mathcal{R}_o.$$

The proof of Theorem 6.2 can be found in Section 6.4.2.

Theorem 6.3 *If $(R_1,R_2) \in \mathcal{R}_o(Q,X_1,X_2)$, then*

$$\big(R_1, R_2 - I(X_1;V_1|U_1',Q)\big) \in \mathcal{R}_i(Q,X_1,X_2,U_1') \subseteq \mathcal{R}_i,$$

where U_1' is defined as in Definition 6.5.

Proof: Let $(R_1,R_2) \in \mathcal{R}_o(Q,X_1,X_2)$. Let the region $\tilde{\mathcal{R}}_o(Q,X_1,X_2)$ be the slight modification of $\mathcal{R}_o(Q,X_1,X_2)$ where we replace the term $H(Y_2|X_1,Q)$ in (6.13) by $H(Y_2|U_1',Q)$. Since by doing so we relax that bound, (R_1,R_2) is also in $\tilde{\mathcal{R}}_o(Q,X_1,X_2)$. After noticing that $I(X_1;V_1|U_1',Q) = H(V_1|U_1',Q) - H(V_1|X_1,Q)$, one can then verify that the pair $(R_1, R_2 - I(X_1;V_1|U_1',Q))$ satisfies all the inequalities (6.7) through (6.12) for $U_1 = U_1'$. □

Corollary 6.1 *If the channel $p_{V_1|X_1}$ is deterministic, then $\mathcal{R}_i = \mathcal{R}_o = \mathcal{R}$.*

Proof: If $p_{V_1|X_1}$ is deterministic, then $U_1' = V_1$, and the gap $I(X_1;V_1|U_1',Q)$ is zero. Thus, we obtain an exact characterization of the capacity region for the structured IMA channel with a deterministic X_1–V_1 channel. □

Consider a Gaussian IMA channel, where all the alphabets are the complex plane and $V_i = b_iX_i + Z_i$ for $i = 1,2$, where b_i is a complex constant and Z_i is complex, circularly symmetric Gaussian noise independent of (X_i,Q). Let $g_1(X_1,V_2) = a_1X_1 + V_2$, and analogously for $g_2(X_2,V_1)$. The resulting channel can be summarized by the equations

$$\begin{aligned} Y_1 &= a_1X_1 + b_2X_2 + Z_2 \\ Y_2 &= b_1X_1 + a_2X_2 + Z_1. \end{aligned}$$

This is equivalent to the Gaussian signal interaction model defined in Definition 2.16. The following is a second corollary to Theorem 6.3.

Corollary 6.2 *For the Gaussian IMA channel, $\mathcal{R}_i$ gives a* 1*-bit approximation of the capacity region.*

Proof: Theorems 6.1 and 6.2 can be extended to channels with continuous alphabets. In this case $U_1' = b_1 X_1 + Z_1'$, where $Z_1' \sim Z_1$ and $Z_1' \perp (X_1, Z_1, Q)$. Denoting by $h(\cdot)$ the differential entropy, we have

$$\begin{aligned} I(X_1; V_1 | U_1', Q) &= h(V_1|U_1', Q) - h(V_1|U_1', X_1, Q) \\ &= h(V_1|U_1', Q) - h(Z_1|Q) \\ &\le h(V_1 - U_1'|Q) - h(Z_1|Q) \\ &= h(Z_1 - Z_1'|Q) - h(Z_1|Q) \\ &= h(Z_1 - Z_1') - h(Z_1) \\ &= \log(2) = 1 \text{ bit.} \end{aligned}$$

This concludes the proof of the corollary. □

The approximation given in Corollary 6.2 can be very useful in the high SNR regime.

6.4 Proof Outlines

6.4.1 Proof of Theorem 6.1

Assume that $(R_1, R_2) \in \mathcal{R}_i(Q, X_1, X_2, U_1)$ for a given joint distribution $p_Q\, p_{X_1|Q}\, p_{X_2|Q}\, p_{U_1|X_1,Q}$. The random variable Q is a time-sharing variable and is assumed to be available at both transmitters and both receivers, modeling a shared pseudo-random number generator.

Through a random coding argument, we show the existence of a code that achieves rates (R_1, R_2). First, generate a sequence $\mathbf{q}$ from $p_Q(\cdot)^T$. Choose some auxiliary rate $B \in [0, R_1]$. Randomly generate a code in the following way. Generate 2^{TB} sequences $\mathbf{u}_1(i)$ from $\prod_{t=1}^{T} p_{U_1|Q}(\cdot|q[t])$, and index them by $i \in \{1, \ldots, 2^{TB}\}$. For each i, generate $2^{T(R_1-B)}$ sequences $\mathbf{x}_1(i,j)$ from $\prod_{t=1}^{T} p_{X_1|U_1,Q}(\cdot|u_1[t](i), q[t])$, and index them by $j \in \{1, \ldots, 2^{T(R_1-B)}\}$. Finally, generate 2^{TR_2} sequences $\mathbf{x}_2(k)$ from $\prod_{t=1}^{T} p_{X_2|Q}(\cdot|q[t])$, and index them by $k \in \{1, \ldots, 2^{TR_2}\}$.

Once this code is generated, it is fixed for all time and used to encode a message pair $(W_1, W_2) = \big((i,j), k\big)$ into transmit vectors $\mathbf{x}_1(i,j)$ and $\mathbf{x}_2(k)$. Receiver 1 declares its estimate $(\hat{W}_1, \hat{W}_2)$ to be the unique triple $\big((i,j), k\big)$ for which $\big(\mathbf{x}_1(i,j), \mathbf{u}_1(i), \mathbf{x}_2(k), \mathbf{y}_1, \mathbf{q}\big) \in \mathcal{T}_\delta(X_1, U_1, X_2, Y_1, Q)$, where $\mathcal{T}_\delta$ is the set of all jointly typical sequences as defined Appendix F. If such a unique triple cannot be found, Receiver 1 declares an error. Receiver 2 declares its estimate $\tilde{W}_2$ to be the unique index k for which there exists at least one index i such that $\big(\mathbf{u}_1(i), \mathbf{x}_2(k), \mathbf{y}_2, \mathbf{q}\big) \in \mathcal{T}_\delta(U_1, X_2, Y_2, Q)$. If such a unique index cannot be found, Receiver 2 declares an error. An error occurs when a receiver declares an error, or when $(\hat{W}_1, \hat{W}_2) \neq (W_1, W_2)$ or $\tilde{W}_2 \neq W_2$.

Without loss of generality, assume that $(W_1, W_2) = \big((1,1), 1\big)$.
As $\big(\mathbf{x}_1(1,1), \mathbf{u}_1(1), \mathbf{x}_2(1), \mathbf{y}_1, \mathbf{y}_2, \mathbf{q}\big) \in \mathcal{T}_\delta(X_1, U_1, X_2, Y_1, Y_2, Q)$ with high probability, to upper bound the probability of error it is sufficient to consider the

events where typicality holds for an index triple $(i, j, k) \neq (1, 1, 1)$. The error events at Receiver 1 are that $\big(\mathbf{x}_1(i, j), \mathbf{u}_1(i), \mathbf{x}_2(k), \mathbf{y}_1, \mathbf{q}\big) \in \mathcal{T}_\delta(X_1, U_1, X_2, Y_1, Q)$ for

- $i \neq 1$, j arbitrary and $k = 1$,
- $i = 1$, $j \neq 1$ and $k = 1$,
- $i = 1$, $j = 1$ and $k \neq 1$,
- $i \neq 1$, j arbitrary and $k \neq 1$,
- $i = 1$, $j \neq 1$ and $k \neq 1$.

The error events at Receiver 2 are that $\big(\mathbf{u}_1(i), \mathbf{x}_2(k), \mathbf{y}_2, \mathbf{q}\big) \in \mathcal{T}_\delta(U_1, X_2, Y_2, Q)$ for

- $i \neq 1$ and $k \neq 1$,
- $i = 1$ and $k \neq 1$.

From standard arguments [11] it follows that the expected probability of all these events (where the expectation is over all random codes and over all messages (W_1, W_2)) can be made arbitrarily small if

$$\begin{aligned} B + (R_1 - B) = R_1 &\leq I(X_1; Y_1 | X_2, Q), && (6.14)\\ R_1 - B &\leq I(X_1; Y_1 | U_1, X_2, Q), && (6.15)\\ R_2 &\leq I(X_2; Y_1 | X_1, Q), && (6.16)\\ B + (R_1 - B) + R_2 = R_1 + R_2 &\leq I(X_1, X_2; Y_1 | Q), && (6.17)\\ (R_1 - B) + R_2 &\leq I(X_1, X_2; Y_1 | U_1, Q), && (6.18) \end{aligned}$$

and

$$\begin{aligned} B + R_2 &\leq I(U_1, X_2; Y_2 | Q), && (6.19)\\ R_2 &\leq I(X_2; Y_2 | U_1, Q). && (6.20) \end{aligned}$$

In (6.14), (6.16) and (6.17), we used the Markov chain $U_1 \multimap X_1 \multimap (X_2, Y_1, Y_2)$ (conditioned on Q) to drop U_1 from the right hand side. The constraint $B \in [0, R_1]$ can be written as

$$\begin{aligned} B - R_1 &\leq 0, && (6.21)\\ -B &\leq 0. && (6.22) \end{aligned}$$

Out of the constraints (6.14) through (6.22), three are lower bounds on B (namely (6.15), (6.18) and (6.22)), whereas two are upper bounds on B ((6.19) and (6.21)). It is clear that an auxiliary rate B that satisfies these 5 bounds exists if and only if every upper bound on B is larger than (or equal to) every lower bound on B. For instance, to be able to find a B that satisfies (6.15) and (6.19), we require

$$R_1 + R_2 \leq I(X_1; Y_1 | U_1, X_2, Q) + I(U_1, X_2; Y_2 | Q), \tag{6.23}$$

and, on the other hand, if such a B exists, than (6.23) is true. Note that (6.23) is the same condition as (6.5) in the claim of the theorem. Analogously, by combining (6.18) with (6.19), we obtain (6.6). All the other combinations of lower and upper bounds on B yield inequalities that are already implied by (6.14), (6.16), (6.17), (6.20), by the non-negativity of the rate R_1 or by the non-negativity of the mutual information. Hence, inequalities (6.14) through (6.22) are true for some B if and only if (6.1) through (6.6) are true (together with $R_1, R_2 \geq 0$). But these 6 inequalities hold since we assume that $(R_1, R_2) \in \mathcal{R}_i(Q, X_1, X_2, U_1)$. Thus, the expected probability of all error events can be made arbitrarily small by choosing T large enough. It follows that there exists at least one code for which all decoding error probabilities are arbitrarily small.

6.4.2 Proof of Theorem 6.2

The proof follows the same lines as the proof of Theorem 1 in [44]. Let a structured IMA channel be given by $p_{V_i|X_i}$ and g_i for $i = 1, 2$. Assume that $(R_1, R_2) \in \mathcal{R}$ and fix an arbitrary $\epsilon > 0$. From the definition of $\mathcal{R}$, there exists a block length T and a (T, ϵ)-code with rates at least $R_1 - \epsilon$ and $R_2 - \epsilon$, respectively. Let W_1 and W_2 be independent messages and let $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{V}_1$, $\mathbf{V}_2$, $\mathbf{Y}_1$, $\mathbf{Y}_2$ be the random T-sequences induced by them, the code and the channel. Generate a sequence $\mathbf{U}_1'$ by passing $\mathbf{X}_1$ through an auxiliary channel described by $p_{V_1|X_1}$. Note that, by construction, $\mathbf{U}_1' \mathrel{-\!\!\circ\!\!-} \mathbf{X}_1 \mathrel{-\!\!\circ\!\!-} (\mathbf{X}_2, \mathbf{V}_1, \mathbf{V}_2, \mathbf{Y}_1, \mathbf{Y}_2)$ forms a Markov chain, and $(\mathbf{U}_1, \mathbf{X}_1)$ has the same distribution as $(\mathbf{V}_1, \mathbf{X}_1)$. In particular, $H(\mathbf{U}_1) = H(\mathbf{V}_1)$ and $H(\mathbf{U}_1|\mathbf{X}_1) = H(\mathbf{V}_1|\mathbf{X}_1)$.

By Fano's inequality we have

$$TR_i \leq I(\mathbf{X}_i; \mathbf{Y}_i) + T\epsilon',$$

for $i = 1, 2$,

$$TR_2 \leq I(\mathbf{X}_2; \mathbf{Y}_1) + T\epsilon',$$

and

$$T(R_1 + R_2) \leq I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}_1) + T\epsilon',$$

where ϵ' can be made arbitrarily small by choosing ϵ sufficiently small. We can then write

$$\begin{aligned} &T(R_1 - \epsilon') \\ &\qquad \leq I(\mathbf{X}_1; \mathbf{Y}_1) \\ &\qquad \leq I(\mathbf{X}_1; \mathbf{Y}_1, \mathbf{X}_2) \\ &\qquad = I(\mathbf{X}_1; \mathbf{Y}_1 | \mathbf{X}_2) \\ &\qquad = H(\mathbf{Y}_1 | \mathbf{X}_2) - H(\mathbf{Y}_1 | \mathbf{X}_1, \mathbf{X}_2) \\ &\qquad = H(\mathbf{Y}_1 | \mathbf{X}_2) - H(\mathbf{V}_2 | \mathbf{X}_2) \\ &\qquad \leq \sum_{t=1}^{T} \big[H(Y_1[t] | X_2[t]) - H(V_2[t] | X_2[t]) \big], \end{aligned} \tag{6.24}$$

where the last inequality holds because $H(\mathbf{Y}_1|\mathbf{X}_2)$ is upper bounded by its single-letter form, and $H(\mathbf{V}_2|\mathbf{X}_2)$ is equal to its single-letter form, because $\mathbf{V}_2$ is the output of a memoryless channel whose input is $\mathbf{X}_2$. Following the same initial steps, we obtain

$$\begin{aligned} &T(R_2 - \epsilon') \\ &\qquad \leq I(\mathbf{X}_2; \mathbf{Y}_1) \\ &\qquad \leq H(\mathbf{Y}_1 | \mathbf{X}_1) - H(\mathbf{Y}_1 | \mathbf{X}_1, \mathbf{X}_2) \\ &\qquad = H(\mathbf{V}_2) - H(\mathbf{V}_2 | \mathbf{X}_2) \\ &\qquad \leq \sum_{t=1}^{T} \big[H(V_2[t]) - H(V_2[t] | X_2[t]) \big]. \end{aligned} \tag{6.25}$$

In a similar manner, we have

$$\begin{aligned} &T(R_1 + R_2 - \epsilon') \\ &\qquad \leq I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}_1) \\ &\qquad = H(\mathbf{Y}_1) - H(\mathbf{V}_2 | \mathbf{X}_2) \\ &\qquad \leq \sum_{t=1}^{T} \big[H(Y_1[t]) - H(V_2[t] | X_2[t]) \big] \end{aligned} \tag{6.26}$$

and

$$\begin{aligned} &T(R_2 - \epsilon') \\ &\qquad \leq I(\mathbf{X}_2; \mathbf{Y}_2) \\ &\qquad \leq I(\mathbf{X}_2; \mathbf{Y}_2, \mathbf{X}_1) \\ &\qquad = I(\mathbf{X}_2; \mathbf{Y}_2 | \mathbf{X}_1) \\ &\qquad = H(\mathbf{Y}_2 | \mathbf{X}_1) - H(\mathbf{V}_1 | \mathbf{X}_1) \\ &\qquad \leq \sum_{t=1}^{T} \big[H(Y_2[t] | X_1[t]) - H(V_1[t] | X_1[t]) \big]. \end{aligned} \tag{6.27}$$

We further obtain

$$\begin{aligned}
&T(R_1 + R_2 - 2\epsilon') \\
&\qquad \le I(\mathbf{X}_1; \mathbf{Y}_1, \mathbf{U}_1', \mathbf{X}_2) + I(\mathbf{X}_2; \mathbf{Y}_2) \\
&\qquad = I(\mathbf{X}_1; \mathbf{U}_1') + \underbrace{I(\mathbf{X}_1; \mathbf{X}_2|\mathbf{U}_1')}_{=0} \\
&\qquad\quad + I(\mathbf{X}_1; \mathbf{Y}_1|\mathbf{U}_1', \mathbf{X}_2) + I(\mathbf{X}_2; \mathbf{Y}_2) \\
&\qquad = \underbrace{H(\mathbf{U}_1')}_{=H(\mathbf{V}_1)} - \underbrace{H(\mathbf{U}_1'|\mathbf{X}_1)}_{=H(\mathbf{V}_1|\mathbf{X}_1)} + H(\mathbf{Y}_1|\mathbf{U}_1', \mathbf{X}_2) \\
&\qquad\quad - H(\mathbf{V}_2|\mathbf{X}_2) + H(\mathbf{Y}_2) - H(\mathbf{V}_1) \\
&\qquad \le \sum_{t=1}^{T} \big[H(Y_2[t]) - H(V_1[t]|X_1[t]) \\
&\qquad\quad + H(Y_1[t]|U_1'[t], X_2[t]) - H(V_2[t]|X_2[t])\big], \qquad (6.28)
\end{aligned}$$

where the indicated term is zero because W_1 and W_2 are independent. Finally,

$$\begin{aligned}
&T(R_1 + 2R_2 - 3\epsilon') \\
&\qquad \le I(\mathbf{X}_1; \mathbf{Y}_1, \mathbf{U}_1') + I(\mathbf{X}_2; \mathbf{Y}_2) \\
&\qquad\quad + I(\mathbf{X}_2; \mathbf{Y}_1, \mathbf{V}_2) \\
&\qquad = I(\mathbf{X}_1; \mathbf{U}_1') + I(\mathbf{X}_1; \mathbf{Y}_1|\mathbf{U}_1') + I(\mathbf{X}_2; \mathbf{Y}_2) \\
&\qquad\quad + I(\mathbf{X}_2; \mathbf{V}_2) + \underbrace{I(\mathbf{X}_2; \mathbf{Y}_1|\mathbf{V}_2)}_{=0} \\
&\qquad = \underbrace{H(\mathbf{U}_1')}_{=H(\mathbf{V}_1)} - H(\mathbf{U}_1'|\mathbf{X}_1) + H(\mathbf{Y}_1|\mathbf{U}_1') - H(\mathbf{V}_2) \\
&\qquad\quad + H(\mathbf{Y}_2) - H(\mathbf{V}_1) + H(\mathbf{V}_2) - H(\mathbf{V}_2|\mathbf{X}_2) \\
&\qquad = H(\mathbf{Y}_2) - H(\mathbf{U}_1'|\mathbf{X}_1) + H(\mathbf{Y}_1|\mathbf{U}_1') - H(\mathbf{V}_2|\mathbf{X}_2) \\
&\qquad \le \sum_{t=1}^{T} \big[H(Y_2[t]) - H(V_1[t]|X_1[t]) \\
&\qquad\quad + H(Y_1[t]|U_1'[t]) - H(V_2[t]|X_2[t])\big]. \qquad (6.29)
\end{aligned}$$

Setting (Q, X_1, X_2) to be random variables with Q uniformly distributed in $\{1, \ldots, T\}$ and $p_{X_i|Q}(x_i|t) = \mathbf{P}(X_i[t] = x_i)$, for $i = 1, 2$, we see that the inequalities (6.24) through (6.29) can be rewritten as $(R_1 - \epsilon', R_2 - \epsilon') \in \mathcal{R}_o(Q, X_1, X_2) \subseteq \mathcal{R}_o$, where ϵ' can be made arbitrarily small. As $\mathcal{R}_o$ is closed, we see that any achievable rate pair is in $\mathcal{R}_o$.

6.5 Secrecy in the Interference-Multiple Access Channel

In the IMA channel, no eavesdroppers are present. However, one can imagine that certain applications require part of the message W_1, which is only intended for Receiver 1, to be kept secret from Receiver 2. In this section, we are concerned with this type of secrecy.

6.5.1 Unintentional Secrecy

In the achievability proof of Section 6.4.1, we show the existence of a superposition code. When using this code, Transmitter 1 combines a "cloud index" i of rate B with a "cloud member index" j of rate $R_1 - B$ to identify a codeword $\mathbf{x}_1$ for transmission. The role of the cloud index i is to help Receiver 2 in the reception of W_2. The proposition given below says that if the auxiliary rate B satisfies certain conditions, then this code, which was chosen without secrecy constraints in mind, guarantees (weak) secrecy for j against Receiver 2. This is why we call this notion "unintentional secrecy".

Proposition 6.1 *Fix a distribution $p_Q\ p_{X_1|Q}\ p_{X_2|Q}\ p_{U_1|X_1,Q}$. Assume that the auxiliary rate B satisfies the following two conditions:*

$$B \leq I(U_1; Y_2|X_2, Q) \quad \textit{and} \tag{6.30}$$

$$B = \min\{I(X_1; Y_2|X_2, Q), I(X_1, X_2; Y_2|Q) - R_2\} - \epsilon_1 \tag{6.31}$$

for some $\epsilon_1 > 0$. Consider the code whose existence we have shown in Section 6.4.1 and let J be the part of the message W_1 that indicates the cloud member. Let $\mathbf{Y}_2$ be the received sequence at Receiver 2. Then, we have that over the probability space of the code,

$$H(J|\mathbf{Y}_2) \geq T\big(R_1 - B - \epsilon_1 - \epsilon'\big)$$

with high probability, where ϵ' can be made arbitrarily small by increasing the reliability of the code.

Intuitively, this proposition holds because of the following. Condition (6.30) implies that Receiver 2 can, with high probability, decode the index I correctly, where I denotes the cloud index. The second condition (6.31) is such that the rate of I occupies almost all the mutual information available for transmission between Transmitter 1 and Receiver 2. Given these two facts, it seems natural that Receiver 2 can only obtain very limited information about the second index J. A formal proof is given in Appendix D.

Note that if ϵ_1 is chosen very small, then (6.30) and (6.31) can imply that $I(X_1; Y_2|X_2, Q)$ and $I(U_1; Y_2|X_2, Q)$ are almost equal. One might ask whether this causes U_1 to be essentially equal to X_1, which would imply that the secrecy rate $R_1 - B$ is very small. However, one can show that there exist random variables U_1, X_1, Y_1, Y_2 for which $U_1 \multimap X_1 \multimap (Y_1, Y_2)$ forms a Markov chain and for which $I(U_1; Y_2) = I(X_1; Y_2)$ but $I(U_1; Y_1) < I(X_1; Y_1)$.

Note however that the conditions (6.30) and (6.31) are somewhat artificial and unlikely to be satisfied "unintentionally". One can also construct IMA codes with secrecy on purpose. This is the subject of the following sub-section.

6.5.2 Intentional Secrecy

Here, we state a result that guarantees that part of W_1 remains (weakly) secret from Receiver 2. We first define a new version of the capacity region $\mathcal{R}$ that takes secrecy into account.

Definition 6.6 *The* ***equivocation-capacity region*** $\mathcal{R}_e$ *of a given IMA channel is the set of all triples* (R_1, R_2, R_e) *such that for any* $\epsilon > 0$*, there exists a block length* T *and a* (T, ϵ)*-code for which we have* $\frac{1}{T}H(W_i) \geq R_i - \epsilon$*,* $i = 1, 2$ *and* $\frac{1}{T}H(W_1|\mathbf{Y}_2) \geq R_e - \epsilon$*, where* $\mathbf{Y}_2$ *is the received sequence at Receiver 2.*

Definition 6.7 *For a given distribution* $p_Q\ p_{X_1|Q}\ p_{X_2|Q}\ p_{U_1|X_1,Q}$*, define* $\mathcal{R}_{e,i}(Q, X_1, X_2, U_1)$ *to be the set of all triples* (R_1, R_2, R_e) *of non-negative real numbers satisfying the 6 constraints of Definition 6.4 and the additional constraint*

$$\begin{aligned} R_e \leq \min\big\{&R_1, I(X_1; Y_1|U_1, X_2, Q), I(X_1, X_2; Y_1|U_1, Q) - R_2\big\} \\ &- \min\big\{I(X_1; Y_2|U_1, X_2, Q), I(X_1, X_2; Y_2|U_1, Q) - R_2\big\}. \end{aligned} \tag{6.32}$$

In addition, define

$$\mathcal{R}_{e,i} \triangleq \cup_{Q,X_1,X_2,U_1} \mathcal{R}_{e,i}(Q, X_1, X_2, U_1),$$

Theorem 6.4 *For any IMA channel given by a set of alphabets and a distribution* $p_{Y_1,Y_2|X_1,X_2}$*, we have*

$$\mathcal{R}_{e,i} \subseteq \mathcal{R}_e.$$

To prove this theorem, we show the existence of a scheme that has the following characteristics. Transmitter 1 uses the same structured code as in the proof of Theorem 6.1, but the private index j is split into two sub-indices j_n and j_s. We are able to show that there exists a code for which j_n is not necessarily secret, but j_s is kept secret from Receiver 2. The bound on the equivocation is derived in a similar way as in Proposition 6.1. A detailed proof is given in Appendix D.

Note that Theorem 6.4 provides weak secrecy from Receiver 2. If a certain (R_1, R_2, R_e) lies in $\mathcal{R}_e$, then the results of Maurer and Wolf, explained in Appendix E, can be used to show that we can achieve *strong* secrecy at a rate arbitrarily close to R_e.

7 Discussion and Future Work

In this thesis, we show that it is possible to guarantee a computable lower bound on the strong perfect secrecy capacity for wireless relay networks with an arbitrary acyclic topology. Our results are for the well-motivated model of Gaussian signal interaction, as well as for the simpler deterministic interaction model. We also provide an upper bound on the perfect secrecy capacity for arbitrary wireless networks. Unfortunately, it seems difficult to characterize the gap between the upper and lower bounds.

The Gaussian result should be viewed as a first step towards an approximate characterization of the perfect secrecy capacity of arbitrary networks. Future efforts should go into finding an upper bound that can be related to the lower bound expression, as well as decreasing the subtractive constants α, β and γ in Theorem 3.2.

The result for deterministic signal interaction can provide valuable insights into possible coding schemes. In addition, it proves to be more easy to handle. Therefore, we believe that it is important to pursue the problem under this model as well. The main aim would again be to find an upper bound on the secrecy capacity that can be related to the lower bound. Towards this aim, one can also try to improve the lower bound by considering additional coding techniques like decode-and-forward and destructive interference. Our results for the fan network show that for special network topologies with discrete memoryless channels, one can obtain characterizations of the perfect secrecy capacity. To find these characterizations, we used intuitions and techniques from our results for general networks.

Feedback from the destination to the source is a promising feature, but studying it for arbitrary wireless networks seems very challenging. Our results for the line network are an encouraging first step. A valuable extension would be a non-trivial upper bound on the secret key capacity for this small network.

During my doctoral studies, I have worked on several other problems that are not contained in this thesis. The most important results are the following:

We considered the problem of lossy compression of a Gaussian source that is to be reproduced by two independent decoders, each having access to a different Gaussian side-information source that is correlated with the data source. We found the rate-distortion region for the case when the side-information sources are physically degraded. In more recent work, we studied lossy compression of a discrete source with erased side-information. We found that this type of setup shares many properties with the Gaussian case, and all our Gaussian results were repeated for this setup. These results can be found in *1.* and *8.* (in the CV at the end of the thesis).

In joint work with Vasudevan (*7.* in the CV), we considered a source coding situation where two encoders, observing correlated sources, can collaborate by means of a rate-limited link from one to the other. The aim is to jointly describe the sources to a decoder. We provided an inner bound on the rate-distortion region and showed tightness for two special cases.

Recently, we presented a result obtained in collaboration with Vasudevan and Vojnović (see *4.* in the CV at the of this thesis). In this contribution, we studied the algorithmic problem of distributed binary consensus in a complete graph. We showed that when each node can use 3 states for memory and signaling, the reliability of the best consensus algorithm improves dramatically as compared to the case when the nodes have binary memory and binary messages. We also showed that the convergence is as fast as for the case without limits on the state.

Together with Rethnakaran Pulikkoonattu, we developed a software called Xitip, which is a C-version of the “Information-Theoretic Inequality Prover” (ITIP) by Yeung and Yan [50]. Our version of this software is faster, more easily portable, and it includes a graphical user interface. Further references about this work are given in the CV at the end of the thesis.

Appendix for Chapter 3 A

A.1 Proof of Lemma 3.1

We have

$$R_{\mathcal{A};j}(p) = \min_{\Omega \in \Lambda(\mathcal{A},j)} H(Y_{\Omega^c}|X_{\Omega^c}) \tag{A.1}$$

$$R_{\mathcal{B};j}(p) = \min_{\Omega \in \Lambda(\mathcal{B},j)} H(Y_{\Omega^c}|X_{\Omega^c}). \tag{A.2}$$

Let $\mathcal{A}$ and $\mathcal{B}$ be subsets of $\mathcal{V}$ such that $\mathcal{A} \subseteq \mathcal{B}$. Then, the monotonicity follows from the fact that $\Lambda(\mathcal{B}, j) \subseteq \Lambda(\mathcal{A}, j)$. We now show the submodularity. Let $\mathcal{A}$ and $\mathcal{B}$ be two subsets of $\mathcal{V}$. Let Ω_1 and Ω_2 be the optimizers of (A.1) and (A.2), respectively. Define the sets

$$\begin{aligned}\Psi_1 &= \Omega_1 \setminus \Omega_2\\ \Psi_2 &= \Omega_2 \setminus \Omega_1\\ \Psi_{12} &= \Omega_1 \cap \Omega_2\\ \Psi_0 &= (\Omega_1 \cup \Omega_2)^c.\end{aligned}$$

Note that the four sets defined above form a partition of $\mathcal{V}$. It follows that

$$\Omega_1 = \Psi_1 \cup \Psi_{12} \tag{A.3}$$

$$\Omega_2 = \Psi_2 \cup \Psi_{12}, \tag{A.4}$$

where both unions are of disjoint sets. It is easy to verify that $\Omega_1 \cup \Omega_2 \in \Lambda(\mathcal{A} \cup \mathcal{B}, j)$ and $\Omega_1 \cap \Omega_2 \in \Lambda(\mathcal{A} \cap \mathcal{B}, j)$. Hence,

$$\begin{aligned}
&R_{\mathcal{A}\cup\mathcal{B};j}(p) + R_{\mathcal{A}\cap\mathcal{B};j}(p) \\
&\quad\leq H(Y_{(\Omega_1\cup\Omega_2)^c}|X_{(\Omega_1\cup\Omega_2)^c}) + H(Y_{(\Omega_1\cap\Omega_2)^c}|X_{(\Omega_1\cap\Omega_2)^c}) \\
&\quad= H(Y_{\Psi_0}|X_{\Psi_0}) + H(Y_{\Psi_1}, Y_{\Psi_2}, Y_{\Psi_0}|X_{\Psi_1}, X_{\Psi_2}, X_{\Psi_0}) \\
&\quad= H(Y_{\Psi_0}|X_{\Psi_0}) + H(Y_{\Psi_1}, Y_{\Psi_0}|X_{\Psi_1}, X_{\Psi_2}, X_{\Psi_0}) \\
&\qquad+ H(Y_{\Psi_2}|Y_{\Psi_1}, Y_{\Psi_0}, X_{\Psi_1}, X_{\Psi_2}, X_{\Psi_0}) \\
&\quad= H(Y_{\Psi_0}|X_{\Psi_0}) + H(Y_{\Psi_1}, Y_{\Psi_0}|X_{\Psi_1}, X_{\Psi_0}) \\
&\qquad- I(X_{\Psi_2}; Y_{\Psi_1}, Y_{\Psi_0}|X_{\Psi_1}, X_{\Psi_0}) \\
&\qquad+ H(Y_{\Psi_2}|Y_{\Psi_1}, Y_{\Psi_0}, X_{\Psi_1}, X_{\Psi_2}, X_{\Psi_0}) \\
&\quad\leq H(Y_{\Psi_0}|X_{\Psi_2}, X_{\Psi_0}) + I(X_{\Psi_2}; Y_{\Psi_0}|X_{\Psi_0}) \\
&\qquad+ H(Y_{\Psi_1}, Y_{\Psi_0}|X_{\Psi_1}, X_{\Psi_0}) \\
&\qquad- I(X_{\Psi_2}; Y_{\Psi_1}, Y_{\Psi_0}|X_{\Psi_1}, X_{\Psi_0}) + H(Y_{\Psi_2}|Y_{\Psi_0}, X_{\Psi_2}, X_{\Psi_0}) \\
&\quad= H(Y_{\Psi_2}, Y_{\Psi_0}|X_{\Psi_2}, X_{\Psi_0}) + H(Y_{\Psi_1}, Y_{\Psi_0}|X_{\Psi_1}, X_{\Psi_0}) \\
&\qquad+ I(X_{\Psi_2}; Y_{\Psi_0}|X_{\Psi_0}) - I(X_{\Psi_2}; Y_{\Psi_1}, Y_{\Psi_0}|X_{\Psi_1}, X_{\Psi_0}), \qquad \text{(A.5)}
\end{aligned}$$

where the first inequality follows from the definition of $R_{\cdot;j}(p)$ and the second inequality follows from dropping part of the conditioning in the last conditional entropy term. We further have

$$\begin{aligned}
I(X_{\Psi_2}; Y_{\Psi_0}|X_{\Psi_0}) &= H(X_{\Psi_2}|X_{\Psi_0}) - H(X_{\Psi_2}|Y_{\Psi_0}, X_{\Psi_0}) \\
&\leq H(X_{\Psi_2}|X_{\Psi_0}, X_{\Psi_1}) - H(X_{\Psi_2}|Y_{\Psi_0}, X_{\Psi_0}, X_{\Psi_1}) \\
&= I(X_{\Psi_2}; Y_{\Psi_0}|X_{\Psi_0}, X_{\Psi_1}) \\
&\leq I(X_{\Psi_2}; Y_{\Psi_1}, Y_{\Psi_0}|X_{\Psi_0}, X_{\Psi_1}), \qquad \text{(A.6)}
\end{aligned}$$

where the first inequality is true because the first term remains unchanged by the independence of the tuple $X_{\mathcal{V}}$, while the second term decreases in absolut value by adding variables in the conditioning. Substituting (A.6) in (A.5), together with (A.3) and (A.4), we find that

$$\begin{aligned}
R_{\mathcal{A}\cup\mathcal{B};j}(p) + R_{\mathcal{A}\cap\mathcal{B};j}(p) &\leq H(Y_{\Psi_2}, Y_{\Psi_0}|X_{\Psi_2}, X_{\Psi_0}) + H(Y_{\Psi_1}, Y_{\Psi_0}|X_{\Psi_1}, X_{\Psi_0}) \\
&= H(Y_{\Omega_1^c}|X_{\Omega_1^c}) + H(Y_{\Omega_2^c}|X_{\Omega_2^c}) \\
&= R_{\mathcal{A};j}(p) + R_{\mathcal{B};j}(p).
\end{aligned}$$

This concludes the proof of the lemma.

A.2 Proof of Lemma 3.2

We have

$$R_{\mathcal{A};j}(p) = \min_{\Omega\in\Lambda(\mathcal{A},j)} I(X_\Omega; Y_{\Omega^c}|X_{\Omega^c}) \qquad \text{(A.7)}$$

$$R_{\mathcal{B};j}(p) = \min_{\Omega\in\Lambda(\mathcal{B},j)} I(X_\Omega; Y_{\Omega^c}|X_{\Omega^c}). \qquad \text{(A.8)}$$

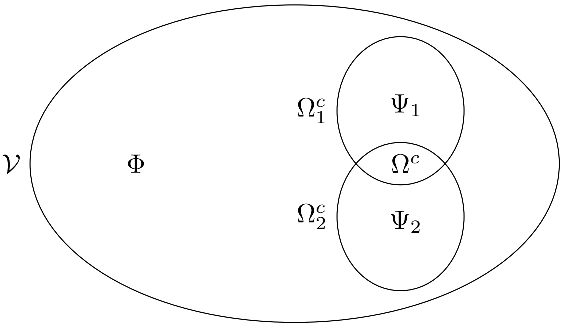

Figure A.1: Illustration of the sets used in the proof of Lemma 3.1.

Let $\mathcal{A}$ and $\mathcal{B}$ be subsets of $\mathcal{V}$ such that $\mathcal{A} \subseteq \mathcal{B}$. Then, the monotonicity follows from the fact that $\Lambda(\mathcal{B}, j) \subseteq \Lambda(\mathcal{A}, j)$. We now show the weak submodularity. Let $\mathcal{A}$ and $\mathcal{B}$ be disjoint subsets of $\mathcal{V}$. Let Ω_1 and Ω_2 be the optimizers of (A.7) and (A.8), respectively. Since $\Omega_1 \in \Lambda(\mathcal{A}, j)$ and $\Omega_2 \in \Lambda(\mathcal{B}, j)$, we have $\mathcal{A} \subseteq \Omega_1$, $\mathcal{B} \subseteq \Omega_2$ and $j \in \Omega_1^c \cap \Omega_2^c$. Hence, $\Omega \triangleq \Omega_1 \cup \Omega_2 \in \Lambda(\mathcal{A} \cup \mathcal{B}, j)$. Note that $\Omega^c = \Omega_1^c \cap \Omega_2^c$ and define

$$\begin{aligned}\Psi_1 &\triangleq \Omega_1^c \setminus \Omega^c\\ \Psi_2 &\triangleq \Omega_2^c \setminus \Omega^c\\ \Phi &\triangleq (\Psi_1 \cup \Psi_2 \cup \Omega^c)^c.\end{aligned}$$

Observe (see Figure A.1) that Ψ_1, Ψ_2, Ω^c and Φ are disjoint sets and such that Ω_1^c is partitioned by Ψ_1 and Ω^c, while Ω_2^c is partitioned by Ψ_2 and Ω^c. Finally, $\mathcal{V}$ is partitioned by Ψ_1, Ψ_2, Ω^c and Φ. We have

$$\begin{aligned}R_{\mathcal{A}\cup\mathcal{B};j}(p) &\leq I(X_\Omega; Y_{\Omega^c}|X_{\Omega^c})\\ &= I(X_\Phi, X_{\Psi_1}, X_{\Psi_2}; Y_{\Omega^c}|X_{\Omega^c})\\ &\overset{(a)}{\leq} I(X_\Phi, X_{\Psi_2}; Y_{\Omega^c}|X_{\Psi_1}, X_{\Omega^c})\\ &\quad + I(X_\Phi, X_{\Psi_1}; Y_{\Omega^c}|X_{\Psi_2}, X_{\Omega^c})\\ &\leq I(X_\Phi, X_{\Psi_2}; Y_{\Omega^c}, Y_{\Psi_1}|X_{\Psi_1}, X_{\Omega^c})\\ &\quad + I(X_\Phi, X_{\Psi_1}; Y_{\Omega^c}, Y_{\Psi_2}|X_{\Psi_2}, X_{\Omega^c})\\ &= I(X_{\Omega_1}; Y_{\Omega_1^c}|X_{\Omega_1^c})\\ &\quad + I(X_{\Omega_2}; Y_{\Omega_2^c}|X_{\Omega_2^c})\\ &= R_{\mathcal{A};j}(p) + R_{\mathcal{B};j}(p),\end{aligned}$$

where (a) is true because for independent random variables A, B, C, D and any random variable Y, we have (see e.g. [46]) $I(A, B, C; Y|D) \leq I(A, B; Y|C, D) + I(A, C; Y|B, D)$.

A.3 Proof of Theorem 3.4

The claim of the theorem is that communication at a rate arbitrarily close to the one given in the optimization (3.8) is possible with perfect secrecy. It suffices to show the existence of a block code that achieves this.

Let θ be such that

$$(R^{\mathrm{BC}}_{m_A}(\theta), R^{\mathrm{BC}}_{m_B}(\theta)) \in \mathcal{R}^{\mathrm{MAC}}_{D} \tag{A.9}$$

and such that

$$\Phi_E \cap \big([0, R^{\mathrm{BC}}_{m_A}(\theta)] \times [0, R^{\mathrm{BC}}_{m_B}(\theta)]\big) \tag{A.10}$$

is non-empty, where

$$\begin{aligned}\Phi_E = \big\{&(R_{j_A}, R_{j_B}) \in \mathcal{R}^{\mathrm{MAC}}_{E} :\\ &R_{j_A} + R_{j_B} = \frac{1}{2}\log(1 + h^2_{AE} + h^2_{BE})\big\}.\end{aligned} \tag{A.11}$$

Then, define $R_D = R^{\mathrm{BC}}_{m_A}(\theta) + R^{\mathrm{BC}}_{m_B}(\theta)$ and $R_E = \frac{1}{2}\log(1 + h^2_{AE} + h^2_{BE})$, and let $R = R_D - R_E - \epsilon_1$. Let W be the information message, uniformly distributed in $\{1, \ldots, 2^{TR}\}$. We split W into two parts, W_A and W_B, of rate R_{w_A} and R_{w_B}, respectively, such that W_k is uniformly distributed in $\{1, \ldots, 2^{TR_{w_k}}\}$, for $k \in \{A, B\}$. Before each transmission block, the source S generates two junk messages J_A and J_B, uniformly distributed in $\{1, \ldots, 2^{TR_{j_A}}\}$ and $\{1, \ldots, 2^{TR_{j_B}}\}$, respectively. Define $M_A = (W_A, J_A)$ and $M_B = (W_B, J_B)$. The junk rates R_{j_A} and R_{j_B} are picked arbitrarily such that

$$(R_{j_A}, R_{j_B}) \in \Phi_E \cap \big([0, R^{\mathrm{BC}}_{m_A}(\theta)] \times [0, R^{\mathrm{BC}}_{m_B}(\theta)]\big).$$

The information rates are chosen as

$$R_{w_k} = R^{\mathrm{BC}}_{m_k}(\theta) - R_{j_k} - \frac{\epsilon_1}{2},$$

for $k \in \{A, B\}$. Note that this choice ensures

$$R_{m_k} = R_{w_k} + R_{j_k} < R^{\mathrm{BC}}_{m_k}(\theta), \tag{A.12}$$

and hence, by the Gaussian broadcast channel achievability result [9], there exists a broadcast code from S to A and B such that A can decode (M_A, M_B) and B can decode M_B, with arbitrarily small error probability.

The relays operate as follows: A discards message M_B and encodes M_A only, while B encodes M_B. The encoding function used at k is a randomly generated mapping from $\{1, \ldots, 2^{TR_{m_k}}\}$ to the set of X_k-typical sequences of length T, for $k = A, B$. This code is generated once at random, and is fixed and deterministic thereafter. From the MAC achievability [9], it follows that the two properties

$$(R_{m_A}, R_{m_B}) \in \mathcal{R}^{\mathrm{MAC}}_{D}$$

and

$$(R_{j_A}, R_{j_B}) \in \mathcal{R}_E^{\mathrm{MAC}}$$

imply that

$$\mathrm{E}\Bigg[\mathbf{P}(\text{wrong decoding of } (M_A, M_B) \text{ at } D)+$$

$$\frac{1}{2^{T(R_{w_A}+R_{w_B})}}\sum_{w_A=1}^{2^{TR_{w_A}}}\sum_{w_B=1}^{2^{TR_{w_B}}}\mathbf{P}(\tilde{A}_{w_A,w_B})\Bigg] \leq \epsilon_0,$$

where $\tilde{A}_{w_A,w_B}$ is the event that E wrongly decodes (J_A, J_B) given that $(W_A, W_B) = (w_A, w_B)$ is available at E. From this, using an adaption of Fano's inequality similar to Lemma A.4, it follows that there exists a code for which the error probability at D can be made small, and for which

$$H(J_A, J_B|W_A, W_B, \mathbf{Y}_E) \leq T\epsilon_2.$$

Now, we are ready to bound the equivocation at E. We have

$$\begin{aligned}
H(W|\mathbf{Y}_E) &= H(W_A, W_B|\mathbf{Y}_E)\\
&\geq I(W_A, W_B; \mathbf{X}_A, \mathbf{X}_B|\mathbf{Y}_E)\\
&= H(\mathbf{X}_A, \mathbf{X}_B|\mathbf{Y}_E) - H(\mathbf{X}_A, \mathbf{X}_B|\mathbf{Y}_E, W_A, W_B)\\
&\overset{(a)}{\geq} H(\mathbf{X}_A, \mathbf{X}_B|\mathbf{Y}_E) - H(M_A, M_B|\mathbf{Y}_E, W_A, W_B)\\
&= H(\mathbf{X}_A, \mathbf{X}_B) - I(\mathbf{X}_A, \mathbf{X}_B; \mathbf{Y}_E) - H(J_A, J_B|\mathbf{Y}_E, W_A, W_B)\\
&\overset{(b)}{\geq} H(M_A, M_B) - T\epsilon_3 - I(\mathbf{X}_A, \mathbf{X}_B; \mathbf{Y}_E) - H(J_A, J_B|\mathbf{Y}_E, W_A, W_B)\\
&\geq T\big(R_{m_A}^{\mathrm{BC}}(\theta) + R_{m_B}^{\mathrm{BC}}(\theta)\big)\\
&- I(\mathbf{X}_A, \mathbf{X}_B; \mathbf{Y}_E) - T\epsilon_1 - T\epsilon_2 - T\epsilon_3\\
&\overset{(c)}{\geq} T\big(R_{m_A}^{\mathrm{BC}}(\theta) + R_{m_B}^{\mathrm{BC}}(\theta)\big)\\
&- T\frac{1}{2}\log(1 + h_{AE}^2 + h_{BE}^2) - T\epsilon_1 - T\epsilon_2 - T\epsilon_3,
\end{aligned}$$

where (a) is true because $(M_A, M_B) = (W_A, W_B, J_A, J_B) \multimap (\mathbf{X}_A, \mathbf{X}_B) \multimap \mathbf{Y}_E$ forms a Markov chain, (b) follows from Fano's inequality, and in (c) we upper bound $I(\mathbf{X}_A, \mathbf{X}_B; \mathbf{Y}_E)$ by the sum-rate capacity of the multiple access channel from (A, B) to E. Therefore, by choosing ϵ_1, ϵ_2 and ϵ_3 small enough, the equivocation $\frac{1}{T}H(W|\mathbf{Y}_E)$ can be made arbitrarily close to $R_D - R_E$. Hence we weakly achieve $(R_D - R_E, R_D - R_E)$. Finally, we use Lemma E.5 to conclude that $R_s = R_D - R_E$ is strongly achievable.

A.4 Proof of Theorem 3.6

Consider an achievable rate tuple $R_{\mathcal{S}}$. Fix an arbitrarily small $\epsilon > 0$. From Theorem 14.10.1 in [9], we know that there exists a distribution $p(\{x_i\}_{i\in\mathcal{V}})$

with the following property. For any cut Ω such that $D \in \Omega^c$, we have

$$\sum_{i \in \Omega \cap \mathcal{S}} R_i - \epsilon' \leq I(X_\Omega; Y_{\Omega^c} | X_{\Omega^c}), \tag{A.13}$$

where $\epsilon' > 0$ can be made arbitrarily small by choosing ϵ small. We can write (A.13) in a slightly different form: For any subset $\mathcal{I}$ of the sources $\mathcal{S}$, and for any cut $\Omega \in \Lambda(\mathcal{I}; D)$, we have

$$\sum_{i \in \mathcal{I}} R_i - \epsilon' \leq I(X_\Omega; Y_{\Omega^c} | X_{\Omega^c}).$$

This implies that for any subset $\mathcal{I} \subseteq \mathcal{S}$, we have

$$\sum_{i \in \mathcal{I}} R_i - \epsilon \leq \min_{\Omega \in \Lambda(\mathcal{I}; D)} I(X_\Omega; Y_{\Omega^c} | X_{\Omega^c}).$$

Therefore, $R_\mathcal{S} \in \mathcal{R}_{\mathcal{S};D}(p)$ for the given distribution p. Since such a distribution can be found for any achievable $R_\mathcal{S}$, the claim of the theorem follows.

A.5 Proof of Theorem 3.7

A.5.1 Code Construction for Layered Networks

First, assume that the network is layered in the sense that for each source node $i \in \mathcal{S}$ and each node $j \in \mathcal{V}$, all the paths from i to j have the same length. Fix any product transmit distribution $\prod_{i \in \mathcal{V}} p(x_i)$. Let $R_\mathcal{S} \in \mathcal{R}_{\mathcal{S};D}(p)$ be given. Fix a block length T. We construct a random block code in the following way. The encoding function f_i at node $i \in \mathcal{V}$ maps each possible sequence $\mathbf{y}_i$ and each possible source symbol $w_i \in \{1, \ldots, 2^{\lfloor TR_i \rfloor}\}$ (if $i \in \mathcal{S}$) to a transmit sequence $\mathbf{x}_i$ that is chosen uniformly at random from the set $\mathcal{T}_\delta(X_i)$ of all robustly X_i-typical sequences of length T (see Appendix F). Hence, for every node $i \in \mathcal{V}$, we know that its predecessors can only transmit robustly typical sequences, and therefore, we know that $\mathbf{Y}_i \in \mathcal{T}_\delta(Y_i)$, where $\mathcal{T}_\delta(Y_i)$ is defined by the channel function $g_i(\cdot)$ and the typical sets of all the predecessors of i, *i.e.*,

$$\mathcal{T}_\delta(Y_i) = g_i(\mathcal{T}_\delta(X_{\text{In}(i)})).$$

The constant δ can be chosen arbitrarily in the interval $(0, 1)$. Once the encoding functions $f_\mathcal{V}$ are constructed in this random way, they are deterministic and fixed for all time.

In this section, $(\mathbf{X}_\mathcal{V}, \mathbf{Y}_\mathcal{V})$ denote sequences that depend on the *same* realization of the source messages $W_\mathcal{S}$. By $w_\mathcal{S}$, we denote a particular realization of the source messages. We illustrate this using the example network in Figure A.2. The message tuple is the pair $W_\mathcal{S} = (W_{S_1}, W_{S_2})$, and the network has three layers of nodes: (S_2, A), (B_1, B_2), and D. The node S_1 encodes W_{S_1} into

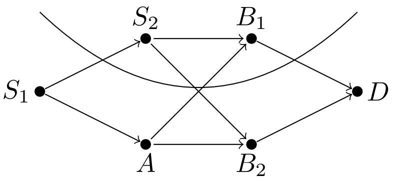

Figure A.2: An example of a network with two source nodes S_1 and S_2.

a sequence $\mathbf{X}_{S_1}$ which it transmits during the first block of T time-slots. During the second block of T time-slots, node S_2 transmits a sequence $\mathbf{X}_{S_2}$ which is a function of $(W_{S_2}, \mathbf{Y}_{S_2})$, where $\mathbf{Y}_{S_2}$ was received during the first block. Similarly, A transmits $\mathbf{X}_A = f_A(\mathbf{Y}_A)$ during the second block. Of course, during this second block, S_1 can already transmit a new message. Communication from B_1 and B_2 to D happens during the third block.

The destination D produces a tuple of messages $\hat{w}_{\mathcal{S}}$ that is jointly typical with the received sequence $\mathbf{Y}_D$. Note that for a fixed code, the received sequences are functions of the message tuple $W_{\mathcal{S}}$. Hence, joint typicality of $\hat{w}_{\mathcal{S}}$ and $\mathbf{y}_D$ simply means that $\mathbf{y}_D$ is *the* output at D produced by transmitting $\hat{w}_{\mathcal{S}}$. If this $\hat{w}_{\mathcal{S}}$ is not unique, we declare an error.

A.5.2 Definitions

Definition A.1 *For $\mathcal{I} \subseteq \mathcal{S}$, we denote by $\mathcal{N}(\mathcal{I})$ the subset of nodes in $\mathcal{V}$ that are* not *in the flow of $\mathcal{I}$. More precisely, $\mathcal{N}(\mathcal{I})$ is the set of nodes $j \in \mathcal{V}$ such that there is no path from any of the nodes in $\mathcal{I}$ to j, i.e., $\mathcal{N}(\mathcal{I}) = \cap_{i \in \mathcal{I}} \mathcal{N}(\{i\})$.*

For instance, in the example of Figure A.2, $\mathcal{N}(\{S_1\}) = \mathcal{N}(\{S_1, S_2\}) = \phi$ while $\mathcal{N}(\{S_2\}) = \{S_1, A\}$.

Definition A.2 *Assume that node i receives $\mathbf{y}_i$ if $w_{\mathcal{S}}$ was sent and $\mathbf{y}'_i$ if $w'_{\mathcal{S}}$ was sent. Then, for the given code, we say that node i **can distinguish** between $w_{\mathcal{S}}$ and $w'_{\mathcal{S}}$ if $\mathbf{y}_i \neq \mathbf{y}'_i$ and that node i **cannot distinguish** between $w_{\mathcal{S}}$ and $w'_{\mathcal{S}}$ if $\mathbf{y}_i = \mathbf{y}'_i$.*

A.5.3 Error Analysis

The execution of the code defined above is governed by two sources of randomness: one is the random construction of the code, the other is the random choice of the message tuple $W_{\mathcal{S}}$. Assume now that a certain message tuple $w_{\mathcal{S}}$ was sent. This eliminates one source of randomness. Hence, the probabilities in this section are over the random choice of the code. We can write the error probability at D as

$$\begin{aligned}\mathbf{P}(\text{err}) &\le \sum_{w'_{\mathcal{S}} \neq w_{\mathcal{S}}} \mathbf{P}\Big(\substack{D \text{ cannot distinguish} \\ w_{\mathcal{S}} \text{ and } w'_{\mathcal{S}}}\Big) \\ &= \sum_{\mathcal{I} \subseteq \mathcal{S}} \sum_{\substack{w'_{\mathcal{S}} : w'_i \neq w_i \forall i \in \mathcal{I} \\ w'_{\mathcal{I}^c} = w_{\mathcal{I}^c}}} \mathbf{P}\Big(\substack{D \text{ cannot distinguish} \\ w_{\mathcal{S}} \text{ and } w'_{\mathcal{S}}}\Big). \end{aligned} \tag{A.14}$$

We now focus on upper bounding each term in the above sum. Fix a subset $\mathcal{I} \subseteq \mathcal{S}$, and fix a $w'_{\mathcal{S}} \neq w_{\mathcal{S}}$ such that $w'_i \neq w_i$ for all $i \in \mathcal{I}$ and $w'_{\mathcal{I}^c} = w_{\mathcal{I}^c}$. We have

$$\begin{aligned}&\mathbf{P}(D \text{ cannot distinguish } w_{\mathcal{S}} \text{ and } w'_{\mathcal{S}}) \\ &= \sum_{\substack{\Omega : \mathcal{I} \subseteq \Omega \\ D \in \Omega^c}} \mathbf{P}\Big(\substack{\Omega \text{ can distinguish } w_{\mathcal{S}} \text{ and } w'_{\mathcal{S}} \\ \text{but } \Omega^c \text{ cannot}}\Big) \\ &\overset{(a)}{=} \sum_{\substack{\Omega : \mathcal{I} \subseteq \Omega \\ \mathcal{N}(\mathcal{I}) \cup \{D\} \subseteq \Omega^c}} \mathbf{P}\Big(\substack{\Omega \text{ can distinguish } w_{\mathcal{S}} \text{ and } w'_{\mathcal{S}} \\ \text{but } \Omega^c \text{ cannot}}\Big)\end{aligned} \tag{A.15}$$

where (a) holds because the terms for which $\mathcal{N}(\mathcal{I}) \nsubseteq \Omega^c$ are zero. The reason for this is the following. Since the nodes in $\mathcal{N}(\mathcal{I})$ are not in the flow of $\mathcal{I}$, and since $w_{\mathcal{I}^c} = w'_{\mathcal{I}^c}$, they receive and send the same sequences under $w_{\mathcal{S}}$ and $w'_{\mathcal{S}}$. Hence, they cannot distinguish between $w_{\mathcal{S}}$ and $w'_{\mathcal{S}}$.

One can show that each term in (A.15) is upper bounded as

$$\mathbf{P}\Big(\substack{\Omega \text{ can distinguish } w_{\mathcal{S}} \text{ and } w'_{\mathcal{S}} \\ \text{but } \Omega^c \text{ cannot}}\Big) \le 2^{-TI(X_\Omega;Y_{\Omega^c}|X_{\Omega^c})}\, 2^{T|\mathcal{V}|\epsilon} \tag{A.16}$$

for an arbitrary $\epsilon > 0$. The proof of this bound can be found in [5]. Here, we illustrate the proof steps using the example network in Figure A.2. Assume that $w_{\mathcal{S}}$ and $w'_{\mathcal{S}}$ are such that $w_{S_1} = w'_{S_1}$ but $w_{S_2} \neq w'_{S_2}$, *i.e.*, $\mathcal{I} = \{S_2\}$. Consider the cut $\Omega = \{S_2, B_1\}$. In Figure A.2, the separation between Ω and Ω^c is indicated by a curve. Let $\mathcal{E}_i$ denote the event that node i cannot distinguish $w_{\mathcal{S}}$ and $w'_{\mathcal{S}}$. We have

$$\begin{aligned}&\mathbf{P}\Big(\substack{\Omega \text{ can distinguish } w_{\mathcal{S}} \text{ and } w'_{\mathcal{S}} \\ \text{but } \Omega^c \text{ cannot}}\Big) \\ &= \mathbf{P}(\mathcal{E}_{S_1}, \mathcal{E}_A, \mathcal{E}^c_{S_2}, \mathcal{E}_{B_2}, \mathcal{E}^c_{B_1}, \mathcal{E}_D) \\ &= \underbrace{\mathbf{P}(\mathcal{E}_{S_1})}_{\le 1} \underbrace{\mathbf{P}(\mathcal{E}_A|\mathcal{E}_{S_1})}_{\le 1} \underbrace{\mathbf{P}(\mathcal{E}^c_{S_2}|\mathcal{E}_{S_1}, \mathcal{E}_A)}_{\le 1} \\ &\quad \mathbf{P}(\mathcal{E}_{B_2}|\mathcal{E}_{S_1}, \mathcal{E}_A, \mathcal{E}^c_{S_2}) \underbrace{\mathbf{P}(\mathcal{E}^c_{B_1}|\mathcal{E}_{S_1}, \mathcal{E}_A, \mathcal{E}^c_{S_2}, \mathcal{E}_{B_2})}_{\le 1} \\ &\quad \mathbf{P}(\mathcal{E}_D|\mathcal{E}_{S_1}, \mathcal{E}_A, \mathcal{E}^c_{S_2}, \mathcal{E}_{B_2}, \mathcal{E}^c_{B_1}) \\ &\le \mathbf{P}(\mathcal{E}_{B_2}|\mathcal{E}_{S_1}, \mathcal{E}_A, \mathcal{E}^c_{S_2}) \\ &\quad \mathbf{P}(\mathcal{E}_D|\mathcal{E}_{S_1}, \mathcal{E}_A, \mathcal{E}^c_{S_2}, \mathcal{E}_{B_2}, \mathcal{E}^c_{B_1}) \\ &= \mathbf{P}(\mathcal{E}_{B_2}|\mathcal{E}_A, \mathcal{E}^c_{S_2})\mathbf{P}(\mathcal{E}_D|\mathcal{E}_{B_2}, \mathcal{E}^c_{B_1}),\end{aligned} \tag{A.17}$$

where the last equality holds because of the causality across the network layers. The four probabilities that we upper bounded by 1 are indeed close to 1, because for nodes S_1 and A that are in $\mathcal{N}(\mathcal{I})$, the probability of not distinguishing $w_{\mathcal{S}}$ and $w'_{\mathcal{S}}$ is 1, while for any node that is not in $\mathcal{N}(\mathcal{I})$, the probability of distinguishing is close to 1. In general, we use this argument to eliminate all the probabilities for nodes in $\Omega \cup \mathcal{N}(\mathcal{I})$. Given that $\mathcal{E}_A$ is true, we know that A sends the same sequence $\mathbf{x}_A = \mathbf{x}'_A$ under $w_{\mathcal{S}}$ and under $w'_{\mathcal{S}}$. Hence, $\mathbf{P}(\mathcal{E}_{B_2}|\mathcal{E}_A, \mathcal{E}^c_{S_2})$ is the probability that two independently chosen sequences $\mathbf{x}_{S_2}$ and $\mathbf{x}'_{S_2}$ lead to the equality $\mathbf{y}'_{B_2} = \mathbf{y}_{B_2}$, or $g_{B_2}(\mathbf{x}_{S_2}, \mathbf{x}_A) = g_{B_2}(\mathbf{x}'_{S_2}, \mathbf{x}_A)$. Hence, we have

$$\begin{aligned}
\mathbf{P}(\mathcal{E}_{B_2}|\mathcal{E}_A, \mathcal{E}^c_{S_2}) &= \mathbf{P}(\mathbf{y}_{B_2} = g_{B_2}(\mathbf{x}'_{S_2}, \mathbf{x}_A)) \\
&= \mathbf{P}\big((\mathbf{x}'_{S_2}, \mathbf{x}_A, \mathbf{y}_{B_2}) \in \mathcal{T}_\delta \mid (\mathbf{x}_A, \mathbf{y}_{B_2}) \in \mathcal{T}_\delta, \mathbf{x}'_{S_2} \perp (\mathbf{x}_A, \mathbf{y}_{B_2})\big) \\
&\overset{(a)}{\leq} 2^{-T(I(X_{S_2};X_A,Y_{B_2})-2\delta H(X_A,Y_{B_2}))} \\
&= 2^{-TI(X_{S_2};X_A,Y_{B_2})}\ 2^{T2\delta H(X_A,Y_{B_2})} \\
&\leq 2^{-TI(X_{S_2};X_A,Y_{B_2})}\ 2^{T\epsilon},
\end{aligned} \tag{A.18}$$

for an arbitrary $\epsilon > 0$ if we choose δ small enough. In (a), we have used Lemma F.4. Similarly,

$$\mathbf{P}(\mathcal{E}_D|\mathcal{E}_{B_2}, \mathcal{E}^c_{B_1}) \leq 2^{-TI(X_{B_1};X_{B_2},Y_D)}\ 2^{T\epsilon}$$

if δ is chosen small enough. Plugging this back into (A.17), we obtain,

$$\begin{aligned}
&\mathbf{P}\big(\substack{\Omega \text{ can distinguish } w_{\mathcal{S}} \text{ and } w'_{\mathcal{S}} \\ \text{but } \Omega^c \text{ cannot}}\big) \\
&\leq 2^{-T(I(X_{S_2};X_A,Y_{B_2})+I(X_{B_1};X_{B_2},Y_D))}\ 2^{T2\epsilon} \\
&\leq 2^{-T(I(X_{S_2};X_A,Y_{B_2})+I(X_{B_1};X_{B_2},Y_D))}\ 2^{T|\mathcal{V}|\epsilon}.
\end{aligned} \tag{A.19}$$

Note that for the given cut, the Markov structure of the layered network tells us that

$$\begin{aligned}
I(X_\Omega; Y_{\Omega^c}|X_{\Omega^c}) &= I(X_{S_2}, X_{B_1}; Y_A, Y_{B_2}, Y_D|X_{S_1}, X_A, X_{B_2}) \\
&= I(X_{S_2}, X_{B_1}; Y_A|X_{S_1}, X_A, X_{B_2}) \\
&\quad + I(X_{S_2}, X_{B_1}; Y_{B_2}|X_{S_1}, X_A, X_{B_2}, Y_A) \\
&\quad + I(X_{S_2}, X_{B_1}; Y_D|X_{S_1}, X_A, X_{B_2}, Y_A, Y_{B_2}) \\
&\overset{(a)}{=} 0 \\
&\quad + I(X_{S_2}; Y_{B_2}|X_A) \\
&\quad + I(X_{B_1}; Y_D|X_{B_2}) \\
&\overset{(b)}{=} I(X_{S_2}; Y_{B_2}, X_A) - \underbrace{I(X_{S_2}; X_A)}_{=0} \\
&\quad + I(X_{B_1}; Y_D, X_{B_2}) - \underbrace{I(X_{B_1}; X_{B_2})}_{=0},
\end{aligned}$$

where (a) follows from the Markov structure of the network layers, and in (b), the indicated terms are zero because of the independence of the $p(x_i)$. Hence, (A.19) is equivalent to (A.16) for this particular cut in our example network. The same type of proof works for any cut and for general networks.

We finally plug (A.16) back into (A.15) and (A.14) and obtain that

$$\begin{aligned}
\mathbf{P}(\mathrm{err}) &\leq \sum_{\mathcal{I}\subseteq\mathcal{S}} \sum_{\substack{w'_{\mathcal{S}}: \\ w'_i \neq w_i \forall i\in\mathcal{I} \\ w'_{\mathcal{I}^c} = w_{\mathcal{I}^c}}} \sum_{\substack{\Omega:\mathcal{I}\subseteq\Omega \\ \mathcal{N}(\mathcal{I})\cup\{D\}\subseteq\Omega^c}} 2^{-TI(X_\Omega;Y_{\Omega^c}|X_{\Omega^c})}\, 2^{T|\mathcal{V}|\epsilon} \\
&\leq \sum_{\mathcal{I}\subseteq\mathcal{S}} \Big(\prod_{i\in\mathcal{I}} 2^{TR_i}\Big) \sum_{\substack{\Omega:\mathcal{I}\subseteq\Omega \\ \mathcal{N}(\mathcal{I})\cup\{D\}\subseteq\Omega^c}} 2^{-TI(X_\Omega;Y_{\Omega^c}|X_{\Omega^c})}\, 2^{T|\mathcal{V}|\epsilon} \\
&\overset{(a)}{\leq} \sum_{\mathcal{I}\subseteq\mathcal{S}} \Big(\prod_{i\in\mathcal{I}} 2^{TR_i}\Big)\, c\, 2^{-T\min_\Omega I(X_\Omega;Y_{\Omega^c}|X_{\Omega^c})}\, 2^{T|\mathcal{V}|\epsilon} \\
&= c \sum_{\mathcal{I}\subseteq\mathcal{S}} 2^{T\sum_{i\in\mathcal{I}} R_i} 2^{-T\min_\Omega I(X_\Omega;Y_{\Omega^c}|X_{\Omega^c})}\, 2^{T|\mathcal{V}|\epsilon},
\end{aligned}$$

where in (a), c is an upper bound on the total number of cuts possible in a network. The constant c is chosen such that it depends only on the number of nodes $|\mathcal{V}|$. Hence, we see that if the rates are such that

$$\sum_{i\in\mathcal{I}} R_i < \min_{\substack{\Omega:\mathcal{I}\subseteq\Omega \\ \{D\}\cup\mathcal{N}(\mathcal{I})\subseteq\Omega^c}} I(X_\Omega;Y_{\Omega^c}|X_{\Omega^c}) - \epsilon|\mathcal{V}| \tag{A.20}$$

for all $\mathcal{I}\subseteq\mathcal{S}$, then the error probability goes to zero exponentially fast with T, where the probability was taken over all randomly generated codes and for a fixed transmit message $w_{\mathcal{S}}$. In this case, the expected error probability over all messages also goes to zero (indeed, exponentially fast). From this, we can conclude that there exists at least one code for which the error probability can be made arbitrarily small by choosing a sufficiently large T. As long as

$$\sum_{i\in\mathcal{I}} R_i < \min_{\substack{\Omega:\mathcal{I}\subseteq\Omega \\ \{D\}\cup\mathcal{N}(\mathcal{I})\subseteq\Omega^c}} I(X_\Omega;Y_{\Omega^c}|X_{\Omega^c}),$$

we can always find an $\epsilon > 0$ for which (A.20) is satisfied. Since $R_{\mathcal{S}} \in \mathcal{R}_{\mathcal{S};D}(p)$, we know that for all $\mathcal{I}\subseteq\mathcal{S}$,

$$\sum_{i\in\mathcal{I}} R_i < \min_{\substack{\Omega:\mathcal{I}\subseteq\Omega \\ D\in\Omega^c}} I(X_\Omega;Y_{\Omega^c}|X_{\Omega^c}).$$

It remains to show that for any distribution $p(\{x_i\}_{i\in\mathcal{V}})$,

$$\min_{\substack{\Omega:\mathcal{I}\subseteq\Omega \\ \{D\}\cup\mathcal{N}(\mathcal{I})\subseteq\Omega^c}} I(X_\Omega;Y_{\Omega^c}|X_{\Omega^c}) = \min_{\substack{\Omega:\mathcal{I}\subseteq\Omega \\ D\in\Omega^c}} I(X_\Omega;Y_{\Omega^c}|X_{\Omega^c}) \tag{A.21}$$

It is clear that the left hand side in (A.21) is larger than the right hand side. Assume now, that the optimizer in the right hand side of (A.21) is a certain cut Ω^* such that

$$\Omega^* \cap \mathcal{N}(\mathcal{I}) \neq \phi.$$

Define $\Omega' = \Omega^* \setminus \mathcal{N}(\mathcal{I})$, and note that Ω' is a member of the minimization domain in the left hand side of (A.21). Note also that $\Omega'^c = \Omega^{*c} \cup \mathcal{N}(\mathcal{I})$. Thus,

$$\begin{aligned}
I(X_{\Omega'}; Y_{\Omega'^c} | X_{\Omega'^c}) &= I(X_{\Omega'}; Y_{\Omega^{*c}} | X_{\Omega'^c}) \\
&\quad + I(X_{\Omega'}; Y_{\mathcal{N}(\mathcal{I}) \setminus \Omega^{*c}} | Y_{\Omega^{*c}}, X_{\Omega'^c}) \\
&\overset{(a)}{=} I(X_{\Omega'}; Y_{\Omega^{*c}} | X_{\Omega'^c}) \\
&\overset{(b)}{=} I(X_{\Omega'}; Y_{\Omega^{*c}} | X_{\Omega^{*c}}, X_{\Omega^* \setminus \Omega'}) \\
&= I(X_{\Omega'}, X_{\Omega^* \setminus \Omega'}; Y_{\Omega^{*c}} | X_{\Omega^{*c}}) \\
&\quad - I(X_{\Omega^* \setminus \Omega'}; Y_{\Omega^{*c}} | X_{\Omega^{*c}}) \\
&\leq I(X_{\Omega^*}; Y_{\Omega^{*c}} | X_{\Omega^{*c}}) \\
&= \min_{\substack{\Omega: \mathcal{I} \subseteq \Omega \\ D \in \Omega^c}} I(X_\Omega; Y_{\Omega^c} | X_{\Omega^c}),
\end{aligned}$$

where (a) is true because of the following. Ω'^c contains $\mathcal{N}(\mathcal{I})$ while Ω' is disjoint from $\mathcal{N}(\mathcal{I})$. Since $\mathcal{N}(\mathcal{I})$ is exclusively in the flow of itself and the random variables X_i, $i \in \mathcal{V}$ are independent, we have the Markov chain $Y_{\mathcal{N}(\mathcal{I})} -\!\!\circ\!\!- X_{\mathcal{N}(\mathcal{I})} -\!\!\circ\!\!- X_{\Omega'}$. The identity (b) is true because $\Omega' \subseteq \Omega^*$ implies Ω'^c is the disjoint union of Ω^{*c} and $\Omega^* \setminus \Omega'$. It follows that the left hand side of (A.21) is smaller than the right hand side. Hence the equality.

Thus, the set of rates $R_{\mathcal{S}}$ is achievable if it is in $\mathcal{R}_{\mathcal{S};D}(p)$.

A.5.4 Non-Layered Networks

For non-layered networks, we use the time-expansion technique described in Appendix A.12. The set $\mathcal{S}$ defined in Appendix A.12 is the set of all source nodes, also denoted by $\mathcal{S}$ in this section. The set $\mathcal{D}$ is of size one and equal to $\{D\}$. Given a non-layered graph $\mathcal{G}$, we construct an unfolded graph $\mathcal{G}_{\text{unf}}^{(K)}$ that is layered. Let $R_{\mathcal{S}} \in \mathcal{R}_{\text{inner}}$ be given. Let $\prod_{i \in \mathcal{V}} p(x_i)$ be a product distribution such that $R_{\mathcal{S}} \in \mathcal{R}_{\mathcal{S};D}(p)$. From Lemma A.7, we know that asymptotically with large K, the rate-tuple $KR_{\mathcal{S}}$ is in $\mathcal{R}_{\mathcal{S};D}^{\text{unf}}(p)$, where $\mathcal{R}_{\mathcal{S};D}^{\text{unf}}(p)$ is defined in Definition A.4. We can apply the above proof to find that $KR_{\mathcal{S}}$ is achievable in $\mathcal{G}_{\text{unf}}^{(K)}$. Hence, we know that there exists a coding scheme for $\mathcal{G}_{\text{unf}}^{(K)}$ that achieves rates $KR_{\mathcal{S}}$. We can implement such a scheme on the original network by operating the scheme during K stages. By doing so, we obtain transmission rates K times smaller than in the virtual network $\mathcal{G}_{\text{unf}}^{(K)}$. It follows that $R_{\mathcal{S}}$ is achievable in $\mathcal{G}$.

A.6 Proof of Theorem 3.8

It is clear that $\mathcal{R}_{\text{inner}} \subseteq \mathcal{R}_{\text{outer}}$. In this proof, we show that the opposite inclusion is also true. Let a tuple $R_{\mathcal{S}} \in \mathcal{R}_{\text{outer}}$ be given. Let p^* be the distribution on $X_{\mathcal{V}}$ such that $R_{\mathcal{S}} \in \mathcal{R}_{\mathcal{S};D}(p^*)$, let $X^*_{\mathcal{V}}$ be the corresponding random variables and let $Y^*_{\mathcal{V}}$ be the received random variables when $X^*_{\mathcal{V}}$ is at the input of the channels of the network. Fix any subset $\mathcal{I} \subseteq \mathcal{S}$. From the definition of $\mathcal{R}_{\mathcal{S};D}(p^*)$, we have that

$$\begin{aligned}\sum_{i\in\mathcal{I}} R_i &\leq \min_{\Omega\in\Lambda(\mathcal{I};D)} I(X^*_{\Omega}; Y^*_{\Omega^c}|X^*_{\Omega^c}) \\ &= \min_{\Omega\in\Lambda(\mathcal{I};D)} H(Y^*_{\Omega^c}|X^*_{\Omega^c}),\end{aligned}$$

where we used the fact that the channels are deterministic, and hence $H(Y^*_{\Omega^c}|X^*_{\mathcal{V}}) = 0$. For any fixed Ω, we have, due to the linear interaction model,

$$\begin{aligned}H(Y^*_{\Omega^c}|X^*_{\Omega^c}) &= H(\mathbf{G}_{\mathcal{V},\Omega^c}X^*_{\mathcal{V}}|X^*_{\Omega^c}) \\ &= H(\mathbf{G}_{\Omega,\Omega^c}X^*_{\Omega} + \mathbf{G}_{\Omega^c,\Omega^c}X^*_{\Omega^c}|X^*_{\Omega^c}) \\ &= H(\mathbf{G}_{\Omega,\Omega^c}X^*_{\Omega}|X^*_{\Omega^c}) \\ &\leq H(\mathbf{G}_{\Omega,\Omega^c}X^*_{\Omega}) \\ &\leq \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}) \\ &= H(\tilde{Y}_{\Omega^c}|\tilde{X}_{\Omega^c}),\end{aligned}$$

where $(\tilde{X}_{\mathcal{V}}, \tilde{Y}_{\mathcal{V}})$ are the random variables that correspond to the i.i.d. uniform distribution on $\mathcal{X}_{\mathcal{V}}$, denoted by $\tilde{p}$. Define

$$\Omega' \triangleq \arg \min_{\Omega\in\Lambda(\mathcal{I};D)} H(\tilde{Y}_{\Omega^c}|\tilde{X}_{\Omega^c}).$$

Combining all of the above, we obtain that for the fixed subset $\mathcal{I}$,

$$\begin{aligned}\sum_{i\in\mathcal{I}} R_i &\leq \min_{\Omega\in\Lambda(\mathcal{I};D)} H(Y^*_{\Omega^c}|X^*_{\Omega^c}) \\ &\leq H(Y^*_{\Omega'^c}|X^*_{\Omega'^c}) \\ &\leq H(\tilde{Y}_{\Omega'^c}|\tilde{X}_{\Omega'^c}) \\ &= \min_{\Omega\in\Lambda(\mathcal{I};D)} H(\tilde{Y}_{\Omega^c}|\tilde{X}_{\Omega^c}).\end{aligned}$$

Since this is true for any subset $\mathcal{I} \subseteq \mathcal{S}$, we find that $R_{\mathcal{S}} \in \mathcal{R}_{\mathcal{S};D}(\tilde{p}) \subseteq \mathcal{R}_{\text{inner}}$. This concludes the proof.

A.7 Proof of Theorem 3.9

For Gaussian signal interaction, we proceed in a similar way as for deterministic signal interaction in Section A.5. However, there are several differences. First of all, we use the i.i.d. normal distribution $\prod_{i \in \mathcal{V}} p^{\mathrm{G}}(x_i)$ for the transmit symbols. Assume that the network is layered and let communication be over blocks of length T. Since the channel adds noise to the transmitted symbols, the received sequence $\mathbf{Y}_i$ is no longer a function of $\mathbf{X}_{\mathrm{In}(i)}$, but $\mathbf{Y}_i$ can take values in a continuum of different received sequences. We assume now that each node i quantizes the received sequence $\mathbf{Y}_i$ using a Gaussian vector quantizer of distortion 1, producing a representation sequence $\hat{\mathbf{Y}}_i$ that takes values in a finite set. The encoding functions $f_{\mathcal{V}}$ are then constructed in the same way as in Section A.5, with the difference that the function f_i takes $\hat{\mathbf{Y}}_i$ as its input.

The quantizers used in this achievability are constructed as follows. For node i, we pick $2^{T(I(Y_i;\hat{Y}_i)+\epsilon')}$ sequences uniformly at random from $\mathcal{T}_\delta(\hat{Y}_i)$, where $\hat{Y}_i = \alpha_i Y_i + \xi_i$ is as defined in Definition 2.21 and $\epsilon' > 0$ can be chosen arbitrarily small. Every sequence $\mathbf{y}_i$ is then mapped to a sequence $\hat{\mathbf{y}}_i$ such that $(\mathbf{y}_i, \hat{\mathbf{y}}_i) \in \mathcal{T}_\delta$. From rate-distortion theory, it is known that such a mapping is a good Gaussian vector quantizer of distortion 1.

The error analysis up to (A.17) is the same as in Section A.5. Then, to compute for instance the probability $\mathbf{P}(\mathcal{E}_D|\mathcal{E}_{B_2}, \mathcal{E}^c_{B_1})$, we note that there are now exponentially many sequences $\mathbf{x}'_{B_1}$ that are plausible under the message tuple $w'_{\mathcal{S}}$ where "plausible" is defined in Definition A.10. Hence, D cannot distinguish $w_{\mathcal{S}}$ and $w'_{\mathcal{S}}$ if there exists at least one sequence $\mathbf{x}'_{B_1}$ such that $(\mathbf{x}'_{B_1}, \mathbf{x}_{B_2}, \hat{\mathbf{y}}_D) \in \mathcal{T}_\delta$. Let $\underline{\mathcal{X}}_i(w_{\mathcal{S}})$ be the set of all sequences $\mathbf{x}_i$ that are plausible under $w_{\mathcal{S}}$. Similarly, let $\underline{\hat{\mathcal{Y}}}_i(w_{\mathcal{S}})$ be the set of quantized sequences at node i that are plausible under $w_{\mathcal{S}}$. We get

$$\begin{aligned}
\mathbf{P}(\mathcal{E}_D|\mathcal{E}_{B_2}, \mathcal{E}^c_{B_1}) &= \mathbf{P}\Big(\cup_{\mathbf{x}'_{B_1} \in \underline{\mathcal{X}}_{B_1}(w'_{\mathcal{S}})} \big\{(\mathbf{x}'_{B_1}, \mathbf{x}_{B_2}, \hat{\mathbf{y}}_D) \in \mathcal{T}_\delta\big\} \big| \\
&\qquad (\mathbf{x}_{B_2}, \hat{\mathbf{y}}_D) \in \mathcal{T}_\delta, \mathbf{x}_{B_2} \in \underline{\mathcal{X}}_{B_2}(w_{\mathcal{S}}), \hat{\mathbf{y}}_D \in \underline{\hat{\mathcal{Y}}}_D(w_{\mathcal{S}})\Big) \\
&\overset{(a)}{\leq} \sum_{\mathbf{x}'_{B_1} \in \underline{\mathcal{X}}_{B_1}(w'_{\mathcal{S}})} 2^{-T\left(I(X_{B_1};X_{B_2},\hat{Y}_D)-\epsilon\right)} \\
&= |\underline{\mathcal{X}}_{B_1}(w'_{\mathcal{S}})| 2^{-T\left(I(X_{B_1};X_{B_2},\hat{Y}_D)-\epsilon\right)} \\
&\leq |\underline{\mathcal{X}}_{\mathcal{V}}(w'_{\mathcal{S}})| 2^{-T\left(I(X_{B_1};X_{B_2},\hat{Y}_D)-\epsilon\right)} \\
&\overset{(b)}{\leq} 2^{T\gamma} 2^{-T\left(I(X_{B_1};X_{B_2},\hat{Y}_D)-\epsilon\right)},
\end{aligned} \tag{A.22}$$

where (a) is derived as in (A.18) and (b) follows from Lemma A.2 in Section A.10. Since there are at most $|\mathcal{V}|$ terms in (A.17), we get a factor of at most $\left(2^{T\gamma}\right)^{|\mathcal{V}|} \leq 2^{T\beta}$ in front the probability of a given cut Ω. More precisely, (A.16) should be replaced by

$$\mathbf{P}\left(\substack{\Omega \text{ can distinguish } w_{\mathcal{S}} \text{ and } w'_{\mathcal{S}} \\ \text{but } \Omega^c \text{ cannot}}\right) \leq 2^{T\beta}\, 2^{-TI(X_\Omega;\hat{Y}_{\Omega^c}|X_{\Omega^c})}\, 2^{T|\mathcal{V}|\epsilon}. \tag{A.23}$$

The remaining steps of the error probability analysis are again as in Section A.5, and we find that the expected error probability goes to zero as long as for all $\mathcal{I} \subseteq \mathcal{S}$,

$$\sum_{i\in\mathcal{I}} R_i < \min_{\substack{\Omega:\mathcal{I}\subseteq\Omega \\ D\in\Omega^c}} I(X_\Omega;\hat{Y}_{\Omega^c}|X_{\Omega^c}) - \beta,$$

which is true for any $R_\mathcal{S}$ in $\mathcal{R}^{\mathrm{G}}_{S;D}(\beta)$. This concludes the proof for layered networks.

For non-layered networks, we use the time-unfolding technique of Section A.12 in a very similar way as we did in the end of Section A.5. The set $\mathcal{S}$ defined in Section A.12 is the set of all source nodes, also denoted by $\mathcal{S}$ in this section. The set $\mathcal{D}$ is of size one and equal to $\{D\}$. Given a non-layered graph $\mathcal{G}$, we construct an unfolded graph $\mathcal{G}^{(K)}_{\mathrm{unf}}$ that is layered. Let $R_\mathcal{S} \in \mathcal{R}^{\mathrm{G}}_{\mathrm{inner}} = \mathcal{R}^{\mathrm{G}}_{S;D}(\beta)$ be given. This implies that for every $\mathcal{I} \subseteq \mathcal{S}$,

$$\sum_{i\in\mathcal{I}} R_i < \min_{\Omega\in\Lambda(\mathcal{I};D)} I(X_\Omega;\hat{Y}_{\Omega^c}|X_{\Omega^c}) - \beta.$$

Lemma A.7 can be adapted to this case to show that this implies that asymptotically with large K, the rate-tuple $KR_\mathcal{S}$ is in $\mathcal{R}^{\mathrm{unf,G}}_{\mathcal{S};D}$, where $\mathcal{R}^{\mathrm{unf,G}}_{\mathcal{S};D}$ is defined as all rate-tuples $R^{\mathrm{unf}}_\mathcal{S}$ such that for every $\mathcal{I} \subseteq \mathcal{S}$,

$$\sum_{i\in\mathcal{I}} R^{\mathrm{unf}}_i < \min_{\Omega_{\mathrm{unf}}\in\Lambda_{\mathrm{unf}}(\mathcal{I};D)} I(X_{\Omega_{\mathrm{unf}}};\hat{Y}_{\Omega^c_{\mathrm{unf}}}|X_{\Omega^c_{\mathrm{unf}}}) - K\beta.$$

We can apply the above proof to find that $KR_\mathcal{S}$ is achievable in $\mathcal{G}^{(K)}_{\mathrm{unf}}$. When doing so, we use Lemma A.3 instead of Lemma A.2 to conclude that (A.22) still holds (with the same order of γ as before). Since the number of nodes in $\mathcal{G}^{(K)}_{\mathrm{unf}}$ is at most $2K|\mathcal{V}|$, we need to modify (A.23) to

$$\mathbf{P}\Big(\substack{\Omega \text{ can distinguish } w_\mathcal{S} \text{ and } w'_\mathcal{S} \\ \text{but } \Omega^c \text{ cannot}}\Big) \leq 2^{TK\beta}\, 2^{-TI(X_\Omega;\hat{Y}_{\Omega^c}|X_{\Omega^c})}\, 2^{T|\mathcal{V}|\epsilon}. \tag{A.24}$$

Hence, we can conclude that $KR_\mathcal{S}$ is indeed achievable in $\mathcal{G}^{(K)}_{\mathrm{unf}}$. Thus, we know that there exists a coding scheme for $\mathcal{G}^{(K)}_{\mathrm{unf}}$ that achieves rates $KR_\mathcal{S}$. We can implement such a scheme on the original network by operating the scheme during K stages. By doing so, we obtain transmission rates K times smaller than in the virtual network $\mathcal{G}^{(K)}_{\mathrm{unf}}$. It follows that $R_\mathcal{S}$ is achievable in $\mathcal{G}$.

A.8 Proof of Theorem 3.10

Let a multisource network with Gaussian signal interaction be given, and let $R_\mathcal{S}$ be any rate-tuple in $\mathcal{R}_{\mathrm{outer}}$. From the definition of $\mathcal{R}_{\mathrm{outer}}$, it follows that there exists a distribution $p^*(\{x_i\}_{i\in\mathcal{V}})$ such that $R_\mathcal{S} \in \mathcal{R}_{\mathcal{S};D}(p^*)$. Fix any $\mathcal{I} \subseteq \mathcal{S}$. From Lemma A.1 in Section A.9, we know that

$$R_{\mathcal{I};D}(p^*) - \alpha \leq R^{\mathrm{G}}_{\mathcal{I};D}. \tag{A.25}$$

Now consider the shifted tuple $R_{\mathcal{S}} - (\alpha+\beta)\mathbf{1}$ and sum the components of this tuple that are indicated by the set $\mathcal{I}$:

$$\begin{aligned}\sum_{i\in\mathcal{I}}(R_i - \alpha - \beta) &\leq \sum_{i\in\mathcal{I}} R_i - \alpha - \beta \\ &\leq R_{\mathcal{I};D}(p^*) - \alpha - \beta \\ &\overset{(a)}{\leq} R^{\mathrm{G}}_{\mathcal{I};D} - \beta,\end{aligned}$$

where (a) holds because of (A.25). We repeat this argument for every possible $\mathcal{I} \subseteq \mathcal{S}$ and we conclude that

$$R_{\mathcal{S}} - (\alpha+\beta)\mathbf{1} \in \mathcal{R}^{\mathrm{G}}_{\mathcal{S};D}(\beta) = \mathcal{R}_{\text{inner}}.$$

This proves the claim of the theorem.

A.9 Relating Definitions 2.20 and 2.21

Lemma A.1 *Let a Gaussian relay network $\mathcal{G}$ be given, and consider a subset $\mathcal{I} \subseteq \mathcal{V}$ and a single node $j \in \mathcal{V}$. Then, for any distribution $p(\{x_i\}_{i\in\mathcal{V}})$ on the transmit alphabets, we have*

$$R_{\mathcal{I};j}(p) \leq R^{G}_{\mathcal{I};j} + \alpha.$$

Proof: Let $(\tilde{X}, \tilde{Y})$ be the random variables that correspond to the fixed distribution $p(\{x_i\}_{i\in\mathcal{V}})$ and the corresponding received symbols. We have

$$\begin{aligned}R_{\mathcal{I};j}(p) &= \min_{\Omega\in\Lambda(\mathcal{I},D)} I(\tilde{X}_{\Omega}; \tilde{Y}_{\Omega^c} | \tilde{X}_{\Omega^c}) \\ &\overset{(a)}{\leq} I(\tilde{X}_{\Omega'}; \tilde{Y}_{\Omega'^c} | \tilde{X}_{\Omega'^c}) \\ &\leq \max_{p(\{x_i\}_{i\in\mathcal{V}})} I(X_{\Omega'}; Y_{\Omega'^c} | X_{\Omega'^c}) \\ &\triangleq \bar{C}_{\Omega'},\end{aligned}$$

where $\bar{C}_{\Omega'}$ is the capacity of the MIMO channel that the cut Ω' creates. In (a), we define Ω' to be the minimizer of $\min_{\Omega\in\Lambda(\mathcal{I},D)} I(X'_{\Omega}; Y'_{\Omega^c} | X'_{\Omega^c})$, where the distribution of $X'_{\mathcal{V}}$ is the i.i.d. complex normal distribution $\prod_{i\in\mathcal{V}} p^{\mathrm{G}}(x_i)$. From Lemma 6.6 in [5] we have that

$$\bar{C}_{\Omega'} \leq I(X'_{\Omega'}; \hat{Y}'_{\Omega'^c} | X'_{\Omega'^c}) + \alpha. \tag{A.26}$$

This concludes the proof. □

Note that the effect of α is comparable to the beamforming gain in a multi-antenna point-to-point channel.

A.10 Plausible Sequences in Wireless Networks

Definition A.3 *Consider a layered Gaussian multisource network and let $w_{\mathcal{S}}$ be a given realization of the message tuple $W_{\mathcal{S}}$. We define a* ***plausible transmit or quantized sequence*** *under $w_{\mathcal{S}}$ as follows.*

- *For a source node $i \in \mathcal{S}$ with no predecessors,* i.e., *such that $In(i) = \phi$, we say that a transmit sequence $\mathbf{x}_i$ is plausible with $w_{\mathcal{S}}$ if $\mathbf{x}_i = f_i(w_i)$.*

- *For any node $i \in \mathcal{V}$, we say that a quantized sequence $\hat{\mathbf{y}}_i$ is plausible with $w_{\mathcal{S}}$ if there is at least one tuple of sequences $\mathbf{x}_{In(i)}$ (transmit sequences at predecessors of i) that is plausible and such that $(\mathbf{x}_{In(i)}, \hat{\mathbf{y}}_i) \in \mathcal{T}_\delta$.*

- *For any node $i \in \mathcal{S}$ with predecessors, we say that a transmit sequence $\mathbf{x}_i$ is plausible with $w_{\mathcal{S}}$ if there is at least one quantized sequence $\hat{\mathbf{y}}_i$ that is plausible and for which $\mathbf{x}_i = f_i(\hat{\mathbf{y}}_i, w_i)$.*

- *For any node $i \in \mathcal{V} \setminus \mathcal{S}$, we say that a transmit sequence $\mathbf{x}_i$ is plausible with $w_{\mathcal{S}}$ if there is at least one quantized sequence $\hat{\mathbf{y}}_i$ that is plausible and for which $\mathbf{x}_i = f_i(\hat{\mathbf{y}}_i)$.*

These four definitions can be applied recursively to find all plausible sequences at a given node i. The sets of all plausible transmit and quantized sequences under $w_{\mathcal{S}}$ at node i are denoted by $\underline{\mathcal{X}}_i(w_{\mathcal{S}})$ and $\underline{\hat{\mathcal{Y}}}_i(w_{\mathcal{S}})$, respectively.

The following lemma is taken from [4] (Lemma 3.4 in that publication). Its proof can be found there.

Lemma A.2 *Consider a layered multisource Gaussian relay network $\mathcal{G}$. The number of sequences $\mathbf{x}_{\mathcal{V}}$ that are plausible under a given message realization $w_{\mathcal{S}}$ is upper bounded as*

$$|\underline{\mathcal{X}}_{\mathcal{V}}(w_{\mathcal{S}})| \leq 2^{T|\mathcal{V}|} \leq 2^{T\gamma},$$

where γ is defined in Definition 3.1.

Lemma A.3 *Let an acyclic, non-layered multisource Gaussian relay network $\mathcal{G}$ be given and consider its unfolded version $\mathcal{G}_{unf}^{(K)}$ over K stages as given in Section A.12. Then, the number of sequences $\mathbf{x}_{\mathcal{V}}$ that are plausible in $\mathcal{G}_{unf}^{(K)}$ under a given message realization $w_{\mathcal{S}}$ is upper bounded as*

$$|\underline{\mathcal{X}}_{\mathcal{V}}(w_{\mathcal{S}})| \leq 2^{TK|\mathcal{V}|} \leq 2^{TK\gamma}, \tag{A.27}$$

where γ is defined in Definition 3.1. However, for any node $i \in \mathcal{V}_{unf}$,

$$|\underline{\mathcal{X}}_{i}(w_{\mathcal{S}})| \leq 2^{Td|\mathcal{V}|} \leq 2^{T\gamma}. \tag{A.28}$$

Note that the second bound in Lemma A.3 does not depend on K.

Proof: The first bound (A.27) is nothing but Lemma A.2 applied to an unfolded network (which is layered by definition), because the unfolded network has roughly $K|\mathcal{V}|$ nodes. We outline the proof of (A.28) for the example network in Figures A.3 and A.4. Note that the nodes $T_1[k]$, $S_1[k]$, $k = 1, \dots, K$ in $\mathcal{G}_{\text{unf}}^{(K)}$ are all connected to S_1 through infinite-capacity links. Hence, we can assume that they all know the message transmitted by S_1 and hence, there is only one plausible transmit sequence for these nodes. The same holds for $T_2[k]$, $S_2[k]$, $k = 1, \dots, K$ and S_2. Consider now a given node i in $\mathcal{V}_{\text{unf}}$. The number of plausible transmit sequences for this node can be determined by analyzing all the paths going from this node backwards until reaching the set $\{S_1[k], S_2[k]\}_{k=1,\dots,K}$. When $\mathcal{G}$ is acyclic, then there are at most $d|\mathcal{V}|$ nodes in this collection of paths. We then apply the proof idea of Lemma A.2 (Lemma 3.4 in [4]) to this "tree" to conclude the proof. □

Note that if $\mathcal{G}$ contained a cycle, then the collection of paths would contain a number of nodes that grows with K, which would be problematic in the proof of Theorem 3.9 for non-layered networks.

A.11 Bounds used in the Equivocation Analysis

A.11.1 Three Lemmas

The following lemmas are used in the proofs of Propositions 3.1 and 3.2.

Lemma A.4 *Consider a block code for a layered network as defined in Definitions 2.2 and 2.3 that has the following properties. A node S encodes a pair of messages (W_1, W_2), that are uniformly distributed in $\{1, \dots, 2^{\lfloor TR_1 \rfloor}\} \times \{1, \dots, 2^{\lfloor TR_2 \rfloor}\}$ into a transmit sequence $\mathbf{X}_S$. A subset $\mathcal{B}$ of the relay nodes encodes respective messages $W_\mathcal{B}$ of respective rates $R_\mathcal{B}$ into respective transmit sequences $\mathbf{X}_\mathcal{B}$. Let T be the block size of the code, and assume that*

$$\frac{1}{2^{\lfloor TR_1 \rfloor}} \sum_{w_1=1}^{2^{\lfloor TR_1 \rfloor}} \mathbf{P}(A_{w_1}) \le \epsilon_0, \tag{A.29}$$

where A_{w_1} is the event that a certain node E makes an error when decoding $(W_2, W_\mathcal{B})$ given that $W_1 = w_1$ and assuming that W_1 is available at E. Then, the following inequality holds for both deterministic and Gaussian signal interaction:

$$H(W_2, W_\mathcal{B} | W_1, \mathbf{Y}_E) \le 1 + T(R_2 + \sum_{i \in \mathcal{B}} R_i)\epsilon_0. \tag{A.30}$$

In addition, if the signal interaction is deterministic, we have

$$H(\mathbf{X}_S, \mathbf{X}_\mathcal{B} | W_1, \mathbf{Y}_E) \le 1 + T(R_2 + \sum_{i \in \mathcal{B}} R_i)\epsilon_0, \tag{A.31}$$

and if the signal interaction is Gaussian, we have

$$H(\mathbf{X}_S, \mathbf{X}_\mathcal{B}|W_1, \mathbf{Y}_E) \leq 1 + T(R_2 + \sum_{i\in\mathcal{B}} R_i)\epsilon_0 + T\gamma. \tag{A.32}$$

Lemma A.5 *Consider a layered network as defined in Definition 2.2, with finite transmit alphabets. Signal interaction can be deterministic or stochastic (discrete). Let $\mathcal{I} \subseteq \mathcal{B}$ be a subset of the noise-inserting nodes and let $E \in \mathcal{V}$ be any node. Let $\prod_{i\in\mathcal{V}} p(x_i)$ be given and let $(\mathbf{X}_\mathcal{V}, \mathbf{Y}_\mathcal{V})$ be the corresponding random transmit and received sequences. The randomness of these sequences comes from the randomness of the code, as well as from the random messages and the channel noise. We assume that the code is a block code of typical sequences (with respect to p) as it was used in Section 3.6.2 for deterministic interaction. We have that for any $\epsilon > 0$, there is a large enough value of T such that*

$$\mathbf{P}\big(I(\mathbf{X}_S, \mathbf{X}_\mathcal{I}; \mathbf{Y}_E|J_{\mathcal{B}\setminus\mathcal{I}}) > TR_{\mathcal{I}\cup\{S\};E}(p)\big) \leq \epsilon.$$

Here, the probability is over all randomly generated codes.

If we compare this lemma with Theorem 2.1, we note that Lemma A.5 provides a "local" cut-set bound on the mutual information between transmitted and received sequences. The bound is local in the sense that the bounding expression $TR_{\mathcal{I}\cup\{S\};j}(p)$ is computed for the *same* product distribution p as the one that was used in the code construction procedure. Such local bounds are for instance also used in [48] and [10].

Lemma A.6 *Consider a layered network as defined in Definitions 2.2 and 2.3 with Gaussian signal interaction. Let $\mathcal{I} \subseteq \mathcal{B}$ be a subset of the noise-inserting nodes and let $E \in \mathcal{V}$ be any node. Let $(\mathbf{X}_\mathcal{V}, \mathbf{Y}_\mathcal{V})$ be the random transmit and received sequences produced by the block code chosen in Section 3.6.4. We have*

$$I(\mathbf{X}_S, \mathbf{X}_\mathcal{I}; \mathbf{Y}_E|J_{\mathcal{B}\setminus\mathcal{I}}) \leq T\big(R^G_{\mathcal{I}\cup\{S\};E} + \alpha\big).$$

Note that Lemma A.6 is not a local bound, because the upper bound given is not a function of the distribution that was used in the code construction procedure. For this reason, the statement of the lemma is not a high probability statement as in Lemma A.5, but it holds for all codes and every run of a code.

A.11.2 Proof of Lemma A.4

Suppose the assumptions of Lemma A.4 are true.

The upper bound on $H(W_2, W_\mathcal{B}|W_1, \mathbf{Y}_E)$ in (A.30) follows directly from Fano's inequality. To see this, let $(W_2, W_\mathcal{B})$ be a random variable X to be estimated, and let $(W_1, \mathbf{Y}_E)$ be the observable Y. Then, (A.29) simply states that $\mathbf{P}(X \neq \hat{X}) \leq \epsilon_0$, where $\hat{X}$ is the estimate of X from Y. Therefore, Fano's Lemma [9] tells us that $H(X|Y) \leq 1 + \epsilon_0 \log(|\mathcal{X}| - 1)$, which yields (A.30).

For completeness, we provide a proof which is the adaption of the Fano inequality proof to this special case. First, define the binary random variable

$$\xi \triangleq \mathbf{1}_{\{A_{W_1}\}},$$

where A_{w_1} is as defined in Lemma A.4. In words, ξ indicates whether $(W_2, W_{\mathcal{B}})$ was wrongly decoded at E, given that W_1 is available at E. We expand the following joint entropy in two different ways:

$$\begin{aligned} &H(\xi, W_2, W_{\mathcal{B}}|W_1, \mathbf{Y}_E) \\ &\quad = H(W_2, W_{\mathcal{B}}|W_1, \mathbf{Y}_E) + H(\xi|W_1, W_2, W_{\mathcal{B}}, \mathbf{Y}_E) \\ &\quad = H(\xi|W_1, \mathbf{Y}_E) + H(W_2, W_{\mathcal{B}}|W_1, \xi, \mathbf{Y}_E). \end{aligned} \tag{A.33}$$

Knowing $\mathbf{Y}_E$ and W_1, we know the decision $(\hat{W}_1, \hat{W}_{\mathcal{B}}) = \text{funct}(\mathbf{Y}_E, W_1)$ made by E, and hence

$$\begin{aligned} H(\xi|W_1, W_2, W_{\mathcal{B}}, \mathbf{Y}_E) &= H(\xi|W_1, W_2, W_{\mathcal{B}}, \hat{W}_2, \hat{W}_{\mathcal{B}}, \mathbf{Y}_E) \\ &= 0. \end{aligned}$$

Also

$$\begin{aligned} H(\xi|W_1, \mathbf{Y}_E) &\leq H(\xi) \\ &= h_b(\mathbf{P}(\xi = 1)) \\ &\leq 1, \end{aligned}$$

where $h_b(\cdot)$ is the entropy function of a binary random variable. Using this, (A.33) implies

$$\begin{aligned}
H(W_2,&W_{\mathcal{B}}|W_1,\mathbf{Y}_E) \le 1 + H(W_2,W_{\mathcal{B}}|\xi,\mathbf{Y}_E,W_1)\\
&= 1 + \sum_{w_1=1}^{2^{\lfloor TR_1\rfloor}} \mathbf{P}(W_1=w_1)H(W_2,W_{\mathcal{B}}|\xi,\mathbf{Y}_E,W_1=w_1)\\
&= 1 + \sum_{w_1=1}^{2^{\lfloor TR_1\rfloor}} \frac{1}{2^{\lfloor TR_1\rfloor}} H(W_2,W_{\mathcal{B}}|\xi,\mathbf{Y}_E,W_1=w_1)\\
&= 1 + \sum_{w_1=1}^{2^{\lfloor TR_1\rfloor}} \frac{1}{2^{\lfloor TR_1\rfloor}} \big[\mathbf{P}(\xi=0|W_1=w_1)\\
&\quad \cdot \underbrace{H(W_2,W_{\mathcal{B}}|\xi=0,\mathbf{Y}_E,W_1=w_1)}_{=0}\\
&\quad + \mathbf{P}(\xi=1|W_1=w_1)\\
&\quad \cdot \underbrace{H(W_2,W_{\mathcal{B}}|\xi=1,\mathbf{Y}_E,W_1=w_1)}_{\le \log(2^{(\lfloor TR_2\rfloor+\sum_{i\in\mathcal{B}}\lfloor TR_i\rfloor)}-1)\le(\lfloor TR_2\rfloor+\sum_{i\in\mathcal{B}}\lfloor TR_i\rfloor)} \big]\\
&\le 1 + (\lfloor TR_2\rfloor + \sum_{i\in\mathcal{B}}\lfloor TR_i\rfloor)\frac{1}{2^{\lfloor TR_1\rfloor}}\sum_{w_1=1}^{2^{\lfloor TR_1\rfloor}}\mathbf{P}(\xi=1|W_1=w_1)\\
&= 1 + (\lfloor TR_2\rfloor + \sum_{i\in\mathcal{B}}\lfloor TR_i\rfloor)\frac{1}{2^{\lfloor TR_1\rfloor}}\sum_{w=1}^{2^{\lfloor TR_1\rfloor}}\mathbf{P}(A_{w_1})\\
&\overset{(a)}{\le} 1 + (\lfloor TR_2\rfloor + \sum_{i\in\mathcal{B}}\lfloor TR_i\rfloor)\epsilon_0\\
&\le 1 + T(R_2 + \sum_{i\in\mathcal{B}} R_i)\epsilon_0,
\end{aligned}$$

where (a) is true because of (A.29).

To show (A.31) and (A.32), we start with the following inequality:

$$\begin{aligned}
H(\mathbf{X}_S,\mathbf{X}_{\mathcal{B}}|W_1,\mathbf{Y}_E) &\le H(\mathbf{X}_S,\mathbf{X}_{\mathcal{B}},W_2,W_{\mathcal{B}}|W_1,\mathbf{Y}_E)\\
&= H(W_2,W_{\mathcal{B}}|W_1,\mathbf{Y}_E)\\
&\quad + H(\mathbf{X}_S,\mathbf{X}_{\mathcal{B}}|W_1,W_2,W_{\mathcal{B}},\mathbf{Y}_E).
\end{aligned}$$

For deterministic signal interaction, the second term above is zero because the transmit sequences $(\mathbf{X}_S,\mathbf{X}_{\mathcal{B}})$ are a function of $(W_1,W_2,W_{\mathcal{B}})$. Combining this with (A.30) yields (A.31). For Gaussian signal interaction, we have

$$\begin{aligned}
H(\mathbf{X}_S,\mathbf{X}_{\mathcal{B}}|W_1,&W_2,W_{\mathcal{B}},\mathbf{Y}_E)\\
&= \underbrace{H(\mathbf{X}_S|W_1,W_2,W_{\mathcal{B}},\mathbf{Y}_E)}_{=0}\\
&\quad + H(\mathbf{X}_{\mathcal{B}}|\mathbf{X}_S,W_1,W_2,W_{\mathcal{B}},\mathbf{Y}_E),
\end{aligned}$$

where the indicated term is zero because $\mathbf{X}_S$ is a function of (W_1, W_2). Since for any $i \in \mathcal{B}$, $\mathbf{X}_i$ is a deterministic function of $\hat{\mathbf{Y}}_i$ and W_i, the data processing inequality tells us that

$$\begin{aligned}
&H(\mathbf{X}_{\mathcal{B}}|\mathbf{X}_S, W_1, W_2, W_{\mathcal{B}}, \mathbf{Y}_E)\\
&\qquad \le H(\hat{\mathbf{Y}}_{\mathcal{B}}|\mathbf{X}_S, W_1, W_2, W_{\mathcal{B}}, \mathbf{Y}_E)\\
&\qquad \le H(\hat{\mathbf{Y}}_{\mathcal{B}}|W_1, W_2, W_{\mathcal{B}})\\
&\qquad \overset{(a)}{\le} \log(|\underline{\hat{\mathcal{Y}}_{\mathcal{B}}}(W_1, W_2, W_{\mathcal{B}})|)\\
&\qquad \le \log(|\underline{\hat{\mathcal{Y}}_{\mathcal{V}}}(W_1, W_2, W_{\mathcal{B}})|)\\
&\qquad \le \log(|\underline{\mathcal{X}_{\mathcal{V}}}(W_1, W_2, W_{\mathcal{B}})|)\\
&\qquad \overset{(b)}{\le} \log(2^{T\gamma})\\
&\qquad = \gamma T,
\end{aligned}$$

where the set $\underline{\hat{\mathcal{Y}}_{\mathcal{B}}}(W_1, W_2, W_{\mathcal{B}})$ used in (a) is the set of all plausible sequences under $(W_1, W_2, W_{\mathcal{B}})$ as defined in Definition A.3 (Section A.10). The bound (b) can be found in Lemma A.2 of Section A.10. Combining all of the above, we obtain

$$H(\mathbf{X}_S, \mathbf{X}_{\mathcal{B}}|W_1, \mathbf{Y}_E) \le H(W_2, W_{\mathcal{B}}|W_1, \mathbf{Y}_E) + T\gamma. \tag{A.34}$$

Combining (A.34) with (A.30) yields (A.32), which concludes the proof of the lemma.

A.11.3 Proof of Lemma A.5

Assume that the network is layered. Note that in the achievability proof of Section 3.6, whenever we write $(\mathbf{X}_{\mathcal{V}}, \mathbf{Y}_{\mathcal{V}})$, what we mean is $(\mathbf{X}_{\mathcal{V}}(M^{(1)}), \mathbf{Y}_{\mathcal{V}}(M^{(1)}))$, where $M^{(1)} = (W^{(1)}, J_1^{(1)}, J_2^{(1)}, J_{\mathcal{B}}^{(1)})$ is the set of messages transmitted during block 1 (information and junk messages combined). If E cannot directly receive the signal transmitted by S, then $\mathbf{X}_S(M^{(1)})$ and $\mathbf{Y}_E(M^{(1)})$ are sequences that occur during different time blocks. If, for instance, there is one layer of relays between S and E, then a relay node A would receive $\mathbf{Y}_A(M^{(1)})$ during the first time block, when S transmits $\mathbf{X}_S(M^{(1)})$. However, A would transmit $\mathbf{X}_A(M^{(1)})$ during the second block of length T, while already receiving $\mathbf{Y}_A(M^{(2)})$. E would therefore start receiving the block $\mathbf{Y}_E(M^{(1)})$ at the beginning of the second time block, after S has finished transmitting $\mathbf{X}_S(M^{(1)})$. Let N be a large positive integer. We assume that transmission takes place over many time blocks, but we restrict our attention to a window of N time blocks, each of length T. We define a new set of vectors $(\underline{\mathbf{X}}_{\mathcal{V}}, \underline{\mathbf{Y}}_{\mathcal{V}})$ of length $(N+L)T$, where L is the number of layers of relay-nodes that connect S to E in the network. Recall that $\mathcal{V}$ is the set of all nodes in the network, and $\mathcal{A}$ is the set of relay nodes. Let $\mathcal{A}_l \subseteq \mathcal{A}$ be the set of relay nodes that lie in layer l.

For $l = 0, \dots, L$, define

$$\underline{\mathbf{X}}_{\mathcal{A}_l}(\underline{M}) \triangleq \begin{bmatrix} \mathbf{X}_{\mathcal{A}_l}(M^{(1-l)}) \\ \vdots \\ \mathbf{X}_{\mathcal{A}_l}(M^{(0)}) \\ \mathbf{X}_{\mathcal{A}_l}(M^{(1)}) \\ \vdots \\ \mathbf{X}_{\mathcal{A}_l}(M^{(N)}) \\ \mathbf{X}_{\mathcal{A}_l}(M^{(N+1)}) \\ \vdots \\ \mathbf{X}_{\mathcal{A}_l}(M^{(N+L-l)}) \end{bmatrix}$$

and

$$\underline{\mathbf{Y}}_{\mathcal{A}_{l+1}}(\underline{M}) \triangleq \begin{bmatrix} \mathbf{Y}_{\mathcal{A}_{l+1}}(M^{(1-l)}) \\ \vdots \\ \mathbf{Y}_{\mathcal{A}_{l+1}}(M^{(0)}) \\ \mathbf{Y}_{\mathcal{A}_{l+1}}(M^{(1)}) \\ \vdots \\ \mathbf{Y}_{\mathcal{A}_{l+1}}(M^{(N)}) \\ \mathbf{Y}_{\mathcal{A}_{l+1}}(M^{(N+1)}) \\ \vdots \\ \mathbf{Y}_{\mathcal{A}_{l+1}}(M^{(N+L-l)}) \end{bmatrix},$$

where we set $\mathcal{A}_0 \triangleq \{S\}$ and $\mathcal{A}_{L+1} \triangleq \{E\}$, and $\underline{M} = \{M^{(1-L)}, \dots, M^{(N+2L)}\}$ is the set of all messages involved during $N+2L$ blocks. Assume that transmission of the first message $M^{(1)}$ starts at $t = 1$. We then have that the t^{th} component of each super-block $\underline{\mathbf{X}}_i$ or $\underline{\mathbf{Y}}_i$, $i \in \mathcal{V}$ is actually received or transmitted at time t. Fix any $\Omega \in \Lambda(\mathcal{I} \cup \{S\}; E)$. We have the following chain of inequalities, which closely follows the proof of Theorem 14.10.1 in [9]. Most of the steps in this chain of inequalities are true for every realization of the code and for every run of that code realization. However, the last convergence step happens only with high probability over the codes. (Note that since the code is randomly chosen, the mutual information terms below are random variables.)

$$
\begin{aligned}
I(\underline{\mathbf{X}}_S, \underline{\mathbf{X}}_{\mathcal{I}}; \underline{\mathbf{Y}}_E | \underline{J}_{\mathcal{B}\setminus\mathcal{I}}) &\overset{(a)}{\leq} I(\underline{\mathbf{X}}_\Omega; \underline{\mathbf{Y}}_{\Omega^c} | \underline{J}_{\mathcal{B}\setminus\mathcal{I}}) \\
&= \sum_{t=1}^{(N+L)T} I(\underline{\mathbf{X}}_\Omega; \underline{Y}_{\Omega^c}[t] | \underline{Y}_{\Omega^c}[1], \ldots, \underline{Y}_{\Omega^c}[t-1], \underline{J}_{\mathcal{B}\setminus\mathcal{I}}) \\
&= \sum_{t=1}^{(N+L)T} \Big[H(\underline{Y}_{\Omega^c}[t] | \underline{Y}_{\Omega^c}[1], \ldots, \underline{Y}_{\Omega^c}[t-1], \underline{J}_{\mathcal{B}\setminus\mathcal{I}}) \\
&\quad - H(\underline{Y}_{\Omega^c}[t] | \underline{Y}_{\Omega^c}[1], \ldots, \underline{Y}_{\Omega^c}[t-1], \underline{\mathbf{X}}_\Omega, \underline{J}_{\mathcal{B}\setminus\mathcal{I}}) \Big] \\
&\overset{(b)}{=} \sum_{t=1}^{(N+L)T} \Big[H(\underline{Y}_{\Omega^c}[t] | \underline{Y}_{\Omega^c}[1], \ldots, \underline{Y}_{\Omega^c}[t-1], \underline{X}_{\Omega^c}[t], \underline{J}_{\mathcal{B}\setminus\mathcal{I}}) \\
&\quad - H(\underline{Y}_{\Omega^c}[t] | \underline{Y}_{\Omega^c}[1], \ldots, \underline{Y}_{\Omega^c}[t-1], \underline{\mathbf{X}}_\Omega, \underline{X}_{\Omega^c}[t], \underline{J}_{\mathcal{B}\setminus\mathcal{I}}) \Big] \\
&\overset{(c)}{\leq} \sum_{t=1}^{(N+L)T} \Big[H(\underline{Y}_{\Omega^c}[t] | \underline{X}_{\Omega^c}[t]) \\
&\quad - H(\underline{Y}_{\Omega^c}[t] | \underline{X}_\Omega[t], \underline{X}_{\Omega^c}[t]) \Big] \\
&= \sum_{t=1}^{(N+L)T} I(\underline{X}_\Omega[t]; \underline{Y}_{\Omega^c}[t] | \underline{X}_{\Omega^c}[t]) \qquad \text{(A.35)} \\
&\overset{(d)}{\simeq} (N+L)T \; I(X_\Omega; Y_{\Omega^c} | X_{\Omega^c}), \qquad \text{(A.36)}
\end{aligned}
$$

as T and N grow large. The steps are justified below:

(a) The mutual information is made larger by adding more super-blocks on each side.

(b) $\underline{X}_{\Omega^c}[t]$ is what is transmitted by all nodes in Ω^c at time t and thus is a function of $J_{\mathcal{B}\cap\Omega^c} = J_{\mathcal{B}\setminus\mathcal{I}}$ and of everything that was received by all nodes in Ω^c up to time $t-1$. Hence, we have equality in both terms.

(c) The first term is increased by dropping a part of the conditioning. For the second term, we use the fact that $\underline{Y}_{\Omega^c}[t]$ depends only on the current transmit values $\underline{X}_\Omega[t], \underline{X}_{\Omega^c}[t]$.

(d) Assume that $t \in \{(n-1)T+1, \ldots, nT\}$ for some $n \in \{1, \ldots, N+L\}$, *i.e.*, assume that t lies in the n-th block of the super-block. Consider the first layer: $(\underline{X}_S[t], \underline{Y}_{\mathcal{A}_1}[t])$ is a fixed component of $(\mathbf{X}_S(M^{(n)}), \mathbf{Y}_{\mathcal{A}_1}(M^{(n)}))$, which is, with high probability, a set of jointly typical sequences with respect to $p_{X_S} p_{Y_{\mathcal{A}_1}|X_S}$, where $\mathbf{X}_S$ was picked uniformly at random from a codebook of size $2^{T(R+B_1+B_2)}$. Hence, as T grows large, the joint distribution of $(\underline{X}_S[t], \underline{Y}_{\mathcal{A}_1}[t])$ converges to $p_{X_S} p_{Y_{\mathcal{A}_1}|X_S}$. This convergence happens with high probability over all codes. We advance one layer in the network: The tuple $(\underline{X}_{\mathcal{A}_1}[t], \underline{Y}_{\mathcal{A}_2}[t])$ is a fixed component of the tuple of sequences $(\mathbf{X}_{\mathcal{A}_1}(M^{(n-1)}), \mathbf{Y}_{\mathcal{A}_2}(M^{(n-1)}))$, which is w.h.p. a member of a collection of jointly typical sequences with respect to $p_{X_{\mathcal{A}_1}} p_{Y_{\mathcal{A}_2}|X_{\mathcal{A}_1}}$. In addition, the tuple of sequences is independent of the tuple in the first layer, because the message $M^{(n-1)}$ is independent of $M^{(n)}$. Nevertheless, it follows that the distribution of $(\underline{X}_{\mathcal{A}_1}[t], \underline{Y}_{\mathcal{A}_2}[t])$ converges w.h.p. to $p_{X_{\mathcal{A}_1}} p_{Y_{\mathcal{A}_2}|X_{\mathcal{A}_1}}$. We do so for each layer, and we find that the distribution of $(\underline{X}_{\mathcal{V}}[t], \underline{Y}_{\mathcal{V}}[t])$ converges to

$$\prod_{l=1}^{L+1} p_{X_{\mathcal{A}_{l-1}}} p_{Y_{\mathcal{A}_l}|X_{\mathcal{A}_{l-1}}} \tag{A.37}$$

with high probability. The product comes from the fact that all involved messages $M^{(n)}$ through $M^{(n+L)}$ are independent. But because $p_{\{X_i\}_{i\in\mathcal{V}}}$ was chosen to be a product distribution, (A.37) is nothing but $p_{X_\mathcal{V}} p_{Y_\mathcal{V}|X_\mathcal{V}}$. Then, because the mutual information is a continuous function of the underlying distribution, it holds that $I(\underline{X}_\Omega[t]; \underline{Y}_{\Omega^c}[t]|\underline{X}_{\Omega^c}[t]) \to I(X_\Omega; Y_{\Omega^c}|X_{\Omega^c})$ for all t with high probability.

Finally, we can write

$$\begin{aligned} I(\underline{\mathbf{X}}_S, \underline{\mathbf{X}}_\mathcal{I}; \underline{\mathbf{Y}}_E | \underline{J}_{\mathcal{B}\setminus\mathcal{I}}) &\overset{(a)}{\geq} I(\mathbf{X}_{\{S\}\cup\mathcal{I}}(M^{(1)}), \dots, \mathbf{X}_{\{S\}\cup\mathcal{I}}(M^{(N)}); \\ &\qquad \mathbf{Y}_E(M^{(1)}), \dots, \mathbf{Y}_E(M^{(N)}) | J^{(1)}_{\mathcal{B}\setminus\mathcal{I}}, \dots, J^{(N)}_{\mathcal{B}\setminus\mathcal{I}}) \\ &\overset{(b)}{=} \sum_{n=1}^{N} I(\mathbf{X}_S(M^{(n)}), \mathbf{X}_\mathcal{I}(M^{(n)}); \mathbf{Y}_E(M^{(n)}) | J^{(n)}_{\mathcal{B}\setminus\mathcal{I}}) \\ &= N I(\mathbf{X}_S(M^{(1)}), \mathbf{X}_\mathcal{I}(M^{(1)}); \mathbf{Y}_E(M^{(1)}) | J^{(1)}_{\mathcal{B}\setminus\mathcal{I}}), \end{aligned} \tag{A.38}$$

where we obtain (a) by dropping all terms in $\underline{\mathbf{X}}_S$ and $\underline{\mathbf{Y}}_E$ that depend on messages different from $M^{(1)}, \dots, M^{(N)}$, and (b) is true by the independence of the $M^{(n)}$. Combining (A.38) with (A.36), we obtain

$$\begin{aligned} I(\mathbf{X}_S(M^{(1)}), &\mathbf{X}_\mathcal{I}(M^{(1)}); \mathbf{Y}_E(M^{(1)}) | J^{(1)}_{\mathcal{B}\setminus\mathcal{I}}) \\ &\leq \frac{N+L}{N} T I(X_\Omega; Y_{\Omega^c} | X_{\Omega^c}) \\ &\to T I(X_\Omega; Y_{\Omega^c} | X_{\Omega^c}) \end{aligned} \tag{A.39}$$

w.h.p. as N grows large.

Since (A.39) is true for any $\Omega \in \Lambda(\mathcal{I} \cup \{S\}; E)$, we have that w.h.p.,

$$I(\mathbf{X}_S(M^{(1)}); \mathbf{Y}_E(M^{(1)})) \leq T \min_{\Omega \in \Lambda(\mathcal{I}\cup\{S\}; E)} I(X_\Omega; Y_{\Omega^c} | X_{\Omega^c}),$$

which proves the claim of the lemma.

A.11.4 Proof of Lemma A.6

For Gaussian signal interaction, we follow exactly the proof of Lemma A.5 given in the previous section, until the expression A.35, which lets us conclude that

$$\begin{aligned}
I(\underline{\mathbf{X}}_S, \underline{\mathbf{X}}_{\mathcal{I}}; \underline{\mathbf{Y}}_E | \underline{J}_{\mathcal{B}\setminus\mathcal{I}}) &\leq \sum_{t=1}^{(N+L)T} I(\underline{X}_\Omega[t]; \underline{Y}_{\Omega^c}[t] | \underline{X}_{\Omega^c}[t]) \\
&\overset{(a)}{=} (N+L)TI(\underline{X}_\Omega[Q]; \underline{Y}_{\Omega^c}[Q] | \underline{X}_{\Omega^c}[Q], Q) \\
&\leq (N+L)T\big(H(\underline{Y}_{\Omega^c}[Q] | \underline{X}_{\Omega^c}[Q]) \\
&\quad - H(\underline{Y}_{\Omega^c}[Q] | \underline{X}_{\mathcal{V}}[Q], Q)\big) \\
&\overset{(b)}{=} (N+L)TI(\underline{X}_\Omega[Q]; \underline{Y}_{\Omega^c}[Q] | \underline{X}_{\Omega^c}[Q]),
\end{aligned}$$

where in (a), we have introduced a time sharing random variable Q, uniformly distributed in $\{1, \dots, (N+L)T\}$, and (b) holds because from the memoryless nature of the channels, Q can be dropped from the conditioning of the second entropy. Using the same steps as in (A.38) and (A.39), we conclude that as N grows large, we have

$$\begin{aligned}
I(\mathbf{X}_S(M^{(1)}),&\mathbf{X}_{\mathcal{I}}(M^{(1)}); \mathbf{Y}_E(M^{(1)}) | J^{(1)}_{\mathcal{B}\setminus\mathcal{I}}) \\
&\leq TI(\underline{X}_\Omega[Q]; \underline{Y}_{\Omega^c}[Q] | \underline{X}_{\Omega^c}[Q]). \qquad \text{(A.40)}
\end{aligned}$$

This is true for every cut $\Omega \in \Lambda(\mathcal{I} \cup \{S\}; E)$. Hence, we have

$$\begin{aligned}
I(\mathbf{X}_S(M^{(1)}),&\mathbf{X}_{\mathcal{I}}(M^{(1)}); \mathbf{Y}_E(M^{(1)}) | J^{(1)}_{\mathcal{B}\setminus\mathcal{I}}) \\
&\leq TR_{\mathcal{I}\cup\{S\};E}(\tilde{p}),
\end{aligned}$$

where $\tilde{p}$ is the distribution of $\underline{X}_{\mathcal{V}}[Q]$. From Lemma A.1 in Section A.9, we know that

$$R_{\mathcal{I}\cup\{S\};E}(\tilde{p}) \leq R^{\mathrm{G}}_{\mathcal{I}\cup\{S\};E} + \alpha,$$

and hence, the claim of the lemma follows.

A.12 From Layered to Non-Layered Networks

In this section, we provide a lemma that plays a central role in the generalization of our results from layered to general networks. The technique used for such generalizations was introduced in [3, 4]. Let $\mathcal{G} = (\mathcal{V}, \mathcal{L})$ be a general wireless network with deterministic or Gaussian signal interaction. Let $\mathcal{S}$ and $\mathcal{D}$ be two disjoint subsets of $\mathcal{V}$. In our applications, $\mathcal{S}$ is a set of nodes that generate or observe random sources, while $\mathcal{D}$ denotes all the destination nodes and eavesdroppers in the network. Any such network can be unfolded over time to create a layered network. The idea is to unfold the network to K stages

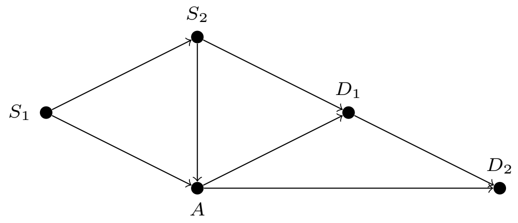

Figure A.3: An example of a non-layered network with $\mathcal{S} = \{S_1, S_2\}$ and $\mathcal{D} = \{D_1, D_2\}$.

such that the k^{th} stage represents what happens in the original network during time slots $(k-1)T+1$ to kT, *i.e.*, during the k^{th} of K blocks. We denote the unfolded network by $\mathcal{G}_{\text{unf}}^{(K)}$. Figure A.3 shows a non-layered network. Here $\mathcal{S} = \{S_1, S_2\}$ and $\mathcal{D} = \{D_1, D_2\}$ are both of size two. There is also a relay node A with no special role. The network is clearly not layered, because there are for instance different paths from S_1 to D_2 of lengths 2, 3 and 4. The corresponding unfolded network is shown in Figure A.4. Each node $i \in \mathcal{V}$ appears at stage $1 \leq k \leq K$ as $i[k]$. Note that the square bracket notation is used here to denote the stages of the unfolding, whereas in the rest of the thesis, a square bracket denotes the time slot. There are additional nodes in Figure A.4: The source node S_1 and its unfolded representatives $S_1[1], \ldots, S_1[K]$ have been augmented by virtual transmitters $T_1[1], \ldots, T_1[K]$ that model the fact that S_1 has access to its source message during all K stages. Such a set of transmitters is added for each node in $\mathcal{S}$. Further, the destination node (or eavesdropper) D_1 and its representatives have been augmented by receivers $R_1[1], \ldots, R_1[K]$ that model the fact that D_1 has access to the information received during all K stages. Again, this is done for every node in $\mathcal{D}$. The communication links connecting to and from the virtual transmitters and receivers are modeled as channels of arbitrarily large throughput. In addition, none of these links (labeled with ∞ in Figure A.4) interferes with a regular received signal and hence, these links impose no constraint on the network. The relay node A is *not* augmented by virtual nodes, modeling the fact that we force A to limit its memory to T received symbols.

For the remainder of this section, we assume that the signal interaction is deterministic. Similar results hold for Gaussian signal interaction.

Definition A.4 *For a given product distribution $\prod_{i\in\mathcal{V}} p(x_i)$, a given subset $\psi \subseteq \mathcal{S}$ and $j \in \mathcal{D}$, we define $\mathcal{R}_{\psi;j}^{unf}(p)$ to be the set of all tuples B_ψ such that for any subset $\mathcal{I} \subseteq \psi$, we have*

$$\sum_{i\in\mathcal{I}} B_i \leq \min_{\Omega_{unf}\in\Lambda_{unf}(\mathcal{I};j)} I(X_{\Omega_{unf}}; Y_{\Omega_{unf}^c} | X_{\Omega_{unf}^c}),$$

where $\Lambda_{unf}(\mathcal{I};j)$ is defined as in Definition 2.19 but for the unfolded network. To compute the mutual information, we use the product distribution

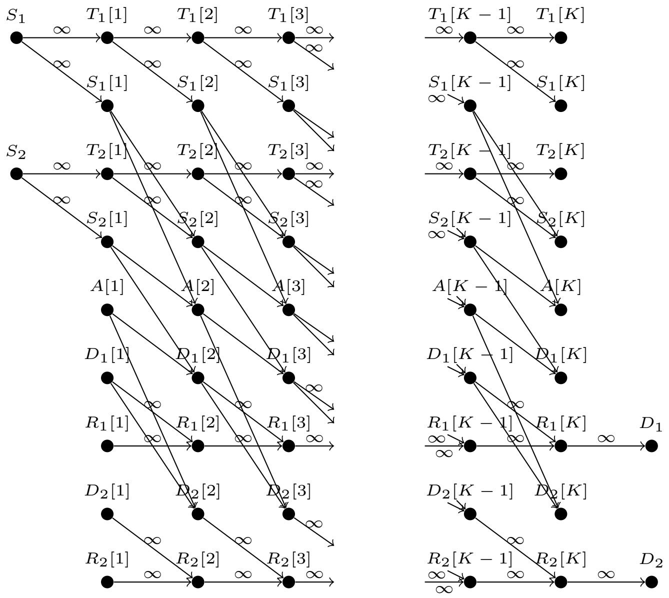

Figure A.4: The unfolded network that corresponds to the example in Fig. A.3.

$\prod_{i\in\mathcal{V}}\prod_{k\in\{1,\ldots,K\}} p(x_i[k])$ *with* $p(x_i[k]) = p(x_i)$.

We are now ready to state the unfolding lemma:

Lemma A.7 *For given* $\prod_{i\in\mathcal{V}} p(x_i)$*,* $\psi \subseteq \mathcal{S}$ *and* $j \in \mathcal{D}$*, the sets* $\frac{1}{K}\mathcal{R}^{unf}_{\psi;j}(p)$ *and* $\mathcal{R}_{\psi;j}(p)$ *are approximately equal for large* K*. More formally, if a tuple* B_ψ *is in* $\mathcal{R}^{unf}_{\psi;j}(p)$*, then* $\frac{1}{K}B_\psi$ *is in* $\mathcal{R}_{\psi;j}(p)$*. On the other hand, if* $\frac{1}{K}B_\psi$ *is in* $\mathcal{R}_{\psi;j}(p)$*, and an arbitrarily small* $\epsilon > 0$ *is given, then there exists an unfolding length* K *such that*

$$B_\psi - \epsilon\mathbf{1} \in \mathcal{R}^{unf}_{\psi;j}(p).$$

Proof: Let $\prod_{i\in\mathcal{V}} p(x_i)$, ψ and j be fixed. Assume that $B_\psi \in \mathcal{R}^{\mathrm{unf}}_{\psi;j}(p)$. It follows that for every $\mathcal{I} \subseteq \psi$,

$$\sum_{i\in\mathcal{I}} B_i \le \min_{\Omega_{\mathrm{unf}}\in\Lambda_{\mathrm{unf}}(\mathcal{I};j)} I(X_{\Omega_{\mathrm{unf}}}; Y_{\Omega^c_{\mathrm{unf}}}|X_{\Omega^c_{\mathrm{unf}}}).$$

Let Ω' be the optimizer of $\min_{\Omega\in\Lambda(\mathcal{I};j)} I(X_\Omega; Y_{\Omega^c}|X_{\Omega^c})$ in the original graph $\mathcal{G}$. There is a corresponding cut Ω'_{unf} in $\mathcal{G}^{(K)}_{\mathrm{unf}}$ such that if a node $i \in \Omega'$, then $i[k] \in \Omega'_{\mathrm{unf}}$ for all $k = 1,\ldots,K$. In addition, the value of the mutual information for the cut Ω'_{unf} in $\mathcal{G}^{(K)}_{\mathrm{unf}}$ is K times the value of the cut Ω' in $\mathcal{G}$.

Hence,

$$\begin{aligned}\min_{\Omega_{\text{unf}}\in\Lambda_{\text{unf}}(\mathcal{I};j)} I(X_{\Omega_{\text{unf}}};Y_{\Omega^c_{\text{unf}}}|X_{\Omega^c_{\text{unf}}}) &\leq I(X_{\Omega'_{\text{unf}}};Y_{{\Omega'_{\text{unf}}}^c}|X_{{\Omega'_{\text{unf}}}^c})\\ &= KI(X_{\Omega'};Y_{\Omega'^c}|X_{\Omega'^c})\\ &= K\min_{\Omega\in\Lambda(\mathcal{I};j)} I(X_\Omega;Y_{\Omega^c}|X_{\Omega^c}).\end{aligned}$$

It follows that

$$\frac{1}{K}\sum_{i\in\mathcal{I}} B_i \leq \min_{\Omega\in\Lambda(\mathcal{I};j)} I(X_\Omega;Y_{\Omega^c}|X_{\Omega^c}).$$

Since this is true for all subsets $\mathcal{I}\subset\psi$, it follows that $\frac{1}{K}B_\psi$ is in $\mathcal{R}_{\psi;j}(p)$.

Now, assume that $\frac{1}{K}B_\psi$ is in $\mathcal{R}_{\psi;j}(p)$. It follows that for every $\mathcal{I}\subseteq\psi$,

$$\frac{1}{K}\sum_{i\in\mathcal{I}} B_i \leq \min_{\Omega\in\Lambda(\mathcal{I};j)} I(X_\Omega;Y_{\Omega^c}|X_{\Omega^c}). \tag{A.41}$$

Lemma 4.2 in [4] applies to this unfolded network, with the slight modification that now there are more than two nodes that have been augmented by virtual receivers or transmitters. That lemma states that

$$\begin{aligned}&\min_{\Omega\in\Lambda(\mathcal{I};j)} I(X_\Omega;Y_{\Omega^c}|X_{\Omega^c})\\ &\qquad\leq \frac{1}{K-c+1}\min_{\Omega_{\text{unf}}\in\Lambda_{\text{unf}}(\mathcal{I};j)} I(X_{\Omega_{\text{unf}}};Y_{\Omega^c_{\text{unf}}}|X_{\Omega^c_{\text{unf}}}),\end{aligned}$$

where $c \triangleq 2^{|\mathcal{V}|-2}$ does not depend on K. Combining this with (A.41), we obtain

$$\begin{aligned}\sum_{i\in\mathcal{I}} B_i - |\mathcal{I}|\epsilon &\leq K\min_{\Omega\in\Lambda(\mathcal{I};j)} I(X_\Omega;Y_{\Omega^c}|X_{\Omega^c}) - |\mathcal{I}|\epsilon\\ &\leq \frac{K}{K-c+1}\min_{\Omega_{\text{unf}}\in\Lambda_{\text{unf}}(\mathcal{I};j)} I(X_{\Omega_{\text{unf}}};Y_{\Omega^c_{\text{unf}}}|X_{\Omega^c_{\text{unf}}})\\ &\quad - |\mathcal{I}|\epsilon\\ &\leq \min_{\Omega_{\text{unf}}\in\Lambda_{\text{unf}}(\mathcal{I};j)} I(X_{\Omega_{\text{unf}}};Y_{\Omega^c_{\text{unf}}}|X_{\Omega^c_{\text{unf}}})\end{aligned}$$

if K is chosen sufficiently large. Since this is true for any $\mathcal{I}\subseteq\psi$, it follows that for a large enough K, $B_\psi - \epsilon\mathbf{1}$ is in $\mathcal{R}^{\text{unf}}_{\psi;j}(p)$. This concludes the proof. □

Appendix for Chapter 4 B

B.1 Detailed Proof of Theorem 4.2

We have already given a proof outline for this theorem in Section 4.4.2. For the interested reader, we give a full proof in this appendix. Fix a distribution p_{X_S}. We wish to show that perfectly secret communication at a rate arbitrarily close to $I(X_S; Y_{\mathcal{A}}) - \max_{E \in \mathcal{E}} I(X_S; Y_{\mathcal{A}_E})$ is possible. We first prove weak perfect secrecy. The extension to strong secrecy is done in the end of this section.

For notational simplicity, we only prove the result for $\mathcal{E} = \{E_1, E_2\}$, *i.e.*, we assume that there are only two eavesdroppers. Assume that the eavesdroppers are labeled such that $I(X_S; Y_{\mathcal{A}_{E_1}}) \geq I(X_S; Y_{\mathcal{A}_{E_2}})$. Define an information rate and two junk rates as follows:

$$\begin{aligned} R &= I(X_S; Y_{\mathcal{A}}) - I(X_S; Y_{\mathcal{A}_{E_1}}) && \text{(B.1)} \\ B_1 &= I(X_S; Y_{\mathcal{A}_{E_1}}) - I(X_S; Y_{\mathcal{A}_{E_2}}) \\ B_2 &= I(X_S; Y_{\mathcal{A}_{E_2}}) - \epsilon_1. \end{aligned}$$

As a result, we have

$$R + B_1 + B_2 = I(X_S; Y_{\mathcal{A}}) - \epsilon_1 \tag{B.2}$$

$$B_1 + B_2 = I(X_S; Y_{\mathcal{A}_{E_1}}) - \epsilon_1 \tag{B.3}$$

$$B_2 = I(X_S; Y_{\mathcal{A}_{E_2}}) - \epsilon_1. \tag{B.4}$$

Let W be the information message, uniformly distributed in $\{1, \ldots, 2^{\lfloor TR \rfloor}\}$ and let J_i be the i^{th} junk message, uniformly distributed in $\{1, \ldots, 2^{\lfloor TB_i \rfloor}\}$ for $i = 1, 2$. Generate a random block code (of block length T) for the broadcast channel in the following way. For every triple (w, j_1, j_2), pick a codeword $\mathbf{x}_S(w, j_1, j_2)$ uniformly at random from $\mathcal{T}_\delta(X_S)$. Once this random code construction process is finished, we fix the code for all times. The decoder at D

uses a typical set decoder [9]. It is well known (see e.g. the proof of Theorem 8.7.1 in [9]) that as long as (B.2) holds, we can achieve

$$\mathrm{E}\left[\mathbf{P}(\text{decoding } (W, J_1, J_2) \text{ wrongly at } D)\right] \leq \frac{\epsilon_0}{3}, \tag{B.5}$$

by choosing T large enough. The above expectation is taken over the (randomly generated) code, and ϵ_0 is an arbitrary constant.

For a fixed $W = w$, the codewords generated at S have the same distribution as a randomly generated code of rate $(B_1 + B_2)$. Thus, from (B.3) and from Theorem 8.7.1 in [9] it follows that for T large enough,

$$\mathrm{E}\left[\mathbf{P}(A_w^{(1)})\right] \leq \frac{\epsilon_0}{3}, \tag{B.6}$$

where $A_w^{(1)}$ is the event that E_1 makes a decoding error when trying to decode (J_1, J_2), assuming that $W = w$ and assuming that W is already available at E_1. Again, the expectation is taken over the randomly generated code. Since this is true for all $w \in \{1, \ldots, 2^{\lfloor TR \rfloor}\}$, it follows that for T large enough,

$$\mathrm{E}\left[\frac{1}{2^{\lfloor TR \rfloor}} \sum_{w=1}^{2^{\lfloor TR \rfloor}} \mathbf{P}(A_w^{(1)})\right] \leq \frac{\epsilon_0}{3}. \tag{B.7}$$

We apply the same argument once more for E_2, using (B.4) and keeping $(W, J_1) = (w, j_1)$ fixed. It follows that

$$\mathrm{E}\left[\frac{1}{2^{\lfloor TR \rfloor + \lfloor TB_1 \rfloor}} \sum_{w=1}^{2^{\lfloor TR \rfloor}} \sum_{j_1=1}^{2^{\lfloor TB_1 \rfloor}} \mathbf{P}(A_{w,j_1}^{(2)})\right] \leq \frac{\epsilon_0}{3}, \tag{B.8}$$

where $A_{w,j_1}^{(2)}$ is the event that E_2 makes a decoding error when trying to decode J_2, assuming that $(W, J_1) = (w, j_1)$ and that (W, J_1) is already available at E_2.

Combining (B.5), (B.7) and (B.8), we obtain

$$\begin{aligned}\mathrm{E}\Big[&\mathbf{P}(\text{decoding } (W, J_1, J_2) \text{ wrongly at } D) + \\ &\frac{1}{2^{\lfloor TR \rfloor}} \sum_{w=1}^{2^{\lfloor TR \rfloor}} \mathbf{P}(A_w^{(1)})\Big] + \\ &\frac{1}{2^{(\lfloor TR \rfloor + \lfloor TB_1 \rfloor)}} \sum_{w=1}^{2^{\lfloor TR \rfloor}} \sum_{j_1=1}^{2^{\lfloor TB_1 \rfloor}} \mathbf{P}(A_{w,j_1}^{(2)})\Big] \leq \epsilon_0.\end{aligned}$$

We conclude that for at least one code,

$$\mathbf{P}(\text{decoding } (W, J_1, J_2) \text{ wrongly at } D) \leq \epsilon_0, \tag{B.9}$$

$$\frac{1}{2^{\lfloor TR \rfloor}} \sum_{w=1}^{2^{\lfloor TR \rfloor}} \mathbf{P}(A_w^{(1)}) \leq \epsilon_0 \tag{B.10}$$

and

$$\frac{1}{2^{(\lfloor TR\rfloor+\lfloor TB_1\rfloor)}}\sum_{w=1}^{2^{\lfloor TR\rfloor}}\sum_{j_1=1}^{2^{\lfloor TB_1\rfloor}}\mathbf{P}(A^{(2)}_{w,j_1})\le\epsilon_0 \tag{B.11}$$

at the same time. From (B.9), we see that this particular code is reliable. Further, (B.10) is nothing but the error probability when estimating (J_1, J_2) from the observable $(W, \mathbf{Y}_{\mathcal{A}_{E_1}})$ for this particular code. The probability is taken over (W, J_1, J_2) and the channel realization. Hence, by Fano's Lemma [9], it follows that

$$H(J_1, J_2|W, \mathbf{Y}_{\mathcal{A}_{E_1}}) \le 1 + T(B_1 + B_2)\epsilon_0. \tag{B.12}$$

Similarly, (B.11) implies that

$$H(J_2|W, J_1, \mathbf{Y}_{\mathcal{A}_{E_2}}) \le 1 + TB_2\epsilon_0. \tag{B.13}$$

Note that the rate R of this code, given in (B.1), is as claimed by the theorem. It remains to analyze the equivocations at E_1 and E_2. We can write

$$\begin{aligned}H(W|\mathbf{Y}_{\mathcal{A}_{E_1}}) &\ge I(W;\mathbf{X}_S|\mathbf{Y}_{\mathcal{A}_{E_1}})\\ &= I(\mathbf{X}_S;W,\mathbf{Y}_{\mathcal{A}_{E_1}})\\ &\quad - I(\mathbf{X}_S;\mathbf{Y}_{\mathcal{A}_{E_1}})\\ &= H(\mathbf{X}_S) - H(\mathbf{X}_S|W,\mathbf{Y}_{\mathcal{A}_{E_1}})\\ &\quad - I(\mathbf{X}_S;\mathbf{Y}_{\mathcal{A}_{E_1}}).\end{aligned} \tag{B.14}$$

We bound the three terms above separately:

$$\begin{aligned}H(\mathbf{X}_S) &\ge I(\mathbf{X}_S;W,J_1,J_2)\\ &= H(W,J_1,J_2) - H(W,J_1,J_2|\mathbf{X}_S)\\ &\overset{(a)}{=} H(W,J_1,J_2) - H(W,J_1,J_2|\mathbf{X}_S,\mathbf{Y}_{\mathcal{A}})\\ &\ge H(W,J_1,J_2) - H(W,J_1,J_2|\mathbf{Y}_{\mathcal{A}})\\ &\overset{(b)}{\ge} H(W,J_1,J_2) - 1 - T(R+B_1+B_2)\epsilon_0\\ &\overset{(c)}{=} T(R+B_1+B_2-\epsilon_2) - 1 - T(R+B_1+B_2)\epsilon_0,\end{aligned} \tag{B.15}$$

where (a) is true because of the Markov chain $(W, J_1, J_2) \multimap \mathbf{X}_S \multimap \mathbf{Y}_{\mathcal{A}}$, while (b) follows from Fano's inequality, together with (B.9). Finally, the small constant ϵ_2 in (c) is introduced because we drop the rounding in $\lfloor TR\rfloor+\lfloor TB_1\rfloor+\lfloor TB_2\rfloor$. As T grows large, ϵ_2 goes to zero.

The second term in (B.14) can be upper bounded as follows:

$$\begin{aligned}H(\mathbf{X}_S,|W,\mathbf{Y}_{\mathcal{A}_{E_1}}) &\le H(\mathbf{X}_S,J_1,J_2|W,\mathbf{Y}_{\mathcal{A}_{E_1}})\\ &= H(J_1,J_2|W,\mathbf{Y}_{\mathcal{A}_{E_1}}) + \underbrace{H(\mathbf{X}_S|W,J_1,J_2,\mathbf{Y}_{\mathcal{A}_{E_1}})}_{=0}\\ &\le 1 + T(B_1+B_2)\epsilon_0,\end{aligned} \tag{B.16}$$

where we used (B.12) in the last inequality and the indicated term is zero because the encoder at S is a deterministic function.

The third term in (B.14) can be bounded as follows:

$$\begin{aligned}
I(\mathbf{X}_S;\mathbf{Y}_{\mathcal{A}_{E_1}}) &\le \sum_{t=1}^{T}\big[H(Y_{\mathcal{A}_{E_1}}[t]|Y_{\mathcal{A}_{E_1}}[1],\dots,Y_{\mathcal{A}_{E_1}}[t-1])\\
&\quad - H(Y_{\mathcal{A}_{E_1}}[t]|Y_{\mathcal{A}_{E_1}}[1],\dots,Y_{\mathcal{A}_{E_1}}[t-1],\mathbf{X}_S)\big]\\
&\overset{(a)}{\le} \sum_{t=1}^{T}\Big(H(Y_{\mathcal{A}_{E_1}}[t]) - H(Y_{\mathcal{A}_{E_1}}[t]|X_S[t])\Big)\\
&= \sum_{t=1}^{T} I(X_S[t];Y_{\mathcal{A}_{E_1}}[t])\\
&\to TI(X_S;Y_{\mathcal{A}_{E_1}}) \qquad \text{(B.17)}
\end{aligned}$$

as T grows, because $\mathbf{X}_S$ is selected randomly among roughly $2^{T(R+B_1+B_2)}$ sequences that lie all in $\mathcal{T}_\delta(X_S)$. Therefore, as T grows large, any component $X_S[t]$ of $\mathbf{X}_S$ has, with high probability, the same distribution as the underlying random variable X_S. The convergence in (B.17) follows from the fact that the mutual information is a continuous function of the underlying distributions. The inequality in (a) follows from the memoryless nature of the channel.

Plugging (B.15), (B.16) and (B.17) into (B.14), we obtain

$$\begin{aligned}
\frac{1}{T}H(W|\mathbf{Y}_{\mathcal{A}_{E_1}}) &\ge R + B_1 + B_2 - \epsilon_2 - \frac{1}{T} - (R+B_1+B_2)\epsilon_0\\
&\quad - \frac{1}{T} - (B_1+B_2)\epsilon_0\\
&\quad - I(X_S;Y_{\mathcal{A}_{E_1}})\\
&\overset{(a)}{=} R + I(X_S;Y_{\mathcal{A}_{E_1}}) - \epsilon_1 - \epsilon_2 - \frac{1}{T} - (R+B_1+B_2)\epsilon_0\\
&\quad - \frac{1}{T} - (B_1+B_2)\epsilon_0\\
&\quad - I(X_S;Y_{\mathcal{A}_{E_1}})\\
&= R - \epsilon_1 - \epsilon_2 - \frac{2}{T} - (R+2B_1+2B_2)\epsilon_0, \qquad \text{(B.18)}
\end{aligned}$$

where the equality (a) follows from (B.3). Hence, by choosing T large enough and ϵ_0, ϵ_1 sufficiently small, the equivocation at eavesdropper E_1 can be made arbitrarily close to R.

For eavesdropper E_2 we proceed in a similar manner. We bound the equivocation as

$$\begin{aligned}
H(W|\mathbf{Y}_{\mathcal{A}_{E_2}}) &\ge H(W|\mathbf{Y}_{\mathcal{A}_{E_2}},J_1)\\
&\ge H(\mathbf{X}_S|J_1) - H(\mathbf{X}_S|W,\mathbf{Y}_{\mathcal{A}_{E_2}},J_1)\\
&\quad - I(\mathbf{X}_S;\mathbf{Y}_{\mathcal{A}_{E_2}}|J_1). \qquad \text{(B.19)}
\end{aligned}$$

We bound the three terms above separately:

$$\begin{aligned}
H(\mathbf{X}_S|J_1) &\geq I(\mathbf{X}_S; W, J_2|J_1) \\
&= H(W, J_2|J_1) - H(W, J_2|\mathbf{X}_S, J_1) \\
&\overset{(a)}{=} H(W, J_2) - H(W, J_2|\mathbf{X}_S, J_1) \\
&= T(R + B_2 - \epsilon_3) - 1 - T(R + B_1 + B_2)\epsilon_0, \qquad \text{(B.20)}
\end{aligned}$$

where (a) is true because (W, J_1, J_2) are independent, and the remaining steps are analogous to the steps used in (B.15). As T grows large, ϵ_3 goes to zero.

The second term in (B.19) can be upper bounded as follows:

$$\begin{aligned}
H(\mathbf{X}_S, |W, \mathbf{Y}_{\mathcal{A}_{E_2}}, J_1) &\leq H(\mathbf{X}_S, J_2|W, \mathbf{Y}_{\mathcal{A}_{E_2}}, J_1) \\
&= H(J_2|W, \mathbf{Y}_{\mathcal{A}_{E_2}}, J_1) + \underbrace{H(\mathbf{X}_S|W, J_1, J_2, \mathbf{Y}_{\mathcal{A}_{E_2}})}_{=0} \\
&\leq 1 + TB_2\epsilon_0, \qquad \text{(B.21)}
\end{aligned}$$

where we used (B.13) in the last inequality and the indicated term is zero because the encoder at S is a deterministic function.

The third term in (B.19) can be bounded as follows:

$$\begin{aligned}
I(\mathbf{X}_S; \mathbf{Y}_{\mathcal{A}_{E_2}}|J_1) &= H(\mathbf{Y}_{\mathcal{A}_{E_2}}|J_1) - H(\mathbf{Y}_{\mathcal{A}_{E_2}}|\mathbf{X}_S, J_1) \\
&\overset{(a)}{\leq} H(\mathbf{Y}_{\mathcal{A}_{E_2}}) - H(\mathbf{Y}_{\mathcal{A}_{E_2}}|\mathbf{X}_S) \\
&= I(\mathbf{X}_S; \mathbf{Y}_{\mathcal{A}_{E_2}}) \\
&\leq TI(X_S; Y_{\mathcal{A}_{E_2}}), \qquad \text{(B.22)}
\end{aligned}$$

where the last inequality holds asymptotically with large T. In (a), we have used the Markov chain $J_1 -\!\!\circ\!\!- \mathbf{X}_S -\!\!\circ\!\!- \mathbf{Y}_{\mathcal{A}_{E_2}}$, and all other steps are analogous to the steps in (B.17).

Plugging (B.20), (B.21) and (B.22) into (B.19), we obtain

$$\begin{aligned}
\frac{1}{T}H(W|\mathbf{Y}_{\mathcal{A}_{E_2}}) &\geq R + B_2 - \epsilon_3 - \frac{1}{T} - (R + B_1 + B_2)\epsilon_0 \\
&\quad - \frac{1}{T} - B_2\epsilon_0 \\
&\quad - I(X_S; Y_{\mathcal{A}_{E_2}}) \\
&\overset{(a)}{=} R + I(X_S; Y_{\mathcal{A}_{E_2}}) - \epsilon_1 - \epsilon_3 - \frac{1}{T} - (R + B_1 + B_2)\epsilon_0 \\
&\quad - \frac{1}{T} - B_2\epsilon_0 \\
&\quad - I(X_S; Y_{\mathcal{A}_{E_2}}) \\
&= R - \epsilon_1 - \epsilon_3 - \frac{2}{T} - (R + B_1 + 2B_2)\epsilon_0, \qquad \text{(B.23)}
\end{aligned}$$

where (a) follows from (B.4). Hence, by choosing T large enough and ϵ_0, ϵ_1 sufficiently small, the equivocation at eavesdropper E_2 can also be made arbitrarily close to R.

This proves that the rate-equivocation pair (R, R) is weakly achievable, where

$$R = I(X_S; Y_{\mathcal{A}}) - \max_{E \in \mathcal{E}} I(X_S; Y_{\mathcal{A}_E}).$$

Thus, it follows from Lemma E.5 that the secrecy rate R is strongly achievable over this fan network. Hence, $\bar{C}_s \geq R$. Since this is true for any transmit distribution p_{X_S}, the claim of the theorem follows.

Appendix for Chapter 5 C

C.1 Proofs of the List Size Lemmas

C.1.1 Proof of Lemma 5.2

First, assume that $\tilde{R} > I(X;Z) + \delta_1$. Fix any $\mathbf{z} \in \mathcal{T}_{\delta-\epsilon}(Z)$, where $\epsilon > 0$ can be chosen arbitrarily small. For a given code $\mathcal{C}$, the set $\mathcal{L}_\delta(X|\mathbf{z})$ is deterministic. However, since the code $\mathcal{C}$ is constructed by picking sequences randomly from $\mathcal{T}_\delta(X)$, we can consider $\mathcal{L}_\delta(X|\mathbf{z})$ to be a random set, and $L_\delta(X|\mathbf{z})$ a random variable. Define $\mathcal{I} \triangleq \{1, \ldots, 2^{T\tilde{R}}\}$ and let $\tilde{\mathbf{X}}(i)$ be the i^{th} sequence in $\mathcal{C}$, where $i \in \mathcal{I}$. We have

$$\begin{aligned}
\mathrm{E}\left[L_\delta(X|\mathbf{z})\right] &= \mathrm{E}\left[\sum_{i=1}^{2^{T\tilde{R}}} \mathbb{I}_{\{(\tilde{\mathbf{X}}(i),\mathbf{z}) \in \mathcal{T}(X,Z)\}}\right] \\
&= \sum_{i=1}^{2^{T\tilde{R}}} \mathbf{P}\big((\tilde{\mathbf{X}}(i),\mathbf{z}) \in \mathcal{T}_\delta(X,Z)\big) && \text{(C.1)} \\
&\overset{\delta_1}{=} 2^{T\tilde{R}} 2^{-TI(X;Z)}, && \text{(C.2)}
\end{aligned}$$

where the last step is true from Lemma F.4.

To prove the lemma, we need to show that $L_\delta(X|\mathbf{z})$ concentrates around its mean as T gets large. We start by computing the second moment of $L_\delta(X|\mathbf{z})$, where the expectation is again over all random codes.

$$\begin{aligned}
\mathrm{E}\left[L_\delta(X|\mathbf{z})^2\right] &= \mathrm{E}\left[\sum_{i=1}^{2^{T\tilde{R}}}\sum_{j=1}^{2^{T\tilde{R}}} \mathbb{I}_{(\tilde{\mathbf{X}}(i),\mathbf{z})\in\mathcal{T}_\delta(X,Z),(\tilde{\mathbf{X}}(j),\mathbf{z})\in\mathcal{T}_\delta(X,Z)}\right] \\
&= \mathrm{E}\left[\sum_{i=1}^{2^{T\tilde{R}}} \mathbb{I}_{\{(\tilde{\mathbf{X}}(i),\mathbf{z})\in\mathcal{T}_\delta(X,Z)\}}\right] \\
&\quad + \mathrm{E}\left[\sum_{i=1}^{2^{T\tilde{R}}}\sum_{\substack{j=1\\ j\neq i}}^{2^{T\tilde{R}}} \mathbb{I}_{\{(\tilde{\mathbf{X}}(i),\mathbf{z})\in\mathcal{T}_\delta(X,Z),(\tilde{\mathbf{X}}(j),\mathbf{z})\in\mathcal{T}_\delta(X,Z)\}}\right] \\
&= \sum_{i=1}^{2^{T\tilde{R}}} \mathbf{P}\big((\tilde{\mathbf{X}}(i),\mathbf{z})\in\mathcal{T}_\delta(X,Z)\big) \\
&\quad + \sum_{i=1}^{2^{T\tilde{R}}}\sum_{\substack{j=1\\ j\neq i}}^{2^{T\tilde{R}}} \mathbf{P}\big((\tilde{\mathbf{X}}(i),\mathbf{z})\in\mathcal{T}_\delta(X,Z),(\tilde{\mathbf{X}}(j),\mathbf{z})\in\mathcal{T}_\delta(X,Z)\big) \\
&\leq \sum_{i=1}^{2^{T\tilde{R}}} \mathbf{P}\big((\tilde{\mathbf{X}}(i),\mathbf{z})\in\mathcal{T}_\delta(X,Z)\big) \\
&\quad + \left(\sum_{i=1}^{2^{T\tilde{R}}} \mathbf{P}\big((\tilde{\mathbf{X}}(i),\mathbf{z})\in\mathcal{T}_\delta(X,Z)\big)\right)^2 \\
&= \mathrm{E}\left[L_\delta(X|\mathbf{z})\right] + \mathrm{E}\left[L_\delta(X|\mathbf{z})\right]^2, \qquad \text{(C.3)}
\end{aligned}$$

where the last inequality is true because $\tilde{\mathbf{X}}(i)$ and $\tilde{\mathbf{X}}(j)$ are independent for $i \neq j$. Thus, from (C.1) and (C.3), the variance of the list size is bounded as

$$\begin{aligned}
\mathrm{Var}\left(L_\delta(X|\mathbf{z})\right) &= \mathrm{E}\left[L_\delta(X|\mathbf{z})^2\right] - \mathrm{E}\left[L_\delta(X|\mathbf{z})\right]^2 \\
&\leq \mathrm{E}\left[L_\delta(X|\mathbf{z})\right].
\end{aligned}$$

From Chebyshev's inequality, we obtain

$$\begin{aligned}
&\mathbf{P}\Big(\big|L_\delta(X|\mathbf{z}) - \mathrm{E}\left[L_\delta(X|\mathbf{z})\right]\big| > \epsilon\mathrm{E}\left[L_\delta(X|\mathbf{z})\right]\Big) \\
&\qquad \leq \frac{\mathrm{Var}\left(L_\delta(X|\mathbf{z})\right)}{\epsilon^2\mathrm{E}\left[L_\delta(X|\mathbf{z})\right]^2} \\
&\qquad \leq \frac{1}{\epsilon^2\mathrm{E}\left[L_\delta(X|\mathbf{z})\right]} \\
&\qquad \overset{(a)}{\leq} \frac{1}{\epsilon^2 2^{T\left(\tilde{R}-I(X;Z)-\delta_1\right)}} \\
&\qquad \to 0,
\end{aligned}$$

as $T \to \infty$, because when $\tilde{R} > I(X;Z) + \delta_1$, then the denominator in (a) has a positive exponent and hence grows exponentially with T. The inequality (a) is true from the δ_1-exponential equality in (C.2). Hence, when $\tilde{R} > I(X;Z) + \delta_1$, the random variable $L_\delta(X|\mathbf{z})$ hardens to its expectation, which is δ_1-equal to $2^{T(\tilde{R}-I(X;Z))}$ for any $\mathbf{z} \in \mathcal{T}_{\delta-\epsilon}(Z)$. Since the set $\mathcal{T}_{\delta-\epsilon}(Z)$ is ϵ-equal to $\mathcal{T}_\delta(Z)$ and $\mathcal{L}_\delta(Z) \subseteq \mathcal{T}_\delta(Z)$, this proves the lemma for the upper regime of $\tilde{R}$, for almost every $\mathbf{z} \in \mathcal{L}_\delta(Z)$.

Assume now that $\tilde{R} < I(X;Z) - \delta_1$. For a fixed $\mathbf{z} \in \mathcal{T}_{\delta-\epsilon}(Z)$, (C.2) still holds, and the expected list size is bounded as

$$\mathrm{E}\left[L_\delta(X|\mathbf{z})\right] \leq 2^{T(\tilde{R}-I(X;Z)+\delta_1)} < 1.$$

Hence, the sets $\mathcal{L}_\delta(X|\mathbf{z})$ cannot be of the same size for all $\mathbf{z} \in \mathcal{T}_{\delta-\epsilon}(Z)$, because a list is of integer size. We can prove that in this case, $L_\delta(X|\mathbf{z}) = 1$ if $\mathbf{z} \in \mathcal{L}_\delta(Z) \cap \mathcal{T}_{\delta-\epsilon}(Z)$. If, however, $\mathbf{z} \notin \mathcal{L}_\delta(Z)$, then $L_\delta(X|\mathbf{z})$ is zero. This second claim follows directly from the fact that by definition, if $\mathbf{z} \notin \mathcal{L}_\delta(Z)$, then there is no $\mathbf{x} \in \mathcal{C}$ for which $(\mathbf{x},\mathbf{z}) \in \mathcal{T}_\delta(X,Z)$. It remains to prove the first part of the claim, namely that $L_\delta(X|\mathbf{z}) = 1$ if $\mathbf{z} \in \mathcal{L}_\delta(Z) \cap \mathcal{T}_{\delta-\epsilon}(Z)$. To see this, recall that $\mathcal{L}_\delta(Z)$ is the set of sequences $\mathbf{z}$ that are jointly typical with at least one codeword $\mathbf{x} \in \mathcal{C}$, and hence, $L_\delta(X|\mathbf{z}) \geq 1$ for $\mathbf{z} \in \mathcal{L}_\delta(Z)$. For a fixed $\mathbf{z} \in \mathcal{L}_\delta(Z) \cap \mathcal{T}_{\delta-\epsilon}(Z)$, we assume without loss of generality that $\tilde{\mathbf{X}}(1)$ is the codeword that is the witness of the fact that $\mathbf{z} \in \mathcal{L}_\delta(Z)$. Then,

$$\begin{aligned}
\mathrm{E}\left[L_\delta(X|\mathbf{z})|\mathbf{z} \in \mathcal{L}_\delta(Z)\right] &= 1 + \sum_{i=2}^{2^{T\tilde{R}}} \mathbf{P}\big((\tilde{\mathbf{X}}(i),\mathbf{z}) \in \mathcal{T}_\delta(X,Z)\big) \\
&\leq 1 + \sum_{i=1}^{2^{T\tilde{R}}} \mathbf{P}\big((\tilde{\mathbf{X}}(i),\mathbf{z}) \in \mathcal{T}_\delta(X,Z)\big) \\
&= 1 + \mathrm{E}\left[L_\delta(X|\mathbf{z})\right] \\
&\leq 1 + 2^{T(\tilde{R}-I(X;Z)+\delta_1)}
\end{aligned}$$

The last inequality follows from (C.2), which is still true. By Markov's inequality, we have that for any $\delta' > 0$,

$$\begin{aligned}
\mathbf{P}\Big(|L_\delta(X|\mathbf{z}) - 1| > 1/2\big|\mathbf{z} \in \mathcal{L}_\delta(Z)\Big) &\leq \frac{\mathrm{E}\left[|L_\delta(X|\mathbf{z}) - 1|\big|\mathbf{z} \in \mathcal{L}_\delta(Z)\right]}{1/2} \\
&= \frac{\mathrm{E}\left[L_\delta(X|\mathbf{z})\big|\mathbf{z} \in \mathcal{L}_\delta(Z)\right] - 1}{1/2} \\
&\leq 2\ 2^{T(\tilde{R}-I(X;Z)+\delta_1)}.
\end{aligned}$$

Since we assumed that $\tilde{R} < I(X;Z) - \delta_1$, the right hand side of the above inequality goes to zero exponentially fast. Hence, with high probability, the list size $L_\delta(X|\mathbf{z})$ lies in the interval $[1, 1.5]$. But since the list size is integer, it follows that with high probability (over the random codebook $\mathcal{C}$), $L_\delta(X|\mathbf{z}) = 1$. This concludes the proof of the list size.

C.1.2 Proof of Lemma 5.3

When $\tilde{R} < I(X;Z) - \delta_1$, then it follows from Lemma 5.2 that for almost all $\mathbf{z} \in \mathcal{L}_\delta(Z)$, $L_\delta(X|\mathbf{z}) = 1$. Hence, there is only one $\mathbf{x}' \in \mathcal{C}$ that is jointly typical with $\mathbf{z}$. Hence, the list $\mathcal{L}_\delta(Y|\mathbf{z})$ is equal to $\mathcal{T}_\delta(Y|\mathbf{x}', \mathbf{z})$ for the $\mathbf{x}'$ and $\mathbf{z}$ in question. By Lemma F.2, the size of this set is δ_2-equal to $2^{TH(Y|X,Z)}$, where $\delta_2 = \delta H(Y|X,Z)$.

Now, assume that $\tilde{R} > I(X;Z)+2\delta_1+\delta_2+4\delta_3$, and fix a sequence $\mathbf{z} \in \mathcal{T}_\delta(Z)$. We have

$$\begin{aligned} L_\delta(Y|\mathbf{z}) &= \sum_{\mathbf{y}\in\mathcal{T}_\delta(Y|\mathbf{z})} \mathbb{I}_{\{\bigcup_{i\in\mathcal{I}}\{\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{y},\mathbf{z})\}\}} && \text{(C.4)} \\ &\le \sum_{\mathbf{y}\in\mathcal{T}_\delta(Y|\mathbf{z})} \sum_{i\in\mathcal{I}} \mathbb{I}_{\{\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{y},\mathbf{z})\}}, && \text{(C.5)} \end{aligned}$$

where $\tilde{\mathbf{X}}(i)$ is defined to be the i^{th} codeword in $\mathcal{C}$. Taking the expectation over the random codebook, we get

$$\begin{aligned} \mathrm{E}\left[L_\delta(Y|\mathbf{z})\right] &= \sum_{\mathbf{y}\in\mathcal{T}_\delta(Y|\mathbf{z})} \mathbf{P}\Big(\bigcup_{i\in\mathcal{I}}\{\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{y},\mathbf{z})\}\Big) \\ &\le \sum_{\mathbf{y}\in\mathcal{T}_\delta(Y|\mathbf{z})} \sum_{i\in\mathcal{I}} \mathbf{P}\big(\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{y},\mathbf{z})\big) \\ &= \sum_{\mathbf{y}\in\mathcal{T}_\delta(Y|\mathbf{z})} \sum_{i\in\mathcal{I}} \tilde{p}_i(\mathbf{y}), && \text{(C.6)} \end{aligned}$$

where we used the definition $\tilde{p}_i(\mathbf{y}) \triangleq \mathbf{P}\big(\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{y},\mathbf{z})\big)$. By Lemma F.4,

$$\tilde{p}_i(\mathbf{y}) \overset{\delta_1}{=} 2^{-TI(X;Y,Z)} \qquad \text{(C.7)}$$

for almost all $\mathbf{y} \in \mathcal{T}_\delta(Y|\mathbf{z})$.

If $\tilde{R} < I(X;Y,Z) - \delta_1$, then we have from (C.7) that

$$\sum_{i\in\mathcal{I}} \tilde{p}_i(\mathbf{y}) \le 2^{T\left(\tilde{R}-I(X;Y,Z)+\delta_1\right)},$$

which goes to zero exponentially with T because the exponent of the right hand side is negative. Furthermore, the events $\{\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{y},\mathbf{z})\}$ are independent for different $i \in \mathcal{I}$. It follows that the union bound used in (C.6) is tight (see Appendix C.2), and thus,

$$\begin{aligned} \mathrm{E}\left[L_\delta(Y|\mathbf{z})\right] &\doteq \sum_{\mathbf{y}\in\mathcal{T}_\delta(Y|\mathbf{z})} \sum_{i\in\mathcal{I}} \tilde{p}_i(\mathbf{y}) && \text{(C.8)} \\ &\overset{\delta_4}{=} 2^{T\left(H(Y|Z)+\tilde{R}-I(X;Y,Z)\right)}, && \text{(C.9)} \end{aligned}$$

where $\delta_4 = \delta_1 + \delta_3$. Note that relative equality (as defined in Section 2.1) is stronger than exponential equality. Hence, we can consider that $\mathrm{E}\left[L_\delta(Y|\mathbf{z})\right] \overset{\delta_4}{=} 2^{T\left(H(Y|Z)+\tilde{R}-I(X;Y,Z)\right)}$.

On the other hand, if $\tilde{R} > I(X;Y,Z) + \delta_1$, then (C.7) implies that

$$\sum_{i \in \mathcal{I}} \tilde{p}_i(\mathbf{y}) \geq 2^{T\left(\tilde{R} - I(X;Y,Z) - \delta_1\right)}$$

grows exponentially with T, because the exponent of the right hand side is positive. Hence, it follows again from Appendix C.2 that the union bound is tight, and that $\mathbf{P}\big(\cup_{i \in \mathcal{I}}\{\tilde{\mathbf{X}}(i) \in \mathcal{T}_\delta(X|\mathbf{y},\mathbf{z})\big) \doteq 1$. Hence,

$$\begin{aligned} \mathrm{E}\left[L_\delta(Y|\mathbf{z})\right] &\doteq |\mathcal{T}_\delta(Y|\mathbf{z})| && \text{(C.10)} \\ &\overset{\delta_3}{=} 2^{TH(Y|Z)}. && \text{(C.11)} \end{aligned}$$

Thus, the expected list size satisfies the exponential equalities stated in the lemma. The rest of the proof is to show that for $\tilde{R} > I(X;Z) + \delta_1$, $L_\delta(Y|\mathbf{z})$ concentrates around its mean $\mathrm{E}\left[L_\delta(Y|\mathbf{z})\right]$ given in (C.9) and (C.11) for large T. Similarly to the proof of Lemma 5.2, this is done by first finding a bound on $\mathrm{Var}\left(L_\delta(Y|\mathbf{z})\right)$ and then using Chebyshev's inequality to show that the probability that $L_\delta(Y|\mathbf{z})$ deviates considerably from $\mathrm{E}\left[L_\delta(Y|\mathbf{z})\right]$ is small.

First, assume that $\tilde{R} < I(X;Y,Z) - \delta_1$. From (C.5),

$$\begin{aligned} L_\delta(Y|\mathbf{z})^2 \leq & \sum_{\mathbf{y}_1 \in \mathcal{T}_\delta(Y|\mathbf{z})} \sum_{\mathbf{y}_2 \in \mathcal{T}_\delta(Y|\mathbf{z})} \sum_{i_1 \in \mathcal{I}} \sum_{i_2 \in \mathcal{I}} \mathbb{I}_{\{\{\tilde{\mathbf{X}}(i_1) \in \mathcal{T}_\delta(X|\mathbf{y}_1,\mathbf{z})\} \cap \{\tilde{\mathbf{X}}(i_2) \in \mathcal{T}_\delta(X|\mathbf{y}_2,\mathbf{z})\}\}} \\ = & \sum_{\mathbf{y} \in \mathcal{T}_\delta(Y|\mathbf{z})} \sum_{i \in \mathcal{I}} \mathbb{I}_{\{\tilde{\mathbf{X}}(i) \in \mathcal{T}_\delta(X|\mathbf{y},\mathbf{z})\}} \\ & + \sum_{\mathbf{y} \in \mathcal{T}_\delta(Y|\mathbf{z})} \sum_{i_1 \in \mathcal{I}} \sum_{\substack{i_2 \in \mathcal{I} \\ i_2 \neq i_1}} \mathbb{I}_{\{\{\tilde{\mathbf{X}}(i_1) \in \mathcal{T}_\delta(X|\mathbf{y},\mathbf{z})\} \cap \{\tilde{\mathbf{X}}(i_2) \in \mathcal{T}_\delta(X|\mathbf{y},\mathbf{z})\}\}} \\ & + \sum_{\mathbf{y}_1 \in \mathcal{T}_\delta(Y|\mathbf{z})} \sum_{\substack{\mathbf{y}_2 \in \mathcal{T}_\delta(Y|\mathbf{z}) \\ \mathbf{y}_2 \neq \mathbf{y}_1}} \sum_{i \in \mathcal{I}} \mathbb{I}_{\{\{\tilde{\mathbf{X}}(i) \in \mathcal{T}_\delta(X|\mathbf{y}_1,\mathbf{z})\} \cap \{\tilde{\mathbf{X}}(i) \in \mathcal{T}_\delta(X|\mathbf{y}_2,\mathbf{z})\}\}} \\ & + \sum_{\mathbf{y}_1 \in \mathcal{T}_\delta(Y|\mathbf{z})} \sum_{\substack{\mathbf{y}_2 \in \mathcal{T}_\delta(Y|\mathbf{z}) \\ \mathbf{y}_2 \neq \mathbf{y}_1}} \sum_{i_1 \in \mathcal{I}} \sum_{\substack{i_2 \in \mathcal{I} \\ i_2 \neq i_1}} \mathbb{I}_{\{\{\tilde{\mathbf{X}}(i_1) \in \mathcal{T}_\delta(X|\mathbf{y}_1,\mathbf{z})\} \cap \{\tilde{\mathbf{X}}(i_2) \in \mathcal{T}_\delta(X|\mathbf{y}_2,\mathbf{z})\}\}}. \end{aligned}$$

Now we take the expectation of $L_\delta(Y|\mathbf{z})^2$ over the random codebooks:

$$\mathrm{E}\left[L_\delta(Y|\mathbf{z})^2\right] \leq \sum_{\mathbf{y}\in\mathcal{T}_\delta(Y|\mathbf{z})} \sum_{i\in\mathcal{I}} \mathbf{P}\big(\tilde{\mathbf{X}}(i) \in \mathcal{T}_\delta(X|\mathbf{y},\mathbf{z})\big) \tag{C.12}$$

$$+ \sum_{\mathbf{y}\in\mathcal{T}_\delta(Y|\mathbf{z})} \sum_{i_1\in\mathcal{I}} \sum_{\substack{i_2\in\mathcal{I}\\ i_2\neq i_1}} \mathbf{P}\big(\{\tilde{\mathbf{X}}(i_1) \in \mathcal{T}_\delta(X|\mathbf{y},\mathbf{z})\} \cap \{\tilde{\mathbf{X}}(i_2) \in \mathcal{T}_\delta(X|\mathbf{y},\mathbf{z})\}\big) \tag{C.13}$$

$$+ \mathrm{E}\Big[\sum_{\mathbf{y}_1\in\mathcal{T}_\delta(Y|\mathbf{z})} \sum_{\substack{\mathbf{y}_2\in\mathcal{T}_\delta(Y|\mathbf{z})\\ \mathbf{y}_2\neq\mathbf{y}_1}} \sum_{i\in\mathcal{I}} \mathbb{I}_{\{\{\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{y}_1,\mathbf{z})\}\cap\{\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{y}_2,\mathbf{z})\}\}}\Big] \tag{C.14}$$

$$+ \sum_{\mathbf{y}_1\in\mathcal{T}_\delta(Y|\mathbf{z})} \sum_{\substack{\mathbf{y}_2\in\mathcal{T}_\delta(Y|\mathbf{z})\\ \mathbf{y}_2\neq\mathbf{y}_1}} \sum_{i_1\in\mathcal{I}} \sum_{\substack{i_2\in\mathcal{I}\\ i_2\neq i_1}} \mathbf{P}\big(\{\tilde{\mathbf{X}}(i_1) \in \mathcal{T}_\delta(X|\mathbf{y}_1,\mathbf{z})\} \cap \{\tilde{\mathbf{X}}(i_2) \in \mathcal{T}_\delta(X|\mathbf{y}_2,\mathbf{z})\}\big). \tag{C.15}$$

We treat each of the terms (C.12) to (C.15) separately. First, consider (C.12):

$$\sum_{\mathbf{y}\in\mathcal{T}_\delta(Y|\mathbf{z})} \sum_{i\in\mathcal{I}} \mathbf{P}\big(\tilde{\mathbf{X}}(i) \in \mathcal{T}_\delta(X|\mathbf{y},\mathbf{z})\big) = \sum_{\mathbf{y}\in\mathcal{T}_\delta(Y|\mathbf{z})} \sum_{i\in\mathcal{I}} \tilde{p}_i(\mathbf{y}) \doteq \mathrm{E}\left[L_\delta(Y|\mathbf{z})\right],$$

where the relative equality was given in (C.8). Consider now the term (C.13). Keeping in mind that different codewords $\tilde{\mathbf{X}}(i_1)$ and $\tilde{\mathbf{X}}(i_2)$ are picked independently from $\mathcal{T}_\delta(X)$, we obtain

$$\begin{aligned}\sum_{\mathbf{y}\in\mathcal{T}_\delta(Y|\mathbf{z})} \sum_{i_1\in\mathcal{I}} \sum_{\substack{i_2\in\mathcal{I}\\ i_2\neq i_1}} &\mathbf{P}\big(\{\tilde{\mathbf{X}}(i_1),\mathbf{y}) \in \mathcal{T}_\delta(X|\mathbf{y},\mathbf{z})\} \cap \{\tilde{\mathbf{X}}(i_2),\mathbf{y}) \in \mathcal{T}_\delta(X|\mathbf{y},\mathbf{z})\}\big)\\ &= \sum_{\mathbf{y}\in\mathcal{T}_\delta(Y|\mathbf{z})} \sum_{i_1\in\mathcal{I}} \sum_{\substack{i_2\in\mathcal{I}\\ i_2\neq i_1}} \tilde{p}_{i_1}(\mathbf{y})\tilde{p}_{i_2}(\mathbf{y})\\ &\leq a2^{T\delta_3}b(b-1)\big(c2^{T\delta_1}\big)^2\\ &\leq ab^2c^2 2^{T(2\delta_1+\delta_3)},\end{aligned}$$

where

$$\begin{aligned} a &\triangleq 2^{TH(Y|Z)},\\ b &\triangleq 2^{T\tilde{R}},\\ c &\triangleq 2^{-TI(X;Y,Z)}.\end{aligned}$$

Let us now consider (C.15):

$$\begin{aligned}
&\sum_{\mathbf{y}_1\in\mathcal{T}_\delta(Y|\mathbf{z})}\sum_{\substack{\mathbf{y}_2\in\mathcal{T}_\delta(Y|\mathbf{z})\\ \mathbf{y}_2\neq\mathbf{y}_1}}\sum_{i_1\in\mathcal{I}}\sum_{\substack{i_2\in\mathcal{I}\\ i_2\neq i_1}} \mathbf{P}\big(\{\tilde{\mathbf{X}}(i_1)\in\mathcal{T}_\delta(X|\mathbf{y}_1,\mathbf{z})\}\\
&\qquad\qquad\qquad\qquad\cap\{\tilde{\mathbf{X}}(i_2)\in\mathcal{T}_\delta(X|\mathbf{y}_2,\mathbf{z})\}\big)\\
&=\sum_{\mathbf{y}_1\in\mathcal{T}_\delta(Y|\mathbf{z})}\sum_{\substack{\mathbf{y}_2\in\mathcal{T}_\delta(Y|\mathbf{z})\\ \mathbf{y}_2\neq\mathbf{y}_1}}\sum_{i_1\in\mathcal{I}}\sum_{\substack{i_2\in\mathcal{I}\\ i_2\neq i_1}} \tilde{p}_{i_1}(\mathbf{y}_1)\tilde{p}_{i_2}(\mathbf{y}_2)\\
&=\sum_{\mathbf{y}_1\in\mathcal{T}_\delta(Y|\mathbf{z})}\sum_{i_1\in\mathcal{I}}\tilde{p}_{i_1}(\mathbf{y}_1)\sum_{\substack{\mathbf{y}_2\in\mathcal{T}_\delta(Y|\mathbf{z})\\ \mathbf{y}_2\neq\mathbf{y}_1}}\sum_{\substack{i_2\in\mathcal{I}\\ i_2\neq i_1}}\tilde{p}_{i_2}(\mathbf{y}_2)\\
&\leq \mathrm{E}\left[L_\delta(Y|\mathbf{z})\right]^2.
\end{aligned}$$

The computation of (C.14) is more involved. We can write

$$\begin{aligned}
&\frac{1}{|\mathcal{T}_\delta(Y|\mathbf{z})|-1}\sum_{\mathbf{y}_1\in\mathcal{T}_\delta(Y|\mathbf{z})}\sum_{\substack{\mathbf{y}_2\in\mathcal{T}_\delta(Y|\mathbf{z})\\ \mathbf{y}_2\neq\mathbf{y}_1}}\sum_{i\in\mathcal{I}}\mathbb{I}_{\{\{\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{y}_1,\mathbf{z})\}\cap\{\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{y}_2,\mathbf{z})\}\}}\\
&=\sum_{\mathbf{y}_1\in\mathcal{T}_\delta(Y|\mathbf{z})}\sum_{i\in\mathcal{I}}\sum_{\substack{\mathbf{y}_2\in\mathcal{T}_\delta(Y|\mathbf{z})\\ \mathbf{y}_2\neq\mathbf{y}_1}}\frac{1}{|\mathcal{T}_\delta(Y|\mathbf{z})|-1}\mathbb{I}_{\{\{\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{y}_1,\mathbf{z})\}\cap\{\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{y}_2,\mathbf{z})\}\}}\\
&\overset{(a)}{=}\sum_{\mathbf{y}_1\in\mathcal{T}_\delta(Y|\mathbf{z})}\sum_{i\in\mathcal{I}}\mathrm{E}_{\mathbf{Y}_2}\left[\mathbb{I}_{\{\{\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{y}_1,\mathbf{z})\}\cap\{\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{Y}_2,\mathbf{z})\}\}}\right]\\
&=\sum_{\mathbf{y}_1\in\mathcal{T}_\delta(Y|\mathbf{z})}\sum_{i\in\mathcal{I}}\mathbf{P}_{\mathbf{Y}_2}\Big(\{\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{y}_1,\mathbf{z})\}\cap\{\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{Y}_2,\mathbf{z})\}\Big)\\
&=\sum_{\mathbf{y}_1\in\mathcal{T}_\delta(Y|\mathbf{z})}\sum_{i\in\mathcal{I}}\mathbf{P}_{\mathbf{Y}_2}(\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{y}_1,\mathbf{z}))\\
&\qquad\qquad\qquad\qquad\mathbf{P}_{\mathbf{Y}_2}(\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{Y}_2,\mathbf{z})|\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{y}_1,\mathbf{z}))\\
&=\sum_{\mathbf{y}_1\in\mathcal{T}_\delta(Y|\mathbf{z})}\sum_{i\in\mathcal{I}}\mathbb{I}_{\{\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{y}_1,\mathbf{z})\}}\mathbf{P}_{\mathbf{Y}_2}(\mathbf{Y}_2\in\mathcal{T}_\delta(Y|\tilde{\mathbf{X}}(i),\mathbf{z})|\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{y}_1,\mathbf{z}))\\
&=\sum_{\mathbf{y}_1\in\mathcal{T}_\delta(Y|\mathbf{z})}\sum_{i\in\mathcal{I}}\mathbb{I}_{\{\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{y}_1,\mathbf{z})\}}\frac{|\mathcal{T}_\delta(Y|\tilde{\mathbf{X}}(i),\mathbf{z})|-1}{|\mathcal{T}_\delta(Y|\mathbf{z})|-1}\\
&\overset{(b)}{\leq}\sum_{\mathbf{y}_1\in\mathcal{T}_\delta(Y|\mathbf{z})}\sum_{i\in\mathcal{I}}\mathbb{I}_{\{\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{y}_1,\mathbf{z})\}}\frac{|\mathcal{T}_\delta(Y|\tilde{\mathbf{X}}(i),\mathbf{z})|}{(1/2)|\mathcal{T}_\delta(Y|\mathbf{z})|}\\
&\leq 2\sum_{\mathbf{y}_1\in\mathcal{T}_\delta(Y|\mathbf{z})}\sum_{i\in\mathcal{I}}\mathbb{I}_{\{\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{y}_1,\mathbf{z})\}}\frac{2^{TH(Y|X,Z)}}{2^{TH(Y|Z)}}2^{T(\delta_2+\delta_3)}\\
&=2\sum_{\mathbf{y}_1\in\mathcal{T}_\delta(Y|\mathbf{z})}\sum_{i\in\mathcal{I}}\mathbb{I}_{\{\tilde{\mathbf{X}}(i)\in\mathcal{T}_\delta(X|\mathbf{y}_1,\mathbf{z})\}}2^{-TI(X;Y|Z)}2^{T(\delta_2+\delta_3)},
\end{aligned}$$

where in (a), we have defined the random variable $\mathbf{Y}_2$ to be a sequence chosen uniformly from $\mathcal{T}_\delta(Y|\mathbf{z}) \setminus \{\mathbf{y}_1\}$ (and independent of everything else). In (b) we assumed that $|\mathcal{T}_\delta(Y|\mathbf{z})| \geq 2$ and hence, $|\mathcal{T}_\delta(Y|\mathbf{z})| - 1$ is larger than $(1/2)|\mathcal{T}_\delta(Y|\mathbf{z})|$. Taking the expectation over all random codebooks yields:

$$
\begin{aligned}
&\mathrm{E}\Big[\sum_{\mathbf{y}_1 \in \mathcal{T}_\delta(Y|\mathbf{z})} \sum_{\substack{\mathbf{y}_2 \in \mathcal{T}_\delta(Y|\mathbf{z}) \\ \mathbf{y}_2 \neq \mathbf{y}_1}} \sum_{i \in \mathcal{I}} \mathbb{I}_{\{\{\tilde{\mathbf{X}}(i) \in \mathcal{T}_\delta(X|\mathbf{y}_1,\mathbf{z})\} \cap \{\tilde{\mathbf{X}}(i) \in \mathcal{T}_\delta(X|\mathbf{y}_2,\mathbf{z})\}\}} \Big] \\
&\leq 2(|\mathcal{T}_\delta(Y|\mathbf{z})| - 1) \sum_{\mathbf{y}_1 \in \mathcal{T}_\delta(Y|\mathbf{z})} \sum_{i \in \mathcal{I}} \mathbf{P}(\tilde{\mathbf{X}}(i) \in \mathcal{T}_\delta(X|\mathbf{y}_1, \mathbf{z})) 2^{-TI(X;Y|Z)} 2^{T(\delta_2+\delta_3)} \\
&\leq 2(a2^{T\delta_3} - 1) a 2^{T\delta_3} bc 2^{T\delta_1} 2^{-TI(X;Y|Z)} 2^{T(\delta_2+\delta_3)} \\
&\leq 2a^2bc 2^{-TI(X;Y|Z)} 2^{T(\delta_1+\delta_2+3\delta_3)}.
\end{aligned}
$$

Combining the four terms (C.12) to (C.15), we obtain

$$
\begin{aligned}
\mathrm{E}\left[L_\delta(Y|\mathbf{z})^2\right] \leq{}& \mathrm{E}\left[L_\delta(Y|\mathbf{z})\right] + ab^2c^2 2^{T(2\delta_1+\delta_3)} \\
&+ \mathrm{E}\left[L_\delta(Y|\mathbf{z})\right]^2 + 2a^2bc2^{-TI(X;Y|Z)} 2^{T(\delta_1+\delta_2+3\delta_3)}.
\end{aligned}
$$

Therefore, the variance of the list size can be bounded as

$$
\begin{aligned}
\mathrm{Var}\left(L_\delta(Y|\mathbf{z})\right) &= \mathrm{E}\left[L_\delta(Y|\mathbf{z})^2\right] - \mathrm{E}\left[L_\delta(Y|\mathbf{z})\right]^2 \\
&\leq \mathrm{E}\left[L_\delta(Y|\mathbf{z})\right] + ab^2c^2 2^{T(2\delta_1+\delta_3)} + 2a^2bc2^{-TI(X;Y|Z)} 2^{T(\delta_1+\delta_2+3\delta_3)}.
\end{aligned}
$$

From Chebyshev's inequality, we obtain

$$
\begin{aligned}
&\mathbf{P}\big(|L_\delta(Y|\mathbf{z}) - \mathrm{E}\left[L_\delta(Y|\mathbf{z})\right]| > \epsilon \mathrm{E}\left[L_\delta(Y|\mathbf{z})\right]\big) \\
&\leq \frac{\mathrm{Var}\left(L_\delta(Y|\mathbf{z})\right)}{\epsilon^2 \mathrm{E}\left[L_\delta(Y|\mathbf{z})\right]^2} \\
&\leq \frac{1}{\epsilon^2 \mathrm{E}\left[L_\delta(Y|\mathbf{z})\right]} + \frac{ab^2c^2 2^{T(2\delta_1+\delta_3)}}{\epsilon^2 \mathrm{E}\left[L_\delta(Y|\mathbf{z})\right]^2} + \frac{2a^2bc2^{-TI(X;Y|Z)} 2^{T(\delta_1+\delta_2+3\delta_3)}}{\epsilon^2 \mathrm{E}\left[L_\delta(Y|\mathbf{z})\right]^2} \\
&\overset{(a)}{\leq} \frac{1}{\epsilon^2 abc 2^{-T(\delta_1+\delta_3)}} + \frac{1}{\epsilon^2 a 2^{-T(3\delta_1+2\delta_3)}} + \frac{2}{\epsilon^2 bc 2^{TI(X;Y|Z)} 2^{-T(2\delta_1+\delta_2+4\delta_3)}} \\
&= \frac{1}{\epsilon^2 2^{T(\tilde{R}-I(X;Y,Z)+H(Y|Z)-\delta_1-\delta_3)}} + \frac{1}{\epsilon^2 2^{T(H(Y|Z)-3\delta_1-2\delta_3)}} \\
&\quad + \frac{2}{\epsilon^2 2^{T(\tilde{R}-I(X;Y,Z)+I(X;Y|Z)-2\delta_1-\delta_2-4\delta_3)}} \\
&= \frac{1}{\epsilon^2 2^{T(\tilde{R}-I(X;Z)+H(Y|X,Z)-\delta_1-\delta_3)}} + \frac{1}{\epsilon^2 2^{T(H(Y|Z)-3\delta_1-2\delta_3)}} \\
&\quad + \frac{2}{\epsilon^2 2^{T(\tilde{R}-I(X;Z)-2\delta_1-\delta_2-4\delta_3)}} \\
&\overset{(b)}{\rightarrow} 0,
\end{aligned}
$$

where (a) is true because from (C.9), $\mathrm{E}\left[L_\delta(Y|\mathbf{z})\right] \geq abc2^{-T\delta_4}$ and $\delta_4 = \delta_1 + \delta_3$. The convergence in (b) happens as $T \rightarrow \infty$, because when $\tilde{R} > I(X;Z) + 2\delta_1 +$

$\delta_2 + 4\delta_3$, the first and third denominator above grow exponentially with T. In addition, when δ satisfies the assumption $\delta < \frac{H(Y|X)}{6H(X)+2H(Y|Z)}$, the exponent of the second denominator is also positive. Hence, all three terms go to zero exponentially fast with T. This proves the hardening of the list size for $\tilde{R} < I(X;Y,Z) - \delta_1$.

Assume now that $\tilde{R} > I(X;Y,Z) + \delta_1$. From (C.10), we get

$$\mathrm{E}\left[L_\delta(Y|\mathbf{z})\right] \doteq |\mathcal{T}_\delta(Y|\mathbf{z})|,$$

which means, by the definition of relative equality, that

$$\frac{\mathrm{E}\left[L_\delta(Y|\mathbf{z})\right]}{|\mathcal{T}_\delta(Y|\mathbf{z})|} \to 1$$

as T grows large. On the other hand, from (C.4),

$$L_\delta(Y|\mathbf{z}) \leq |\mathcal{T}_\delta(Y|\mathbf{z})|,$$

implying that

$$\begin{aligned}\mathrm{E}\left[L_\delta(Y|\mathbf{z})^2\right] &\leq \mathrm{E}\left[|\mathcal{T}_\delta(Y|\mathbf{z})|^2\right]\\ &= |\mathcal{T}_\delta(Y|\mathbf{z})|^2,\end{aligned}$$

because the expectation is over all codebooks. Hence, the variance of the list size goes to zero with T:

$$\begin{aligned}\frac{\mathrm{Var}\left(L_\delta(Y|\mathbf{z})\right)}{|\mathcal{T}_\delta(Y|\mathbf{z})|^2} &= \frac{\mathrm{E}\left[L_\delta(Y|\mathbf{z})^2\right] - \mathrm{E}\left[L_\delta(Y|\mathbf{z})\right]^2}{|\mathcal{T}_\delta(Y|\mathbf{z})|^2}\\ &\leq \frac{|\mathcal{T}_\delta(Y|\mathbf{z})|^2 - \mathrm{E}\left[L_\delta(Y|\mathbf{z})\right]^2}{|\mathcal{T}_\delta(Y|\mathbf{z})|^2}\\ &= 1 - \left(\frac{\mathrm{E}\left[L_\delta(Y|\mathbf{z})\right]}{|\mathcal{T}_\delta(Y|\mathbf{z})|}\right)^2\\ &\to 1 - 1 = 0.\end{aligned}$$

This concludes the proof of the size of $\mathcal{L}_\delta(Y|\mathbf{z})$.

C.1.3 Proof of Lemma 5.4

We consider the following random experiment: $\mathcal{C}$ is fixed, and a uniformly chosen codeword from $\mathcal{C}$ is transmitted over the channel. Let $\mathbf{X}$ be that random codeword. From Lemma F.2, it follows that the received sequences $(\mathbf{Y}, \mathbf{Z})$ are highly likely to be jointly typical with $\mathbf{X}$, and therefore, $\mathbf{Y} \in \mathcal{L}_\delta(Y)$ and $\mathbf{Z} \in \mathcal{L}_\delta(Z)$ w.h.p.. Let $\mathbf{z}$ be a particular sequence in $\mathcal{L}_\delta(Z)$. Given that $\mathbf{Z} = \mathbf{z}$, we also have that $\mathbf{Y} \in \mathcal{L}_\delta(Y|\mathbf{z})$ w.h.p.. We are looking for the probability $\mathbf{P}(\mathbf{Y} = \mathbf{y}|\mathbf{Z} = \mathbf{z})$, for any given $\mathbf{y} \in \mathcal{L}_\delta(Y|\mathbf{z})$.

We start by computing $\mathbf{P}(\mathbf{Z}=\mathbf{z})$ for a fixed $\mathbf{z} \in \mathcal{L}_\delta(Z)$.

$$\begin{aligned}
\mathbf{P}(\mathbf{Z}=\mathbf{z}) &= \sum_{\mathbf{x}\in\mathcal{C}} \mathbf{P}(\mathbf{X}=\mathbf{x})\mathbf{P}(\mathbf{Z}=\mathbf{z}|\mathbf{X}=\mathbf{x}) \\
&\overset{\delta_5}{=} \sum_{\mathbf{x}\in\mathcal{C}} 2^{-T\tilde{R}} \begin{cases} 2^{-TH(Z|X)} & \text{if } \mathbf{z}\in\mathcal{T}_\delta(Z|\mathbf{x}) \\ 0 & \text{otherwise} \end{cases} \\
&= \sum_{\mathbf{x}\in\mathcal{C}\cap\mathcal{T}_\delta(X|\mathbf{z})} 2^{-T\tilde{R}}2^{-TH(Z|X)} \\
&= L_\delta(X|\mathbf{z})2^{-T(\tilde{R}+H(Z|X))} \\
&\left\{ \begin{array}{ll} = 1 & \text{if } \tilde{R} < I(X;Z)-\delta_1 \\ \overset{\delta_1}{=} 2^{T(\tilde{R}-I(X;Z))} & \text{if } I(X;Z)+\delta_1 < \tilde{R} < H(X) \end{array} \right\} 2^{-T(\tilde{R}+H(Z|X))} \\
&= \begin{cases} 2^{-T(\tilde{R}+H(Z|X))} & \text{if } \tilde{R} < I(X;Z)-\delta_1 \\ 2^{-TH(Z)} & \text{if } I(X;Z)+\delta_1 < \tilde{R} < H(X) \end{cases}
\end{aligned}$$

for almost every possible code $\mathcal{C}$, where we defined $\delta_5 \triangleq \delta H(Z|X)$.

Using this, we compute $\mathbf{P}(\mathbf{X}=\mathbf{x}|\mathbf{Z}=\mathbf{z})$ for given $\mathbf{x}\in\mathcal{C}$ and $\mathbf{z}\in\mathcal{L}_\delta(Z)$. When $(\mathbf{x},\mathbf{z})$ are not jointly typical, then this probability is 0. Otherwise,

$$\begin{aligned}
\mathbf{P}(\mathbf{X}=\mathbf{x}|\mathbf{Z}=\mathbf{z}) &= \frac{\mathbf{P}(\mathbf{X}=\mathbf{x})\ \mathbf{P}(\mathbf{Z}=\mathbf{z}|\mathbf{X}=\mathbf{x})}{\mathbf{P}(\mathbf{Z}=\mathbf{z})} \\
&\overset{\delta_5}{=} \frac{2^{-T\tilde{R}}\ 2^{-TH(Z|X)}}{\mathbf{P}(\mathbf{Z}=\mathbf{z})} \\
&\left\{ \begin{array}{ll} \overset{\delta_5}{=} 1 & \text{if } \tilde{R} < I(X;Z)-\delta_1 \\ \overset{\delta_1+\delta_5}{=} 2^{-T(\tilde{R}-I(X;Z))} & \text{if } I(X;Z)+\delta_1 < \tilde{R}, \end{array} \right.
\end{aligned} \tag{C.16}$$

The above derivations show that the distribution of $\mathbf{Z}$ in $\mathcal{L}_\delta(Z)$, and the conditional distribution of $\mathbf{X}$ in $\mathcal{L}_\delta(X|\mathbf{z})$ given $\mathbf{Z}=\mathbf{z}$ can be made arbitrarily close to uniform by choosing a small δ.

Assume that $\tilde{R} < I(X;Y,Z)-\delta_1$. Then, by Lemma 5.3, $\mathcal{L}_\delta(Y|\mathbf{z})$ is δ_4-close to the disjoint union

$$\cup_{\mathbf{x}\in\mathcal{L}_\delta(X|\mathbf{z})}\mathcal{T}_\delta(Y|\mathbf{x},\mathbf{z}).$$

One can show that for a given $(\mathbf{x},\mathbf{z})$, the distribution of $\mathbf{Y}$ can be made arbitrarily close to uniform on $\mathcal{T}_\delta(Y|\mathbf{x},\mathbf{z})$. From the above derivation, it also follows that given $\mathbf{Z}=\mathbf{z}$, $\mathbf{X}$ can be made close to uniformly distributed in $\mathcal{L}_\delta(X|\mathbf{z})$. Hence, $\mathbf{Y}$ can be made close to uniformly distributed in $\mathcal{L}_\delta(Y|\mathbf{z})$.

Now, assume that $\tilde{R} > I(X;Y,Z)+\delta_1$. For a fixed jointly typical $(\mathbf{x},\mathbf{y},\mathbf{z})$, we compute

$$\begin{aligned}\mathbf{P}(\mathbf{Y}=\mathbf{y}|\mathbf{X}=\mathbf{x},\mathbf{Z}=\mathbf{z}) &= \frac{\mathbf{P}(\mathbf{Y}=\mathbf{y},\mathbf{Z}=\mathbf{z}|\mathbf{X}=\mathbf{x})}{\mathbf{P}(\mathbf{Z}=\mathbf{z}|\mathbf{X}=\mathbf{x})}\\ &\overset{\delta_5+\delta_6}{=} \frac{2^{-TH(Y,Z|X)}}{2^{-TH(Z|X)}}\\ &= 2^{T(-H(Z|X)-H(Y|X,Z)+H(Z|X))}\\ &= 2^{-TH(Y|X,Z)}, \qquad \text{(C.17)}\end{aligned}$$

where $\delta_6 = \delta H(Y,Z|X)$. Then, for jointly typical $(\mathbf{y},\mathbf{z})$, we can compute

$$\begin{aligned}\mathbf{P}(\mathbf{Y}=\mathbf{y}|\mathbf{Z}=\mathbf{z}) &= \sum_{\mathbf{x}\in\mathcal{C}} \mathbf{P}(\mathbf{X}=\mathbf{x}|\mathbf{Z}=\mathbf{z})\ \mathbf{P}(\mathbf{Y}=\mathbf{y}|\mathbf{X}=\mathbf{x},\mathbf{Z}=\mathbf{z})\\ &\overset{\delta_1+2\delta_5}{=} \sum_{\mathbf{x}\in\mathcal{L}_\delta(X|\mathbf{z})} 2^{-T(\tilde{R}-I(X;Z))}\ \mathbf{P}(\mathbf{Y}=\mathbf{y}|\mathbf{X}=\mathbf{x},\mathbf{Z}=\mathbf{z})\\ &\overset{\delta_5+\delta_6}{=} \sum_{\mathbf{x}\in\mathcal{L}_\delta(X|\mathbf{z})} 2^{-T(\tilde{R}-I(X;Z))} \begin{cases} 0 & \text{if } \mathbf{y}\notin\mathcal{T}_\delta(Y|\mathbf{x})\\ 2^{-TH(Y|X,Z)} & \text{if } \mathbf{y}\in\mathcal{T}_\delta(Y|\mathbf{x})\end{cases}\\ &= \sum_{\mathbf{x}\in\mathcal{L}_\delta(X|\mathbf{y},\mathbf{z})} 2^{-T(\tilde{R}-I(X;Z))}\ 2^{-TH(Y|X,Z)}\\ &= \big|\mathcal{L}_\delta(X|\mathbf{y},\mathbf{z})\big| 2^{-T(\tilde{R}-I(X;Z))}\ 2^{-TH(Y|X,Z)}\\ &\overset{\delta_1}{=} 2^{T(\tilde{R}-I(X;Y,Z))}\ 2^{-T(\tilde{R}-I(X;Z)+H(Y|X,Z))}\\ &= 2^{T(-I(X;Y|Z)-H(Y|X,Z))}\\ &= 2^{-TH(Y|Z)}, \qquad \text{(C.18)}\end{aligned}$$

where in the $(\delta_1+2\delta_5)$-equality we have used (C.16) and the fact that $\tilde{R} > I(X;Y,Z)+\delta_1 \geq I(X;Z)+\delta_1$, in the $(\delta_5+\delta_6)$-equality, we used (C.17), and in the δ_1-equality, we have used Lemma 5.2 with (Y,Z) playing the role of Z. From (C.18), it follows that $\mathbf{P}(\mathbf{Y}=\mathbf{y}|\mathbf{Z}=\mathbf{z})$ is $\tilde{\delta}$-close to $2^{-TH(Y|Z)}$, where $\tilde{\delta} = 2\delta_1 + 3\delta_5 + \delta_6$ is a scaled version of δ. This concludes the proof.

C.2 Tightness of the Union Bound for Independent Events

Let $\{\mathcal{E}_i\}_{i=1,\dots,n}$ be a collection of events. The well-known union bound states that

$$\mathbf{P}\Big(\cup_{i=1}^n \mathcal{E}_i\Big) \leq \min\Big\{1, \sum_{i=1}^n \mathbf{P}(\mathcal{E}_i)\Big\}.$$

When the events $\{\mathcal{E}_i\}_{i=1,\ldots,n}$ are independent, then we can also derive the following lower bound on the probability of the union:

$$\begin{aligned}
\mathbf{P}\Big(\cup_{i=1}^n \mathcal{E}_i \Big) &= 1 - \mathbf{P}\Big(\cap_{i=1}^n \mathcal{E}_i^c \Big) \\
&= 1 - \prod_{i=1}^n (1 - \mathbf{P}(\mathcal{E}_i)) \\
&\geq 1 - \prod_{i=1}^n \exp\big(- \mathbf{P}(\mathcal{E}_i)\big) \\
&= 1 - \exp\Big(- \sum_{i=1}^n \mathbf{P}(\mathcal{E}_i)\Big). \qquad \text{(C.19)}
\end{aligned}$$

As $\sum_{i=1}^n \mathbf{P}(\mathcal{E}_i)$ grows large, (C.19) goes to 1, and hence the fraction of the two bounds goes to 1. This implies that

$$\mathbf{P}\Big(\cup_{i=1}^n \mathcal{E}_i \Big) \doteq 1,$$

where "$\doteq$" denotes relative equality.

As x goes to 0, the slope of $1 - \exp(-x)$ goes to 1, and hence, by the rule of Bernoulli-De l'Hôpital, $\frac{1-\exp(-x)}{x} \to 1$, which implies that

$$\mathbf{P}\Big(\cup_{i=1}^n \mathcal{E}_i \Big) \doteq \sum_{i=1}^n \mathbf{P}(\mathcal{E}_i)$$

as $\sum_{i=1}^n \mathbf{P}(\mathcal{E}_i)$ goes to 0.

Appendix for Chapter 6 D

D.1 Proof of Proposition 6.1

Let $p_Q\, p_{X_1|Q}\, p_{X_2|Q}\, p_{U_1|X_1,Q}$ be given, and consider the random code construction outlined in Section 6.4.1. The additional constraint (6.30) makes sure that in addition to the error events listed in Section 6.4.1, the expected probability of the following event goes to zero with T: $(\mathbf{u}_1(i), \mathbf{x}_2(k), \mathbf{y}_2, \mathbf{q}) \in \mathcal{T}_\delta(U_1, X_2, Y_2, Q)$ for

- $i \neq 1$ and $k = 1$,

where we assumed that $(i, j, k) = (1, 1, 1)$ was transmitted. Hence, if we define $p_e \triangleq \mathbf{P}(i \text{ decoded wrongly at Receiver 2})$, we get

$$\mathrm{E}\left[p_e\right] \leq \epsilon_0,$$

where $\epsilon_0 > 0$ can be chosen arbitrarily small. Recall that the expectation is over the random choice of the codebooks. From Markov's inequality, we get

$$\begin{aligned}\mathbf{P}(p_e \geq \sqrt{\epsilon_0}) &\leq \frac{\mathrm{E}\left[p_e\right]}{\sqrt{\epsilon_0}} \\ &\leq \frac{\epsilon_0}{\sqrt{\epsilon_0}} \leq \sqrt{\epsilon_0}.\end{aligned}$$

Hence, by choosing ϵ_0 arbitrarily small, the randomly chosen code has, with arbitrarily high probability, an arbitrarily small probability of decoding i wrongly at Receiver 2. Thus, with high probability, the particular code whose existence was shown in Section 6.4.1 also has this property. Let $W_1 = (I, J)$ be the message that is transmitted by Transmitter 1. Since $\mathbf{X}_1$ is a function of (I, J), $\mathbf{X}_2$ is a function of K, and since Receiver 2 can with high probability decode

(I, K) correctly, Fano's inequality implies that

$$H(\mathbf{X}_1|J, \mathbf{Y}_2, \mathbf{X}_2, \mathbf{Q}) \leq T\epsilon_2 \tag{D.1}$$

and

$$H(\mathbf{X}_1, \mathbf{X}_2|J, \mathbf{Y}_2, \mathbf{Q}) \leq T\epsilon_3, \tag{D.2}$$

where ϵ_2 and ϵ_3 go to zero as ϵ_0 goes to zero and T grows. We analyze the equivocation at Receiver 2 about the message J:

$$\begin{aligned}
H(J|\mathbf{Y}_2, \mathbf{Q}) &\geq H(J|\mathbf{Y}_2, \mathbf{X}_2, \mathbf{Q}) \\
&\geq I(J; \mathbf{X}_1|\mathbf{Y}_2, \mathbf{X}_2, \mathbf{Q}) \\
&= H(\mathbf{X}_1|\mathbf{Y}_2, \mathbf{X}_2, \mathbf{Q}) - H(\mathbf{X}_1|J, \mathbf{Y}_2, \mathbf{X}_2, \mathbf{Q}) \\
&\overset{(a)}{\geq} H(\mathbf{X}_1|\mathbf{Y}_2, \mathbf{X}_2, \mathbf{Q}) - T\epsilon_2 \\
&= H(\mathbf{X}_1|\mathbf{X}_2, \mathbf{Q}) - I(\mathbf{X}_1; \mathbf{Y}_2|\mathbf{X}_2, \mathbf{Q}) - T\epsilon_2 \\
&\overset{(b)}{\geq} TR_1 - T\epsilon_4 - I(\mathbf{X}_1; \mathbf{Y}_2|\mathbf{X}_2, \mathbf{Q}) - T\epsilon_2 \\
&\overset{(c)}{\geq} T\bigl(R_1 - I(X_1; Y_2|X_2, Q) - \epsilon_2 - \epsilon_4\bigr).
\end{aligned} \tag{D.3}$$

In the above derivation, we used (D.1) in (a), and we obtained (b) because $H(\mathbf{X}_1|\mathbf{X}_2, \mathbf{Q}) = H(\mathbf{X}_1|\mathbf{Q}) \geq H(\mathbf{X}_1|\mathbf{Y}_1, \mathbf{Q}) \geq H(W) - \epsilon_4$ for some arbitrarily small ϵ_4, where the last step follows from Fano's inequality. Finally, (c) is obtained from a local upper bound much like the one used in equation (B.17) of Appendix B.

We derive a second bound on the equivocation:

$$\begin{aligned}
H(J|\mathbf{Y}_2, \mathbf{Q}) &\geq I(J; \mathbf{X}_1, \mathbf{X}_2|\mathbf{Y}_2, \mathbf{Q}) \\
&= H(\mathbf{X}_1, \mathbf{X}_2|\mathbf{Y}_2, \mathbf{Q}) - H(\mathbf{X}_1, \mathbf{X}_2|J, \mathbf{Y}_2, \mathbf{Q}) \\
&\geq H(\mathbf{X}_1, \mathbf{X}_2|\mathbf{Q}) - I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}_2|\mathbf{Q}) - T\epsilon_3 \\
&\geq T(R_1 + R_2 - I(X_1, X_2; Y_2|Q) - \epsilon_3 - \epsilon_5),
\end{aligned} \tag{D.4}$$

where all the steps are analogous to (D.3) and ϵ_5 is an arbitrarily small constant. Combining (D.3) and (D.4), we obtain

$$\begin{aligned}
\frac{1}{T}H(J|\mathbf{Y}_2, \mathbf{Q}) \geq R_1 &- \min\{I(X_1; Y_2|X_2, Q), I(X_1, X_2; Y_2|Q) - R_2\} \\
&- \max\{\epsilon_2 + \epsilon_4, \epsilon_3 + \epsilon_5\} \\
= R_1 &- B - \epsilon_1 - \max\{\epsilon_2 + \epsilon_4, \epsilon_3 + \epsilon_5\},
\end{aligned}$$

where we used (6.30) in the last step. This proves the claim of Proposition 6.1.

D.2 Proof of Theorem 6.4

Assume that $(R_1, R_2, R_e) \in \mathcal{R}_{e,i}(Q, X_1, X_2, U_1)$ for a given joint distribution p. First, we choose an auxiliary rate B such that the inequalities (6.14) through

(6.22) given in the proof of Theorem 6.1 are all satisfied. It follows that the expected probability of all the error events described in that proof can be made arbitrarily small. This remains true even if we split the index j into a non-secret part j_n of rate A and a secret part j_s of rate $R_1 - B - A$. The code construction is the same as before, with $j = (j_s, j_n)$. We fix the rate of the non-secret message j_n to be

$$A = \min\big\{I(X_1; Y_2|U_1, X_2, Q), I(X_1, X_2; Y_2|U_1, Q) - R_2\big\} - \epsilon_1,$$

where $\epsilon_1 > 0$ can be chosen arbitrarily small. Let $J = (J_s, J_n)$, I and K be the uniformly distributed random messages. The rates of the indices J_n and K satisfy

$$\begin{aligned} A &< I(X_1; Y_2|X_2, U_1, Q) \\ R_2 &< I(X_2; Y_2|U_1, Q) \\ A + R_2 &< I(X_1, X_2; Y_2|U_1, Q). \end{aligned}$$

It follows from standard arguments that if we assume that a genie makes the secret message J_s and the sequence $\mathbf{U}_1$ available to Receiver 2, then the expected probability (over all randomly generated codes) of wrongly decoding (J_n, K) at the genie-aided Receiver 2 can be made arbitrarily small. Hence, there exists a code for which all the error events listed in the proof of Theorem 6.1 have small probability, and in addition, from Fano's inequality,

$$H(\mathbf{X}_1|J_s, \mathbf{U}_1, \mathbf{Y}_2, \mathbf{X}_2, \mathbf{Q}) \leq T\epsilon_2$$

and

$$H(\mathbf{X}_1, \mathbf{X}_2|J_s, \mathbf{U}_1, \mathbf{Y}_2, \mathbf{Q}) \leq T\epsilon_3,$$

where ϵ_2 and ϵ_3 go to zero as ϵ_1 decreases and T increases. Using these two inequalities, we obtain bounds on the equivocation of the code:

$$\begin{aligned} H(J_s|\mathbf{Y}_2, \mathbf{Q}) &\geq H(J_s|\mathbf{U}_1, \mathbf{Y}_2, \mathbf{X}_2, \mathbf{Q}) \\ &\geq I(J_s; \mathbf{X}_1|\mathbf{U}_1, \mathbf{Y}_2, \mathbf{X}_2, \mathbf{Q}) \\ &= H(\mathbf{X}_1|\mathbf{U}_1, \mathbf{Y}_2, \mathbf{X}_2, \mathbf{Q}) - H(\mathbf{X}_1|J_s, \mathbf{U}_1, \mathbf{Y}_2, \mathbf{X}_2, \mathbf{Q}) \\ &\geq H(\mathbf{X}_1|\mathbf{U}_1, \mathbf{Y}_2, \mathbf{X}_2, \mathbf{Q}) - T\epsilon_2 \\ &= H(\mathbf{X}_1|\mathbf{U}_1, \mathbf{X}_2, \mathbf{Q}) - I(\mathbf{X}_1; \mathbf{Y}_2|\mathbf{U}_1, \mathbf{X}_2, \mathbf{Q}) - T\epsilon_2 \\ &\geq T(R_1 - B) - T\epsilon_4 - I(\mathbf{X}_1; \mathbf{Y}_2|\mathbf{U}_1, \mathbf{X}_2, \mathbf{Q}) - T\epsilon_2 \\ &\geq T\big(R_1 - B - I(X_1; Y_2|U_1, X_2, Q) - \epsilon_2 - \epsilon_4\big), \end{aligned}$$

and

$$\begin{aligned} H(J_s|\mathbf{Y}_2, \mathbf{Q}) &\geq H(J_s|\mathbf{U}_1, \mathbf{Y}_2, \mathbf{Q}) \\ &\geq I(J_s; \mathbf{X}_1, \mathbf{X}_2|\mathbf{U}_1, \mathbf{Y}_2, \mathbf{Q}) \\ &= H(\mathbf{X}_1, \mathbf{X}_2|\mathbf{U}_1, \mathbf{Y}_2, \mathbf{Q}) - H(\mathbf{X}_1, \mathbf{X}_2|J_s, \mathbf{U}_1, \mathbf{Y}_2, \mathbf{Q}) \\ &\geq H(\mathbf{X}_1, \mathbf{X}_2|\mathbf{U}_1, \mathbf{Q}) - I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}_2|\mathbf{U}_1, \mathbf{Q}) - T\epsilon_3 \\ &\geq T(R_1 - B + R_2 - I(X_1, X_2; Y_2|U_1, Q) - \epsilon_3 - \epsilon_5), \end{aligned}$$

where all the steps are justified analogously to (D.3) and (D.4) in the previous section of this appendix. It follows that for the given code,

$$\begin{aligned}\frac{1}{T}H(J_s|\mathbf{Y}_2,\mathbf{Q}) &\geq R_1 - B \\ &\quad - \min\big\{I(X_1;Y_2|U_1,X_2,Q), I(X_1,X_2;Y_2|U_1,Q) - R_2\big\} - \epsilon_6,\end{aligned}$$

where $\epsilon_6 = \max\{\epsilon_2 + \epsilon_4, \epsilon_3 + \epsilon_5\}$ can be made arbitrarily small by choosing ϵ_1 small and T large. Therefore, as long as

$$R_e \leq R_1 - B - \min\big\{I(X_1;Y_2|U_1,X_2,Q), I(X_1,X_2;Y_2|U_1,Q) - R_2\big\}, \quad \text{(D.5)}$$

the equivocation of this code is at least $R_e - \epsilon_6$. Hence, $(R_1, R_2, R_e) \in \mathcal{R}_e$ if they satisfy (6.14) through (6.22) and (D.5) for some auxiliary rate B. Note that (D.5) is an upper bound on B. Now, we combine each lower bound on B with each upper bound on B exactly as it was done in Section 6.4.1. We find that an auxiliary rate B that satisfies (6.14) through (6.22) and (D.5) exists if and only if (R_1, R_2, R_e) satisfy the seven conditions given in Definition 6.7. This concludes the proof.

E From Weak to Strong Secrecy

In this section, we outline how a code that provides weak secrecy at rate R can be turned into a code that provides strong secrecy at the same rate R using extractors. This technique was introduced by Maurer and Wolf in [29] and the contents of this appendix are basically a repetition of Section 3.3 in that publication. There is a small difference, however. In [29], the authors showed the existence of a new code that uses the given (weak) code several times to produce a strongly secret key that is shared between S and D. In our setup, the shared secret between S and D should not be created by the code, but it should be equal to some message W which is observed by S *before* applying the code. To overcome this problem, we use linear extractors that are in some sense invertible. Thanks to this, the strongly secret message can be taken to be the output of a source, the inverse of an extractor is applied to it, and then the procedure described in [29] is used to establish a reliable and strongly secret estimate of the message at the destination.

E.1 Auxiliary Lemmas

We start by defining what extractors are and stating a theorem that guarantees their existence. We start by two auxiliary definitions.

Definition E.1 *Let X be a discrete random variable, taking values in an alphabet $\mathcal{X}$. The* ***min-entropy*** *of X is defined as*

$$H_\infty(X) \triangleq -\log \max_{x\in\mathcal{X}} p_X(x).$$

Definition E.2 *Let X and Y be discrete random variables, taking values in the same alphabet $\mathcal{X}$. The* ***variational distance*** *(or statistical distance) between*

X and Y is defined as

$$d(X,Y) \triangleq \frac{1}{2}\sum_{x\in\mathcal{X}} |p_X(x) - p_Y(x)|.$$

Definition E.3 *A function $e : \{0,1\}^N \times \{0,1\}^d \to \{0,1\}^r$ is called a seeded (k', ϵ')-**extractor** if for any random variable X with range $\mathcal{X} \subseteq \{0,1\}^N$ and min-entropy $H_\infty(X) \geq k'$, the variational distance $d([V, e(X,V)], U_{d+r})$ is at most ϵ' when V is independent of X and uniformly distributed in $\{0,1\}^d$. Here, U_n denotes a random variable uniformly distributed on $\{0,1\}^n$.*

An important property is that extractors can be applied to a random object X without knowing its distribution. The only knowledge that is required about X is a lower bound on its min-entropy.

The following lemma is a special case of Theorem 1 in [37] which guarantees the existence of linear extractors with parameters chosen in the same way as in Lemma 8 of [29] in order to suit the strong secrecy proof.

Lemma E.1 *For every choice of the parameters N, $0 < k' < N$, $\epsilon' > 0$ and $\tilde{\epsilon} > 0$, there are explicit linear (k', ϵ')-extractors $e : \{0,1\}^N \times \{0,1\}^d \to \{0,1\}^r$, where $d = \mathcal{O}\left((\log \frac{N}{\epsilon'})^2\right)$ and $r = k' - \tilde{\epsilon}$.*

Proof: The statement of the lemma follows from Theorem 1 in [37] by setting $k = k'$ and $m = k - \tilde{\epsilon}$, which yields the desired value of r. To determine the order of d, we compute the auxiliary parameter γ to be $\gamma = \frac{2k-\tilde{\epsilon}-1}{k-\tilde{\epsilon}-1}$. Hence, as N grows large, γ converges to 2. It follows that $d = \mathcal{O}\left((\log \frac{N}{\epsilon'})^2\right)$. The linearity of the extractors can be shown by inspecting the explicit construction provided in the proof of Theorem 1 in [37]. This observation was also made in [7]. □

The following lemma is equivalent to Lemma 9 in [29] and follows from Lemma E.1.

Lemma E.2 *Let $k', \delta_1, \delta_2 > 0$ be constants. Then there exists, for a sufficiently large N, a linear extractor $e : \{0,1\}^N \times \{0,1\}^d \to \{0,1\}^r$, where $d \leq \delta_1 N$ and $r \geq k' - \delta_2 N$, such that for all random variables X with $\mathcal{X} \subseteq \{0,1\}^N$ and $H_\infty(X) \geq k'$, we have*

$$H\big(e(X,V)|V\big) \geq r - 2^{-N^{1/2-o(1)}}.$$

The essential point of Lemma E.2 is that the function e does not depend on the distribution of X, and the output of e is essentially independent of the seed V and uniformly distributed in $\{0,1\}^r$.

Proof: The statement of the lemma follows from Lemma E.1 by setting $\epsilon'(N) \triangleq 2^{-\sqrt{N}/\log N}$. This yields that for a large enough N, there exists a (k', ϵ')-extractor e, mapping $\{0,1\}^N \times \{0,1\}^d$ to $\{0,1\}^r$. Note that

$d = \mathcal{O}\left((\log \frac{N}{\epsilon'})^2\right)$, where

$$\begin{aligned}(\log \frac{N}{\epsilon'(N)})^2 &= (\log N + \frac{\sqrt{N}}{\log N})^2 \\ &= (\log N)^2 + 2\sqrt{N} + \frac{N}{(\log N)^2}.\end{aligned}$$

Hence, for large enough N, $\frac{d}{N}$ can be made arbitrarily small, and $d \leq \delta_1 N$ is satisfied. It is further clear that $r \geq k' - \tilde{\epsilon} \geq k' - \delta_2 N$ for large enough N. The rest of the proof is the same as for Lemma 9 in [29]. □

The following lemma relates the conditional min-entropy of a sequence of i.i.d. random variables to the conditional Shannon entropy of one component of the sequence. The lemma can be found in [29] (Lemma 6).

Lemma E.3 *Let μ_{XZ} be the joint probability measure of two random variables X and Z, where X is discrete and Z is discrete or Gaussian. Let $0 < \delta \leq \frac{1}{2}$ be fixed and let N be an integer. Let $\mathbf{X}$ and $\mathbf{Z}$ be sequences of length N, with distribution $\mu^N_{X,Y}(\cdot)$. Let $\mathcal{F}(\delta)$ be the event that $(\mathbf{X},\mathbf{Z}) \in A_\delta(X,Z)$, where $A_\delta(X,Z)$ is the set of all weakly typical sequences (see e.g. [9]). Then, we have*

$$\mathbf{P}\big(\overline{\mathcal{F}(\delta)}\big) \to 0$$

as $N \to \infty$, and

$$H_\infty(\mathbf{X}|\mathbf{Z}=\mathbf{z},\mathcal{F}(\delta)) \geq N(1-\delta)H(X|Z).$$

The statement of the lemma provided here differs slightly from Lemma 6 in [29] because the min-entropy is easier to deal with than the Rényi-entropy. In addition, this version of the lemma is also valid when Z is Gaussian (as opposed to discrete).

Proof: The fact that $(\mathbf{X},\mathbf{Z})$ are weakly jointly typical with high probability follows from Theorem 9.2.2 in [9]. The lower bound on the min-entropy is obtained as follows. The definition of weak typicality implies that for all $\mathbf{x} \in \mathcal{X}^N, \mathbf{z} \in \mathcal{Z}^N$ such that $(\mathbf{x},\mathbf{z})$ are jointly typical, we have

$$p_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}) \leq 2^{-N(1-\delta)H(X|Z)}.$$

Hence, we obtain

$$\begin{aligned}H_\infty(\mathbf{X}|\mathbf{Z}=\mathbf{z},\mathcal{F}(\delta)) &= -\log \max_{\mathbf{x}\in A^{(N)}_\delta(X|\mathbf{z})} p_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}) \\ &\geq -\log 2^{-N(1-\delta)H(X|Z)} \\ &= N(1-\delta)H(X|Z).\end{aligned}$$

□

We state one last auxiliary lemma before stating and proving the main lemma of this appendix.

Lemma E.4 *Let X and Q be discrete random variables, and let $s > 0$. Then the set of values $q \in \mathcal{Q}$ for which $H_\infty(X) - H_\infty(X|Q = q) \leq \log|\mathcal{Q}| + s$ has probability at least $1 - 2^{-s}$.*

Proof: We start with the following inequality:

$$\begin{aligned}
\mathrm{E}_Q\left[2^{-H_\infty(X|Q=Q)}\right] &= \sum_{q\in\mathcal{Q}} p_Q(q) 2^{-H_\infty(X|Q=q)} \\
&= \sum_{q\in\mathcal{Q}} p_Q(q) \max_{x\in\mathcal{X}} p_{X|Q}(x|q) \\
&= \sum_{q\in\mathcal{Q}} \max_{x\in\mathcal{X}} \underbrace{p_{X,Q}(x,q)}_{\leq p_X(x)} \\
&\leq |\mathcal{Q}| \max_{x\in\mathcal{X}} p_X(x) \\
&= |\mathcal{Q}| 2^{-H_\infty(X)}. \qquad \text{(E.1)}
\end{aligned}$$

Then, using Markov's inequality, we obtain

$$\begin{aligned}
&\mathbf{P}_Q(-H_\infty(X|Q = Q) \geq \log|\mathcal{Q}| + s - H_\infty(X)) \\
&\quad = \mathbf{P}_Q(2^{-H_\infty(X|Q=Q)} \geq 2^s|\mathcal{Q}|2^{-H_\infty(X)}) \\
&\quad \leq \frac{\mathrm{E}_Q\left[2^{-H_\infty(X|Q=Q)}\right]}{2^s|\mathcal{Q}|2^{-H_\infty(X)}} \\
&\quad \leq \frac{|\mathcal{Q}|2^{-H_\infty(X)}}{2^s|\mathcal{Q}|2^{-H_\infty(X)}} \\
&\quad = 2^{-s},
\end{aligned}$$

where we used (E.1) in the last inequality. This concludes the proof. □

E.2 Main Lemma

Lemma E.5 *Let a memoryless wireless network or channel be given, and assume that a rate-equivocation pair $(R, R - \rho)$ is weakly achievable over this network, where $\rho \geq 0$ is an arbitrary constant. Then, the secrecy rate $R - \rho$ is strongly achievable over this network.*

Proof: To show that $R - \rho$ is strongly achievable, we proceed as follows. Let an arbitrary $\bar{\epsilon} > 0$ be given. We will show that there exists a block length $\bar{T}$ and a $(\bar{T}, \bar{\epsilon})$-code of rate at least $R - \rho - \bar{\epsilon}$ and equivocation at least $\frac{1}{\bar{T}}(H(\bar{W}) - \bar{\epsilon})$, where $\bar{W}$ is the secret message.

Since $(R, R - \rho)$ is weakly achievable, we know that for every $\epsilon > 0$ there exists, for a large enough block length T, a (T, ϵ)-code that achieves rate at least $R - \epsilon$ and equivocation at least $R - \rho - \epsilon$. We run this code M times over M subsequent blocks of T time-slots. This takes a total of TM uses of the network channels. Let $W^{(m)}$, $\hat{W}^{(m)}$, $\mathbf{Y}_D^{(m)}$ and $\mathbf{Y}_E^{(m)}$ be, respectively, the message, the message estimate at D, the sequence of length T received at D and at E during block m, for $m \in \{1, \dots, M\}$. Note that E can be any eavesdropper in $\mathcal{E}$. We define

$$\begin{aligned}
\underline{W} &\triangleq (W^{(1)}, \dots, W^{(M)}) \\
\underline{\hat{W}} &\triangleq (\hat{W}^{(1)}, \dots, \hat{W}^{(M)}) \\
\underline{\mathbf{Y}_D} &\triangleq (\mathbf{Y}_D^{(1)}, \dots, \mathbf{Y}_D^{(M)}) \\
\underline{\mathbf{Y}_E} &\triangleq (\mathbf{Y}_E^{(1)}, \dots, \mathbf{Y}_E^{(M)})
\end{aligned}$$

to be the tuples that regroup the quantities of all M blocks. Since we use the same (T, ϵ)-code for each block, we have that for all $m \in \{1, \dots, M\}$,

$$H(W^{(m)}) \geq T(R - \epsilon)$$

and

$$H(W^{(m)}|\mathbf{Y}_E^{(m)}) \geq T(R - \rho - \epsilon)$$

for all $E \in \mathcal{E}$.

Information reconciliation: Note that the error probability of each block m satisfies $\mathbf{P}(W^{(m)} \neq \hat{W}^{(m)}) \leq \epsilon$. It follows that

$$\begin{aligned}
\mathbf{P}(\underline{W} = \underline{\hat{W}}) &\geq (1 - \epsilon)^M \\
&\to 0
\end{aligned}$$

with increasing M. Hence, we have no guarantee that $\underline{\hat{W}}$ equals $\underline{W}$ with high probability. The first phase of the protocol proposed in [29], called "information reconciliation", establishes a new estimate $\underline{\breve{W}}$ at the destination D such that $\mathbf{P}(\underline{\breve{W}} \neq \underline{W}) \leq \epsilon_1$, where ϵ_1 can be made arbitrarily small by choosing M large enough. This is done by using a Slepian-Wolf lossless compression scheme [42] on the i.i.d. sequence $(W^{(m)}, \hat{W}^{(m)})$. From Fano's inequality, $H(W^{(m)}|\hat{W}^{(m)}) \leq \epsilon(RT - \epsilon) + 1$. We know from [42] that it is sufficient to transmit $(1 + \epsilon_2)MH(W^{(1)}|\hat{W}^{(1)}) \leq (1 + \epsilon_2)M(\epsilon(RT - \epsilon) + 1)$ additional bits from S to D in order to produce such a new estimate $\underline{\breve{W}}$, where ϵ_2 can be made arbitrarily small by choosing M large enough. Let $f_{\mathrm{SW}}(\underline{W})$ be this sequence of additional bits. We transmit $f_{\mathrm{SW}}(\underline{W})$ over the network using a reliable source-destination code. Such a code exists for all networks considered in this thesis.

Privacy Amplification: We now describe the second phase of the procedure introduced in [29]. Assume that for all $m \in \{1, \dots, M\}$, $W^{(m)}$ is uniformly distributed in $\{0, 1\}^{T(R-\epsilon)}$, then it follows that $\underline{W}$ is uniformly distributed in $\{0, 1\}^{MT(R-\epsilon)}$. We would like to show that there exists an extractor which acts on the random object $\underline{W}$ and a random seed V and produces a random variable

$\bar{W}$ that is uniformly distributed in $\{0,1\}^r$, where r can be made arbitrarily close to $MT(R-\rho)$. Moreover, the conditional entropy of $\bar{W}$ given the information at any eavesdropper should be arbitrarily close to $H(\bar{W})$, hence making $\bar{W}$ strongly secret. We provide the details hereafter.

Fix any eavesdropper $E \in \mathcal{E}$. We apply Lemma E.3 to the sequence of random variables $(W^{(m)}, \mathbf{Y}_E^{(m)})_{m\in\{1,\dots,M\}}$ and obtain that

$$\begin{aligned} H_\infty(\underline{W}|\underline{\mathbf{Y}_E} = \underline{\mathbf{y}_E}, \mathcal{F}(\delta)) &\geq M\big(H(W^{(1)}|\mathbf{Y}_E^{(1)}) - \delta'\big) \\ &\geq MT(R-\rho-\epsilon) - M\delta', \end{aligned}$$

where $\delta' = \delta H(W^{(1)}|\mathbf{Y}_E^{(1)})$ can be made arbitrarily small by choosing δ arbitrarily. Furthermore, $\mathcal{F}(\delta)$ happens with high probability as M grows large. Fix an arbitrary constant $s > 0$. By Lemma E.4, we have with probability $1 - 2^{-s}$ that

$$\begin{aligned} &H_\infty\big(\underline{W}|\underline{\mathbf{Y}_E} = \underline{\mathbf{y}_E}, \mathcal{F}(\delta), f_{\mathrm{SW}}(\underline{W}) = i\big) \\ &\quad\geq H_\infty(\underline{W}|\underline{\mathbf{Y}_E} = \underline{\mathbf{y}_E}, \mathcal{F}(\delta)) - \log(\#\text{ possible SW-messages}) - s \\ &\quad\geq MT(R-\rho-\epsilon) - M\delta' - (1+\epsilon_2)(\epsilon(RT-\epsilon)+1) - s \\ &\quad= MT\big(R-\rho-\epsilon-\frac{\delta'}{T} - (1+\epsilon_2)(\epsilon(\frac{R}{M} - \frac{\epsilon}{MT}) + \frac{1}{MT}) - \frac{s}{MT}\big) \\ &\quad\geq MT(R-\rho-\epsilon_3), \end{aligned} \tag{E.2}$$

where ϵ_3 can be made arbitrarily small by choosing M and T sufficiently large. This holds for any eavesdropper $E \in \mathcal{E}$ and any realization of the received sequence $\mathbf{Y}_E$ of that eavesdropper.

It follows by Lemma E.2 that there is an explicit linear extractor $e : \{0,1\}^N \times \{0,1\}^d \to \{0,1\}^r$ with $N = MT(R-\epsilon)$, where $d \leq \delta_1 MT(R-\epsilon)$ and $r \geq k'$, where $k' = MT(R-\rho-\epsilon_3)$ as given in (E.2). Let us apply the extractor e to $\underline{W}$ and to an independent uniform seed V and let $\bar{W} = e(\underline{W}, V)$. From (E.2) and Lemma E.2, we obtain that

$$\begin{aligned} &H(\bar{W}|\underline{\mathbf{Y}_E} = \underline{\mathbf{y}_E}, \mathcal{F}(\delta), f_{\mathrm{SW}}(\underline{W}) = i, V) \\ &\qquad\geq r - 2^{-N^{1/2-o(1)}} \end{aligned} \tag{E.3}$$

for all $E \in \mathcal{E}$ and all $\underline{\mathbf{y}_E}$. We send the seed V reliably over the network together with $f_{\mathrm{SW}}(\underline{W})$. To describe V, we need to transmit at most $\delta_1 MT(R-\epsilon)$ additional bits. Let $\hat{V}$ be the estimate of V produced at the destination D. Then D can apply the same extractor e to $(\check{W}, \hat{V})$, which is equal to $(\underline{W}, V)$ with high probability. Hence, D can correctly reproduce $\bar{W}$ with high probability. Since we do not worry about secrecy during the transmission of the seed V, we must assume that all the eavesdroppers have knowledge of V.

However, using (E.3), we find that for every $E \in \mathcal{E}$,

$$\begin{aligned} &H(\bar{W}|\underline{\mathbf{Y}_E}, f_{\mathrm{SW}}(\underline{W}), V) \\ &\quad \geq \mathbf{P}(\mathcal{F}(\delta))\big(r - 2^{-M^{1/2-o(1)}}\big) + \mathbf{P}(\overline{\mathcal{F}(\delta)})0 \\ &\quad \geq (1 - \epsilon_\delta^{(M)})\big(r - 2^{-M^{1/2-o(1)}}\big) \\ &\quad \geq r - \epsilon_5, \end{aligned}$$

where $\epsilon_\delta^{(M)}$ is as defined in Lemma F.1 in Appendix F, but for the random variables $(W^{(1)}, \mathbf{Y}_E^{(1)})$. Since $\epsilon_\delta^{(M)}$ goes to zero with M, the constant ϵ_5 can be made arbitrarily small by the choice of M. In addition,

$$H(\bar{W}) \geq r - 2^{-M^{1/2-o(1)}} \geq r - \epsilon_6,$$

where ϵ_6 can also be made arbitrarily small. This means that $\bar{W}$ is almost uniformly distributed in $\{0,1\}^r$ and $H(\bar{W}|\underline{\mathbf{Y}_E}, f_{\mathrm{SW}}(\underline{W}), V)$ can be made arbitrarily close to $H(\bar{W})$.

Analysis of the number of network uses: We analyze the total number times the above procedure uses the network. The M repetitions of the (T, ϵ)-code require MT network uses. For any network that we consider in this thesis, a certain reliable communication rate between S and D can be guaranteed. This rate R_{SD} is measured in bits per uses of the network. Hence, to transmit $(f_{\mathrm{SW}}(\underline{W}), V)$ reliably from S to D, we require at most

$$\begin{aligned} &\frac{(1+\epsilon_2)M(\epsilon(RT-\epsilon)+1) + \delta_1 MT(R-\epsilon)}{R_{SD}} \\ &= MT\Big(\frac{(1+\epsilon_2)\epsilon R}{R_{SD}} + \frac{(1+\epsilon_2)(1-\epsilon^2)}{TR_{SD}} + \frac{\delta_1(R-\epsilon)}{R_{SD}}\Big) \\ &\leq MT\epsilon_7 \end{aligned}$$

network uses. By choosing M and T sufficiently large, ϵ_7 can be made arbitrarily small. Hence, the *total* number of network uses of the scheme is upper bounded by $(1+\epsilon_7)MT$ and can be made arbitrarily close to MT.

Inverting the extractor at S: In [29] it is shown how the procedure described above, consisting of information reconciliation followed by privacy amplification, can be used to establish a shared, strongly secret key between S and D. To do so, one selects $\underline{W}$ uniformly at random from $\{0,1\}^{MT(R-\epsilon)}$ and applies the above procedure. The result is a key $\bar{W}$ that is shared between S and D and strongly secret with high probability.

For secret communication of a message coming from a given data source, we slightly modify the procedure. Note that the extractor e whose existence is given by Lemma E.2 is linear. It follows that for a fixed seed $V = v$, the preimage of $\bar{w} = e(\cdot, v)$ is of the same size for each possible output $\bar{w} \in \{0,1\}^r$.

Hence, we can "invert" the extractor in the following sense. Let $\bar{W}$ be the binary representation of a source message of length r. The source node S generates a random uniform seed V in $\{0,1\}^d$ and then selects a sequence $\underline{W}$ uniformly at random from the pre-image of $e(\cdot, V)$ for the output $\bar{W}$. Note that since all possible pre-images of $e(\cdot, V)$ have the same size, $\underline{W}$ is uniformly distributed in $\{0,1\}^N$. We then apply the procedure described above to the sequence $\underline{W}$. As a result, the message $\bar{W}$ is reliably transmitted to D. This procedure uses the network $\bar{T} \triangleq MT(1+\epsilon_7)$ times, where ϵ_7 can be made arbitrarily small. The probability of wrongly reproducing $\bar{W}$ at D can also be made arbitrarily small. The communication rate of the procedure is

$$\begin{aligned}\frac{1}{\bar{T}} H(\bar{W}) &= \frac{r}{MT(1+\epsilon_7)}\\ &\geq \frac{R-\rho-\epsilon_3}{1+\epsilon_7}\\ &\geq R-\rho-\epsilon_8,\end{aligned}$$

where ϵ_8 can be made arbitrarily small. For every eavesdropper, the equivocation rate is

$$\begin{aligned}&\frac{1}{\bar{T}} H(\bar{W}|\underline{\mathbf{Y}_E}, f_{\text{SW}}(\underline{W}), V)\\ &\quad\geq \frac{1}{\bar{T}}(r-\epsilon_5)\\ &\quad= \frac{1}{\bar{T}}(H(\bar{W})-\epsilon_5),\end{aligned}$$

where ϵ_5 can be made arbitrarily small. Hence, the procedure described in this proof is a $(\bar{T}, \bar{\epsilon})$-code that strongly achieves secrecy rate $R-\rho$. □

F Robust Typicality

In this thesis, we use a variant of strong typicality that was introduced by Orlitsky and Roche in [35].

Definition F.1 *Let* $\mathbf{x} = (x[1], \ldots, x[T])$ *be a sequence taking values in* $\mathcal{X}^T$*, where* $\mathcal{X}$ *is a finite alphabet. The* ***empirical frequency*** *of* $x \in \mathcal{X}$ *in* $\mathbf{x}$ *is defined as*

$$\nu_{\mathbf{x}}(x) \triangleq \frac{1}{T}|\{t : x[t] = x\}|.$$

Definition F.2 *Let* X *be a random variable taking values in a discrete set* $\mathcal{X}$*. Let* $p_X(\cdot)$ *be the probability mass function of* X*. A sequence* $\mathbf{x} \in \mathcal{X}^T$ *is* ***robustly*** δ***-typical*** *for* $0 < \delta < 1$ *if for all* $x \in \mathcal{X}$*,*

$$|\nu_{\mathbf{x}}(x) - p_X(x)| \leq \delta p_X(x).$$

The set of all δ*-typical sequences is denoted by* $\mathcal{T}_\delta(X)$*. When is clear from the context which random variable we refer to, we use the short form* $\mathcal{T}_\delta$*.*

In this thesis, we often refer to robust δ-typicality simply by "typicality". The main advantages of robust typicality compared to strong typicality (see [11]) are the following. In strong typicality, one needs to state that zero-probability symbols should not appear in a typical sequence. The notion of robust typicality already makes sure that this is the case. An even more useful property of robust typicality is that $(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_\delta(X, Y)$ implies $\mathbf{x} \in \mathcal{T}_\delta(X)$, where both typical sets are defined for the *same* δ. This is not the case for strong typicality.

Definition F.3 *When* X *is a tuple of random variables, Definition F.2 is still valid. If two sequences* $\mathbf{x}$ *and* $\mathbf{y}$ *satisfy* $(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_\delta$*, we say that* $\mathbf{x}$ *and* $\mathbf{y}$ *are* ***jointly typical****. The set of all sequences* $\mathbf{x}$ *that are jointly typical with a fixed sequence* $\mathbf{y}$ *is denoted by* $\mathcal{T}_\delta(X|\mathbf{y})$*.*

Note that if (X, Y) are such that $Y = g(X)$, where $g(\cdot)$ is a deterministic function, then two sequences $(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_\delta(X, Y)$ if and only if $\mathbf{x} \in \mathcal{T}_\delta(X)$ and $\mathbf{y} = (y[1], \ldots, y[T]) = \big(g(x[1]), \ldots, g(x[T])\big)$.

We state a few lemmas that were proved in [35] and that are used in this thesis.

Lemma F.1 *For any given $0 < \delta < 1$, and a given random variable X, we have*

$$(1 - \epsilon_\delta^{(T)}) 2^{T(1-\delta)H(X)} \leq |\mathcal{T}_\delta(X)| \leq 2^{T(1+\delta)H(X)},$$

with $\epsilon_\delta^{(T)} \triangleq 2|\mathcal{S}_X| e^{-T\delta^2 \mu_X/3}$, where we define $\mathcal{S}_X \triangleq \{x \in \mathcal{X} : p_X(x) > 0\}$ to be the support of X and $\mu_X \triangleq \min_{x \in \mathcal{S}_X} p_X(x)$ to be the smallest non-zero value of $p_X(x)$. Furthermore, if $\mathbf{X}$ is a random sequence of length T generated in an i.i.d. manner from $p_X(x)$, then

$$\mathbf{P}(\mathbf{X} \in T_\delta(X)) \geq 1 - \epsilon_\delta^{(T)}.$$

These facts are given in Lemma 19 and Corollary 2 in [35] and their proofs can be found there.

Lemma F.2 *Let random variables (X, Y) be given. For any $0 < \delta_1 < \delta_2 < 1$ and for a fixed $\mathbf{x} \in \mathcal{T}_{\delta_1}(X)$, we have*

$$(1 - \epsilon_{\delta_1,\delta_2}^{(T)}) 2^{T(1-\delta_2)H(Y|X)} \leq |\mathcal{T}_{\delta_2}(Y|\mathbf{x})| \leq 2^{T(1+\delta_2)H(Y|X)},$$

where $\epsilon_{\delta_1,\delta_2}^{(T)} = 2|\mathcal{S}_{X,Y}| e^{-\frac{T}{3}\frac{(\delta_2-\delta_1)^2}{(1+\delta_1)}\mu_{X,Y}}$. Here, $\mathcal{S}_{X,Y}$ is the support of (X, Y) and $\mu_{X,Y}$ is the smallest non-zero value of $p_{XY}(x, y)$. Furthermore, if $\mathbf{Y}$ is a random sequence of length T generated from $\prod_{t=1}^T p_{Y|X}(\cdot|x[t])$, then

$$\mathbf{P}(\mathbf{Y} \in \mathcal{T}_{\delta_2}(Y|\mathbf{x})) \geq 1 - \epsilon_{\delta_1,\delta_2}^{(T)}.$$

These facts are stated and proved in Lemmas 22 and 24 in [35].

Lemma F.3 *For all $(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_\delta(X, Y)$,*

$$2^{-T(1+\delta)H(Y|X)} \leq p_{Y|X}(\mathbf{y}|\mathbf{x}) \leq 2^{-T(1-\delta)H(Y|X)}.$$

This fact is stated and proved in Lemma 20 in [35].

Lemma F.4 *Let random variables (X, Y) be given. For any $0 < \delta_1 < \delta_2 < 1$ and for a fixed $\mathbf{x} \in \mathcal{T}_{\delta_1}(X)$, we have the following. If $\mathbf{Y}'$ is a random sequence of length T generated from $p_Y(\cdot)^T$ independently of the value of $\mathbf{x}$, then*

$$(1 - \epsilon_{\delta_1,\delta_2}^{(T)}) 2^{-T(I(X;Y)+2\delta_2 H(Y))} \leq \mathbf{P}(\mathbf{Y} \in \mathcal{T}_{\delta_2}(Y|\mathbf{x})) \leq 2^{-T(I(X;Y)-2\delta_2 H(Y))},$$

where $\epsilon_{\delta_1,\delta_2}^{(T)}$ is defined in Lemma F.2.

This fact and its proof can be found as Lemma 25 in [35].

Bibliography

[1] 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, Jul. 1993.

[2] A. Avestimehr, S. Diggavi, and D. Tse, "A deterministic approach to wireless relay networks," in *Proc. of the Allerton Conf. on Commun., Control and Computing*, Illinois, USA, Sep. 2007, see: http://licos.epfl.ch/index.php?p=research_projWNC.

[3] ——, "Wireless network information flow," in *Proc. of the Allerton Conf. on Commun., Control and Computing*, Illinois, USA, Sep. 2007, see: http://licos.epfl.ch/index.php?p=research_projWNC.

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

[5] ——, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inform. Theory*, 2009, submitted, http://arxiv.org/abs/0906.5394.

[6] N. Cai and R. Yeung, "Secure network coding," in *Proc. of the IEEE Int. Symposium on Inform. Theory*, Lausanne, Switzerland, Jun. 2002.

[7] M. Cheraghchi, F. Didier, and A. Shokrollahi, "Invertible extractors and wiretap protocols," in *Proc. of the IEEE Int. Symposium on Inform. Theory*, Seoul, Korea, 2009, http://arxiv.org/abs/0901.2120.

[8] H. Chong, M. Motani, H. K. Garg, and H. E. Gamal, "On the Han-Kobayashi region for the interference channel," *IEEE Trans. Inform. Theory*, Aug. 2006, submitted.

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

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

[11] ——, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic Press, 1981.

[12] W. Diffie and M. E. Hellman, "New directions in cryptography," *IEEE Transactions on Information Theory*, vol. 22, no. 6, pp. 644–654, 1976.

[13] R. Etkin, D. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," in *Proc. of the IEEE Inform. Theory Workshop*, Chengdu, China, Oct. 2006.

[14] A. E. Gamal and M. H. Costa, "The capacity region of a class of deterministic interference channels," *IEEE Trans. Inform. Theory*, vol. 28, no. 2, pp. 343–346, Mar. 1982.

[15] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[16] T. Han and K. Kobayashi, "A new achievable region for the interference channel," *IEEE Trans. Inform. Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.

[17] A. Khisti and G. Wornell, "The MIMOME channel," in *Proc. of the Allerton Conf. on Commun., Control and Computing*, Allerton, USA, Oct. 2007, http://arxiv.org/abs/0710.1325.

[18] ——, "Secure transmission with multiple antennas: The MISOME wiretap channel," *IEEE Trans. Inform. Theory*, Aug. 2007, submitted, http://arxiv.org/abs/0708.4219.

[19] A. Khisti, G. Wornell, A. Wiesel, and Y. Eldar, "On the Gaussian MIMO wiretap channel," in *Proc. of the IEEE Int. Symposium on Inform. Theory*, Nice, France, Jun. 2007.

[20] L. Lai and H. E. Gamal, "Cooperative secrecy: The relay-eavesdropper channel," in *Proc. of the IEEE Int. Symposium on Inform. Theory*, Nice, France, Jun. 2007.

[21] ——, "The relayeavesdropper channel: Cooperation for secrecy," *IEEE Trans. Inform. Theory*, vol. 54, no. 9, pp. 4005–4019, Sep. 2008.

[22] Y. Liang, G. Kramer, H. Poor, and S. Shamai, "Compound wire-tap channels," in *Proc. of the Allerton Conf. on Commun., Control and Computing*, Allerton, USA, 2007.

[23] Y. Liang and V. Poor, "Generalized multiple access channels with confidential messages," in *Proc. of the IEEE Int. Symposium on Inform. Theory*, Seattle, USA, Jul. 2006, http://arxiv.org/abs/cs/0605014.

[24] Y. Liang, A. Somekh-Baruch, H. Poor, S. Shamai, and S. Verdu, "Capacity of cognitive interference channels with and without secrecy," *IEEE Trans. Inform. Theory*, vol. 55, no. 2, pp. 604 – 619, Feb. 2009.

[25] R. Liu, I. Marić, R. Yates, and P. Spasojević, "The discrete memoryless multiple access channel with confidential messages," in *Proc. of the IEEE Int. Symposium on Inform. Theory*, Seattle, USA, Jul. 2006, http://arxiv.org/abs/cs/0605005.

[26] T. Liu, V. Prabhakaran, and S. Vishwanath, "The secrecy capacity of a class of parallel gaussian compound wiretap channels," in *Proc. of the IEEE Int. Symposium on Inform. Theory*, Toronto, Canada, Jul. 2008.

[27] T. Liu and S. Shamai, "A note on the secrecy capacity of the multi-antenna wiretap channel," *IEEE Trans. Inform. Theory*, Oct. 2007, submitted, http://arxiv.org/abs/0710.4105.

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

[29] U. Maurer and S. Wolf, "Information-theoretic key agreement: From weak to strong secrecy for free," *EUROCRYPT, LNCS 1807*, pp. 351–368, 2000, Springer.

[30] ——, "Secret key agreement over a non-authenticated channel — Part I: Definitions and bounds," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 822–831, Apr. 2003.

[31] ——, "Secret key agreement over a non-authenticated channel — Part II: The simulatability condition," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 832–838, Apr. 2003.

[32] ——, "Secret key agreement over a non-authenticated channel — Part III: Privacy amplification," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 839–851, Apr. 2003.

[33] F. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," in *Proc. of the IEEE Int. Symposium on Inform. Theory*, Toronto, Canada, Jun. 2008.

[34] Y. Oohama, "Coding for relay channels with confidential messages," in *Proc. of the IEEE Inform. Theory Workshop*, Cairns, Australia, Sep. 2001.

[35] A. Orlitsky and J. Roche, "Coding for computing," *IEEE Trans. Inform. Theory*, vol. 47, no. 3, pp. 903–917, Mar. 2001.

[36] A. Özgür, O. Lévêque, and D. Tse, "Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3549–3572, 2007.

[37] R. Raz, O. Reingold, and S. Vadhan, "Extracting all the randomness and reducing the error in Trevisan's extractors," *Journal of Computer and System Sciences*, vol. 65, pp. 97–128, 2002.

[38] R. Rivest, A. Shamir, and L. Adleman, “A method for obtaining digital signatures and public-key cryptosystems,” *Communications of the ACM*, vol. 21, no. 2, pp. 120–126, 1978.

[39] C. E. Shannon, “Communication theory of secrecy systems,” *Bell System Tech. J.*, vol. 28, pp. 656–715, Oct. 1949.

[40] C. Shannon, “Two-way communication channels,” in *Proc. of the 4th Berkeley Symp. on Mathematical Statistics and Probability*, vol. 1, Berkeley, USA, 1961, pp. 611–644.

[41] P. Shor, “Algorithms for quantum computation: Discrete logarithms and factoring,” in *Proc. 35th Annual Symposium on Foundations of Computer Science*, 1994, pp. 124–134.

[42] D. Slepian and J. Wolf, “Noiseless coding of correlated information sources,” *IEEE Trans. Inform. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.

[43] E. Tekin and A. Yener, “The general Gaussian multiple-access and two-way wiretap channels: Achievable rates and cooperative jamming,” *IEEE Trans. Inform. Theory*, vol. 54, no. 6, pp. 2735–2751, Jun. 2008.

[44] E. Telatar and D. Tse, “Bounds on the capacity region of a class of interference channels,” in *Proc. of the IEEE Int. Symposium on Inform. Theory*, Nice, France, 2007.

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

[46] D. Tse and S. Hanly, “Multiaccess fading channels. I. Polymatroid structure, optimal resource allocation and throughput capacities,” *IEEE Trans. Inform. Theory*, vol. 44, no. 7, pp. 2796–2815, Nov. 1998.

[47] 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.

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

[49] L.-L. Xie and P. Kumar, “A network information theory for wireless communication: Scaling laws and optimal operation,” *IEEE Trans. Inform. Theory*, vol. 50, pp. 748–767, May 2004.

[50] R. Yeung and Y. Yan, “Information theoretic inequality prover,” N.T., Hong Kong, China, http://user-www.ie.cuhk.edu.hk/˜ITIP.

Curriculum Vitae – Etienne Perron

Education:

- **École Polytechnique Fédérale de Lausanne (EPFL)**
  Ph.D. in computer and communication sciences
  research topic: network rate-distortion theory /
  information theoretic secrecy in wireless networks
  2004 - 2009

- **EPFL**
  Swiss federal diploma of engineering
  communication systems division
  1998 - 2004

Professional Experience:

- **École Polytechnique Fédérale de Lausanne (EPFL)**
  research assistant; research in small teams in applied mathematics,
  teaching, supervision of semester projects
  2004 - 2009

- **University of California**, Berkeley, USA
  master thesis of 6 months; research in applied statistics / sensor networks
  2004

- **Institute for Telecommunications Research**, Adelaide, Australia
  occupational training of 5 months; work in a theoretical research group
  2002

- **Lightning Instrumentation**, Lausanne, Switzerland
  internship of 6 weeks; adaption of software for a Linux-based router
  2002

- **ELCA Informatique**, Lausanne, Switzerland
  internship of 4 weeks; creation of firm-wide guidelines for web-applications
  2001

- **Innpublish**, Celerina, Switzerland
  part-time web-programmer; development of web-applications
  2000-2003

Conference Publications:

1. Etienne Perron, Suhas Diggavi, Emre Telatar, "*Lossy Source Coding with Gaussian or Erased Side-Information*", to be presented at the IEEE International Symposium on Information Theory, Seoul, Korea, July 2009
2. Etienne Perron, Suhas Diggavi, Emre Telatar, "*The Interference-Multiple-Access Channel*", to be presented at the IEEE International Conference on Communications, Dresden, Germany, June 2009
3. Etienne Perron, Suhas Diggavi, Emre Telatar, "*On Cooperative Wireless Network Secrecy*", Proc. of IEEE Infocom, Rio de Janeiro, Brazil, April 2009
4. Etienne Perron, Dinkar Vasudevan, Milan Vojnović, "*Using Three States for Binary Consensus on Complete Graphs*", Proc. of IEEE Infocom, Rio de Janeiro, Brazil, April 2009
5. Etienne Perron, Suhas Diggavi, Emre Telatar, "*On Noise Insertion Strategies for Wireless Network Secrecy*", Proc. of the Information Theory and Applications Workshop, San Diego, USA, February 2009
6. Etienne Perron, Suhas Diggavi, Emre Telatar, "*Wireless Network Secrecy with Public Feedback*", Proc. of the Allerton Conference on Communication, Control, and Computing, Illinois, USA, September 2008
7. Dinkar Vasudevan, Etienne Perron, "*Cooperative Source Coding with Encoder Breakdown*", Proc. of the IEEE International Symposium on Information Theory, Nice, France, June 2007
8. Etienne Perron, Suhas Diggavi, Emre Telatar, "*On the Role of Encoder Side-Information in Source Coding for Multiple Decoders*", Proc. of the IEEE International Symposium on Information Theory, Seattle, USA, July 2006
9. Etienne Perron, Mohammad Rezaeian, Alex Grant, "*The On-Off Fading Channel*", Proc. of the IEEE International Symposium on Information Theory, Yokohama, Japan, July 2003

Software:

1. *Xitip:* X-Windows Information-Theoretic Inequality Prover, joint work with Rethnakarn Pulikkoonattu, written in C using lex and yacc;
   builds upon the software ITIP by Yeung and Yan;
   available versions: C source code, Linux precompiled binary file, Windows self-extracting executable;
   webpage: *http://xitip.epfl.ch/*