On delay and security in network coding

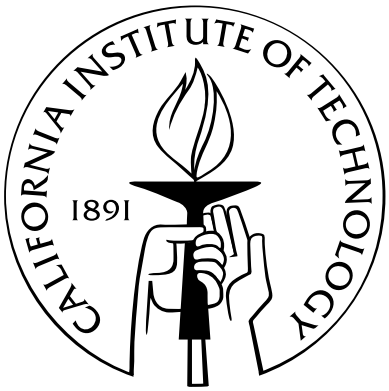
Thesis by

Theodoros K. Dikaliotis

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2013

(Defended May 31, 2012)

To my parents Kwsta and Vasiliki Dikalioti

Acknowledgements

The journey of getting a PhD is long and filled with bright moments and many more disappointing ones. There were plenty of times I felt this journey would never end and certainly that would be the case without the help of many bright and talented people I met during my stay at Caltech. Most important among all I am indebted to my advisor, Tracey Ho, who with her endless kindness, patience, and continuous encouragement, helped me tackle one of the most difficult endeavors in my life. The ease with which she solved seemingly intractable problems will always be an inspiration for me.

My sincere gratitude goes to Michelle Effros for her help and insights in many parts of this thesis. Her pedagogical way of explaining even the most obscure notions of information theory has helped me countless times. I would like to thank Sid Jaggi for being an excellent host in Hong Kong and proving to me that research can sometimes be fun. A great part of this work would not be possible without the help of my friend and occasional research mentor Alex Dimakis. His intuition and passion for research always amazed me, but more importantly I am lucky to know him personally and consider him as one of my best friends. I would like to thank my thesis committee members– Babak Hassibi, Steven Low, and P. P. Vaidyanathan–for the valuable feedback and comments on the presentation of this thesis.

I am extremely thankfully to all my friends that made my stay at Caltech bearable. In particular, I would like to thank: Derek Leong for the countless philosophical discussions in and out of our office and his help with any kind of IT problem that might have arisen during all these years. Hongyi Yao for numerous brainstorming sessions and introducing me to Chinese cuisine. Kwstas Zafeiropoulos and Stavros Gonidakis for being the best company during many nights in Hollywood and day trips to Laguna. Without their presence this work would have been better and finished much faster but I regret none of the moments we have spent together. My dear friend Makis Fellios for hours of endless conversations and the support he has shown me whenever I needed it the most. Svitlana Vyetrenko for being the brightest part of my life during the PhD years; without her my stay at Caltech would have been unimaginable.

Last, but not least, I would like to thank my family for their unconditional love and support: My uncle Elias and Jeanne for being my family away from home and for having their house always open

to me, I will forever be grateful for all their help during my years at Caltech. My beloved parents Kwstas and Vasiliki for cherishing me with love and sacrificing everything for me and my sisters; for believing in me even in those moments I was losing faith in myself. Their appreciation for education and their constant encouragements shaped the person I am today. My sisters Stavroula and Marina for patiently listening and occasionally advising me on all my "adventures" but thankfully never taking me too seriously. To all the people above I will always be thankful for making this journey happen and worth taking.

Abstract

In this thesis, delay and security issues in network coding are considered. First, we study the delay incurred in the transmission of a fixed number of packets through acyclic networks comprised of erasure links. The two transmission schemes studied are routing with hop-by-hop retransmissions, where every node in the network simply stores and forwards its received packets, and linear coding, where nodes mix their packets by forwarding linear combinations of all their previously received packets. We show that even though the achievable rates of coding and routing are the same, network coding can have an increasingly better performance than routing as the number of packets increases.

Secondly, we investigate the security benefits of network coding. We investigate the achievable secrecy rate region in a general network of noisy wiretap channels with general communication demands. The eavesdropper has access to an unknown set of links, and on the wiretapped links observes a degraded version of the intended receiver's observation. While characterizing the capacity in general is an open problem, in the noise-free case there exist inner and outer bounds. In the noisy case, we show how one can change any of the wiretap channels to a noiseless degraded broadcast channel, so that the derived network's rate region bounds, and under certain conditions is equivalent, to that of the initial network. Specifically, we showed that in case the eavesdropper can choose only a single link to wiretap at each time, then one can change all the links in the network with corresponding noiseless ones, creating an equivalent noiseless secrecy problem. In the case where the eavesdropper can wiretap multiple links simultaneously, we derive upper and lower bounding noiseless network problems.

Finally, we consider design practical code design for the detection of adversarial errors in a distributed storage system. We build on work of functions that can fool linear polynomials to create and communicate hash functions of the data in order to detect with high probability the maliciously attacked nodes in the system.

Contents

List of Figures

Chapter 1

Introduction

This thesis studies problems related to transmission delay and security in network scenarios with coding. It builds on classical information theory as well as the more recent field of network coding. Both these fields consider the use of coding operations in order to characterize and approach fundamental limits of performance in various communication scenarios, the former focusing on general noisy channels and the latter on networks of simple channels. Among the performance objectives of interest in network communications, throughput, latency and security are some of the most important. While throughput (capacity) has been extensively studied for various channel models and network structures, relatively less is known so far about latency and security in networks.

The first problem in this thesis analyzes file transmission delay on unreliable networks, which is a more detailed performance metric compared to throughput. Specifically, in a network of erasure links with a single source and a single destination we examine the expected time it takes to send a fixed number of packets from the source to the destination. Previous work has shown that when hop by hop feedback is unavailable, network coding is necessary to achieve the maximum throughput in the network [1]. Although in the case of a single unicast transmission both network coding and routing with hop-by-hop feedback and retransmissions achieve capacity, we show that in terms of latency, network coding outperforms routing, and give an analytical characterization of the performance gap.

The other two problems we investigate deal with the security benefits of network coding. Long gone, are the days where networks had purely academic interest and all users were legitimate. In todays world, computer network fall victims of malicious attacks from adversaries ranging in power and computation resources from a single user to rogue nations. In network that carry sensitive information it is of paramount importance to send the maximum possible information while ensuring that there is no leakage to any potential wiretapper, and in the first of the two problems we are concerned with the secure rate region of general network with generic demands in the presence of an eavesdropper.

In the second problem we deal with the detection of malicious attacks in a distributed storage system using coding. In recent years, the demand for large-scale data storage has increased signifi-

cantly, with applications like social networks, file, and video sharing demanding access and security for storing massive amounts of data. Distributed storage systems introduce redundancy to increase reliability and use coding since it has reduced storage requirements than simple replication [2]. However when coding is used, the distributed storage system is particularly vulnerable to malicious attacks. Due to the mixing operations used for the reconstruction of the initial information, even a small number of attacked nodes can compromise the whole system, making the detection of those contagious nodes of paramount importance. We derive a detection scheme that uses hash functions and the communication of a logarithmic number of bits to identify compromised nodes.

1.1 Background and Related Work

1.1.1 Network Coding

Ahlswede et al. [3] introduced network coding for the class of *multicast* problems, where one source wishes to transmit the same information to all receivers in the network. They showed that the traditional approach of storing and forwarding packets might not to be sufficient to achieve the multicast rate when there are multiple receivers. Instead intermediate nodes should in general forward functions of their incoming packets in order to achieve the min-cut of the network, which is the maximum information rate that can be achieved. Work by [4] showed that propagating linear combinations (linear network coding) suffices to achieve the same rate region and an algebraic framework for linear network coding was presented in [5]. Authors in [6] gave the first polynomial time design algorithms for linear network codes.

Further Ho et al. [7] showed that with a sufficiently large field size the linear combinations of network coding can be chosen randomly in a distributed manner, and achieve the maximum flow capacity with high probability in a practical and decentralized manner with low design complexity. Practical network coding protocols have further been developed [8, 9]. Creating random linear combinations of packets in a network have been proved to be robust against packet losses [1] as mixing packets together act as an erasure code introducing redundancy in the network. Authors in [10] used vector spaces spanned by the transmitted packets as codewords and developed an error-correction code for a noncoherent network model in the form of rank metric codes.

Dimakis et al. [2] used network coding to develop distributed storage codes with efficient storage requirements and optimal repair characteristics that by far surpass simple replication. Further research [11, 12] was devoted to the development of regenerating codes, or the exact repair of the just the systematic part of the code [13]. Related to distributed storage, network coding has been proved to have supreme characteristics in file sharing systems [14, 15].

1.1.2 Information Theoretic Security

Shannon [16] set the foundations for the mathematical treatment of secrecy and cryptography and was the first to define the notion of information theoretic security. Later Wyner [17] found the maximum secure achievable rate for a point-to-point channel in the presence of a wiretapper. Much work have been devoted since then, with varying degrees of success, to derive results on more complicated networks involving multiple nodes interconnected together. Csiszar and Korner [18] analyzed the case of a broadcast channel with both a confidential and a public message; a variation of this problem appears in [19]. A natural generalization of these settings was the model for secret-key agreement [20].

Relating network coding to secrecy and network security, Cai and Yeung [21] studied the maximum rate at which information can be reliably sent without information leakage to an eavesdropper that observes all messages traversing a limited but unknown subset of links. A similar problem where only the input vector is modified is considered in [22]. In [23, 24] Cui, Ho and Kliewer show that finding the capacity of the secure network communication problem is NP hard for the case of unequal capacity links. Different notions of security are introduced in [25, 26].

1.2 Thesis Outline and Contributions

The thesis outline and contributions are as follows. In Chapter 2 we look at the expected delay to send a given number n of packets through networks of erasure channels. Unlike existing results [1, 27, 28] on network coding over lossy networks that focus on the rate region achieved by network coding, the expected time to send a number of packets through a network has not been completely characterized. The expected time to complete a transmission reveals sublinear characteristics of the network that are not shown in the achievable rates. We show that in the case of a parallel path network the gap between the expected transmission time of network coding versus hop-by-hop retransmissions can grow as $\sqrt{n}$ despite that the two transmission schemes achieve the same rates. Moreover closely related work on delay in queueing theory [29, 30] assumes Poisson arrivals and their results pertain to the delay of individual packets in steady state, our analysis does not involve any assumptions about reaching steady state.

In Chapter 3 we connect two separate but related bodies of literature: that of the wiretap channels by Wyner [17] and the field of secure network coding introduced by Cai and Yeung [21]. Building on the equivalence approach of [31, 32], we derive upper and lower bounds for the secrecy capacity region of noisy networks composed of degraded "simultaneously maximizeable" wiretap channels. Simultaneously maximizeable channels are those for which the same input distribution maximizes the mutual information to the intended receiver and the eavesdropper. For the case where the eavesdropper can choose to wiretap at most one channel at a time, the upper and lower

bounds are tight, and thus an equivalence holds between the noisy network and a network where each eavesdropped link is replaced by a set of noiseless channels. This may come as a surprise since in [31, 32] tight bounds cannot be found for networks with multiuser channels. One reason for this is that unlike [31, 32] the receiver of the degraded output is an eavesdropper and has no decoding requirements. For the case where the eavesdropper can access multiple channels at the same time, the upper and lower bounds are shown to be generally loose and it is still an open problem whether one can find noiseless channels that would emulate a noisy channel in the case of multiple eavesdropped channels. Both the upper and the lower bounds apply to general networks with general multiple multicast demands, even when the actual capacity region is unknown.

In Chapter 4 we study the security of a distributed storage system and derive a detection scheme to detect compromised nodes. We use pseudorandom small-bias generators used in [33] for a general field size and create almost perfectly random sequences that can fool any linear function. Effectively we take projections of the data stored in the distributed storage system with random vectors and convey the result to a central node that uses the existing redundancy of the code to detect all attacked nodes. Our work makes no assumption for the computational capabilities of the attacker, and we ensure detection with high probability as long as the number of attacked nodes are within the capabilities of the code used.

Chapter 2

On the Delay Advantage of Coding in Packet Erasure Networks

This chapter considers the block delay for unicasting a file consisting of n packets over a packet erasure network with probabilistic erasures. Such networks have been extensively studied from the standpoint of capacity. Various schemes involving coding or retransmissions have been shown to be capacity-achieving for unicasting in networks with packet erasures, e.g. [28, 1, 34, 35]. For a capacity-achieving strategy, the expected block delay for transmitting n packets is $\frac{n}{C} + D(n)$ where C is the minimum cut capacity and the delay function $D(n)$ is sublinear in n but differs in general for different strategies. In general networks, the optimal $D(n)$ is achieved by random linear network coding, in that decoding succeeds with high probability for any realization of packet erasure events for which the corresponding minimum cut capacity is n[1]. However, relatively little has been known previously about the behavior of the delay function $D(n)$ for coding or retransmission strategies.

In this paper, we analyze the delay function $D(n)$ for random linear network coding (coding for short) as well as an uncoded hop-by-hop retransmission strategy (routing for short) where only one copy of each packet is kept in intermediate node buffers. Schemes such as [36, 35] ensure that there is only one copy of each packet in the network; without substantial non-local coordination or feedback, it is complicated for an uncoded topology-independent scheme to keep track of multiple copies of packets at intermediate nodes and prevent capacity loss from duplicate packet transmissions. We also assume that the routing strategy fixes how many packets will traverse each route a priori based on link statistics, without adjusting to link erasure realizations. While routing strategies could dynamically re-route packets under atypical realizations, this would not be practical if the min-cut links are far from the source. On the other hand, network coding allows redundant packets to be transmitted efficiently in a topology-independent manner, without feedback or coordination, except for an acknowledgment from the destination when it has received the entire file. As such, network

[1]The field size and packet length are assumed in this paper to be sufficiently large so that the probability of rank-deficient choices of coding coefficients can be neglected, along with the fractional overhead of specifying the random coding vectors.

coding can fully exploit variable link realizations. These differences result in a coding advantage in delay function $D(n)$ which, as we will show, can be unbounded with increasing n.

A major technical challenge in the analysis of the delay function for the routing strategy involves computing the expectation of the maximum of two independent negative binomial random variables. This problem has been previously studied in [37], where authors explain in detail why it is complicated[2] and derive an approximate solution to the problem. Our analysis addresses this open problem by finding an exact expression and showing that it grows to infinity at least as the square root of n.

Related work on queuing delay in uncoded [29, 30] and coded [38] systems has considered the case of random arrivals and their results pertain to the delay of individual packets in steady state. This differs from our work which considers the delay for communicating a fixed size batch of n packets that are initially present at the source.

2.1 Main Results

For a line network, the capacity is given by the worst link. We show a finite bound on the delay function that applies to both coding and the routing scheme when there is a single worst link.

Theorem 1. *Consider n packets communicated through a line network of ℓ links with erasure probabilities $p_1, p_2, \ldots, p_\ell$ where there is a unique worst link:*

$$p_m := \max_{1\le i\le \ell} p_i, \quad p_i < p_m < 1 \quad \forall\, i \neq m.$$

The expected time $\mathbb{E}T_n$ to send all n packets either with coding or routing is:

$$\mathbb{E}T_n = \frac{n}{1 - \max\limits_{1\le i\le \ell} p_i} + D(n, p_1, p_2, \ldots, p_\ell), \tag{2.1}$$

where the delay function $D(n, p_1, p_2, \ldots, p_\ell)$ is non-decreasing in n and upper bounded by:

$$\bar{D}(p_1, p_2, \ldots, p_\ell) := \sum_{i=1, i\neq m}^{\ell} \frac{p_m}{p_m - p_i}.$$

If on the other hand there are two links that take the worst value, then the delay function is not bounded but still exhibits sublinear behavior. Pakzad et al. [27] show that in the case of a line network with identical links, the optimal delay function grows as $\sqrt{n}$. This is achieved by both

[2] Authors in [37] deal with the expected maximum of any number of negative binomial distributions but the difficulty remains even for two negative binomial distributions.

coding and the routing strategy[3].

In contrast, for parallel path networks, we show that the delay function behaves quite differently for the coded and uncoded schemes.

Theorem 2. *The expected time $\mathbb{E}T_n^c$ taken to send n packets using coding over a k-parallel path multi-hop network is*

$$\mathbb{E}T_n^c = \frac{n}{k - \sum_{i=1}^{k} \max_{1 \le j \le \ell} p_{ij}} + D_n^c$$

where the delay function D_n^c depends on all the erasure probabilities p_{ij}, for $i \in \{1, \ldots, k\}$, $1 \le j \le \ell$. In the case where there is single worst link in each path D_n^c is bounded, i.e. *$D_n^c \in \mathsf{O}(1)$ whereas if there are multiple worst links in at least one path then $D_n^c \in \mathsf{O}(\sqrt{n})$. The result holds regardless of any statistical dependence between erasure processes on different paths.*

Theorem 3. *The expected time $\mathbb{E}T_n^r$ taken to send n packets through a k-parallel path network by routing is*

$$\mathbb{E}T_n^r = \frac{n}{k - \sum_{i=1}^{k} \max_{1 \le j \le \ell} p_{ij}} + D_n^r \tag{2.2}$$

where the delay function D_n^r depends on all the erasure probabilities p_{ij}, for $i \in \{1, \ldots, k\}$, $1 \le j \le \ell$ and grows at least as $\sqrt{n}$, i.e. *$D_n^r \in \Omega(\sqrt{n})$.*

The above results on parallel path networks generalize to arbitrary topologies. We define *single-bottleneck networks* as networks that have a single min-cut.

Theorem 4. *In a network of erasure channels with a single source $\mathcal{S}$ and a single receiver $\mathcal{T}$ the expected time $\mathbb{E}T_n^r$ taken to send n packets by routing is*

$$\mathbb{E}T_n^r = \frac{n}{C} + \hat{D}_n^r$$

where C is the capacity of the network and $\hat{D}_n^r \in \Omega(\sqrt{n})$. In the case of network coding the expected time $\mathbb{E}T_n^c$ taken to send n packets is

$$\mathbb{E}T_n^r = \frac{n}{C} + \hat{D}_n^r$$

where $\hat{D}_n^c \in \mathsf{O}(1)$ for single-bottleneck networks.

[3]The result in [27] is derived for the routing strategy which is delay-optimal in a line network; as discussed above, coding in a sufficiently large field is delay-optimal in any network.

We also prove the following concentration result:

Theorem 5. *The time T_n^c for n packets to be transmitted from a source to a sink over a network of erasure channels using network coding is concentrated around its expected value with high probability. In particular for sufficiently large n:*

$$\mathbb{P}\mathrm{r}\left[|T_n^c - \mathbb{E}T_n^c| > \epsilon_n\right] \leq \frac{2C}{n} + o\left(\frac{1}{n}\right), \tag{2.3}$$

where C is the capacity of the network and ϵ_n represents the corresponding deviation and is equal to $\epsilon_n = n^{1/2+\delta}/C$, $\delta \in (0, 1/2)$.

Since $\mathbb{E}T_n^{\mathrm{c}}$ grows linearly in n and the deviations ϵ_n are sublinear, T_n^{c} is tightly concentrated around its expectation for large n with probability approaching one. Subsequent to our initial conference publications [39, 40], further results on delay for line networks have been obtained by [41, 42].

2.2 Model

We consider a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ denotes the set of nodes and $\mathcal{E} = \mathcal{V} \times \mathcal{V}$ denotes the set of edges or links. We assume a discrete time model, where at each time step each node $v \in \mathcal{V}$ can transmit one packet on its outgoing edges. For every edge $e \in \mathcal{E}$ each transmission succeeds with probability $1 - p_e$ or the transmitted packet gets erased with probability p_e; erasures across different edges and time steps are assumed to be independent. In our model, in case of a success the packet is assumed to be transmitted to the next node instantaneously, i.e. we ignore the transmission delay along the links. We assume that no edge fails with probability 1 (i.e. $p_e < 1$ for all $e \in \mathcal{E}$) since in such a case we can remove that edge from the network.

Within network $\mathcal{G}$ there is a single source $\mathcal{S} \in \mathcal{V}$ that wishes to transmit n packets to a single destination $\mathcal{T}$ in $\mathcal{G}$. We investigate the expected time it takes for the n packets to be received by $\mathcal{T}$ under two transmission schemes, network coding and routing. When network coding is employed, each packet transmitted by a node $v \in \mathcal{V}$ is a random linear combination of all previously received packets at the node v. The destination node $\mathcal{T}$ decodes once it has received n linearly independent combinations of the initial packets. When routing is employed, the number of packets transmitted in each path is fixed ahead of the transmission, in such a way that the expected time for all n packets to reach destination $\mathcal{T}$ is minimized.

All nodes in the network are assumed to have sufficiently large buffers to store the necessary number of packets to accommodate the transmission scheme. In the case of routing, we assume an automatic repeat request (ARQ) scheme with instantaneous feedback available on each hop. Thus, a node can drop a packet that has been successfully received by the next node. For the case of coding,

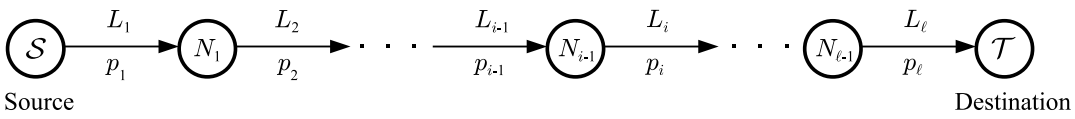


Figure 2.1: Multi-hop line network

as explained in [43], information travels through the network in the form of innovative packets, where a packet is innovative for a node v if it is not in the linear span of packets previously received by v. For simplicity of analysis, we assume that a node can store up to n linearly independent packets; smaller buffers can be used in practice[4]. Feedback is not needed except when the destination $\mathcal{T}$ receives all the information and signals the end of transmission to all nodes. Our results hold without any restrictions on the number of packets n or the number of edges in the network, and there is no requirement for the network to reach steady state.

2.3 Line Networks

The line network under consideration is depicted in Figure 2.1. The network consists of ℓ links L_i, $1 \leq i \leq \ell$ and $\ell + 1$ nodes N_j, $0 \leq j \leq \ell$. Node N_j, $0 \leq j \leq \ell - 1$ is connected to node N_{j+1} to its right through the erasure link L_{j+1}, where we assume that the source $\mathcal{S}$ and the destination $\mathcal{T}$ are also defined as nodes N_0 and N_ℓ respectively. The probability of transmission failure on each link L_i is denoted by p_i.

For the case of a line network there is no difference between network coding and routing in the expected time it takes to transmit a fixed number of packets. Note that coding at each hop (network coding) is needed to achieve minimum delay in the absence of feedback, whereas coding only at the source is suboptimal in terms of throughput and delay [1].

Proof of Theorem 1. By using the interchangeability result on service station from Weber [45], we can interchange the position of any two links without affecting the departure process of node $N_{\ell-1}$ and therefore the delay function. Consequently, we can interchange the worst link in the queue (which is unique from the assumptions of Theorem 1) with the first link, and thus we will assume that the first link is the worst link ($p_2, p_3, \ldots, p_\ell < p_1 < 1$).

Note that in a line network, under coding the subspace spanned by all packets received so far at a node N_i contains that of its next hop node N_{i+1}, similarly to the case of routing where the set of packets received at a node N_i is a superset of that of its next hop node N_{i+1}. Let the random variable $R_i^n, 1 \leq i \leq \ell - 1$, denote the rank difference between node N_i and node N_{i+1}, at the moment packet n arrives at N_1. This is exactly the number of packets present at node N_i that

[4]By the results of [44], the buffer size needed for coding is no larger than that needed for routing.

are innovative for N_{i+1} (which for brevity we refer to simply as innovative packets at node N_i in this proof) at the random time when packet n arrives at N_1. For any realization of erasures, the evolution of the number of innovative packets at each node is the same under coding and routing.

The time T_n taken to send n packets from the source node $\mathcal{S}$ to the destination $\mathcal{T}$ can be expressed as the sum of time $T_n^{(1)}$ required for all the n packets to cross the first link and the time τ_n required for all the remaining innovative packets $R_1^n, \ldots, R_{\ell-1}^n$ at nodes $N_1, \ldots, N_{\ell-1}$ respectively to reach the destination node $\mathcal{T}$:

$$T_n = T_n^{(1)} + \tau_n.$$

All the quantities in the equation above are random variables and we want to compute their expected values. Due to the linearity of the expectation

$$\mathbb{E}T_n = \mathbb{E}T_n^{(1)} + \mathbb{E}\tau_n \tag{2.4}$$

and by defining $X_j^{(1)}, 1 \leq j \leq n$ to be the time taken for packet j to cross the first link, we get:

$$\mathbb{E}T_n^{(1)} = \sum_{j=1}^{n} \mathbb{E}X_j^{(1)} = \frac{n}{1-p_1} \tag{2.5}$$

since $X_j^{(1)}, 1 \leq j \leq n$, are all geometric random variables ($\mathbb{P}\mathrm{r}\left(X_j^{(1)} = k\right) = (1-p_1) \cdot p_1^{k-1}, k \geq 1$). Therefore combining equations (2.4) and (2.5) we get:

$$\mathbb{E}T_n^{(1)} = \frac{n}{1-p_1} + \mathbb{E}\tau_n. \tag{2.6}$$

Equations (2.1), (2.6) give

$$D(n, p_1, p_2, \ldots, p_\ell) = \mathbb{E}\tau_n$$

which is the expected time taken for all the remaining innovative packets at nodes $N_1, \ldots, N_{\ell-1}$ to reach the destination. For the simplest case of a two-hop network ($\ell = 2$) we can derive recursive formulas for computing this expectation for each n. Table 2.3 has closed-form expressions for the delay function $D(n, p_1, p_2)$ for $n = 1, \ldots, 4$. It is seen that as n grows, the number of terms in the above expression increases rapidly, making these exact formulas impractical, and as expected for larger values of ℓ (≥ 3) the situation only worsens. Our subsequent analysis derives tight upper bounds on the delay function $D(n, p_1, p_2, \ldots, p_\ell)$ for any ℓ which do not depend on n.

The $(\ell-1)$-tuple $Y_n = (R_1^n, \ldots, R_{\ell-1}^n)$ representing the number of innovative packets remaining at nodes $N_1, \ldots, N_{\ell-1}$ at the moment packet n arrives at node N_1 (including packet n) is a multi-dimensional Markov process with state space $E \subset \mathbb{N}^{\ell-1}$ (the state space is a proper subset of $\mathbb{N}^{\ell-1}$

Table 2.1: The delay function $D(n, p_1, p_2)$ for different values of n

n	$D(n, p_1, p_2)$
1	$\frac{1}{1-p_2}$
2	$\frac{2}{1-p_2} - \frac{1}{1-p_1p_2}$
3	$\frac{1+p_2\big(2-p_1\big(6-p_1+(2-5p_1)p_2+(1-3(1-p_1)p_1)p_2^2\big)\big)}{(1-p_2)(1-p_1p_2)^3}$
4	$\frac{\left\{\begin{array}{c} 1+p_2(3-p_1(11+4p_1^4p_2^4+p_2(5+(5-p_2)p_2)+p_1^3p_2(1-p_2(5+2p_2(5+3p_2))) \\ -p_1(4+p_2(15+p_2(21-(1-p_2)p_2)))+p_1^2(1-p_2(1-p_2(31+p_2(5+4p_2)))))) \end{array}\right\}}{(1-p_2)(1-p_1p_2)^5}$

since Y_n can never take the values $(0, *, \ldots, *)$ where the $*$ represents any integer value). Using the coupling method [46] and an argument similar to the one given at Proposition 2 in [47] it can be shown that Y_n is a stochastically increasing function of n (meaning that as n increases there is a higher probability of having more innovative packets at nodes $N_1, \ldots, N_{\ell-1}$).

Proposition 1. *The Markov process $Y_n = (R_1^n, \ldots, R_{\ell-1}^n)$ is $\preceq_{st}$-increasing.*

Proof. Given in Appendix A along with the necessary definitions. □

A direct result of Proposition 1 is that the expected time taken $\mathbb{E}\tau_n$ for the remaining innovative packets at nodes $N_1, \ldots, N_{\ell-1}$ to reach the destination is a non-decreasing function of n:

$$\mathbb{E}\tau_n \leq \mathbb{E}\tau_{n+1} \leq \lim_{n\to\infty} \mathbb{E}\tau_n \tag{2.7}$$

where the second inequality is meaningful when the limit exists.

Innovative packets travelling in the network from node N_1 to the destination node $\mathcal{T}$ can be viewed as customers travelling through a network of service stations in tandem. Indeed, each innovative packet (customer) arrives at the first station (node N_1) with a geometric arrival process and the transmission (service) time is also geometrically distributed. Once an innovative packet has been transmitted (serviced) it leaves the current node (station) and arrives at the next node (station) waiting for its next transmission (service).

It is helpful to assume the first link to be the worst one in order to use the results of Hsu and Burke in [48]. The authors proved that a tandem network with geometrically distributed service times and a geometric input process, reaches steady state as long as the input process is slower than any of the service times. Our line network is depicted in Figure 2.1 and the input process (of innovative packets) is the geometric arrival process at node N_1 from the source $\mathcal{S}$. Since $p_2, p_3, \ldots, p_\ell < p_1$ the arrival process is slower than any service process (transmission of the innovative packet to the next hop) and therefore the network in Figure 2.1 reaches steady state.

Sending an arbitrarily large number of packets ($n \to \infty$) makes the problem of estimating

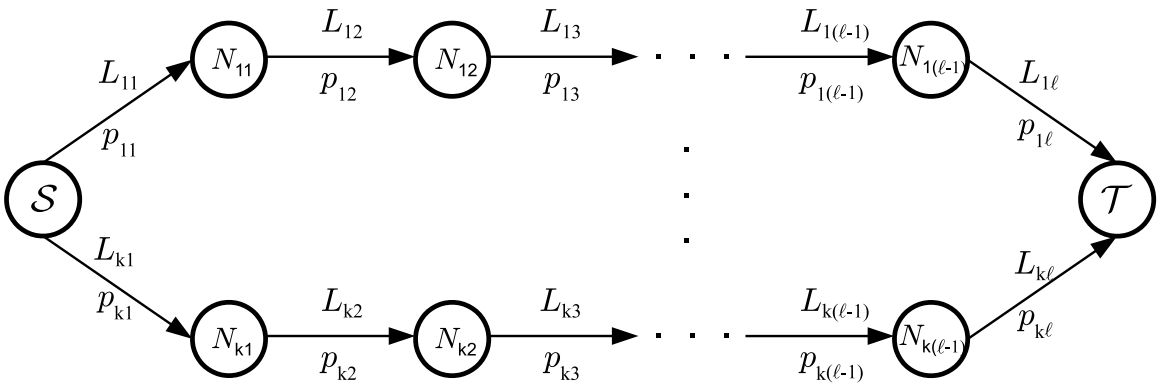


Figure 2.2: Two parallel multi-hop line networks having links with different erasure probabilities

$\lim_{n\to\infty} \mathbb{E}\tau_n$[5] the same as calculating the expected time taken to send all the remaining innovative packets at nodes $N_1, \ldots, N_{\ell-1}$ to reach the destination $\mathcal{T}$ at steady state. This is exactly the expected end-to-end delay for a single customer in a line network that has reached equilibrium. This quantity has been calculated in [49] (page 67, Theorem 4.10) and is equal to

$$\lim_{n\to\infty} \mathbb{E}\tau_n = \sum_{i=2}^{\ell} \frac{p_1}{p_1 - p_i}. \tag{2.8}$$

Combining equations (2.7) and (2.8) and changing p_1 to $p_m := \max p_i < 1$ concludes the proof of Theorem 1. □

2.4 k-parallel Path Network

We define the *k-parallel path network* as the network depicted in Figure 2.2. This network consists of k parallel multi-hop line networks (paths) with $k\ell$ nodes and $k\ell$ links, with ℓ links in each path (our results are readily extended to networks with different number of links in each path). Each node $N_{i(j-1)}$ is connected to the node N_{ij} on its right by a link L_{ij}, for $i \in \{1, \ldots, k\}$ and $1 \leq j \leq \ell$ where for consistency we assume that the source $\mathcal{S}$ and the destination $\mathcal{T}$ are defined as nodes N_{i0} and $N_{i\ell}$, $i \in \{1, \ldots, k\}$, respectively.

For the case of routing with retransmissions, the source $\mathcal{S}$ divides the n packets between the different paths so that the time taken to send all the packets is minimized in expectation. This is accomplished by having the number of packets that cross each path to be proportional to the capacity of the path. Indeed, if the source $\mathcal{S}$ sends $n_1, \ldots, n_k$ number of packets though each path then according to Theorem 1 the expected time to send these packets is $\frac{n_i}{1-p_{1i}} + D_{n_i}$, $i \in \{1, \ldots, k\}$,

[5] If the network was not reaching a steady state the above limit would diverge.

where D_{n_i} are bounded delay functions. The values n_i are chosen so that the linear terms of the above expected values are equal, *i.e.* $\frac{n_1}{1-p_{11}} = \ldots = \frac{n_k}{1-p_{k1}}$ and $n_1 + \ldots + n_k = n$. Therefore the choice of

$$n_i = \frac{n(1-p_{i1})}{k - \sum_{i=1}^{k} p_{i1}},\ i \in \{1, \ldots, k\} \tag{2.9}$$

minimizes the expected time to send the n packets. Therefore from now on, when routing is performed, source $\mathcal{S}$ is assumed to send $n(1-p_{i1})/(k-\sum_{i=1}^{k} p_{i1})$ over each path i.[6]

2.4.1 Coding Strategy

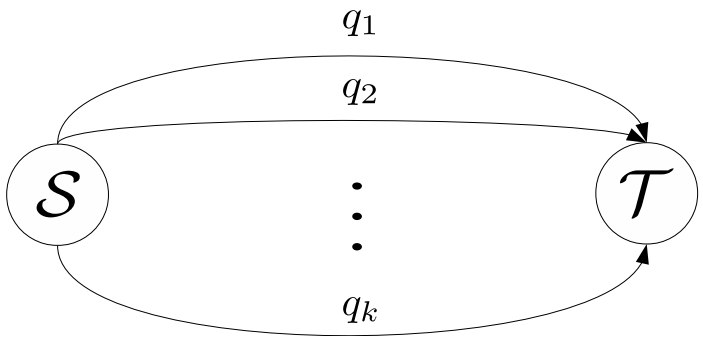


Figure 2.3: A network of k parallel erasure links with erasure probabilities $q_1, \ldots, q_k$ connecting source $\mathcal{S}$ and destination $\mathcal{T}$.

Before we analyze the expected time $\mathbb{E}T_n^{\mathrm{c}}$ taken to send n packets through the network in Figure 2.2 using coding (where the c superscript stands for coding), we prove the following proposition that holds for the simplified network of k parallel erasure links connecting the source to the destination as in Figure 2.3.

Proposition 2. *The expected time $\mathbb{E}\hat{T}_n^c$ taken to send by coding n packets from source $\mathcal{S}$ to destination $\mathcal{T}$ through k parallel erasure links with erasure probabilities $q_1, \ldots, q_k$ respectively is*

$$\mathbb{E}\hat{T}_n^c = \frac{n}{k - \sum_{i=1}^{k} q_i} + B_n$$

where B_n is a bounded term. This relation holds regardless of any statistical dependence between the erasure processes on different links.

Proof. We define $A_0, A_1, \ldots, A_k$ to be the probabilities of having $0, 1, \ldots, k$ links succeed at a specific

[6]To simplify the notation we will assume that all numbers $n(1-p_{i1})/(k-\sum_{i=1}^{k} p_{i1})$ are integers. Our results extend to the case that those numbers are not integers by rounding them to the closest integer.

time instance. The recursive formula for $\mathbb{E}\hat{T}_n^c$ is:

$$\begin{aligned} \mathbb{E}\hat{T}_n &= A_0 \cdot (\mathbb{E}\hat{T}_n^c + 1) + A_1 \cdot (\mathbb{E}\hat{T}_{n-1}^c + 1) + \ldots + A_k \cdot (\mathbb{E}\hat{T}_{n-k}^c + 1) \\ \Leftrightarrow (1 - A_0) \cdot \mathbb{E}\hat{T}_n &= A_1 \cdot \mathbb{E}\hat{T}_{n-1} + \ldots + A_k \cdot \mathbb{E}\hat{T}_{n-k} + 1 \end{aligned} \tag{2.10}$$

where $\mathbb{E}\hat{T}_m = 0$ for $m \leq 0$ and the last term in (2.10) is obtained from the relation $\sum_{i=0}^{k} A_i = 1$.

The general solution of (2.10) is given by the sum of a homogeneous solution and a special solution. A special solution for the non-homogeneous recursive equation (2.10) is linear $D \cdot n$ where after some algebra $D = 1/(A_1 + 2A_2 + \ldots + kA_k)$, which is the inverse of the expected number of links succeeding in a given instant. Therefore $D = 1/(k - \sum_{i=1}^{k} q_i)$, independent of any statistical dependence between erasures on different links.

The homogeneous solution of linear recurrence relation with constant coefficients (2.10) can be expressed in terms of the roots of the characteristic equation $p(x) = (1-A_0)x^k - A_1x^{k-1} - \ldots - A_k$ [50, Section 3.2]. We will prove that the characteristic equation has $x = 1$ as a root and all the other roots have absolute value less than 1. Indeed since $A_0 + \ldots + A_k = 1 \Rightarrow (1 - A_0) - A_1 - \ldots - A_k = 0$, therefore $x = 1$ is a root of $p(x)$; now assume that $x = 1$ is a multiple root of $p(x)$. Then

$$\begin{aligned} p'(1) = 0 &\Leftrightarrow k(1 - A_0) - (k-1)A_1 - \ldots - A_{k-1} = 0 \\ &\Leftrightarrow k(1 - A_0) - (k-1)A_1 - \ldots - (k - (k-1))A_{k-1} = 0 \\ &\Leftrightarrow k = k(A_0 + A_1 + \ldots + A_{k-1}) - A_1 - 2A_2 - \ldots - (k-1)A_{k-1} \\ &\Leftrightarrow k = k(1 - A_k) - A_1 - 2A_2 - \ldots - (k-1)A_{k-1} \\ &\Leftrightarrow k = k - (A_1 + 2A_2 + \ldots + kA_k) \\ &\Leftrightarrow 0 = A_1 + 2A_2 + \ldots + kA_k \\ &\Leftrightarrow k = p_1 + p_2 + \ldots + p_k \end{aligned}$$

This implies that all links fail with probability 1, which contradicts the assumption from Section 2.2 that no link fails with probability 1. Assume now that characteristic equation $p(x)$ has a complex root $x = r \cdot e^{i \cdot \phi}$ where $|x| > 1$ or equivalently $r > 1$. Define $f(x) = x^k$ and $g(x) = A_0x^k + A_1x^{k-1} + \ldots + A_k$ then $p(x) = 0$ is equivalent to $f(x) = g(x)$ but this last equality cannot hold since $|g(x)| < |f(x)|$ for $|x| > 1$. Indeed $|g(x)| \leq A_0|x|^k + A_1|x|^{k-1} + \ldots + A_k = A_0r^k + A_1r^{k-1} + \ldots + A_k < (A_0 + A_1 + \ldots + A_k)r^k = r^k = |f(x)|$.

Let $R = \{r : p(r) = 0\}$ be the set of all roots of $p(x)$. The general solution for recursion formula (2.10) is

$$\mathbb{E}\hat{T}_n^c = \frac{n}{k - \sum_{i=1}^{k} q_i} + \sum_{r_j \in R} F_j r_j^n cos(n \cdot \phi_j) + \sum_{r_j \in R} G_j r_j^n sin(n \cdot \phi_j).$$

We can set

$$B_n = \sum_{r_j \in R} F_j r_j^n cos(n \cdot \phi_j) + \sum_{r_j \in R} G_j r_j^n sin(n \cdot \phi_j) \tag{2.11}$$

and since $|B_n| \leq \sum_{r_j \in R} |F_j| + |G_j|$ this concludes our proof. □

Now we are ready to prove the following theorem for the k-parallel path network shown in Figure 2.2.

Proof of Theorem 2. As discussed in the proof of Theorem 1, by using the results of [45] we can interchange the position of the first link of each path with one of the worst links of the path without affecting the arrival process at the receiver $\mathcal{T}$. Therefore without loss of generality we will assume that the first link in each path is one of the worst links in the path. Also, as in the proof of Theorem 1, for brevity we refer to packets present at a node N_i that are innovative for the next hop node N_{i+1} as innovative packets at node N_i.

The time T_n^{c} taken to send n packets from source $\mathcal{S}$ to the destination $\mathcal{T}$ in Figure 2.2 can be expressed as the sum of the time $\hat{T}_n^{\mathrm{c}}$ required for all n packets to reach one of nodes $N_{11}, \ldots, N_{k1}$ and the remaining time $\tilde{T}_n^{\mathrm{c}}$ required for all innovative packets remaining in the network to reach the destination $\mathcal{T}$, *i.e.*

$$T_n^{\mathrm{c}} = \hat{T}_n^{\mathrm{c}} + \tilde{T}_n^{\mathrm{c}}. \tag{2.12}$$

As in the proof of Theorem 1 all quantities in equation (2.12) are random variables and we want to compute their expected values. Due to linearity of expectation,

$$\mathbb{E}T_n^{\mathrm{c}} = \mathbb{E}\hat{T}_n^{\mathrm{c}} + \mathbb{E}\tilde{T}_n^{\mathrm{c}}, \tag{2.13}$$

where by Proposition 2,

$$\mathbb{E}\hat{T}_n^{\mathrm{c}} = \frac{n}{k - \sum_{i=1}^{k} p_{i1}} + B_n \tag{2.14}$$

where B_n is bounded. This holds regardless of any statistical dependence between the erasure processes on the first link of each path, and the remainder of the proof is unaffected by any statistical dependence between erasure processes on different paths.

The time $\mathbb{E}\tilde{T}_n^{\mathrm{c}}$ required to send all the remaining innovative packets at nodes N_{ij} ($i \in \{1, \ldots, k\}$, $j \in \{2, \ldots, \ell - 1\}$) to the destination is less than the expected time $\mathbb{E}\tilde{\tau}$ it would have taken if all the remaining innovative packets were returned back to the source $\mathcal{S}$ and sent to the destination $\mathcal{T}$

using only the first path. Let R_{ij} denote the number of remaining innovative packets at node N_{ij} at the moment the n^{th} packet has arrived at one of the k nodes $N_{11}, \ldots, N_{k1}$. Then the total number of remaining innovative packets R is $R = \sum_{i=1}^{k} \sum_{j=1}^{\ell-1} R_{ij}$ and the expected time $\mathbb{E}\tilde{\tau}$ is upper bounded by

$$\mathbb{E}\tilde{\tau} = \mathbb{E}\left[\mathbb{E}\left(\tilde{\tau}|R\right)\right] \leq \sum_{j=1}^{\ell} \frac{\mathbb{E}R}{1-p_{1j}}. \tag{2.15}$$

where $\mathbb{E}R/(1-p_{1j})$ is the expected time taken for R packets to cross the j^{th} hop in the first path.

By combining the fact that $\mathbb{E}\tilde{T}_n^{\text{c}} \leq \mathbb{E}\tilde{\tau}$ with equations (2.13) and (2.14) we get

$$\mathbb{E}T_n^{\text{c}} = \frac{n}{k - \sum_{i=1}^{k} p_{i1}} + D_n^{\text{c}} \tag{2.16}$$

where D_n^{c} is upper bounded by

$$D_n^{\text{c}} \leq B_n + \sum_{j=1}^{\ell} \frac{\mathbb{E}R}{1-p_{1j}}.$$

By Proposition 1, the number of remaining innovative packets at each node of each path is a stochastically increasing random variable with respect to n. Therefore, the expected number of remaining packets is an increasing function of n. Consequently one can find an upper bound on $\mathbb{E}R_{ij}$ by examining the line network in steady state, or equivalently, as $n \to +\infty$. For the case where the first link of each path is the unique worst link of the path, as shown in [48], each line network will reach steady state and consequently $E(R) \in \mathsf{O}(1)$. If there are multiple worst links in at least one path, then $\mathbb{E}R \in \mathsf{O}(\sqrt{n})$. This can be seen by interchanging the positions of links such that the worst links of each path are positioned at the start. By the results of [27], the number of innovative packets remaining at nodes positioned between two such worst links is $\mathsf{O}(\sqrt{n})$. By the results of [48], the number of innovative packets remaining at other intermediate network nodes is $\mathsf{O}(1)$.

Substituting p_{1i} with $\max_{1\leq j\leq \ell} p_{ij}$ for $i \in \{1, \ldots, k\}$ in equation (2.16) concludes the proof. □

2.4.2 Routing Strategy

In this section we analyze the expected time $\mathbb{E}T_n^{\text{r}}$ taken to send n packets through the parallel path network in Figure 2.2 using routing (where the r superscript stands for routing). We first prove the following two propositions.

Proposition 3. *For $a, b, c_1, c_2 \in \mathbb{N}^+$ with $a < b$ the sum $\sum_{m=a}^{b} \frac{c_1 - m}{c_2 + m}$ is equal to:*

$$\sum_{m=a}^{b} \frac{c_1 - m}{c_2 + m} = a - b - 1 + (c_1 + c_2)\left(H_{c_2+b} - H_{c_2+a-1}\right) \tag{2.17}$$

where H_n is the n^{th} Harmonic number, i.e. $H_n = \sum_{i=1}^{n} \frac{1}{i}$.

Proof.

$$\sum_{m=a}^{b} \frac{c_1 - m}{c_2 + m} = c_1 \sum_{m=a}^{b} \frac{1}{c_2 + m} - \sum_{m=a}^{b} \frac{m}{c_2 + m} = c_1\left(H_{c_2+b} - H_{c_2+a-1}\right) - \sum_{m=a}^{b} \frac{m}{c_2 + m} \tag{2.18}$$

Where $\sum_{m=a}^{b} \frac{m}{c_2 + m}$ can be evaluated as follows:

$$\begin{aligned} b - a + 1 &= \sum_{m=a}^{b} \frac{c_2 + m}{c_2 + m} \\ \Leftrightarrow b - a + 1 &= c_2 \sum_{m=a}^{b} \frac{1}{c_2 + m} + \sum_{m=a}^{b} \frac{m}{c_2 + m} \\ \Leftrightarrow \sum_{m=a}^{b} \frac{m}{c_2 + m} &= b - a + 1 - c_2\left(H_{c_2+b} - H_{c_2+a-1}\right) \end{aligned} \tag{2.19}$$

So from equations (2.18) and (2.19) we conclude that:

$$\sum_{m=a}^{b} \frac{c_1 - m}{c_2 + m} = a - b - 1 + (c_1 + c_2)\left(H_{c_2+b} - H_{c_2+a-1}\right)$$

□

Consider the network of Figure 2.3 with $k = 2$ parallel erasure links. As shown in equation (2.9) in order to minimize the expected completed time the routing strategy sends $\frac{n(1-q_1)}{2-q_1-q_2}$ packets over the first link and $\frac{n(1-q_2)}{2-q_1-q_2}$ packets over the second link. Proposition 4 examines this expected transmission time under routing.

Proposition 4. *The expected time $\mathbb{E}\hat{T}_n^r$ taken to send by routing n packets from the source to the destination through two parallel erasure links with probabilities of erasure q_1 and q_2 respectively is*

$$\mathbb{E}\hat{T}_n^r = \frac{n}{2 - q_1 - q_2} + U_n^{q_1,q_2}$$

where $U_n^{q_1,q_2}$ is an unbounded term that grows at least as square root of n. The term routing means

that out of the n packets, $\frac{n(1-q_1)}{2-q_1-q_2}$ *packets are transmitted through the link with* q_1 *probability of erasure and* $\frac{n(1-q_2)}{2-q_1-q_2}$ *packets through the link with* q_2 *probability of erasure.*

Proof. Denote by $A_{i,j}$ the expected time to send i packets over the link with erasure probability q_1 and j packets over the link with erasure probability q_2. Clearly $\mathbb{E}\hat{T}_n^{\mathrm{r}} = A_{i,j}$ with $i = \frac{n(1-q_1)}{2-q_1-q_2}$, $j = \frac{n(1-q_2)}{2-q_1-q_2}$. $A_{i,j}$ satisfies the following two dimensional recursion formula:

$$\left\{ \begin{array}{c} A_{i,j} = q_1q_2[A_{i,j}+1] + (1-q_1)q_2[A_{i-1,j}+1] \\ +q_1(1-q_2)[A_{i,j-1}+1] + (1-q_1)(1-q_2)[A_{i-1,j-1}+1] \\ A_{i,0} = \frac{i}{1-q_1}, \quad A_{0,j} = \frac{j}{1-q_2}, \quad A_{0,0} = 0 \end{array} \right\}$$

or equivalently

$$\left\{ \begin{array}{c} (1-q_1q_2)A_{i,j} = (1-q_1)q_2A_{i-1,j} + q_1(1-q_2)A_{i,j-1} \\ +(1-q_1)(1-q_2)A_{i-1,j-1} + 1 \\ A_{i,0} = \frac{i}{1-q_1}, \quad A_{0,j} = \frac{j}{1-q_2}, \quad A_{0,0} = 0 \end{array} \right\}. \tag{2.20}$$

The two dimensional recursion formula in (2.20) has a specific solution $\frac{i}{2(1-q_1)} + \frac{j}{2(1-q_2)}$ and a general solution $B_{i,j}$ where

$$\left\{ \begin{array}{c} (1-q_1q_2)B_{i,j} = (1-q_1)q_2B_{i-1,j} + q_1(1-q_2)B_{i,j-1} \\ +(1-q_1)(1-q_2)B_{i-1,j-1}, i,j \geq 1 \\ B_{i,0} = \frac{i}{2(1-q_1)}, \quad B_{0,j} = \frac{j}{2(1-q_2)}, \quad B_{0,0} = 0 \end{array} \right\}. \tag{2.21}$$

In order to solve equation (2.21) we will use the Z–transform with respect to i. More specifically we define the Z–transform as:

$$\hat{B}_{z,j} = \sum_{i=0}^{\infty} B_{i,j} \cdot z^i. \tag{2.22}$$

By multiplying all terms in equation (2.21) by z^i and summing over i we get:

$$\begin{aligned} (1-q_1q_2)\sum_{i=1}^{\infty} B_{i,j} \cdot z^i &= (1-q_1)q_2\sum_{i=1}^{\infty} B_{i-1,j} \cdot z^i + q_1(1-q_2)\sum_{i=1}^{\infty} B_{i,j-1} \cdot z^i \\ &\quad +(1-q_1)(1-q_2)\sum_{i=1}^{\infty} B_{i-1,j-1} \cdot z^i \\ \Leftrightarrow (1-q_1q_2)\left[\hat{B}_{z,j} - B_{0,j}\right] &= z(1-q_1)q_2\hat{B}_{z,j} + q_1(1-q_2)\left[\hat{B}_{z,j-1} - B_{0,j-1}\right] \\ &\quad +z(1-q_1)(1-q_2)\hat{B}_{z,j-1}. \end{aligned}$$

Table 2.2: Some pairs of functions along with their Z–transforms

Sequence	Z–transform
1	$\frac{1}{1-z}$
i	$\frac{z}{(1-z)^2}$
$\frac{\binom{i+j-t-1}{j-1}}{b^{i+j-t}}$	$\frac{z^t}{(b-z)^j}$, for $t \leq j$

Since $B_{0,j} = \frac{j}{2(1-q_2)}$ the above equation becomes:

$$\left\{ \begin{array}{c} \left[(1-q_1q_2) - z(1-q_1)q_2\right] \hat{B}_{z,j} = \left[q_1(1-q_2) + z(1-q_1)(1-q_2)\right] \hat{B}_{z,j-1} \\ +j\frac{1-q_1}{2(1-q_2)} + \frac{q_1}{2} \\ \hat{B}_{z,0} = \sum_{i=0}^{\infty} B_{i,0} z^i = \sum_{i=0}^{\infty} \frac{i}{2(1-q_1)} z^i = \frac{z}{2(1-q_1)(1-z)^2} \end{array} \right\} \tag{2.23}$$

where equation (2.23) is an one dimensional recursion formula with the following general solution [50, Section 3.2]:

$$\begin{aligned} \hat{B}_{z,j} = \frac{z}{(1-q_1)(1-z)^2} & \left[\frac{q_1(1-q_2) + z(1-q_1)(1-q_2)}{1-q_1q_2 - z(1-q_1)q_2}\right]^j \\ & + \frac{j}{2(1-q_2)(1-z)} - \frac{z}{2(1-q_1)(1-z)^2}. \end{aligned} \tag{2.24}$$

Equation (2.24) can be written in a compact form

$$\hat{B}_{z,j} = \hat{a}(z) \cdot \hat{b}(j,z) + \hat{d}(j,z) \tag{2.25}$$

by defining the functions $\hat{a}(z)$, $\hat{b}(z,j)$ and $\hat{d}(z,j)$ as follows:

$$\begin{aligned} \hat{a}(z) &= \frac{z}{(1-q_1)(1-z)^2} \\ \hat{b}(z,j) &= \left[\frac{q_1(1-q_2) + z(1-q_1)(1-q_2)}{1-q_1q_2 - z(1-q_1)q_2}\right]^j \\ \hat{d}(z,j) &= \frac{j}{2(1-q_2)(1-z)} - \frac{z}{2(1-q_1)(1-z)^2}. \end{aligned}$$

Now we are ready to compute the inverse Z–transform of $\hat{B}_{z,j}$. Using Table 2.2 along with equation (2.25):

$$B_{i,j} = Z^{-1}\left\{\hat{a}(z) \cdot \hat{b}(z,j)\right\} + Z^{-1}\left\{\hat{d}(z,j)\right\}$$

$$\Leftrightarrow B_{i,j} = \sum_{m=0}^{i} a(i-m)\cdot b(m,j) + \frac{j}{2(1-q_2)} - \frac{i}{2(1-q_1)}$$

where $a(i)$ and $b(i,j)$ are the inverse Z–transforms of $\hat{a}(z)$ and $\hat{b}(z,j)$ respectively. From Table 2.2 $a(i) = \frac{i}{1-q_1}$ and therefore the equation above becomes

$$B_{i,j} = \sum_{m=0}^{i} \frac{i-m}{1-q_1} b(m,j) + \frac{j}{2(1-q_2)} - \frac{i}{2(1-q_1)}. \tag{2.26}$$

The remaining step in order to compute $B_{i,j}$ is to evaluate $b(i,j)$:

$$\begin{aligned}
b(i,j) &= Z^{-1}\left\{\left[\frac{q_1(1-q_2)+z(1-q_1)(1-q_2)}{1-q_1q_2-z(1-q_1)q_2}\right]^j\right\} \\
&= \frac{1}{[(1-q_1)q_2]^j}\cdot Z^{-1}\left\{\frac{\sum_{t=0}^{j}\binom{j}{t} z^t(1-q_1)^t(1-q_2)^t[q_1(1-q_2)]^{j-t}}{\left(\frac{1-q_1q_2}{(1-q_1)q_2}-z\right)^j}\right\} \\
&= \left[\frac{q_1(1-q_2)}{q_2(1-q_1)}\right]^j \sum_{t=0}^{j}\binom{j}{t}\cdot\left(\frac{1-q_1}{q_1}\right)^t \cdot Z^{-1}\left\{\frac{z^t}{\left(\frac{1-q_1q_2}{(1-q_1)q_2}-z\right)^j}\right\} \\
&= \frac{(q_1(1-q_2))^j((1-q_1)q_2)^i}{(1-q_1q_2)^{i+j}}\sum_{t=0}^{j}\binom{j}{t}\binom{i+j-t-1}{j-1}\left(\frac{1-q_1q_2}{q_1q_2}\right)^t.
\end{aligned}$$

Therefore equation (2.26) becomes

$$B_{i,j} = \left[\frac{q_1(1-q_2)}{1-q_1q_2}\right]^j \sum_{m=0}^{i}\sum_{t=0}^{j}\frac{i-m}{1-q_1}\left[\frac{(1-q_1)q_2}{1-q_1q_2}\right]^m\binom{j}{t}\binom{m+j-t-1}{j-1}\left(\frac{1-q_1q_2}{q_1q_2}\right)^t + \frac{j}{2(1-q_2)} - \frac{i}{2(1-q_1)}$$

and since the expected time $A_{i,j} = B_{i,j} + \frac{i}{2(1-q_1)} + \frac{j}{2(1-q_2)}$ then

$$A_{i,j} = \left[\frac{q_1(1-q_2)}{1-q_1q_2}\right]^j \sum_{m=0}^{i}\sum_{t=0}^{j}\frac{i-m}{1-q_1}\left[\frac{(1-q_1)q_2}{1-q_1q_2}\right]^m\binom{j}{t}\binom{m+j-t-1}{j-1}\left(\frac{1-q_1q_2}{q_1q_2}\right)^t + \frac{j}{1-q_2}. \tag{2.27}$$

We are interested in evaluating $\mathbb{E}\hat{T}_n^{\mathrm{r}} = A_{i,j}$ for $i = \frac{n(1-q_1)}{2-q_1-q_2}$ and $j = \frac{n(1-q_2)}{2-q_1-q_2}$ and therefore from equation (2.27) we get

$$\mathbb{E}\hat{T}_n^{\mathrm{r}} = \frac{n}{2-q_1-q_2} + U_n^{q_1,q_2}$$

where

$$U_n^{q_1,q_2}=\left[\frac{q_1(1-q_2)}{1-q_1q_2}\right]^{\frac{n(1-q_2)}{2-q_1-q_2}}\sum_{m=0}^{\frac{n(1-q_1)}{2-q_1-q_2}}\sum_{t=0}^{\frac{n(1-q_2)}{2-q_1-q_2}}\frac{\frac{n(1-q_1)}{2-q_1-q_2}-m}{1-q_1}\left[\frac{(1-q_1)q_2}{1-q_1q_2}\right]^m\binom{\frac{n(1-q_2)}{2-q_1-q_2}}{t}\binom{m+\frac{n(1-q_2)}{2-q_1-q_2}-t-1}{\frac{n(1-q_2)}{2-q_1-q_2}-1}$$

with $\binom{m}{w}=0$ if $m<w$. If we define $W=\frac{(1-q_1)q_2}{1-q_1q_2}$, $E=\frac{q_1(1-q_2)}{1-q_1q_2}$ and $F=\frac{1-q_1q_2}{q_1q_2}$, then the above expression can be written more compactly as

$$U_n^{q_1,q_2}=E^{\frac{n(1-q_2)}{2-q_1-q_2}}\sum_{m=0}^{\frac{n(1-q_1)}{2-q_1-q_2}}\sum_{t=0}^{\frac{n(1-q_2)}{2-q_1-q_2}}\frac{\frac{n(1-q_1)}{2-q_1-q_2}-m}{1-q_1}\binom{\frac{n(1-q_2)}{2-q_1-q_2}}{t}\binom{\frac{n(1-q_2)}{2-q_1-q_2}+m-t-1}{\frac{n(1-q_2)}{2-q_1-q_2}-1}W^mF^t.$$

In order to prove that function $U_n^{q_1,q_2}$ is unbounded we will prove that $U_n^{q_1,q_2}$ is larger than another simpler to analyze function that goes to infinity and therefore $U_n^{q_1,q_2}$ also increases to infinity. Indeed the equation above can be written as

$$\begin{aligned}U_n^{q_1,q_2}&=E^{\frac{n(1-q_2)}{2-q_1-q_2}}\sum_{m=0}^{\frac{n(1-q_1)}{2-q_1-q_2}}\sum_{t=0}^{\frac{n(1-q_2)}{2-q_1-q_2}}\frac{\frac{n(1-q_1)}{2-q_1-q_2}-m}{1-q_1}\binom{\frac{n(1-q_2)}{2-q_1-q_2}}{t}\binom{\frac{n(1-q_2)}{2-q_1-q_2}+m-t}{\frac{n(1-q_2)}{2-q_1-q_2}}\frac{\frac{n(1-q_2)}{2-q_1-q_2}}{\frac{n(1-q_2)}{2-q_1-q_2}+m-t}W^mF^t\\&>\frac{n(1-q_2)E^{\frac{n(1-q_2)}{2-q_1-q_2}}}{(1-q_1)(2-q_1-q_2)}\sum_{m=0}^{\frac{n(1-q_1)}{2-q_1-q_2}}\sum_{t=0}^{\frac{n(1-q_2)}{2-q_1-q_2}}\binom{\frac{n(1-q_2)}{2-q_1-q_2}}{t}\binom{\frac{n(1-q_2)}{2-q_1-q_2}+m-t}{\frac{n(1-q_2)}{2-q_1-q_2}}\frac{\frac{n(1-q_1)}{2-q_1-q_2}-m}{\frac{n(1-q_2)}{2-q_1-q_2}+m}W^mF^t\end{aligned}$$

and since all terms in the above double sum are non-negative we can disregard as many terms as we wish without violating direction of the inequality, specifically

$$U_n^{q_1,q_2}>\frac{n(1-q_2)E^{\frac{n(1-q_2)}{2-q_1-q_2}}}{(1-q_1)(2-q_1-q_2)}\sum_{m\in J,t\in G}\binom{\frac{n(1-q_2)}{2-q_1-q_2}}{t}\binom{\frac{n(1-q_2)}{2-q_1-q_2}+m-t}{\frac{n(1-q_2)}{2-q_1-q_2}}\frac{\frac{n(1-q_1)}{2-q_1-q_2}-m}{\frac{n(1-q_2)}{2-q_1-q_2}+m}W^mF^t \quad (2.28)$$

where $J=\{\lceil\frac{n(1-q_1)}{2-q_1-q_2}(1-\frac{1}{\sqrt{n}})\rceil,\ldots,\frac{n(1-q_1)}{2-q_1-q_2}\}$, $G=\{\lceil(1-q_1)\frac{n(1-q_2)}{2-q_1-q_2}(1-\frac{1}{\sqrt{n}})\rceil,\ldots,\lfloor(1-q_1)\frac{n(1-q_2)}{2-q_1-q_2}\rfloor\}$ and $\lfloor x\rfloor$, $\lceil x\rceil$ are the floor and the ceiling functions respectively.

By using the lower and upper Stirling-based bound [51]:

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n<n!<\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n}},\quad n\geq 1$$

one can find that

$$\binom{n}{\beta n}>\frac{1}{\sqrt{2\pi\beta(1-\beta)n}}\cdot 2^{nH(\beta)}\cdot e^{-\frac{1}{12n\beta(1-\beta)}},\ \beta\in(0,1)$$

and

$$\binom{\bar{\beta}n}{n} > \sqrt{\frac{\bar{\beta}}{2\pi(\bar{\beta}-1)n}} \cdot 2^{n\bar{\beta}H\left(\frac{1}{\bar{\beta}}\right)} \cdot e^{-\frac{\bar{\beta}}{12n(\bar{\beta}-1)}},\ \bar{\beta} > 1$$

where $H(\beta) = -\beta\log_2(\beta) - (1-\beta)\log_2(1-\beta)$ is the entropy function and therefore using inequality (2.28) we can derive:

$$U_n^{q_1,q_2} > \frac{1}{2\pi(1-q_1)} \sum_{m\in J, t\in G} \frac{\frac{n(1-q_1)}{2-q_1-q_2} - m}{\frac{n(1-q_2)}{2-q_1-q_2} + m} f\left(\frac{m}{M}, \frac{t}{T}\right) e^{-\frac{2-q_1-q_2}{12n(1-q_2)} h\left(\frac{m}{M}, \frac{t}{T}\right)} 2^{\frac{n(1-q_2)}{2-q_1-q_2} g\left(\frac{m}{M}, \frac{t}{T}\right)} \tag{2.29}$$

where $M = \frac{n(1-q_1)}{2-q_1-q_2}$, $T = \frac{n(1-q_2)}{2-q_1-q_2}$, $f(\alpha,\beta) = \sqrt{\frac{1+\alpha\frac{1-q_1}{1-q_2}-\beta}{\beta(1-\beta)(\alpha\frac{1-q_1}{1-q_2}-\beta)}}$, $h(\alpha,\beta) = \frac{1+\alpha\frac{1-q_1}{1-q_2}-\beta}{\alpha\frac{1-q_1}{1-q_2}-\beta} + \frac{1}{\beta(1-\beta)}$ and

$$g(\alpha,\beta) = \log_2(E) + \alpha\frac{1-q_1}{1-q_2}\log_2(W) + H(\beta) + (1+\alpha\frac{1-q_1}{1-q_2}-\beta)H\left(\frac{1}{1+\alpha\frac{1-q_1}{1-q_2}-\beta}\right) + \beta\log_2(F).$$

Since $1-\frac{1}{\sqrt{n}} \le \frac{m}{M} \le 1$ and $(1-q_1)-\frac{1}{\sqrt{n}} \le \frac{t}{T} \le (1-q_1)$ we define functions $f(\alpha,\beta)$, $h(\alpha,\beta)$ and $g(\alpha,\beta)$ within the region $N = \left[1-\frac{1}{\sqrt{n}}, 1\right] \times \left[1-q_1-\frac{1}{\sqrt{n}}, 1-q_1\right]$. Moreover we are only concerned with large enough n so that $0 < \beta < \alpha$ and region N looks like the one in Figure 2.4. For large values of n, $f(\alpha,\beta) > \sqrt{\frac{1}{2q_1(1-q_1)}}$ and $h(\alpha,\beta) < 1 + \frac{2(1-q_2)}{(1-q_1)q_2} + \frac{2}{q_1(1-q_1)}$ within region N and therefore from inequality (2.29) we get:

$$\begin{aligned} U_n^{q_1,q_2} &> \frac{1}{\sqrt{8\pi^2 q_1(1-q_1)^3}} e^{-\frac{2-q_1-q_2}{12n(1-q_2)}\left(1+\frac{2(1-q_2)}{(1-q_1)q_2}+\frac{2}{q_1(1-q_1)}\right)} \sum_{m\in J, t\in G} \frac{\frac{n(1-q_1)}{2-q_1-q_2} - m}{\frac{n(1-q_2)}{2-q_1-q_2} + m} 2^{\frac{n(1-q_2)}{2-q_1-q_2} g\left(\frac{m}{M}, \frac{t}{T}\right)} \\ &> \frac{e^{-1}}{\sqrt{8\pi^2 q_1(1-q_1)^3}} \sum_{m\in J, t\in G} \frac{\frac{n(1-q_1)}{2-q_1-q_2} - m}{\frac{n(1-q_2)}{2-q_1-q_2} + m} 2^{\frac{n(1-q_2)}{2-q_1-q_2} g\left(\frac{m}{M}, \frac{t}{T}\right)} \end{aligned} \tag{2.30}$$

for large enough n.

Function $g(\alpha,\beta)$ satisfies the following three conditions:

1. $\frac{\partial g}{\partial \alpha} = \frac{1-q_1}{1-q_2}\log_2\left(W\frac{\alpha(1-q_1)+(1-\beta)(1-q_2)}{\alpha(1-q_1)-\beta(1-q_2)}\right)$ and $\frac{\partial g}{\partial \beta} = \log_2\left(\frac{F(1-\beta)[\alpha(1-q_1)-\beta(1-q_2)]}{\beta[\alpha(1-q_1)+(1-\beta)(1-q_2)]}\right)$
2. $\frac{\partial^2 g}{\partial \alpha^2} = -\frac{(1-q_1)^2}{[\alpha(1-q_1)-\beta(1-q_2)][\alpha(1-q_1)+(1-\beta)(1-q_2)]\ln 2} < 0$
3. $\frac{\partial^2 g}{\partial \alpha^2}\cdot\frac{\partial^2 g}{\partial \beta^2} - \frac{\partial^2 g}{\partial \alpha \partial \beta}\cdot\frac{\partial^2 g}{\partial \beta \partial \alpha} = \frac{(1-q_1)^2}{\beta(1-\beta)[\alpha(1-q_1)+(1-\beta)(1-q_2)][\alpha(1-q_1)-\beta(1-q_2)](\ln 2)^2} > 0$

It's easy to see from condition 1 that $\left.\frac{\partial g(\alpha,\beta)}{\partial \alpha}\right|_{(1,1-q_1)} = 0$ and $\left.\frac{\partial g(\alpha,\beta)}{\partial \beta}\right|_{(1,1-q_1)} = 0$. Moreover conditions 2 and 3 show the concavity of $g(\alpha,\beta)$ within region N and along with condition 1 it is proved that function $g(\alpha,\beta)$ achieves a maximum at point $(\alpha,\beta) = (1, 1-q_1)$. Therefore $g(\alpha,\beta) \le g(1,1-q_1) = 0$ making the exponent of 2 in (2.30) non-positive guaranteeing an exponential decay of each term in the sum. Since region N is compact (closed and convex) and function $g(\alpha,\beta)$ is

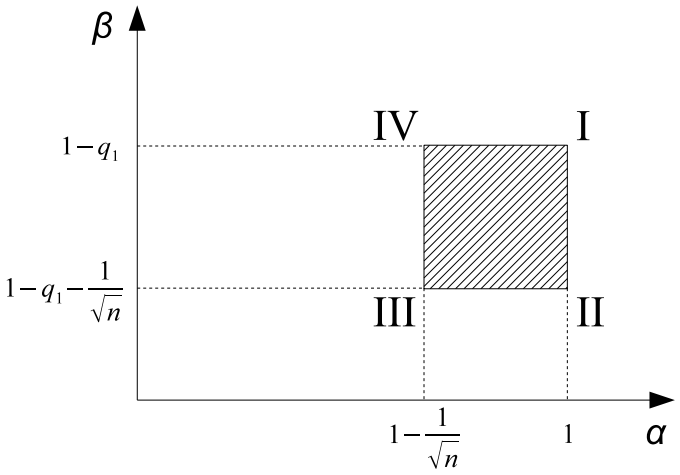


Figure 2.4: The region N where function $g(\alpha, \beta)$ is defined on.

concave, and therefore it will achieve its minimum on the boundary of N. It's not difficult to show that $\frac{\partial g(\alpha,1-q_1)}{\partial \alpha} \geq 0$ for $\alpha \leq 1$ and therefore function $g(\alpha, 1-q_1)$ decreases in value from point I to point IV. Similarly $\frac{\partial g(1,\beta)}{\partial \beta} \geq 0$ for $\beta \leq 1-q_1$ and therefore function $g(1,\beta)$ decreases in value from point I to point II. Since $\frac{\partial g(\alpha,1-q_1-1/\sqrt{n})}{\partial \alpha} \geq 0$ for $a \leq 1$ and $\frac{\partial g(1-1/\sqrt{n},\beta)}{\partial \beta} \geq 0$ for $\beta \leq 1-q_1$ with similar arguments as above we show that the minimum value for $g(\alpha, \beta)$ within N is achieved at point $C \equiv (\alpha_m, \beta_m) = (1-\frac{1}{\sqrt{n}}, 1-q_1-\frac{1}{\sqrt{n}})$. Therefore $g\left(\frac{k}{n}, \frac{i}{n}\right) \geq g\left(\alpha_m, \beta_m\right)$ or else from equation (2.30):

$$U_n^{q_1,q_2} > \frac{e^{-1}(1-q_2)\sqrt{n}}{(2-q_1-q_2)\sqrt{8\pi^2 q_1(1-q_1)}} 2^{\frac{n(1-q_2)}{2-q_1-q_2} g(a_m,\beta_m)} \sum_{m \in J} \frac{\frac{n(1-q_1)}{2-q_1-q_2} - m}{\frac{n(1-q_2)}{2-q_1-q_2} + m}$$

Using the Taylor expansion of function $r(x) = g(1-x, 1-q_1-x)$ around $x = 0$ we get the following expression:

$$f(x) = \frac{q_1^2(q_2-q_1) - q_2(1-q_1^2)}{(1-q_1)q_1q_2(1-q_1q_2)\ln 2} x^2 + \mathsf{O}(x^3).$$

For $x = \frac{1}{\sqrt{n}}$ we get

$$\frac{n(1-q_2)}{2-q_1-q_2} g\left(\alpha_m, \beta_m\right) = \frac{(1-q_2)\left(q_1^2(q_2-q_1) - q_2(1-q_1^2)\right)}{(2-q_1-q_2)(1-q_1)q_1q_2(1-q_1q_2)\ln 2} + \mathsf{O}\left(\frac{1}{\sqrt{n}}\right)$$

where along with Proposition 3 we get

$$U_n^{q_1,q_2} > \frac{e^{-1}(1-q_2)\sqrt{n}}{(2-q_1-q_2)\sqrt{8\pi^2 q_1(1-q_1)}} 2^{\frac{(1-q_2)\left(q_1^2(q_2-q_1)-q_2(1-q_1^2)\right)}{(2-q_1-q_2)(1-q_1)q_1q_2(1-q_1q_2)\ln 2} + \frac{c}{\sqrt{n}} t(n)} \tag{2.31}$$

where $t(n) = n\left(H_n - H_{n-k(n)-1}\right) - k(n) - 1$ and $k(n) = A\sqrt{n}$ with $A = \frac{(1-q_1)}{2-q_1-q_2}$. The above expression can be simplified by using the bounds proved by Young in [52]:

$$\ln n + \gamma + \frac{1}{2(n+1)} < H_n < \ln n + \gamma + \frac{1}{2n}$$

where γ is the Euler's constant. We obtain from (2.31):

$$U_n^{q_1,q_2} > \frac{e^{-1}(1-q_2)\sqrt{n}}{(2-q_1-q_2)\sqrt{8\pi^2 q_1(1-q_1)}} 2^{\frac{(1-q_2)\left(q_1^2(q_2-q_1)-q_2(1-q_1^2)\right)}{(2-q_1-q_2)(1-q_1)q_1q_2(1-q_1q_2)\ln 2}+\frac{c}{\sqrt{n}}}\phi(n) \tag{2.32}$$

where $\phi(n) = n\ln\left(\frac{n}{n-k(n)-1}\right) - \frac{n}{2(n+1)}\frac{k(n)+2}{n-k(n)-1} - k(n) - 1$. It can be easily proved that function $\omega(n) = n\ln\left(\frac{n}{n-k(n)-1}\right) - k(n) - 1$ is greater than $\frac{A^2}{2}$ for $n > 1$. Indeed

$$\omega''(n) = \frac{A(A^2+3)n + 2(A^2+2)\sqrt{n} + A}{4(n - A\sqrt{n} - 1)^2 n^{3/2}} > 0 \quad \text{for} \quad n > 1 \tag{2.33}$$

and since $\lim_{n\to+\infty}\omega'(n) = 0$ it means that $\omega'(n) < 0$ for $n > 1$ and therefore $\omega(n)$ is a decreasing function of $n > 1$. Moreover

$$\lim_{n\to+\infty}\omega(n) = \lim_{n\to+\infty}\frac{\ln\left(\frac{n}{n-k(n)-1}\right) - \frac{k(n)}{n}}{\frac{1}{n}} - 1 \overset{\text{L'Hospital}}{=} \lim_{n\to+\infty}\frac{\frac{k(n)}{n^2} + \frac{k^2(n)}{n^2} + \frac{2}{n}}{-\frac{1}{n^2}(2+2k(n)-2n)} - 1 = \frac{A^2}{2}$$

and therefore $\omega(n) > \frac{A^2}{2}$ for $n > 1$. Finally inequality (2.32) becomes

$$U_n^{q_1,q_2} > \frac{e^{-1}(1-q_2)\sqrt{n}}{(2-q_1-q_2)\sqrt{8\pi^2 q_1(1-q_1)}} 2^{\frac{(1-q_2)\left(q_1^2(q_2-q_1)-q_2(1-q_1^2)\right)}{(2-q_1-q_2)(1-q_1)q_1q_2(1-q_1q_2)\ln 2}+\frac{c}{\sqrt{n}}}\left(\frac{1}{2}\left(\frac{1-q_1}{2-q_1-q_2}\right)^2 - \frac{n}{2(n+1)}\frac{k(n)+2}{n-k(n)-1}\right)$$

Clearly the above function is unbounded and $U_n^{q_1,q_2}$ increases with respect to n at least as $\sqrt{n}$. □

Now we have all the necessary tools to prove the following theorem for k-parallel path multi-hop networks as shown in Figure 2.2.

Proof of Theorem 3. Without loss of generality due to [45] we can interchange the first link of each of the k line networks with the worst link of the line network. The first term in equation (2.2) is due to the capacity of the k parallel multi-hop line network. The second term D_n^{r} is sublinear in n; what is left to prove is that term D_n^{r} grows as $\Omega(\sqrt{n})$. This follows from Proposition 4. The number of packets transmitted on the first two paths is $n_1 = n\Big(1 - \max_{1\le i\le \ell} p_{1i}\Big)\Big/\Big(k - \sum_{i=1}^{k}\max_{1\le j\le \ell} p_{ij}\Big)$ and $n_2 = n\Big(1 - \max_{1\le i\le \ell} p_{2i}\Big)\Big/\Big(k - \sum_{i=1}^{k}\max_{1\le j\le \ell} p_{ij}\Big)$ respectively. The time T_n^{r} taken to send n packets through the k-parallel path multi-hop network is greater than the time $\hat{T}_n^{\mathrm{r}}$ taken for n_1 packets to

reach node N_{11} and n_2 packets to reach node N_{21}. Therefore from Proposition 4

$$\mathbb{E}T_n^{\mathrm{r}} > \frac{n}{k - \sum_{i=1}^{k} \max_{1\le j\le \ell} p_{ij}} + U_{n'}^{\max_{1\le i\le \ell} p_{1i}, \max_{1\le j\le \ell} p_{2j}}.$$

where $n' = n\Big(2 - \max_{1\le i\le \ell} p_{1i} - \max_{1\le i\le \ell} p_{2i}\Big)\Big/\Big(k - \sum_{i=1}^{k} \max_{1\le j\le \ell} p_{ij}\Big)$ is proportional to n. By Proposition 4, $U_{n'}^{\max_{1\le i\le \ell} p_{1i}, \max_{1\le j\le \ell} p_{2j}}$ grows as $\Omega(\sqrt{n'})$. Thus, D_n^{r} grows as $\Omega(\sqrt{n})$. □

2.5 General Network Topologies

We next consider networks with general topologies.

Lemma 1. *In a single-bottleneck network, there exists a max-flow subgraph comprising paths each of which has a single worst link.*

Proof. Given a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with a single minimum cut, let $(v_1, w_1), \ldots, (v_k, w_k)$ be the edges crossing the minimum cut. Let $\mathcal{G}'$ be a max flow subgraph. Consider the network $\mathcal{G} - \mathcal{G}'$ obtained from $\mathcal{G}$ by reducing the capacity of each link $(i, j) \in \mathcal{E}$ by the capacity of the corresponding link in $\mathcal{G}'$ if any. There is a path from the source to each node $v_i, 1 \le i \le k$ (which may not all be distinct), otherwise this would contradict the assumption that there is a single minimum cut. Thus, we can find a subgraph $\mathcal{G}''$ comprising a set of paths of nonzero and nonoverlapping capacity from the source to each distinct node $v_i, 1 \le i \le k$. Similarly, we can find a subgraph $\mathcal{G}'''$ comprising a set of paths of nonzero and nonoverlapping capacity from each distinct node $w_i, 1 \le i \le k$, to the sink. We can then decompose the union of subgraphs $\mathcal{G}' + \mathcal{G}'' + \mathcal{G}'''$ (obtained by adding the capacities of corresponding links) into a sufficiently large number of paths each of which has a single worst link corresponding to the min cut of the original network. □

Proof of Theorem 4. The expected time $\mathbb{E}T_n^{\mathrm{r}}$ required to send all n packets by routing through network $\mathcal{G}$ from source $\mathcal{S}$ to destination $\mathcal{T}$ is greater than the time $\mathbb{E}\breve{T}_n^{\mathrm{r}}$ it would take the n packets to cross the mincut of the network by routing. Specifically if we assume that all nodes on the source's side of the cut are collapsed into a super source node and all nodes on the sink's side of the cut are collapsed into a super destination node then the network becomes a parallel erasure links network as shown in Figure 2.3. Then

$$\mathbb{E}T_n^{\mathrm{r}} \ge \mathbb{E}\breve{T}_n^{\mathrm{r}} = \frac{n}{C} + D_n^{\mathrm{r}}$$

where $D_n^{\mathrm{r}} \in \Omega(\sqrt{n})$ by Theorem 3.

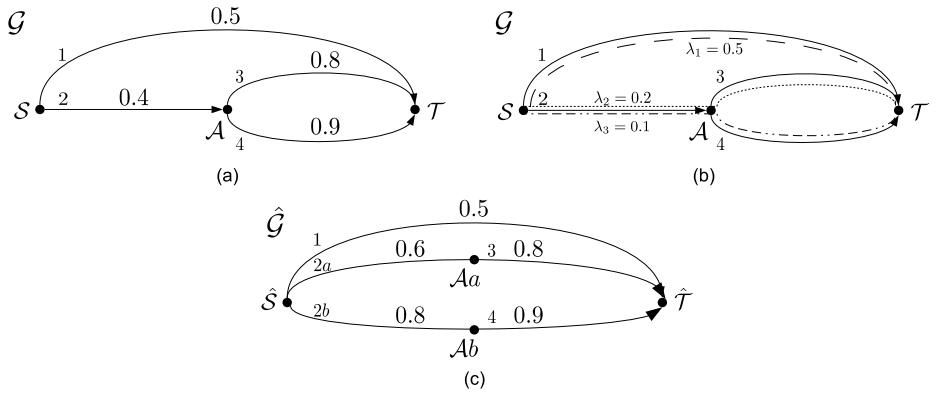


Figure 2.5: (a) Network $\mathcal{G}$ with a single source $\mathcal{S}$, a single destination $\mathcal{T}$, an intermediate node $\mathcal{A}$, and four erasure links 1, 2, 3, and 4 with probabilities of erasure 0.5, 0.4, 0.8, 0.9 respectively. (b) The solution of the linear program on network $\mathcal{G}$ would give us three rates $\lambda_1 = 0.5$, $\lambda_2 = 0.2$, and $\lambda_3 = 0.1$. (*c*) Network $\hat{\mathcal{G}}$ derived from the solution of the linear program

For the case of coding on a network $\mathcal{G}$, for any max-flow subgraph (composed of flows on paths from source $\mathcal{S}$ to destination $\mathcal{T}$), one can construct a parallel path network $\hat{\mathcal{G}}$ that requires at least as much time to send the n packets from the source to the destination.

Denote by $\mathcal{F}$ the set of source-sink flows in the max-flow subgraph. For each flow $f \in \mathcal{F}$, let λ_f denote the flow rate and let $\mathcal{P}_f$ denote the path of flow f. For each node $v \in \mathcal{V}$ in network $\mathcal{G}$, let $\mathcal{K}_v$ denote the set of flows passing through node v, where $\mathcal{K}_\mathcal{S}$ and $\mathcal{K}_\mathcal{T}$ are equal to the sets of all flows in network $\mathcal{G}$. For each edge $e \in \mathcal{E}$ let $\mathcal{F}_e$ denote the set of flows passing through edge e. For the example in Figure 2.5(b), $\mathcal{F} = \{1, 2, 3\}$, $\lambda_1 = 0.5$, $\lambda_2 = 0.2$, $\lambda_3 = 0.1$, $\mathcal{P}_1 = \mathcal{S} \to \mathcal{T}$ for flow 1, $\mathcal{P}_2 = \mathcal{S} \to \mathcal{A} \to \mathcal{T}$ for flow 2, and $\mathcal{P}_3 = \mathcal{S} \to \mathcal{A} \to \mathcal{T}$ for flow 3, $\mathcal{K}_\mathcal{A} = \{2, 3\}$, and $\mathcal{F}_1 = \{1\}$, $\mathcal{F}_2 = \{2, 3\}$, $\mathcal{F}_3 = \{2\}$, and $\mathcal{F}_4 = \{3\}$.

The process of creating network $\hat{\mathcal{G}} = (\hat{\mathcal{V}}, \hat{\mathcal{E}})$ from $\mathcal{G}$ is the following.

1. For every node $v \in \mathcal{G}$, create a set of nodes $\hat{\mathcal{V}}_v = \{\hat{v}_f : f \in \mathcal{K}_v\}$. The set of nodes $\hat{\mathcal{V}}$ is defined as $\bigcup_{v \in \mathcal{V}} \hat{\mathcal{V}}_v$.

2. The edges of network $\hat{\mathcal{G}}$ are created as follows. For each flow $f \in \mathcal{F}$ and for each edge (u, v) in path $\mathcal{P}_f$ of flow f, create an edge in network $\hat{\mathcal{G}}$ from $\hat{u}_f$ to $\hat{v}_f$ with probability of erasure

$$\hat{p}_{(\hat{u}_f, \hat{v}_f)} = 1 - \frac{\lambda_f}{\sum_{w \in \mathcal{F}_{(u,v)}} \lambda_w}(1 - p_{(u,v)})$$

   where $p_{(u,v)}$ is the probability of erasure of link (u, v) in network $\mathcal{G}$. Define a function $H\big((\hat{u}_f, \hat{v}_f)\big) = \frac{\lambda_f}{\sum_{w \in \mathcal{F}_{u,v)}} \lambda_w}$.

3. Collapse all nodes of set $\mathcal{V}_\mathcal{S}$ to a single node $\hat{\mathcal{S}}$ that denotes the source in network $\hat{\mathcal{G}}$, and

collapse all nodes of set $\mathcal{V}_{\mathcal{T}}$ to a single node $\hat{\mathcal{T}}$ that denotes the destination in network $\hat{\mathcal{G}}$.

The process above splits every node $v \in \mathcal{V}$ into K_v separate nodes and splits every edge $e \in \mathcal{E}$ into $|\mathcal{F}_e|$ separate edges. The sum of capacities of all edges that edge e is split into is equal to the capacity of edge e. The result of applying this procedure to network $\mathcal{N}$ of Figure 2.5(b) is shown in Figure 2.5(c). In network $\hat{\mathcal{G}}$ erasure events on different links are not independent but correlated as follows. For every edge $(u,v) \in \mathcal{E}$, denote by $\mathcal{C}_{(u,v)} = \left\{(\hat{u},\hat{v}) \in \hat{\mathcal{E}} : \hat{v} \in \mathcal{K}_u, \hat{m} \in \mathcal{K}_v\right\}$ the set of edges in $\hat{\mathcal{G}}$ that are derived from edge $(u,v) \in \mathcal{E}$. The erasures on all edges in set $\mathcal{C}_{(u,v)}$ are not independent but correlated as follows. At each time step, with probability $1 - p_{(u,v)}$ one edge in set $\mathcal{C}_{(u,v)}$ succeeds, or all fail with probability $p_{(u,v)}$. In the case of a success, edge $\hat{e} \in \mathcal{C}_{(u,v)}$ is the single successful edge with probability $H_{\hat{e}}$.

The time taken $\hat{T}_n^{\mathrm{c}}$ for the n packets to travel through network $\hat{\mathcal{G}}$ by coding is at least as large as the time T_n^{c} taken in network $\mathcal{G}$, i.e.

$$\mathbb{E}T_n^{\mathrm{c}} \leq \mathbb{E}\hat{T}_n^{\mathrm{c}}. \tag{2.34}$$

Indeed network $\hat{\mathcal{G}}$ can be emulated by network $\mathcal{G}$ if each node $v \in \mathcal{G}$ has $|K_v|$ different buffers and packets between different buffers are not mixed. By construction, networks $\mathcal{G}$ and $\hat{\mathcal{G}}$ have the same capacity and since $\hat{\mathcal{G}}$ is a parallel path network, the mincut of network $\hat{\mathcal{G}}$ passes through the worst link of each path. According to Theorem 2

$$\mathbb{E}\hat{T}_n^{\mathrm{c}} = \frac{n}{C} + \hat{D}_n^{\mathrm{c}} \tag{2.35}$$

where $\hat{D}_n^{\mathrm{c}} \in \Omega(\sqrt{n})$ when there are multiple worst links in at least one path or $\hat{D}_n^{\mathrm{c}} \in \mathsf{O}(1)$ when there is a single worst link at each path. For a single-bottleneck network, by Lemma 1, one can construct a max-flow subgraph comprising paths each of which has a single worst link, so $\hat{D}_n^{\mathrm{c}} \in \mathsf{O}(1)$. Equations (2.34), (2.35) conclude our proof. □

2.6 Proof of Concentration

Here we present a martingale concentration argument. In particular we prove a slightly stronger version of Theorem 5:

Theorem 6 (Extended version of Theorem 5)**.** *The time T_n^c for n packets to be transmitted from a source to a sink over a network of erasure channels using network coding is concentrated around its expected value with high probability. In particular for sufficiently large n:*

$$\mathbb{P}\mathrm{r}[|T_n^c - \mathbb{E}T_n^c| > \epsilon_n] \leq \frac{2C}{n} + \frac{2Cn^{2\delta}}{n^2 - n^{1+2\delta}}.$$

where C is the capacity of the network and ϵ_n represents the corresponding deviation and is equal to $\epsilon_n = n^{1/2+\delta}/C$, $\delta \in (0, 1/2)$.

Proof. The main idea of the proof is to use the method of Martingale bounded differences [53]. This method works as follows: first we show that the random variable we want to show is concentrated is a function of a finite set of independent random variables. Then we show that this function is Lipschitz with respect to these random variables, *i.e.* it cannot change its value too much if only one of these variables is modified. Using this function we construct the corresponding Doob martingale and use the Azuma-Hoeffding [53] inequality to establish concentration. See also [54, 55] for related concentration results using similar martingale techniques. Unfortunately however this method does not seem to be directly applicable to T_n^{c} because it cannot be naturally expressed as a function of a *bounded number* of independent random variables. We use the following trick of showing concentration for another quantity first and then linking that concentration to the concentration of T_n^{c}.

Specifically, we define R_t to be the number of innovative (linearly independent) packets received at the destination node $\mathcal{T}$ after t time steps. R_t is linked with T_n^{c} through the equation:

$$T_n^{\mathrm{c}} = \arg_t(R_t = n). \tag{2.36}$$

The number of received packets is a well defined function of the link states at each time step. If there are L number of links in network $\mathcal{G}$, then:

$$R_t = g(z_{11}, ..., z_{1L}, \ldots, z_{t1}, ..., z_{tL}).$$

The random variables z_{ij},$1 \leq i \leq t$ and $1 \leq j \leq L$, are equal to 0 or 1 depending on whether link j is OFF or ON at time i. If a packet is sent on a link that is ON, it is received successfully; if sent on a link that is OFF, it is erased. It is clear that this function satisfies a bounded Lipschitz condition with a bound equal to 1:

$$|g(z_{11}, ..., z_{1L}, ..., z_{ij}, ..., z_{t1}, ..., z_{tL}) - g(z_{11}, ..., z_{1L}, ..., z_{ij}^{'}, ..., z_{t1}, ..., z_{tL})| \leq 1.$$

This is because if we look at the history of all the links failing or succeeding at all the t time slots, changing one of these link states in one time slot can at most influence the received rank by one. We note that we assume that coding is performed over a very large field to ensure that every packet that could potentially be innovative due to connectivity, indeed is.

Using the Azuma-Hoeffding inequality (see the Appendix Theorem 7) on the Doob martingale

constructed by $R_t = g(z_{11}, ..., z_{1L}, ..., z_{t1}, ..., z_{tL})$ we get following the concentration result:

Proposition 5. *The number of received innovative packets R_t is a random variable concentrated around its mean value:*

$$\mathbb{P}_{\mathbb{r}}(|R_t - \mathbb{E}R_t| \geq \varepsilon_t) \leq \frac{1}{t} \quad \textit{where} \quad \varepsilon_t \doteq \sqrt{\frac{tL}{2}\ell n(2t)}. \tag{2.37}$$

Proof. Given in Appendix B. □

Using this concentration and the relation (2.36) between T_n^{c} and R_t we can show that deviations of the order $\varepsilon_t \doteq \sqrt{\frac{tL}{2}\ell n(2t)}$ for R_t translate to deviations of the order of $\epsilon_n = n^{1/2+\delta}/C$ for T_n^{c}. In Theorem 6 smaller values δ give tighter bounds that hold for larger n. Define the events:

$$H_t = \{|R_t - \mathbb{E}R_t| < \varepsilon_t\}$$

and

$$\overline{H}_t = \{|R_t - \mathbb{E}R_t| \geq \varepsilon_t\}$$

and further define t_n^u (u stands for upper bound) to be some t, ideally the smallest t, such that $\mathbb{E}R_t - \varepsilon_t \geq n$ and t_n^l (l stands for lower bound) to be some t, ideally the largest t, such that $\mathbb{E}R_t + \varepsilon_t \leq n$. Then we have:

$$\begin{aligned} \mathbb{P}_{\mathbb{r}}(T_n^{\mathrm{c}} \geq t_n^u) &= \mathbb{P}_{\mathbb{r}}(T_n^{\mathrm{c}} \geq t_n^u | H_{t_n^u}) \cdot \mathbb{P}_{\mathbb{r}}(H_{t_n^u}) \\ &+ \mathbb{P}_{\mathbb{r}}(T_n^{\mathrm{c}} \geq t_n^u | \overline{H}_{t_n^u}) \cdot \mathbb{P}_{\mathbb{r}}(\overline{H}_{t_n^u}) \end{aligned}$$

where:

- $\mathbb{P}_{\mathbb{r}}(T_n^{\mathrm{c}} \geq t_n^u | H_{t_n^u}) = 0$ since at time $t = t_n^u$ the destination has already received more than n innovative packets. Indeed given that $H_{t_n^u}$ holds: $n \leq \mathbb{E}R_{t_n^u} - \varepsilon_{t_n^u} < R_{t_n^u}$ where the first inequality is due to the definition of t_n^u.
- $\mathbb{P}_{\mathbb{r}}(H_{t_n^u}) \leq 1$
- $\mathbb{P}_{\mathbb{r}}(T_n^{\mathrm{c}} \geq t_n^u | \overline{H}_{t_n^u}) \leq 1$
- $\mathbb{P}_{\mathbb{r}}(\overline{H}_{t_n^u}) \leq \frac{1}{t_n^u}$ due to equation (2.37).

Therefore:

$$\mathbb{P}_{\mathbb{r}}(T_n^{\mathrm{c}} \geq t_n^u) \leq \frac{1}{t_n^u}. \tag{2.38}$$

Similarly:

$$\begin{aligned}\mathbb{P}\mathrm{r}(T_n^{\mathrm{c}} \geq t_n^l) &= \mathbb{P}\mathrm{r}(T_n^{\mathrm{c}} \geq t_n^l | H_{t_n^l}) \cdot \mathbb{P}\mathrm{r}(H_{t_n^l}) \\ &+ \mathbb{P}\mathrm{r}(T_n^{\mathrm{c}} \geq t_n^l | \overline{H}_{t_n^l}) \cdot \mathbb{P}\mathrm{r}(\overline{H}_{t_n^l})\end{aligned}$$

where:

- $\mathbb{P}\mathrm{r}(T_n^{\mathrm{c}} \leq t_n^l | H_{t_n^l}) = 0$ since at time $t = t_n^l$ the destination has already received less than n innovative packets. Indeed given that $H_{t_n^l}$ holds: $R_{t_n^u} < \mathbb{E}R_{t_n^u} + \varepsilon_{t_n^u} < n$ where the last inequality is due to the definition of t_n^l.
- $\mathbb{P}\mathrm{r}(H_{t_n^l}) \leq 1$
- $\mathbb{P}\mathrm{r}(T_n^{\mathrm{c}} \leq t_n^l | \overline{H}_{t_n^l}) \leq 1$
- $\mathbb{P}\mathrm{r}(\overline{H}_{t_n^l}) \leq \frac{1}{t_n^l}$ due to equation (2.37).

Therefore:

$$\mathbb{P}\mathrm{r}(T_n^{\mathrm{c}} \leq t_n^l) \leq \frac{1}{t_n^l}. \tag{2.39}$$

Equations (2.38) and (2.39) show that the random variable T_n^{c} representing the time required for n packets to travel across network $\mathcal{G}$ exhibits some kind of concentration between t_n^l and t_n^u, which are both functions of n. As shown in Lemma 2 in Appendix B, for large enough n a legitimate choice for t_n^l and t_n^u is the following:

$$t_n^u = (n + n^{1/2+\delta'})/C,\ \delta' \in (0, 1/2) \tag{2.40}$$

$$t_n^l = (n - n^{1/2+\delta'})/C,\ \delta' \in (0, 1/2) \tag{2.41}$$

From both (2.38) and (2.39):

$$\begin{aligned}\mathbb{P}\mathrm{r}(t_n^l \leq T_n^{\mathrm{c}} \leq t_n^u) &= 1 - \mathbb{P}\mathrm{r}(T_n^{\mathrm{c}} \leq t_n^l) - \mathbb{P}\mathrm{r}(T_n^{\mathrm{c}} \geq t_n^u) \\ &\geq 1 - \frac{1}{t_n^l} - \frac{1}{t_n^u}\end{aligned} \tag{2.42}$$

and by substituting in (2.42) the t_n^u, t_n^l from equations (2.40) and (2.41) we get:

$$\mathbb{P}\mathrm{r}(-\frac{n^{1/2+\delta'}}{C} \leq T_n^{\mathrm{c}} - \frac{n}{C} \leq \frac{n^{1/2+\delta'}}{A}) \geq 1 - \frac{C}{n - n^{1/2+\delta'}} - \frac{C}{n + n^{1/2+\delta'}}$$

and since $\mathbb{E}T_n^{\mathrm{c}} = \frac{n}{C} + \mathsf{O}(\sqrt{n})$ we have:

$$\mathbb{P}\mathbb{r}(|T_n^{\mathrm{c}} - \mathbb{E}T_n^{\mathrm{c}}| \leq \frac{n^{1/2+\delta}}{C}) \geq 1 - \frac{2C}{n} - \frac{2Cn^{2\delta}}{n^2 - n^{1+2\delta}}$$

or

$$\mathbb{P}\mathbb{r}(|T_n^{\mathrm{c}} - \mathbb{E}T_n^{\mathrm{c}}| > \frac{n^{1/2+\delta}}{C}) \leq \frac{2C}{n} + \frac{2Cn^{2\delta}}{n^2 - n^{1+2\delta}}$$

where $\delta > \delta'$ and this concludes the proof. □

Appendix A

Proof of Proposition 1

Definition 1. *A binary relation $\preceq$ defined on a set P is called a preorder if it is reflexive and transitive, i.e. $\forall a, b, c \in P$:*

$$a \preceq a \qquad \textit{(reflexivity)} \tag{A.1}$$

$$(a \preceq b) \wedge (b \preceq c) \Rightarrow a \preceq c \qquad \textit{(transitivity)} \tag{A.2}$$

Definition 2. *On the set $\mathbb{N}^{\ell-1}$ of all integer $(\ell-1)$-tuples we define the regular preorder $\preceq$ that is $\forall a, b \in \mathbb{N}^{\ell-1}$ $a \preceq b$ iff $a_1 \leq b_1, \ldots, a_{\ell-1} \leq b_{\ell-1}$ where $a = (a_1, \ldots, a_{\ell-1})$ and $b = (b_1, \ldots, b_{\ell-1})$. Similarly we can define the preorder $\succeq$.*

Definition 3. *A random vector $X \in \mathbb{N}^{\ell-1}$ is said to be stochastically smaller in the usual stochastic order than a random vector $Y \in \mathbb{N}^{\ell-1}$, (denoted by $X \preceq_{st} Y$) if: $\forall \omega \in \mathbb{N}^{\ell-1}$, $\mathbb{Pr}(X \succeq \omega) \leq \mathbb{Pr}(Y \succeq \omega)$.*

Definition 4. *A family of random variables $\{Y_n\}_{n\in\mathbb{N}}$ is called stochastically increasing ($\preceq_{st}$-increasing) if $Y_k \preceq_{st} Y_n$ whenever $k \leq n$.*

Proof of Proposition 1. Markov process $\{Y_n, n \geq 1\}$, is a multidimensional process on $E = \mathbb{N}^{\ell-1}$ representing the number of innovative packets at nodes $N_1, \ldots, N_{\ell-1}$ when packet n arrives at N_1. To prove that the Markov process $\{Y_n, n \geq 1\}$ is stochastically increasing we introduce two other processes $\{X_n, n \geq 1\}$ and $\{Z_n, n \geq 1\}$ having the same state space and transition probabilities as $\{Y_n, n \geq 1\}$.

More precisely, Markov process $\{Y_n, n \geq 1\}$ is effectively observing the evolution of the number of innovative packets present at every node of the tandem queue. We define the two new processes $\{X_n, n \geq 1\}$ and $\{Z_n, n \geq 1\}$ to observe the evolution of two other tandem queues having the same link failure probabilities as the queue of $\{Y_n, n \geq 1\}$.

As seen in Figure A.1, at each time step and at every link, the queues for $\{X_n, n \geq 1\}$ and $\{Z_n, n \geq 1\}$ either both succeed or a fail together. Moreover the successes or failures on each link

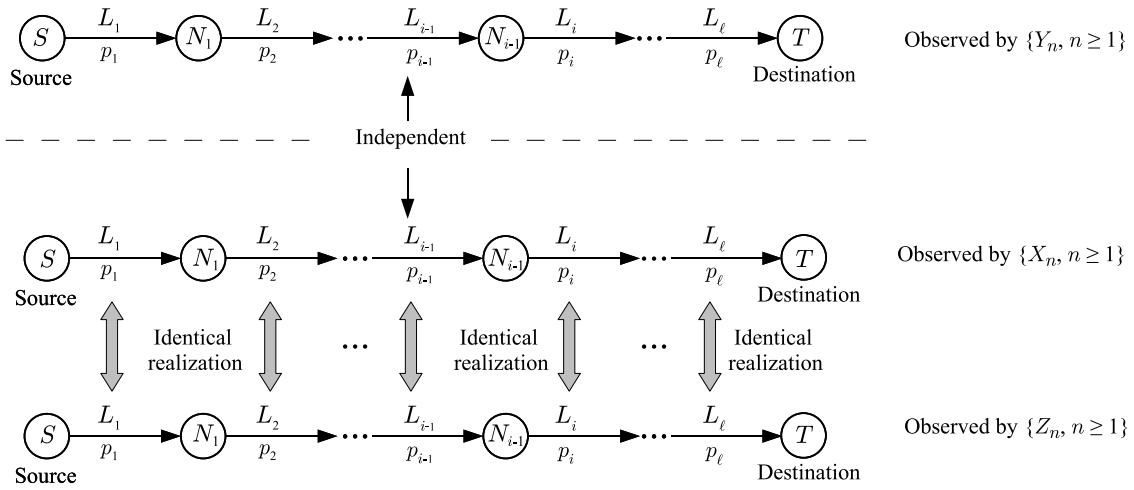


Figure A.1: Multi-hop network with the corresponding Markov chains

on the queues observed by $\{X_n, n \geq 1\}$ and $\{Z_n, n \geq 1\}$ are independent of the successes or failures on the queue observed by $\{Y_n, n \geq 1\}$. Formally the joint process $\{(X_n, Z_n), n \geq 1\}$ constitute a coupling meaning that marginally each one of $\{X_n, n \geq 1\}$ and $\{Z_n, n \geq 1\}$ have the transition matrix $\mathbb{P}\mathrm{r}_Y$ of $\{Y_n, n \geq 1\}$. If Markov processes $\{X_n, n \geq 1\}$ and $\{Z_n, n \geq 1\}$ have different initial conditions then the following relation holds:

$$X_1 \preceq Z_1 \Rightarrow X_n \preceq Z_n \tag{A.3}$$

The proof of the above statement is very similar to the proof of Proposition 2 in [47]. Essentially relation (A.3) states that since at both queues all links succeed or fail together the queue that holds more packets at each node initially ($n = 1$) will also hold more packets subsequently ($n > 1$) at every node.

The initial state Y_1 of Markov process $\{Y_n, n \geq 1\}$ is state $\alpha = (1, 0, \ldots, 0)$ that is also called the minimal state since any other state is greater than the minimal state. To prove Proposition 1 we set both processes $\{Y_n, n \geq 1\}$ and $\{X_n, n \geq 1\}$ to start from the minimal state ($Y_1 \stackrel{\mathcal{D}}{=} \delta_\alpha, X_1 \stackrel{\mathcal{D}}{=} \delta_\alpha$ where $\stackrel{\mathcal{D}}{=}$ means equality in distribution), whereas process $\{Z_n, n \geq 1\}$ has initial distribution μ that is the distribution of process $\{Y_n, n \geq 1\}$ after $(n-k)$ steps ($\mu = \mathbb{P}\mathrm{r}_Y^{n-k}\delta_\alpha$ and $Z_1 \stackrel{\mathcal{D}}{=} \mu$). Then for every ω in the state space of $\{Y_n, n \geq 1\}$ we get:

$$\mathbb{P}\mathrm{r}(X_n \succeq \omega) = \mathbb{P}\mathrm{r}(Y_n \succeq \omega) = \mathbb{P}\mathrm{r}(Z_k \succeq \omega) \tag{A.4}$$

where the first equality holds since the two processes have the same distribution–both start from

the minimal element and have the same transition matrices–and the second equality holds since

$$Z_k \overset{\mathcal{D}}{=} \mathbb{P}\mathrm{r}_Y^k \mu \equiv \mathbb{P}\mathrm{r}_Y^k (\mathbb{P}\mathrm{r}_Y^{n-k} \delta_\alpha) = \mathbb{P}\mathrm{r}_Y^n \delta_\alpha \overset{\mathcal{D}}{=} Y_n.$$

Moreover due to the definition of the minimal element, $X_1 \preceq Z_1$ and using (A.3) we get $X_n \preceq Z_n$. Therefore

$$\mathbb{P}\mathrm{r}(Z_k \succeq \omega) \geq \mathbb{P}\mathrm{r}(X_k \succeq \omega) = \mathbb{P}\mathrm{r}(Y_k \succeq \omega). \tag{A.5}$$

The last equality follows from the fact that the two distributions have the same law. Equations (A.4) and (A.5) conclude the proof. □

Appendix B

Proof of Proposition 5

Definition 5. *A sequence of random variables $V_0, V_1, \ldots$ is said to be a* ***martingale with respect to*** *another sequence $U_0, U_1, \ldots$ if, for all $n \geq 0$, the following conditions hold:*

- $\mathbb{E}[|V_n|] < \infty$
- $\mathbb{E}[V_{n+1}|U_0, \ldots, U_n] = V_n$

A sequence of random variables $V_0, V_1, \ldots$ is called ***martingale*** *when it is a martingale with respect to itself. That is:*

- $\mathbb{E}[|V_n|] < \infty$
- $\mathbb{E}[V_{n+1}|V_0, ..., V_n] = V_n$

Theorem 7. *(Azuma-Hoeffding Inequality): Let X_0, X_1,...,X_n be a martingale such that*

$$B_k \leq X_k - X_{k-1} \leq B_k + d_k$$

for some constants d_k and for some random variables B_k that may be a function of $X_0, ..., X_{k-1}$. Then for all $t \geq 0$ and any $\lambda > 0$,

$$\mathbb{Pr}(|X_t - X_0| \geq \lambda) \leq 2 \exp\left(-\frac{2\lambda^2}{\sum_{i=1}^{t} d_i^2}\right)$$

Proof. Theorem 12.6 in [53] □

Proof of Proposition 5. The proof is based on the fact that from a sequence of random variables $U_1, U_2, \ldots, U_n$ and any function f it's possible to define a new sequence $V_0, \ldots, V_n$

$$\begin{cases} V_0 = \mathbb{E}[f(U_1, \ldots, U_n)] \\ V_i = \mathbb{E}[f(U_1, \ldots, U_n)|U_1, \ldots, U_i] \end{cases}$$

that is a martingale (*Doob* martingale). Using the identity $\mathbb{E}[V|W] = \mathbb{E}[\mathbb{E}[V|U,W]|W]$ it's easy to verify that the above sequence $V_0, \ldots, V_n$ is indeed a martingale. Moreover if function f is *c-Lipschitz* and $U_1, \ldots, U_n$ are independent it can be proved that the differences $V_i - V_{i-1}$ are restricted within bounded intervals [53] (pages 305-306).

Function $R_t = g(z_{11}, ..., z_{tL})$ has a bounded expectation, is *1-Lipschitz* and the random variables z_{ij} are independent and therefore all the requirements of the above analysis hold. Specifically by setting

$$G_h = \mathbb{E}[g(z_{11}, ..., z_{tL}) \mid \underbrace{z_{11}, ..., z_{kr}}_{h\text{-terms in total}}]$$

we can apply the Azuma-Hoeffding inequality on the $G_0, ..., G_{tL}$ martingale and we get the following concentration result

$$\mathbb{P}\mathrm{r}[|G_{tL} - G_0| \geq \lambda] = \mathbb{P}\mathrm{r}[|R_t - \mathbb{E}[R_t]| \geq \lambda] \leq 2\exp\{-\frac{2\lambda^2}{tL}\}. \tag{B.1}$$

The equality above holds since

- $G_0 = \mathbb{E}[R_t]$
- $G_{tL} = R_t$ (the random variable itself)

and by substituting on (B.1) λ with $\varepsilon_t \doteq \sqrt{\frac{tL}{2}\ell n(2t)}$

$$\mathbb{P}\mathrm{r}[|R_t - \mathbb{E}[R_t]| \geq \varepsilon_t] \leq \frac{1}{t}$$

□

Lemma 2. *A legitimate choice for t_n^u and t_n^l is:*

$$t_n^u = (n + n^{1/2+\delta'})/C,\ \delta' \in (0, 1/2)$$
$$t_n^l = (n - n^{1/2+\delta'})/C,\ \delta' \in (0, 1/2)$$

Proof. For any $t \leq n/C$, the expected number of received packets $\mathbb{E}R_t$ is given by $\mathbb{E}R_t = Ct - r(t)$, where C is the capacity of the network and $r(t)$ can be bounded as follows. Letting $n_t = Ct \leq n$, we have

$$E(T_{n_t}^c) \quad = \quad E(E(T_{n_t}^c | r(t)))$$

$$\begin{aligned} &= E(t + \mathsf{O}(r(t))) \\ &= t + \mathsf{O}(r(t)) \end{aligned}$$

which by Theorem 4 implies that $r(t)$ should be $\mathsf{O}(\sqrt{n_t}) \leq \mathsf{O}(\sqrt{n})$.

The only requirement for t_n^u is that it is a t such that $\mathbb{E}R_t - \epsilon_t \geq n$. This is indeed true for large enough n if we substitute t_n^u with $(n + n^{1/2+\delta'})/C$:

$$\begin{aligned} \mathbb{E}[R_{t_n^u}] - \epsilon_{t_n^u} \geq n \Rightarrow Ct_n^u - r(t_n^u) - \epsilon_{t_n^u} \geq n \Rightarrow Ct_n^u - r(t_n^u) - \sqrt{\frac{Lt_n^u}{2}\ln(2t_n^u)} \geq n \\ \Rightarrow C \cdot \frac{n + n^{1/2+\delta'}}{C} - r(t_n^u) - \sqrt{\frac{L(n + n^{1/2+\delta'})}{2C}\ln(\frac{2(n + n^{1/2+\delta'})}{C})} \geq n. \end{aligned} \tag{B.2}$$

Since $r(t) \in \mathsf{O}(\sqrt{n})$ there is a constant $B > 0$ such that $r(t) \leq B\sqrt{n}$ and therefore in order for (B.2) to hold it is sufficient if

$$\begin{aligned} n + n^{1/2+\delta'} - B\sqrt{n} - \sqrt{\frac{L(n + n^{1/2+\delta'})}{2C}\ln(\frac{2(n + n^{1/2+\delta'})}{C})} \geq n \\ \Rightarrow n^{1/2+\delta'} \geq \sqrt{\frac{L(n + n^{1/2+\delta'})}{2C}\ln(\frac{2(n + n^{1/2+\delta'})}{C})} + B\sqrt{n} \\ \Rightarrow n^{1/2+\delta'} \geq \sqrt{n}\sqrt{\frac{L(1 + n^{\delta'-1/2})}{2C}\ln(\frac{2(n + n^{1/2+\delta'})}{C})} + B\sqrt{n} \\ \Rightarrow n^{\delta'} \geq \sqrt{\frac{L(1 + n^{\delta'-1/2})}{2C}\ln(\frac{2(n + n^{1/2+\delta'})}{C})} + B \end{aligned}$$

where the last equation holds for large enough n.

Similarly it can be proved that t_n^l can be substituted with $(n - n^{1/2+\delta'})/C$ such that for large n, $\mathbb{E}R_t + \epsilon_t \leq n$. □

Chapter 3

Network Equivalence in the Presence of an Eavesdropper

3.1 Introduction

The problem of secure (secret) communication in the presence of an eavesdropper has been studied using a variety of approaches. One body of literature studies the secure capacity of the wiretap channel introduced by Wyner in [17]. In this model, the eavesdropper's observation is a degraded version of the legitimate receiver's observation, and the goal of secure communication is to maximize the rate at which information can be reliably delivered to the legitimate receiver without information leakage to the eavesdropper. Another body of literature investigates the secure capacity of networks of noise-free links. Under this model, introduced by Cai and Yeung in [21], an eavesdropper perfectly observes all messages traversing a restricted but unknown subset of links. The goal again is to maximize the rate at which information can be reliably delivered to the intended receiver(s) without information leakage to the eavesdropper. The results of [21] treat multicast networks with equal capacity links; various extensions appear in [22, 56]. In [23, 24] Cui, Ho, and Kliewer show that finding the capacity of the secure network communication problem is NP hard for the case of unequal capacity links; the paper therefore gives some achievable coding strategies without proving whether those bounds are tight.

This paper aims to build a bridge between the wiretap channel and secure network coding literatures. While conceptually related, the two fields have evolved largely independently. The first paper on the capacity of a network of wiretap channels is [57], which finds upper and lower bounds on the unicast capacity of a network of independent erasure channels when the output observed by the eavesdropper equals that of the intended receiver on all wiretapped channels. This paper also considers the problem of secure communication over a network of independent wiretap channels. For our work, the channels are physically degraded, "simultaneously maximizable" wiretap channels (see Definition 6 in Section 3.2), which include the erasure channels of [57] as a special case. We further

generalize the model from [57] by broadening the focus from unicast capacity to a consideration of complete capacity regions. The "capacity region," defined formally in Section 3.2, refers to the set of all vectors of simultaneously achievable rates under all possible combinations of unicast and multicast connections. In our model, the eavesdropper wiretaps a limited but unknown subset of channels and for each eavesdropped link overhears a possibly degraded version of the received channel output at the intended link output. Our central result shows that the secure capacity region for any network of such channels can be bounded from above and below by the secure network coding capacity regions of a corresponding pair of noiseless networks. In the case where the eavesdropper has access to only one link, the identity of which is unknown to the code designer, the upper and lower bounds on the secure network coding capacity are identical. This result gives an equivalence in the sense that the capacity regions of the noiseless and noisy networks are identical for all possible topologies and all possible connection types that can be established across the network. When the eavesdropper has access to more than one link, the upper and lower bounds differ, giving new achievability results and converses for cases where the secure network coding problem can be solved. Using these bounding and equivalence results, secure network coding capacity bounds can be applied to bound the secure capacity of a network of wiretap channels.[1] The bounding relationship between the secrecy capacity region for noisy wiretap networks and noise-free wiretap networks is derived by generalizing the techniques developed by Koetter, Effros, and Medard in [31, 32], which show similar capacity bounds in the absence of secrecy constraints.

3.2 Network Model

The following description defines terminology and notation for a network of independent wiretap channels and its secure capacity under a given restricted adversarial model. Consider a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V} \times \mathbb{N}$ is a set of directed "edges" between pairs of nodes in the network. For the purposes of this work, edge (i, j, k) represents the k^{th} wiretap channel through which node i communicates to node j and through which an eavesdropper may or may not be listening. The total number of nodes in the network is m, and each node $i \in \mathcal{V}$ transmits, at each time step t, a random variable $X_t^{(i)} \in \mathcal{X}^{(i)}$ and receives a random variable $Y_t^{(i)} \in \mathcal{Y}^{(i)}$. The sets $\mathcal{X}^{(i)}$ and $\mathcal{Y}^{(i)}$ are the input and output alphabets of the outgoing and incoming channels, respectively, at node i; we consider both discrete and continuous alphabets. The indegree $d_{\text{in}}(i)$ and

[1] The definitions of reliability and security used in the prior literature vary from "strong reliability" and "strong security," where the receiver's error probability is precisely zero and the eavesdropper learns nothing about the transmitted message (mutual information precisely equals 0) to "weak reliability" and "weak security," where the constraints on error probability and mutual information are asymptotic in nature. We use the latter definitions, as defined formally in Section 3.2.

outdegree $d_{\text{out}}(i)$ of node i in graph $\mathcal{G}$ are defined as

$$\mathcal{E}_{\text{in}}(i) = \{(u,v,w) \in \mathcal{E} : v = i\}\,,\ d_{\text{in}}(i) = |\mathcal{E}_{\text{in}}(i)|$$
$$\mathcal{E}_{\text{out}}(i) = \{(u,v,w) \in \mathcal{E} : u = i\}\,,\ d_{\text{out}}(i) = |\mathcal{E}_{\text{out}}(i)|\,.$$

If node i has indegree or outdegree larger than one (that is, if node $i \in \mathcal{V}$ receives outputs from more than one channel or transmits inputs to more than one channel) then

$$\mathcal{X}^{(i)} = \prod_{e \in \mathcal{E}_{\text{out}}(i)} \mathcal{X}^{(e)} \quad \text{and} \quad \mathcal{Y}^{(i)} = \prod_{e \in \mathcal{E}_{\text{in}}(i)} \mathcal{Y}^{(e)}.$$

The channel inputs and outputs for node i at time t are given by

$$X_t^{(i)} = \left(X_t^{(e)} : e \in \mathcal{E}_{\text{out}}(i)\right) \quad \text{and} \quad Y_t^{(i)} = \left(Y_t^{(e)} : \mathcal{E}_{\text{in}}(i)\right).$$

Here $X_t^{(e)}$ and $Y_t^{(e)}$ denote the input to and the output from edge e, at time t, and $\mathcal{X}^{(e)}$ and $\mathcal{Y}^{(e)}$ denote their alphabets.

Let $\mathcal{P}(\mathcal{E})$ denote the power set of the set of all edges. We define a secure communication problem by defining an adversarial set $A \subseteq \mathcal{P}(\mathcal{E})$. Each set $E \in A$ describes a subset of channels over which an eavesdropper may be listening. The goal of code design is to build a code that is secure against eavesdropping on the set of channels E for every $E \in A$. When the eavesdropper listens to edge $e = (i,j,k)$, the eavesdropper receives, at each time t, a degraded version $Z_t^{(e)}$ of the channel output $Y_t^{(e)}$ observed by the intended recipient, which is the output node j of edge $e = (i,j,k)$. If the eavesdropper has eavesdropping set $E \in A$, then at time t it receives the set of random variables $\left(Z_t^{(e)} : e \in E\right)$, which we compactly write as $Z_t^{(E)}$. The vector $\left(Z_1^{(E)}, \ldots, Z_n^{(E)}\right)$ of observations from all edges $e \in E$ over time steps $t \in \{1, \ldots, n\}$ is denoted by $\left(Z^{(E)}\right)^n$. Similarly we define $\left(X^{(E)}\right)^n = \left(X_1^{(E)}, \ldots, X_n^{(E)}\right)$ and $\left(Y^{(E)}\right)^n = \left(Y_1^{(E)}, \ldots, Y_n^{(E)}\right)$ where $X_t^{(E)} = \left(X_t^{(e)} : e \in E\right)$ and $Y_t^{(E)} = \left(Y_t^{(e)} : e \in E\right)$.

For each $e \in \mathcal{E}$, channel e is a memoryless, time-invariant, physically degraded wiretap channel described by a conditional distribution

$$p(y^{(e)}, z^{(e)} | x^{(e)}) = p(y^{(e)} | x^{(e)}) \cdot p(z^{(e)} | y^{(e)}).$$

All wiretap channels are independent by assumption, giving

$$p\big(y^{(\mathcal{E})}, z^{(\mathcal{E})} | x^{(\mathcal{E})}\big) = \prod_{e \in \mathcal{E}} p\big(y^{(e)}, z^{(e)} | x^{(e)}\big) = \prod_{e \in \mathcal{E}} p\big(y^{(e)} | x^{(e)}\big) p\big(z^{(e)} | y^{(e)}\big).$$

We further restrict our attention to channels that are "simultaneously maximizable," as defined

below.

Definition 6. *Wiretap channel e is called simultaneously maximizable if* $\arg\left[\max_{p(x)} I(X^{(e)};Y^{(e)})\right] = \arg\left[\max_{p(x)} I(X^{(e)};Z^{(e)})\right]$.

The maximization in Definition 6 is subject to any constraints on the channel input (*e.g.*, the power constraint at the input to a Gaussian channel) associated with the communication network of interest. Examples of simultaneously maximizable wiretap channels include weakly symmetric channels and Gaussian channels. For all simultaneously maximizable wiretap channels, the following property holds.

Lemma 3. *[58, Proposition* 3.4.4[2]*] Given a simultaneously maximizable wiretap channel e,*

$$\max_{p(x^{(e)})} \left[I(X^{(e)};Y^{(e)}) - I(X^{(e)};Z^{(e)})\right] = \max_{p(x^{(e)})} I(X^{(e)};Y^{(e)}) - \max_{p(x^{(e)})} I(X^{(e)};Z^{(e)}).$$

Intuitively, restriction to simultaneously maximizable channels simplifies our analysis since the same input distribution maximizes an individual wiretap channel's capacity to both its intended and unintended receivers. This property is employed in the derivations of Theorems 9 and 10 in Section 3.4.

A code of blocklength n operates over n time steps with the goal of reliably communicating, for each $i \in \mathcal{V}$ and non-empty $\mathcal{B} \subseteq \mathcal{V}\backslash\{i\}$, message

$$W^{(i\to\mathcal{B})} \in \mathcal{W}^{(i\to\mathcal{B})} \stackrel{\text{def}}{=} \{1,\ldots,2^{nR^{(i\to\mathcal{B})}}\}$$

from source node $i \in \mathcal{V}$ to set $\mathcal{B} \subseteq \mathcal{V}\backslash\{i\}$ of sink nodes in a manner that guarantees information theoretic security in the presence of any eavesdropper $E \in A$. This message delivery constitutes a unicast connection if $|\mathcal{B}| = 1$ and a multicast connection if $|\mathcal{B}| > 1$. Constant $R^{(i\to\mathcal{B})}$ is called the transmission rate from source i to sink set $\mathcal{B}$. By setting $R^{(i\to\mathcal{B})} = 0$ for some subset of $(i,\mathcal{B})$ pairs, we can obtain both a single unicast connection (as in [57]) and a single multicast connection (as in [21]) as special cases of this framework. The vector of all rates $R^{(i\to\mathcal{B})}$ is denoted by $R = \left(R^{(i\to\mathcal{B})} : i \in \mathcal{V}, \mathcal{B} \in \mathcal{B}^{(i)}\right)$, where set $\mathcal{B}^{(i)} = \{\mathcal{B} : \mathcal{B} \subseteq \mathcal{V}\backslash\{i\}, \mathcal{B} \neq \emptyset\}$ is the set of non-empty receiver sets to which node i may wish to transmit. Similarly, the vector of all messages is denoted by $W = \left(W^{(i\to\mathcal{B})} : i \in \mathcal{V}, \mathcal{B} \in \mathcal{B}^{(i)}\right)$. In an m-node network, vectors R and W have dimension $m(2^{m-1}-1)$ since nodes send no messages to themselves or to empty sets.

In addition to all outgoing messages $\left(W^{(i\to\mathcal{B})} : \mathcal{B} \in \mathcal{B}^{(i)}\right)$, each node $i \in \mathcal{V}$ is assumed to have access to a random variable $T^{(i)} \in \mathcal{T}^{(i)} \stackrel{\text{def}}{=} \{1,\ldots,2^{nC^{(i)}}\}$ for use in establishing random keys to

[2]The result is stated in [58, Proposition 3.4.4] for weakly symmetric channels, but the proof given there applies without change to all simultaneously maximizable channels, as noted in [58, Remark 3.4.6].

protect information from the adversary. Each $T^{(i)}$ is uniformly distributed on its alphabet and independent of all messages and channel noise. Here

$$C^{(i)} = \sum_{e \in \mathcal{E}_{\text{out}}(i)} \max_{p(x^{(e)})} I\big(X^{(e)}; Y^{(e)}\big)$$

is the sum of the outgoing channel capacities from node i. Node i needs at most $nC^{(i)}$ bits of randomness in a code of blocklength n since it could not hope to transmit any more than this even if it dedicated all outgoing links exclusively to that communication.

Definition 7. *Let a network*

$$\mathcal{N} \stackrel{def}{=} \left(\prod_{e \in \mathcal{E}} \mathcal{X}^{(e)}, \prod_{e \in \mathcal{E}} \left(p(y^{(e)}|x^{(e)}) p\left(z^{(e)}|y^{(e)}\right) \right), \prod_{e \in \mathcal{E}} \left(\mathcal{Y}^{(e)} \times \mathcal{Z}^{(e)} \right) \right)$$

be given corresponding to a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. A blocklength n solution $\mathcal{S}(\mathcal{N})$ to network $\mathcal{N}$ is defined as a set of encoding and decoding functions

$$X_t^{(i)} : \left(\mathcal{Y}^{(i)}\right)^{t-1} \times \prod_{\mathcal{B} \in \mathcal{B}^{(i)}} \mathcal{W}^{(i \to \mathcal{B})} \times \mathcal{T}^{(i)} \longrightarrow \mathcal{X}^{(i)}$$

$$\breve{W}^{(j \to \mathcal{K}, i)} : \left(\mathcal{Y}^{(i)}\right)^{n} \times \prod_{\mathcal{B} \in \mathcal{B}^{(i)}} \mathcal{W}^{(i \to \mathcal{B})} \times \mathcal{T}^{(i)} \longrightarrow \mathcal{W}^{(j \to \mathcal{K})}$$

mapping $\left(Y_1^{(i)}, \ldots, Y_{t-1}^{(i)}, \left(W^{(i \to \mathcal{B})} : \mathcal{B} \in \mathcal{B}^{(i)}\right), T^{(i)}\right)$ to $X_t^{(i)}$ for each $i \in \mathcal{V}$ and $t \in \{1, \ldots, n\}$ and mapping $\left(Y_1^{(i)}, \ldots, Y_n^{(i)}, \left(W^{(i \to \mathcal{B})} : \mathcal{B} \in \mathcal{B}^{(i)}\right), T^{(i)}\right)$ to $\breve{W}^{(j \to \mathcal{K}, i)}$ for each $j \in \mathcal{V}$, $\mathcal{K} \in \mathcal{B}^{(j)}$, and $i \in \mathcal{K}$. The solution $\mathcal{S}(\mathcal{N})$ of blocklength n is called a $(\lambda, \varepsilon, A, R)$–solution, denoted $(\lambda, \varepsilon, A, R)$–$\mathcal{S}(\mathcal{N})$, if the specified encoding and decoding functions imply $\Pr\left(\breve{W}^{(j \to \mathcal{K}, i)} \neq W^{(j \to \mathcal{K})}\right) < \lambda$ for every $j \in \mathcal{V}$, $\mathcal{K} \in \mathcal{B}^{(j)}$, and $i \in \mathcal{K}$ and $I\left(\left(Z^E\right)^n; W\right) < n\varepsilon$ for every $E \in A$.

Definition 8. *The A–secure rate region $\mathcal{R}(\mathcal{N}, A) \subseteq \mathbb{R}_+^{m(2^{m-1}-1)}$ of a network $\mathcal{N}$ is the closure of all rate vectors R such that for any $\lambda > 0$ and $\varepsilon > 0$, a solution $(\lambda, \varepsilon, A, R)$–$\mathcal{S}(\mathcal{N})$ exists.*

Given an arbitrary network $\mathcal{N}$ and some channel $\bar{e} \in \mathcal{E}$, the model $\mathcal{N}_{\bar{e}}(R_c, R_p)$ for $\mathcal{N}$, defined similarly to [31, 32], is used in the equivalence and bounding results proved in the following sections.

Definition 9. *Given a network $\mathcal{N} \stackrel{def}{=} \left(\prod_{e \in \mathcal{E}} \mathcal{X}^{(e)}, \prod_{e \in \mathcal{E}} \left(p\left(y^{(e)}|x^{(e)}\right) p\left(z^{(e)}|y^{(e)}\right) \right), \prod_{e \in \mathcal{E}} \left(\mathcal{Y}^{(e)} \times \mathcal{Z}^{(e)} \right) \right)$ and some channel $\bar{e} \in \mathcal{E}$, $\mathcal{N}_{\bar{e}}(R_c, R_p)$ replaces arbitrary degraded wiretap channel*

$$\mathcal{C}_{\bar{e}} = \left(\mathcal{X}^{(\bar{e})}, p(y^{(\bar{e})}|x^{(\bar{e})}) p(z^{(\bar{e})}|y^{(\bar{e})}), \mathcal{Y}^{(\bar{e})} \times \mathcal{Z}^{(\bar{e})}\right)$$

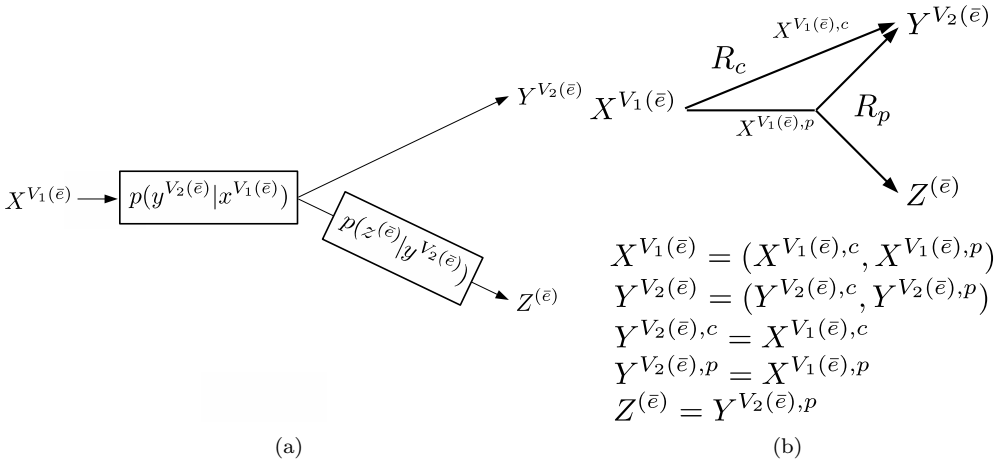


Figure 3.1: (a) A noisy degraded broadcast channel $\bar{e}$. (b) A noiseless degraded broadcast channel with rates $R_c + R_p$ and R_p toward the regular and degraded output respectively.

with the noiseless degraded wiretap channel

$$\mathcal{C}(R_c, R_p) = \Big(\{0,1\}^{R_c+R_p}, \delta\big(y^{(\bar{e})} - (x^{(\bar{e}),c}, x^{(\bar{e}),p})\big)\delta\big(z^{(\bar{e})} - y^{(\bar{e}),p}\big), \{0,1\}^{R_c+R_p} \times \{0,1\}^{R_p}\Big)$$

that delivers the rate-R_c confidential portion $x^{(\bar{e}),c}$ of channel input $x^{(\bar{e})} = (x^{(\bar{e}),c}, x^{(\bar{e}),p})$ to the intended receiver and the rate-R_p public portion $x^{(\bar{e}),p}$ of that input to both the intended receiver and the eavesdropper. The resulting network is given by

$$\mathcal{N}_{\bar{e}}(R_c, R_p) \stackrel{def}{=} \Bigg(\{0,1\}^{R_c+R_p} \times \prod_{e \in \mathcal{E}\setminus\{\bar{e}\}} \mathcal{X}^{(e)}, \delta\big(y^{(\bar{e})} - (x^{(\bar{e}),c}, x^{(\bar{e}),p})\big)\delta\big(z^{(\bar{e})} - y^{(\bar{e}),p}\big)$$
$$\cdot \prod_{e \in \mathcal{E}\setminus\{\bar{e}\}} \Big(p(y^{(e)}|x^{(e)})\, p\Big(z^{(e)}|y^{(e)}\Big)\Big), \{0,1\}^{R_c+R_p} \times \{0,1\}^{R_p} \times \prod_{e \in \mathcal{E}\setminus\{\bar{e}\}} \Big(\mathcal{Y}^{(e)} \times \mathcal{Z}^{(e)}\Big) \Bigg).$$

Figure 3.1 illustrates wiretap channel $\mathcal{C}_{\bar{e}}$ and noiseless model $\mathcal{C}_{\bar{e}}(R_c, R_p)$ from Definition 9. The given noiseless wiretap channel is physically degraded since wiretap output $Z^{(\bar{e})} = Y^{(\bar{e}),p}$ is conditionally independent of input $X^{(\bar{e})} = \big(X^{(\bar{e}),c}, X^{(\bar{e}),p}\big)$ given intended output $Y^{(\bar{e})} = (Y^{(\bar{e}),c}, Y^{(\bar{e}),p})$. It is also simultaneously maximizable since independently maximizing the entropies of components of $X^{(\bar{e}),c}$ and $X^{(\bar{e}),p}$ of the channel input maximizes the mutual information for both the intended receiver and the wiretap output. As in [31, 32], we allow non-integer values of R_c and R_p to denote noiseless bit pipes that require multiple channel uses to deliver some integer number of bits.

Many of the proofs in the sections that follow rely on the notion of a "stacked network" introduced in [31, 32]. The stacked network defined here simply adds an eavesdropper to the stacked network introduced in [31, 32]. Informally, the N-fold stacked network $\underline{\mathcal{N}}$ contains N copies of network $\mathcal{N}$.

The N copies of each node $i \in \mathcal{V}$ use the outgoing messages and channel outputs from all N layers of the network to form the channel inputs in each layer of the stack. Likewise, each node uses the channel outputs and messages from all layers in the stack in building its message reconstructions. An eavesdropper $E \in A$ overhears all copies of channel e for each $e \in E$.

As defined formally below, a solution for N-fold stacked network $\underline{\mathcal{N}}$ must securely and reliably transmit, for each $i \in \mathcal{V}$ and $\mathcal{B} \in \mathcal{B}^{(i)}$, N independent messages $\underline{W}^{(i\to\mathcal{B})}(1), \ldots, \underline{W}^{(i\to\mathcal{B})}(N)$ from node i to all the receivers in set $\mathcal{B}$. Following [31, 32] we underline the variable names from $\mathcal{N}$ to obtain variables for the stacked network $\underline{\mathcal{N}}$. Therefore $\underline{W}^{(i\to\mathcal{B})} \in \underline{\mathcal{W}}^{(i\to\mathcal{B})} \stackrel{\text{def}}{=} \left(\mathcal{W}^{(i\to\mathcal{B})}\right)^N$, $\underline{T}^{(i)} \in \underline{\mathcal{T}}^{(i)} \stackrel{\text{def}}{=} \left(\mathcal{T}^{(i)}\right)^N$, $\underline{X}_t^{(i)} \in \underline{\mathcal{X}}^{(i)} \stackrel{\text{def}}{=} \left(\mathcal{X}^{(i)}\right)^N$, $\underline{Y}_t^{(i)} \in \underline{\mathcal{Y}}^{(i)} \stackrel{\text{def}}{=} \left(\mathcal{Y}^{(i)}\right)^N$, and $\underline{Z}_t^{(e)} \in \underline{\mathcal{Z}}^{(e)} \stackrel{\text{def}}{=} \left(\mathcal{Z}^{(e)}\right)^N$ denote N-dimensional vectors of messages, channel inputs, channel outputs, and eavesdropper outputs corresponding to $W^{i\to\mathcal{B}}$, $X_t^{(i)}$, $Y_t^{(i)}$, and $Z_t^{(e)}$, respectively, in network $\mathcal{N}$. The variables in the ℓ^{th} layer of the stack are denoted by an argument ℓ. For example $\underline{X}_t^{(i)}(\ell)$ is the layer-ℓ channel input from node i at time t. Finally, following [31, 32], we define the rate $R^{(i\to\mathcal{B})}$ for a stacked network to be $(\log_2 |\underline{\mathcal{W}}^{(i\to\mathcal{B})}|)/(nN)$ since any solution of blocklength n for N-fold stacked network $\underline{\mathcal{N}}$ can be operated as a rate-R solution of blocklength nN for network $\mathcal{N}$ under this definition [31, Theorem 1]. A similar argument, given in Theorem 8 below, justifies the security constraint imposed in the definition that follows.

Definition 10. *Let a network*

$$\mathcal{N} \stackrel{def}{=} \left(\prod_{e\in\mathcal{E}} \mathcal{X}^{(e)}, \prod_{e\in\mathcal{E}} \left(p_e\left(y^{(e)}|x^{(e)}\right) p_e\left(z^{(e)}|y^{(e)}\right)\right), \prod_{e\in\mathcal{E}} \left(\mathcal{Y}^{(e)} \times \mathcal{Z}^{(e)}\right)\right)$$

be given corresponding to a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, and let an eavesdropper set $A \subseteq P(\mathcal{E})$ be defined on network $\mathcal{N}$. Let $\underline{\mathcal{N}}$ be the N-fold stacked network for $\mathcal{N}$. A blocklength-n solution $\mathcal{S}(\underline{\mathcal{N}})$ to this network is defined as a set of encoding and decoding functions

$$\underline{X}_t^{(i)} : \left(\underline{\mathcal{Y}}^{(i)}\right)^{t-1} \times \prod_{\mathcal{B}\in\mathcal{B}^{(i)}} \underline{\mathcal{W}}^{(i\to\mathcal{B})} \times \underline{\mathcal{T}}^{(i)} \longrightarrow \underline{\mathcal{X}}^{(i)}$$
$$\breve{\underline{W}}^{(j\to\mathcal{K},i)} : \left(\underline{\mathcal{Y}}^{(i)}\right)^{n} \times \prod_{\mathcal{B}\in\mathcal{B}^{(i)}} \underline{\mathcal{W}}^{(i\to\mathcal{B})} \times \underline{\mathcal{T}}^{(i)} \longrightarrow \underline{\mathcal{W}}^{(j\to\mathcal{K})}$$

mapping $\left(\underline{Y}_1^{(i)}, \ldots, \underline{Y}_{t-1}^{(i)}, \left(\underline{W}^{(i\to\mathcal{B})} : \mathcal{B} \in \mathcal{B}^{(i)}\right), \underline{T}^{(i)}\right)$ to $\underline{X}_t^{(i)}$ for each $i \in \mathcal{V}$ and $t \in \{1, \ldots, n\}$ and mapping $\left(\underline{Y}_1^{(i)}, \ldots, \underline{Y}_n^{(i)}, \left(\underline{W}^{(i\to\mathcal{B})} : \mathcal{B} \in \mathcal{B}^{(i)}\right), \underline{T}^{(i)}\right)$ to $\breve{\underline{W}}^{(j\to\mathcal{K},i)}$ for each $j \in \mathcal{V}$, $\mathcal{K} \in \mathcal{B}^{(j)}$, and $i \in \mathcal{K}$. The solution $\mathcal{S}(\underline{\mathcal{N}})$ is called a $(\lambda, \varepsilon, A, R)$–solution for stacked network $\underline{\mathcal{N}}$, denoted $(\lambda, \varepsilon, A, R)$–$\mathcal{S}(\underline{\mathcal{N}})$, if $\left(\log_2 \left|\underline{W}^{(i\to\mathcal{B})}\right|\right)/(nN) = R^{(i\to\mathcal{B})}$, $I\left(\left(\underline{Z}^{(E)}\right)^n; \underline{W}\right) < nN\varepsilon$ for every $E \in A$, and the specified encoding and decoding functions imply $\Pr\left(\breve{\underline{W}}^{(j\to\mathcal{K},i)} \neq \underline{W}^{(j\to\mathcal{K})}\right) < \lambda$.

Definition 11. *The A-secure rate region $\mathcal{R}(\underline{\mathcal{N}}, A) \subseteq \mathbb{R}_+^{m(2^{m-1}-1)}$ of stacked network $\underline{\mathcal{N}}$ is the*

closure of all rate vectors R such that for any $\lambda > 0$ and any $\varepsilon > 0$, a solution $(\lambda, \varepsilon, A, R)$-$\mathcal{S}(\underline{\mathcal{N}})$ exists for sufficiently large N.

Like [31, Theorem 1], Theorem 8 shows that the capacity regions for a network $\mathcal{N}$ and its corresponding stacked version $\underline{\mathcal{N}}$ are identical and that a stacked solution yields error probability decaying exponentially to zero with the number of layers N; in this case the capacity of interest is the secure capacity. The definition of a stacked solution follows [31, Definition 5].

Definition 12. *Let a network*

$$\mathcal{N} \stackrel{def}{=} \left(\prod_{e \in \mathcal{E}} \mathcal{X}^{(e)}, \prod_{e \in \mathcal{E}} \left(p_e \left(y^{(e)} | x^{(e)} \right) p_e \left(z^{(e)} | y^{(e)} \right) \right), \prod_{e \in \mathcal{E}} \mathcal{Y}^{(e)} \times \prod_{e \in \mathcal{E}} \mathcal{Z}^{(e)} \right)$$

be given corresponding to a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Fix positive integers n and N to serve as the blocklength and stack size, respectively in the definition that follow. For each $i \in \mathcal{V}$ and $\mathcal{B} \in \mathcal{B}^{(i)}$, let $R^{(i \to \mathcal{B})}$ and $\tilde{R}^{(i \to \mathcal{B})}$ be constants with $\tilde{R}^{(i \to \mathcal{B})} \geq R^{(i \to \mathcal{B})}$. Define $W^{(i \to \mathcal{B})} = \{1, \ldots, 2^{nR^{(i \to \mathcal{B})}}\}$ and $\tilde{W}^{(i \to \mathcal{B})} = \{1, \ldots, 2^{n\tilde{R}^{(i \to \mathcal{B})}}\}$. Let $\underline{\mathcal{N}}$ be the N-fold stacked network for $\mathcal{N}$. A blocklength-n stacked solution $\underline{\mathcal{S}}(\underline{\mathcal{N}})$ to this network is defined as a set of mappings

$$\begin{aligned}
\underline{\tilde{W}}^{(i \to \mathcal{B})} &: \underline{\mathcal{W}}^{(i \to \mathcal{B})} \to \underline{\tilde{\mathcal{W}}}^{(i \to \mathcal{B})} \\
X_t^{(i)} &: \left(\mathcal{Y}^{(i)} \right)^{t-1} \times \prod_{\mathcal{B} \in \mathcal{B}^{(i)}} \tilde{\mathcal{W}}^{(i \to \mathcal{B})} \times \mathcal{T}^{(i)} \longrightarrow \mathcal{X}^{(i)} \\
\breve{\tilde{W}}^{(j \to \mathcal{K}, i)} &: \left(\mathcal{Y}^{(i)} \right)^{n} \times \prod_{\mathcal{B} \in \mathcal{B}^{(i)}} \tilde{\mathcal{W}}^{(i \to \mathcal{B})} \times \mathcal{T}^{(i)} \longrightarrow \tilde{\mathcal{W}}^{(j \to \mathcal{K})} \\
\underline{\breve{W}}^{(j \to \mathcal{K}, i)} &: \underline{\tilde{\mathcal{W}}}^{(j \to \mathcal{K})} \to \underline{\mathcal{W}}^{(j \to \mathcal{K})}
\end{aligned}$$

such that channel encoder $\underline{\tilde{W}}^{(i \to \mathcal{B})}(\cdot)$ encodes message $\underline{W}^{(i \to \mathcal{B})}$ to $\underline{\tilde{W}}^{(i \to \mathcal{B})}$, encoder $X_t^{(i)}(\cdot)$ independently encodes each dimension $\ell \in \{1, \ldots, N\}$ of outgoing messages $\underline{\tilde{W}}^{(i \to \mathcal{B})}$, received channel outputs $\underline{Y}_1^{(i)}, \ldots, \underline{Y}_{t-1}^{(i)}$, and random keys $\underline{T}^{(i)}$ to channel input $\underline{X}_t^{(i)}$, node decoder $\breve{\tilde{W}}^{(j \to \mathcal{K}, i)}(\cdot)$ independently decodes each dimension of the reconstruction $\underline{\breve{\tilde{W}}}^{(j \to \mathcal{K}, i)}$ of $\underline{\tilde{W}}^{(j \to \mathcal{K})}$ at node i, and channel decoder $\underline{\breve{W}}^{(j \to \mathcal{K}, i)}(\cdot)$ reconstructs message vector $\underline{W}^{(j \to \mathcal{K})}$ as a function of $\underline{\tilde{W}}^{(j \to \mathcal{K})}$, giving

$$\begin{aligned}
\underline{\tilde{W}}^{(i \to \mathcal{B})} &= \underline{\tilde{W}}^{(i \to \mathcal{B})}\left(\underline{W}^{(i \to \mathcal{B})} \right) \\
\underline{X}_t^{(i)}(\ell) &= X_t^{(i)} \left(\underline{Y}_1^{(i)}(\ell), \ldots, \underline{Y}_{t-1}^{(i)}(\ell), \left(\underline{\tilde{W}}^{(i \to \mathcal{B})}(\ell) : \mathcal{B} \in \mathcal{B}^{(i)} \right), \underline{T}^{(i)}(\ell) \right) \\
\underline{\breve{\tilde{W}}}^{(j \to \mathcal{K}, i)}(\ell) &= \breve{\tilde{W}}^{(j \to \mathcal{K}, i)} \left(\underline{Y}_1^{(i)}(\ell), \ldots, \underline{Y}_n^{(i)}(\ell), \left(\underline{\tilde{W}}^{(i \to \mathcal{B})}(\ell) : \mathcal{B} \in \mathcal{B}^{(i)} \right), \underline{T}^{(i)}(\ell) \right) \\
\underline{\breve{W}}^{(j \to \mathcal{K}, i)} &= \underline{\breve{W}}^{(j \to \mathcal{K}, i)}(\underline{\breve{\tilde{W}}}^{(j \to \mathcal{K}, i)}).
\end{aligned}$$

Theorem 8. *The rate regions $\mathcal{R}(\mathcal{N}, A)$ and $\mathcal{R}(\underline{\mathcal{N}}, A)$ are identical. Further, there exists a sequence of $(2^{-N\delta}, \varepsilon, A, R)$–$\underline{\mathcal{S}}(\underline{\mathcal{N}})$ stacked solutions for the stacked network $\underline{\mathcal{N}}$ for some $\delta > 0$.*

Proof. The argument to show $\mathcal{R}(\underline{\mathcal{N}}, A) \subseteq \mathcal{R}(\mathcal{N}, A)$ is identical to that of [31, Theorem 1]: given any $R \in \text{int}(\mathcal{R}(\underline{\mathcal{N}}, A))$, a blocklength-$n$ $(\lambda, \varepsilon, A, R) - \mathcal{S}(\underline{\mathcal{N}})$ solution for network $\underline{\mathcal{N}}$ is unraveled across time to achieve a blocklength-nN solution for network $\mathcal{N}$. Since the given code satisfies the causality constraints and precisely implements the operations of $\mathcal{S}(\underline{\mathcal{N}})$, the solution $\mathcal{S}(\mathcal{N})$ achieves the same rate, error probability, and secrecy on $\mathcal{N}$ as the solution $\mathcal{S}(\underline{\mathcal{N}})$ achieves on $\underline{\mathcal{N}}$, which gives the forward result.

The converse likewise follows [31, Theorem 2]. Again, fix $\varepsilon > 0$, and for any $R \in \text{int}\big(\mathcal{R}(\mathcal{N}, A)\big)$ choose $\tilde{R} \in \text{int}\big(\mathcal{R}(\mathcal{N}, A)\big)$ with $\tilde{R}^{(i\to\mathcal{B})} > R^{(i\to\mathcal{B})}$ for all $(i, \mathcal{B})$ with $R^{(i\to\mathcal{B})} > 0$. Define $\rho = \min_{i\in\mathcal{V}} \min_{\mathcal{B}\in\mathcal{B}^{(i)}} \big(\tilde{R}^{(i\to\mathcal{B})} - R^{(i\to\mathcal{B})}\big)$ and choose constant $\lambda > 0$ satisfying

$$\max_{i\in\mathcal{V}} \max_{\mathcal{B}\in\mathcal{B}^{(i)}} \tilde{R}^{(i\to\mathcal{B})}\lambda + h(\lambda) < \rho.$$

This is possible by choosing λ small enough so that $\lambda < \rho/(3\max_{i\in\mathcal{V}} \max_{\mathcal{B}\in\mathcal{B}^{(i)}} \tilde{R}^{(i\to\mathcal{B})})$ and $h(\lambda) < \rho/(3\rho)$. Since $\tilde{R}^{(i\to\mathcal{B})} > R^{(i\to\mathcal{B})}$, there exists a blocklength n such that a $(\lambda, \frac{\varepsilon}{3}, A, \tilde{R})$–$\mathcal{S}(\mathcal{N})$ single-layer solution exists. A stacked solution is built using this same $(\lambda, \frac{\varepsilon}{3}, A, R)$–$\mathcal{S}(\mathcal{N})$ single-layer solution in each layer and a randomly chosen channel code across the layers of the stack, as described in Definition 12. Precisely, for each $W^{(i\to\mathcal{B})} \in \mathcal{W}^{(i\to\mathcal{B})}$, codeword $\tilde{W}^{(i\to\mathcal{B})}$ is chosen independently and uniformly at random from $\tilde{\mathcal{W}}^{(i\to\mathcal{B})}$. The argument proving the asymptotic decay in the expected error probability for each intended receiver ($\mathbb{E}_{\mathcal{C}}[P_e^{(n)}] \leq 2^{-N\delta'}$) [31, Theorem 2] remains unchanged; the expectation is here taken with respect to the random channel code designs for all messages $(i, \mathcal{B})$. All that remains to be done, then, is to demonstrate the security of the earlier algorithm. Towards this end, we next show that since the solution used in each layer of the stack has mutual information leakage no greater than $\frac{n\varepsilon}{3}$ for each $E \in A$, the expected value of the mutual information $\mathbb{E}_{\mathcal{C}}\left[I\left(\left(\underline{\tilde{Z}}^{(E)}\right)^n; \underline{W}\right)\right]$ using an independent randomly chosen channel code for each message $(i, \mathcal{B})$ is no greater than $\frac{nN\varepsilon}{3}$ for each $E \in A$. The eavesdropper's observation $\left(\underline{\tilde{Z}}^{(E)}\right)^n$ is denoted with a tilde since each single-layer solution is applied to a channel-coded message $\underline{\tilde{W}}(\ell) = \big(\underline{\tilde{W}}^{(i\to\mathcal{B})}(\ell) : i \in \mathcal{V}, \mathcal{B} \in \mathcal{B}^{(i)}\big)$, where $\underline{\tilde{W}}^{(i\to\mathcal{B})} = \underline{\tilde{W}}^{(i\to\mathcal{B})}(\underline{W}^{(i\to\mathcal{B})})$ as described in Definition 12. Again, expectation $\mathbb{E}_{\mathcal{C}}$ denotes the expectation with respect to the random channel code design. A specific instance of each channel code is chosen later in the argument that follows. The mutual information for a given $E \in A$ is bounded as

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}}\left[I\left(\left(\underline{\tilde{Z}}^{(E)}\right)^n; \underline{W}\right)\right] \\
&\overset{(a)}{\leq} \mathbb{E}_{\mathcal{C}}\left[I\left(\left(\underline{\tilde{Z}}^{(E)}\right)^n; \underline{\tilde{W}}\right)\right]
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}_{\mathcal{C}}\left[I\left(\left(\underline{\tilde{Z}}^{(E)}(1)\right)^n, \ldots, \left(\underline{\tilde{Z}}^{(E)}(N)\right)^n; \underline{\tilde{W}}(1), \ldots, \underline{\tilde{W}}(N)\right)\right] \\
&= \mathbb{E}_{\mathcal{C}}\left[H\left(\left(\underline{\tilde{Z}}^{(E)}(1)\right)^n, \ldots, \left(\underline{\tilde{Z}}^{(E)}(N)\right)^n\right) - H\left(\left(\underline{\tilde{Z}}^{(E)}(1)\right)^n, \ldots, \left(\underline{\tilde{Z}}^{(E)}(N)\right)^n \middle| \underline{\tilde{W}}(1), \ldots, \underline{\tilde{W}}(N)\right)\right] \\
&\overset{(b)}{\leq} \mathbb{E}_{\mathcal{C}}\left[\sum_{\ell=1}^{N} H\left(\left(\underline{\tilde{Z}}^{(E)}(\ell)\right)^n\right) - \sum_{\ell=1}^{N} H\left(\left(\underline{\tilde{Z}}^{(E)}(\ell)\right)^n \middle| \underline{\tilde{W}}(1), \ldots, \underline{\tilde{W}}(N), \left(\underline{\tilde{Z}}^{(E)}(1)\right)^n, \ldots, \left(\underline{\tilde{Z}}^{(E)}(\ell-1)\right)^n\right)\right] \\
&\overset{(c)}{=} \mathbb{E}_{\mathcal{C}}\left[\sum_{\ell=1}^{N} H\left(\left(\underline{\tilde{Z}}^{(E)}(\ell)\right)^n\right) - \sum_{\ell=1}^{N} H\left(\left(\underline{\tilde{Z}}^{(E)}(\ell)\right)^n \middle| \underline{\tilde{W}}(\ell)\right)\right] \\
&= \sum_{\ell=1}^{N} \mathbb{E}_{\mathcal{C}}\left[I\left(\left(\underline{\tilde{Z}}^{(E)}(\ell)\right)^n; \underline{\tilde{W}}(\ell)\right)\right] \overset{(d)}{<} \frac{nN\varepsilon}{3},
\end{aligned}
$$

where (a) holds due to the data processing inequality since $\underline{W} \to \underline{\tilde{W}} \to \left(\underline{\tilde{Z}}^{(E)}\right)^n$ forms a Markov chain, (b) holds by the chain rule and the fact that conditioning reduces entropy, (c) holds by the independence of the copies of network $\mathcal{N}$ in the N layers of N-fold stacked network $\underline{\mathcal{N}}$ and the independent application of solution $\mathcal{S}(\mathcal{N})$ in each layer, and (d) holds since $\mathcal{S}(\mathcal{N})$ is a $(\lambda, \varepsilon/3, A, R)$ secure solution.

Thus $\mathbb{E}_{\mathcal{C}}\left[P_e^{(n)}\right] \leq 2^{-N\delta'}$ and $\mathbb{E}_{\mathcal{C}}\left[I\left(\left(\underline{\tilde{Z}}^{(E)}\right)^n; \underline{W}\right)\right] \leq \frac{nN\varepsilon}{3}$. It remains to show that there is a specific instance for the choice of each channel code such that both the probability of error and the mutual information are not too large. We prove this using Markov's inequality [59, Section 8.1] to show that the probability, under the random channel code design, that $P_e^{(n)} \geq 3 \cdot 2^{-N\delta'}$ or $I\big((\underline{Z}^{(E)})^n; \underline{W}\big) \geq nN\varepsilon$ is strictly less than 1. Precisely,

$$
\begin{aligned}
&\Pr\left(\left\{P_e^{(n)} \geq 3 \cdot 2^{-N\delta'}\right\} \cup \left\{I\left(\left(\underline{\tilde{Z}}^{(E)}\right)^n; \underline{W}\right) \geq nN\varepsilon\right\}\right) \\
&\overset{(a)}{\leq} \Pr\left(P_e^{(n)} \geq 3 \cdot 2^{-N\delta'}\right) + \Pr\left(I\left(\left(\underline{\tilde{Z}}^{(E)}\right)^n; \underline{W}\right) \geq nN\varepsilon\right) \\
&\overset{(b)}{\leq} \frac{\mathbb{E}_{\mathcal{C}}\left[P_e^{(n)}\right]}{3 \cdot 2^{-N\delta'}} + \frac{\mathbb{E}_{\mathcal{C}}\left[I\left(\left(\underline{\tilde{Z}}^{(E)}\right)^n; \underline{W}\right)\right]}{nN\varepsilon} \\
&\overset{(c)}{\leq} \frac{2}{3} < 1,
\end{aligned} \tag{3.1}
$$

where inequality (a) is the union bound, (b) is Markov's inequality, and (c) applies our earlier bounds on $\mathbb{E}_{\mathcal{C}}\left[P_e^{(n)}\right]$ and $\mathbb{E}_{\mathcal{C}}\left[I\left(\left(\underline{\tilde{Z}}^{(E)}\right)^n; \underline{W}\right)\right]$. Therefore, for sufficiently large N there must be at least one instance of the collection of codes with error probability no greater than $3 \cdot 2^{-N\delta'} < 2^{-N\delta}$ ($\delta = \delta'/2$) and mutual information no greater than $nN\varepsilon$. □

3.3 Intuition and Summary of Results

We derive lower and upper bounds on the secrecy capacity region of a noisy network of wiretap channels in the presence of an eavesdropper that has access to the degraded outputs from an unknown

subset $E \in A$ of the wiretap channels in the network. In some cases, these lower and upper bounds are identical, showing equivalence of secure capacity between noisy and noiseless wiretap networks. We derive these results using an approach from [31, 32], which shows that the capacity of a network $\mathcal{N}_A$ is a subset of the capacity of network $\mathcal{N}_B$ by showing that any solution for $\mathcal{N}_A$ can be modified to obtain a solution for $\mathcal{N}_B$ with similar performance.

In Theorem 9, we show that for any network $\mathcal{N}$ of wiretap channels and any edge $\bar{e} \in \mathcal{E}$, replacing channel $\mathcal{C}_{\bar{e}}$ with a noiseless degraded wiretap channel $\mathcal{C}_{\bar{e}}(R_c, R_p)$, with $R_c > \max_{p(x^{(\bar{e})})} I(X^{(\bar{e})}; Y^{(\bar{e})}) - \max_{p(x^{(\bar{e})})} I(X^{(\bar{e})}; Z^{(\bar{e})})$ and $R_p > \max_{p(x^{(\bar{e})})} I(X^{(\bar{e})}; Z^{(\bar{e})})$, as shown in Figure 3.1(b), yields a new network $\mathcal{N}_{\bar{e}}(R_c, R_p)$ whose capacity region is a superset of the secure capacity region of $\mathcal{N}$. Theorem 9 is similar to [32, Theorem 5], which shows that for traditional (rather than secrecy) capacity, replacing a noisy degraded broadcast channel with the same noiseless counterpart yields an upper bounding network. The proof of Theorem 9, which extends the argument of [32, Theorem 5] from traditional to secure capacity, appears in Section 3.4.1.

Theorem 9. *Consider a network $\mathcal{N}$ and an adversarial set $A \subseteq \mathcal{P}(\mathcal{E})$. If*

$$R_c > \max_{p(x^{(\bar{e})})} I(X^{(\bar{e})}; Y^{(\bar{e})}) - \max_{p(x^{(\bar{e})})} I(X^{(\bar{e})}; Z^{(\bar{e})})$$
$$R_p > \max_{p(x^{(\bar{e})})} I(X^{(\bar{e})}; Z^{(\bar{e})}),$$

then $\mathcal{R}(\mathcal{N}, A) \subseteq \mathcal{R}(\mathcal{N}_{\bar{e}}(R_c, R_p), A)$.

Theorem 10 proves that the upper bound shown in Theorem 9 is tight in both the case where $\bar{e}$ is a secure link ($\bar{e} \notin E$ for all $E \in A$) and the case where link $\bar{e}$ is not simultaneously eavesdropped with any other link ($\bar{e} \in E$ implies $|E| = 1$). The proof of Theorem 10 appears in Section 3.4.2.

Theorem 10. *Consider a network $\mathcal{N}$, an adversarial set $A \subseteq \mathcal{P}(\mathcal{E})$, and a single link $\bar{e} \in \mathcal{E}$. Let*

$$R_c = \max_{p(x^{(\bar{e})})} I(X^{(\bar{e})}; Y^{(\bar{e})}) - \max_{p(x^{(\bar{e})})} I(X^{(\bar{e})}; Z^{(\bar{e})})$$
$$R_p = \max_{p(x^{(\bar{e})})} I(X^{(\bar{e})}; Z^{(\bar{e})}).$$

If $\bar{e}$ is invulnerable to wiretapping ($\bar{e} \notin E$ for all $E \in A$) or is not simultaneously wiretapped with other links ($\bar{e} \in E$ implies $|E| = 1$), then $\mathcal{R}(\mathcal{N}, A) = \mathcal{R}(\mathcal{N}_{\bar{e}}(R_c, R_p), A)$.

Example 1 demonstrates the applications of Theorem 9 and 10 in an example. While Theorem 9 seems to be tight on many small examples, it is not always tight when the replaced link appears in one or more eavesdropping sets of size greater than 1 ($\bar{e} \in E$ for some $E \in A$ such that $|E| > 1$), as illustrated by Example 1. It remains an open problem whether one can find tight noiseless network models in this case.

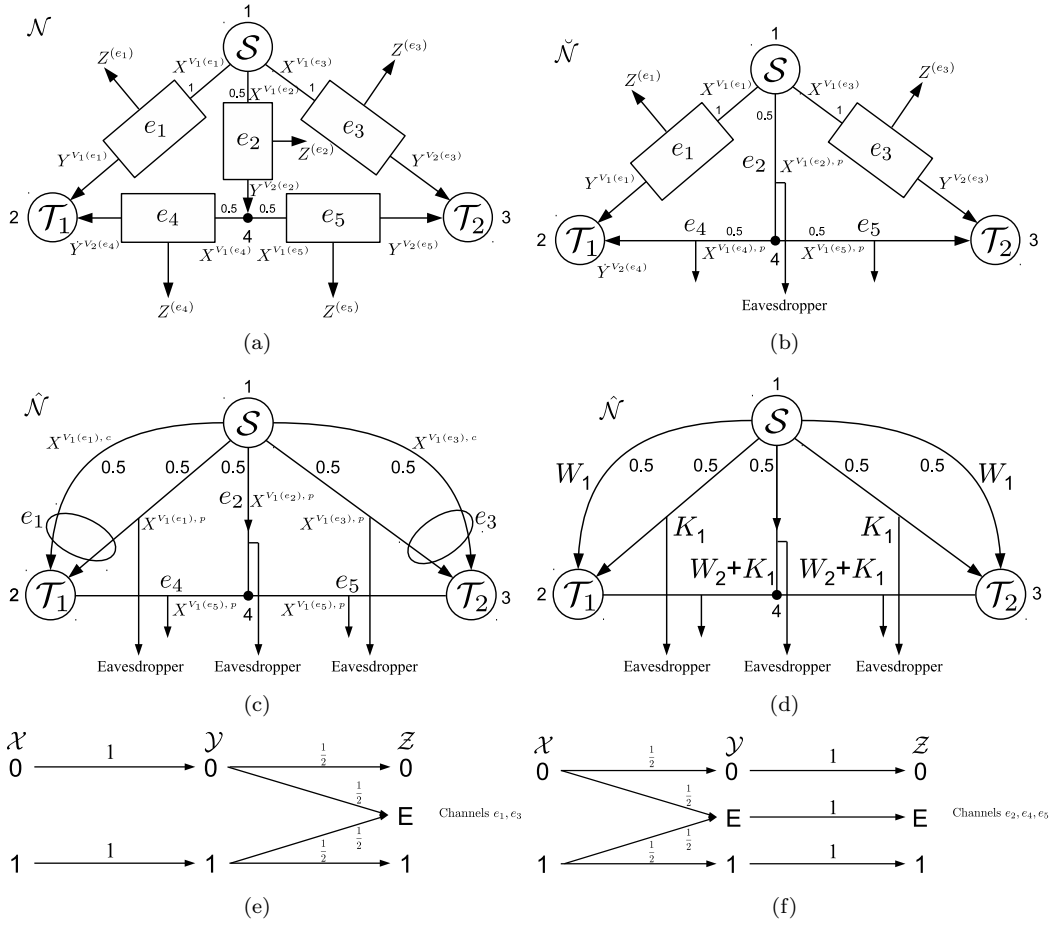


Figure 3.2: (a) The network for Example 1 and (b) its equivalent model by replacing channels e_2, e_4, and e_5 by their equivalent noiseless links by Theorem 10 (rate-0 links are omitted from the model). (c) The noiseless model of (a) by applying Theorem 9 and (d) the secrecy capacity achieving code for the network in (c). (e), (f) The channel distributions for independent degraded wiretap channels e_1, e_3 and e_2, e_4, e_5 respectively.

Example 1. *Figure 3.2(a) shows a network. Channels* $e_1 = (1, 2, 1)$, $e_2 = (1, 4, 1)$, $e_3 = (1, 3, 1)$, $e_4 = (4, 2, 1)$, *and* $e_5 = (4, 3, 1)$ *are independent degraded binary wiretap channels. Channels* e_1 *and* e_3 *have erasure probability* 0 *at each intended receiver and erasure probability* $\frac{1}{2}$ *at each wiretap output, as shown in Figure 3.2(e). Channels* e_2, e_4, *and* e_5 *have erasure probability* $\frac{1}{2}$, *with identical outputs for their intended and eavesdropped outputs, as shown in Figure 3.2(f). We wish to employ the network to securely transmit a single multicast from source* $\mathcal{S}$ *at node* 1 *to terminals* $\mathcal{T}_1$ *and* $\mathcal{T}_2$ *at nodes* 2 *and* 3. *We therefore set* $R^{(i\to\mathcal{B})} = 0$ *for all* $(i, \mathcal{B}) \neq (1, \{2, 3\})$ *and then consider the point* $R \in \mathcal{R}(\mathcal{N}, A)$ *that maximizes* $R^{(1\to\{2,3\})}$ *subject to these constraints. The eavesdropper can listen in on either both* e_1 *and* e_3 *or just* e_2, *giving* $A = \big\{\{e_1, e_3\}, \{e_2\}\big\}$. *When the eavesdropper overhears* e_1 *and* e_3, *it has access to the degraded output of these links. Since* $Y^{(e_2)} = Z^{(e_2)}$ *with probability* 1, *when the eavesdropper overhears link* e_2, *it receives everything heard by the intended receiver over this link. The network* $\breve{\mathcal{N}}$ *shown in Figure 3.2(b) has secrecy capacity under adversarial set* $A = \big\{\{e_1, e_3\}, \{e_2\}\big\}$ *identical to that of the network in Figure 3.2(a)* $\big(\mathcal{R}(\mathcal{N}, A) = \mathcal{R}(\breve{\mathcal{N}}, A)\big)$ *and is obtained by three applications of Theorem 9. Here channel* $\mathcal{C}_{e_4}$ *and* $\mathcal{C}_{e_5}$ *have been replaced by channel* $\mathcal{C}(\frac{1}{2}, 0)$ *since channels* e_4 *and* e_5 *are invulnerable to eavesdropping* ($e_4, e_5 \notin E$ *for all* $E \in A$). *Likewise* $\mathcal{C}_{e_2}$ *has been replaced by* $\mathcal{C}(0, \frac{1}{2})$ *since* e_2 *cannot be simultaneously eavesdropped with any other channel* ($e_2 \in E$ *implies* $|E| = 1$) *and has* 0 *confidential bits. The noiseless network* $\hat{\mathcal{N}}$ *is an upper bounding model for the network in Figure 3.2(b) (and therefore also an upper bounding model for the network in Figure 3.2(a), giving* $\mathcal{R}(\mathcal{N}, A) = \mathcal{R}(\breve{\mathcal{N}}, A) \subseteq \mathcal{R}(\hat{\mathcal{N}}, A)$), *and is obtained by two applications of Theorem 9. These applications replace channels* e_1 *and* e_3 *by their upper bounding models. We therefore bound the maximal rate* $R^{1\to\{2,3\}}$ *achievable in* $\mathcal{N}$ *and* $\breve{\mathcal{N}}$ *by finding the corresponding maximal multicast rate in* $\hat{\mathcal{N}}$.

*A rate-*1 *blocklength-*2 *code for network* $\hat{\mathcal{N}}$ *is shown in Figure 3.2(d). The message* $W^{(1\to\{2,3\})} \in \{0, 1\}^2$ *is broken into a pair of messages* $W^{(1\to\{2,3\})} = \big(W_1, W_2\big) \in \{0, 1\}^2$ *with* $H\big(W_1\big) = H\big(W_2\big) = 1$ *and* $H\big(W_1, W_2\big) = 2$. *Random key* $K_1 \in \{0, 1\}$ *is chosen uniformly at random and independently of* $\big(W_1, W_2\big)$. *The code is secure since* $I\big(W_1, W_2; K_1\big) = 0$ *and* $I\big(W_1, W_2; W_2 + K_1\big) = 0$. *This code achieves the secure multicast capacity from* $\mathcal{S}$ *to* $\{\mathcal{T}_1, \mathcal{T}_2\}$ *of network* $\hat{\mathcal{N}}$ *by Lemma 6 in Appendix C. Lemma 7 in the same appendix proves that the noisy network* $\mathcal{N}$ *of Figure 3.2(a) has multicast secrecy capacity at most* 0.875.

To build some intuition about the result, notice that our capacity-achieving code for $\hat{\mathcal{N}}$ transmits the same key over a pair of noiseless links (e_1 and e_3 in $\hat{\mathcal{N}}$). Direct emulation of this solution over the corresponding noisy links in $\breve{\mathcal{N}}$ network in Figure 3.2(a) fails to maintain security. Specifically, if the same input is transmitted over channels e_1 and e_3 ($X_t^{(e_1)} = X_t^{(e_3)}$ for all $t \in \{1, \ldots, n\}$), then an eavesdropper accessing $E = \{e_1, e_3\}$ sees independent channel outputs $Z_t^{(e_1)}$ and $Z_t^{(e_3)}$ resulting from the same channel input $X_t^{(e_1)} = X_t^{(e_3)}$ at each time t. Since each transmitted bit is erased with probability $\frac{1}{2}$ and the erasure events are independent by assumption, an eavesdropper that wiretaps

both e_1 and e_3 is expected to receive roughly 75% of the transmitted information bits. Consequently, a key of rate 0.5 is not enough to completely protect $W^{(1\to\{2,3\})}$ from the eavesdropper in this case. While it is possible to avoid this problem on a single eavesdropped link by removing redundancy before transmission, the problem is more difficult to address in the case where the eavesdropper has access to multiple channels simultaneously. The problem here is that transmitting correlated information on multiple channels may be necessary to achieve the secure capacity in the noiseless case, but the same strategy may fail in the noisy case since the eavesdropper may be able to take advantage of the correlation between different channels' inputs.

Theorems 11 and 12 provide two different lower bounds for the case of multiple wiretapped channels. These bounds are designed to guarantee that the links to the eavesdropper are filled to capacity.

Lower bound – Model 1

The first lower bound results from removing the public portion of the upper bounding model. The lower bound is achievable since it is always possible to simply avoid the transmission of any rate on channel $\bar{e}$ that can be overheard by the eavesdropper. The proof of Theorem 11, appears in Section 3.4.

Theorem 11. *Consider a network $\mathcal{N}$, an adversarial set $A \subseteq \mathcal{P}(\mathcal{E})$, and a single link $\bar{e} \in \mathcal{E}$. If*

$$R_c < \max_{p(x^{(\bar{e})})} I(X^{(\bar{e})}; Y^{(\bar{e})}) - \max_{p(x^{(\bar{e})})} I(X^{(\bar{e})}; Z^{(\bar{e})})$$

then $\mathcal{R}(\mathcal{N}_{\bar{e}}(R_c, 0), A) \subseteq \mathcal{R}(\mathcal{N}, A)$.

The lower bound in Lemma 11 is not tight in general. As a result, we do not use it to bound all channels but instead apply it to a selective sequence of channels from $\mathcal{E}$. Notice that the model $\mathcal{C}_{\bar{e}}(R_c, 0)$ for channel $\mathcal{C}_{\bar{e}}$ in Theorem 11 sets the public rate R_p to zero. This effectively removes $\bar{e}$ from all eavesdropping sets $E \in A$, giving a new adversarial set $A' = \big\{E \backslash \{\bar{e}\} : E \in A\big\}$. Repeated application of Theorem 11 on a carefully chosen sequence of channels enable us to reduce all eavesdropping sets to size at most one. Once this is accomplished, we can use the equivalence result of Theorem 10 to replace the remaining noisy channels.

Lower bound – Model 2

In this model we bound the secrecy capacity region of network $\mathcal{N}$ with adversarial set $A \subseteq \mathcal{P}(\mathcal{E})$ by deriving a relationship between that secrecy capacity and the traditional capacity of a noiseless communication network called the A-enhanced network $\mathcal{N}(A)$ defined below and illustrated by Figure 3.3.

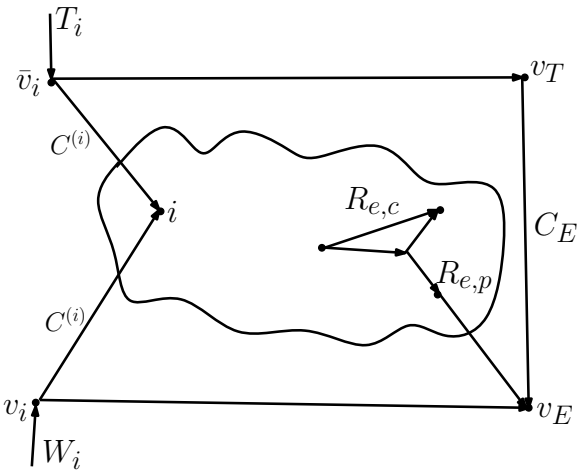


Figure 3.3: The A-enhanced network $\mathcal{N}(A)$.

Definition 13. *Consider network $\mathcal{N}$ on graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Define rate vector $\check{R}_{c,p} = \big((\check{R}_{e,c}, \check{R}_{e,p}) : e \in \mathcal{E}\big)$, and fix an adversarial set $A \subseteq \mathcal{P}(\mathcal{E})$. The A-enhanced network $\mathcal{N}(\check{R}_{c,p}, A)$ on graph $\check{\mathcal{G}} = (\check{\mathcal{V}}, \check{\mathcal{E}})$ is defined as follows:*

1. *$\check{\mathcal{V}} = \mathcal{V} \cup \big\{v_i : i \in \mathcal{V}\big\} \cup \big\{\bar{v}_i : i \in \mathcal{V}\big\} \cup \big\{v_E : E \in A\big\} \cup \{v_T\}$. For each $i \in \mathcal{V}$ we call v_i and $\bar{v}_i$ the i^{th} message node and random key node of network $\mathcal{N}(\check{R}_{c,p}, A)$. For each $E \in A$, node v_E is called an eavesdropper node. Node v_T is called the overall key node.*

2. *$\check{\mathcal{E}} = \big\{h_i : i \in \mathcal{V}\big\} \cup \big\{\bar{h}_i : i \in \mathcal{V}\big\} \cup \mathcal{E} \cup \big\{h_e : e \in \mathcal{E}\big\} \cup \big\{(v_T, v_E, 1) : E \in A\big\}$.*

For each $i \in \mathcal{V}$, h_i and $\bar{h}_i$ are noiseless hyperarcs of capacity

$$\check{C}^{(i)} = \sum_{e \in \mathcal{E}_{out}(i)} (\check{R}_{e,c} + \check{R}_{e,p}).$$

Hyperarc h_i noiselessly delivers the same rate-$\check{C}^{(i)}$ description from node v_i to all of the nodes in $\big\{i\big\} \cup \big\{v_E : E \in A\big\}$. Hyperarc $\bar{h}_i$ delivers the same rate-$\check{C}^{(i)}$ description from node $\bar{v}_i$ to both of the nodes in $\big\{i, v_T\big\}$. For each $e = (i, j, k) \in \mathcal{E}$, channel $\check{\mathcal{C}}_e$ in network is a bit pipe of capacity $R_{e,c}$ from node i to node j, and hyperarc h_e is a noiseless hyperarc of capacity $R_{e,p}$ from node i to all of the nodes in $\big\{j\big\} \cup \big\{v_E : E \in A, e \in E\big\}$; set $\big\{v_E : E \in A, e \in E\big\}$ is empty if edge e of graph $\mathcal{G}$ is invulnerable to eavesdropping. For every $E \in A$ channel $\mathcal{C}_{(v_T, v_E, 1)}$ is noiseless bit pipe of capacity

$$C_E = \sum_{e \in \mathcal{E}} (\check{R}_{e,c} + \check{R}_{e,p}) - \sum_{e \in E} \check{R}_{e,p}$$

from node v_T to node v_E.

The A-enhanced network is used for traditional (rather than secure) communication with a collection of reconstruction constraints that depend on both $\mathcal{N}$ and A. Specifically, for each $i \in \mathcal{V}$

and $\mathcal{B} \in \mathcal{B}^{(i)}$, a solution for A-enhanced network $\mathcal{N}(\check{R}_{c,p}, A)$ must deliver message $W^{(v_i \to \mathcal{B})}$ from node v_i to all of the nodes in $\mathcal{B} \in \mathcal{B}^{(i)}$, where $\mathcal{B}^{(i)}$ is the receivers set for node $i \in \mathcal{V}$ in network $\mathcal{N}$ (rather than network $\mathcal{N}(\check{R}_{c,p}, A)$)[3]. In addition, a solution for network $\mathcal{N}(\check{R}_{c,p}, A)$ must deliver random keys $T^{(i)} \in \mathcal{T}^{(i)} = \{1, \ldots, 2^{n\check{C}^{(i)}}\}$ from node $\bar{v}_i$ to nodes $\{v_E : E \in A\}$. We therefore define a solution $\mathcal{S}(\mathcal{N}(\check{R}_{c,p}, A))$ for an A-enhanced network $\mathcal{N}(\check{R}_{c,p}, A)$ as follows

Definition 14. *Let $\mathcal{N}(\check{R}_{c,p}, A)$ be the A-enhanced network for network $\mathcal{N}$ and adversarial set $A \subseteq \mathcal{P}(\mathcal{E})$. A blocklength-$n$ solution $\mathcal{S}(\mathcal{N}(\check{R}_{c,p}, A))$ to network $\mathcal{N}(\check{R}_{c,p}, A)$ is defined as a set of encoding and decoding functions*

$$
\begin{aligned}
(X^{(v_i)})^n &: \prod_{\mathcal{B} \in \mathcal{B}^{(i)}} \mathcal{W}^{(v_i \to \mathcal{B})} \longrightarrow (\mathcal{X}^{(v_i)})^n \\
(X^{(\bar{v}_i)})^n &: \mathcal{T}^{(i)} \longrightarrow (\mathcal{X}^{(\bar{v}_i)})^n \\
X_t^{(i)} &: \left(\mathcal{Y}^{(h_i)}\right)^{t-1} \times \left(\mathcal{Y}^{(\bar{h}_i)}\right)^{t-1} \times \prod_{e \in \mathcal{E}_{in}(i)} \left(\left(\mathcal{Y}^{(e)}\right)^{t-1} \times \left(\mathcal{Y}^{(h_e)}\right)^{t-1}\right) \longrightarrow \mathcal{X}^{(i)} \times \mathcal{X}^{(h_e)} \\
X_t^{(v_T)} &: \prod_{i \in \mathcal{V}} \left(\mathcal{Y}^{(\bar{h}_i)}\right)^{t-1} \longrightarrow \prod_{E \in A} \mathcal{X}^{(v_T, v_E, 1)} \\
\check{W}^{(v_j \to \mathcal{K}, i)} &: \left(\mathcal{Y}^{(i)}\right)^n \times \left(\mathcal{Y}^{(h_i)}\right)^n \times \left(\mathcal{Y}^{(\bar{h}_i)}\right)^n \times \prod_{e \in \mathcal{E}_{in}(i)} \left(\mathcal{Y}^{(h_e)}\right)^n \longrightarrow \mathcal{W}^{(v_j \to \mathcal{K})} \\
\check{T}_E &: \prod_{i \in \mathcal{V}} \left(\mathcal{Y}^{(h_i)}\right)^n \times \prod_{e \in E} \left(\mathcal{Y}^{(h_e)}\right)^n \times \left(\mathcal{Y}^{(v_T, v_E, 1)}\right)^n \longrightarrow \prod_{i \in \mathcal{V}} \mathcal{T}^{(i)} \\
\check{W}_E &: \prod_{i \in \mathcal{V}} \left(\mathcal{Y}^{(h_i)}\right)^n \times \prod_{e \in E} \left(\mathcal{Y}^{(h_e)}\right)^n \times \left(\mathcal{Y}^{(v_T, v_E, 1)}\right)^n \longrightarrow \prod_{i \in \mathcal{V}} \prod_{\mathcal{B} \in \mathcal{B}(i)} \mathcal{W}^{(v_i \to \mathcal{B})}.
\end{aligned}
$$

For each $i \in \mathcal{V}$, encoder $(X^{(v_i)})^n$ at node v_i maps $\left(W^{(v_i \to \mathcal{B})} : \mathcal{B} \in \mathcal{B}^{(i)}\right)$ to $(X^{(v_i)})^n = (X^{(h_i)})^n$ (since node v_i has a single output to noiseless hyperarc h_i), while encoder $(X^{(\bar{v}_i)})^n$ at node $\bar{v}_i$ maps $T^{(i)}$ to $(X^{(\bar{v}_i)})^n = (X^{(\bar{h}_i)})^n$ (since node $\bar{v}_i$ has a single output to noiseless hyperarc $\bar{h}_i$). For each $i \in \mathcal{V}$, encoder $X_t^{(i)}$ at node i maps past network outputs $\left((Y^{(h_i)})^{t-1}, (Y^{(\bar{h}_i)})^{t-1}, \left((Y^{(e)})^{t-1}, (Y^{(h_e)})^{t-1} : e \in \mathcal{E}_{in}(i)\right)\right)$ to $X_t^{(i)}$ and $X_t^{(h_e)}$. Encoder $X_t^{(v_T)}$ at node v_T maps past network outputs $(Y^{(h_i)})^{t-1}$ to $\left(X_t^{(v_T, v_E, 1)} : E \in A\right)$. For each $j \in \mathcal{V}$, $\mathcal{K} \in \mathcal{B}^{(j)}$ and $i \in \mathcal{K}$, decoder $\check{W}^{(v_j \to \mathcal{K}, i)}$ maps:

$$\left((Y^{(i)})^n, (Y^{(h_i)})^n, (Y^{(\bar{h}_i)})^n, \left((Y^{(h_e)})^n : e \in \mathcal{E}_{in}(i)\right)\right)$$

to $\check{W}^{(v_j \to \mathcal{K}, i)}$. For each $E \in A$, decoders $\check{T}_E$ and $\check{W}_E$ map $\left((Y^{(h_i)})^n, (Y^{(h_i)})^n, (Y^{(v_T, v_E, 1)})^n\right)$ to reproductions $\left(T^{(1)}, \ldots, T^{(m)}\right)$ and $\left(W^{(v_i \to \mathcal{B})} : i \in \mathcal{V}, \mathcal{B} \in \mathcal{B}^{(i)}\right)$. Given a rate vector $R = \left(R^{(i \to \mathcal{B})} : i \in \mathcal{V}, \mathcal{B} \in \mathcal{B}^{(i)}\right)$, the solution $\mathcal{S}(\mathcal{N}(R, A))$ of blocklength n is called a (λ, R)–solution, denoted (λ, R)-$\mathcal{S}(\mathcal{N}(\check{R}_{c,p}, A))$, if $\log_2\left(\left|W^{(v_i \to \mathcal{B})}\right|\right)/n = R^{(i \to \mathcal{B})}$, and the specified encoding and decoding functions imply $\Pr\left(\check{W}^{(v_j \to \mathcal{K}, i)} \neq W^{(v_j \to \mathcal{K})}\right) < \lambda$ for every $j \in \mathcal{V}$, $\mathcal{K} \in \mathcal{B}^{(j)}$, and $i \in \mathcal{K}$ and

[3]The use of rate $R^{(i \to \mathcal{B})}$ for message $W^{(v_i \to \mathcal{B})}$ (*i.e.*, $W^{(i \to \mathcal{B})} \in \mathcal{W}^{(i \to \mathcal{B})} = \{1, \ldots, 2^{nR^{(i \to \mathcal{B})}}\}$) is used to relate the capacity region $\mathcal{R}\big(\mathcal{N}(\check{R}_{c,p}, A)\big)$ for $\mathcal{N}(\check{R}_{c,p}, A)$ to the A-secure capacity region $\mathcal{R}(\mathcal{N}, A)$ for $\mathcal{N}$ in Theorem 12.

$\Pr\left(\breve{T}_E \neq T \cup \breve{W}_E \neq W\right) < \lambda$ *for every* $E \in A$.

Definition 15. *The rate region* $\mathcal{R}(\mathcal{N}(R_{c,p}, A)) \subseteq \mathbb{R}_+^{m(2^{m-1}-1)}$ *of the* A*-enhanced network* $\mathcal{N}(\breve{R}_{c,p}, A)$ *of network* $\mathcal{N}$ *is the closure of all rate vectors* R *such that for any* $\lambda > 0$*, a solution* (λ, R)*–*$\mathcal{S}(\mathcal{N}(\breve{R}_{c,p}, A))$ *exists.*

Theorem 12. *Consider network* $\mathcal{N}$ *on graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ *and an adversarial set* $A \subseteq \mathcal{P}(\mathcal{E})$*. Let* $\mathcal{N}(R_{c,p}, A)$ *be the* A*-enhanced network of network* $\mathcal{N}$*. If for every* $e \in \mathcal{E}$

$$\breve{R}_{e,p} < \max_{p(x)} I(X^{(e)}; Z^{(e)})$$
$$\breve{R}_{e,c} < \max_{p(x)} I(X^{(e)}; Y^{(e)}) - \max_{p(x)} I(X^{(e)}; Z^{(e)}),$$

then $\mathcal{R}(\mathcal{N}(\breve{R}_{c,p}, A)) \subseteq \mathcal{R}(\mathcal{N}, A)$.

Unlike the rest of the results, where changing a single wiretap channel $\mathcal{C}_{\bar{e}}$ to its noiseless counterpart $\mathcal{C}_{\bar{e}}(R_c, R_p)$ results in an equivalent or bounding network, Theorem 12 requires all wiretap channels in the noisy network $\mathcal{N}$ to be changed to noiseless channels in order to obtain a lower bounding network. Intuitively, this is because our construction requires the eavesdropper $E \in A$ to decode all sources of randomness in the network, which is not possible generally for noisy networks where the entropy of the noise can be potentially infinite. If we wish to replace only some noisy channels by their noiseless counterparts then Theorem 11 should be used. When all channels are to be replaced Theorem 12 can be used, potentially leading to a tighter bound.

3.4 Proofs

In the proofs following, for notational convenience we shorten notation as $X = X^{(\bar{e})}$, $Y = Y^{(\bar{e})}$, and $Z = Z^{(\bar{e})}$.

3.4.1 Proof of Theorem 9

Proof of Theorem 9. By Theorem 8 it suffices to prove $\mathcal{R}(\underline{\mathcal{N}}, A) \subseteq \mathcal{R}(\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p), A)$. Fix any rate vector R in the relative interior of the A–secure rate region of network $\mathcal{N}$, *i.e.* $R \in \operatorname{int}(\mathcal{R}(\mathcal{N}, A))$. Choose some $\tilde{R} \in \operatorname{int}(\mathcal{R}(\mathcal{N}, A))$ for which $\tilde{R}^{(i\to\mathcal{B})} > R^{(i\to\mathcal{B})}$ for all $i \in \mathcal{V}$ and $\mathcal{B} \in \mathcal{B}^{(i)}$ with $R^{(i\to\mathcal{B})} > 0$. Then for any $\lambda > 0$ and $\varepsilon > 0$ there exists a $(\lambda, \varepsilon, A, \tilde{R})$–$\mathcal{S}(\mathcal{N})$ solution for network $\mathcal{N}$; let n be the blocklength of that solution. Then $\Pr\left(\breve{\tilde{W}}^{(j\to\mathcal{K},i)} \neq \tilde{W}^{(j\to\mathcal{K})}\right) < \lambda$ for all $(j, \mathcal{K})$ with $\mathcal{K} \in \mathcal{B}^{(j)}$, $i \in \mathcal{K}$ and $R^{(j\to\mathcal{K})} > 0$, and $I\big((\tilde{Z}^E)^n; \tilde{W}\big) < n\varepsilon$ for all $E \in A$. We use the single-layer solution $(\lambda, \varepsilon, A, \tilde{R})$–$\mathcal{S}(\mathcal{N})$ for network $\mathcal{N}$ and a carefully chosen λ to build a random N-layer stacked solution $(2^{-N\delta}, \varepsilon, A, R)$–$\underline{\mathcal{S}}(\underline{\mathcal{N}})$ for network $\underline{\mathcal{N}}$ as described in the proof of Theorem 8. As in that

proof of Theorem 8, the error probability and secrecy bounds are calculated in expectation over a random code choice and then the existence of at least one single good code is proved.

Theorem 5 of Section V in [32] shows that in the communication, rather than secrecy capacity problem, we can build a sequence of rate-R random codes for network $\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p)$ with error probability approaching zero. We apply the same code construction here; the code construction combines the random stacked code $(2^{-N\delta}, \varepsilon, A, R)$–$\underline{\mathcal{S}}(\underline{\mathcal{N}})$ across network $\underline{\mathcal{N}}$ with the aid of $2n$ random emulation code encoders $\big\{\big(\alpha_{t,N}^{(p)}, \alpha_{t,N}^{(c)}\big)\big\}_{t=1}^{n}$ and $2n$ corresponding decoders $\big\{\big(\beta_{t,N}^{(p)}, \beta_{t,N}^{(c)}\big)\big\}_{t=1}^{n}$ of blocklength N. The random codes $\big\{\big(\alpha_{t,N}^{(p)}, \alpha_{t,N}^{(c)}\big), \big(\beta_{t,N}^{(p)}, \beta_{t,N}^{(c)}\big)\big\}_{t=1}^{n}$ are constructed as follows. The random decoder $\beta_{t,N}^{(p)} : \{0,1\}^{NR_p} \to \underline{\tilde{\mathcal{Z}}}$ maps each sequence of NR_p bits to a codeword drawn i.i.d. according to distribution $\prod_{\ell=1}^{N} p\big(\tilde{\underline{z}}(\ell)\big)$. The tilde superscript on $\underline{\tilde{Z}}$ denotes the fact that the underlying stacked code $(2^{-N\delta}, \varepsilon, A, R)$–$\underline{\mathcal{S}}(\underline{\mathcal{N}})$ operates on every layer of the stacked network at rate $\tilde{R}$. For each $\tilde{b}^{(p)} \in \{0,1\}^{NR_p}$, the random design of decoder $\beta_{t,N}^{(c)} : \{0,1\}^{NR_c} \times \{0,1\}^{NR_p} \to \underline{\tilde{\mathcal{Y}}}$ draws codewords $\beta_{t,N}^{(c)}(1, \tilde{b}^{(p)}), \ldots, \beta_{t,N}^{(c)}(2^{NR_c}, \tilde{b}^{(p)})$ i.i.d. according to distribution $\prod_{\ell=1}^{N} p\big(\underline{\tilde{y}}(\ell) | \beta_{t,N}^{(p)}(\tilde{b}^{(p)}, \ell)\big)$, where $\beta_{t,N}^{(p)}(\tilde{b}^{(p)}, \ell)$ denotes the ℓ^{th} component of N-vector $\beta_{t,N}^{(p)}(\tilde{b}^{(p)})$. For each $\underline{\tilde{X}}_t \in \underline{\tilde{\mathcal{X}}}$ random encoder $\alpha_{t,N}^{(p)} : \underline{\tilde{\mathcal{X}}} \to \{0,1\}^{NR_p}$ chooses index $\alpha_{tN}^{(p)}(\underline{\tilde{x}})$ uniformly at random from those $\tilde{b}^{(p)} \in \{0,1\}^{NR_p}$ for which $\big(\underline{\tilde{x}}, \beta_{tN}^{(p)}(\tilde{b}^{(p)})\big) \in \hat{A}_{\epsilon_1,t}^{(N)}(\tilde{X}, \tilde{Z})$, whereas for each $\tilde{b}^{(p)} \in \{0,1\}^{NR_p}$ encoder $\alpha_{t,N}^{(c)} : \underline{\tilde{\mathcal{X}}} \times \{0,1\}^{NR_p} \to \{0,1\}^{NR_c}$ chooses an index $\alpha_{tN}^{(p)}\big(\underline{\tilde{x}}, \tilde{b}^{(p)}\big)$ uniformly at random from those $\tilde{b}^{(c)} \in \{0,1\}^{NR_c}$ such that $\big(\underline{\tilde{x}}, \beta_{tN}^{(c)}(\tilde{b}^{(c)}, \alpha_{t,N}^{(p)}(\underline{\tilde{x}})), \beta_{tN}^{(p)}(\alpha_{t,N}^{(p)}(\underline{\tilde{x}}))\big) \in \hat{A}_{\epsilon_2,t}^{(N)}(\tilde{X}, \tilde{Y}, \tilde{Z})$, where $\hat{A}_{\epsilon_1,t}^{(N)}(\tilde{X}, \tilde{Z})$ and $\hat{A}_{\epsilon_2,t}^{(N)}(\tilde{X}, \tilde{Y}, \tilde{Z})$ are restricted typical sets, whose definitions are given in equations (D.1) and (D.2) of Appendix D.

Formally, $\mathcal{S}(\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p))$ for stacked network $\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p)$ is derived from the stacked solution $\big(2^{-N\delta}, \varepsilon, A, R\big)$–$\underline{\mathcal{S}}(\underline{\mathcal{N}})$ of stacked network $\underline{\mathcal{N}}$ as follows. Let $e = (i, j, k)$, then each component $\hat{Y}_t^{(e)}$, $e \in \mathcal{E}_{\text{in}}(\nu)$, of the network output $\underline{\hat{Y}}_t^{(\nu)}$ at time t in stacked network $\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p)$ is channel decoded to obtain

$$\underline{\tilde{Y}}_t^{(e)} = \begin{cases} \beta_{tN}^{(c)}\big(\underline{\hat{Y}}_t^{(e)}\big) & \text{if } \nu = j \\ \underline{\hat{Y}}_t^{(e)} & \text{otherwise} \end{cases} .$$

Subsequently the encoding functions $\underline{\tilde{X}}_t^{(\nu)}$ for each $\nu \in \mathcal{V}$ of code $\mathcal{S}(\underline{\mathcal{N}})$ for stacked network $\underline{\mathcal{N}}$ are applied to give

$$\underline{\tilde{X}}_t^{(\nu)} = \underline{\tilde{X}}_t^{(\nu)}\left(\underline{\tilde{Y}}_1^{(\nu)}, \ldots, \underline{\tilde{Y}}_{t-1}^{(\nu)}, \big(\underline{\tilde{W}}^{(\nu \to \mathcal{B})} : \mathcal{B} \in \mathcal{B}^{(\nu)}\big), \underline{\tilde{T}}^{(\nu)}\right).$$

Then each component $\underline{\tilde{X}}_t^{(e)}$, $e \in \mathcal{E}_{\text{out}}(\nu)$, of the network input $\underline{\tilde{X}}_t^{(\nu)}$ is encoded (if necessary) using the emulation code's encoder to give

$$\underline{\hat{X}}_t^{(e)} = \begin{cases} \big(\alpha_{tN}^{(c)}(\underline{\tilde{X}}_t^{(i)}, \alpha_{tN}^{(p)}(\underline{\tilde{X}}_t^{(i)})), \alpha_{tN}^{(p)}(\underline{\tilde{X}}_t^{(i)})\big) & \text{if } e = \bar{e} \\ \underline{\tilde{X}}_t^{(e)} & \text{otherwise} \end{cases}$$

thereby giving the inputs for all channels in network $\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p)$. If there is an adversarial set $E \in A$ such that $\bar{e} \in E$ then eavesdropper overhearing edge $\bar{e}$ in stacked network $\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p)$ is receiving bits $\tilde{B}_t^{(p)} = \alpha_{tN}^{(p)}(\underline{\tilde{X}}_t^{V_1(\bar{e})})$ and not $\underline{\tilde{Z}}_t^{(\bar{e})} = \beta_{tN}^{(p)}(\tilde{B}_t^{(p)})$. We will prove that by looking at the bits $(\tilde{B}^{(p)})^n$ instead of $(\underline{\tilde{Z}}^{(\bar{e})})^n$ the eavesdropper has no gain in terms of mutual information with the message. Indeed

$$\underline{W} \to \underline{\tilde{W}} \to \left((\underline{\tilde{X}}^{V_1(\bar{e})})^n, (\underline{\tilde{X}}^{V_1(E\setminus\{\bar{e}\})})^n\right) \to \left((\underline{\tilde{Z}}^{(\bar{e})})^n, (\underline{\tilde{Z}}^{(E\setminus\{\bar{e}\})})^n\right) \overset{(a)}{\to} \left((\underline{\tilde{B}}^{(p)})^n, (\underline{\tilde{Z}}^{(E\setminus\{\bar{e}\})})^n\right)$$

where (a) holds since when multiple bit sequences correspond to the $(\underline{\tilde{Z}}^{(\bar{e})})^n$ chosen then one of the bit sequences is chosen at random. Therefore

$$\mathbb{E}_{\mathcal{C}}\left[I_{\hat{p}}\left(\underline{W}; (\underline{\tilde{B}}^{(p)})^n, (\underline{\tilde{Z}}^{(E\setminus\{\bar{e}\})})^n\right)\right] \leq \mathbb{E}_{\mathcal{C}}\left[I_{\hat{p}}\left(\underline{\tilde{W}}; (\underline{\tilde{B}}^{(p)})^n, (\underline{\tilde{Z}}^{(E\setminus\{\bar{e}\})})^n\right)\right] \leq \mathbb{E}_{\mathcal{C}}\left[I_{\hat{p}}\left(\underline{\tilde{W}}; (\underline{\tilde{Z}}^{(\bar{e})})^n, \underline{\tilde{Z}}^{(E\setminus\{\bar{e}\})})^n\right)\right]$$

where subscript $\hat{p}$ is used to stress that the mutual informations are computed with respect to the probability distribution $\hat{p}$ induced on network $\mathcal{N}_{\bar{e}}(R_c, R_p)$ through solution $\mathcal{S}(\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p))$ described above, and consequently without loss of generality we will do our analysis as if the eavesdropper overhearing edge $\bar{e}$ receives $(\underline{\tilde{Z}}^{(\bar{e})})^n$ and not $(\tilde{B}^{(p)})^n$.

Define the indicator function J as

$$J = \begin{cases} 1, & \text{There exists } t \in \{1, \ldots, n\} \text{ such that } (\underline{\tilde{X}}_t, \underline{\tilde{Z}}_t) \notin \hat{A}_{\epsilon_1,t}^{(N)}(\tilde{X}, \tilde{Z}) \text{ or} \\ & (\underline{\tilde{X}}_t, \underline{\tilde{Y}}_t, \underline{\tilde{Z}}_t) \notin \hat{A}_{\epsilon_2,t}^{(N)}(\tilde{X}, \tilde{Y}, \tilde{Z}) \\ 0, & \text{otherwise} \end{cases} . \tag{3.2}$$

It is proved in Lemma 15 of [32] that $p_t\left[\left(\hat{A}_{\epsilon_1,t}^{(N)}(\tilde{X}, \tilde{Z})\right)^c\right] \leq 2^{-Nc_1(\epsilon_1,t)}$ and $p_t\left[\left(\hat{A}_{\epsilon_2,t}^{(N)}(\tilde{X}, \tilde{Y}, \tilde{Z})\right)^c\right] \leq 2^{-Nc_2(\epsilon_2,t)}$ for sufficiently large N where $c_1(\epsilon_1, t) > 0$ and $c_2(\epsilon_2, t) > 0$. Due to the union bound $\Pr(J = 1) \leq \sum_{t=1}^n p_t\left[\left(\hat{A}_{\epsilon_1,t}^{(N)}(\tilde{X}, \tilde{Z})\right)^c\right] + \sum_{t=1}^n p_t\left[\left(\hat{A}_{\epsilon_2,t}^{(N)}(\tilde{X}, \tilde{Y}, \tilde{Z})\right)^c\right]$ and since blocklength n is fixed one can choose N sufficiently large so that $\Pr(J = 1) \leq 2^{-Nc(\epsilon_1,\epsilon_2)}$ where $c(\epsilon_1, \epsilon_2) = \frac{1}{2}\min_t\{c_1(\epsilon_1, t), c_2(\epsilon_2, t)\}$. By changing the noisy channel of edge $\bar{e}$ of stacked network $\underline{\mathcal{N}}$ to the noiseless bit pipes of network $\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p)$ and applying the stacked solution $\underline{\mathcal{S}}(\underline{\mathcal{N}})$ along with the set of encoders/decoders $\{a_{tN}^{(0)}, a_{tN}^{(1)}, \beta_{tN}^{(1)}, \beta_{tN}^{(2)}\}_{t=1}^n$ we effectively change the distribution on $\underline{\tilde{Y}}_t^{V_2(\bar{e})}$ and therefore the distribution of $\underline{\tilde{X}}_t^{V_1(\bar{e})}$ for all channels e that are downstream of edge $\bar{e}$.

It was proved in Step 2 of the proof of Theorem 4 in [32] that provided that the three inequalities below hold

$$\begin{aligned} 2a_1(\epsilon_1, t) + \epsilon_1 &< R_p - I(\tilde{X}_t; \tilde{Z}_t) \; \forall t \in \{1, \ldots, n\} \\ 4a_2(\epsilon_2, t) &< R_c - (I(\tilde{X}_t; \tilde{Y}_t) - I(\tilde{X}_t; \tilde{Z}_t)) \; \forall t \in \{1, \ldots, n\} \\ 4a_1(\epsilon_1, t) + 3\epsilon_1 &< c_2(\epsilon_2, t) \; \forall t \in \{1, \ldots, n\} \end{aligned} \tag{3.3}$$

for all $t \in \{1,\ldots,n\}$ where $a_1(\epsilon_1,t)$ and $a_2(\epsilon_2,t)$ are defined in Appendix D, along with

$$\begin{aligned}\sum_{t'=1}^{t-1}\nu(t') &< \eta_t(\nu(t))/2 \quad \forall t \in \{1,\ldots,n\}\\ \sum_{t'=1}^{n}\nu(t') &< \delta/2\end{aligned} \tag{3.4}$$

where $\nu = 4a_1(\epsilon_1,t)+3\epsilon_1+8a_2(\epsilon_2,t)$, then $\hat{\mathbb{P}}\mathrm{r}\big((\tilde{\underline{x}}_t,\tilde{\underline{z}}_t) \notin \hat{A}^{(N)}_{\epsilon_1,t}(\tilde{X},\tilde{Z}) \bigcup (\tilde{\underline{x}}_t,\tilde{\underline{y}}_t,\tilde{\underline{z}}_t) \notin \hat{A}^{(N)}_{\epsilon_2,t}(\tilde{X},\tilde{Y},\tilde{Z})\big) \leq 2^{-N\hat{c}'(\epsilon_1,\epsilon_2,t)}$ for some $\hat{c}'(\epsilon_1,\epsilon_2,t) > 0$ where probability $\hat{\mathbb{P}}\mathrm{r}$ is computed with the new distribution induced in network $\underline{\mathcal{N}}_{\bar{e}}(R_c,R_p)$ with the use of random encoders/decoders $\big\{a^{(0)}_{tN}, a^{(1)}_{tN}, \beta^{(1)}_{tN}, \beta^{(2)}_{tN}\big\}_{t=1}^{n}$. Therefore by the union bound $\hat{\mathbb{P}}\mathrm{r}(J=1) \leq \sum_{t=1}^{n} 2^{-N\hat{c}'(\epsilon_1,\epsilon_2,t)} \leq 2^{-N\hat{c}(\epsilon_1,\epsilon_2)}$ for $\hat{c}(\epsilon_1,\epsilon_2) = \frac{1}{2}\min_t \hat{c}'(\epsilon_1,\epsilon_2,t)$ and sufficiently large N. For reasons that will become evident shortly we will use N large enough so that $\Pr(I=0) \geq \frac{1}{2}$ and $\hat{\mathbb{P}}\mathrm{r}(J=0) \geq \frac{1}{2}$. The definitions of $a_1(\epsilon_1,t)$ and $a_2(\epsilon_2,t)$ are given in Appendix D and they both tend to zero as $\epsilon_1(t)$ and $\epsilon_2(t)$ tend to zero.

To explore the security of code $\mathcal{S}(\underline{\mathcal{N}}_{\bar{e}}(R_c,R_p))$, we first investigate the probability that the emulated channel outputs $\tilde{\underline{Z}}^{(E)}_1,\ldots,\tilde{\underline{Z}}^{(E)}_n$ to an eavesdropper E at times $1,\ldots,n$ that are jointly typical with the message vector $\tilde{\underline{W}}$ under stacked solution $\underline{\mathcal{S}}(\underline{\mathcal{N}})$ on network $\underline{\mathcal{N}}$. Define typical set $A^{(N)}_{\epsilon',E}$ for each eavesdropper $E \in A$ as

$$\begin{aligned}A^{(N)}_{\epsilon',E} = \Bigg\{(\tilde{\underline{w}},\tilde{\underline{z}}^E_1,\ldots,\tilde{\underline{z}}^E_n) \in \underline{\tilde{\mathcal{W}}} \times \underline{\tilde{\mathcal{Z}}}^E \times \ldots \times \underline{\tilde{\mathcal{Z}}}^E :\ &\left|-\frac{1}{N}\log p(\tilde{\underline{w}}) - H(\tilde{W})\right| \leq \epsilon',\\ &\left|-\frac{1}{N}\log p(\tilde{\underline{z}}^E_1,\ldots,\tilde{\underline{z}}^E_n) - H(\tilde{Z}^E_1,\ldots,\tilde{Z}^E_n)\right| \leq \epsilon',\\ &\left|-\frac{1}{N}\log p(\tilde{\underline{z}}^E_1,\ldots,\tilde{\underline{z}}^E_n,\tilde{\underline{w}}) - H(\tilde{Z}^E_1,\ldots,\tilde{Z}^E_n,\tilde{W})\right| \leq \epsilon'\Bigg\}.\end{aligned} \tag{3.5}$$

For each $E \in A$

$$\Pr\left((A^{(N)}_{\epsilon',E})^c\right) \leq 2^{-Nf(\epsilon',E)} \overset{(a)}{\leq} 2^{-Nf(\epsilon')} \tag{3.6}$$

for some $f(\epsilon',E) > 0$ by Lemma 8 in [31], which follows from the Chernoff bound. Inequality (a) holds by setting $f(\epsilon') = \min_{E\in A} f(\epsilon',E)$.

Let $I_{\hat{p}}\big(\tilde{\underline{W}};(\tilde{\underline{Z}}^{(E)})^n\big)$ be the mutual information between message $\tilde{\underline{W}}$ and eavesdropped output $(\tilde{\underline{Z}}^{(E)})^n$ with respect to the probability distribution $\hat{p}$ induced at the solution $\mathcal{S}(\underline{\mathcal{N}}_{\bar{e}}(R_c,R_p))$ for stacked network $\underline{\mathcal{N}}_{\bar{e}}(R_c,R_p)$. Then,

$$\begin{aligned}&\mathbb{E}_{\mathcal{C}}\left[I_{\hat{p}}\big(\tilde{\underline{W}};(\tilde{\underline{Z}}^{(E)})^n\big)\right]\\ &\leq \mathbb{E}_{\mathcal{C}}\left[I_{\hat{p}}\big(\tilde{\underline{W}};(\tilde{\underline{Z}}^{(E)})^n,J\big)\right]\\ &= \mathbb{E}_{\mathcal{C}}\left[I_{\hat{p}}\big(\tilde{\underline{W}};J\big)\right] + \mathbb{E}_{\mathcal{C}}\left[I_{\hat{p}}\big(\tilde{\underline{W}};(\tilde{\underline{Z}}^{(E)})^n|J\big)\right]\end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{\leq} 1+\mathbb{E}_{\mathcal{C}}\left[I_{\hat{p}}\big(\tilde{\underline{W}};(\tilde{\underline{Z}}^{(E)})^n|J\big)\right]\\
&=1+\hat{\mathbb{P}}\mathrm{r}(J=1)\cdot\mathbb{E}_{\mathcal{C}}\left[I_{\hat{p}}\big(\tilde{\underline{W}};(\tilde{\underline{Z}}^{(E)})^n|J=1\big)\right]+\hat{\mathbb{P}}\mathrm{r}(J=0)\cdot\mathbb{E}_{\mathcal{C}}\left[I_{\hat{p}}\big(\tilde{\underline{W}};(\tilde{\underline{Z}}^{(E)})^n|J=0\big)\right]\\
&\overset{(b)}{\leq} 1+2^{-N\hat{c}(\epsilon_1,\epsilon_2)}nN\sum_{i\in\mathcal{V}}\sum_{\mathcal{B}\in\mathcal{B}^{(i)}}\tilde{R}^{(i\to\mathcal{B})}+\mathbb{E}_{\mathcal{C}}\left[I_{\hat{p}}\big(\underline{W};(\tilde{\underline{Z}}^{(E)})^n|J=0\big)\right]\\
&\overset{(c)}{\leq} 2+\mathbb{E}_{\mathcal{C}}\left[I_{\hat{p}}\big(\tilde{\underline{W}};(\tilde{\underline{Z}}^{(E)})^n|J=0\big)\right] \qquad (3.7)
\end{aligned}
$$

where (a) follows since $\mathbb{E}_{\mathcal{C}}\left[I_{\hat{p}}(\tilde{\underline{W}};J)\right]\leq\mathbb{E}_{\mathcal{C}}[H(J)]\leq 1$ since J is a binary variable, (b) follows since $\hat{\mathbb{P}}\mathrm{r}(J=1)\leq 2^{-N\hat{c}(\epsilon_1,\epsilon_2)}$, $\mathbb{E}_{\mathcal{C}}\left[I_{\hat{p}}\big(\tilde{\underline{W}};(\tilde{\underline{Z}}^{(E)})^n|J=1\big)\right]\leq H\big(\tilde{\underline{W}}|J=1\big)\leq nN\sum_{i\in\mathcal{V}}\sum_{\mathcal{B}\in\mathcal{B}^{(i)}}\tilde{R}^{(i\to\mathcal{B})}\leq nNC^{(i)}$, and $\Pr(J=0)\leq 1$, and (c) holds for N sufficiently large.

To bound $\mathbb{E}_{\mathcal{C}}\left[I_{\hat{p}}\big(\tilde{\underline{W}};(\tilde{\underline{Z}}^{(E)})^n|J=0\big)\right]$, note that

$$
\begin{aligned}
\mathbb{E}_{\mathcal{C}}\left[I_{\hat{p}}\big(\tilde{\underline{W}};(\tilde{\underline{Z}}^{E})^n|J=0\big)\right]&=\sum_{(\tilde{\underline{w}},(\tilde{\underline{z}}^E)^n)\in(A^{(N)}_{\epsilon',E})^c}\hat{p}(\tilde{\underline{w}},(\tilde{\underline{z}}^E)^n|J=0)\log\frac{\hat{p}(\tilde{\underline{w}},(\tilde{\underline{z}}^E)^n|J=0)}{\hat{p}(\tilde{\underline{w}}|J=0)\hat{p}((\tilde{\underline{z}}^E)^n|J=0)}\\
&\quad+\sum_{(\tilde{\underline{w}},(\tilde{\underline{z}}^E)^n)\in A^{(N)}_{\epsilon',E}}\hat{p}(\tilde{\underline{w}},(\tilde{\underline{z}}^E)^n|J=0)\log\frac{\hat{p}(\tilde{\underline{w}},(\tilde{\underline{z}}^E)^n|J=0)}{\hat{p}(\tilde{\underline{w}}|J=0)\hat{p}((\tilde{\underline{z}}^E)^n|J=0)}. \qquad (3.8)
\end{aligned}
$$

To bound the first term of (3.8), note that for N sufficiently large

$$
\begin{aligned}
&\left\{\sum_{(\tilde{\underline{w}},(\tilde{\underline{z}}^E)^n)\in(A^{(N)}_{\epsilon',E})^c}\hat{p}(\tilde{\underline{w}},(\tilde{\underline{z}}^E)^n|J=0)\log\frac{\hat{p}(\tilde{\underline{w}},(\tilde{\underline{z}}^E)^n|J=0)}{\hat{p}(\tilde{\underline{w}})\hat{p}((\tilde{\underline{z}}^E)^n|J=0)}\right\}\\
&\leq\sum_{(\tilde{\underline{w}},(\tilde{\underline{z}}^E)^n)\in(A^{(N)}_{\epsilon',E})^c}\hat{p}(\tilde{\underline{w}},(\tilde{\underline{z}}^E)^n|J=0)\log\frac{\hat{p}((\tilde{\underline{z}}^E)^n|J=0)}{\hat{p}(\tilde{\underline{w}})\hat{p}((\tilde{\underline{z}}^E)^n|J=0)}\\
&\overset{(a)}{=}\sum_{(\tilde{\underline{w}},(\tilde{\underline{z}}^E)^n)\in(A^{(N)}_{\epsilon',E})^c}\hat{p}(\tilde{\underline{w}},(\tilde{\underline{z}}^E)^n|J=0)\log\frac{1}{1/2^{nN\sum_{i\in\mathcal{V}}\sum_{\mathcal{B}\in\mathcal{B}^{(i)}}|\tilde{W}^{(i\to\mathcal{B})}|}}\\
&=\Big(nN\sum_{i\in\mathcal{V}}\sum_{\mathcal{B}\in\mathcal{B}^{(i)}}|\tilde{W}^{(i\to\mathcal{B})}|\Big)\sum_{(\tilde{\underline{w}},(\tilde{\underline{z}}^E)^n)\in(A^{(N)}_{\epsilon',E})^c}\hat{p}(\tilde{\underline{w}},(\tilde{\underline{z}}^E)^n|J=0)\\
&\leq\Big(nN\sum_{i\in\mathcal{V}}\sum_{\mathcal{B}\in\mathcal{B}^{(i)}}|\tilde{W}^{(i\to\mathcal{B})}|\Big)\sum_{(\tilde{\underline{w}},(\tilde{\underline{z}}^E)^n)\in(A^{(N)}_{\epsilon',E})^c}\frac{\hat{p}(\tilde{\underline{w}},(\tilde{\underline{z}}^E)^n,J=0)}{\hat{\Pr}(J=0)}\\
&\overset{(b)}{\leq}2\Big(nN\sum_{i\in\mathcal{V}}\sum_{\mathcal{B}\in\mathcal{B}^{(i)}}|\tilde{W}^{(i\to\mathcal{B})}|\Big)\sum_{(\tilde{\underline{w}},(\tilde{\underline{z}}^E)^n)\in(A^{(N)}_{\epsilon',E})^c}\hat{p}(\tilde{\underline{w}},(\tilde{\underline{z}}^E)^n,J=0)\\
&\overset{(c)}{\leq}\Big(nN\sum_{i\in\mathcal{V}}\sum_{\mathcal{B}\in\mathcal{B}^{(i)}}|\tilde{W}^{(i\to\mathcal{B})}|\Big)2^{N\sum_{t=1}^{n}b(\epsilon_1,\epsilon_2,t)}\sum_{(\tilde{\underline{w}},(\tilde{\underline{z}}^E)^n)\in(A^{(N)}_{\epsilon',E})^c}p(\tilde{\underline{w}},(\tilde{\underline{z}}^E)^n,J=0)\\
&\leq\Big(nN\sum_{i\in\mathcal{V}}\sum_{\mathcal{B}\in\mathcal{B}^{(i)}}|\tilde{W}^{(i\to\mathcal{B})}|\Big)2^{N\sum_{t=1}^{n}b(\epsilon_1,\epsilon_2,t)}\sum_{(\tilde{\underline{w}},(\tilde{\underline{z}}^E)^n)\in(A^{(N)}_{\epsilon',E})^c}p(\tilde{\underline{w}},(\tilde{\underline{z}}^E)^n)
\end{aligned}
$$

$$\overset{(d)}{\leq} \Big(nN \sum_{i\in\mathcal{V}} \sum_{\mathcal{B}\in\mathcal{B}^{(i)}} |\tilde{W}^{(i\to\mathcal{B})}| \Big) 2^{N\sum_{t=1}^{n} b(\epsilon_1,\epsilon_2,t)} 2^{-Nf(\epsilon')}. \tag{3.9}$$

where (a) holds since all messages are equiprobable and therefore $\hat{p}(\underline{\tilde{w}}) = 2^{-nN\sum_{i\in\mathcal{V}}\sum_{\mathcal{B}\in\mathcal{B}^{(i)}}|\tilde{W}^{(i\to\mathcal{B})}|}$, (b) holds since $\hat{\mathbb{P}}\mathrm{r}(J=0) \leq \frac{1}{2}$ for N sufficiently large, (c) replaces $\hat{p}$ by p using Lemma 8 proved in Appendix E where $b(t)$ is defined in Appendix E to be $b(\epsilon_1,\epsilon_2,t) = 4a_1(\epsilon_1,t) + 8a_2(\epsilon_2,t) + 2\epsilon_1(t) + 2/N$, and (d) follows from inequality (3.6). In order to upper bound the term in equation (3.9) we need to choose parameters $\epsilon_1(1),\ldots,\epsilon_1(n)$ and $\epsilon_2(1),\ldots,\epsilon_2(n)$ so that the exponent of $2^{-N\left(f(\epsilon')-\sum_{t=1}^{n} b(\epsilon_1,\epsilon_2,t)\right)}$ is negative. We first choose parameter ϵ' of the typical set defined in equation (3.5) to be equal to parameter ε used in the bound $I\big((\underline{\tilde{Z}}^E)^n;\underline{W}\big) \leq n\varepsilon$ on the rate which mutual information is revealed to the eavesdropper. We then choose parameters $\epsilon_1(n)$ and $\epsilon_2(n)$ so that

$$\nu(n) < \min\left\{\frac{\delta}{4n}, \frac{f(\varepsilon)}{4n}, \frac{\varepsilon}{4n}\right\} \tag{3.10}$$

and all the subsequent $\epsilon_1(t)$ and $\epsilon_2(t)$ for $t \in \{n-1,\ldots,1\}$ such that

$$\nu(t) \;<\; \min\left\{\frac{\delta}{4n}, \frac{f(\varepsilon)}{4n}, \frac{\varepsilon}{4n}, \min_{t'>t}\left[\frac{\eta_{t'}(\nu(t'))}{4t'}\right]\right\} \;\forall t \in \{n-1,\ldots,1\} \tag{3.11}$$

and this guarantees that inequalities (3.3) and (3.4) are satisfied. Parameter $b(\epsilon_1,\epsilon_2,t)$ can be written as $b(\epsilon_1,\epsilon_2,t) = \nu(t) + \frac{2}{N} - \epsilon_1(t)$ and therefore once all $\epsilon_1(1),\ldots,\epsilon_1(n)$ have been chosen to satisfy equations (3.10) and (3.11) we use a sufficiently large N such that $\frac{2}{N} < \min_t \epsilon_1(t)$, giving $b(\epsilon_1,\epsilon_2,t) < \nu(t)$ for all $t \in \{1,\ldots,n\}$ and therefore

$$\sum_{t=1}^{n} b(\epsilon_1,\epsilon_2,t) < \sum_{t=1}^{n} \nu(t) \overset{(a)}{<} \frac{1}{4}\min\big\{\delta, f(\varepsilon), \varepsilon\big\}; \tag{3.12}$$

here inequality (a) follows from (3.10) and (3.11). Consequently, combining the inequality above and (3.9) we get

$$\sum_{(\underline{\tilde{w}},(\underline{\tilde{z}}^E)^n)\in(A^{(N)}_{\epsilon',E})^c} \hat{p}(\underline{\tilde{w}},(\underline{\tilde{z}}^E)^n|J=0)\log\frac{\hat{p}(\underline{\tilde{w}},(\underline{\tilde{z}}^E)^n|J=0)}{\hat{p}(\underline{\tilde{w}})\hat{p}((\underline{\tilde{z}}^E)^n|J=0)} \leq \Big(nN\sum_{i\in\mathcal{V}}\sum_{\mathcal{B}\in\mathcal{B}^{(i)}}|\tilde{W}^{(i\to\mathcal{B})}|\Big)2^{-\frac{3}{4}Nf(\delta)} \leq 1$$

for sufficiently large N.

To bound the second term of (3.8), note that

$$p\big((\underline{\tilde{z}}^E)^n, J=0\big) = p\big((\underline{\tilde{z}}^E)^n\big)\Pr\big(J=0|(\underline{\tilde{z}}^E)^n\big).$$

To bound this probability, define set

$$G^{(N)} = \left\{ (\tilde{\underline{z}}^E)^n : \Pr\left(J=1 \,|\, (\tilde{\underline{z}}^E)^n\right) < \frac{1}{2} \right\} \tag{3.13}$$

and therefore

$$p\big((\tilde{\underline{z}}^E)^n, J=0\big) \geq \frac{1}{2} p\big((\tilde{\underline{z}}^E)^n\big),\ \forall (\tilde{\underline{z}}^E)^n \in G^{(N)}. \tag{3.14}$$

The probability of observing a vector $(\tilde{\underline{z}}^E)^n$ outside of set $G^{(N)}$ is exponentially small

$$\sum_{(\tilde{\underline{z}}^E)^n \in \left(G^{(N)}\right)^c} p((\tilde{\underline{z}}^E)^n) < 2 \cdot \Pr(J=1) < 2 \cdot 2^{-Nc(\epsilon_1,\epsilon_2)}. \tag{3.15}$$

Thus we bound the second term of (3.8) as

$$\begin{aligned}
&\left\{ \sum_{(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n) \in A^{(N)}_{\epsilon',E}} \hat{p}(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n|J=0) \log \frac{\hat{p}(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n|J=0)}{\hat{p}(\underline{\tilde{w}})\hat{p}((\tilde{\underline{z}}^E)^n|J=0)} \right\} \\
&= \sum_{(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n) \in A^{(N)}_{\epsilon',E}} \hat{p}(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n|J=0) \log \frac{\hat{p}(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n, J=0)}{\hat{p}(\underline{\tilde{w}})\hat{p}((\tilde{\underline{z}}^E)^n, J=0)} \\
&\overset{(a)}{\leq} \sum_{(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n) \in A^{(N)}_{\epsilon',E}} \hat{p}(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n|J=0) \log \frac{p(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n, J=0) 2^{2\sum_{t=1}^n b(\epsilon_1,\epsilon_2,t)}}{p(\underline{\tilde{w}})p((\tilde{\underline{z}}^E)^n, J=0)} \\
&= \sum_{(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n) \in A^{(N)}_{\epsilon',E}} \hat{p}(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n|J=0) \log \frac{p(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n, J=0)}{p(\underline{\tilde{w}})p((\tilde{\underline{z}}^E)^n, J=0)} \\
&\quad + \sum_{(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n) \in A^{(N)}_{\epsilon',E}} \hat{p}(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n|J=0) \left(2\sum_{t=1}^n b(\epsilon_1,\epsilon_2,t) \right) \\
&\leq \sum_{(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n) \in A^{(N)}_{\epsilon',E} \bigwedge (\tilde{\underline{z}}^E)^n \in G^{(N)}} \hat{p}(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n|J=0) \log \frac{p(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n, J=0)}{p(\underline{\tilde{w}})p((\tilde{\underline{z}}^E)^n, J=0)} \\
&\quad + \sum_{(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n) \in A^{(N)}_{\epsilon',E} \bigwedge (\tilde{\underline{z}}^E)^n \in \left(G^{(N)}\right)^c} \hat{p}(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n|J=0) \log \frac{p(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n, J=0)}{p(\underline{\tilde{w}})p((\tilde{\underline{z}}^E)^n, J=0)} \\
&\quad + 2\sum_{t=1}^n b(\epsilon_1,\epsilon_2,t) \\
&\overset{(b)}{\leq} \sum_{(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n) \in A^{(N)}_{\epsilon',E} \bigwedge (\tilde{\underline{z}}^E)^n \in G^{(N)}} \hat{p}(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n|J=0) \log \frac{p(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n, J=0)\, 2}{p(\underline{\tilde{w}})p((\tilde{\underline{z}}^E)^n)} \\
&\quad + \sum_{(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n) \in A^{(N)}_{\epsilon',E} \bigwedge (\tilde{\underline{z}}^E)^n \in \left(G^{(N)}\right)^c} \hat{p}(\underline{\tilde{w}},(\tilde{\underline{z}}^E)^n|J=0) \log \frac{p((\tilde{\underline{z}}^E)^n, J=0)}{p(\underline{\tilde{w}})p((\tilde{\underline{z}}^E)^n, J=0)} \\
&\quad + \varepsilon
\end{aligned} \tag{3.16}$$

$$
\begin{aligned}
&\leq \sum_{(\underline{\tilde{w}},(\underline{\tilde{z}}^E)^n)\in A^{(N)}_{\epsilon',E}\bigwedge(\underline{\tilde{z}}^E)^n\in G^{(N)}} \hat{p}(\underline{\tilde{w}},(\underline{\tilde{z}}^E)^n|J=0)\log\frac{p(\underline{\tilde{w}},(\underline{\tilde{z}}^E)^n)\,2}{p(\underline{\tilde{w}})p((\underline{\tilde{z}}^E)^n)}\\
&+\sum_{(\underline{\tilde{w}},(\underline{\tilde{z}}^E)^n)\in A^{(N)}_{\epsilon',E}\bigwedge(\underline{\tilde{z}}^E)^n\in\left(G^{(N)}\right)^c} \hat{p}(\underline{\tilde{w}},(\underline{\tilde{z}}^E)^n|J=0)\log\frac{1}{1/2^{nN\sum_{i\in\mathcal{V}}\sum_{\mathcal{B}\in\mathcal{B}^{(i)}}|\tilde{W}^{(i\to\mathcal{B})}|}}\\
&+\varepsilon\\
&=\sum_{(\underline{\tilde{w}},(\underline{\tilde{z}}^E)^n)\in A^{(N)}_{\epsilon',E}\bigwedge(\underline{\tilde{z}}^E)^n\in G^{(N)}} \hat{p}(\underline{\tilde{w}},(\underline{\tilde{z}}^E)^n|J=0)\log\frac{p(\underline{\tilde{w}},(\underline{\tilde{z}}^E)^n)}{p(\underline{\tilde{w}})p((\underline{\tilde{z}}^E)^n)}\\
&+\sum_{(\underline{\tilde{w}},(\underline{\tilde{z}}^E)^n)\in A^{(N)}_{\epsilon',E}\bigwedge(\underline{\tilde{z}}^E)^n\in G^{(N)}} \hat{p}(\underline{\tilde{w}},(\underline{\tilde{z}}^E)^n|J=0)\log 2\\
&+\Big(nN\sum_{i\in\mathcal{V}}\sum_{\mathcal{B}\in\mathcal{B}^{(i)}}|\tilde{W}^{(i\to\mathcal{B})}|\Big)\sum_{(\underline{\tilde{w}},(\underline{\tilde{z}}^E)^n)\in A^{(N)}_{\epsilon',E}\bigwedge(\underline{\tilde{z}}^E)^n\in\left(G^{(N)}\right)^c} p(\underline{\tilde{w}},(\underline{\tilde{z}}^E)^n|J=0)\\
&+\varepsilon\\
&\overset{(c)}{\leq}\sum_{(\underline{\tilde{w}},(\underline{\tilde{z}}^E)^n)\in A^{(N)}_{\epsilon',E}\bigwedge(\underline{\tilde{z}}^E)^n\in G^{(N)}} \hat{p}(\underline{\tilde{w}},(\underline{\tilde{z}}^E)^n|J=0)\log\frac{2^{-N\left(H(\tilde{W},\tilde{Z}^E)-\epsilon'\right)}}{2^{-N\left(H(\tilde{W})+H(\tilde{Z}^E)+2\epsilon'\right)}}\\
&+\Big(nN\sum_{i\in\mathcal{V}}\sum_{\mathcal{B}\in\mathcal{B}^{(i)}}|\tilde{W}^{(i\to\mathcal{B})}|\Big)\sum_{(\underline{\tilde{w}},(\underline{\tilde{z}}^E)^n)\in A^{(N)}_{\epsilon',E}\bigwedge(\underline{\tilde{z}}^E)^n\in\left(G^{(N)}\right)^c}\frac{p(\underline{\tilde{w}},(\underline{\tilde{z}}^E)^n,J=0)}{\Pr(J=0)}\\
&+1+\varepsilon\\
&\overset{(d)}{\leq}\sum_{(\underline{\tilde{w}},(\underline{\tilde{z}}^E)^n)\in A^{(N)}_{\epsilon',E}\bigwedge(\underline{\tilde{z}}^E)^n\in G^{(N)}} \hat{p}(\underline{\tilde{w}},(\underline{\tilde{z}}^E)^n|J=0)\log 2^{N\left(I(\tilde{W};\tilde{Z}^E)+3\epsilon'\right)}\\
&+2\Big(nN\sum_{i\in\mathcal{V}}\sum_{\mathcal{B}\in\mathcal{B}^{(i)}}|\tilde{W}^{(i\to\mathcal{B})}|\Big)\sum_{(\underline{\tilde{w}},(\underline{\tilde{z}}^E)^n)\in A^{(N)}_{\epsilon',E}\bigwedge(\underline{\tilde{z}}^E)^n\in\left(G^{(N)}\right)^c} p(\underline{\tilde{w}},(\underline{\tilde{z}}^E)^n)\\
&+1+\varepsilon\\
&\leq N\left(I(\tilde{W};\tilde{Z}^E)+3\epsilon'\right)+1+\varepsilon\\
&+2\Big(nN\sum_{i\in\mathcal{V}}\sum_{\mathcal{B}\in\mathcal{B}^{(i)}}|\tilde{W}^{(i\to\mathcal{B})}|\Big)\sum_{(\underline{\tilde{z}}^E)^n\in\left(G^{(N)}\right)^c} p((\underline{\tilde{z}}^E)^n)\\
&\overset{(e)}{\leq} N\left(I(\tilde{W};\tilde{Z}^E)+3\epsilon'\right)+1+\varepsilon\\
&+4\Big(nN\sum_{i\in\mathcal{V}}\sum_{\mathcal{B}\in\mathcal{B}^{(i)}}|\tilde{W}^{(i\to\mathcal{B})}|\Big)2^{-Nc(\epsilon_1,\epsilon_2)}\\
&\overset{(f)}{\leq} N\left(I(\tilde{W};\tilde{Z}^E)+3\varepsilon\right)+2+\varepsilon. \qquad (3.17)
\end{aligned}
$$

Here (a) follows from the fact that $p(\underline{\tilde{w}})=\hat{p}(\underline{\tilde{w}})$ and the bounds proved in Lemma 8 of Appendix E; (b) follows from inequalities (3.12) and (3.14), (c) follows from the definition of the typical set in (3.5); (d) holds since we choose N large enough so that $\Pr(J=0)\geq\frac{1}{2}$; (e) follows from (3.15) and

finally inequality (f) holds since $\epsilon' = \varepsilon$ and N is chosen large enough so that the last term is less than 1. The solution $(\lambda, \varepsilon, A, \tilde{R})$–$\mathcal{S}(\mathcal{N})$ for network $\mathcal{N}$ is secure, *i.e.* $I(\tilde{W}; \tilde{Z}^E) < n\varepsilon$ and therefore by combining inequalities (3.8) and (3.17) we get

$$\mathbb{E}_{\mathcal{C}}\left[I_{\hat{p}}\big(\underline{\tilde{W}}; (\underline{\tilde{Z}}^E)^n | J = 0\big)\right] \leq N(n\varepsilon + 3\varepsilon) + 3 + \varepsilon$$

and therefore by inequality (3.7)

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\left[I_{\hat{p}}\big(\underline{\tilde{W}}; (\underline{\tilde{Z}}^E)^n\big)\right] &\leq N(n\varepsilon + 3\varepsilon) + 5 + \varepsilon \\ &\leq nN(\varepsilon + \frac{3\varepsilon}{n} + \frac{5+\varepsilon}{nN}) \leq 5nN\varepsilon\end{aligned}$$

for sufficiently large N.

Therefore we have constructed a random code $(\lambda, 5\varepsilon, A, R)$–$\mathcal{S}(\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p))$ for network $\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p)$. To conclude the proof one should prove that there is at least one code instance where both the probability of error and the mutual information between the message and the eavesdropper is not large. One can follow an analysis identical to the one used in equation (3.1) of Theorem 8 to prove that there is indeed at least one deterministic code solution $(3\lambda, 15\varepsilon, A, R)$–$\mathcal{S}(\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p))$ for network $\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p)$. □

3.4.2 Proof of Theorem 10

In order to prove Theorem 10 we first prove Lemma 4 that provides a lower bounding network for the case where one replaces a noisy degraded wiretap channel $\mathcal{C}_{\bar{e}}$ with noiseless channel $\mathcal{C}_{\bar{e}}(R_c, R_p)$ for the case where channel $\bar{e}$ is secure ($\bar{e} \notin E$ for all $E \in A$) or when channel $\bar{e}$ is not simultaneously eavesdropped with any other link in the network ($|E| = 1$ if $e \in E$). We then prove in Lemma 5 a continuity result on the rate region $\mathcal{R}(\mathcal{N}_{\bar{e}}(R_c, R_p), A)$ with respect to (R_c, R_p) when $R_c > 0$ and $R_p > 0$. The lower bounding network of Lemma 4 along with an application of the continuity result of Lemma 5 will lead to the proof of Theorem 10.

Lemma 4. *Consider a network $\mathcal{N}$, an adversarial set $A \subseteq \mathcal{P}(\mathcal{E})$, and a single link $\bar{e} \in \mathcal{E}$. Let*

$$\begin{aligned}R_c &< \max_{p(x)} I(X^{(\bar{e})}; Y^{(\bar{e})}) - \max_{p(x)} I(X^{(\bar{e})}; Z^{(\bar{e})}) \\ R_p &< \max_{p(x)} I(X^{(\bar{e})}; Z^{(\bar{e})}).\end{aligned}$$

If $\bar{e}$ is invulnerable to wiretapping ($\bar{e} \notin E$ for all $E \in A$) or is not simultaneously wiretapped with other links ($|E| = 1$ if $\bar{e} \in E$), then $\mathcal{R}(\mathcal{N}_{\bar{e}}(R_c, R_p), A) \subseteq \mathcal{R}(\mathcal{N}, A)$.

Proof. We will prove Lemma 4 for a network $\mathcal{N}_{\bar{e}}(R_c, R_p)$ where channel $\bar{e}$ is eavesdropped but not

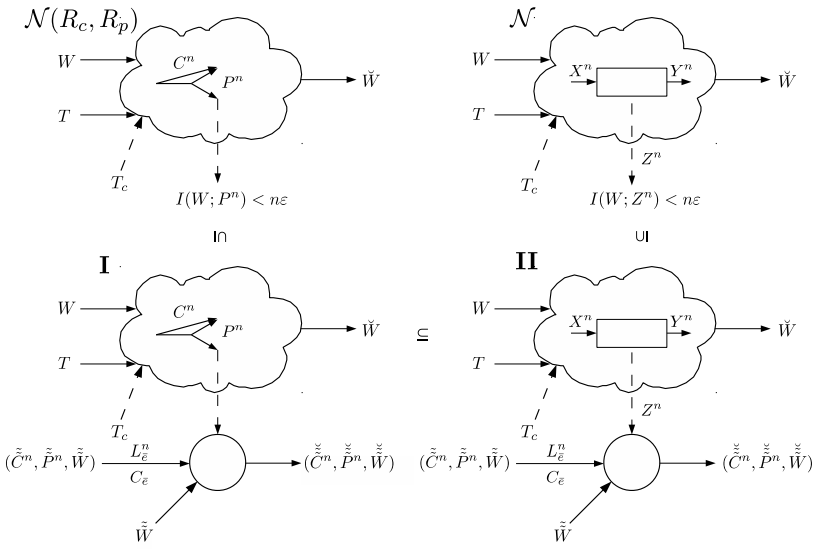


Figure 3.4: Network $\mathcal{N}_{\bar{e}}(R_c, R_p)$ along with networks **I**, **II** and $\mathcal{N}$ that assist proving Lemma 4. In the proof of Lemma 4 network **I** is operated in a stack of N_1 layers where network **II** is operated in a double stack of $N_2 N_1$ layers.

simultaneously with any other channel, that is $\exists \bar{E} \in A$ such that $\bar{E} = \{\bar{e}\}$ and $\bar{e} \notin E$ for all $E \in A \backslash \bar{E}$. The proof of Lemma 4 when channel $\bar{e}$ is invulnerable to wiretapping ($\bar{e} \notin E$ for all $E \in A$) is a simple version of the proof below and we will outline it.

Denote by C_t and P_t the rate R_c and rate R_p transmissions across the confidential and public links, respectively, of edge $\bar{e} \in \mathcal{E}$ at time t. Let $C^n = (C_1, \ldots, C_n)$, $P^n = (P_1, \ldots, P_n)$ and denote by C_j^i and P_j^i for any $j < i$ the vectors $C_j^i = (C_j, C_{j+1}, \ldots, C_i)$ and $P_j^i = (P_j, P_{j+1}, \ldots, P_i)$. The secrecy is achieved by having some independent source of randomness T (secret keys) injected at one or more locations within the network. Moreover there is the randomness corresponding to all the noisy channels in the network and collectively denoted by T_c where both T and T_c are depicted in Figure 3.4. The rest of the proof shows how to achieve any point R inside the secrecy rate region of network $\mathcal{N}_{\bar{e}}(R_c, R_p)$ in network $\mathcal{N}$. In particular we will take a secure code of rate R for network $\mathcal{N}_{\bar{e}}(R_c, R_p)$ and construct a secure code for network $\mathcal{N}$. To assist in the proof above we will make use of networks **I** and **II** shown in Figure 3.4 that are identical to networks $\mathcal{N}_{\bar{e}}(R_c, R_p)$ and $\mathcal{N}$ respectively with the addition of a noiseless side channel of capacity $C_{\bar{e}}$ from a "super-source" that has access to (W, C^n, P^n) to the eavesdropper wiretapping channel $\bar{e}$ where the exact value of capacity $C_{\bar{e}}$ will be specified in equation (3.22). In network **I** (**II**) the eavesdropper is required to decode the message W from all the sources, the public bits P^n and the confidential bits C^n with the use of the side channel, the eavesdropper's information P^n (Z^n) and the message W.

The outline of the proof is that one applies a secure code of blocklength n for network $\mathcal{N}_{\bar{e}}(R_c, R_p)$ to every layer of a stacked version of network **I**. The number of layers in the stacked version of network **I** is denoted by N_1. The reason for creating a stacked version of network **I** is so that we can use the

law of large number and typical sequences in order to prove that the eavesdropper can decode the message W along with the confidential C^n and public bits P^n by having access to W, P^n and the side channel of capacity $C_{\bar{e}}$ from the "super source". The constructed coded for the stacked version of network **I** can be seen as a code of blocklength $n_1 = nN_1$ for the non-stacked version of network **I**. To move the proof from network **I** to network **II** we use a stacked version of network **II** with N_2 number of layers where $N_2 \neq N_1$ in general. The code used at each layer of the stacked version of network **II** is the code of blocklength n_1 constructed above. We need to use a stacked version of network **II** to use a channel code at edge $\bar{e}$ of network **II** to emulate the noiseless edge $\bar{e}$ of network **I**. Below follow the details of the proof described above.

Choose any $\lambda > 0$, $\varepsilon > 0$ and $R \in \text{int}\big(\mathcal{R}(\mathcal{N}_{\bar{e}}, A)\big)$ in the relative interior of rate region $\mathcal{R}(\mathcal{N}_{\bar{e}}, A)$. We will show how to construct a $(\lambda, 12\varepsilon, A, R)$–$\mathcal{S}(\mathcal{N})$ for network $\mathcal{N}$.

$\mathcal{N}_{\bar{\mathbf{e}}}(\mathbf{R_c}, \mathbf{R_p})$ to I: Network **I** is identical to network $\mathcal{N}_{\bar{e}}(R_c, R_p)$ with the addition of a noiseless bit pipe of capacity $C_{\bar{e}}$ to the eavesdropper of channel $\bar{e}$ as well as an infinite capacity link from the source messages W to the eavesdropper effectively making W available to the eavesdropper of channel $\bar{e}$. In network **I** all receivers are required to decode their messages with small probability of error and moreover the eavesdropper of channel $\bar{e}$ is required to decode the confidential and public bits along with all the source message W by having access to the side channel of capacity $C_{\bar{e}}$, the public bits and all the source messages W. Assume that we take any rate tuple R in the relative interior of the A–secure rate region of network $\mathcal{N}_{\bar{e}}(R_c, R_p)$, *i.e.* $R \in \text{int}\left(\mathcal{R}(\mathcal{N}_{\bar{e}}(R_c, R_p), A)\right)$ and we will show how to construct a code of the same rate for network $\mathcal{N}$. For reasons that will become evident later we will choose two rates $\tilde{R} \in \text{int}\left(\mathcal{R}(\mathcal{N}_{\bar{e}}(R_c, R_p), A)\right)$ and $\tilde{\tilde{R}} \in \text{int}\left(\mathcal{R}(\mathcal{N}_{\bar{e}}(R_c, R_p), A)\right)$ such that $R^{(i\to\mathcal{B})} < \tilde{R}^{(i\to\mathcal{B})} < \tilde{\tilde{R}}^{(i\to\mathcal{B})}$ for all $i \in \mathcal{V}$ and $\mathcal{B} \in \mathcal{B}^{(i)}$ with $R^{(i\to\mathcal{B})} > 0$. As in the proof of Theorem 8 set $\tilde{\tilde{\rho}} = \min\limits_{R^{(i\to\mathcal{B})}>0} (\tilde{\tilde{R}}^{(i\to\mathcal{B})} - \tilde{R}^{(i\to\mathcal{B})})$ and $\tilde{\rho} = \min\limits_{R^{(i\to\mathcal{B})}>0} (\tilde{R}^{(i\to\mathcal{B})} - R^{(i\to\mathcal{B})})$ and find constants $\tilde{\tilde{\lambda}}$ and $\tilde{\lambda}$ satisfying

$$\max_{(i,\mathcal{B}):R^{(i\to\mathcal{B})}>0} \tilde{\tilde{R}}^{(i\to\mathcal{B})}\tilde{\tilde{\lambda}} + h(\tilde{\tilde{\lambda}}) < \tilde{\tilde{\rho}} \tag{3.18}$$

$$\max_{(i,\mathcal{B}):R^{(i\to\mathcal{B})}>0} \tilde{R}^{(i\to\mathcal{B})}\tilde{\lambda} + h(\tilde{\lambda}) < \tilde{\rho} \tag{3.19}$$

Then there exists a blocklength n secrecy code $(\tilde{\lambda}, \varepsilon, A, \tilde{\tilde{R}})$–$\mathcal{S}(\mathcal{N}_{\bar{e}}(R_c, R_p), A)$ of rate $\tilde{\tilde{R}}$ for network $\mathcal{N}_{\bar{e}}(R_c, R_p)$ such that $\Pr\left(\tilde{\tilde{W}}^{(j\to\mathcal{K},i)} \neq \tilde{\tilde{W}}^{(j\to\mathcal{K})}\right) < \tilde{\tilde{\lambda}}$ for $\mathcal{K} \in \mathcal{B}^{(j)}$ such that $i \in \mathcal{K}$ and $R^{(j\to\mathcal{K})} > 0$ and $I(\tilde{\tilde{P}}^n; \tilde{\tilde{W}}) < \frac{n\varepsilon}{12}$, where the double tilde on the public bits, and the message refers to the fact that the code operates at rate $\tilde{\tilde{R}}$.

Without loss of generality, we assume that the rates R_p, R_c of the confidential and public bit

pipes satisfy the following inequalities

$$\begin{aligned} R_c &\geq \max_{p(x)} I(X;Y) - \max_{p(x)} I(X;Z) - \frac{\varepsilon}{2} \\ R_p &\geq \max_{p(x)} I(X;Z) - \frac{\varepsilon}{2}. \end{aligned} \tag{3.20}$$

If not, we can replace the bit pipes with larger ones satisfying the inequalities above. Clearly a code that worked for the old network can be applied to the new network after the bit pipes are changed and the additional capacity has been added. Similar to Lemma 9 of Appendix F, we can assume without loss of generality that the public bit pipe is filled to capacity for the code of rate $\tilde{\tilde{R}}$; that is the public bit pipe $\tilde{\tilde{P}}^n$ carries a number of independent bits per transmission that can be made arbitrarily close to R_p. Specifically we can assume that

$$H(\tilde{\tilde{P}}^n) \geq n(R_p - \frac{\varepsilon}{2}). \tag{3.21}$$

by choosing parameter μ of Lemma 9 as $\mu = \frac{\varepsilon}{2}$.

Since (3.18) holds, it was shown in the proof of Theorem 8 that a stacked solution of N_1 layers for network $\mathbf{I}$ can be build using code $(\tilde{\lambda}, \varepsilon, A, \tilde{\tilde{R}})$–$\mathcal{S}(\mathcal{N}_{\bar{e}}(R_c, R_p), A)$ so that all receiving nodes $j \in \mathcal{N}$ at stacked network $\mathbf{I}$ in Figure 3.4 can receive all $\underline{\tilde{W}}^{(i\to\mathcal{B})}$ with rate $\tilde{R}^{(i\to\mathcal{B})}$ and probability of error $\Pr\left(\underline{\breve{W}}^{(i\to\mathcal{B})} \neq \underline{\tilde{W}}^{(i\to\mathcal{B})}\right) \leq \tilde{\lambda}/2$ for sufficiently large N_1.

We set the capacity $C_{\bar{e}}$ of the side channel equal to

$$C_{\bar{e}} = \frac{1}{n} H(\tilde{\tilde{C}}^n | \tilde{\tilde{P}}^n, \tilde{\tilde{W}}) + \varepsilon. \tag{3.22}$$

Therefore since the eavesdropper has access to message $\underline{\tilde{\tilde{W}}}$ and public bits $\underline{\tilde{\tilde{P}}}^n$ then the side channel has the necessary capacity to transfer enough bits so that the eavesdropper is able to decode the confidential bits $\underline{\tilde{\tilde{C}}}^n$. Indeed assume that we define the notion of typical set for the tuple $(\underline{\tilde{\tilde{W}}}, \underline{\tilde{\tilde{P}}}^n)$ of the message and the public bits as

$$\begin{aligned} A_\nu^{(N_1)}(\tilde{\tilde{W}}, \tilde{\tilde{P}}^n) = \Big\{ (\underline{\tilde{\tilde{w}}}, \underline{\tilde{\tilde{p}}}^n) : & \left| -\frac{1}{N_1} \log p(\underline{\tilde{\tilde{w}}}) - H(\tilde{\tilde{W}}) \right| \leq \nu, \\ & \left| -\frac{1}{N_1} \log p(\underline{\tilde{\tilde{p}}}^n) - H(\tilde{\tilde{P}}^n) \right| \leq \nu, \\ & \left| -\frac{1}{N_1} \log p(\underline{\tilde{\tilde{w}}}, \underline{\tilde{\tilde{p}}}^n) - H(\tilde{\tilde{W}}, \tilde{\tilde{P}}^n) \right| \leq \nu \Big\}. \end{aligned} \tag{3.23}$$

It can be proved that the probability of observing an atypical tuple $(\underline{\tilde{\tilde{w}}}, \underline{\tilde{\tilde{p}}}^n) \notin A_\nu^{(N_1)}(\tilde{\tilde{W}}, \tilde{\tilde{P}}^n)$ drops exponentially fast with increasing N_1. Indeed similar to Lemma 8 of [31] one can use the Chernoff bound and prove that $\Pr\left[\left(A_\nu^{(N_1)}(\tilde{\tilde{W}}, \tilde{\tilde{P}}^n)\right)^c\right] \leq 2^{-N_1 u(\nu)}$ with $u(\nu) > 0$, for large enough N_1. We will consider the case where we encounter an atypical $(\underline{\tilde{\tilde{w}}}, \underline{\tilde{\tilde{p}}}^n)$ tuple as an error event. Moreover

the conditional typical set $A_\nu^{(N_1)}(\tilde{\tilde{C}}^n|\underline{\tilde{\tilde{w}}},\underline{\tilde{\tilde{p}}}^n)$ with respect to a specific typical sequence $(\underline{\tilde{\tilde{w}}},\underline{\tilde{\tilde{p}}}^n) \in A_\nu^{(N_1)}(\tilde{\tilde{W}},\tilde{\tilde{P}}^n)$ is defined as

$$A_\nu^{(N_1)}(\tilde{\tilde{C}}^n|\underline{\tilde{\tilde{w}}},\underline{\tilde{\tilde{p}}}^n) = \left\{\underline{\tilde{\tilde{c}}}^n : \left|-\frac{1}{N_1}\log p(\underline{\tilde{\tilde{w}}},\underline{\tilde{\tilde{p}}}^n,\underline{\tilde{\tilde{c}}}^n) - H(\tilde{\tilde{W}},\tilde{\tilde{P}}^n,\tilde{\tilde{C}}^n)\right| \leq \nu\right\}. \tag{3.24}$$

By using the Chernoff bound as in Lemma 8 of [31] it can be proved that for all $(\underline{\tilde{\tilde{w}}},\underline{\tilde{\tilde{p}}}^n) \in A_\nu^{(N_1)}(\tilde{\tilde{W}},\tilde{\tilde{P}}^n)$ the probability of observing a $(\underline{\tilde{\tilde{c}}}^n)$ tuple so that $\underline{\tilde{\tilde{c}}}^n \notin A_\lambda^{(N_1)}(\tilde{\tilde{C}}^n|\underline{\tilde{\tilde{w}}},\underline{\tilde{\tilde{p}}}^n)$ is dropping exponentially fast, *i.e.* $\Pr\left[\left(A_\nu^{(N_1)}(\tilde{\tilde{C}}^n|\underline{\tilde{\tilde{w}}},\underline{\tilde{\tilde{p}}}^n)\right)^c\right] \leq 2^{-N_1 u'(\nu)}$ with $u'(\nu) > 0$, for large enough N_1. As before we will only consider the case of typical $\underline{\tilde{\tilde{c}}}^n$ given a typical $(\underline{\tilde{\tilde{w}}},\underline{\tilde{\tilde{p}}}^n)$ since the observing an atypical tuple will be regarded as an error event.

The size of the conditional typical set can be shown to be upper bounded by $|A_\nu^{(N_1)}(\tilde{\tilde{C}}^n \mid \underline{\tilde{\tilde{w}}},\underline{\tilde{\tilde{p}}}^n)| \leq 2^{N_1(H(\tilde{\tilde{C}}^n|\tilde{\tilde{W}},\tilde{\tilde{P}}^n)+2\nu)}$ for every $(\underline{\tilde{\tilde{w}}},\underline{\tilde{\tilde{p}}}^n) \in A_\nu^{(N_1)}(\tilde{\tilde{W}},\tilde{\tilde{P}}^n)$. Indeed

$$\begin{aligned}
1 &\geq \sum_{\underline{\tilde{\tilde{c}}}^n \in A_\nu^{(N_1)}(\tilde{\tilde{C}}^n \mid \underline{\tilde{\tilde{w}}},\underline{\tilde{\tilde{p}}}^n)} \frac{p(\underline{\tilde{\tilde{w}}},\underline{\tilde{\tilde{p}}}^n,\underline{\tilde{\tilde{c}}}^n)}{p(\underline{\tilde{\tilde{w}}},\underline{\tilde{\tilde{p}}}^n)} \\
&\overset{(a)}{\geq} \sum_{\underline{\tilde{\tilde{c}}}^n \in A_\nu^{(N_1)}(\tilde{\tilde{C}}^n \mid \underline{\tilde{\tilde{w}}},\underline{\tilde{\tilde{p}}}^n)} \frac{2^{-N_1(H(\tilde{\tilde{W}},\tilde{\tilde{P}}^n,\tilde{\tilde{C}}^n)+\nu)}}{2^{-N_1(H(\tilde{\tilde{W}},\tilde{\tilde{P}}^n)-\nu)}} \\
&= \left|A_\nu^{(N_1)}(\tilde{\tilde{C}}^n \mid \underline{\tilde{\tilde{w}}},\underline{\tilde{\tilde{p}}}^n)\right| 2^{-N_1(H(\tilde{\tilde{C}}^n|\tilde{\tilde{W}},\tilde{\tilde{P}}^n)+2\nu)}
\end{aligned}$$

where inequality (a) is based on the definitions (3.23) and (3.24) of the typical sets. Therefore the size of the conditional typical set $\left|A_\nu^{(N_1)}(\tilde{\tilde{C}}^n \mid \underline{\tilde{\tilde{w}}},\underline{\tilde{\tilde{p}}}^n)\right|$ is upper bounded by

$$\left|A_\nu^{(N_1)}(\tilde{\tilde{C}}^n \mid \underline{\tilde{\tilde{w}}},\underline{\tilde{\tilde{p}}}^n)\right| \leq 2^{N_1(H(\tilde{\tilde{C}}^n|\tilde{\tilde{W}},\tilde{\tilde{P}}^n)+2\nu)} \tag{3.25}$$

Therefore for each $(\underline{\tilde{\tilde{w}}},\underline{\tilde{\tilde{p}}}^n) \in A_\nu^{(N_1)}(\tilde{\tilde{W}},\tilde{\tilde{P}}^n)$ all conditionally typical $\underline{\tilde{\tilde{c}}}^n \in A_\nu^{(N_1)}(\tilde{\tilde{C}}^n \mid \underline{\tilde{\tilde{w}}},\underline{\tilde{\tilde{p}}}^n)$ are assigned a unique bit sequence. Due to the size of the conditional typical set $A_\nu^{(N_1)}(\tilde{\tilde{C}}^n \mid \underline{\tilde{\tilde{w}}},\underline{\tilde{\tilde{p}}}^n)$ one need to use at most $N_1(H(\tilde{\tilde{C}}^n|\tilde{\tilde{W}},\tilde{\tilde{P}}^n)+2\nu)+1$ bits to uniquely identify each $\underline{\tilde{\tilde{c}}}^n$ inside the set. This one-to-one mapping between all $\underline{\tilde{\tilde{c}}}^n \in A_\nu^{(N_1)}(\tilde{\tilde{C}}^n \mid \underline{\tilde{\tilde{w}}},\underline{\tilde{\tilde{p}}}^n)$ for all typical $(\underline{\tilde{\tilde{w}}},\underline{\tilde{\tilde{p}}}^n)$ is revealed to the eavesdropper and since the eavesdropper has access to both the message $\underline{\tilde{\tilde{w}}}$ transmitted and the public bits $\underline{\tilde{\tilde{p}}}^n$, the super-source only needs to transmit through the noiseless channel of capacity $C_{\bar{e}}$ the $N_1(H(\tilde{\tilde{C}}^n|\tilde{\tilde{W}},\tilde{\tilde{P}}^n)+2\nu)+1$ bits that identify $\underline{\tilde{\tilde{c}}}^n$ given $(\underline{\tilde{\tilde{w}}},\underline{\tilde{\tilde{p}}}^n)$. Through the noiseless side channel one can transfer error free $C_{\bar{e}}$ bits per use or else $N_1(H(\tilde{\tilde{C}}^n|\tilde{\tilde{W}},\tilde{\tilde{P}}^n)+n\varepsilon)$ after nN_1 uses of the N_1 layers of the stacked network. Therefore if $\nu = \varepsilon/4$ then the error free channel has enough capacity to transfer all $N_1(H(\tilde{\tilde{C}}^n|\tilde{\tilde{W}},\tilde{\tilde{P}}^n)+2\nu)+1$ bits needed for the decoding of the $\underline{\tilde{\tilde{c}}}^n$ tuple for large enough N_1.

The two sources of error on the code for network **I** is when the one of the messages $\underline{\tilde{W}}^{(i\to j)}$ is de-

coded erroneously and this happens with probability at most $\tilde{\lambda}/2$, or when $(\underline{\tilde{\tilde{w}}}, \underline{\tilde{\tilde{p}}}^n) \notin A_\nu^{(N_1)}(\tilde{\tilde{W}}, \tilde{\tilde{P}}^n)$ or if for a typical $(\underline{\tilde{\tilde{w}}}, \underline{\tilde{\tilde{p}}}^n) \notin A_\nu^{(N_1)}(\tilde{\tilde{W}}, \tilde{\tilde{P}}^n)$ then the confidential bits are atypical *i.e.* $\underline{\tilde{\tilde{c}}}^n \notin A_\lambda^{(N_1)}(\tilde{\tilde{C}}^n | \underline{\tilde{\tilde{w}}}, \underline{\tilde{\tilde{p}}}^n)$. By taking the numbers of layers N_1 of the stacked version of network **I** large enough one can ensure that both events happen with probability less than $\tilde{\lambda}/4$ and therefore we have devised a code for a stacked version of network **I** of rate $\tilde{R}$ and overall probability of error less than or equal to λ. This code from now one will be viewed as a code of blocklength $n_1 = nN_1$ for the non-stacked version of network **I** where for notational convenience the public bits $\underline{\tilde{\tilde{P}}}^{n_1}$, the confidential bits $\underline{\tilde{\tilde{C}}}^{n_1}$ and the messages $\underline{\tilde{\tilde{W}}}$ and $\underline{\tilde{W}}$ of rate $\tilde{\tilde{R}}$ and $\tilde{R}$ are denoted as $\tilde{\tilde{P}}^{n_1}$, $\tilde{\tilde{C}}^{n_1}$ $\tilde{\tilde{W}}^{n_1}$, $\tilde{W}^{n_1}$ respectively.

For all E the eavesdroppers observation $(\underline{\tilde{\tilde{Z}}}_{\mathbf{I}}^{(E)})^{n_1}$ in network **I** is the same as the eavesdropper's observation is network $\mathcal{N}_{\bar{e}}(R_c, R_p)$ and similar to the proof of Theorem 8 where each layer is independent of the others we get

$$\mathbb{E}_{\mathcal{C}}\left[I\big(\tilde{\tilde{W}}^{n_1}; (\underline{\tilde{\tilde{Z}}}_{\mathbf{I}}^{(E)})^{n_1}\big)\right] \leq \frac{n_1 \varepsilon}{12} \ \forall E \in A. \tag{3.26}$$

I to II: We use a stacked version of network **I** with N_2 layers (that is in general different from N_1) where on each layer of the stack we apply the code of blocklength n_1 and rate $\tilde{R}$ with probability of error $\Pr\left(\breve{\tilde{W}}^{n_1} \neq \tilde{W}^{n_1} \cup (\breve{\tilde{\tilde{C}}}^{n_1}, \breve{\tilde{\tilde{P}}}^{n_1}, \breve{\tilde{\tilde{W}}}^{n_1}) \neq (\tilde{\tilde{C}}^{n_1}, \tilde{\tilde{P}}^{n_1}, \tilde{\tilde{W}}^{n_1})\right) \leq \tilde{\lambda}$. Networks **I** and **II** are identical except edge $\bar{e}$ where the noiseless bit-pipes of network **I** have been replaced by a broadcast channel in network **II**. The stacked code of network **I** is transformed into a code for the stacked version of network **II** with N_2 layers by adding a channel code at edge $\bar{e}$ of network **II** so that the confidential and public bits are transmitted through the channel that replaced the noiseless bit pipes of $\bar{e}$ of network **I**. It will be proved that the eavesdropper at network **II** can decode the public bits $\underline{\tilde{\tilde{P}}}^{n_1}$ (where the underscore refers to the N_2 layers of the stacked) by overhearing the noisy transmission $\underline{\tilde{\tilde{Z}}}^{n_1}$, the noiseless bits $\underline{\tilde{\tilde{L}}}^{n_1}$ that go through the side link of capacity $C_{\bar{e}}$ and the message $\underline{\tilde{\tilde{W}}}^{n_1}$ and therefore the eavesdropper can apply the same code as that was used for network **I** and additionally decode the confidential bits $\underline{\tilde{\tilde{C}}}^{n_1}$. Below we will explain in details and give the proofs of all the above steps.

Since (3.19) holds as discussed in the proof of Theorem 8 that a stacked solution of N_2 layers for network **I** can be build using the random code of blocklength n_1 and rate $\tilde{R}$ that was designed for network **I**. For sufficiently large N_2 all messages can be delivered with rate $R^{(i \to \mathcal{B})}$ and probability of error $\Pr\left(\breve{\underline{W}}^{(i \to \mathcal{B})} \neq \underline{W}^{(i \to \mathcal{B})}\right) \leq \lambda/6$.

For edge $\bar{e}$ in network **II** at every time step $t \in \{1, \ldots, n_1\}$ once both the public $\underline{\tilde{\tilde{P}}}_t$ and the confidential $\underline{\tilde{\tilde{C}}}_t$ bits have been received a random channel encoder/decoder pair $\{a_{N_2,t}, b_{N_2,t}\}_{t=1}^{n_1}$ is applied to transfer these bits through edge $\bar{e}$. At each time step $t \in \{1, \ldots, n_1\}$ there are $N_2(R_p + R_c)$ bits delivered at the N_2 layers of edge $\bar{e}$ and these bits have to be conveyed to the receiver through the noisy channel that replaced the noiseless bits of edge $\bar{e}$. The $N_2(R_p + R_c)$ bits at each time step

correspond to $2^{N_2(R_p+R_c)}$ incoming messages $m_t(i)$, $i \in \left\{1,\ldots,2^{N_2(R_p+R_c)}\right\}$ at edge $\bar{e}$. The channel encoders $a_{N_2,t}$ at each time step t have assigned to each one of the $2^{N_2(R_p+R_c)}$ incoming messages $m_t(i)$ a random N_2–tuple $\underline{x}_t(i) = (x_{t1}(i),\ldots,x_{tN_2}(i))$ where $x_{tj}(i)$, $j \in \{1,\ldots,N_2\}$ are chosen from the distribution $p(x)$ that gives rise to the corresponding mutual informations $\max\limits_{p(x)} I(X;Y)$ and $\max\limits_{p(x)} I(X;Z)$ of the noisy channel $\bar{e}$ for network **II**. This mapping is revealed to both the eavesdropper and the output of $\bar{e}$.

Once the $N_2(R_p+R_c)$ public and confidential bits at time t are ready for transmission then the corresponding N_2–tuple is transmitted through the N_2 noisy channels across the N_2 layers and the intended receiver gets $\underline{y}_t$ while the eavesdropper gets $\underline{z}_t$. From the received N_2–tuple $\underline{y}_t$ the decoder finds the N_2–tuple $\underline{x}_t(i)$ corresponding to message $m_t(i)$ so that $(\underline{x}_t(i),\underline{y}_t)$ are jointly typical, *i.e.* find message $m_t(i)$ such that $(\underline{x}_t(i),\underline{y}_t) \in A_{t,\gamma}^{(N_2)}(X,Y)$ where

$$\begin{aligned} A_{t,\gamma}^{(N_2)}(X,Y) = \Bigg\{(\underline{x},\underline{y}) \in \underline{\mathcal{X}}^{V_1(\bar{e})} \times \underline{\mathcal{Y}}^{V_2(\bar{e})} : \ & \left|-\frac{1}{N_2}\log p_t(\underline{x}) - H_t(X)\right| \leq \gamma, \\ & \left|-\frac{1}{N_2}\log p_t(\underline{y}) - H_t(Y)\right| \leq \gamma, \\ & \left|-\frac{1}{N_2}\log p_t(\underline{x},\underline{y}) - H_t(X,Y)\right| \leq \gamma\Bigg\} \end{aligned} \tag{3.27}$$

where $p_t(x,y) = p(y|x)p_t(x)$. Similar to (3.2) define indicator function I as

$$I = \begin{cases} 1, & \text{There is } t \in \{1,\ldots,n_1\} \text{ such that the encode/decoder pair } \{a_{N_2,t}, b_{N_2,t}\} \text{ fails} \\ 0, & \text{otherwise} \end{cases} \tag{3.28}$$

The encoder/decoder pair fail if the received sequence $\underline{y}_t$ is not jointly typical with what was sent and this probability is upper bounded by $2^{-N_2 g_t(\gamma)}$ for some $g_t(\gamma) > 0$ [60, Chapter 7] or if there are more than one sequences $\underline{x}_t(j)$, $\underline{x}_t(k)$ with $j \neq k$ that are jointly typical with the received sequence $\underline{y}_t$. According to Theorem 7.6.1 of [60] the probability that $\underline{y}_t$ is jointly typical with some $\underline{x}_t(j)$ where $m_t(j)$ is different from the message $m_t(i)$ that was sent is upper bounded by $2^{-N_2(\max_{p(x)} I(X;Y)-3\gamma)}$ and therefore the overall average probability of error at each time step t for the channel code is upper bounded by

$$2^{-N_2 g_t(\gamma)} + 2^{N_2(R_c+R_p)} 2^{-N_2(\max_{p(x)} I(X;Y)-3\gamma)} \leq 2^{-N_2 g_t(\gamma)} + 2^{-N_2\left[\left(\max_{p(x)} I(X;Y)-(R_c+R_p)\right)-3\gamma\right]}.$$

The above probability is the average probability or error but since there is no guarantee that the messages $m_t(i)$ that needs to transmitted with the channel code are equiprobable then we have to worry about the expected probability of error instead of the average. In Appendix I of [32] it was proved that by a careful choice of the channel code's index assignments, each channel code for every t can have an expected error probability no greater than the average probability. So by

choosing $\gamma = \frac{1}{6}\left(\max_{p(x)} I(X;Y) - (R_c + R_p)\right)$ we can get that the channel code fails at time t with an expected probability upper bounded by

$$2^{-N_2 g_t(\gamma)} + 2^{-\frac{N_2}{2}\left(\max_{p(x)} I(X;Y) - (R_c+R_p)\right)}, \tag{3.29}$$

and by using the union bound across the n_1 time steps we can upper bound the probability of event $(I = 1)$ as

$$\Pr(I = 1) \leq 2^{-N_2 \delta_I}. \tag{3.30}$$

Before we continue we need to give a few definitions that will be used in equation (3.35). Let $\mathcal{S}$ be a subset of set $\{1, \ldots, n_1\}$, *i.e.* $\mathcal{S} \subseteq \{1, \ldots, n_1\}$, including the empty set as well as the whole set, and $\mathcal{S}^c$ to be the complement of $\mathcal{S}$, *i.e.* $\mathcal{S}^c = \{1, \ldots, n_1\} \backslash \mathcal{S}$. We define $F_{\mathcal{S}}$ to be the set $\{F_i : i \in \mathcal{S}\}$ and for reasons that will become obvious from equations (3.36), (3.37) and (G.1) we need to compute an upper bound on quantity $H(\tilde{\tilde{P}}_{\mathcal{S}^c}, \tilde{\tilde{C}}_{\mathcal{S}^c} | \tilde{\tilde{L}}_{\bar{e}}^{n_1}, \tilde{\tilde{W}}^{n_1}, \tilde{\tilde{P}}_{\mathcal{S}}, \tilde{\tilde{C}}_{\mathcal{S}})$. For the code of blocklength n_1 for network $\mathbf{I}$ the eavesdropper receives $\tilde{\tilde{L}}_{\bar{e}}^{n_1}$ that are the bits sent by the link of capacity $C_{\bar{e}}$, the public bits $\tilde{\tilde{P}}^{n_1}$ and the message $\tilde{\tilde{W}}^{n_1}$ and can decode the confidential bits $\tilde{\tilde{C}}^{n_1}$ with probability of error at most λ. Therefore due to Fano's inequality [60] we get

$$H(\tilde{\tilde{C}}^{n_1} | \tilde{\tilde{P}}^{n_1}, \tilde{\tilde{L}}_{\bar{e}}^{n_1}, \tilde{\tilde{W}}^{n_1}) \leq h(\lambda) + n_1 R_c \lambda. \tag{3.31}$$

where we remind that $h(p) = -p \log_2 p - (1-p)\log_2(1-p)$ is the binary entropy function. Therefore from the above inequality we get

$$\begin{aligned} H(\tilde{\tilde{P}}^{n_1}, \tilde{\tilde{L}}_{\bar{e}}^{n_1}, \tilde{\tilde{W}}^{n_1}, \tilde{\tilde{C}}^{n_1}) &= H(\tilde{\tilde{P}}^{n_1}, \tilde{\tilde{L}}_{\bar{e}}^{n_1}, \tilde{\tilde{W}}^{n_1}) + H(\tilde{\tilde{C}}^{n_1} | \tilde{\tilde{P}}^{n_1}, \tilde{\tilde{L}}_{\bar{e}}^{n_1}, \tilde{\tilde{W}}^{n_1}) \\ &\leq H(\tilde{\tilde{P}}^{n_1}, \tilde{\tilde{L}}_{\bar{e}}^{n_1}, \tilde{\tilde{W}}^{n_1}) + h(\lambda) + n_1 R_c \lambda \end{aligned}$$

or else by applying the chain rule for entropies on both sides of the above inequality

$$\begin{aligned} &H(\tilde{\tilde{L}}_{\bar{e}}^{n_1}, \tilde{\tilde{W}}^{n_1}, \tilde{\tilde{P}}_{\mathcal{S}}, \tilde{\tilde{C}}_{\mathcal{S}}) + H(\tilde{\tilde{P}}_{\mathcal{S}^c}, \tilde{\tilde{C}}_{\mathcal{S}^c} | \tilde{\tilde{L}}_{\bar{e}}^{n_1}, \tilde{\tilde{W}}^{n_1}, \tilde{\tilde{P}}_{\mathcal{S}}, \tilde{\tilde{C}}_{\mathcal{S}}) \leq H(\tilde{\tilde{L}}_{\bar{e}}^{n_1}, \tilde{\tilde{W}}^{n_1}, \tilde{\tilde{P}}_{\mathcal{S}}) + H(\tilde{\tilde{P}}_{\mathcal{S}^c} | \tilde{\tilde{L}}_{\bar{e}}^{n_1}, \tilde{\tilde{W}}^{n_1}, \tilde{\tilde{P}}_{\mathcal{S}}) + h(\lambda) + n_1 R_c \lambda \\ \Rightarrow\ & H(\tilde{\tilde{C}}_{\mathcal{S}} | \tilde{\tilde{L}}_{\bar{e}}^{n_1}, \tilde{\tilde{W}}^{n_1}, \tilde{\tilde{P}}_{\mathcal{S}}) + H(\tilde{\tilde{P}}_{\mathcal{S}^c}, \tilde{\tilde{C}}_{\mathcal{S}^c} | \tilde{\tilde{L}}_{\bar{e}}^{n_1}, \tilde{\tilde{W}}^{n_1}, \tilde{\tilde{P}}_{\mathcal{S}}, \tilde{\tilde{C}}_{\mathcal{S}}) \leq H(\tilde{\tilde{P}}_{\mathcal{S}^c} | \tilde{\tilde{L}}_{\bar{e}}^{n_1}, \tilde{\tilde{W}}^{n_1}, \tilde{\tilde{P}}_{\mathcal{S}}) + h(\lambda) + n_1 R_c \lambda \\ &\Rightarrow H(\tilde{\tilde{P}}_{\mathcal{S}^c}, \tilde{\tilde{C}}_{\mathcal{S}^c} | \tilde{\tilde{L}}_{\bar{e}}^{n_1}, \tilde{\tilde{W}}^{n_1}, \tilde{\tilde{P}}_{\mathcal{S}}, \tilde{\tilde{C}}_{\mathcal{S}}) \leq H(\tilde{\tilde{P}}_{\mathcal{S}^c}) + h(\lambda) + n_1 R_c \lambda \end{aligned}$$

and since $H(\tilde{\tilde{P}}_{\mathcal{S}^c}) \leq (n_1 - |\mathcal{S}|) R_p$ we get

$$\begin{aligned} H(\tilde{\tilde{P}}_{\mathcal{S}^c}, \tilde{\tilde{C}}_{\mathcal{S}^c} | \tilde{\tilde{L}}_{\bar{e}}^{n_1}, \tilde{\tilde{W}}^{n_1}, \tilde{\tilde{P}}_{\mathcal{S}}, \tilde{\tilde{C}}_{\mathcal{S}}) &\leq (n_1 - |\mathcal{S}|) R_p + h(\lambda) + n_1 R_c \lambda \\ &\leq (n_1 - |\mathcal{S}|)\left(R_p + h(\lambda) + n_1 R_c \lambda\right) \end{aligned}$$

where the last inequality holds since we choose set $\mathcal{S}$ so that $|\mathcal{S}| \leq n_1 - 1$. This is done so that $\mathcal{S}^c \neq \emptyset$ in order $\tilde{\tilde{P}}_{\mathcal{S}^c}$ and $\tilde{\tilde{C}}_{\mathcal{S}^c}$ to have a non-trivial meaning. Moreover by choosing λ small enough so that $h(\lambda) \leq \frac{\max_{p(x)} I(X;Z)-R_p}{4}$ and $\lambda \leq \frac{\max_{p(x)} I(X;Z)-R_p}{4n_1 R_c}$ we get

$$H(\tilde{\tilde{P}}_{\mathcal{S}^c}, \tilde{\tilde{C}}_{\mathcal{S}^c} | \tilde{\tilde{L}}_{\tilde{e}}^{n_1}, \tilde{\tilde{W}}^{n_1}, \tilde{\tilde{P}}_{\mathcal{S}}, \tilde{\tilde{C}}_{\mathcal{S}}) \leq \frac{n_1 - |\mathcal{S}|}{2} \left(R_p + \max_{p(x)} I(X;Z) \right). \tag{3.32}$$

As we discussed above the public and the confidential bits are transmitted through the channel with the use of a random code. We define the notion of typicality for each time step $t \in \{1, \ldots, n_1\}$ with respect to the eavesdropper's channel as the following set

$$\begin{aligned} A_{t,\beta}^{(N_2)}(X,Z) = \Bigg\{ (\underline{x}, \underline{z}) \in \underline{\mathcal{X}} \times \underline{\mathcal{Z}} : & \left| -\frac{1}{N_2} \log p_t(\underline{x}) - H_t(X) \right| \leq \beta, \\ & \left| -\frac{1}{N_2} \log p_t(\underline{z}) - H_t(Z) \right| \leq \beta, \\ & \left| -\frac{1}{N_2} \log p_t(\underline{x}, \underline{z}) - H_t(X,Z) \right| \leq \beta \Bigg\}. \end{aligned} \tag{3.33}$$

Upon the reception of the information from the degraded channel $(\underline{z}_1, \ldots, \underline{z}_{n_1})$, where $\underline{z}_t \in \mathcal{Z}^{N_2}$ $\forall t \in \{1, \ldots, n_1\}$, the eavesdropper tries to find the an n_1–tuple of indexes $(i_1, \ldots, i_{n_1})$ so that the transmitted codeword $(\underline{x}_1(i_1), \ldots, \underline{x}_{n_1}(i_{n_1}))$ has the property $(\underline{x}_t(i_t), \underline{z}_t) \in A_{t,\beta}^{(N_2)}$ for all $t \in \{1, \ldots, n_1\}$ (then we call that the two sequences are jointly typical). An error occurs if the received sequence $(\underline{z}_1, \ldots, \underline{z}_{n_1})$ is not jointly typical with what was sent and the probability of this event is upper bounded by $n_1 2^{-N_2 k'(\beta)}$ for some $k'(\beta) > 0$ where the term n_1 comes from the union bound over all the n_1 time steps and the term $2^{-N_2 k'(\beta)}$ comes from an argument identical to the one used in Lemma 8 of [31]. For sufficiently large N_2 and $k(\beta) = \frac{1}{2} k'(\beta)$ we get that the probability sequence $(\underline{z}_1, \ldots, \underline{z}_{n_1})$ is not jointly typical with what was sent is upper bounded by $2^{-N_2 k(\beta)}$. A decoding error can also occur if there are more than one sequences $(\underline{x}_1(i_1), \ldots, \underline{x}_{n_1}(i_{n_1}))$ that are jointly typical with the received sequence $(\underline{z}_1, \ldots, \underline{z}_{n_1})$. The probability of this event is computed in detail in the following. Since we consider a random code meaning that we will compute the average probability of error of all codes created i.i.d. by the distribution $p(x)$ that maximizes $I(X;Y)$ and $I(X;Z)$ then without loss of generality we will assume that for all time steps $t \in \{1, \ldots, n_1\}$ N_2–tuple $\underline{x}_t(1)$ was the one transmitted. The eavesdropper will find that $(\underline{x}_1(1), \ldots, \underline{x}_{n_1}(1))$ (compactly written as $\underline{x}^{n_1}(1)$) is jointly typical with what was received $(\underline{z}_1, \ldots, \underline{z}_{n_1})$ (compactly written as $\underline{z}^{n_1}$) and now we will prove that the probability of having any other $(\underline{x}_1(i_1), \ldots, \underline{x}_{n_1}(i_{n_1}))$ (compactly written as $\underline{x}^{n_1}(i^{n_1})$) with $i^{n_1} = (i_1, \ldots, i_{n_1}) \neq (1, \ldots, 1)$ that is typical with $(\underline{z}_1, \ldots, \underline{z}_{n_1})$ is exponentially small. Specifically if all i_j for $j \in \{1, \ldots, n_1\}$ are different from 1 then the probability (averaging over all codes $\mathcal{C}$) that $(\underline{x}_1(i_1), \ldots, \underline{x}_{n_1}(i_{n_1}))$ is jointly typical with $(\underline{z}_1, \ldots, \underline{z}_{n_1})$ is upper bounded by

$2^{-N_2 n_1\left(\max_{p(x)} I(X;Z)+3\beta\right)}$. Indeed

$$\Pr\left(\forall t:(\underline{x}_t(i_t),\underline{z}_t)\in A_{t,\beta}^{(N_2)}(X,Z)\right)=\sum_{\mathcal{C}}\Pr\left\{\forall t:(\underline{x}_t(i_t),\underline{z}_t)\in A_{t,\beta}^{(N_2)}(X,Z)\,|\mathcal{C}\right\}\Pr(\mathcal{C})$$

where $\mathcal{C}$ represents the code. For the specific expression above the part of the code that is of interest is $\underline{x}^{n_1}(1)$ and $\underline{x}^{n_1}(i^{n_1})$ and therefore the right hand side of the above expression becomes

$$\begin{aligned}
&\sum_{\underline{x}^{n_1}(1),\underline{x}^{n_1}(i^{n_1})}\mathbb{P}\mathrm{r}\left\{\forall t:(\underline{x}_t(i_t),\underline{z}_t)\in A_{t,\beta}^{(N_2)}(X,Z)\,|\underline{x}^{n_1}(1),\underline{x}^{n_1}(i^{n_1})\right\}\mathbb{P}\mathrm{r}\left(\underline{x}^{n_1}(1),\underline{x}^{n_1}(i^{n_1})\right)\\
&\overset{(a)}{=}\sum_{\underline{x}^{n_1}(1),\underline{x}^{n_1}(i^{n_1}),\underline{z}^{n_1}}\mathbb{P}\mathrm{r}\left\{\forall t:(\underline{x}_t(i_t),\underline{z}_t)\in A_{t,\beta}^{(N_2)}(X,Z)\cap\underline{z}^{n_1}\,|\underline{x}^{n_1}(1),\underline{x}^{n_1}(i^{n_1})\right\}\mathbb{P}\mathrm{r}\left(\underline{x}^{n_1}(1)\right)\mathbb{P}\mathrm{r}\left(\underline{x}^{n_1}(i^{n_1})\right)\\
&=\sum_{\underline{x}^{n_1}(1),\underline{x}^{n_1}(i^{n_1}),\underline{z}^{n_1}}\mathbb{P}\mathrm{r}\left\{\forall t:(\underline{x}_t(i_t),\underline{z}_t)\in A_{t,\beta}^{(N_2)}(X,Z)\,|\underline{x}^{n_1}(1),\underline{x}^{n_1}(i^{n_1}),\underline{z}^{n_1}\right\}\mathbb{P}\mathrm{r}\left(\underline{z}^{n_1}\,|\underline{x}^{n_1}(1),\underline{x}^{n_1}(i^{n_1})\right)\mathbb{P}\mathrm{r}\left(\underline{x}^{n_1}(1)\right)\mathbb{P}\mathrm{r}\left(\underline{x}^{n_1}(i^{n_1})\right)\\
&\overset{(b)}{=}\sum_{\underline{x}^{n_1}(1),\underline{x}^{n_1}(i^{n_1}),\underline{z}^{n_1}}\Big(\prod_{t\in\{1,\ldots,n_1\}}\mathbb{P}\mathrm{r}\left\{(\underline{x}_t(i_t),\underline{z}_t)\in A_{t,\beta}^{(N_2)}(X,Z)\,|\underline{x}_t(i_t),\underline{z}_t\right\}\Big)\mathbb{P}\mathrm{r}\left(\underline{z}^{n_1}\,|\underline{x}^{n_1}(1)\right)\mathbb{P}\mathrm{r}\left(\underline{x}^{n_1}(1)\right)\mathbb{P}\mathrm{r}\left(\underline{x}^{n_1}(i^{n_1})\right)\\
&=\sum_{\underline{x}^{n_1}(i^{n_1}),\underline{z}^{n_1}}\Big(\prod_{t\in\{1,\ldots,n_1\}}\mathbb{P}\mathrm{r}\left\{(\underline{x}_t(i_t),\underline{z}_t)\in A_{t,\beta}^{(N_2)}(X,Z)\,|\underline{x}_t(i_t),\underline{z}_t\right\}\Big)\mathbb{P}\mathrm{r}\left(\underline{x}^{n_1}(i^{n_1})\right)\sum_{\underline{x}^{n_1}(1)}\mathbb{P}\mathrm{r}\left(\underline{z}^{n_1},\underline{x}^{n_1}(1)\right)\\
&=\sum_{\underline{x}^{n_1}(i^{n_1}),\underline{z}^{n_1}}\Big(\prod_{t\in\{1,\ldots,n_1\}}\mathbb{P}\mathrm{r}\left\{(\underline{x}_t(i_t),\underline{z}_t)\in A_{t,\beta}^{(N_2)}(X,Z)\,|\underline{x}_t(i_t),\underline{z}_t\right\}\Big)\mathbb{P}\mathrm{r}\left(\underline{x}^{n_1}(i^{n_1})\right)\mathbb{P}\mathrm{r}\left(\underline{z}^{n_1}\right)\\
&\overset{(c)}{=}\sum_{\underline{x}^{n_1}(i^{n_1}),\underline{z}^{n_1}}\prod_{t\in\{1,\ldots,n_1\}}\mathbb{P}\mathrm{r}\left\{(\underline{x}_t(i_t),\underline{z}_t)\in A_{t,\beta}^{(N_2)}(X,Z)\,|\underline{x}_t(i_t),\underline{z}_t\right\}\mathbb{P}\mathrm{r}\left(\underline{x}_t(i_t)\right)\mathbb{P}\mathrm{r}\left(\underline{z}_t\right)\\
&=\prod_{t\in\{1,\ldots,n_1\}}\sum_{\underline{x}_t(i_t),\underline{z}_t}\mathbb{P}\mathrm{r}\left\{(\underline{x}_t(i_t),\underline{z}_t)\in A_{t,\beta}^{(N_2)}(X,Z)\,|\underline{x}_t(i_t),\underline{z}_t\right\}\mathbb{P}\mathrm{r}\left(\underline{x}_t(i_t)\right)\mathbb{P}\mathrm{r}\left(\underline{z}_t\right)\\
&=\prod_{t\in\{1,\ldots,n_1\}}\sum_{(\underline{x}_t(i_t),\underline{z}_t)\in A_{t,\beta}^{(N_2)}}\mathbb{P}\mathrm{r}\left(\underline{x}_t(i_t)\right)\mathbb{P}\mathrm{r}\left(\underline{z}_t\right)\\
&\overset{(d)}{\leq}\prod_{t\in\{1,\ldots,n_1\}}2^{N(H_t(X,Z)+\beta)}2^{-N(H_t(X)-\beta)}2^{-N(H_t(Z)-\beta)}\\
&\overset{(f)}{=}2^{-n_1N(\max_{p(x)}I(X;Z)-3\beta)}
\end{aligned}\tag{3.34}$$

where (a) and (c) hold since the codebook is chosen independently for different time steps, (b) holds since $\underline{z}^{n_1}$ is conditionally independent of $\underline{x}^{n_1}(i^{n_1})$ when $\underline{x}^{n_1}(1)$ is given and (d) holds since the codebook is created by the distribution $p(x)$ that maximizes the mutual information $I(X;Z)$. In the inequality above we assumed that all indexes $i_1,\ldots,i_{n_1}$ are different from 1. If some of the indexes are equal to 1 then the probability of joint typicality is larger. Specifically define the subset $\mathcal{S}\subseteq\{1,\ldots,n_1\}$ so that $\forall t\in\mathcal{S}\ i_t=1$ whereas $\forall t\in\mathcal{S}^c\ i_t\neq 1$. Then the probability that $(\underline{x}_1(i_1),\ldots,\underline{x}_{n_1}(i_{n_1}))$ is jointly typical with $(\underline{z}_1,\ldots,\underline{z}_{n_1})$ is upper bounded by $2^{-(n_1-|\mathcal{S}|)N_2\left(\max_{p(x)}I(X;Z)-3\beta\right)}$. Indeed we just need to do the analysis that derived equation (3.34) only for the indexes in the set $\mathcal{S}^c$ and upper bound the probability that $(\underline{x}_t(i_t),\underline{z}_t)\in A_{t,\beta}^{(N_2)}(X,Z)$

for $t \in \mathcal{S}$ by 1. Therefore

$$\Pr\left(\forall t : (\underline{x}_t(i_t), \underline{z}_t) \in A_{t,\beta}^{(N_2)}(X,Z)\right) \leq 2^{-(n_1-|\mathcal{S}|)N_2\left(\max_{p(x)} I(X;Z)-3\beta\right)} \tag{3.35}$$

if $\forall t \in \mathcal{S}\ i_t = 1$ and $\forall t \in \mathcal{S}^c\ i_t \neq 1$ and remember that we assume that $\underline{x}^n(1)$ was the message transmitted.

In Appendix G we define the notion of typicality for vector $(\tilde{\underline{L}}^{n_1}, \tilde{\underline{W}}^{n_1}, \tilde{\underline{P}}^{n_1}, \tilde{\underline{C}}^{n_1})$ and the notion of conditional typicality for vector $(\tilde{\underline{P}}^{n_1}, \tilde{\underline{C}}^{n_1})$ given a typical vector $(\tilde{\underline{\ell}}^{n_1}, \tilde{\underline{w}}^{n_1})$. We also showed that with very high probability vectors $(\tilde{\underline{L}}^{n_1}, \tilde{\underline{W}}^{n_1}, \tilde{\underline{P}}^{n_1}, \tilde{\underline{C}}^{n_1})$ will be typical and therefore the event that this vector is atypical is regarded as an error event and consequently the eavesdropper will only search among the typical $(\tilde{\underline{P}}^{n_1}, \tilde{\underline{C}}^{n_1})$ given $(\tilde{\underline{\ell}}^{n_1}, \tilde{\underline{w}}^{n_1})$ to find the one transmitted. Once the bits from the side link $\tilde{\underline{\ell}}^{n_1}$, and the message $\tilde{\underline{w}}^{n_1}$ have been received the number of conditionally typical public and confidential bits $(\tilde{\underline{P}}^{n_1}, \tilde{\underline{C}}^{n_1})$ given the specific $\tilde{\underline{\ell}}^{n_1}$ and $\tilde{\underline{w}}^{n_1}$ is upper bounded by $2^{\frac{n_1 N_2}{2}\left(R_p+\max_{p(x)} I(X;Z)\right)+2N_2\omega}$ according to equation (G.1). Therefore the probability that there is another $(\underline{x}_1(i_1), \ldots, \underline{x}_{n_1}(i_{n_1}))$ (other than $(\underline{x}_1(1), \ldots, \underline{x}_{n_1}(1))$) with, $i_j \neq 1, \forall j \in \{1, \ldots, n_1\}$ that is jointly typical with $(\underline{z}_1, \ldots, \underline{z}_{n_1})$ is upper bounded by

$$\begin{aligned}
\Pr(E_\emptyset) &\overset{(a)}{\leq} \left|A_\omega^{(N_2)}\left(\tilde{\underline{P}}^{n_1}, \tilde{\underline{C}}^{n_1} \mid \tilde{\underline{\ell}}^{n_1}, \tilde{\underline{w}}^{n_1}\right)\right| 2^{-nN_2\left(\max_{p(x)} I(X;Z)-3\beta\right)} \\
&\overset{(b)}{\leq} 2^{\frac{n_1 N_2}{2}\left(R_p+\max_{p(x)} I(X;Z)\right)+2N\omega}\, 2^{-n_1 N_2\left(\max_{p(x)} I(X;Z)-3\beta\right)} \\
&= 2^{-\frac{n_1 N_2}{2}\left(\max_{p(x)} I(X;Z)-R_p\right)+2N_2\omega+3n_1N_2\beta} \\
&\leq 2^{-\frac{N_2}{2}\left(\max_{p(x)} I(X;Z)-R_p\right)+2N\omega+3n_1N_2\beta}.
\end{aligned} \tag{3.36}$$

where inequalities (a) and (b) follow from (3.34) and (G.1) respectively and the meaning subscript $\emptyset$ will become obvious from the following equation. In general the probability that there is another $(\underline{x}_1(i_1), \ldots, \underline{x}_{n_1}(i_{n_1}))$ with $i_t = 1$ for $t \in \mathcal{S} \subseteq \{1, \ldots, n_1\}$ and $i_t \neq 1$ for $t \in \mathcal{S}^c$ such that it is jointly typical with $(\underline{z}_1, \ldots, \underline{z}_{n_1})$ is upper bounded by

$$\begin{aligned}
\Pr(E_\mathcal{S}) &\overset{(a)}{\leq} \left|A_\omega^{(N_2)}\left(\tilde{\underline{P}}_{\mathcal{S}^c}, \tilde{\underline{C}}_{\mathcal{S}^c} \mid \tilde{\underline{\ell}}^{n_1}, \tilde{\underline{w}}^{n_1}, \tilde{\underline{p}}_\mathcal{S}, \tilde{\underline{c}}_\mathcal{S}\right)\right| 2^{-(n_1-|\mathcal{S}|)N_2\left(\max_{p(x)} I(X;Z)-3\beta\right)} \\
&\overset{(b)}{\leq} 2^{\frac{(n_1-|\mathcal{S}|)N_2}{2}\left(R_p+\max_{p(x)} I(X;Z)\right)+2N_2\omega}\, 2^{-(n_1-|\mathcal{S}|)N_2\left(\max_{p(x)} I(X;Z)-3\beta\right)} \\
&= 2^{-\frac{(n_1-|\mathcal{S}|)N_2}{2}\left(\max_{p(x)} I(X;Z)-R_p\right)+2N_2\omega+3n_1N_2\beta} \\
&\leq 2^{-\frac{N_2}{2}\left(\max_{p(x)} I(X;Z)-R_p\right)+2N_2\omega+3n_1N_2\beta}
\end{aligned} \tag{3.37}$$

where inequalities (a) and (b) hold due to (3.35) and (G.1) respectively and $(\tilde{\underline{p}}_t, \tilde{\underline{c}}_t)$ are set to values corresponding to message $m_t(1)$ for all $t \in \mathcal{S}$. If ω is set to $\omega = \frac{1}{16}\left(\max_{p(x)} I(X;Z) - R_p\right)$ and $\beta = \frac{1}{24n_1}\left(\max_{p(x)} I(X;Z) - R_p\right)$ then the probability that there is at least one $(\underline{x}_1(i_1), \ldots, \underline{x}_{n_1}(i_{n_1}))$

other than $(\underline{x}_1(1),\ldots,\underline{x}_{n_1}(1))$ that is jointly typical with $(\underline{z}_1(1),\ldots,\underline{z}_{n_1}(1))$ is bounded above according to equations (3.36) and (3.37) by

$$\Pr\left(E_{\mathcal{S}}\right) \leq 2^{n_1} 2^{-\frac{n_1 N_2}{4}\left(\max_{p(x)} I(X;Z)-R_p\right)} \leq 2^{-\frac{n_1 N_2}{8}\left(\max_{p(x)} I(X;Z)-R_p\right)}$$

where the term 2^{n_1} in the first inequality above is coming from the union bound over all subsets $\mathcal{S}$ and the last inequality holds for sufficiently large N_2. Since $R_p < \max_{p(x)} I(X;Z)$ then this probability drops exponentially fast to zero and therefore since the eavesdropper can decode the public bits $\underline{\tilde{\tilde{P}}}^{n_1}$ it can use the code of network **I** and decode $\underline{\tilde{\tilde{C}}}^{n_1}$. By choosing N_2 sufficiently large we can make the overall probability of error, that is either one of the messages $\underline{W}^{(i\to j)}$ is decoded erroneously at some node $j \in \mathcal{V}$ or the public bits $\underline{\tilde{\tilde{P}}}^{n_1}$, confidential bits $\underline{\tilde{\tilde{C}}}^{n_1}$ or the message $\underline{\tilde{\tilde{W}}}^{n_1}$ erroneously decode on the eavesdropper, upper bounded by $\frac{\lambda}{3}$. The code created for the stack version of network **II** with N_2 layers can be viewed as a code of blocklength $n_2 = N_2 n_1 = N_2 N_1 n$ for the non-stacked version of network **II**.

II to $\mathcal{N}$: So far we have started with a code for network $\mathcal{N}_{\bar{e}}(R_c, R_p)$ of blocklength n and rate $\tilde{\tilde{R}}$ and we constructed a code for network **II** of blocklength $n_2 = N_2 N_1 n$ and rate R and we will now prove that this is also a secure code. To do that we need to prove that the for all eavesdropping sets E not containg $\bar{e}$, *i.e.* $I(\underline{\tilde{\tilde{W}}}^{n_1}; (\underline{\tilde{\tilde{Z}}}^{(E)})^{n_1})$ is small and that $I(\underline{\tilde{\tilde{W}}}^{n_1}; (\underline{\tilde{\tilde{Z}}})^{n_1})$ is small too.

For term $I(\underline{\tilde{\tilde{W}}}^{n_1}; (\underline{\tilde{\tilde{Z}}}^{(E)})^{n_1})$, similar to (3.7) and for N_2 sufficiently large,

$$\mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{\tilde{W}}}^{n_1}; (\underline{\tilde{\tilde{Z}}}^{(E)})^{n_1}\big)\right] \leq 2 + \mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{\tilde{W}}}^{n_1}; (\underline{\tilde{\tilde{Z}}}^{(E)})^{n_1} | I=0\big)\right]$$

Note that $\mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{\tilde{W}}}^{n_1}; (\underline{\tilde{\tilde{Z}}}^{(E)})^{n_1} | I=0\big)\right] = \mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{\tilde{W}}}^{n_1}; (\underline{\tilde{\tilde{Z}}}^{(E)}_{\mathbf{I}})^{n_1} | I=0\big)\right] = \mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{\tilde{W}}}^{n_1}; (\underline{\tilde{\tilde{Z}}}^{(E)}_{\mathbf{I}})^{n_1}\big)\right]$ where $(\underline{\tilde{\tilde{Z}}}^{(E)}_{\mathbf{I}})^{n_1}$ is the eavesdropper's observation on E links of the stacked version of network **I** with N_2 layers, since $(\underline{\tilde{\tilde{Z}}}^{(E)})^{n_1}$ and $(\underline{\tilde{\tilde{Z}}}^{(E)}_{\mathbf{I}})^{n_1}$ are identical when the channel code on edge $\bar{e}$ do not fail for any time $t \in \{1,\ldots,n_1\}$, which is the condition indicated by $I=0$. Moreover the last equality holds since the event $I=0$ on network **I** is independent of the messages $\underline{\tilde{\tilde{W}}}^{n_1}$ and overheard signal $(\underline{\tilde{\tilde{Z}}}^{(E)}_{\mathbf{I}})^{n_1}$ on network **I**. Thus

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{\tilde{W}}}^{n_1}; (\underline{\tilde{\tilde{Z}}}^{(E)})^{n_1}\big)\right] &\leq 2 + \mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{\tilde{W}}}^{n_1}; (\underline{\tilde{\tilde{Z}}}^{(E)}_{\mathbf{I}})^{n_1}\big)\right] \\ &\leq 2 + N_2 \mathbb{E}_{\mathcal{C}}\left[I\big(\tilde{\tilde{W}}^{n_1}; (\tilde{\tilde{Z}}^{(E)}_{\mathbf{I}})^{n_1}\big)\right] \overset{(a)}{\leq} 2 + N_2 n_1 \varepsilon \overset{(b)}{\leq} n_1 N_2 \varepsilon \end{aligned} \tag{3.38}$$

where (a) follows from (3.26), and (b) holds for sufficiently large N_2.

For term $I(\underline{\tilde{\tilde{W}}}^{n_1}; (\underline{\tilde{\tilde{Z}}})^{n_1})$, we first need to upper bound $H(\underline{\tilde{\tilde{P}}}^{n_1}, \underline{\tilde{\tilde{C}}}^{n_1}, \underline{\tilde{\tilde{W}}}^{n_1} | \underline{\tilde{\tilde{Z}}}^{n_1}, \underline{\tilde{\tilde{L}}}^{n_1}_{\bar{e}}, \underline{\tilde{\tilde{W}}}^{n_1})$. Since the public bits $\underline{\tilde{\tilde{P}}}^{n_1}$ and the confidential bits $\underline{\tilde{\tilde{C}}}^{n_1}$ can be decoded with probability of error at most

λ then due to Fano's inequality [60, Theorem 2.10.1]

$$\mathbb{E}_{\mathcal{C}}\left[H(\underline{\tilde{\tilde{P}}}^{n_1},\underline{\tilde{\tilde{C}}}^{n_1},\underline{\tilde{\tilde{W}}}^{n_1}|\underline{\tilde{\tilde{Z}}}^{n_1},\underline{\tilde{\tilde{L}}}_{\bar{e}}^{n_1},\underline{\tilde{\tilde{W}}}^{n_1})\right]=\mathbb{E}_{\mathcal{C}}\left[H(\underline{\tilde{\tilde{P}}}^{n_1},\underline{\tilde{\tilde{C}}}^{n_1}|\underline{\tilde{\tilde{Z}}}^{n_1},\underline{\tilde{\tilde{L}}}_{\bar{e}}^{n_1},\underline{\tilde{\tilde{W}}}^{n_1})\right]\leq h(\lambda)+n_2(R_p+R_c)\lambda \tag{3.39}$$

From the definition of mutual information we get

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}}\left[I(\underline{\tilde{\tilde{P}}}^{n_1},\underline{\tilde{\tilde{C}}}^{n_1},\underline{\tilde{\tilde{W}}}^{n_1};\underline{\tilde{\tilde{Z}}}^{n_1},\underline{\tilde{\tilde{L}}}_{\bar{e}}^{n_1},\underline{\tilde{\tilde{W}}}^{n_1})\right]\\
&=\mathbb{E}_{\mathcal{C}}\left[H(\underline{\tilde{\tilde{P}}}^{n_1},\underline{\tilde{\tilde{C}}}^{n_1},\underline{\tilde{\tilde{W}}}^{n_1})\right]-\mathbb{E}_{\mathcal{C}}\left[H(\underline{\tilde{\tilde{P}}}^{n_1},\underline{\tilde{\tilde{C}}}^{n_1},\underline{\tilde{\tilde{W}}}^{n_1}|\underline{\tilde{\tilde{Z}}}^{n_1},\underline{\tilde{\tilde{L}}}_{\bar{e}}^{n_1},\underline{\tilde{\tilde{W}}}^{n_1})\right]\\
&\overset{(a)}{\geq}\mathbb{E}_{\mathcal{C}}\left[H(\underline{\tilde{\tilde{P}}}^{n_1},\underline{\tilde{\tilde{C}}}^{n_1},\underline{\tilde{\tilde{W}}}^{n_1})\right]-h(\lambda)-n_2(R_p+R_c)\lambda\\
&\overset{(b)}{\geq}\mathbb{E}_{\mathcal{C}}\left[H(\underline{\tilde{\tilde{P}}}^{n_1},\underline{\tilde{\tilde{W}}}^{n_1})\right]+\mathbb{E}_{\mathcal{C}}\left[H(\underline{\tilde{\tilde{C}}}^{n_1}|\underline{\tilde{\tilde{P}}}^{n_1},\underline{\tilde{\tilde{W}}}^{n_1})\right]-n_2\varepsilon\\
&\overset{(c)}{=}N_2N_1\mathbb{E}_{\mathcal{C}}\left[H(\tilde{\tilde{P}}^{n},\tilde{\tilde{W}})\right]+N_2N_1\mathbb{E}_{\mathcal{C}}\left[H(\tilde{\tilde{C}}^{n}|\tilde{\tilde{P}}^{n},\tilde{\tilde{W}})\right]-n_2\varepsilon\\
&\overset{(d)}{=}N_2N_1\mathbb{E}_{\mathcal{C}}\left[H(\tilde{\tilde{P}}^{n},\tilde{\tilde{W}})\right]+nN_2N_1(C_{\bar{e}}-\varepsilon)-n_2\varepsilon\\
&=N_2N_1\left(\mathbb{E}_{\mathcal{C}}\left[H(\tilde{\tilde{W}})\right]+\mathbb{E}_{\mathcal{C}}\left[H(\tilde{\tilde{P}}^{n}|\tilde{\tilde{W}})\right]\right)+nN_2N_1(C_{\bar{e}}-2\varepsilon)\\
&\overset{(e)}{\geq}N_2N_1\left(\mathbb{E}_{\mathcal{C}}\left[H(\tilde{\tilde{W}})\right]+\mathbb{E}_{\mathcal{C}}\left[H(\tilde{\tilde{P}}^{n})\right]\right)+nN_2N_1(C_{\bar{e}}-3\varepsilon)
\end{aligned}\tag{3.40}$$

where inequality (a) is Fano's inequality expressed in (3.39), inequality (b) holds since we choose λ small enough so that $h(\lambda)+n_1(R_p+R_c)\lambda<n_2\varepsilon$, equality (c) holds since the information bits on the different layers of the stacked network are independent, (d) follows from equation (3.22), and (e) holds since the code of blocklength n and rate $\tilde{\tilde{R}}$ is secure and therefore

$$\begin{aligned}
n\epsilon&\geq\mathbb{E}_{\mathcal{C}}\left[I(\tilde{\tilde{W}};\tilde{\tilde{P}}^{n})\right]\\
\Leftrightarrow n\epsilon&\geq\mathbb{E}_{\mathcal{C}}\left[H(\tilde{\tilde{P}}^{n})\right]-\mathbb{E}_{\mathcal{C}}\left[H(\tilde{\tilde{P}}^{n}|\tilde{\tilde{W}})\right]\\
\Leftrightarrow\mathbb{E}_{\mathcal{C}}\left[H(\tilde{\tilde{P}}^{n}|\tilde{\tilde{W}})\right]&\geq\mathbb{E}_{\mathcal{C}}\left[H(\tilde{\tilde{P}}^{n})\right]-n\varepsilon
\end{aligned}\tag{3.41}$$

Moreover by the chain rule we get

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}}\left[I(\underline{\tilde{\tilde{P}}}^{n_1},\underline{\tilde{\tilde{C}}}^{n_1},\underline{\tilde{\tilde{W}}}^{n_1};\underline{\tilde{\tilde{Z}}}^{n_1},\underline{\tilde{\tilde{L}}}_{\bar{e}}^{n_1},\underline{\tilde{\tilde{W}}}^{n_1})\right]\\
&=\mathbb{E}_{\mathcal{C}}\left[I(\underline{\tilde{\tilde{W}}}^{n_1};\underline{\tilde{\tilde{P}}}^{n_1},\underline{\tilde{\tilde{C}}}^{n_1},\underline{\tilde{\tilde{W}}}^{n_1})\right]+\mathbb{E}_{\mathcal{C}}\left[I(\underline{\tilde{\tilde{L}}}_{\bar{e}}^{n_1};\underline{\tilde{\tilde{P}}}^{n_1},\underline{\tilde{\tilde{C}}}^{n_1},\underline{\tilde{\tilde{W}}}^{n_1}|\underline{\tilde{\tilde{W}}}^{n_1})\right]+\mathbb{E}_{\mathcal{C}}\left[I(\underline{\tilde{\tilde{Z}}}^{n_1};\underline{\tilde{\tilde{P}}}^{n_1},\underline{\tilde{\tilde{C}}}^{n_1},\underline{\tilde{\tilde{W}}}^{n_1}|\underline{\tilde{\tilde{W}}}^{n_1},\underline{\tilde{\tilde{L}}}_{\bar{e}}^{n_1})\right]\\
&\leq\mathbb{E}_{\mathcal{C}}\left[H(\underline{\tilde{\tilde{W}}}^{n_1})\right]+\mathbb{E}_{\mathcal{C}}\left[H(\underline{\tilde{\tilde{L}}}_{\bar{e}}^{n_1})\right]+\mathbb{E}_{\mathcal{C}}\left[I(\underline{\tilde{\tilde{Z}}}^{n_1};\underline{\tilde{\tilde{P}}}^{n_1},\underline{\tilde{\tilde{C}}}^{n_1}|\underline{\tilde{\tilde{W}}}^{n_1},\underline{\tilde{\tilde{L}}}_{\bar{e}}^{n_1})\right]\\
&=N_2N_1\mathbb{E}_{\mathcal{C}}\left[H(\tilde{\tilde{W}})\right]+n_1N_2N_1C_{\bar{e}}+\mathbb{E}_{\mathcal{C}}\left[I(\underline{\tilde{\tilde{Z}}}^{n_1};\underline{\tilde{\tilde{P}}}^{n_1},\underline{\tilde{\tilde{C}}}^{n_1}|\underline{\tilde{\tilde{W}}}^{n_1},\underline{\tilde{\tilde{L}}}_{\bar{e}}^{n_1})\right]
\end{aligned}\tag{3.42}$$

Therefore $N_2N_1\left(\mathbb{E}_{\mathcal{C}}\left[H(\tilde{\tilde{W}})\right]+\mathbb{E}_{\mathcal{C}}\left[H(\tilde{\tilde{P}}^{n})\right]\right)+nN_2N_1(C_{\bar{e}}-3\varepsilon)\leq N_2N_1\mathbb{E}_{\mathcal{C}}\left[H(\tilde{\tilde{W}})\right]+n_1N_2N_1C_{\bar{e}}+$

$\mathbb{E}_{\mathcal{C}}\left[I(\tilde{\underline{Z}}^{n_1};\tilde{\underline{P}}^{n_1},\tilde{\underline{C}}^{n_1}|\tilde{\underline{W}}^{n_1},\tilde{\underline{L}}_{\bar{e}}^{n_1})\right]$ or equivalently

$$N_2N_1\mathbb{E}_{\mathcal{C}}\left[H(\tilde{\underline{P}}^{n})\right] \leq \mathbb{E}_{\mathcal{C}}\left[I(\tilde{\underline{Z}}^{n_1};\tilde{\underline{P}}^{n_1},\tilde{\underline{C}}^{n_1}|\tilde{\underline{W}}^{n_1},\tilde{\underline{L}}_{\bar{e}}^{n_1})\right] + 3nN_2N_1\varepsilon \tag{3.43}$$

Moreover since $\tilde{\underline{W}}^{n_1},\tilde{\underline{L}}_{\bar{e}}^{n_1},\tilde{\underline{P}}^{n_1},\tilde{\underline{C}}^{n_1} \to \tilde{\underline{X}}^{n_1} \to \tilde{\underline{Z}}^{n_1}$ we get from the data processing inequality that $\mathbb{E}_{\mathcal{C}}\left[I(\tilde{\underline{W}}^{n_1},\tilde{\underline{L}}_{\bar{e}}^{n_1},\tilde{\underline{P}}^{n_1},\tilde{\underline{C}}^{n_1};\tilde{\underline{Z}}^{n_1})\right] \leq \mathbb{E}_{\mathcal{C}}\left[I(\tilde{\underline{X}}^{n_1};\tilde{\underline{Z}}^{n_1})\right]$. By using the chain rule in the left hand side of the inequality we get

$$\mathbb{E}_{\mathcal{C}}\left[I(\tilde{\underline{W}}^{n_1};\tilde{\underline{Z}}^{n_1})\right] + \mathbb{E}_{\mathcal{C}}\left[I(\tilde{\underline{L}}_{\bar{e}}^{n_1};\tilde{\underline{Z}}^{n_1}|\tilde{\underline{W}}^{n_1})\right] + \mathbb{E}_{\mathcal{C}}\left[I(\tilde{\underline{P}}^{n_1},\tilde{\underline{C}}^{n_1};\tilde{\underline{Z}}^{n_1}|\tilde{\underline{W}}^{n_1},\tilde{\underline{L}}_{\bar{e}}^{n_1})\right] \leq \mathbb{E}_{\mathcal{C}}\left[I(\tilde{\underline{X}}^{n_1};\tilde{\underline{Z}}^{n_1})\right]$$
$$\leq nN_2N_1\max_{p(x)} I(X;Z)$$

or else $\mathbb{E}_{\mathcal{C}}\left[I(\tilde{\underline{W}}^{n_1};\tilde{\underline{Z}}^{n_1})\right]$ is upper bounded by

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\left[I(\tilde{\underline{W}}^{n_1};\tilde{\underline{Z}}^{n_1})\right] &\leq nN_2N_1\max_{p(x)} I(X;Z) - \mathbb{E}_{\mathcal{C}}\left[I(\tilde{\underline{P}}^{n_1},\tilde{\underline{C}}^{n_1};\tilde{\underline{Z}}^{n_1}|\tilde{\underline{W}}^{n_1},\tilde{\underline{L}}_{\bar{e}}^{n_1})\right] \\ &\overset{(a)}{\leq} nN_2N_1\max_{p(x)} I(X;Z) - N_2N_1\mathbb{E}_{\mathcal{C}}\left[H(\tilde{\underline{P}}^{n})\right] + 3nN_2N_1\varepsilon \end{aligned} \tag{3.44}$$

where inequality (a) follows from (3.43). By combining inequalities (3.20) and (3.21) we get

$$H(\tilde{\underline{P}}^{n}) \geq n\left(\max_{p(x)} I(X;Z) - \varepsilon\right)$$

and therefore by combining the inequality above with (3.44) we get

$$\mathbb{E}_{\mathcal{C}}\left[I(\tilde{\underline{W}}^{n_1};\tilde{\underline{Z}}^{n_1})\right] \leq 4nN_2N_1\varepsilon$$

and since $\underline{W}^{n_1} \to \tilde{\underline{W}}^{n_1} \to \tilde{\underline{Z}}^{n_1}$, where $\underline{W}^{n_1}$ si the message of rate R. Due to the data processing inequality

$$\mathbb{E}_{\mathcal{C}}\left[I(\underline{W}^{n_1};\tilde{\underline{Z}}^{n_1})\right] \leq \mathbb{E}_{\mathcal{C}}\left[I(\tilde{\underline{W}}^{n_1};\tilde{\underline{Z}}^{n_1})\right] \leq 4nN_2N_1\varepsilon. \tag{3.45}$$

From inequalities (3.38), (3.45) and since $n_1 = nN_1$ we have created a random code that for any $E \in A$

$$\mathbb{E}_{\mathcal{C}}\left[I\big(\tilde{\underline{W}}^{n_1};(\tilde{\underline{Z}}^{(E)})^{n_1}\big)\right] \leq 4nN_1N_2\varepsilon \tag{3.46}$$

and probability of error less than $\frac{\lambda}{3}$. One can follow an analysis identical to the one used in equation (3.1) of Theorem 8 to prove that there is at least one deterministic code $(\lambda, 12\varepsilon, R)$–$\mathcal{S}(\mathcal{N}, A)$ for network $\mathcal{N}$. □

Lemma 5. *The secrecy rate region* $\mathcal{R}\big(\mathcal{N}_{\bar{e}}(R_c, R_p), A\big)$ *is continuous in* (R_c, R_p) *for all* $R_c > 0$, $R_p > 0$ *and* $A \in \mathcal{P}(\mathcal{E})$.

Proof. The proof is almost identical to the proof of Lemma 2 of [31]. The only difference is that now the codes are secure codes, and we need to check the security constraint for the code that results from the prior code construction. Define for any $R_c > 0$, $R_p > 0$, $\mu_c < R_c$ and $\mu_p < R_p$

$$\zeta(\mu_c, \mu_p) \stackrel{\text{def}}{=} \max_{\hat{R} \in \mathcal{R}(\mathcal{N}_{\bar{e}}(R_c+\mu_c, R_p+\mu_p), A)} \min_{\check{R} \in \mathcal{R}(\mathcal{N}_{\bar{e}}(R_c-\mu_c, R_p-\mu_p), A)} ||\hat{R} - \check{R}||_\infty,$$

which is the worst-case ℓ_∞-norm between any point in $\mathcal{R}\big(\underline{\mathcal{N}}_{\bar{e}}(R_c + \mu_c, R_p + \mu_p), A\big)$ and its closest point in $\mathcal{R}\big(\underline{\mathcal{N}}_{\bar{e}}(R_c - \mu_c, R_p - \mu_p), A\big)$. To prove continuity, we show that for any $\psi > 0$, there exists a $\mu_c > 0$ and $\mu_p > 0$ such that $\zeta(\mu_c, \mu_p) \leq \psi$.

Fix any $\mu_c > 0$, $\mu_p > 0$ and $\hat{R} \in \mathcal{R}\big(\underline{\mathcal{N}}_{\bar{e}}(R_c + \mu_c, R_p + \mu_p), A\big)$. Then for any λ and $\hat{N}$ sufficiently large there is a solution $(\lambda, \varepsilon, \hat{R})$–$\mathcal{S}\big(\underline{\mathcal{N}}_{\bar{e}}(R_c + \mu_c, R_p + \mu_p), A\big)$ for network $\underline{\mathcal{N}}_{\bar{e}}(R_c + \mu_c, R_p + \mu_p)$. The same solution can be applied to the stacked network $\underline{\mathcal{N}}_{\bar{e}}(R_c - \mu_c, R_p - \mu_p)$ with $\check{N}$ number of layers in the stack as long as number $\check{N}$ is chosen large enough so that

$$\check{N}(R_c - \mu_c) \geq \hat{N}(R_c + \mu_c)$$
$$\check{N}(R_p - \mu_p) \geq \hat{N}(R_p + \mu_p).$$

Indeed this can be accomplished by operating the solution $\mathcal{S}\big(\underline{\mathcal{N}}_{\bar{e}}(R_c + \mu_c, R_p + \mu_p), A\big)$ unchanged across the first $\hat{N}$ copies of the stacked network $\underline{\mathcal{N}}_{\bar{e}}(R_c - \mu_c, R_p - \mu_p)$ since networks $\underline{\mathcal{N}}_{\bar{e}}(R_c + \mu_c, R_p + \mu_p)$ and $\underline{\mathcal{N}}_{\bar{e}}(R_c - \mu_c, R_p - \mu_p)$ are identical apart from edge $\bar{e}$, and sending the $\hat{N}(R_c + \mu_c)$ and $\hat{N}(R_p + \mu_p)$ bits intended for transmission across the $\hat{N}$ bit pipes of rate $R_c + \mu_c$ and $R_p + \mu_p$ respectively of edge $\bar{e}$ across the $\check{N}$ copies of the rate $R_c - \mu_c$ and $R_c + \mu_c$ respectively of bit pipe $\bar{e}$ in network $\underline{\mathcal{N}}_{\bar{e}}(R_c - \mu_c, R_p - \mu_p)$. Setting

$$\check{N} = \left\lceil \hat{N} \max\left(\frac{R_c + \mu_c}{R_c - \mu_c}, \frac{R_p + \mu_p}{R_p - \mu_p}\right) \right\rceil \tag{3.47}$$

and

$$q = \arg\max\left(\frac{R_c + \mu_c}{R_c - \mu_c}, \frac{R_p + \mu_p}{R_p - \mu_p}\right).$$

The rate $\check{R}$ for the resulting code in network $\underline{\mathcal{N}}_{\bar{e}}(R_c - \mu_c, R_p - \mu_p)$ is

$$\check{R} = \frac{\hat{R}\hat{N}}{\check{N}} \geq \hat{R}\frac{\hat{N}(R_q - \mu_q)}{\hat{N}(R_q + \mu_q) + R_q - \mu_q}$$

which approaches $\hat{R}$ as $\hat{N}$ grows without bound. The new code is secure since $I\big((\underline{\check{Z}}^E)^n; \underline{W}\big) =$

$I\big((\underline{\hat{Z}}^E)^n;\underline{W}\big) < n\hat{N}\varepsilon < n\check{N}\varepsilon$ for every $E \in A$. □

Now we have all the tools necessary to prove Theorem 10 saying that one can change a link $\bar{e}$, that is either not eavesdropped or if eavesdropped it is not wiretapped simultaneously with any other link, with a noiseless degraded broadcast channel without affecting the secrecy rate region of the initial network. Applying Theorem 10 for every channel in a network where $|E| = 1$ for every $E \in A$, proves the optimality of the separation between channel network coding over unreliable channels in secure communication networks.

Proof of Theorem 10. From Lemma 4 and Theorem 9 we get $\mathcal{R}(\mathcal{N}_{\bar{e}}(R_c-\mu, R_p-\mu), A) \subseteq \mathcal{R}(\mathcal{N}, A) \subseteq \mathcal{R}(\mathcal{N}_{\bar{e}}(R_c+\mu, R_p+\mu), A)\ \ \forall \mu > 0$. Therefore by taking μ to zero and using the continuity proved in Lemma 5 we have $\mathcal{R}(\mathcal{N}_{\bar{e}}(R_c, R_p), A) = \mathcal{R}(\mathcal{N}, A)$ where $R_c = \max\limits_{p(x)} I(X^{(\bar{e})}; Y^{(\bar{e})}) - \max\limits_{p(x)} I(X^{(\bar{e})}; Z^{(\bar{e})})$ and $R_p = \max\limits_{p(x)} I(X^{(\bar{e})}; Z^{(\bar{e})})$. □

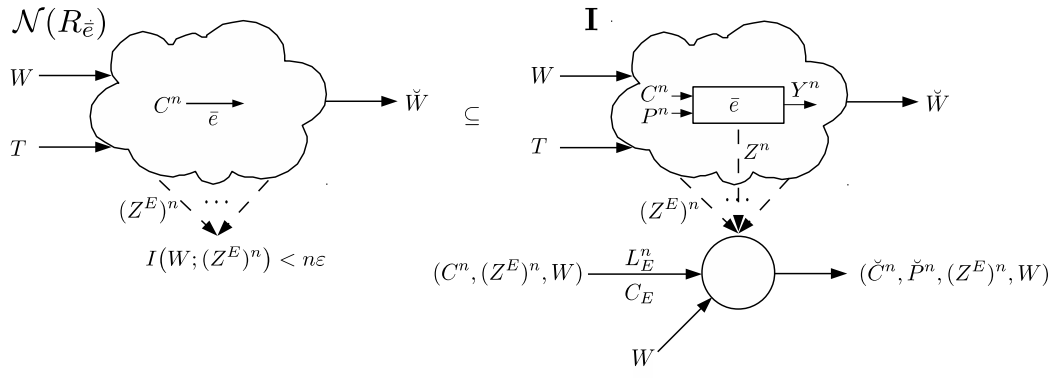


Figure 3.5: The "equivalent" secret and communication networks

3.4.3 Proof of Theorem 11

Proof of Theorem 11. The idea of the proof is very similar to the proof of Lemma 4. The essence of the proof is that we will start with a secure code for network $\mathcal{N}_{\bar{e}}(R_c, 0)$ and we will construct a secure code for network $\mathcal{N}$. Specifically we will show how to achieve any point R inside the secrecy rate region of network $\mathcal{N}_{\bar{e}}(R_c, 0)$ at network $\mathcal{N}$. The two networks $\mathcal{N}_{\bar{e}}(R_c, 0)$ and $\mathcal{N}$ are identical except from edge $\bar{e}$ where at network $\mathcal{N}_{\bar{e}}(R_c, 0)$ edge $\bar{e}$ is a noiseless bit pipe that can carry up to R_c bits per channel use whereas edge $\bar{e}$ in network $\mathcal{N}$ is a noisy degraded broadcast channel with the property $R_c < \max_{p(x)} I(X,Y) - \max_{p(x)} I(X,Z)$. This property is essential so that one can apply a secure channel such as the one derived by Wyner [17] and carry the information traveling edge $\bar{e}$ securely.

Specifically the information bits C^n transversing edge $\bar{e}$ of network $\mathcal{N}_{\bar{e}}(R_c, 0)$ will be mixed with some random bits P^n in network $\mathcal{N}$ to ensure the security of the code for network $\mathcal{N}$. To prove that an eavesdropper gets a very small mutual information with the message by choosing any eavesdropping set $E \in A$ we would make use of an auxiliary network $\mathbf{I}$ shown in Figure 3.5. Network $\mathbf{I}$ is similar to network $\mathcal{N}$ with the addition that for every eavesdropping set E where $\bar{e} \in E$ we add a receiver that has access to the whole message W, the eavesdropper's observation $(Z^{E\setminus\{\bar{e}\}})^n$, the degraded output Z^n from the channel in edge $\bar{e}$ and a noiseless bit of capacity C_E carrying bits (the value of C_E is defined in equation (3.49)) from a "super source" that has access to (C^n, P^n) the information transversing edge $\bar{e}$ at network $\mathcal{N}$, message W and the eavesdropper's observation $(Z^{E\setminus\{\bar{e}\}})^n$. The additional receiver corresponding to eavesdropping set E demands message W, the eavesdropping observations $(Z^{E\setminus\{\bar{e}\}})^n$ and the information (P^n, C^n) input to edge $\bar{e}$. These additional receivers assist in the proof of the secrecy of the code for those eavesdropping set $E \in A$ such that $\bar{e} \in E$ in the following manner: the sum of capacities of $(W, (Z^{E\setminus\{\bar{e}\}})^n, Z^n, L_E^n)$ (where L_E^n are the bits in the noiseless bit pipe of capacity C_E) that are all the incoming links to the auxiliary receivers is almost equal to the entropy of $(P^n, C^n, W, (Z^{E\setminus\{\bar{e}\}})^n)$ that correspond to the decoded message at the auxiliary receivers and therefore all links are filled up to capacity. Therefore there is no spare

capacity at links $((Z^{E\setminus\{\bar{e}\}})^n, Z^n)$ to carry any information about message W and therefore the code is secure. The proof of secrecy for eavesdropping sets $E \in A$ such that $\bar{e} \notin E$ is derived identical to equation (3.38) of the proof of Lemma4 and therefore is skipped. The details of the proof for those $E \in A$ such that $\bar{e} \in E$ follow below.

Fix any rate R inside the secrecy rate region of network $\mathcal{N}_{\bar{e}}(R_c, 0)$, *i.e.* $R \in \text{int}(\mathcal{R}(\mathcal{N}(R_{\bar{e}}), A))$ and choose some $\tilde{R} \in \text{int}(\mathcal{R}(\mathcal{N}(R_{\bar{e}}), A))$ such that $\tilde{R}^{(i\to\mathcal{B})} > R^{(i\to\mathcal{B})}$ for all $i \in \mathcal{V}$ and $\mathcal{B} \in \mathcal{B}^{(i)}$ with $R^{(i\to\mathcal{B})} > 0$. Then for any $\lambda > 0$ and $\varepsilon > 0$ there is a solution of blocklength n and rate $\tilde{R}$ for network $\mathcal{N}_{\bar{e}}(R_c, 0)$ such that $\Pr\left(\tilde{\tilde{W}}^{(j\to\mathcal{K},i)} \neq \tilde{W}^{(j\to\mathcal{K})}\right) < \lambda$ for all $(j, \mathcal{K})$ with $\mathcal{K} \in \mathcal{B}^{(j)}$, $i \in \mathcal{K}$ and $R^{(j\to\mathcal{K})} > 0$, and $I\big((\tilde{Z}^E)^n; \tilde{W}\big) < n\varepsilon$ for all $E \in A$, where the tilde on the eavesdropper's observation $(\underset{\sim}{\tilde{Z}}^E)^n$ and the message $\tilde{W}$ refers to the fact that the code operates at rate $\tilde{R}$. By carefully choosing parameter λ it was shown in the proof of Theorem 8 one can create a $(2^{-N\delta}, \varepsilon, R)$–$\mathcal{S}(\mathcal{N}_{\bar{e}}(R_c, 0), A)$ solution for the stacked network $\underline{\mathcal{N}}_{\bar{e}}(R_c, 0)$. One can use code $(2^{-N\delta}, \varepsilon, R)$–$\mathcal{S}(\underline{\mathcal{N}}_{\bar{e}}(R_c, 0), A)$ along with some channel encoder/decoder $\{a_{Nt}, b_{Nt}\}_{t=1}^N$ pair and find a secure solution for the stacked network $\mathcal{N}$ where the noiseless bit pipe of edge $\bar{e}$ have been replaced by a noisy degraded broadcast channel. The channel code works as follows: At each time step $t \in \{1, \ldots, n\}$ once the bits $\underline{\tilde{C}}_t$ on edge $\bar{e}$ of network $\mathcal{N}(R_{\bar{e}})$ have been received they are transmitted through the noisy channel along with $N(\max_{p(x)} I(X; Z) - \varepsilon)$ random bits denoted by $\underline{\tilde{P}}_t$. The random bits $\underline{\tilde{P}}_t$ are generated independently across different times t and independent of the message W and the noise randomness. Therefore the entropy $H\big(\underline{\tilde{P}}^n\big)$ of the random bits across the n time steps is equal to

$$H\big(\underline{\tilde{P}}^n\big) = nN\left(\max_{p(x)} I(X; Z) - \varepsilon\right) \tag{3.48}$$

and the rate R_p of the random bits equal to $R_p = \max_{p(x)} I(X; Z) - \varepsilon$.

Therefore at each time step t there is a total of $N(R_c + R_p)$ bits delivered at the input of edge $\bar{e}$ and have to be conveyed through the degraded broadcast channel that replaced the noiseless bit pipe of rate R_c of network $\mathcal{N}_{\bar{e}}(R_c, 0)$. The $N(R_c + R_p)$ bits at each time step correspond to $2^{N(R_c+R_p)}$ incoming messages $m_t(i)$, $i \in \big\{1, \ldots, 2^{N(R_c+R_p)}\big\}$ at edge $\bar{e}$. The channel encoders $a_{N,t}$ at each time step t have assigned to each one of the $2^{N(R_c+R_p)}$ incoming messages $m_t(i)$ a random N–tuple $\underline{\tilde{X}}_t(i) = \Big(\tilde{X}_{t1}(i), \ldots, \tilde{X}_{tN}(i)\Big)$ where $\tilde{X}_{tj}(i)$, $j \in \{1, \ldots, N\}$ are chosen from the distribution $p(x)$ that gives rise to the corresponding mutual informations $\max_{p(x)} I(X; Y)$ and $\max_{p(x)} I(X; Z)$ of the noisy degraded broadcast channel $\bar{e}$. This mapping is revealed to both the eavesdropper and $V_2(\bar{e})$ that is the output of $\bar{e}$. Once the $N(R_c + R_p)$ public and confidential bits at time t are ready for transmission then the corresponding N–tuple is transmitted through the N noisy channels across the N layers and the intended receiver gets $\underline{\tilde{Y}}_t$. From the received N–tuple $\underline{\tilde{Y}}_t$ the decoder finds the N–tuple $\underline{\tilde{X}}_t(i)$ corresponding to the transmitted message $m_t(i)$ so that $(\underline{\tilde{X}}_t(i), \underline{\tilde{Y}}_t)$ are jointly typical, *i.e.* find message $m_t(i)$ such that $(\underline{\tilde{X}}_t(i), \underline{\tilde{Y}}_t) \in A_{t,\gamma}^{(N)}(\tilde{X}, \tilde{Y})$ where the definition

$A_{t,\gamma}^{(N)}(\tilde{X},\tilde{Y})$ is given in (3.27). As shown in equation (3.29) one can choose γ appropriately so that the probability that the channel code fails drops exponentially fast. At a later point in the proof it is going to be useful to mention that if one has access to the bits $\underline{\tilde{C}}_t$ transversing edge $\bar{e}$ and the degraded output $\underline{\tilde{Z}}_t$ of edge $\bar{e}$ then one can decode the random bits $\underline{\tilde{P}}_t$. Indeed from the received N–tuple $\underline{\tilde{Z}}_t$ one can find N–tuple $\underline{\tilde{X}}_t(i)$ corresponding to the transmitted message $m_t(i)$ so that $(\underline{\tilde{X}}_t(i),\underline{\tilde{Z}}_t)$ are jointly typical, *i.e.* find message $m_t(i)$ such that $(\underline{\tilde{X}}_t(i),\underline{\tilde{Z}}_t)\in A_{t,\beta}^{(N)}(\tilde{X},\tilde{Z})$ where the definition $A_{t,\beta}^{(N)}(\tilde{X},\tilde{Z})$ is given in (3.33). The probability that there are more than one sequences $\underline{\tilde{X}}_t(j)$, $\underline{\tilde{X}}_t(k)$ with $j\neq k$ that are jointly typical with the received sequence $\underline{\tilde{Z}}_t$ is upper bounded by $2^{NR_p}2^{-N(\max_{p(x)} I(X;Z)-3\beta)}$ and by choosing $\beta=\frac{1}{6}\max_{p(x)} I(X;Z)$ makes the probability of error to drop exponentially fast with the number of layers N.

By setting the capacity C_E of the side channel to each auxiliary receiver equal to

$$C_E=\frac{1}{n}H\big(\tilde{C}^n|(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n,\tilde{W}\big)+\varepsilon. \tag{3.49}$$

then the side channel has the necessary capacity to transfer enough bits so that the auxiliary receiver is able to decode bits $\underline{\tilde{C}}^n$. Indeed since each auxiliary receiver has access to message $\underline{\tilde{W}}$ and eavesdropper's observation $(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n$ by defining the notion of typical set for tuple $(\underline{\tilde{w}},(\underline{\tilde{z}}^{E\setminus\{\bar{e}\}})^n)$ as

$$\begin{aligned}A_{\nu}^{(N)}\big(\tilde{W},(\tilde{Z}^{E\setminus\{\bar{e}\}})^n\big)=\ \Big\{(\underline{\tilde{w}},(\underline{\tilde{z}}^{E\setminus\{\bar{e}\}})^n):\ &\Big|-\frac{1}{N}\log p(\underline{\tilde{w}})-H(\tilde{W})\Big|\leq\nu,\\ &\Big|-\frac{1}{N}\log p((\underline{\tilde{z}}^{E\setminus\{\bar{e}\}})^n)-H\big((\tilde{Z}^{E\setminus\{\bar{e}\}})^n\big)\Big|\leq\nu,\\ &\Big|-\frac{1}{N}\log p(\underline{\tilde{w}},(\underline{\tilde{z}}^{E\setminus\{\bar{e}\}})^n)-H\big(\tilde{W},(\tilde{Z}^{E\setminus\{\bar{e}\}})^n\big)\Big|\leq\nu\Big\}.\end{aligned}$$

and the notion of conditionally typical $\underline{\tilde{c}}^n$ given a typical $(\underline{\tilde{w}},(\underline{\tilde{z}}^{E\setminus\{\bar{e}\}})^n)$ as

$$A_{\nu}^{(N)}(\tilde{C}^n|\underline{\tilde{w}},(\underline{\tilde{z}}^{E\setminus\{\bar{e}\}})^n)=\ \Big\{\underline{\tilde{c}}^n:\Big|-\frac{1}{N}\log p(\underline{\tilde{c}}^n,\underline{\tilde{w}},(\underline{\tilde{z}}^{E\setminus\{\bar{e}\}})^n)-H\big(\tilde{C}^n,\tilde{W},(\tilde{Z}^{E\setminus\{\bar{e}\}})^n\big)\Big|\leq\nu\Big\}.$$

one can prove similarly to (3.25) that the size of the conditionally typical set $A_{\nu}^{(N)}(\tilde{C}^n|\underline{\tilde{w}},(\underline{\tilde{z}}^{E\setminus\{\bar{e}\}})^n)$ is upper bounded by

$$\Big|A_{\nu}^{(N)}(\tilde{C}^n|\underline{\tilde{w}},(\underline{\tilde{z}}^{E\setminus\{\bar{e}\}})^n)\Big|\leq 2^{N[H(\tilde{C}^n|\tilde{W},(\tilde{Z}^{E\setminus\{\bar{e}\}})^n)+2\nu]}.$$

By using the Chernoff bound as in Lemma 8 of [31] it can be proved that the probability of observing an atypical $\underline{\tilde{w}},(\underline{\tilde{z}}^{E\setminus\{\bar{e}\}})^n$ or a conditionally atypical $\underline{\tilde{c}}^n$ for a typical $(\underline{\tilde{w}},(\underline{\tilde{z}}^{E\setminus\{\bar{e}\}})^n)$ drops exponentially fast. Similar to the proof of Lemma 4 we only consider the case of typical $\underline{\tilde{c}}^n$ given a typical $(\underline{\tilde{w}},(\underline{\tilde{z}}^{E\setminus\{\bar{e}\}})^n)$ since the observing an atypical tuple has exponential small probability and will be

regarded as an error event. By setting $\nu = \varepsilon/4$ the error free channel has enough capacity to transfer all $N[H(\tilde{C}^n|\tilde{W}, (\tilde{Z}^{E\setminus\{\bar{e}\}})^n) + 2\nu]$ bits needed to describe all typical $\underline{\tilde{c}}^n$ during the nN channel uses of the noiseless bit pipe and therefore allowing the auxiliary receiver to decode $\underline{\tilde{c}}^n$. So far we have started with a rate $\tilde{R}$ secure code for network $\mathcal{N}_{\bar{e}}(R_c, 0)$ and constructed code for a stacked version of network **I** in Figure 3.5. The sources of error for the constructed code of network **I** are when the code for the stack network $\underline{\mathcal{N}}_{\bar{e}}(R_c, 0)$ fails or when the channel code at edge $\bar{e}$ fails or when tuple $(\underline{\tilde{w}}, (\underline{\tilde{z}}^{E\setminus\{\bar{e}\}})^n)$ or $\underline{\tilde{c}}^n$ are atypical. The overall probability of decoding error for the stacked version of network **I** can be made less than value λ by choosing the number of layers N of the stack large enough.

In order to prove the security of the constructed code for network $\underline{\mathcal{N}}$ we first need to upper bound $H\big(\underline{\tilde{P}}^n, \underline{\tilde{C}}^n, (\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n, \underline{\tilde{W}}|\underline{\tilde{Z}}^n, (\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n, \underline{\tilde{L}}_E^n, \underline{\tilde{W}}\big)$. Since the message $\underline{\tilde{W}}$, the public bits $\underline{\tilde{P}}^n$, the confidential bits $\underline{\tilde{C}}^n$ and the eavesdropper's observation $(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n$ can be decoded with probability of error at most λ then due to Fano's inequality [60]

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\left[H\big(\underline{\tilde{P}}^n, \underline{\tilde{C}}^n, (\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n, \underline{\tilde{W}}|\underline{\tilde{Z}}^n, (\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n, \underline{\tilde{L}}_E^n, \underline{\tilde{W}}\big)\right] &= \mathbb{E}_{\mathcal{C}}\left[H\big(\underline{\tilde{P}}^n, \underline{\tilde{C}}^n|\underline{\tilde{Z}}^n, (\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n, \underline{\tilde{L}}_E^n, \underline{\tilde{W}}\big)\right] \\ &\leq h(\lambda) + nN(R_c + R_p)\lambda \end{aligned} \tag{3.50}$$

From the definition of mutual information we get

$$\begin{aligned}&\mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{P}}^n, \underline{\tilde{C}}^n, (\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n, \underline{\tilde{W}}; \underline{\tilde{Z}}^n, (\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n, \underline{\tilde{L}}_E^n, \underline{\tilde{W}}\big)\right] \\ &= \mathbb{E}_{\mathcal{C}}\left[H\big(\underline{\tilde{P}}^n, \underline{\tilde{C}}^n, (\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n, \underline{\tilde{W}}\big)\right] - \mathbb{E}_{\mathcal{C}}\left[H\big(\underline{\tilde{P}}^n, \underline{\tilde{C}}^n, (\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n, \underline{\tilde{W}}|\underline{\tilde{Z}}^n, (\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n, \underline{\tilde{L}}_E^n, \underline{\tilde{W}}\big)\right] \\ &\overset{(a)}{\geq} \mathbb{E}_{\mathcal{C}}\left[H\big(\underline{\tilde{P}}^n, \underline{\tilde{C}}^n, (\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n, \underline{\tilde{W}}\big)\right] - h(\lambda) - nN(R_p + R_c)\lambda \\ &\overset{(b)}{\geq} \mathbb{E}_{\mathcal{C}}\left[H\big((\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n, \underline{\tilde{W}}\big)\right] + \mathbb{E}_{\mathcal{C}}\left[H\big(\underline{\tilde{C}}^n|(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n, \underline{\tilde{W}}\big)\right] + \mathbb{E}_{\mathcal{C}}\left[H\big(\underline{\tilde{P}}^n|\underline{\tilde{C}}^n, (\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n, \underline{\tilde{W}}\big)\right] - nN\varepsilon \\ &\overset{(c)}{=} \mathbb{E}_{\mathcal{C}}\left[H\big((\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n, \underline{\tilde{W}}\big)\right] + \mathbb{E}_{\mathcal{C}}\left[H\big(\underline{\tilde{C}}^n|(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n, \underline{\tilde{W}}\big)\right] + \mathbb{E}_{\mathcal{C}}\left[H\big(\underline{\tilde{P}}^n\big)\right] - nN\varepsilon \\ &\overset{(d)}{=} \mathbb{E}_{\mathcal{C}}\left[H\big((\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n, \underline{\tilde{W}}\big)\right] + N\mathbb{E}_{\mathcal{C}}\left[H\big(\tilde{C}^n|(\tilde{Z}^E)^n, \tilde{W}\big)\right] + \mathbb{E}_{\mathcal{C}}\left[H\big(\underline{\tilde{P}}^n\big)\right] - nN\varepsilon \\ &\overset{(e)}{=} \mathbb{E}_{\mathcal{C}}\left[H\big((\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n, \underline{\tilde{W}}\big)\right] + nN(C_E - \varepsilon) + \mathbb{E}_{\mathcal{C}}\left[H\big(\underline{\tilde{P}}^n\big)\right] - nN\varepsilon \\ &= \mathbb{E}_{\mathcal{C}}\left[H\big((\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n, \underline{\tilde{W}}\big)\right] + nN(C_E - 2\varepsilon) + \mathbb{E}_{\mathcal{C}}\left[H\big(\underline{\tilde{P}}^n\big)\right] \end{aligned} \tag{3.51}$$

where (a) holds due to (3.50), (b) holds since we choose λ small enough so that $h(\lambda)+n(R_p+R_c)\lambda < n\varepsilon$, (c) holds since the public bits $\underline{\tilde{P}}^n$ are chosen independently of the message $\underline{\tilde{W}}$, the confidential bits $\underline{\tilde{C}}^n$ and the noise inserted in the network, (d) holds since the information bits on the different layers of the stacked network are independent, and (e) follows from equation (3.49).

By applying the chain rule on mutual information $I\big(\underline{\tilde{P}}^n, \underline{\tilde{C}}^n, (\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n, \underline{\tilde{W}}; \underline{\tilde{Z}}^n, (\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n, \underline{\tilde{L}}_E^n, \underline{\tilde{W}}\big)$

we get similarly to inequality (3.42)

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{P}}^n,\underline{\tilde{C}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n,\underline{\tilde{W}};\underline{\tilde{Z}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n,\underline{\tilde{L}}_E^n,\underline{\tilde{W}}\big)\right]\\
&\leq \mathbb{E}_{\mathcal{C}}\left[H\big((\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n,\underline{\tilde{W}}\big)\right]+\mathbb{E}_{\mathcal{C}}\left[H\big(\underline{\tilde{L}}_E^n\big)\right]+\mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{P}}^n,\underline{\tilde{C}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n,\underline{\tilde{W}};\underline{\tilde{Z}}^n|(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n,\underline{\tilde{L}}_E^n,\underline{\tilde{W}}\big)\right]\\
&\leq \mathbb{E}_{\mathcal{C}}\left[H\big((\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n,\underline{\tilde{W}}\big)\right]+nNC_E+\mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{P}}^n,\underline{\tilde{C}}^n;\underline{\tilde{Z}}^n|(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n,\underline{\tilde{L}}_E^n,\underline{\tilde{W}}\big)\right]
\end{aligned}$$

Therefore $\mathbb{E}_{\mathcal{C}}\left[H\big((\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n,\underline{\tilde{W}}\big)\right]+nN(C_E-2\varepsilon)+\mathbb{E}_{\mathcal{C}}\left[H\big(\underline{\tilde{P}}^n\big)\right]\leq \mathbb{E}_{\mathcal{C}}\left[H\big((\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n,\underline{\tilde{W}}\big)\right]+nNC_E+\mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{P}}^n,\underline{\tilde{C}}^n;\underline{\tilde{Z}}^n|(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n,\underline{\tilde{L}}_E^n,\underline{\tilde{W}}\big)\right]$ or equivalently

$$\mathbb{E}_{\mathcal{C}}\left[H\big(\underline{\tilde{P}}^n\big)\right]\leq \mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{P}}^n,\underline{\tilde{C}}^n;\underline{\tilde{Z}}^n|(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n,\underline{\tilde{L}}_E^n,\underline{\tilde{W}}\big)\right]+2nN\varepsilon. \tag{3.52}$$

Moreover since $\underline{\tilde{W}},\underline{\tilde{L}}_E^n,\underline{\tilde{P}}^n,\underline{\tilde{C}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n\to\underline{\tilde{X}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n\to\underline{\tilde{Z}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n$ we get from the data processing inequality that $I\big(\underline{\tilde{W}},\underline{\tilde{L}}_E^n,\underline{\tilde{P}}^n,\underline{\tilde{C}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n;\underline{\tilde{Z}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n\big)\leq I\big(\underline{\tilde{X}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n;\underline{\tilde{Z}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n\big)$ and by using the chain rule in the left hand side of this inequality we get

$$\begin{aligned}
&I\big(\underline{\tilde{W}};\underline{\tilde{Z}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n\big)+I\big(\underline{\tilde{L}}_E^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n;\underline{\tilde{Z}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n|\underline{\tilde{W}}\big)+I\big(\underline{\tilde{P}}^n,\underline{\tilde{C}}^n;\underline{\tilde{Z}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n|\underline{\tilde{W}},\underline{\tilde{L}}_E^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n\big)\\
&\leq I\big(\underline{\tilde{X}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n;\underline{\tilde{Z}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n\big).
\end{aligned}\tag{3.53}$$

The right hand side of inequality (3.53) can be expanded as

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{X}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n;\underline{\tilde{Z}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n\big)\right]&=\mathbb{E}_{\mathcal{C}}\left[H\big(\underline{\tilde{Z}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n\big)\right]-\mathbb{E}_{\mathcal{C}}\left[H\big(\underline{\tilde{Z}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n|\underline{\tilde{X}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n\big)\right]\\
&\overset{(a)}{=}\mathbb{E}_{\mathcal{C}}\left[H\big(\underline{\tilde{Z}}^n\big)\right]+\mathbb{E}_{\mathcal{C}}\left[H\big((\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n|\underline{\tilde{Z}}^n\big)\right]-\mathbb{E}_{\mathcal{C}}\left[H\big(\underline{\tilde{Z}}^n|\underline{\tilde{X}}^n\big)\right]\\
&\leq \mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{X}}^n;\underline{\tilde{Z}}^n\big)\right]+\mathbb{E}_{\mathcal{C}}\left[H\big((\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n\big)\right]\\
&=nN\max_{p(x)}I(X;Z)+\mathbb{E}_{\mathcal{C}}\left[H\big((\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n\big)\right].
\end{aligned}\tag{3.54}$$

where equality (a) holds since $\underline{\tilde{Z}}^n$ is conditionally independent of $(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n$ given $\underline{\tilde{X}}^n$. Due to the fact that $I\big(\underline{\tilde{P}}^n,\underline{\tilde{C}}^n;\underline{\tilde{Z}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n|\underline{\tilde{W}},\underline{\tilde{L}}_E^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n\big)=I\big(\underline{\tilde{P}}^n,\underline{\tilde{C}}^n;\underline{\tilde{Z}}^n|\underline{\tilde{W}},\underline{\tilde{L}}_E^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n\big)$ and combining (3.53), (3.54) we get

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{W}};\underline{\tilde{Z}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n\big)\right]+\mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{L}}_E^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n;\underline{\tilde{Z}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n|\underline{\tilde{W}}\big)\right]+\mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{P}}^n,\underline{\tilde{C}}^n;\underline{\tilde{Z}}^n|\underline{\tilde{W}},\underline{\tilde{L}}_E^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n\big)\right]\\
&\leq nN\max_{p(x)}I(X;Z)+\mathbb{E}_{\mathcal{C}}\left[H\big((\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n\big)\right]
\end{aligned}$$

or since $H\big((\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n|\underline{\tilde{W}}\big)=I\big(\underline{\tilde{L}}_E^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n;(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n|\underline{\tilde{W}}\big)\leq I\big(\underline{\tilde{L}}_E^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n;\underline{\tilde{Z}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n|\underline{\tilde{W}}\big)$

$$\mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{W}};\underline{\tilde{Z}}^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n\big)\right]+\mathbb{E}_{\mathcal{C}}\left[H\big((\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n|\underline{\tilde{W}}\big)\right]+\mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{P}}^n,\underline{\tilde{C}}^n;\underline{\tilde{Z}}^n|\underline{\tilde{W}},\underline{\tilde{L}}_E^n,(\underline{\tilde{Z}}^{E\setminus\{\bar{e}\}})^n\big)\right]$$

$$\leq nN \max_{p(x)} I(X;Z) + \mathbb{E}_{\mathcal{C}}\left[H\big((\tilde{\underline{Z}}^{E\setminus\{\bar{e}\}})^n\big)\right]$$

or else $\mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{W}};\tilde{\underline{Z}}^n,(\tilde{\underline{Z}}^{E\setminus\{\bar{e}\}})^n\big)\right]$ is upper bounded by

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{W}};\tilde{\underline{Z}}^n,(\tilde{\underline{Z}}^{E\setminus\{\bar{e}\}})^n\big)\right]\\
&\leq nN \max_{p(x)} I(X;Z) + \mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{W}};(\tilde{\underline{Z}}^{E\setminus\{\bar{e}\}})^n\big)\right] - \mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{P}}^n,\underline{\tilde{C}}^n;\underline{\tilde{Z}}^n|\underline{\tilde{W}},\underline{\tilde{L}}_E^n,(\tilde{\underline{Z}}^{E\setminus\{\bar{e}\}})^n\big)\right]\\
&\overset{(a)}{\leq} nN \max_{p(x)} I(X;Z) + \mathbb{E}_{\mathcal{C}}\left[I\big(\underline{\tilde{W}};(\tilde{\underline{Z}}^{E\setminus\{\bar{e}\}})^n\big)\right] - \mathbb{E}_{\mathcal{C}}\left[H\big(\underline{\tilde{P}}^n\big)\right] + 2nN\varepsilon\\
&\overset{(b)}{\leq} nN \max_{p(x)} I(X;Z) + N\mathbb{E}_{\mathcal{C}}\left[I\big(\tilde{W};(\tilde{Z}^{E\setminus\{\bar{e}\}})^n\big)\right] - nN\big(\max_{p(x)} I(X;Z) - \varepsilon\big) + 2nN\varepsilon\\
&\overset{(c)}{\leq} 4nN\varepsilon
\end{aligned}$$

where inequality (a) follows from (3.52) and inequality (b) holds since the messages $\tilde{W}$ are independent across different layers and equation (3.49) and inequality (c) holds since the code for network $\mathcal{N}_{\bar{e}}(R_c,0)$ of rate $\tilde{R}$ is secure, *i.e.* $\mathbb{E}_{\mathcal{C}}\left[I\big(\tilde{W};(\tilde{Z}^{E\setminus\{\bar{e}\}})^n\big)\right] \leq n\varepsilon$.

Since $\underline{W} \to \underline{\tilde{W}} \to \tilde{\underline{Z}}^n, (\tilde{\underline{Z}}^{E\setminus\{\bar{e}\}})^n$ and the data processing inequality we get that we have constructed a random code with $\mathbb{E}_{\mathcal{C}}\left[I(\underline{W};\tilde{\underline{Z}}^n,(\tilde{\underline{Z}}^{E\setminus\{\bar{e}\}})^n)\right] \leq 4nN\varepsilon$. One can follow an analysis identical to the one used in equation (3.1) of Theorem 8 to prove that there is at least one deterministic code $(3\lambda, 12\varepsilon, R)$–$\mathcal{S}(\mathcal{N},A)$ for network $\mathcal{N}$. □

3.4.4 Proof of Theorem 12

Proof of Theorem 12. The proof is very similar to the proof of Lemma 4 and therefore we will only give an outline. Similar to the proof of Lemma 4 we assume any rate R in the relative interior of the rate region of network $\mathcal{N}(A)$, *i.e.* $R \in \operatorname{int}(\mathcal{R}(\mathcal{N}(A)))$, and choose another rate $\tilde{R} \in \operatorname{int}(\mathcal{R}(\mathcal{N}(A)))$ such that $\tilde{R}^{(i\to\mathcal{B})} > R^{(i\to\mathcal{B})}$ for all $i \in \mathcal{V}$ and $\mathcal{B} \in \mathcal{B}^{(i)}$ with $R^{(i\to\mathcal{B})} > 0$. Then by a carefully selected $\tilde{\lambda} > 0$ and for every $\varepsilon > 0$ we can construct a stacked solution of rate R for stacked network $\underline{\mathcal{N}}(A)$ with probability of error less than $\lambda/3$. The number of layers in the stack will be denoted by N.

We make use of an auxiliary network **I** which is the same as the A-enhanced network except that the noiseless bit pipes in $\{\check{\mathcal{C}}_e : e \in \mathcal{E}\} \cup \{h_e : e \in \mathcal{E}\}$ are changed back to the original noisy channels. To guarantee the security of the code used in network **I** we need to fill up to capacity the degraded broadcast channels of network **I**. Specifically when the broadcast noiseless channels of network $\underline{\mathcal{N}}(A)$ are changed to their noisy counterparts in network **I** the capacity of the eavesdropper's channel increases from $R_{e,p}$ to $\max_{p(x)} I(X^{(e)},Z^{(e)})$. To fill up the capacity to the eavesdropper for every broadcast channel $e \in E$ at each time step $t \in \{1,\ldots,n\}$ we add to the public bits some random bits $\underline{F}_t^{(e)}$ such that $H\big(\underline{F}_t^{(e)}\big) = N\big(\max_{p(x)} I(X^{(e)},Z^{(e)}) - R_{e,p}\big)$. Since the random bits are independent

across different time steps

$$H\big((\underline{F}^{(e)})^n\big) = nN\big(\max_{p(x)} I(X^{(e)}, Z^{(e)}) - R_{e,p}\big). \tag{3.55}$$

In the rest of the proof we will assume that the public bits $(\underline{\bar{P}}^{(e)})^n$ at each channel e of network $\mathbf{I}$ of Figure 3.6 denote both the public bits $(\underline{P}^{(e)})^n$ of network $\underline{\mathcal{N}}(A)$ and the random bits $(\underline{F}^{(e)})^n$ added to fill up the eavesdropper's pipe, *i.e.* $(\underline{\bar{P}}^{(e)})^n = \big((\underline{P}^{(e)})^n, (\underline{F}^{(e)})^n\big)$

We will show that rate R is achievable in network $\mathbf{I}$ of Figure 3.6 where all noiseless links are changed to noisy ones. Rate R is indeed achievable since we can apply a channel code at every channel $e \in \mathcal{E}$ and since $R_e < \max_{p(x)} I(X^{(e)}, Y^{(e)})$ and $\bar{R}_{e,p} + R_{e,c} < \max_{p(x)} I(X^{(e)}, Y^{(e)})$ for every point to point and degraded channel respectively the channel codes can transfer the bits previously transfered by the noiseless links to the output of channel e for all $e \in \mathcal{E}$ with small probability of error where rate $\bar{R}_{e,p}$ correspond to the rate of public bits $(\underline{\bar{P}}^{(e)})^n$. The $|E|$ additional receivers of the stacked network $\underline{\mathcal{N}}(A)$ can decode messages $(\underline{W}, \underline{T})$ by having access to message $\underline{W}$, the bits $\underline{L}_E^n$ through the side channel, and the public bits $(\underline{P}^E)^n$ where $(\underline{P}^E)^n$ denotes $((\underline{P}^{(e_1)})^n, \ldots, (\underline{P}^{(e_{j_E})})^n)$ for all $e_r \in E$ with small probability of error. Since by decoding $(\underline{W}, \underline{T})$ one knows every bit transversing the noiseless network $\underline{\mathcal{N}}(A)$ and therefore the confidential bits $(\underline{C}^E)^n$ by Fano's inequality [60, Theorem 2.10.1] we get

$$H\big((\underline{C}^E)^n | \underline{W}, \underline{L}_E^n, (\underline{P}^E)^n\big) \leq h(\lambda) + nN\lambda \sum_{e \in E} R_{e,c}.$$

This equation is similar to equation (3.31) in the proof of Lemma 4. By an analysis identical to the one following equation (3.31) it can be proved that the additional receivers can decode $(\underline{W}, \underline{T})$ and therefore rate R is achievable at network $\mathbf{I}$ of Figure 3.6.

It only remains to prove that the code above designed for network $\mathbf{I}$ of Figure 3.6 is secure when used in network $\mathcal{N}$. Indeed by Fano's inequality and by choosing λ small enough on can get

$$\begin{aligned} I\big(\underline{T}; \underline{L}_E^n, (\underline{Z}^{(E)})^n | \underline{W}\big) &= H\big(\underline{T}|\underline{W}\big) - H\big(\underline{T}|\underline{W}, \underline{L}_E^n, (\underline{Z}^{(E)})^n\big) \\ &\geq H\big(\underline{T}\big) - nN\varepsilon. \end{aligned}$$

and since $I\big(\underline{T}; \underline{L}_E^n, (\underline{Z}^{(E)})^n | \underline{W}\big) = I\big(\underline{T}; (\underline{Z}^{(E)})^n | \underline{W}\big) + I\big(\underline{T}; \underline{L}_E^n | \underline{W}, (\underline{Z}^{(E)})^n\big)$ we get $I\big(\underline{T}; (\underline{Z}^{(E)})^n | \underline{W}\big) \geq H\big(\underline{T}\big) - I\big(\underline{T}; \underline{L}_E^n | \underline{W}, (\underline{Z}^{(E)})^n\big) - nN\varepsilon$ or else

$$\begin{aligned} -I\big(\underline{T}; (\underline{Z}^{(E)})^n | \underline{W}\big) &\leq I\big(\underline{T}; \underline{L}_E^n | \underline{W}, (\underline{Z}^{(E)})^n\big) - H\big(\underline{T}\big) + n\varepsilon \\ &\leq H\big(\underline{L}_E^n\big) - H\big(\underline{T}\big) + nN\varepsilon. \end{aligned} \tag{3.56}$$

Moreover since $\underline{W}, \underline{T}, (\underline{F}^E)^n \to (\underline{X}^E)^n \to (\underline{Z}^{(E)})^n$ we have

$$I\big(\underline{W}, \underline{T}, (\underline{F}^E)^n; (\underline{Z}^{(E)})^n\big) \leq I\big((\underline{X}^E)^n; (\underline{Z}^{(E)})^n\big) \leq nN \sum_{e \in E} \max_{p(x)} I(X^{(e)}, Z^{(e)})$$

and since $I\big(\underline{W}, \underline{T}, (\underline{F}^E)^n; (\underline{Z}^{(E)})^n\big) = I\big(\underline{W}, \underline{T}; (\underline{Z}^{(E)})^n\big) + I\big((\underline{F}^E)^n; (\underline{Z}^{(E)})^n | \underline{W}, \underline{T}\big)$ the inequality above becomes

$$\begin{aligned}
I\big(\underline{W}; (\underline{Z}^{(E)})^n\big) &\leq nN \sum_{e \in E} \max_{p(x)} I(X^{(e)}, Z^{(e)}) - I\big(\underline{T}; (\underline{Z}^{(E)})^n | \underline{W}\big) - I\big((\underline{F}^E)^n; (\underline{Z}^{(E)})^n | \underline{W}, \underline{T}\big) \\
&\overset{(a)}{=} nN \sum_{e \in E} \max_{p(x)} I(X^{(e)}, Z^{(e)}) - I\big(\underline{T}; (\underline{Z}^{(E)})^n | \underline{W}\big) - H\big((\underline{F}^E)^n\big) + nN\varepsilon \\
&\overset{(b)}{=} nN \sum_{e \in E} \max_{p(x)} I(X^{(e)}, Z^{(e)}) + H\big(\underline{L}_E^n\big) - H\big(\underline{T}\big) - H\big((\underline{F}^E)^n\big) + 2nN\varepsilon \\
&\overset{(c)}{\leq} nN \sum_{e \in E} \max_{p(x)} I(X^{(e)}, Z^{(e)}) + nNC_E - H\big(\underline{T}\big) - H\big((\underline{F}^E)^n\big) + 2nN\varepsilon \\
&\overset{(d)}{\leq} 3nN\varepsilon
\end{aligned}$$

where inequality (a) holds since $H\big((\underline{F}^E)^n | \underline{W}, \underline{T}\big) = H\big((\underline{F}^E)^n\big)$ due to the fact the random bits $((\underline{F}^E)^n)$ are independent from $(\underline{W}, \underline{T})$ and moreover $H\big((\underline{F}^E)^n | \underline{W}, \underline{T}, (\underline{Z}^{(E)})^n\big) \leq nN\varepsilon$ since $(\underline{F}^E)^n$ can be decoded with small probability of error by having access to $\big(\underline{W}, \underline{T}, (\underline{Z}^{(E)})^n\big)$. Inequality (b) follows from (3.56), inequality (c) holds since it upper bounds the entropy of bits transversing the side channel by its capacity and finally inequality (d) follows from the fact that $H\big(\underline{T}\big) = nNR_T$ the definition of rate R_T and capacity C_E and equation (3.55). We have proved therefore that the constructed random code for network $\mathcal{N}$ is secure and one can follow an analysis identical to the one used in equation (3.1) of Theorem 8 to prove that there is at least one deterministic code $(\lambda, 9\varepsilon, R)$–$\mathcal{S}(\mathcal{N}, A)$ for network $\mathcal{N}$. □

Appendix C

Secrecy Capacity for Networks $\hat{\mathcal{N}}$ in Figure 3.2(c) and $\breve{\mathcal{N}}$ in Figure 3.2(b)

Lemma 6 derives the secure multicast capacity for noiseless network $\hat{\mathcal{N}}$ in Figure 3.2(c).

Lemma 6. *Given the noiseless network $\hat{\mathcal{N}}$ in Figure 3.2(c) and the adversarial set $A = \big\{\{e_1, e_3\}, \{e_2\}\big\}$, $R^{(1\to\{2,3\})} \leq 1$ for all $R \in \mathcal{R}(\hat{\mathcal{N}}, A)$, and the given bound is achieved with equality when $R^{(i\to\mathcal{B})} = 0$ for all $(i, \mathcal{B}) \neq (1, \{2, 3\})$.*

Proof. The achievability of rate vector R with $R^{(1\to\{2,3\})} = 1$ and $R^{(i\to\mathcal{B})} = 0$ for all other $(i, \mathcal{B})$ is proven by the code in Figure 3.2(d), as described in Example 1. It therefore remains only to derive the converse. The definition of secure capacity requires that for any $\lambda > 0$ and $\varepsilon > 0$ there exists a blocklength n such that $I\big(W^{(1\to\{2,3\})}; (Z^{(e_2)})^n\big) < n\varepsilon$ and $\Pr\big(\breve{W}^{(1\to\{2,3\})} \neq W^{(1\to\{2,3\})}\big) < \lambda$. For any such code,

$$
\begin{aligned}
nR^{(1\to\{2,3\})} &= H\big(W^{(1\to\{2,3\})}\big) \\
&< H\big(W^{(1\to\{2,3\})} | (Z^{(e_2)})^n\big) + n\varepsilon \\
&\overset{(a)}{\leq} H\big(W^{(1\to\{2,3\})} | (Z^{(e_2)})^n\big) + n\varepsilon - H\big(W^{(1\to\{2,3\})} | (Y^{(e_1)})^n, (Y^{(e_2)})^n\big) + h(\lambda) + n\lambda R^{(1\to\{2,3\})} \\
&\overset{(b)}{=} I\big(W^{(1\to\{2,3\})}; (X^{(e_1),c})^n, (X^{(e_1),p})^n | (Z^{(e_2)})^n\big) + n\varepsilon + h(\lambda) + n\lambda R^{(1\to\{2,3\})} \\
&\leq H\big((X^{(e_1),c})^n\big) + H\big((X^{(e_1),p})^n\big) + n\big(\varepsilon + h(\lambda)/n + \lambda R^{(1\to\{2,3\})}\big) \\
&\overset{(c)}{\leq} n\big(1 + \varepsilon + h(\lambda)/n + \frac{3}{2}\lambda\big).
\end{aligned}
$$

where equality (a) follows from Fano's inequality [60, Theorem 2.10.1] since receiver $\mathcal{T}_1$ can decode message $W^{(1\to\{2,3\})}$ with error probability no greater than λ by observing the outputs of channels e_1 and e_2, (b) follows since $(Y^{(e_1)})^n = \big((X^{(e_1),c})^n, (X^{(e_1),p})^n\big)$ and $(Y^{(e_2)})^n = (Z^{(e_2)})^n$, and (c) follows from the capacities of the confidential and public parts of channel e_1 and the fact that $R^{(1\to\{2,3\})} \leq \frac{3}{2}$,

which is the multicast capacity without the secrecy constraint. Therefore $R^{(1\to\{2,3\})} \leq 1$ for network $\hat{\mathcal{N}}$ since the inequality must hold for all ε and λ greater than 0. $\square$

Lemma 7 bounds the maximal secure multicast capacity for network $\mathcal{N}$ of Figure 3.2(a). The following definitions are used in its proof. For any (possibly empty) subset L of $\{1,\ldots,n\}$ let L^c denote the compliment ($L^c = \{1,\ldots,n\}\backslash L$), and X_L be the vector $(X_t : t \in L)$ with elements ordered according to their time indices; for example, $X_{\{1,5\}} = (X_1, X_5)$. Let $L_j = \{t \in \{1,\ldots,n\} : Z_t^{(e_j)} = X_t^{(e_j)}\}$; that is, L_j is the subset of time steps left unerased by channel e_j. Note that L_j is a deterministic function of $(Z^{(e_j)})^n$ and therefore $H\big((Z^{(e_j)})^n\big) = H\big(L_j, (Z^{(e_j)})^n\big)$. Now we have all the notation necessary to prove Lemma 7.

Lemma 7. *Given the network $\breve{\mathcal{N}}$ of Figure 3.2(b) and adversarial set $A = \big\{\{e_1, e_3\}, \{e_2\}\big\}$, $R^{(1\to\{2,3\})} \leq 0.875$ for all $R \in \mathcal{R}(\breve{\mathcal{N}}, A)$.*

Proof. The proof is by contradiction. The outline is as follows: First, we assume the existence of a secure code that violates the given bound. We then show that the security of this solution implies the lower bound on the entropy $H\big((Z^{(e_1)})^n, (Z^{(e_3)})^n | W^{(1\to\{2,3\})}\big)$ given in equation (C.2); the entropy in the degraded outputs observed by eavesdropper $\{e_1, e_3\}$ given message $W^{(1\to\{2,3\})}$ results from a combination of noise and random keys. Finally, we show that reliable decoding at receivers $\mathcal{T}_1$ and $\mathcal{T}_2$ implies an upper bound on the same conditional entropy, as described in equation (C.6). Since the lower bound is higher than the upper bound, we have a contradiction; the result follows.

Formally, suppose that there exists a rate vector $R \in \text{int}(\mathcal{R}(\breve{\mathcal{N}}, A))$ in the relative interior of rate region $\mathcal{R}(\breve{\mathcal{N}}, A)$ with $R^{(1\to\{2,3\})} > 0.875$. Then for any $\lambda > 0$ and $\varepsilon > 0$ there is a blocklength n for which a $(\lambda, \varepsilon, A, R)$–$\mathcal{S}(\breve{\mathcal{N}})$ solution exists. Fix such a solution; then

$$\begin{aligned} \Pr\big(\breve{W}^{(1\to\{2,3\})} \neq W^{(1\to\{2,3\})}\big) &< \lambda \\ I\big(W^{(1\to\{2,3\})}; (Z^{(e_1)})^n, (Z^{(e_3)})^n\big) &< n\varepsilon \\ I\big(W^{(1\to\{2,3\})}; (Z^{(e_2)})^n\big) &< n\varepsilon. \end{aligned} \tag{C.1}$$

We begin by bounding entropy $H\big((Z^{(e_1)})^n, (Z^{(e_3)})^n | W^{(1\to\{2,3\})}\big)$ from below as

$$\begin{aligned} &H\big((Z^{(e_1)})^n, (Z^{(e_3)})^n | W^{(1\to\{2,3\})}\big) \\ &\overset{(a)}{>} H\big((Z^{(e_1)})^n, (Z^{(e_3)})^n\big) - n\varepsilon \\ &= H\big((Z^{(e_1)})^n\big) + H\big((Z^{(e_3)})^n | (Z^{(e_1)})^n\big) - n\varepsilon \\ &= H\big(L_1\big) + H\big((Z^{(e_1)})^n | L_1\big) + H\big(L_3 | (Z^{(e_1)})^n\big) + H\big((Z^{(e_3)})^n | (Z^{(e_1)})^n, L_1, L_3\big) - n\varepsilon \\ &\overset{(b)}{=} n(2-\varepsilon) + \frac{1}{2^n} \sum_{l_1 \subseteq \{1,\ldots,n\}} \Big[H\big((Z^{(e_1)})^n | L_1 = l_1\big) + \frac{1}{2^n} \sum_{l_3 \subseteq \{1,\ldots,n\}} H\big((Z^{(e_3)})^n | (Z^{(e_1)})^n, L_1 = l_1, L_3 = l_3\big) \Big] \end{aligned}$$

$$
\begin{aligned}
&= n(2-\varepsilon) + \frac{1}{2}\cdot\frac{1}{2^n}\sum_{l_1\subseteq\{1,\ldots,n\}}\Big[H\big((Z^{(e_1)})^n|L_1=l_1\big)+H\big((Z^{(e_1)})^n|L_1=l_1^c\big)\Big]\\
&+\frac{1}{4}\cdot\frac{1}{2^{2n}}\sum_{l_1\subseteq\{1,\ldots,n\}}\sum_{l_3\subseteq\{1,\ldots,n\}}\Big[H\big((Z^{(e_3)})^n|(Z^{(e_1)})^n,L_1=l_1,L_3=l_3\big)+H\big((Z^{(e_3)})^n|(Z^{(e_1)})^n,L_1=l_1,L_3=l_3^c\big)\\
&+H\big((Z^{(e_3)})^n|(Z^{(e_1)})^n,L_1=l_1^c,L_3=l_3\big)+H\big((Z^{(e_3)})^n|(Z^{(e_1)})^n,L_1=l_1^c,L_3=l_3^c\big)\Big]\\
&= n(2-\varepsilon) + \frac{1}{2^{n+1}}\sum_{l_1\subseteq\{1,\ldots,n\}}\Big[H\big(X_{l_1}^{(e_1)}|L_1=l_1\big)+H(X_{l_1^c}^{(e_1)}|L_1=l_1^c)\Big]\\
&+\frac{1}{2^{2n+2}}\sum_{l_1\subseteq\{1,\ldots,n\}}\sum_{l_3\subseteq\{1,\ldots,n\}}\Big[H\big(X_{l_3}^{(e_3)}|X_{l_1}^{(e_1)},L_1=l_1,L_3=l_3\big)+H\big(X_{l_3}^{(e_3)}|X_{l_1}^{(e_1)},L_1=l_1,L_3=l_3^c\big)\\
&+H\big(X_{l_3^c}^{(e_3)}|X_{l_1^c}^{(e_1)},L_1=l_1^c,L_3=l_3\big)+H\big(X_{l_3^c}^{(e_3)}|X_{l_1^c}^{(e_1)},L_1=l_1^c,L_3=l_3^c\big)\Big]\\
&= n(2-\varepsilon) + \frac{1}{2^{n+1}}\sum_{l_1\subseteq\{1,\ldots,n\}}\Big[H\big(X_{l_1}^{(e_1)}\big)+H\big(X_{l_1^c}^{(e_1)}\big)\Big]+\frac{1}{2^{2n+2}}\sum_{l_1\subseteq\{1,\ldots,n\}}\sum_{l_3\subseteq\{1,\ldots,n\}}\Big[H\big(X_{l_3}^{(e_3)}|X_{l_1}^{(e_1)}\big)\\
&+H\big(X_{l_3^c}^{(e_3)}|X_{l_1}^{(e_1)}\big)+H\big(X_{l_3}^{(e_3)}|X_{l_1^c}^{(e_1)}\big)+H\big(X_{l_3^c}^{(e_3)}|X_{l_1^c}^{(e_1)}\big)\Big]\\
&\geq n(2-\varepsilon) + \frac{1}{2^{n+1}}\sum_{l_1\subseteq\{1,\ldots,n\}}H\big(X_{l_1}^{(e_1)},X_{l_1^c}^{(e_1)}\big)+\frac{1}{2^{2n+2}}\sum_{l_1\subseteq\{1,\ldots,n\}}\sum_{l_3\subseteq\{1,\ldots,n\}}\Big[H\big(X_{l_3}^{(e_3)},X_{l_3^c}^{(e_3)}|X_{l_1}^{(e_1)}\big)\\
&+H\big(X_{l_3}^{(e_3)},X_{l_3^c}^{(e_3)}|X_{l_1^c}^{(e_1)}\big)\Big]\\
&= n(2-\varepsilon) + \frac{1}{2^{n+1}}\sum_{l_1\subseteq\{1,\ldots,n\}}H\big((X^{(e_1)})^n\big)+\frac{1}{2^{2n+2}}\sum_{l_1\subseteq\{1,\ldots,n\}}\sum_{l_3\subseteq\{1,\ldots,n\}}\Big[H\big((X^{(e_3)})^n|X_{l_1}^{(e_1)}\big)+H\big((X^{(e_3)})^n|X_{l_1^c}^{(e_1)}\big)\Big]\\
&= n(2-\varepsilon) + \frac{1}{2}H\big((X^{(e_1)})^n\big)+\frac{1}{2^{n+2}}\sum_{l_1\subseteq\{1,\ldots,n\}}\Big[2H\big((X^{(e_3)})^n\big)-I\big((X^{(e_3)})^n;X_{l_1}^{(e_1)}\big)-I\big((X^{(e_3)})^n;X_{l_1^c}^{(e_1)}\big)\Big]\\
&\overset{(c)}{\geq} n(2-\varepsilon) + \frac{1}{2}H\big((X^{(e_1)})^n\big)+\frac{1}{2^{n+2}}\sum_{l_1\subseteq\{1,\ldots,n\}}\Big[2H\big((X^{(e_3)})^n\big)-|l_1|-|l_1^c|\Big]\\
&= n(2-\varepsilon) + \frac{1}{2}H\big((X^{(e_1)})^n\big)+\frac{1}{2^{n+2}}\sum_{l_1\subseteq\{1,\ldots,n\}}\Big[2H\big((X^{(e_3)})^n\big)-n\Big]\\
&= n\big(\frac{7}{4}-\varepsilon\big)+\frac{1}{2}\Big[H\big((X^{(e_1)})^n\big)+H\big((X^{(e_3)})^n\big)\Big], \qquad \text{(C.2)}
\end{aligned}
$$

where (a) follows from the second inequality of (C.1); (b) holds since $H\big(L_1\big)=H\big(L_3|(Z^{(e_1)})^n\big)=n$ and by the definition of conditional entropy, and (c) holds since $I\big((X^{(e_3)})^n;X_A^{(e_1)}\big)\leq H\big(X_A^{(e_1)}\big)\leq|A|$ for any $A\subseteq\{1,\ldots,n\}$.

To bound $H\big((Z^{(e_1)})^n,(Z^{(e_3)})^n|W^{(1\to\{2,3\})}\big)$ from above, recall that $(Y^{(e_2)})^n=(Z^{(e_2)})^n$. Therefore,

$$
\begin{aligned}
&H\big((Z^{(e_1)})^n,(Z^{(e_3)})^n|W^{(1\to\{2,3\})}\big)\\
&\leq H\big((Z^{(e_1)})^n,(Z^{(e_3)})^n,(Z^{(e_2)})^n|W^{(1\to\{2,3\})}\big)\\
&= H\big((Z^{(e_2)})^n|W^{(1\to\{2,3\})}\big)+H\big((Z^{(e_1)})^n|(Z^{(e_2)})^n,W^{(1\to\{2,3\})}\big)+H\big((Z^{(e_3)})^n|(Z^{(e_1)})^n,(Z^{(e_2)})^n,W^{(1\to\{2,3\})}\big)\\
&\leq H\big((Z^{(e_2)})^n\big)+H\big((Z^{(e_1)})^n|(Y^{(e_2)})^n,W^{(1\to\{2,3\})}\big)+H\big((Z^{(e_3)})^n|(Y^{(e_2)})^n,W^{(1\to\{2,3\})}\big)
\end{aligned}
$$

$$\overset{(a)}{\leq} \frac{1}{2}n + H\big((Z^{(e_1)})^n|(Y^{(e_2)})^n, W^{(1\to\{2,3\})}\big) + H\big((Z^{(e_3)})^n|(Y^{(e_2)})^n, W^{(1\to\{2,3\})}\big) \tag{C.3}$$

where (a) follows since the adversary's output $(Z^{(e_2)})^n$ of the noiseless wiretap channel $\mathcal{C}_{e_2}(o, \frac{1}{2})$ has maximal entropy $\frac{n}{2}$. To bound the second term in (C.3), note that

$$\begin{aligned}
&H\big((Z^{(e_1)})^n|(Y^{(e_2)})^n, W^{(1\to\{2,3\})}\big)\\
&\leq H\big((X^{(e_1)})^n, (Z^{(e_1)})^n|(Y^{(e_2)})^n, W^{(1\to\{2,3\})}\big)\\
&= H\big((X^{(e_1)})^n, (Z^{(e_1)})^n, (Y^{(e_2)})^n, W^{(1\to\{2,3\})}\big) - H\big(W^{(1\to\{2,3\})}\big) - H\big((Y^{(e_2)})^n|W^{(1\to\{2,3\})}\big)\\
&= H\big((X^{(e_1)})^n\big) + H\big(Z^{(e_1)})^n|(X^{(e_1)})^n\big) + H\big((Y^{(e_2)})^n|(X^{(e_1)})^n, (Z^{(e_1)})^n\big)\\
&+ H\big(W^{(1\to\{2,3\})}|(X^{(e_1)})^n, (Z^{(e_1)})^n, (Y^{(e_2)})^n\big) - nR^{(1\to\{2,3\})} - H\big((Y^{(e_2)})^n|W^{(1\to\{2,3\})}\big)\\
&\overset{(a)}{\leq} H\big((X^{(e_1)})^n\big) + \big(1 + \frac{h(\lambda)}{n} + \lambda R^{(1\to\{2,3\})} - R^{(1\to\{2,3\})}\big)n + H\big((Y^{(e_2)})^n|(X^{V_1(e_1)})^n, (Z^{(e_1)})^n\big) - H\big((Y^{(e_2)})^n\big) + n\varepsilon\\
&= H\big((X^{(e_1)})^n\big) + \big(1 + \frac{h(\lambda)}{n} + (\lambda-1)R^{(1\to\{2,3\})} + \varepsilon\big)n - \Big(H\big((Y^{(e_2)})^n\big) - H\big((Y^{(e_2)})^n|(X^{(e_1)})^n, (Z^{(e_1)})^n\big)\Big)\\
&\overset{(b)}{\leq} H\big((X^{(e_1)})^n\big) + \big(1 + \frac{h(\lambda)}{n} + (\lambda-1)R^{(1\to\{2,3\})} + \varepsilon\big)n.
\end{aligned} \tag{C.4}$$

Here (a) follows from $H\big((Z^{(e_1)})^n|(X^{(e_1)})^n\big) = n$, $H\big(W^{(1\to\{2,3\})}|(X^{(e_1)})^n, (Z^{(e_1)})^n, (Y^{(e_2)})^n\big) \leq h(\lambda) + n\lambda R^{(1\to\{2,3\})}$ by Fano's inequality [60, Theorem 2.10.1], and $H\big((Y^{(e_2)})^n|W^{(1\to\{2,3\})}\big) \geq H\big((Y^{(e_2)})^n\big) - n\varepsilon$ by the third inequality of (C.1); (b) holds since conditioning reduces entropy. The bound for the third term in (C.3) is derived the same way, giving

$$H\big((Z^{(e_3)})^n|(Y^{(e_2)})^n, W^{(1\to\{2,3\})}\big) < H\big((X^{(e_3)})^n\big) + \big(1 + \frac{h(\lambda)}{n} + (\lambda-1)R^{(1\to\{2,3\})} + \varepsilon\big)n. \tag{C.5}$$

Therefore combining (C.3) with (C.4) and (C.5) gives

$$H\big((Z^{(e_1)})^n, (Z^{(e_3)})^n|W^{(1\to\{2,3\})}\big) \leq H\big((X^{(e_1)})^n\big) + H\big((X^{(e_3)})^n\big) + \big(\frac{5}{2} + 2\big(\frac{h(\lambda)}{n} + (\lambda-1)R^{(1\to\{2,3\})} + \varepsilon\big)\big)n \tag{C.6}$$

Satisfying both (C.2) and (C.6) requires that

$$\begin{aligned}
&H\big((X^{(e_1)})^n\big) + H\big((X^{(e_3)})^n\big) + \big(\frac{5}{2} + 2\big(\frac{h(\lambda)}{n} + (\lambda-1)R^{(1\to\{2,3\})} + \varepsilon\big)\big)n\\
&> \frac{1}{2}\Big(H\big((X^{(e_1)})^n\big) + H\big((X^{(e_3)})^n\big)\Big) + \big(\frac{7}{4} - \varepsilon\big)n
\end{aligned}$$

which is equivalent to

$$H\big((X^{(e_1)})^n\big) + H\big((X^{(e_3)})^n\big) > \big(4R^{(1\to\{2,3\})} - \frac{3}{2} - 4\frac{h(\lambda)}{n} - 4\lambda R^{(1\to\{2,3\})} - 6\varepsilon\big)n.$$

Since $R^{(1\to\{2,3\})} > 0.875$ by assumption, achieving the above bound for all $\lambda, \varepsilon > 0$ requires

$$H\big((X^{(e_1)})^n\big) + H\big((X^{(e_3)})^n\big) > 2n,$$

which is the desired contradiction since $(X^{(e_1)})^n$ and $(X^{(e_3)})^n$ are binary vectors of dimension n, giving $H\big((X^{(e_1)})^n\big) + H\big((X^{(e_3)})^n\big) \leq 2n$. □

Appendix D

Defining Typical Sets $\hat{A}_{\epsilon_1,t}^{(N)}(\tilde{X},\tilde{Z})$ and $\hat{A}_{\epsilon_2,t}^{(N)}(\tilde{X},\tilde{Y},\tilde{Z})$

The following analysis is taken out of [32] pages $34-38$. We reproduce it here and then extend it to investigate the impact of having an eavesdropper within the network. The following definitions are used in the proof of Theorem 9 in Section 3.4. Given any vectors $\epsilon_1 = (\epsilon_1(1),\ldots,\epsilon_1(n))$ and $\epsilon_2 = (\epsilon_2(1),\ldots,\epsilon_2(n))$ with $\epsilon_1(t) > 0$ and $\epsilon_2(t) > 0$ for all $t \in \{1,\ldots,n\}$, let

$$A_{\epsilon_1,t}^{(N)}(\tilde{X},\tilde{Z}) = \Bigg\{ (\underline{\tilde{x}},\underline{\tilde{z}}) \in \underline{\tilde{\mathcal{X}}} \times \underline{\tilde{\mathcal{Z}}} : \left| -\frac{1}{N}\log p_t(\underline{\tilde{x}}) - H_t(\tilde{X}) \right| \leq \epsilon_1(t) \\ \left| -\frac{1}{N}\log p_t(\underline{\tilde{z}}) - H_t(\tilde{Z}) \right| \leq a_1(\epsilon_1,t) \\ \left| -\frac{1}{N}\log p_t(\underline{\tilde{x}},\underline{\tilde{z}}) - H_t(\tilde{X},\tilde{Z}) \right| \leq a_1(\epsilon_1,t) \Bigg\},$$

where

$$a_1(\epsilon_1,t) \stackrel{\text{def}}{=} (1+\epsilon_1(t)) \cdot \inf \Bigg\{ \epsilon' > 0 : \text{ for all } N \text{ sufficiently large,} \\ \Pr\left(\left| -\frac{1}{N}\log p_t(\underline{\tilde{z}}) - H_t(\tilde{Z}) \right| > \epsilon' \cup \left| -\frac{1}{N}\log p_t(\underline{\tilde{x}},\underline{\tilde{z}}) - H_t(\tilde{X},\tilde{Z}) \right| > \epsilon' \right) \leq 2^{-6N\epsilon_1(t)} \Bigg\}.$$

As in [32] the restricted typical set $\hat{A}_{\epsilon_1,t}^{(N)}(\tilde{X},\tilde{Z})$, which is henceforth called simply the typical set, is defined as

$$\hat{A}_{\epsilon_1,t}^{(N)}(\tilde{X},\tilde{Z}) = \left\{ (\underline{\tilde{x}},\underline{\tilde{z}}) \in A_{\epsilon_1,t}^{(N)}(\tilde{X},\tilde{Z}) : p_t\left((A_{\epsilon_1,t}^{(N)}(\tilde{X},\tilde{Z}))^c \,|x \right) \leq 2^{-3N\epsilon_1(t)} \right\} \tag{D.1}$$

for each $t \in \{1,\ldots,n\}$. Similarly, let

$$A_{\epsilon_2,t}^{(N)}(\tilde{X},\tilde{Y},\tilde{Z}) = \Bigg\{ \left(\underline{\tilde{x}},\underline{\tilde{y}},\underline{\tilde{z}}\right) \in \underline{\mathcal{X}} \times \underline{\mathcal{Y}} \times \underline{\mathcal{Z}} : \left| -\frac{1}{N}\log p_t(\underline{\tilde{z}}) - H_t(\tilde{Z}) \right| \leq a_2(\epsilon_2,t)$$

$$\left|-\frac{1}{N}\log p_t(\underline{\tilde{x}},\underline{\tilde{z}}) - H_t(\tilde{X},\tilde{Z})\right| \le a_2(\epsilon_2,t)$$
$$\left|-\frac{1}{N}\log p_t(\underline{\tilde{y}},\underline{\tilde{z}}) - H_t(\tilde{Y},\tilde{Z})\right| \le a_2(\epsilon_2,t)$$
$$\left|-\frac{1}{N}\log p_t(\underline{\tilde{x}},\underline{\tilde{y}},\underline{\tilde{z}}) - H_t(\tilde{X},\tilde{Y},\tilde{Z})\right| \le a_2(\epsilon_2,t)\Bigg\},$$

where

$$a_2(\epsilon_2,t) \stackrel{\text{def}}{=} (1+\epsilon_2(t))\cdot \inf\Bigg\{\epsilon' > 0: \text{ for all } N \text{ sufficiently large,}$$
$$\Pr\Bigg(\left|-\frac{1}{N}\log p_t(\underline{\tilde{z}}) - H_t(\tilde{Z})\right| > \epsilon' \cup \left|-\frac{1}{N}\log p_t(\underline{\tilde{y}},\underline{\tilde{z}}) - H_t(\tilde{Y},\tilde{Z})\right| > \epsilon' \cup$$
$$\left|-\frac{1}{N}\log p_t(\underline{\tilde{x}},\underline{\tilde{z}}) - H_t(\tilde{X},\tilde{Z})\right| > \epsilon' \cup \left|-\frac{1}{N}\log p_t(\underline{\tilde{x}},\underline{\tilde{y}},\underline{\tilde{z}}) - H_t(\tilde{X},\tilde{Y},\tilde{Z})\right| > \epsilon'\Bigg) \le 2^{-6N\epsilon_2(t)}\Bigg\}.$$

and define the restricted typical set, henceforth simply called the typical set, as

$$\hat{A}^{(N)}_{\epsilon_2,t}(\tilde{X},\tilde{Y},\tilde{Z}) = \left\{(\underline{\tilde{x}},\underline{\tilde{y}},\underline{\tilde{z}}) \in A^{(N)}_{\epsilon_2,t}(\tilde{X},\tilde{Y},\tilde{Z}) : p_t\left((A^{(N)}_{\epsilon_2,t}(\tilde{X},\tilde{Y},\tilde{Z}))^c\,|\underline{x}\right) \le 2^{-3N\epsilon_2(t)}\right\}. \tag{D.2}$$

Appendix E

Bounds on the Probabilities of $\underline{\hat{\mathcal{N}}}_{\bar{e}}(R_c, R_p)$ Network

Lemma 8. *If* $(\underline{\tilde{x}}, \underline{\tilde{y}}) \in \hat{A}^{(N)}_{\epsilon_1,t}(\tilde{X}, \tilde{Y})$ *and* $(\underline{\tilde{x}}, \underline{\tilde{y}}, \underline{\tilde{z}}) \in \hat{A}^{(N)}_{\epsilon_2,t}(\tilde{X}, \tilde{Y}, \tilde{Z})$ *for all* $t \in \{1, \ldots, n\}$ *then,*

$$p(\underline{\tilde{w}}, (\underline{\tilde{z}}^E)^n, J = 0)2^{-N\sum_{t=1}^{n} b(\epsilon_1,\epsilon_2,t)} \leq \hat{p}(\underline{\tilde{w}}, (\underline{\tilde{z}}^E)^n, J = 0) \leq p(\underline{\tilde{w}}, (\underline{\tilde{z}}^E)^n, J = 0)2^{N\sum_{t=1}^{n} b(\epsilon_1,\epsilon_2,t)}$$

and

$$p((\underline{\tilde{z}}^E)^n, J = 0)2^{-N\sum_{t=1}^{n} b(\epsilon_1,\epsilon_2,t)} \leq \hat{p}((\underline{\tilde{z}}^E)^n, J = 0) \leq p((\underline{\tilde{z}}^E)^n, J = 0)2^{N\sum_{t=1}^{n} b(\epsilon_1,\epsilon_2,t)}$$

where the definition of variable J *is given in equation (3.2) and* $b(\epsilon_1, \epsilon_2, t) = 4a_1(\epsilon_1, t) + 8a_2(\epsilon_2, t) + 2\epsilon_1(t) + 2/N$.

Proof. In the rest of the proof we need the following definitions

$$K_t(\underline{\tilde{x}}, \underline{\tilde{z}}) = \begin{cases} 1, & \text{if } (\underline{\tilde{x}}, \underline{\tilde{z}}) \in \hat{A}^{(N)}_{\epsilon_1,t}(\tilde{X}, \tilde{Z}) \\ 0, & \text{otherwise} \end{cases}$$

$$K_t(\underline{\tilde{x}}, \underline{\tilde{y}}, \underline{\tilde{z}}) = \begin{cases} 1, & \text{if } (\underline{\tilde{x}}_t, \underline{\tilde{z}}_t) \in \hat{A}^{(N)}_{\epsilon_1,t}(\tilde{X}, \tilde{Z}) \text{ and } (\underline{\tilde{x}}_t, \underline{\tilde{y}}_t, \underline{\tilde{z}}_t) \in \hat{A}^{(N)}_{\epsilon_2,t}(\tilde{X}, \tilde{Y}, \tilde{Z}) \\ 0, & \text{otherwise} \end{cases}$$

and

$$q_t(\underline{\tilde{x}}) = \sum_{\underline{\tilde{z}} \in \underline{\mathcal{Z}}} K_t(\underline{\tilde{x}}, \underline{\tilde{z}}) p_t(\underline{\tilde{z}})$$

$$q_t(\underline{\tilde{x}}, \underline{\tilde{z}}) = \sum_{\underline{\tilde{y}} \in \underline{\mathcal{Y}}} K_t(\underline{\tilde{x}}, \underline{\tilde{y}}, \underline{\tilde{z}}) p_t(\underline{\tilde{y}} | \underline{\tilde{z}}).$$

Finally we define

$$F_t^{(1)}(\underline{\tilde{x}}) = \left\{ \underline{\tilde{z}} : (\underline{\tilde{x}}, \underline{\tilde{z}}) \in \hat{A}_{\epsilon_1,t}^{(N)}(\tilde{X}, \tilde{Z}) \right\}$$

for all $\underline{\tilde{x}} \in \underline{\mathcal{X}}$ such that $\left| -\frac{1}{N} \log p_t(\underline{\tilde{x}}) - H_t(\tilde{X}) \right| \leq \epsilon_1(t)$ and $p_t \left[\left(A_{\epsilon_1,t}^{(N)}(\tilde{X}, \tilde{Z}) \right)^c \middle| \underline{\tilde{x}} \right] < 2^{-3N\epsilon_1(t)}$. Similarly

$$F_t^{(2)}(\underline{\tilde{x}}, \underline{\tilde{z}}) = \left\{ \underline{\tilde{y}} : (\underline{\tilde{x}}, \underline{\tilde{y}}, \underline{\tilde{z}}) \in \hat{A}_{\epsilon_2,t}^{(N)}(\tilde{X}, \tilde{Y}, \tilde{Z}) \right\}$$

for all $(\underline{\tilde{x}}, \underline{\tilde{z}}) \in \hat{A}_{\epsilon_1,t}^{(N)}(\tilde{X}, \tilde{Z})$ and $p_t \left[\left(\hat{A}_{\epsilon_2,t}^{(N)}(\tilde{X}, \tilde{Y}, \tilde{Z}) \right)^c \middle| \underline{\tilde{x}}, \underline{\tilde{z}} \right] \leq 2^{-3N\epsilon_2(t)}$. Then it is easy to get

$$\begin{aligned}
&\sum_{\underline{\tilde{z}} \in F_t^{(1)}(\underline{\tilde{x}})} p(\underline{\tilde{z}}|\underline{\tilde{x}}) \leq 1 \\
\Rightarrow &\sum_{\underline{\tilde{z}} \in F_t^{(1)}(\underline{\tilde{x}})} \frac{p_t(\underline{\tilde{x}}, \underline{\tilde{z}})}{p_t(\underline{\tilde{x}})} \leq 1 \\
\overset{(a)}{\Rightarrow} &\sum_{\underline{\tilde{z}} \in F_t^{(1)}(\underline{\tilde{x}})} \frac{2^{-N(H_t(\tilde{X},\tilde{Z}) + a_1(\epsilon_1,t))}}{2^{-N(H_t(\tilde{X}) - \epsilon_1(t))}} \leq 1 \\
\Rightarrow &|F_t^{(1)}(\underline{\tilde{x}})| 2^{-N(H_t(\tilde{Z}|\tilde{X}) + a_1(\epsilon_1,t) + \epsilon_1(t))} \leq 1 \\
\Rightarrow &|F_t^{(1)}(\underline{\tilde{x}})| \leq 2^{N(H_t(\tilde{Z}|\tilde{X}) + a_1(\epsilon_1,t) + \epsilon_1(t))}
\end{aligned} \tag{E.1}$$

where inequality (a) follows from (D.1) that gives the definition of set $\hat{A}_{\epsilon_1,t}^{(N)}(\tilde{X}, \tilde{Z})$. Moreover

$$\begin{aligned}
q_t(\underline{\tilde{x}}) \overset{\text{def}}{=} \sum_{\underline{\tilde{z}} \in F_t^{(1)}(\underline{\tilde{x}})} p_t(\underline{\tilde{y}}) &\overset{(a)}{\leq} |F_t^{(1)}(\underline{\tilde{x}})| 2^{-N(H_t(\tilde{Y}) - a_1(\epsilon_1,t))} \overset{(b)}{\leq} 2^{-N(I_t(\tilde{X};\tilde{Y}) - 2a_1(\epsilon_1,t) - \epsilon_1(t))} \\
&\Rightarrow \frac{1}{q_t(\underline{\tilde{x}})} \geq 2^{N(I_t(\tilde{X};\tilde{Y}) - 2a_1(\epsilon_1,t) - \epsilon_1(t))}
\end{aligned} \tag{E.2}$$

where (a) follows from (D.1) and (b) holds due to (E.1). In order to find a lower bound on $\hat{p}_t(\underline{\tilde{z}}|\underline{\tilde{x}})$ we use the expression for $\hat{p}_t(\underline{\tilde{z}}|\underline{\tilde{x}})$ given by Lemma 7 of [32] and by following the steps below

$$\begin{aligned}
&\hat{p}_t(\underline{\tilde{z}}|\underline{\tilde{x}}) = p_t(\underline{\tilde{y}}) \frac{1 - (1 - q_t(\underline{\tilde{x}}))^{2^{NR_p}}}{q_t(\underline{\tilde{x}})} \overset{(a)}{\geq} \frac{1}{2} \frac{p_t(\underline{\tilde{z}})}{q_t(\underline{\tilde{x}})} \\
\Rightarrow &\hat{p}_t(\underline{\tilde{z}}|\underline{\tilde{x}}) \geq \frac{1}{2} p(\underline{\tilde{y}}|\underline{\tilde{x}}) \frac{p_t(\underline{\tilde{x}})}{p_t(\underline{\tilde{x}}, \underline{\tilde{z}})} \frac{p_t(\underline{\tilde{z}})}{q_t(\underline{\tilde{x}})} \\
\Rightarrow &\hat{p}_t(\underline{\tilde{z}}|\underline{\tilde{x}}) \overset{(b)}{\geq} \frac{1}{2} p(\underline{\tilde{z}}|\underline{\tilde{x}}) \frac{2^{-N(H_t(\tilde{X}) + \epsilon_1(t))} 2^{-N(H_t(\tilde{Y}) + a_1(\epsilon_1,t))}}{2^{-N(H_t(\tilde{X},\tilde{Y}) - a_1(\epsilon_1,t))} q_t(\underline{\tilde{x}})} \\
\Rightarrow &\hat{p}_t(\underline{\tilde{z}}|\underline{\tilde{x}}) \overset{(c)}{\geq} \frac{1}{2} p(\underline{\tilde{z}}|\underline{\tilde{x}}) 2^{N(-I_t(\tilde{X};\tilde{Y}) - \epsilon_1(t) - 2a_1(\epsilon_1,t))} 2^{N(I_t(\tilde{X};\tilde{Y}) - \epsilon_1(t) - 2a_1(\epsilon_1,t))} \\
\Rightarrow &\hat{p}_t(\underline{\tilde{z}}|\underline{\tilde{x}}) \geq p(\underline{\tilde{z}}|\underline{\tilde{x}}) 2^{-N(4a_1(\epsilon_1,t) + 2\epsilon_1(t) + 1/N))}
\end{aligned}$$

where inequality (a) holds for sufficiently large N, and inequalities (b), (c) hold due to (D.1) and (E.2) respectively. Following a similar approach it was proved at Lemma 16 of [32] the other side of the above inequality and finally we have

$$p(\underline{\tilde{z}}|\underline{\tilde{x}})\,2^{-N(4a_1(\epsilon_1,t)+2\epsilon_1(t)+1/N))} \leq \hat{p}_t(\underline{\tilde{z}}|\underline{\tilde{x}}) \leq p(\underline{\tilde{z}}|\underline{\tilde{x}})\,2^{N(4a_1(\epsilon_1,t)+2\epsilon_1(t)+1/N))} \tag{E.3}$$

for all $(\underline{\tilde{x}},\underline{\tilde{z}}) \in \hat{A}^{(N)}_{\epsilon_1,t}(\tilde{X},\tilde{Z})$ and for sufficiently large N.

We can derive a similar expression for $\hat{p}(\underline{\tilde{y}}|\underline{\tilde{x}},\underline{\tilde{z}})$, indeed

$$\begin{aligned}
&\sum_{\underline{\tilde{y}}\in F_t^{(2)}(\underline{\tilde{x}},\underline{\tilde{z}})} p(\underline{\tilde{y}}|\underline{\tilde{x}},\underline{\tilde{z}}) \leq 1\\
\Rightarrow &\sum_{\underline{\tilde{y}}\in F_t^{(2)}(\underline{\tilde{x}},\underline{\tilde{z}})} \frac{p_t(\underline{\tilde{x}},\underline{\tilde{y}},\underline{\tilde{z}})}{p_t(\underline{\tilde{x}},\underline{\tilde{z}})} \leq 1\\
\overset{(a)}{\Rightarrow} &\sum_{\underline{\tilde{y}}\in F_t^{(2)}(\underline{\tilde{x}},\underline{\tilde{z}})} \frac{2^{-N(H_t(\tilde{X},\tilde{Y},\tilde{Z})+a_2(\epsilon_2,t))}}{2^{-N(H_t(\tilde{X},\tilde{Z})-a_2(\epsilon_2,t))}} \leq 1\\
\Rightarrow &|F_t^{(2)}(\underline{\tilde{x}},\underline{\tilde{z}})|2^{-N(H_t(\tilde{Y}|\tilde{X},\tilde{Z})+2a_2(\epsilon_2,t))} \leq 1\\
\Rightarrow &|F_t^{(2)}(\underline{\tilde{x}},\underline{\tilde{z}})| \leq 2^{N(H_t(\tilde{Y}|\tilde{X},\tilde{Z})+2a_2(\epsilon_2,t))}
\end{aligned} \tag{E.4}$$

where inequality (a) follows from (D.2) that gives the definition of set $\hat{A}^{(N)}_{\epsilon_2,t}(\tilde{X},\tilde{Y},\tilde{Z})$. Moreover

$$\begin{aligned}
q_t(\underline{\tilde{x}},\underline{\tilde{z}}) &\overset{\text{def}}{=} \sum_{\underline{\tilde{y}}\in F_t^{(2)}(\underline{\tilde{x}},\underline{\tilde{z}})} p_t(\underline{\tilde{y}}|\underline{\tilde{z}}) \leq \sum_{\underline{\tilde{y}}\in F_t^{(2)}(\underline{\tilde{x}},\underline{\tilde{z}})} p_t(\underline{\tilde{y}}|\underline{\tilde{z}}) \leq \sum_{\underline{\tilde{y}}\in F_t^{(2)}(\underline{\tilde{x}},\underline{\tilde{z}})} \frac{p_t(\underline{\tilde{y}},\underline{\tilde{z}})}{p_t(\underline{\tilde{z}})} \overset{(a)}{\leq} \sum_{\underline{\tilde{y}}\in F_t^{(2)}(\underline{\tilde{x}},\underline{\tilde{z}})} \frac{2^{-N(H_t(\tilde{Y},\tilde{Z})-a_2(\epsilon_2,t))}}{2^{-N(H_t(\tilde{Z})+a_2(\epsilon_2,t))}}\\
&\leq |F_t^{(2)}(\underline{\tilde{x}},\underline{\tilde{z}})|2^{-N(H_t(\tilde{Y}|\tilde{Z})-2a_2(\epsilon_2,t))} \overset{(b)}{\leq} 2^{N(H_t(\tilde{Y}|\tilde{X},\tilde{Z})+2a_2(\epsilon_2,t))}2^{-N(H_t(\tilde{Y}|\tilde{Z})-2a_2(\epsilon_2,t))}\\
&\Rightarrow \frac{1}{q_t(\underline{\tilde{x}},\underline{\tilde{z}})} \geq 2^{N(I(\tilde{X};\tilde{Y}|\tilde{Z})-4a_2(\epsilon_2,t))}
\end{aligned} \tag{E.5}$$

where (a) follows from (D.2) and (b) holds due to (E.4). In order to find a lower bound on $\hat{p}_t(\underline{\tilde{y}}|\underline{\tilde{x}},\underline{\tilde{z}})$ we use an expression for $\hat{p}_t(\underline{\tilde{y}}|\underline{\tilde{x}},\underline{\tilde{z}})$ similar to the one proved in Lemma 7 of [31] and by following the steps below

$$\begin{aligned}
&\hat{p}_t(\underline{\tilde{y}}|\underline{\tilde{x}},\underline{\tilde{z}}) = p_t(\underline{\tilde{y}}|\underline{\tilde{z}})\frac{1-(1-q_t(\underline{\tilde{x}},\underline{\tilde{z}}))^{2^{NR_c}}}{q_t(\underline{\tilde{x}},\underline{\tilde{z}})} \overset{(a)}{\geq} \frac{1}{2}\frac{p_t(\underline{\tilde{y}}|\underline{\tilde{z}})}{q_t(\underline{\tilde{x}},\underline{\tilde{z}})}\\
\Rightarrow &\hat{p}_t(\underline{\tilde{y}}|\underline{\tilde{x}},\underline{\tilde{z}}) \geq \frac{1}{2}p(\underline{\tilde{y}}|\underline{\tilde{x}},\underline{\tilde{z}})\frac{p_t(\underline{\tilde{y}},\underline{\tilde{z}})\,p_t(\underline{\tilde{x}},\underline{\tilde{z}})}{p_t(\underline{\tilde{z}})\,p_t(\underline{\tilde{x}},\underline{\tilde{y}},\underline{\tilde{z}})}\frac{1}{q_t(\underline{\tilde{x}},\underline{\tilde{z}})}\\
\Rightarrow &\hat{p}_t(\underline{\tilde{y}}|\underline{\tilde{x}},\underline{\tilde{z}}) \overset{(b)}{\geq} \frac{1}{2}p(\underline{\tilde{y}}|\underline{\tilde{x}},\underline{\tilde{z}})\frac{2^{-N(H_t(\tilde{Y},\tilde{Z})+a_2(\epsilon_2,t))}2^{-N(H_t(\tilde{X},\tilde{Z})+a_2(\epsilon_2,t))}}{2^{-N(H_t(\tilde{Z})-a_2(\epsilon_2,t))}2^{-N(H_t(\tilde{X},\tilde{Y},\tilde{Z})-a_2(\epsilon_2,t))}q_t(\underline{\tilde{x}},\underline{\tilde{z}})}\\
\Rightarrow &\hat{p}_t(\underline{\tilde{y}}|\underline{\tilde{x}},\underline{\tilde{z}}) \overset{(c)}{\geq} \frac{1}{2}p(\underline{\tilde{y}}|\underline{\tilde{x}},\underline{\tilde{z}})2^{N(-I_t(\tilde{X};\tilde{Y}|\tilde{Z})-4a_2(\epsilon_2,t))}2^{N(I_t(\tilde{X};\tilde{Y}|\tilde{Z})-4a_2(\epsilon_2,t))}\\
\Rightarrow &\hat{p}_t(\underline{\tilde{y}}|\underline{\tilde{x}},\underline{\tilde{z}}) \geq p(\underline{\tilde{y}}|\underline{\tilde{x}},\underline{\tilde{z}})2^{-N(8a_2(\epsilon_2,t)+1/N))}
\end{aligned}$$

where inequality (a) holds for sufficiently large N, and inequalities (b), (c) hold due to (D.2) and (E.5) respectively. Following a similar approach it was proved at Lemma 16 of [32] the other side of the above inequality and finally we have

$$p(\underline{\tilde{y}}|\underline{\tilde{x}},\underline{\tilde{z}})\,2^{-N(8a_2(\epsilon_2,t)+1/N))} \leq \hat{p}_t(\underline{\tilde{y}}|\underline{\tilde{x}},\underline{\tilde{z}}) \leq (\underline{\tilde{y}}|\underline{\tilde{x}},\underline{\tilde{z}})\,2^{N(8a_2(\epsilon_2,t)+1/N))} \tag{E.6}$$

for all $(\underline{\tilde{x}},\underline{\tilde{y}},\underline{\tilde{z}}) \in \hat{A}^{(N)}_{\epsilon_2,t}(\tilde{X},\tilde{Y},\tilde{Z})$ and for large enough N.

Assume that we denote by $\underline{\tilde{x}}^t$, $\underline{\tilde{y}}^t$, and $\underline{\tilde{z}}^t$ all channel inputs, channel outputs, and eavesdropper outputs for all channels in the network for all times up to t. Then by an analysis similar to the one in Step 4 of [32] we get

$$p(\underline{w},\underline{\tilde{x}}^t,\underline{\tilde{y}}^t,\underline{\tilde{z}}^t) = p(\underline{w})\left[\prod_{t'=1}^{t} p(\underline{\tilde{x}}_{t'}|\underline{\tilde{y}}^{t'-1},\underline{w})\right]\left[\prod_{t'=1}^{t}\hat{p}(\underline{\tilde{z}}_{t'}|\underline{\tilde{x}}_{t'})\right]\left[\prod_{t'=1}^{t} p(\underline{\tilde{y}}_{t'}|\underline{\tilde{x}}_{t'})\right]. \tag{E.7}$$

for the random N-layer stacked solution $(2^{-N\delta},\varepsilon,R)$–$\underline{\mathcal{S}}(\underline{\mathcal{N}},A)$ for network $\underline{\mathcal{N}}$. Similarly for network $\underline{\mathcal{N}}_{\bar{e}}(R_c,R_p)$ we have

$$\hat{p}(\underline{w},\underline{\tilde{x}}^t,\underline{\tilde{y}}^t,\underline{\tilde{z}}^t) = p(\underline{w})\left[\prod_{t'=1}^{t} p(\underline{\tilde{x}}_{t'}|\underline{\tilde{y}}^{t'-1},\underline{w})\right]\left[\prod_{t'=1}^{t} p(\underline{\tilde{z}}_{t'}|\underline{\tilde{x}}_{t'})\right]\left[\prod_{t'=1}^{t}\hat{p}(\underline{\tilde{y}}_{t'}|\underline{\tilde{x}}_{t'})\right]. \tag{E.8}$$

that is different from equation (E.7) in terms $\hat{p}(\underline{\tilde{z}}_{t'}|\underline{\tilde{x}}_{t'})$ and $\hat{p}(\underline{\tilde{y}}_{t'}|\underline{\tilde{x}}_{t'})$ since due to the emulation the conditional distributions at the emulated channel change.

Therefore if $\forall t \in \{1,\ldots,n\}$ $(\underline{\tilde{x}}_t,\underline{\tilde{y}}_t,\underline{\tilde{z}}_t) \in \hat{A}^{(N)}_{\epsilon_2,t}(\tilde{X},\tilde{Y},\tilde{Z})$ and $(\underline{\tilde{x}}_t,\underline{\tilde{z}}_t) \in \hat{A}^{(N)}_{\epsilon_1,t}(\tilde{X},\tilde{Z})$ then by inequalities (E.3), (E.6), (E.7) and (E.8) we get

$$p(\underline{w},\underline{\tilde{x}}^t,\underline{\tilde{y}}^t,\underline{\tilde{z}}^t)2^{-\sum_{t'=1}^{t} b(\epsilon_1,\epsilon_2,t')} \leq \hat{p}(\underline{w},\underline{\tilde{x}}^t,\underline{\tilde{y}}^t,\underline{\tilde{z}}^t) \leq p(\underline{w},\underline{\tilde{x}}^t,\underline{\tilde{y}}^t,\underline{\tilde{z}}^t)2^{\sum_{t'=1}^{t} b(\epsilon_1,\epsilon_2,t')}$$

by defining $b(\epsilon_1,\epsilon_2,t) = 4a_1(\epsilon_1,t)+8a_2(\epsilon_2,t)+2\epsilon_1(t)+2/N$. Thus by setting $t=n$ and by summing over all $(\underline{\tilde{x}}^t,\underline{\tilde{y}}^t)$ with $(\underline{\tilde{x}}_t,\underline{\tilde{z}}_t) \in \hat{A}^{(N)}_{\epsilon_1(t),t}(\tilde{X},\tilde{Z})$ and $(\underline{\tilde{x}}_t,\underline{\tilde{y}}_t,\underline{\tilde{z}}_t) \in \hat{A}^{(N)}_{\epsilon_2(t),t}(\tilde{X},\tilde{Y},\tilde{Z})$ for all $t \in \{1,\ldots,n\}$ and over all $\underline{\tilde{z}}^{(v)}$ where $v \notin E$ (nodes that do not get eavesdropped) we get

$$p(\underline{w},\underline{\tilde{z}}^E,I=0)2^{-\sum_{t=1}^{n} b(\epsilon_1,\epsilon_2,t)} \leq \hat{p}(\underline{w},\underline{\tilde{z}}^E,I=0) \leq p(\underline{w},\underline{\tilde{z}}^E,I=0)2^{\sum_{t=1}^{n} b(\epsilon_1,\epsilon_2,t)} \tag{E.9}$$

where the definition of variable I is given in equation (3.2) and similarly

$$p(\underline{\tilde{z}}^E,I=0)2^{-\sum_{t=1}^{n} b(\epsilon_1,\epsilon_2,t)} \leq \hat{p}(\underline{\tilde{z}}^E,I=0) \leq p(\underline{\tilde{z}}^E,I=0)2^{\sum_{t=1}^{n} b(\epsilon_1,\epsilon_2,t)}. \tag{E.10}$$

and this concludes the proof. □

Appendix F

Filling up the Bit Pipes

Lemma 9. *Fix any rate vector $R \in int\big(\mathcal{R}(\mathcal{N}_{\bar{e}}(R_c, R_p), A)\big)$ in the relative interior of rate region $\mathcal{R}(\mathcal{N}_{\bar{e}}(R_c, R_p), A)$ of network $\mathcal{N}_{\bar{e}}(R_c, R_p)$. Then for any $\mu < R_p$, $\lambda > 0$ and $\varepsilon > 0$ there exists a $(\lambda, \varepsilon, A, R)$–$\mathcal{S}(\mathcal{N}_{\bar{e}}(R_c, R_p))$ solution of blocklength n such that $H\big(P^n\big) \geq n(R_p - \mu)$ where P^n are the public bits through channel $\mathcal{C}_{\bar{e}}(R_c, R_p)$.*

Proof. Since $R \in \text{int}\big(\mathcal{R}(\mathcal{N}_{\bar{e}}(R_c, R_p), A)\big)$, for any $\lambda > 0$ and $\varepsilon > 0$, there exists a solution $(\lambda, \varepsilon, A, R)$–$\mathcal{S}(\mathcal{N}_{\bar{e}}(R_c, R_p))$ for network $\mathcal{N}_{\bar{e}}(R_c, R_p)$ under adversarial set A. If the entropy bound is satisfied, then the result is immediate. Otherwise for any we construct a new code of the same rate that satisfies the given entropy bounds. The code construction first builds a stacked solution $\underline{\mathcal{S}}(\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p))$ for N_1-fold stacked network $\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p)$, and then modifies that solution to build a new solution $\mathcal{S}(\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p))$ for $\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p)$ by adding extra randomness on the public part of channel of $\bar{e}$ to fill up the bit pipe. This new solution with the added randomness can be considered as a random code for network $\mathcal{N}_{\bar{e}}(R_c, R_p)$ with blocklength $n_1 = nN$ with small expected error, expected rate of information leakage and expected entropy on the public bhigh enough (the exact values of the expected error, information leakage and entropy of public bits will be chosen later). In order to prove the existence of at least one good code, we would create one more stack of N_2 layers from the random solution of blocklength n_1. We create this stack with each layer having high expected entropy on the public bit pipe of channel $\bar{e}$, to use the Chernoff bound (since each layer is independent) and prove the existence of at least a single good code with high entropy on the public part of channel $\bar{e}$ and small probability of error and information leakage. The details of the proof follow.

Choose two rates $\tilde{R} \in \text{int}\left(\mathcal{R}(\mathcal{N}_{\bar{e}}(R_c, R_p), A)\right)$ and $\tilde{\tilde{R}} \in \text{int}\left(\mathcal{R}(\mathcal{N}_{\bar{e}}(R_c, R_p), A)\right)$ such that $R^{(i\to\mathcal{B})} < \tilde{R}^{(i\to\mathcal{B})} < \tilde{\tilde{R}}^{(i\to\mathcal{B})}$ for all $i \in \mathcal{V}$ and $\mathcal{B} \in \mathcal{B}^{(i)}$ with $R^{(i\to\mathcal{B})} > 0$ and $R^{(i\to\mathcal{B})} = \tilde{R}^{(i\to\mathcal{B})} = 0$ for all $R^{(i\to\mathcal{B})} = 0$. As in the proof of Theorem 8 set $\tilde{\tilde{\rho}} = \min\limits_{R^{(i\to\mathcal{B})}>0} (\tilde{\tilde{R}}^{(i\to\mathcal{B})} - \tilde{R}^{(i\to\mathcal{B})})$ and $\tilde{\rho} = \min\limits_{R^{(i\to\mathcal{B})}>0} (\tilde{R}^{(i\to\mathcal{B})} - R^{(i\to\mathcal{B})})$ and find constants $\tilde{\tilde{\lambda}}$ and $\tilde{\lambda}$ satisfying

$$\max_{(i,\mathcal{B}):R^{(i\to\mathcal{B})}>0} \tilde{\tilde{R}}^{(i\to\mathcal{B})}\tilde{\tilde{\lambda}} + h(\tilde{\tilde{\lambda}}) < \tilde{\tilde{\rho}} \tag{F.1}$$

$$\max_{(i,\mathcal{B}):R^{(i\to\mathcal{B})}>0} \tilde{R}^{(i\to\mathcal{B})}\tilde{\lambda} + h(\tilde{\lambda}) < \tilde{\rho} \tag{F.2}$$

Then for $\tilde{\tilde{\lambda}} > 0$ and there is a blocklength n' secrecy code of rate $\tilde{\tilde{R}}$ for network $\mathcal{N}_{\bar{e}}(R_c, R_p)$ such that $\Pr\left(\tilde{\tilde{W}}^{(j\to\mathcal{K},i)} \neq \tilde{\tilde{W}}^{(j\to\mathcal{K})}\right) < \tilde{\tilde{\lambda}}$ for $\mathcal{K} \in \mathcal{B}^{(j)}$ such that $i \in \mathcal{K}$ and $R^{(j\to\mathcal{K})} > 0$ and $I(\tilde{\tilde{P}}^n; \tilde{\tilde{W}}) < \frac{n\varepsilon}{4}$, where the double tilde on the public bits, and the message refers to the fact that the code operates at rate $\tilde{\tilde{R}}$. It was proved in Theorem 8, that used Theorem 1 of [31] that since (F.1) holds we can use the single-layer solution $(\tilde{\tilde{\lambda}}, \frac{\varepsilon}{4}, A, \tilde{\tilde{R}})$–$\mathcal{S}\big(\mathcal{N}_{\bar{e}}(R_c, R_p)\big)$ for network $\mathcal{N}_{\bar{e}}(R_c, R_p)$ to build a random N_1 stacked solution $(2^{-N_1\delta_1}, \frac{\varepsilon}{4}, A, \tilde{R})$–$\underline{\mathcal{S}}\big(\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p)\big)$ for stacked network $\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p)$ where $\delta_1 > 0$.

Let $p\big(\underline{\tilde{\tilde{P}}}^{n'}\big)$ be the inputs of public bits for channel $\bar{e}$ induced by the random stacked solution $(\lambda'', \varepsilon', A, \tilde{R})$–$\underline{\mathcal{S}}\big(\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p)\big)$. These inputs are i.i.d. under the random code design giving

$$p\big(\underline{\tilde{\tilde{P}}}^{n'}\big) = \prod_{\ell=1}^{N} p\big(\underline{\tilde{\tilde{P}}}^{n'}(\ell)\big)$$

For each $t \in \{1, \ldots, n'\}$, we define the set of typical inputs for edge $\bar{e}$ as

$$A_{\mu'}^{(N_1)}(\tilde{\tilde{P}}^t) = \left\{\underline{\tilde{\tilde{P}}}^t : \left|-\frac{1}{N_1}\log p\big(\underline{\tilde{\tilde{P}}}^{t'}\big) - H\big(\tilde{\tilde{P}}^{t'}\big)\right| \leq \mu',\ 1 \leq t' \leq t\right\}, \tag{F.3}$$

and event J as

$$J = \begin{cases} 1, & \text{if } \underline{\tilde{\tilde{P}}}^{n'} \notin A_{\mu'}^{(N_1)}(\tilde{\tilde{P}}^{n'}) \\ 0, & \text{otherwise} \end{cases}.$$

Similar to the proof of Lemma 8 in [31], when $\big(\underline{\tilde{\tilde{P}}}^{n'}(1), \ldots, \underline{\tilde{\tilde{P}}}^{n'}(N_1)\big)$ are drawn i.i.d. from distribution $p\big(\underline{\tilde{\tilde{P}}}^{n'}\big)$, the probability of observing an atypical input $\underline{\tilde{\tilde{P}}}^{n'}$ in stacked solution $\underline{\mathcal{S}}\big(\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p)\big)$ drops exponentially in N_1. We choose N_1 large enough so that

$$\left\{\left(\Pr(J=1) < \min\big(\frac{\tilde{\lambda}}{2}, \frac{1}{n'R_p}\big)\right) \wedge \left(2^{-N_1\delta_1} < \frac{\tilde{\lambda}}{2}\right)\right\} \tag{F.4}$$

and treat the situation of observing an atypical $\underline{\tilde{\tilde{P}}}^{n'}$ as an error event.

For $2 \leq t \leq n'$ and every $\underline{\tilde{\tilde{P}}}^{t-1} \in A_{\mu'}^{(N_1)}\big(\underline{\tilde{\tilde{P}}}^{t-1}\big)$, define the conditional typical set as

$$A_{\mu'}^{(N_1)}\big(\underline{\tilde{\tilde{P}}}_t | \underline{\tilde{\tilde{P}}}^{t-1}\big) \stackrel{\text{def}}{=} \left\{\underline{\tilde{\tilde{P}}}_t : \underline{\tilde{\tilde{P}}}^t \in A_{\mu'}^{(N_1)}\big(\underline{\tilde{\tilde{P}}}^t\big)\right\}. \tag{F.5}$$

Then the cardinality of the conditional typical set $A_{\mu'}^{(N_1)}\big(\underline{\tilde{\tilde{P}}}_t | \underline{\tilde{\tilde{P}}}^{t-1}\big)$ for any $\underline{\tilde{\tilde{P}}}^{t-1} \in A_{\mu'}^{(N_1)}\big(\underline{\tilde{\tilde{P}}}^{t-1}\big)$ is

upper bounded as

$$\begin{aligned}
1 &\geq \sum_{\underline{\tilde{\tilde{P}}}_t \in A_{\mu'}^{(N_1)}\left(\tilde{\tilde{P}}_t|\tilde{\tilde{P}}^{t-1}\right)} p\left(\underline{\tilde{\tilde{P}}}_t|\underline{\tilde{\tilde{P}}}^{t-1}\right) \\
&= \sum_{\underline{\tilde{P}}_t \in A_{\mu'}^{(N_1)}\left(\tilde{\tilde{P}}_t|\tilde{\tilde{P}}^{t-1}\right)} \frac{p\left(\underline{\tilde{\tilde{P}}}^{t}\right)}{p\left(\underline{\tilde{\tilde{P}}}^{t-1}\right)} \\
&\overset{(a)}{\geq} \sum_{\underline{\tilde{P}}_t \in A_{\mu'}^{(N_1)}\left(\tilde{\tilde{P}}_t|\tilde{\tilde{P}}^{t-1}\right)} \frac{2^{-N_1[H(\tilde{\tilde{P}}^t)+\mu']}}{2^{-N[H(\tilde{\tilde{P}}^{t-1})-\mu']}} \\
&= \left|A_{\mu'}^{(N_1)}\left(\tilde{\tilde{P}}_t|\tilde{\tilde{P}}^{t-1}\right)\right| 2^{-N_1[H(\tilde{\tilde{P}}_t|\tilde{\tilde{P}}^{t-1})+2\mu']}
\end{aligned}$$

where inequality (a) applies follows from the definitions of the typical and conditional typical sets from (F.3), and (F.5); therefore

$$\left|A_{\mu'}^{(N_1)}\left(\tilde{\tilde{P}}_t|\tilde{\tilde{P}}^{t-1}\right)\right| \leq 2^{N_1[H(\tilde{\tilde{P}}_t|\tilde{\tilde{P}}^{t-1})+2\mu']} \tag{F.6}$$

Since $\left|A_{\mu'}^{(N_1)}\left(\tilde{\tilde{P}}_1\right)\right| \leq 2^{N_1[H(\tilde{\tilde{P}}_1)+\mu']} \leq 2^{N_1[H(\tilde{\tilde{P}}_1)+2\mu']}$, and therefore (F.6) holds for all $1 \leq t \leq n'$, where for notational convenience $\left(\tilde{\tilde{P}}\right)^0 = \emptyset$.

To build $\mathcal{S}(\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p))$ that has filled up the public bits pipe of channel $\bar{e}$, from solution $\underline{\mathcal{S}}(\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p))$ we replace any typical $\underline{\tilde{\tilde{P}}}_t \in \{0,1\}^{N_1}$ with a unique sequence of $n_t = \left\lceil N_1\left(H(\tilde{\tilde{P}}_t|\tilde{\tilde{P}}^{t-1}) + 2\mu'\right)\right\rceil$ bits that uniquely describe $\underline{\tilde{\tilde{P}}}_t$, and any atypical channel input $\underline{\tilde{\tilde{P}}}_t$ by a binary string chosen uniformly at random from $\{0,1\}^{n_t}$. The n_t bits are followed by $N_1 R_p - n_t$ independent random bits denoted by $\underline{\tilde{\tilde{F}}}_t$ that fill up the public bit pipe of channel $\bar{e}$. Let $\underline{\tilde{\tilde{Q}}}^{n'} = \left(G(\underline{\tilde{\tilde{P}}}^{n'}), \underline{\tilde{\tilde{F}}}^{n'}\right)$ denote the resulting bits where $G : \{0,1\}^{n' N_1} \to \{0,1\}^{\sum_{t=1}^{n'} n_t}$ is the function mapping $\underline{\tilde{\tilde{P}}}^{n'}$ to $\sum_{t=1}^{n'} n_t$ bit sequences and this function G is one-to-one for typical for typical inputs. Under solution $\mathcal{S}(\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p))$ instead of bits $\tilde{\tilde{P}}^{n'}$, bits $\underline{\tilde{\tilde{Q}}}^{n'}$ will be transmitted through the public bit pipe of channel $\bar{e}$. Code $\mathcal{S}(\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p))$ works identical to stacked code $\underline{\mathcal{S}}(\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p))$ anywhere in stacked network $\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p)$ apart from the public bit pipe of channel $\bar{e}$, where for each time step $1 \leq t \leq n'$, the receiver of bits $\underline{\tilde{\tilde{Q}}}_t$ reconstructs $\underline{\tilde{\tilde{P}}}_t$ by the looking at the first n_t bits of $\underline{\tilde{\tilde{Q}}}_t$; the case where $\underline{\tilde{\tilde{P}}}_t$ is not typical and the reconstruction will fail and this is considered as an error event.

We have to prove how the transmission of bits $\underline{\tilde{\tilde{Q}}}^{n'}$ instead of $\underline{\tilde{\tilde{P}}}^{n'}$ does not compromise the security of code $\mathcal{S}(\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p))$ compute its probability of error and entropy of the bits transversing the public bit pipe of channel $\bar{e}$. Indeed, the security of code $\mathcal{S}(\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p))$ is as good as that of stacked the solution $\underline{\mathcal{S}}(\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p))$ since $\underline{\tilde{W}} \to \tilde{\tilde{P}}^{n'} \to \tilde{\tilde{Q}}^{n'}$ and therefore due to the data processing inequality $I(\underline{\tilde{W}}; \tilde{\tilde{Q}}^{n'}) \leq I(\underline{\tilde{W}}; \tilde{\tilde{P}}^{n'}) \leq nN\varepsilon'$ (note that message $\underline{\tilde{W}}$ is of rate $\tilde{R}$ and that is why it has a single tilde). To compute the probability of error, we note that code $\mathcal{S}(\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p))$ works

identical to stacked code $(\frac{\tilde{\lambda}}{2}, \frac{\varepsilon}{4}, A, \tilde{R})$–$\underline{\mathcal{S}}(\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p))$ (where $\frac{\tilde{\lambda}}{2}$ holds due to (F.4)) anywhere in stacked network $\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p)$ apart from the public bit pipe of channel $\bar{e}$, where for each time step $1 \leq t \leq n'$, the receiver of bits $\underline{\tilde{\tilde{Q}}}_t$ reconstructs $\underline{\tilde{\tilde{P}}}_t$ by the looking at the first n_t bits of $\underline{\tilde{\tilde{Q}}}_t$. According to (F.4), with probability at most equal to $\frac{\tilde{\lambda}}{2}$, $\underline{\tilde{\tilde{P}}}_t$ is not typical and the reconstruction will fail. That is considered an error event and therefore the overall probability of failure for code $\mathcal{S}(\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p))$ is less than $\tilde{\lambda}$.

Now we will compute the entropy of bits $\underline{\tilde{\tilde{Q}}}^{n'}$ that transverse the public bit pipe of channel $\bar{e}$ on $\mathcal{S}(\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p))$. The expected entropy of typical $\underline{\tilde{\tilde{P}}}^{n'} \in A_{\mu'}^{(N_1)}(\tilde{\tilde{P}}^t)$ is bounded by

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}}\left[\Pr(J=0)\cdot H\big(\underline{\tilde{\tilde{P}}}^{n'}|J=0\big)\right] \\
&= \mathbb{E}_{\mathcal{C}}\left[H\big(\underline{\tilde{\tilde{P}}}^{n'}|J\big) - \Pr(J=1)H\big(\underline{\tilde{\tilde{P}}}^{n'}|J=1\big)\right] \\
&\overset{(a)}{\geq} \mathbb{E}_{\mathcal{C}}\big[H\big(\underline{\tilde{\tilde{P}}}^{n'}|J\big)\big] - \Pr(J=1)nR_p \\
&\overset{(b)}{\geq} \mathbb{E}_{\mathcal{C}}\big[H\big(\underline{\tilde{\tilde{P}}}^{n'}, J\big) - H\big(J\big)\big] - 1 \\
&\overset{(c)}{=} \mathbb{E}_{\mathcal{C}}\big[H\big(\underline{\tilde{\tilde{P}}}^{n'}\big)\big] - 2 \\
&\overset{(c)}{=} N_1\mathbb{E}_{\mathcal{C}}\big[H\big(\underline{\tilde{\tilde{P}}}^{n'}\big)\big] - 2
\end{aligned} \tag{F.7}$$

where (a) holds due to the capacity of the public bit pipe of channel $\bar{e}$, (b) follows from (F.4), and (c) holds since $H(J) \leq 1$ and $H\big(J|\underline{\tilde{\tilde{P}}}^{n'}\big) = 0$. The expected entropy of the random bits $\underline{\tilde{\tilde{F}}}^{n'}$ is

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}}\big[H(\underline{\tilde{\tilde{F}}}^{n'})\big] &= \sum_{t=1}^{n'} \mathbb{E}_{\mathcal{C}}\left[H(\underline{\tilde{\tilde{F}}}^{n'})\right] \\
&= \sum_{t=1}^{n'} \Big(N_1R_p - \big\lceil N_1\big(\mathbb{E}_{\mathcal{C}}\big[H(\underline{\tilde{\tilde{P}}}_t|\underline{\tilde{\tilde{P}}}^{t-1})\big] + 2\mu'\big)\big\rceil\Big) \\
&\overset{(a)}{\geq} \sum_{t=1}^{n'} \Big(N_1R_p - N_1\big(\mathbb{E}_{\mathcal{C}}\big[H(\underline{\tilde{\tilde{P}}}_t|\underline{\tilde{\tilde{P}}}^{t-1})\big] + 2\mu'\big) - 1\Big) \\
&= N_1n'(R_p - 2\mu') - N_1\mathbb{E}_{\mathcal{C}}\big[H(\underline{\tilde{\tilde{P}}}^{n'})\big] - n' \\
&\overset{(b)}{\geq} N_1n'(R_p - 3\mu') - N_1\mathbb{E}_{\mathcal{C}}\big[H(\underline{\tilde{\tilde{P}}}^{n'})\big]
\end{aligned} \tag{F.8}$$

where (a) follows since $\lceil x \rceil \leq x + 1$, and (b) holds for sufficiently large N_1.

Finally, for all N_1 sufficiently large, the expected entropy $\mathbb{E}_{\mathcal{C}}\big[H(\underline{\tilde{\tilde{Q}}}^{n'})\big]$ of the overall input $\underline{\tilde{\tilde{Q}}}^{n'} = \big(G(\underline{\tilde{\tilde{P}}}^{n'}), \underline{\tilde{\tilde{F}}}^{n'}\big)$ to the public bit pipe of channel $\bar{e}$ for code $\mathcal{S}(\underline{\mathcal{N}}_{\bar{e}}(R_c, R_p))$ is

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}}\big[H\big(\underline{\tilde{\tilde{Q}}}^{n'}\big)\big] \\
&= \mathbb{E}_{\mathcal{C}}\big[H\big(G(\underline{\tilde{\tilde{P}}}^{n'}), \underline{\tilde{\tilde{F}}}^{n'}\big)\big]
\end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{=} \mathbb{E}_{\mathcal{C}}\big[H\big(G(\underline{\tilde{\tilde{P}}}^{n'})\big)\big] + \mathbb{E}_{\mathcal{C}}\big[H\big(\underline{\tilde{\tilde{F}}}^{n'}\big)\big] \\
&\geq \mathbb{E}_{\mathcal{C}}\big[H\big(G(\underline{\tilde{\tilde{P}}}^{n'})|J\big)\big] + \mathbb{E}_{\mathcal{C}}\big[H\big(\underline{\tilde{\tilde{F}}}^{n'}\big)\big] \\
&\geq \mathbb{E}_{\mathcal{C}}\big[\Pr(J=0)\cdot H\big(G(\underline{\tilde{\tilde{P}}}^{n'})|J=0\big)\big] + \mathbb{E}_{\mathcal{C}}\big[H\big(\underline{\tilde{\tilde{F}}}^{n'}\big)\big] \\
&\overset{(b)}{=} \mathbb{E}_{\mathcal{C}}\big[\Pr(J=0)\cdot H\big(\underline{\tilde{\tilde{P}}}^{n'}|J=0\big)\big] + \mathbb{E}_{\mathcal{C}}\big[H\big(\underline{\tilde{\tilde{F}}}^{n'}\big)\big] \\
&\overset{(c)}{\geq} N_1\mathbb{E}_{\mathcal{C}}\big[H\big(\underline{\tilde{\tilde{P}}}^{n'}\big)\big] - 2 + N_1 n'(R_p - 3\mu') - N_1\mathbb{E}_{\mathcal{C}}\big[H\big(\underline{\tilde{\tilde{P}}}^{n'}\big)\big] \\
&\overset{(d)}{\geq} n'N_1(R_p - 4\mu')
\end{aligned}
\tag{F.9}
$$

where (a) holds since random bits $\underline{\tilde{\tilde{F}}}^{n'}$ are independent of $G(\underline{\tilde{\tilde{P}}}^{n'})$, (b) holds since when $J=0$ function G is one-to-one, (c) follows from (F.7) and (F.8), and (d) holds for sufficiently large N_1. Therefore code $\mathcal{S}(\underline{\mathcal{N}}_{\bar{e}}(R_c,R_p))$ has the public bit pipe of channel $\bar{e}$ filled up close to capacity.

So far we have constructed a random code $(\tilde{\lambda}, \frac{\varepsilon}{4}, A, \tilde{R})$–$\mathcal{S}(\underline{\mathcal{N}}_{\bar{e}}(R_c,R_p))$ of blocklength n' for stacked network $\underline{\mathcal{N}}_{\bar{e}}(R_c,R_p)$. Unraveling this blocklength n' solution across time as described in [31, Theorem 1] yields a blocklength $n_1 = n'N_1$ solution of rate $\tilde{R}$ for network $\mathcal{N}_{\bar{e}}(R_c,R_p)$.

By applying the blocklength $n_1 = n'N_1$ solution on a stacked version network $\mathcal{N}_{\bar{e}}(R_c,R_p)$ and due to (F.2) one can construct another random solution $(\frac{\lambda}{4}, \frac{\varepsilon}{4}, A, R)$–$\mathcal{S}(\underline{\mathcal{N}}_{\bar{e}}(R_c,R_p))$ for network $\underline{\mathcal{N}}_{\bar{e}}(R_c,R_p)$ of blocklength n_1 with for sufficiently large number of layers N_2. The expected probability of error and information leakage of this code is

$$
\mathbb{E}_{\mathcal{C}}[P_e^{(n)}] \leq \frac{\lambda}{4} \tag{F.10}
$$

$$
\mathbb{E}_{\mathcal{C}}\big[I\big((\underline{\tilde{Z}}^{(E)})^{n_1};\underline{W}\big)\big] \leq \frac{n'N_1N_2\varepsilon}{4} \tag{F.11}
$$

respectively. The entropy transversing the public bits of channel $\bar{e}$ is

$$
H\big(\underline{\tilde{\tilde{Q}}}^{n_1}(1),\ldots,\underline{\tilde{\tilde{Q}}}^{n_1}(N_2)\big) = \sum_{\ell=1}^{N_2} H\big(\underline{\tilde{\tilde{Q}}}^{n_1}(\ell)\big) \tag{F.12}
$$

since all the layers in the stack are independent. The expected value of each term in the sum is computed in (F.9) to be greater than $n'N_1(R_p-4\mu')$ and therefore the expected entropy for all layers is

$$
\mathbb{E}_{\mathcal{C}}[H(\underline{\tilde{\tilde{Q}}}^{n_1})] \geq n'N_1N_2(R_p-4\mu') \tag{F.13}
$$

Similar to Theorem 8 it remains to prove that there is a specific instance for the choice of each channel code such that the probability of error, the information leakage are not too large and the

entropy of the public bit pipe of channel $\bar{e}$ is not too low. Precisely,

$$\begin{aligned}
&\Pr\left(\left\{P_e^{(n)} \geq \lambda\right\} \cup \left\{I\big((\underline{\tilde{Z}}^{(E)})^{n_1}; \underline{W}\big) \geq n' N_1 N_2 \varepsilon\right\} \cup \left\{H\big(\underline{\tilde{\tilde{Q}}}^{n_1}\big) \leq n' N_1 N_2 (R_p - 5\mu')\right\}\right) \\
&\overset{(a)}{\leq} \Pr\left(P_e^{(n)} \geq \lambda\right) + \Pr\left(I\big((\underline{\tilde{Z}}^{(E)})^{n_1}; \underline{W}\big) \geq n' N_1 N_2 \varepsilon\right) + \Pr\left(H\big(\underline{\tilde{\tilde{Q}}}^{n_1}\big) \leq n' N_1 N_2 (R_p - 5\mu')\right) \\
&\overset{(b)}{\leq} \frac{\mathbb{E}_{\mathcal{C}}\big[P_e^{(n)}\big]}{\lambda} + \frac{\mathbb{E}_{\mathcal{C}}\big[I\big((\underline{\tilde{Z}}^{(E)})^{n_1}; \underline{W}\big)\big]}{n' N_1 N_2 \varepsilon} + \Pr\left(\sum_{\ell=1}^{N_2} H\big(\underline{\tilde{\tilde{Q}}}^{n_1}(\ell)\big) \leq n' N_1 N_2 (R_p - 5\mu')\right) \\
&\overset{(c)}{\leq} \frac{\lambda}{4\lambda} + \frac{n' N_1 N_2 \varepsilon}{4 n' N_1 N_2 \varepsilon} + 2^{-N_2 \gamma(\mu')} \\
&= \frac{1}{2} + 2^{-N_2 \gamma(\mu')} \overset{(d)}{<} \frac{2}{3},
\end{aligned}$$

where inequality (a) is the union bound, (b) is Markov's inequalitya along with (F.12), (c) applies bounds (F.10), (F.11) and equation (F.12), (c) holds is derived by using the Chernoff bound similar to Lemma 8 [31], and (d) holds for large enough N_2. Therefore, for sufficiently large N_2 there must be at least one instance of code $\mathcal{S}(\underline{\mathcal{N}}_{\bar{e}}(R - c, R_p))$ with error probability no greater than λ, information leakage no more than $n' N_1 N_2 \varepsilon$ and entropy of information on the public bit pipe greater than $n' N_1 N_2 (R_p - \mu')$. By setting $\mu' = \frac{\mu}{5}$ and unraveling this code in time to become a code of blocklength $n = n' N_1 N_2$ for network $\mathcal{N}_{\bar{e}}(R_c, R_p)$ concludes our proof. $\square$

Appendix G

Conditional Typical Sequences

Definition 16. *The typical set* $A_\omega^{(N_2)}$ *is the set of* $\left(\tilde{\underline{\ell}}^{n_1}, \tilde{\underline{w}}^{n_1}, \tilde{\underline{p}}^{n_1}, \tilde{\underline{c}}^{n_1}\right)$ *tuples with the property*

$$A_\omega^{(N_2)} = \left\{ \left(\tilde{\underline{\ell}}^{n_1}, \tilde{\underline{w}}^{n_1}, \tilde{\underline{p}}^{n_1}, \tilde{\underline{c}}^{n_1}\right), \forall \mathcal{S} \subseteq \{1, \ldots, n_1\} : \left| -\frac{1}{N_2} \log \Pr(\tilde{\underline{\ell}}^{n_1}, \tilde{\underline{w}}^{n_1}, \tilde{\underline{p}}_{\mathcal{S}}, \tilde{\underline{c}}_{\mathcal{S}}) - H(\tilde{L}^{\tilde{n}}, \tilde{W}^{\tilde{n}}, \tilde{P}_{\mathcal{S}}, \tilde{C}_{\mathcal{S}}) \right| \leq \omega \right\}.$$

Definition 17. *For every* $\mathcal{S} \subseteq \{1, \ldots, n_1\}$ *the conditional typical set* $A_\omega^{(N_2)}\left(\tilde{\underline{P}}_{\mathcal{S}^c}, \tilde{\underline{C}}_{\mathcal{S}^c} \mid \tilde{\underline{\ell}}^{n_1}, \tilde{\underline{w}}^{n_1}, \tilde{\underline{p}}_{\mathcal{S}}, \tilde{\underline{p}}_{\mathcal{S}}\right)$ *of* $(\tilde{\underline{P}}_{\mathcal{S}^c}, \tilde{\underline{C}}_{\mathcal{S}^c})$ *with respect to a specific sequence* $(\tilde{\underline{\ell}}^{n_1}, \tilde{\underline{w}}^{n_1}, \tilde{\underline{p}}_{\mathcal{S}}, \tilde{\underline{p}}_{\mathcal{S}})$ *that*

$$\left| -\frac{1}{N_2} \log \Pr(\tilde{\underline{\ell}}^{n_1}, \tilde{\underline{w}}^{n_1}, \tilde{\underline{p}}_{\mathcal{S}}, \tilde{\underline{p}}_{\mathcal{S}}) - H(\tilde{L}^{n_1}, \tilde{W}^{n_1}, \tilde{P}_{\mathcal{S}}, \tilde{C}_{\mathcal{S}}) \right| \leq \omega$$

is defined as

$$A_\omega^{(N_2)}\left(\tilde{P}_{\mathcal{S}^c}, \tilde{C}_{\mathcal{S}^c} \mid \tilde{\underline{\ell}}^{n_1}, \tilde{\underline{w}}^{n_1}, \tilde{\underline{p}}_{\mathcal{S}}, \tilde{\underline{c}}_{\mathcal{S}}\right) = \left\{ (\tilde{\underline{p}}_{\mathcal{S}^c}, \tilde{\underline{c}}_{\mathcal{S}^c}) : \ (\tilde{\underline{\ell}}^{n_1}, \tilde{\underline{w}}^{n_1}, \tilde{\underline{p}}^{n_1}, \tilde{\underline{c}}^{n_1}) \in A_\omega^{(N_2)} \right\}.$$

Lemma 10. *The probability of a random* $\left(\tilde{\underline{\ell}}^{n_1}, \tilde{\underline{w}}^{n_1}, \tilde{\underline{p}}^{n_1}, \tilde{\underline{c}}^{n_1}\right)$ *being outside the typical set is exponentially small,* i.e.

$$\Pr\left[\left(\tilde{\underline{\ell}}^{n_1}, \tilde{\underline{w}}^{n_1}, \tilde{\underline{p}}^{n_1}, \tilde{\underline{c}}^{n_1}\right) \notin A_\omega^{(N_2)}\right] \leq 2^{-Nb(\omega)}.$$

for some $b(\omega) > 0$ *and* N_2 *sufficiently large.*

Proof. Similar to Lemma 6 of [31] using the Chernoff bound. □

Lemma 11. *For every* $(\tilde{\underline{\ell}}^{n_1}, \tilde{\underline{w}}^{n_1}, \tilde{\underline{p}}_{\mathcal{S}}, \tilde{\underline{p}}_{\mathcal{S}})$ *with* $\left| -\frac{1}{N_2} \log \Pr(\tilde{\underline{\ell}}^{n_1}, \tilde{\underline{w}}^{n_1}, \tilde{\underline{p}}_{\mathcal{S}}, \tilde{\underline{p}}_{\mathcal{S}}) - H(\tilde{L}^{n_1}, \tilde{W}^{n_1}, \tilde{P}_{\mathcal{S}}, \tilde{C}_{\mathcal{S}}) \right| \leq$ ω *the size of the conditional typical is upper bounded by*

$$|A_\omega^{(N_2)}(\tilde{\underline{P}}_{\mathcal{S}^c}, \tilde{\underline{C}}_{\mathcal{S}^c} \mid (\tilde{\underline{\ell}}^{n_1}, \tilde{\underline{w}}^{n_1}, \tilde{\underline{p}}_{\mathcal{S}}, \tilde{\underline{p}}_{\mathcal{S}}))| \leq 2^{N_2(H(\tilde{P}_{\mathcal{S}^c}, \tilde{C}_{\mathcal{S}^c} | \tilde{L}^{n_1}, \tilde{W}^{n_1}, \tilde{P}_{\mathcal{S}}, \tilde{C}_{\mathcal{S}}) + 2\omega)}$$

Proof.

$$
\begin{aligned}
1 &\geq \sum_{(\underline{p}_{\mathcal{S}^c},\underline{c}_{\mathcal{S}^c})\in A_\omega^{(N_2)}\left(\tilde{\tilde{P}}_{\mathcal{S}^c},\tilde{\tilde{C}}_{\mathcal{S}^c}|\,(\underline{\tilde{\tilde{\ell}}}^{n_1},\tilde{\tilde{w}}^{n_1},\underline{\tilde{\tilde{p}}}_{\mathcal{S}},\underline{\tilde{\tilde{p}}}_{\mathcal{S}})\right)} \frac{\Pr\left(\underline{\tilde{\tilde{\ell}}}^{n_1},\tilde{\tilde{w}}^{n_1},\underline{\tilde{\tilde{p}}}^{n_1},\underline{\tilde{\tilde{c}}}^{n_1}\right)}{\Pr\left(\underline{\tilde{\tilde{\ell}}}^{n_1},\tilde{\tilde{w}}^{n_1},\underline{\tilde{\tilde{p}}}_{\mathcal{S}},\underline{\tilde{\tilde{p}}}_{\mathcal{S}}\right)} \\
&\overset{(a)}{\geq} \sum_{(\underline{p}_{\mathcal{S}^c},\underline{c}_{\mathcal{S}^c})\in A_\omega^{(N_2)}\left(\tilde{\tilde{P}}_{\mathcal{S}^c},\tilde{\tilde{C}}_{\mathcal{S}^c}|\,(\underline{\tilde{\tilde{\ell}}}^{n_1},\tilde{\tilde{w}}^{n_1},\underline{\tilde{\tilde{p}}}_{\mathcal{S}},\underline{\tilde{\tilde{p}}}_{\mathcal{S}})\right)} \frac{2^{-N_2(H(\tilde{\tilde{L}}^{n_1},\tilde{\tilde{W}}^{n_1},P^n,C^n)+\omega)}}{2^{-N_2(H(\tilde{\tilde{L}}^{n_1},\tilde{\tilde{W}}^{n_1},\tilde{\tilde{P}}_{\mathcal{S}},\tilde{\tilde{C}}_{\mathcal{S}})-\omega)}} \\
&= \left|A_\omega^{(N_2)}\left(\tilde{\tilde{P}}_{\mathcal{S}^c},\tilde{\tilde{C}}_{\mathcal{S}^c}\mid \underline{\tilde{\tilde{\ell}}}^{n_1},\tilde{\tilde{w}}^{n_1},\underline{\tilde{\tilde{p}}}_{\mathcal{S}},\underline{\tilde{\tilde{p}}}_{\mathcal{S}}\right)\right| 2^{-N_2(H(\tilde{\tilde{P}}_{\mathcal{S}^c},\tilde{\tilde{C}}_{\mathcal{S}^c}|\tilde{\tilde{L}}^{n_1},\tilde{\tilde{W}}^{n_1},\tilde{\tilde{P}}_{\mathcal{S}},\tilde{\tilde{C}}_{\mathcal{S}})+2\omega)}
\end{aligned}
$$

where (a) holds due to Definition 17 of the conditionally typical set and therefore

$$
\left|A_\omega^{(N_2)}\left(\tilde{\tilde{P}}_{\mathcal{S}^c},\tilde{\tilde{C}}_{\mathcal{S}^c}\mid \underline{\tilde{\tilde{\ell}}}^{n_1},\tilde{\tilde{W}}^{n_1},\underline{\tilde{\tilde{p}}}_{\mathcal{S}},\underline{\tilde{\tilde{p}}}_{\mathcal{S}}\right)\right| \leq 2^{N_2(H(\tilde{\tilde{P}}_{\mathcal{S}^c},\tilde{\tilde{C}}_{\mathcal{S}^c}|\tilde{\tilde{L}}^{n_1},\tilde{\tilde{W}}^{n_1},\tilde{\tilde{P}}_{\mathcal{S}},\tilde{\tilde{C}}_{\mathcal{S}})+2\omega)}.
$$

□

By using the upper bound derived for $H(\tilde{\tilde{P}}_{\mathcal{S}^c},\tilde{\tilde{C}}_{\mathcal{S}^c}|\tilde{\tilde{L}}^{n_1},\tilde{\tilde{W}}^{n_1},\tilde{\tilde{P}}_{\mathcal{S}},\tilde{\tilde{C}}_{\mathcal{S}})$ in (3.32) the above inequality becomes

$$
\left|A_\omega^{(N_2)}\left(\tilde{\tilde{P}}_{\mathcal{S}^c},\tilde{\tilde{C}}_{\mathcal{S}^c}\mid \underline{\tilde{\tilde{\ell}}}^{n_1},\tilde{\tilde{W}}^{n_1},\underline{\tilde{\tilde{p}}}_{\mathcal{S}},\underline{\tilde{\tilde{p}}}_{\mathcal{S}}\right)\right| \leq 2^{\frac{(n_1-|\mathcal{S}|)N_2}{2}\left(R_p+\max_{p(x)} I(X;Z)\right)+2N_2\omega}. \tag{G.1}
$$

Appendix H

Entropy of the Public Bits

Since messages $(\underline{W}, \underline{T})$ can be decoded with probability of error λ by accessing $\underline{L}_E^n$ the bits through the side link of capacity C_E, the public bits $(\underline{P}^E)^n$ and the message $\underline{W}$ and by applying Fano's inequality we get $H\big(\underline{W}, \underline{T} | \underline{L}_E^n, (\underline{P}^E)^n, \underline{W}\big) \leq h(\lambda) + nN\lambda R_T$. Since

$$H\big(\underline{W}, \underline{T} | \underline{L}_E^n, (\underline{P}^E)^n, \underline{W}\big) = H\big(\underline{W}, \underline{T}, \underline{L}_E^n, (\underline{P}^E)^n\big) - H\big(\underline{L}_E^n, (\underline{P}^E)^n \underline{W}\big)$$

and by choosing a small enough λ so that $h(\lambda) + nN\lambda R_T \leq nN\varepsilon$ we get

$$H\big(\underline{W}, \underline{T}, \underline{L}_E^n, (\underline{P}^E)^n\big) \leq H\big(\underline{L}_E^n, (\underline{P}^E)^n, \underline{W}\big) + nN\varepsilon.$$

Since $\underline{L}_E^n, (\underline{P}^E)^n$ are given by a deterministic function of $(\underline{W}, \underline{T})$ the equation above becomes

$$\begin{aligned} H\big(\underline{W}, \underline{T}\big) &\leq H\big(\underline{L}_E^n, (\underline{P}^E)^n, \underline{W}\big) + nN\varepsilon \\ &\leq H\big((\underline{P}^E)^n\big) + H\big(\underline{L}_E^n\big) + H\big(\underline{W}\big) + nN\varepsilon \\ &\leq H\big((\underline{P}^E)^n\big) + nNC_E + H\big(\underline{W}\big) + nN\varepsilon. \end{aligned} \tag{H.1}$$

Equation (H.1) and the fact that $H(\underline{T}) = nNR_T$ gives

$$H\big((\underline{P}^E)^n\big) \geq nN \sum_{e \in E} R_{e,p} - 2nN\varepsilon \tag{H.2}$$

Chapter 4

Security in Distributed Storage Systems by Communicating a Logarithmic Number of Bits

4.1 Introduction

We study the security and data integrity of distributed storage systems that use coding for redundancy. It is well known that maximum distance separable (MDS) codes can offer maximum reliability for a given storage overhead and can be used for distributed storage in data centers and peer-to-peer storage systems like OceanStore [61], Total Recall [62], and FS2You [63], that use nodes across the Internet for distributed file storage and sharing. In this chapter we are interested in dealing with errors in the encoded representation. The errors could be introduced either through (unlikely) hard drive undetected failures or through a malicious or compromised server in the storage network.

This second threat is much more eminent when the system uses network coding to maintain the redundancy of the encoded system as proposed recently [2]. To illustrate this consider a large data object that has size $\mathcal{M}$ bits. If this object is to be stored on n servers, depending on the desired redundancy, an (n, k) linear MDS code can be used, dividing the object into k packets of size $\mathcal{M}/k$ each, and storing an encoded packet at each server. Assuming the code is over a finite field $\mathbb{F}_q$, requiring $\log q$ bits to represent each symbol, each server will also need to keep a header denoting the coding coefficients of the linear combinations stored on the server (see section 4.2 for the details) and the size of this header is larger than the size of the useful data if the code is used only once. For this reason it was proposed that the same code is used several times [8] by dividing each packet into N symbols of $\log q$ bits and repeating the same code N times. If $N >> n$ the overhead of storing the coefficients becomes negligible. We refer to this as the N–extended version of an MDS code, shown in Figure 4.2 for the $(4, 2)$ code used in Figure 4.1.

Observe that in this example, each node is storing two linear combinations, (rows) as opposed

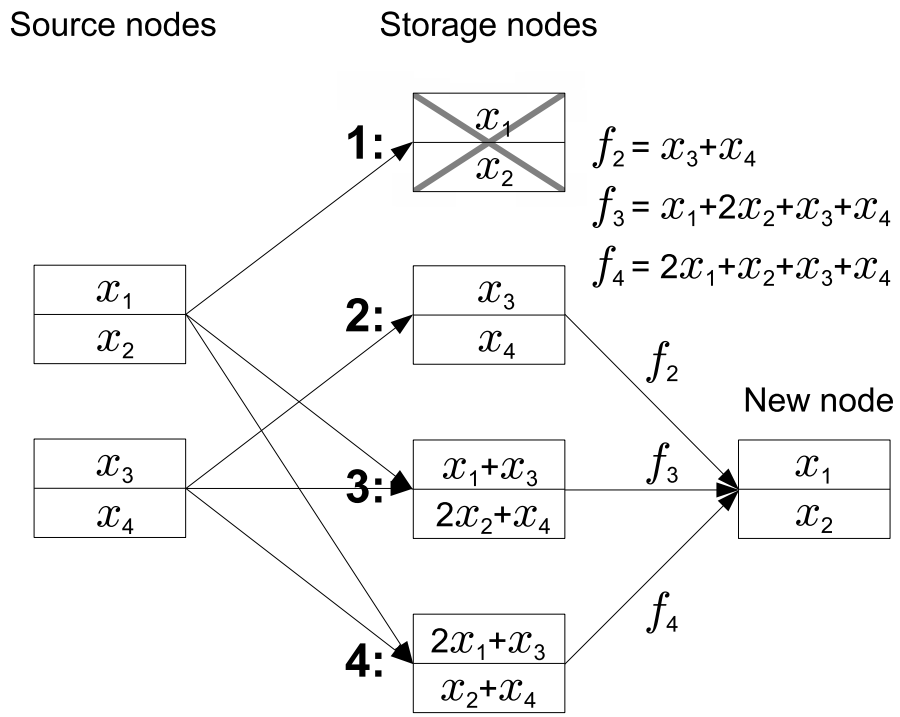


Figure 4.1: A $(4, 2)$ MDS code along with the repair of the first storage node. Each node stores two packets and any two nodes contain enough information to recover all four data packets. In this example the first node leaves the system and a new node is formed by communicating linear combinations f_2, f_3, f_4 which can be used to solve for x_1, x_2 at the new node.

to one. This *sub-packetization* is performed to facilitate *repair* through network coding as proposed in [2]. The problem of repair consists of constructing a new encoded node by accessing as little information from existing encoded nodes. In the example of Figure 4.1, we assume that the first storage node failed and the redundancy of the system needs to be refreshed. This is achieved by communicating "small" linear combinations f_2, f_3, f_4 of the encoded packets from nodes $2, 3$, and 4 each of size $1/2$ of what each node is storing, which as proven in [2], is information theoretically minimal. As storage nodes leave the system and new ones are added, this forms a dynamic storage network that keeps a fixed redundancy and reliability by building new encoded packets from already existing ones. The problem of security should now be clear: even if a single node in this storage system is compromised and participates in this repair process, then it can send incorrect linear combinations that will create erroneous packets at the new nodes. All new nodes using these linear equations will have incorrect data and soon the whole system will be contaminated with nodes having erroneous data.

Our contribution: Since the problem of repairing a code is equivalent to wireline network coding [2], existing techniques for network error correction can be used to detect and correct the errors [64, 65]. These techniques are designed to work for general networks and always guarantee a

transmission rate of $C - 2z$, where C is the min-cut capacity from the source to the destination and z is the number of links contaminated by the adversary. Our approach, that is, creating and communicating small linear hashes which preserve the structure of the code, allows the detection of errors and achieves a transmission rate that can be asymptotically equal to C (by having the receiver connecting to all the non-erroneous nodes) since it takes advantage of the specific structure of the network and the set of links an adversary can contaminate.

To explain our scheme, consider the $(4, 2)$ MDS code of Figure 4.1 and assume one of the four nodes contains errors (say in both rows). A trusted verifier that communicates with all four nodes can find this error by getting the 8 equations contained in each of the $\begin{pmatrix} 4 \\ 2 \end{pmatrix} = 6$ node pairs. Since this is a $(4, 2)$ MDS code, the combinations of equations that come from error-free nodes will be full rank and give a consistent solution whereas the other sets will give different solutions (or might not even be full rank). This is, of course, just using the error-correction capability of the code to detect an error. Our contribution involves using this idea to the N-extended version of a code, by creating a *linear projection (hash) of each row on the same random vector*. The key observation is that if the same random projection is used, this creates an error-correcting code for the hashes which can be communicated to the verifier. The benefit is that each hash has size only $1/N$ of the data in each row reducing the amount of communication to the verifier. One complication is that each node needs to project its data on the same random vector of length N, which requires $N \log q$ bits of common randomness. Subsequently the problem at the verifier is to decode an error-correcting code under adversarial errors. This decoding task can be computationally inefficient but we do not address this issue here, assuming that the verifier can detect the errors if they are within the error-correcting capabilities of the code as dictated by the minimum distance (half the minimum distance). Our analysis investigates under which conditions the small projected hash code will detect any error in the large amount of data stored at the nodes. In particular, we prove the following

Theorem 13. *In a distributed storage system storing a total of $\mathcal{M}$ bits, using an N–extended (n, k) MDS code over $\mathbb{F}_q$, with the n storage nodes sharing $O(\mathcal{M})$ bits of common randomness, our random hashing scheme can detect up to $t \leq t_1 \equiv \lfloor (n-k)/2 \rfloor$ errors by communicating a total of $n(n-k)(\log \mathcal{M} + \log t_1)$ bits to a verifier, with probability of failure*

$$\mathbb{P}\mathrm{r}[F] \leq \frac{1}{\mathcal{M}}.$$

One important weakness of the previous result is the large common randomness required which is comparable to the total size of the data object stored ($1/k(n-k)$ fraction of the $\mathcal{M}$ bits). Note that these bits do not have to be a secret, they only need to be realized after the error has been introduced to the new disk. Their large number, however, makes it impractical to generate them at

one node and then communicate them to the others. Our second contribution involves showing how to use only $O(\log \mathcal{M})$ bits of common randomness to achieve almost the same performance:

Theorem 14. *In a distributed storage system storing a total of $\mathcal{M}$ bits, using an N–extended (n, k) MDS code over $\mathbb{F}_q$, with the n storage nodes sharing $O(\log \mathcal{M})$ bits of common randomness, our pseudorandom hashing scheme can detect up to $t_1 = \lfloor (n-k)/2 \rfloor$ errors by communicating a total of $O(n(n-k)\log \mathcal{M})$ bits to the verifier, with probability of failure*

$$\mathbb{P}\mathrm{r}[F'] \leq \frac{1}{\mathcal{M}}.$$

If there is no common randomness, the verifier can generate the $O(\log \mathcal{M})$ random bits and communicate these to all the nodes requiring a total of $O(n \log \mathcal{M})$ extra communicated bits.

Notice that in this case the total number of bits communicated scales only logarithmically in $\mathcal{M}$, to achieve a probability of failure that scales like $1/\mathcal{M}$. Our construction relies on the pseudorandom small-bias generator used in [33] which can expand $\log N$ random symbols of $\mathbb{F}_q$ (which require $\log N \log q$ random bits to generate), into N pseudorandom symbols that can "fool" any linear function[1]. The only modification to our algorithm is projecting each stored row on this pseudorandom vector to generate each hash and this induces only a small addition to the probability of error. Notice that our work does not rely on any cryptographic assumptions and guarantees that errors inserted in the distributed storage system will be detected with high probability if they are within the capabilities of the code used.

Using the error-correction capability of the code for distributed storage has been suggested before as a way to detect errors [67, 68] and identify "free riders" within the network. A different approach to find errors injected in distributed storage and content distribution systems is the use of signatures and hash functions. Reference [69] introduced the use of homomorphic hashing functions that enables a nodes to perform on-the-fly verification of erasure-encoded blocks. Gkantsidis et al. [15] used the computationally less expensive secure random checksums to detect polluted packets in content distribution system that use network coding while [70, 71] used a method of subspace signatures based on different cryptographic primitives. See also [72, 73, 74] for other related work on security and distributed storage.

4.2 Model

As stated, we consider a data object of size $\mathcal{M}$ bits that is divided into k pieces (of size $\mathcal{M}/k$ bits each) and these are coded into n ($> k$) encoded pieces through a linear (n, k) maximum distance

[1] First introduced by Naor and Naor in [66] for linear functions in $\mathbb{F}_2$.

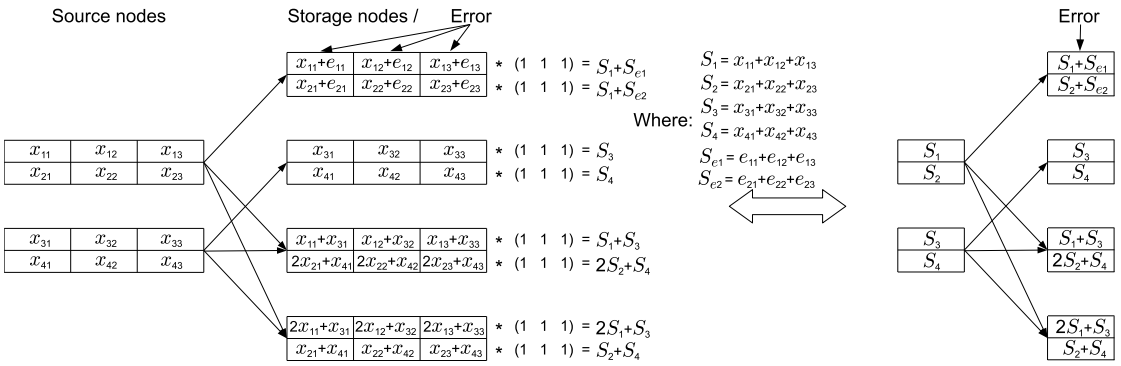


Figure 4.2: Illustration of the 3–extended version of the $(4, 2)$ MDS code shown in Figure 4.1. Each of the three columns stored on the source nodes is coded by repeatedly using the $(4, 2)$ MDS of Figure 4.1. During verification, each row is projected on the vector $r^{\mathrm{T}} = (1\ 1\ 1)$ and the corresponding products $S_1, \ldots, S_4$ form a codeword of the initial $(4, 2)$ MDS code. For example, the errors at the first row of the first node will not be absorbed by the projection as long as $(e_{11}\ e_{12}\ e_{13}) * (1\ 1\ 1)^{\mathrm{T}} \neq 0$

separable (MDS) code. These encoded pieces are stored on n distinct storage nodes along with a header denoting the exact linear combination saved at all the storage nodes. Since the size of the code (n, k) will be much smaller than N, the overhead of storing the code description everywhere (including the verifier) is minimal. This simplifies the model and we can now assume that the errors occur only at the data, since an error at the header would be immediately detected.

We assume that the original information (of size $\mathcal{M}$ bits) is organized into a matrix X with $k(n-k)$ rows and N columns. The elements of this matrix are elements of the finite field $\mathbb{F}_q$, *i.e.*, $X \in \mathbb{F}_q^{k(n-k)\times N}$ where q is a prime or an integer power of a prime. Each column $X_i^c \in \mathbb{F}_q^{k(n-k)\times 1}$ $(i \in \{1, \ldots, N\})$ of matrix X will be separately encoded with the use of an (n, k) MDS code with generator matrix $G \in \mathbb{F}_q^{n(n-k)\times k(n-k)}$ and all the columns $GX_i^c \in \mathbb{F}_q^{n(n-k)\times 1}$ derived by this encoding will be stored on the n different storage nodes of the distributed storage system. We will call this code applied to the N different columns of matrix X as the N–extended MDS code. The overall effect that the N–extended MDS code has upon the information matrix X is captured by the matrix multiplication GX. Figure 4.2 shows such a code for $N = 3$ where the MDS code used is the same as the one shown in Figure 4.1.

The storage nodes of the distributed storage system are assumed to have limited computational capabilities allowing them only to perform inexpensive operations over the finite field $\mathbb{F}_q$. Some of these storage nodes are assumed to store erroneous information, where these errors might be either random due to hardware failures or inserted adversarially by a malicious user. The malicious user can be computationally unbounded, have knowledge of all the information stored on the distributed storage system and can insert errors to any t of the storage nodes.

We assume the existence of a special node called the *verifier* that is assigned to check the integrity of the data stored on different storage nodes. The verifier does not have access to the initial data

object (other than the description of the code) and therefore has to rely on the communicated information to check which nodes contain errors.

4.3 Random Hashes

4.3.1 Illustrating Example

Assume that in the distributed storage system shown in Figure 4.2 with four storage nodes it is known that one of them (the first in this example) stores erroneous information. The goal of the verifier that overlooks the state of the whole system is first to find the erroneous disk with the minimum data exchange and second to repair it by using the information stored on the other disks. Since all three columns stored on the distributed storage system are codewords of a $(4, 2)$ MDS code with at most one error (some columns might be error free) and minimum distance $d = 3$, the naïve approach to find the erroneous disk is to download all data from different disks and then by using minimum distance decoding on each separate column one would be able to find the erroneous disk.

The naïve approach would certainly find the faulty disk but it would require the transfer of double the size of the file stored ($\frac{n}{k}\mathcal{M}$ bits of information in general). So as the size of the file increases this approach will become prohibitively expensive in bandwidth. Instead of transmitting all the information stored on the distributed storage system, the central node could choose a vector with each component chosen independently and uniformly at random from $\mathbb{F}_q$ and have each storage node transmit the inner product (called the hash product) between the randomly chosen vector and each of the rows stored at the disks. In the absence of errors, these hash products will form a codeword of the MDS code used to encode the different columns of the information matrix. In case there are errors, as in the case of the first node in Figure 4.2, the multiplication with the random vector will not obscure these errors unless $S_{ei} = 0 \Leftrightarrow e_{i1} + e_{i2} + e_{i3} = 0$, for $i = \{1, 2\}$. The reason why the chosen vector should be random is so that the adversary cannot deliberately choose the error values in order to make them "disappear" after the vector multiplication.

4.3.2 General Case

The initial information matrix $X \in \mathbb{F}_q^{n(n-k)\times N}$ is coded with the use of an N-extended MDS code with generator matrix $G \in \mathbb{F}_q^{n(n-k)\times k(n-k)}$. Some of the storage nodes contain errors and therefore what is actually stored on the distributed storage system is $Y = GX + E$ where $Y, E \in \mathbb{F}_q^{n(n-k)\times N}$ and E is the error matrix. The verifier wants to identify all erroneous disks by sending hash product requests to all nodes. Then the following theorem holds:

Proof of Theorem 13. All storage nodes share $N \log q$ bits of common randomness and therefore they can create the same random vector $r \in \mathbb{F}_q^{N\times 1}$ with each component of vector r drawn uniformly

at random from $\mathbb{F}_q$. After the random vector r is computed, each storage node calculates the hash product–inner product–between the random vector r and its content on every row. These $n(n-k)$ hash products are equal to:

$$H = Yr = (GX + E)r \Leftrightarrow H = G(Xr) + e \tag{4.1}$$

where $e = Er \in \mathbb{F}_q^{n(n-k)\times 1}$ is a column vector with at most t_m non-zero components representing the erroneous disks (these non-zero components must correspond to the position of at most t_m storage nodes with errors). The key observation is that the projection will not identify an error pattern at a specific row if vector r is orthogonal to that row of E. Intuitively, a randomly selected r will be non-orthogonal to an arbitrary row of E with high probability and this is the probability we need to analyze.

From equation (4.1) it can be seen that the order of applying the MDS encoding on the different columns of the information matrix X and the calculation of the hash products can be interchanged ($(GX)r = G(Xr)$) making the process of identifying the erroneous disks equivalent to finding the error positions in a regular MDS code that is guaranteed to succeed if the minimum distance of the code $(n-k+1)$ is larger than twice the number of errors $2t$ (that is indeed satisfied by the assumptions of Theorem 13).

The set of indices that correspond to the components of vector e that come from disk i is $R_i = \{(i-1)(n-k)+1, \ldots, i(n-k)\}$. We are interested in vector e since this gives us the positions of the faulty disks. One complication that might arise is the fact that disk i might contain an error, meaning that rows $\{E_j^r, j \in R_i\}$ of the error matrix E are not all zero whereas the corresponding components of vector e ($\{e_j, j \in R_i\}$) turn out to be zero and therefore our scheme fails to detect that error. Assume that the set of erroneous disks is $W \subset \{1, 2, \ldots, n\}$ and define $\mathbb{Pr}[F]$ to be the probability of failing to detect some errors. We get

$$\begin{aligned} \mathbb{Pr}[F] &= \mathbb{Pr}\left[\bigcup_{i\in W}\left\{\bigcap_{j\in R_i}\left(E_j^r\, r = 0\right)\right\}\right] \\ \leq \sum_{i\in W}\mathbb{Pr}\left[\bigcap_{j\in R_i}\left(E_j^r\, r = 0\right)\right] &\overset{*}{\leq} \sum_{i\in W}\frac{1}{q} \leq \frac{\lfloor\frac{n-k}{2}\rfloor}{q} \equiv \frac{t_1}{q} \end{aligned} \tag{4.2}$$

where inequality $(*)$ holds due to the fact that the probability that some storage node with errors produce zero hash products is less than $1/q^f$ where f is the number of linearly independent errors rows saved at its disk. So by assuming that the adversary has produced linearly dependent errors would only increase the probability of failure.

If the adversary has saved error vectors at storage node i with rank 1 then the probability

$\mathbb{Pr}[\bigcap_{j\in R_i}(E_j^r\, r=0)]$ in equation (4.2) reduces to an equation for a single row (assuming row k):

$$\mathbb{Pr}\left[\sum_{e_{kf}\neq 0} e_{kf}\, r_f = 0\right] = \mathbb{Pr}\left[r_f = -\sum_{e_{kf'}\neq 0} \frac{e_{kf'}}{e_{kf}}\, r_{f'}\right] = \frac{1}{q}$$

where we only took the terms with a non-zero error coefficient e_{kf}. The numbers $(e_{kf'}/e_{kf})\, r_{f'}$ (e_{kf} is any non-zero error element from the k^{th} row) are independent and uniform over $\mathbb{F}_q$ and so is their sum according to Lemma 12. So the last equality holds since two independent uniformly distributed over $\mathbb{F}_q$ random numbers are equal with probability $1/q$.

When the errors have rank $f > 1$ then the probability $\mathbb{Pr}[\bigcap_{j\in R_i}(E_j^r\, r=0)]$ can be evaluated by disregarding the linearly dependent rows. By looking only at the linearly independent ones and by choosing f columns we can formulate an invertible submatrix $\hat{E}_i \in \mathbb{F}_q^{f\times f}$ and similarly to the previous analysis we have that $\mathbb{Pr}[\bigcap_{j\in R_i}(E_j^r\, r=0)] = \mathbb{Pr}[\hat{E}_i\,\hat{r} = \hat{b}]$ where $\hat{r}, \hat{b} \in \mathbb{F}^{\,f\times 1}$ where $\hat{r}$ are the components of the random vector that correspond to the columns where the submatrix $\hat{E}_i$ was formed. Since $\hat{b}$ is uniformly random, due to the previous analysis $\mathbb{Pr}[\hat{E}\,\hat{r} = b] = 1/q^f$.

Each of the n storage nodes has to convey to the verifier the result of the hash product from all its $(n-k)$ rows, so that the total size of the hash communicated is $\mathcal{H} = n(n-k)\log q$, whereas the size of the file $\mathcal{M} = k(n-k)N\log q$. By substituting the field q equal to $\lfloor \frac{n-k}{2} \rfloor \mathcal{M}$ we conclude the proof of Theorem 13. □

Lemma 12. *The sum of any number of independent uniformly distributed random variables gives a uniformly distributed random variable.*

Proof. Without loss of generality we will prove Lemma 12 only for the case of two random variables. Assume that $x, y \in \mathbb{F}_q$ are two independent and uniformly distributed random variables. We will prove that $x + y$ is also uniformly distributed, indeed $\forall t_1, t_2 \in \mathbb{F}_q$:

$$\begin{aligned}\mathbb{Pr}[x+y=t_1] &= \sum_{t_2\in\mathbb{F}_q} \mathbb{Pr}[x = t_1 - y | y = t_2]\mathbb{Pr}[y = t_2] \\ &\overset{(*)}{=} \sum_{t_2\in\mathbb{F}_q} \mathbb{Pr}[x = t_1 - t_2]\cdot\frac{1}{q} = \sum_{t_2\in\mathbb{F}_q}\frac{1}{q}\cdot\frac{1}{q} = q\cdot\frac{1}{q^2} = \frac{1}{q}\end{aligned}$$

where equality (*) holds due to the independence between x and y. □

Before we continue to prove Theorem 14 we need to give the following definition (extension of Definition 2.1 in [33] to non-prime numbers):

Definition 18. *a) Let q be a prime or an integer power of a prime. For a random variable X with*

values in $\mathbb{F}_q$*, let the* bias *of* X *be defined by*

$$bias(X) = (q-1)\mathbb{Pr}[X=0] - \mathbb{Pr}[X \neq 0]$$

A random variable $X \in \mathbb{F}_q$ *is* ϵ-biased *if* $|bias(X) \leq \epsilon|$.
b) The sample space $\mathcal{S} \subseteq \mathbb{F}_q^\ell$ *is* ϵ-biased *if for all* $c \in \mathbb{F}_q$ *and each sequence* $\beta = (\beta_1, ..., \beta_\ell) \in \mathbb{F}_q^n \backslash \{0^\ell\}$ *the following is valid: if a sequence* $X = (x_1, \ldots, x_\ell) \in \mathcal{S}$ *is chosen uniformly at random from* $\mathcal{S}$*, then the random variable* $(\sum_{i=1}^{\ell} \beta_i x_i + c)$ *is* ϵ*-biased.*

Proof of Theorem 14. All storage nodes execute the algorithm described in Proposition 4.1[2] of [33] and produces a pseudorandom vector $r' \in \mathbb{F}_q^{N\times 1}$ with N components. The quantity m in the algorithm (and consequently the field size $\mathbb{F}_{q^m}$ too) is chosen so that the bias $(q-1)(N-1)/q^m$ is equal to 1 and therefore $q^m = (q-1)(N-1)$ or $m = O(\log N)$. The size of the necessary seed that needs to be provided at all the storage nodes so that they can start the algorithm is two elements from $\mathbb{F}_{q^m}$ chosen uniformly at random or equivalently $2m \log q \equiv O(\log N)$ random bits.

Once all storage nodes have constructed the same pseudorandom vector r' they compute the inner product between vector r' and the content stored on each row of the storage nodes. These pseudorandom products are all sent to the verifier to identify the erroneous disks. The whole analysis is identical to the proof of Theorem 13 with one major difference in the calculation of failure probability $\mathbb{Pr}[F']$. For the case of a pseudorandom vector r', using the same notation as in the proof of Theorem 13:

$$\begin{aligned}
\mathbb{Pr}[F'] &= \mathbb{Pr}\left[\bigcup_{i\in W}\left\{\bigcap_{j\in R_i}\left(E_j^r\, r' = 0\right)\right\}\right] \\
\leq \sum_{i\in W}\mathbb{Pr}\left[\bigcup_{j\in R_i}\left(E_j^r\, r' = 0\right)\right] &\leq \sum_{i\in W}\sum_{j\in R_i}\mathbb{Pr}\left(E_j^r\, r' = 0\right) \\
&\overset{*}{\leq}(n-k)\lfloor\frac{n-k}{2}\rfloor\frac{2}{q} \equiv \frac{2(n-k)t_1}{q}
\end{aligned}$$

where inequality $(*)$ holds since $\mathbb{Pr}\left(E_j^r\, r' = 0\right) = 2/q$. Indeed the bias of the space constructed by the pseudorandom procedure is 1 that means:

$$\begin{aligned}
&\left|(q-1)\mathbb{Pr}\left(E_j^r\, r' = 0\right) - \mathbb{Pr}\left(E_j^r\, r' \neq 0\right)\right| \leq 1 \\
\Leftrightarrow &\left|(q-1)\mathbb{Pr}\left(E_j^r\, r' = 0\right) - \left[1 - \mathbb{Pr}\left(E_j^r\, r' = 0\right)\right]\right| \leq 1 \\
\Leftrightarrow &\left|q\,\mathbb{Pr}\left(E_j^r\, r' = 0\right) - 1\right| \leq 1 \Rightarrow \mathbb{Pr}\left(E_j^r\, r' = 0\right) \leq \frac{2}{q}
\end{aligned}$$

By setting $q = 2(n-k)t_1\mathcal{M}$ we conclude the proof. □

[2]This algorithm is described for q prime but it is readily extensible to q equal to an integer power of a prime.

We would like to underline here that both theorems above exhibit the same behavior on the probability. In Theorem 14 the size of the required common randomness is decreased in the expense of an increased field size. Moreover the use of pseudorandom generators incurs the additional computational cost at each storage node of $O(Nm^2)$ or $O(\mathcal{M} \log \mathcal{M})$ operations in $\mathbb{F}_q$ to generate the pseudorandom vector r'.

Chapter 5

Summary and Future Work

5.1 Summary

In this thesis we examined network transmission delay and security of network coding. With respect to latency we studied acyclic networks comprised of erasure links, where the source has a fixed number of packets it wishes to deliver to the destination. We discussed how network coding compares with traditional routing and showed that even in the unicast case where coding and routing with hop-by-hop retransmissions can achieve the same transmission rate, network coding provides superior performance with respect to transmission delay. The building network for our analysis was a line network. Viewing it as a tandem of queues and using tools developed in queuing theory, we found a tight upper bound on the transmission delay. On the other hand, in networks containing multiple paths, we showed that unlike the line network, network coding and routing have different performance with respect to delay. Specifically we showed that the difference in transmission time between network coding and routing can grow as the square root of the number of packets. Finally, with an analysis based on Azuma-Hoeffding inequality, we proved that time it takes for network coding to complete a transmission of a fixed number of packets is well concentrated around its expected value.

With respect to security of network coding we study the secrecy rate region of a general network with multiple multicast demands in the presence of an eavesdropper that has access to an unknown number set of links and receives a degraded version of what the intended receiver gets. All the channels in the network are assumed to be simultaneously maximizable, meaning that for each channel the same distribution maximizes the mutual information towards the intended receiver and the eavesdropper. We show how to change any of the channels in the network to a corresponding noiseless degraded broadcast channel so that the derived network after the change has a secrecy capacity region that bounds the secrecy capacity region of the initial network. We provide both upper and lower bounds that are independent of the specific network topology and demands. In the case where the eavesdropper cannot wiretap multiple links simultaneously then the bounds are tight leading to an equivalence result. By applying the equivalence transformation from a noisy channel to

a noiseless one for each channel in the network, one can map a noisy network to a noiseless with the same secrecy capacity region; where the noiseless network problem is in general more tractable. In the case where the eavesdropper can wiretap multiple channels at the same time then the equivalence does not hold since the bounds presented in this thesis in this case are loose.

Finally, this thesis closes with the design of a practical code for the detection of maliciously attacked nodes in a distributed storage system. We use pseudorandom generator functions that can fool linear polynomials to create and communicate hash functions of the data in order to detect with high probability all the maliciously attacked nodes in the distributed storage system.

5.2 Future Work

Further work includes extension of the delay analysis and comparison between network coding and routing to general networks with multiple sources and receivers. An interesting avenue is to find tight bounds on the expected transmission time when coding or routing is used in general networks. It is also interesting to analyze the performance of network coding versus routing under different constraints on parameters such as finite field size or buffer size.

On the security side, one area of further work is finding a noiseless model that is equivalent to a noisy wiretap channel in the case where the eavesdropper can wiretap multiple channels simultaneously. Our models, fail to be tight and up until now, it is even an open question whether a noiseless tight model exists.

Bibliography

[1] D. S. Lun, M. Médard, and M. Effros, "On coding for reliable communication over packet networks," in *In Proc. 42nd Annual Allerton Conference on Communication, Control, and Computing, Invited paper*, September–October 2004.

[2] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. O. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," *Submitted for journal publication. Preliminary version appeared in Infocom 2007.*

[3] R. Ahlswede, N. Cai, S.-Y. Li, and R. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000.

[4] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Transactions on Information Theory*, vol. 49, pp. 371–381, 2003.

[5] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Transactions on Networking*, vol. 11, no. 5, pp. 782–795, October 2003.

[6] S. Jaggi, P. Sanders, P. A. Chou, M. Effros, S. Egner, K. Jain, and L. M. G. M. Tolhuizen, "Polynomial time algorithms for multicast network code construction," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1973–1982, June 2005.

[7] T. Ho, M. Médard, R. Koetter, D. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4413–4430, October 2006.

[8] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding," in *Proc. 41st Annual Allerton Conference on Communication, Control, and Computing*, 2003. [Online]. Available: http://research.microsoft.com/ pachou/pubs/ChouWJ03.pdf

[9] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, "Xors in the air: practical wireless network coding," *SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 4, pp. 243–254, 2006.

[10] R. Koetter and F. R. Kschischang, "Coding for errors and erasures in random network coding," in *Information Theory, 2007. ISIT 2007. IEEE International Symposium on*, June 2007, pp. 791 –795.

[11] Y. Wu, R. Dimakis, and K. Ramchandran, "Deterministic regenerating codes for distributed storage," in *Proc. of Allerton Conference on Communication, Control, and Computing*, 2007.

[12] Y. Wu and A. G. Dimakis, "Reducing repair traffic for erasure coding-based storage via interference alignment," in *Proceedings of the 2009 IEEE international conference on Symposium on Information Theory–Volume 4*, ser. ISIT'09, 2009, pp. 2276–2280.

[13] Y. Wu, "A construction of systematic mds codes with minimum repair bandwidth," *Information Theory, IEEE Transactions on*, vol. 57, no. 6, pp. 3738–3741, June 2011.

[14] C. Gkantsidis and P. Rodriguez, "Network coding for large scale content distribution," 2005.

[15] ——, "Cooperative security for network coding file distribution," *IEEE INFOCOM*, 2006.

[16] C. E. Shannon, "Communication theory of secrecy systems," *Bell System Technical Journal, Vol 28, pp. 656715*, October 1949.

[17] A. Wyner, "The wire-tap channel," *Bell Systems Technical Journal*, vol. 54, no. 8, pp. 1355–1387, October 1975.

[18] I. Csiszar and J. Korner, "Broadcast channels with confidential messages," *Information Theory, IEEE Transactions on*, vol. 24, no. 3, pp. 339–348, May 1978.

[19] J. Xu and B. Chen, "Broadcast confidential and public messages," in *Conference on Information Sciences and Systems*, 2008.

[20] R. Alswede and I. Csiszar, "Common randomness in information theory and cryptography–part i: Secret sharing," *Information Theory, IEEE Transactions on*, vol. 39, no. 4, pp. 1121–1132, July 1993.

[21] N. Cai and R. W. Yeung, "Secure network coding," in *Proc. 2002 IEEE Int. Symp. Information Theory (ISIT 2002)*, Lausanne, Switzerland, June 2002, p. 323.

[22] J. Feldman, T. Malkin, R. Servedio, and C. Stein, "On the capacity of secure network coding," in *Proc. of Allerton Conference on Communication, Control, and Computing*, September 2004.

[23] T. Cui, T. Ho, and J. Kliewer, "Achievable strategies for secure network coding for general networks," in *Information Theory and Applications Workshop*, 2010.

[24] ——, "On secure network coding with unequal link capacities and restricted wiretapping sets," in *IEEE Information Theory Workshop (ITW)*, 2010.

[25] K. Bhattad and K. R. Nayayanan, "Weakly secure network coding," in *Proc. of NetCod*, Riva del Garda, Italy, April 2005.

[26] K. Jain, "Security based on network topology against the wiretapping attack," *Wireless Communications, IEEE*, vol. 11, no. 1, pp. 68–71, February 2004.

[27] P. Pakzad, C. Fragouli, and A. Shokrollahi, "Coding schemes for line networks," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, September 2005, pp. 1853–1857.

[28] A. F. Dana, R. Gowaikar, R. Palanki, B. Hassibi, and M. Effros, "Capacity of wireless erasure networks," *IEEE Transactions on Information Theory*, vol. 52, pp. 789–804, 2006.

[29] I. Rubin, "Communication networks: Message path delays," *IEEE Trans. Inf. Theory*, vol. 20, no. 6, pp. 738–745, November 1974.

[30] M. Shalmon, "Exact delay analysis of packet-switching concentrating networks," *IEEE Trans. Commun.*, vol. 35, no. 12, pp. 1265–1271, December 1987.

[31] R. Koetter, M. Effros, and M. Médard, "A Theory of Network Equivalence–Part I: Point-to-Point Channels," *Information Theory, IEEE Transactions on*, vol. 57, no. 2, pp. 972 –995, February 2011.

[32] ——, "A theory of network equivalence, Part II: Multiterminal Channels." [Online]. Available: http://arxiv.org/pdf/1007.1033v2.pdf

[33] C. Bertram-Kretzberg and H. Lefmann, "MOD_p-tests, almost independence and small probability spaces," *Random Structures & Algorithms*, vol. 16, no. 4, pp. 293–313, 2000.

[34] B. Smith and B. Hassibi, "Wireless erasure networks with feedback," 2008, http://arxiv.org/pdf/0804.4298v1.

[35] M. J. Neely and R. Urgaonkar, "Optimal backpressure routing for wireless networks with multi-receiver diversity," *Ad Hoc Netw.*, vol. 7, no. 5, pp. 862–881, 2009.

[36] S. Biswas and R. Morris, "Opportunistic routing in multi-hop wireless networks," in *Proc. Second Workshop on Hot Topics in Networks (HotNets-II)*. Cambridge, Massachusetts: ACM SIGCOMM, November 2003.

[37] P. J. Grabner and H. Prodinger, "Maximum statistics of n random variables distributed by the negative binomial distribution," *Combinatorics, Probability and Computing*, vol. 6, no. 2, pp. 179–183, 1997.

[38] B. Shrader and A. Ephremides, "On the queueing delay of a multicast erasure channel," in *Proceedings of the IEEE Information Theory Workshop*, 2006.

[39] T. H. Theodoros K. Dikaliotis, Alexandros Dimakis, "On the delay of network coding over line networks," in *Proceedings of the IEEE International Symposium on Information Theory*, 2009.

[40] T. Dikaliotis, G. A. Dimakis, T. Ho, and M. Effros, "On the delay advantage of coding in packet erasure networks," in *IEEE ITW*, 2010.

[41] A. Heidarzadeh and A. H. Banihashemi, "Coding delay analysis of chunked codes over line networks," in *International Symposium on Network Coding (NetCod)*, 2012.

[42] ——, "How fast can dense codes achieve the min-cut capacity of line networks?" in *International Symposium on Information Theory*, 2012.

[43] J. Sundararajan, D. Shah, and M. Médard, "On queueing in coded networks queue size follows degrees of freedom," *Information Theory for Wireless Networks, 2007 IEEE Information Theory Workshop on*, pp. 1–6, July 2007.

[44] B. Haeulper, M. Kim, and M. Médard, "Optimality of network coding with buffers," in *Information Theory Workshop*, 2011.

[45] R. R. Weber, "The interchangeability of tandem queues with heterogeneous customers and dependent service times," *Adv. Appl. Probability*, vol. 24, no. 3, pp. 727–737, September 1992.

[46] T. Lindvall, *Lectures on the Coupling Method*. Courier Dover Publications, 2002.

[47] H. Castel-Taleb, L. Mokdad, and N. Pekergin, "Aggregated bounding Markov processes applied to the analysis of tandem queues," in *ValueTools '07: Proceedings of the 2nd international conference on Performance evaluation methodologies and tools*. ICST, 2007, pp. 1–10.

[48] J. Hsu and P. Burke, "Behavior of tandem buffers with geometric input and Markovian output," *IEEE Trans. Commun.*, vol. 24, no. 3, pp. 358–361, March 1976.

[49] H. Daduna, *Queueing Networks with Discrete Time Scale*. New York: Springer-Verlag, 2001.

[50] V. K. Balakrishnan, *Introductory Discrete Mathematics*. Dover Publications, 2008.

[51] P. R. Beesack, "Improvements of Stirling's formula by elementary methods," *Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. Fiz.*, no. 274–301, pp. 17–21, 1969.

[52] R. M. Young, "Euler's constant," *Mathematical Gazette*, vol. 472, pp. 187–190, 1991.

[53] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.

[54] Z. Kong, S. A. Aly, E. Soljanin, E. M. Yeh, and A. Klappenecker, "Network coding capacity of random wireless networks under a sinr model," 2008. [Online]. Available: http://www.citebase.org/abstract?id=oai:arXiv.org:0804.4284

[55] S. A. Aly, V. Kapoor, and J. Meng, "Bounds on the network coding capacity for wireless random networks," in *In Proc. 3rd Workshop on Network Coding, Theory, and Applications*, 2007.

[56] S. Y. El Rouayheb and E. Soljanin, "On wiretap networks II," in *Proc. of IEEE ISIT*, Nice, France, June 2007, pp. 551–555.

[57] A. Mills, B. Smith, T. Clancy, E. Soljanin, and S. Vishwanath, "On secure communication over wireless erasure networks," in *Proc. of IEEE ISIT*, July 2008, pp. 161–165.

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

[59] W. J. Steward, *Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modeling*. Princeton University Press, 2009.

[60] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. Wiley, 2006.

[61] S. Rhea, C. Wells, P. Eaton, D. Geels, B. Zhao, H. Weatherspoon, and J. Kubiatowicz, "Maintenance-free global data storage," *IEEE Internet Computing*, vol. 5, no. 5, pp. 40–49, 2001.

[62] R. Bhagwan, K. Tati, Y.-C. Cheng, S. Savage, and G. M. Voelker, "Total recall: System support for automated availability management," in *Proc. of the Network Systems Design and Implementation*, 2004, pp. 337–350.

[63] Y. Sun, F. Liu, B. Li, Li, B. Li, and X. Zhang, "Fs2you: Peer-assisted semi-persistent online storage at a large scale," in *Proc. of the IEEE Conference on Computer Communications*, April 2009.

[64] S. Jaggi, M. Langberg, S. Katti, T. Ho, D. Katabi, and M. Médard, "Resilient network coding in the presence of byzantine adversaries," in *Proc. IEEE INFOCOM 2007*, Anchorage, AK, May 2007.

[65] R. Koetter and F. R. Kschischang, "Coding for errors and erasures in random network coding," *IEEE Transactions on Information Theory*, 2007, submitted.

[66] J. Naor and M. Naor, "Small-bias probability spaces: Efficient constructions and applications," *SIAM Journal on Computing*, vol. 22, pp. 838–856, 1993.

[67] T. S. J. Schwarz and E. L. Miller, "Store, forget, and check: Using algebraic signatures to check remotely administered storage," in *Proc. of ICDCS*, 2006.

[68] W. Litwin and T. Schwarz, "Algebraic signatures for scalable distributed data structures," in *Proc. of the IEEE International Conference on Data Engineering (ICDE 04)*, 2002, pp. 412–423.

[69] M. Krohn, M. Freedman, and D. Mazieres, "On-the-fly verification of rateless erasure codes for efficient content distribution," in *Proc. of the IEEE Symposium on Security and Privacy*, May 2004, pp. 226–240.

[70] F. Zhao, T. Kalker, M. Médard, and K. J. Han, "Signatures for content distribution with network coding," in *Proc. of the IEEE International Symposium on Information Theory (ISIT)*, 2007.

[71] D. Charles, K. Jain, and K. Lauter, "Signatures for network coding," in *Annual Conference on Information Sciences and Systems*, March 2006, pp. 857–863.

[72] J. P. Vilela, L. Lima, and J. Barros, "Lightweight security for network coding," in *Proc. of the IEEE International Conference on Communications*, May 2008.

[73] E. Kehdi and B. Li, "Null keys: Limiting malicious attacks via null space properties of network coding," in *Proc. of the IEEE Conference on Computer Communications*, April 2009.

[74] A. Le and A. Markopoulou, "Locating byzantine attackers in intra-session network coding using spacemac," in *Proc. of NetCod*, June 2010.