May 2016

Nonparametric Anomaly Detection and Secure Communication

Shaofeng Zou
Syracuse University

ABSTRACT

Two major security challenges in information systems are detection of anomalous data patterns that reflect malicious intrusions into data storage systems and protection of data from malicious eavesdropping during data transmissions. The first problem typically involves design of statistical tests to identify data variations, and the second problem generally involves design of communication schemes to transmit data securely in the presence of malicious eavesdroppers. The main theme of this thesis is to exploit information theoretic and statistical tools to address the above two security issues in order to provide information theoretically provable security, i.e., anomaly detection with vanishing probability of error and guaranteed secure communication with vanishing leakage rate at eavesdroppers.

First, the anomaly detection problem is investigated, in which typical and anomalous patterns (i.e., distributions that generate data) are unknown *a priori*. Two types of problems are investigated. The first problem considers detection of the existence of anomalous geometric structures over networks, and the second problem considers the detection of a set of anomalous data streams out of a large number of data streams. In both problems, anomalous data are assumed to be generated by a distribution q, which is different from a distribution p generating typical samples. For both problems, kernel-based tests are proposed, which are based on maximum mean discrepancy (MMD) that measures the distance between mean embeddings of distributions into a reproducing kernel Hilbert space. These tests are nonparametric without exploiting the information about p and q and are universally applicable to arbitrary p and q. Furthermore, these tests are shown to be statistically consistent under certain conditions on the parameters of the problems. These conditions are further shown to be necessary or nearly necessary, which implies that the MMD-based tests are order level optimal or nearly order level optimal. Numerical results are provided to demonstrate the performance of the proposed tests.

The secure communication problem is then investigated, for which the focus is on degraded broadcast channels. In such channels, one transmitter sends messages to multiple receivers, the channel quality of which can be ordered. Two specific models are studied. In the first model, layered decoding and layered secrecy are required, i.e., each receiver decodes one more message than the receiver with one level worse channel quality, and this message should be kept secure from all receivers with worse channel qualities. In the second model, secrecy only outside a bounded range is required, i.e., each message is required to be kept secure from the receiver with two-level worse channel quality. Communication schemes for both models are designed and the corresponding achievable rate regions (i.e., inner bounds on the capacity region) are characterized. Furthermore, outer bounds on the capacity region are developed, which match the inner bounds, and hence the secrecy capacity regions are established for both models.

NONPARAMETRIC ANOMALY DETECTION AND SECURE COMMUNICATION

By

Shaofeng Zou
B. E., Shanghai Jiao Tong University, Shanghai, China, 2011

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Electrical and Computer Engineering

Syracuse University
May 2016

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor Prof. Yingbin Liang for her support and supervision during my PhD study. She has taught me, both consciously and unconsciously, how good research is done. Her enthusiasm and passion for research have been always inspiring me, especially during the tough time in the PhD pursuit. She has always made herself available to advise me despite of her tight schedule. I have learned a lot from her unique perspective on research, her sharp insight on the problem, and immense knowledge.

I would like to thank Prof. Venugopal V. Veeravalli for providing me the great opportunity to visit his group at University of Illinois at Urbana Champaign, which has greatly broadened my scope of knowledge and enriched my research interests. His insightful thoughts have always contributed to a better understanding of the problems.

I would like to thank my collaborators, Prof. Lifeng Lai, Prof. Shlomo Shamai(Shitz), Prof. H. Vincent Poor for their continuously instructive suggestions and comments. They always have a broader view of problems which has provided me more comprehensive understanding of the problems.

I would like to thank my doctoral defense committee: Prof. Biao Chen, Prof. Pramod K. Varshney, Prof. Qinru Qiu, Prof. Venugopal V. Veeravalli, Prof. Lixin Shen for their time and efforts on my thesis. I am also grateful to my lab mates: Jiayao Hu, Ruchen Duan, Weiguang Wang, Anhong He, Huishuai Zhang, Yunhao Sun, Yi Zhou, Zhe Wang for their help in study and life. I also thank my friends, who made my life enjoyable.

I would like to thank my parents and my sister for their love and unconditional sup-

port. I am very indebted to Ying, who supported me in every possible way to see the completion of this work.

I acknowledge the support from Syracuse University and NSF grants CCF-12-18451 and CCF-10-26565.

TABLE OF CONTENTS

List of Tables

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Two major security challenges in information systems are detection of anomalous data patterns that reflect malicious intrusions into data storage systems and protection of data from malicious eavesdropping during data transmissions. The first problem typically involves design of statistical tests to identify data variations, and the second problem generally involves design of communication schemes to transmit data securely in the presence of malicious eavesdroppers. The main theme of this thesis is to exploit information theoretic and statistical tools to address the above two security issues in order to provide information theoretically provable security, i.e., guaranteed anomaly detection with vanishing probability of error and guaranteed secure communication with vanishing leakage rate at eavesdroppers.

1.1 Nonparametric Anomaly Detection

Anomaly detection is an important problem that has attracted intensive interest in various research areas and application domains. The goal of anomaly detection problems is to identify data patterns (typically captured by statistical distributions of data) that do not conform to a certain expected pattern in a dataset. In order to solve the anomaly detection problem of interest, most previous studies have focused on parametric scenarios, assuming the typical and anomalous distributions are known

a priori, although practical applications typically provide only raw data. Implicitly, these studies assumed that the distributions are learned from the data. However, since the ultimate goal is to detect anomalies, learning the distributions first and then constructing the detection rules may not yield optimal performance. It is thus desirable to design data-driven nonparametric tests, which directly perform anomaly detection using data without estimating the distributions as an intermediate step. Furthermore, since such tests do not exploit any information about the distributions, they can be designed to provide universal performance guarantee for arbitrary distributions.

In the first part of the thesis, we investigate nonparametric anomaly detection problems in which both typical and anomalous distributions are unknown, and can be arbitrary. More specifically, we study two classes of anomaly detection problems: 1) detection of an anomalous geometric structure over a network and 2) detection of anomalous data streams out of a large number of data streams.

1.1.1 Anomalous Geometric Structure Detection

For the anomalous geometric structure detection problem, each node in the network observes a random sample. An anomalous structure, if it exists, corresponds to a cluster of nodes in the network that take samples generated by a distribution q. All other nodes in the network take samples generated by a distribution p that is different from q. If there does not exist an anomalous structure, then all nodes receive samples generated by p. The distributions p and q are *arbitrary* and *unknown a priori*. Designed tests are required to distinguish between the null hypothesis (i.e., no anomalous structure exists) and the alternative hypothesis (i.e., there exists an anomalous structure). Due to the fact that the anomalous structure may be one of a number of candidate structures in the network, this is a composite hypothesis testing problem.

Such a problem models a variety of applications. For example, in sensor networks, sensors are deployed over a large range of space. These sensors take measurements from the environment in order to determine whether or not there is intrusion of an anomalous object. Such intrusion typically activates only a few sensors that cover a certain geometric area. An alarm is then triggered if

the network detects an occurrence of intrusion based on the sensorsŠ measurements. Other applications can arise in detecting an anomalous segment of DNA sequences, detecting virus infection of computer networks, and detecting anomalous spot in images.

As an interesting topic, detecting existence of an anomalous geometric structure in networks has been intensively studied in the literature. A number of studies focused on networks with nodes embedded in a lattice such as one dimensional line and square. In [1], the network is assumed to be embedded in a d-dimensional cube, and geometric structures such as line segments, disks, rectangles and ellipsoids associated with nonzero-mean Gaussian random variables need to be detected out of other nodes associated with zero-mean Gaussian noise variables. A multi-scale approach was proposed and its optimality was analyzed. In [2], detection of spatial clusters under the Bernoulli model over a two-dimensional space was studied, and a new calibration of the scan statistic was proposed, which results in optimal inference for spatial clusters. In [3], the problem of identifying a cluster of nodes with nonzero-mean values from zero-mean noise variables over a random field was studied.

Further generalization of the problem has also been studied, when network nodes are associated with a graph structure, and existence of an anomalous cluster or an anomalous subgraph of nodes needs to be detected. In [4], an unknown path corresponding to nonzero-mean variables needs to be detected out of zero-mean variables in a network with nodes connected in a graph. In [5], for various combinatorial and geometric structures of anomalous objects, conditions were established under which testing is possible or hopeless with a small risk. In [6], the cluster of anomalous nodes can either take certain geometric shapes or be connected as subgraphs. Such structures associated with nonzero-mean Gaussian variables need to be detected out of zero-mean variables. In [7] and [8], network properties of anomalous structures such as small cut size were incorporated in order to assist successful detection. More recently, in [9], the problem of detecting connected subgraph with elevated mean out of zero-mean Gaussian random variables was studied. An algorithm was proposed to characterize the family of all connected sub-graphs in terms of linear matrix inequalities. The minimax optimality of such an approach was further established in [10] for

exponential family on lattice networks.

However, previous studies focused on *parametric* or *semiparametric* models, which assume that samples are generated by known distributions such as Gaussian or Bernoulli distributions, or the two distributions are known to have mean shift. Such parametric models may not always hold in real applications. In many cases, distributions can be arbitrary, and may not be Gaussian or Bernoulli. They may not differ in mean either. The distributions may not even be known in advance. Hence, it is desirable to develop nonparametric tests that are universally applicable to arbitrary distributions.

In contrast to previous studies, we study the *nonparametric* problem of detecting an anomalous structure, in which distributions can be *arbitrary* and *unknown a priori*. In order to deal with nonparametric models, we apply maximum mean discrepancy (MMD) as a distance metric. This approach has been applied to solving the two sample problem in [11], in which the quantity of MMD was used as a metric of distance between mean embeddings of two distributions.

We are interested in the asymptotic scenario in which the network size goes to infinity and the number of candidate anomalous structures scales with the network size. Thus, the number of sub-hypotheses under the alternative hypothesis also increases, which causes the composite hypothesis testing problem to be difficult. On the other hand, since the distributions can be arbitrary, it is in general difficult to exploit properties of the distributions such as mean shift to detect existence of an anomalous structure. Furthermore, as the network size becomes large, in contrast to parametric models in which the mean shift can scale with the network size, here it is necessary that the numbers of samples within and outside of each anomalous structure should scale with the network size fast enough in order to provide more accurate information about both distributions p and q and guarantee asymptotically small probability of error. Thus, the problem amounts to characterize how the minimum and maximum sizes of all candidate anomalous structures should scale with the network size in order to consistently detect the existence of an anomalous structure.

We list our main contributions as follows.

(1) We construct MMD-based distribution-free tests for various networks.

(2) We analyze the performance guarantee for the proposed MMD-based test. For the problem of detecting an anomalous interval in a line network, we show that as the network size n goes to infinity, if the minimum size $I_{\min}$ of candidate anomalous intervals satisfies $I_{\min} = \Omega(\log n)$ [1], and the maximum size $I_{\max}$ of candidate anomalous intervals satisfies $n - I_{\max} = \Omega(\underbrace{\log \cdots \cdots \log}_{\text{arbitrary k number of log}} n)$, then the proposed test is consistent, i.e., the probability of error is asymptotically small.

(3) We further derive necessary conditions on $I_{\min}$ and $I_{\max}$ that any test must satisfy in order to be universally consistent for arbitrary p and q. Comparison of sufficient and necessary conditions yields that the MMD-based test is order level optimal in terms of $I_{\min}$ and nearly order level optimal in terms of $I_{\max}$ for the line network.

(4) We further generalize such analysis to other networks and obtain similar type of results. Our results also demonstrate the impact of geometric structures on performance guarantee of tests.

Our technical analysis is very different from that for parametric problems. The obvious difference is due to significantly different approaches applied to the two types of problems. Furthermore, the nonparametric nature also affects the asymptotic formulation of the problem. The lower and upper bounds (such as $I_{\min}$ and $I_{\max}$ in line network) on the sizes of all candidate anomalous structures must scale with the network size in order to guarantee enough samples in and outside the anomalous structure if it occurs. This is significantly different from parametric models where problems can still be well posed even with a single node or the entire network being anomalous, so long as a certain distribution parameter (such as mean shift between the two distributions) scales with the network size. Consequently, the asymptotic analysis for the nonparametric problem requires considerable new technical developments.

Although the kernel-based approach has been used to solve various machine learning problems, it is not widely applied to solving detection problems with only few exceptions such as the two

[1] We adopt the following notations to express asymptotic scaling of quantities with n:
- $f(n) = O(g(n))$: there exist $k, n_0 > 0$ s.t. for all $n > n_0$, $|f(n)| \leq k|g(n)|$;
- $f(n) = \Omega(g(n))$: there exist $k, n_0 > 0$ s.t. for all $n > n_0$, $f(n) \geq kg(n)$;
- $f(n) = \Theta(g(n))$: there exist $k_1, k_2, n_0 > 0$ s.t. for all $n > n_0$, $k_1 g(n) \leq f(n) \leq k_2 g(n)$;
- $f(n) = o(g(n))$: for all $k > 0$, there exists $n_0 > 0$ s.t. for all $n > n_0$, $|f(n)| \leq kg(n)$;
- $f(n) = \omega(g(n))$: for all $k > 0$, there exists $n_0 > 0$ s.t. for all $n > n_0$, $|f(n)| \geq k|g(n)|$.

sample problem [11]. Since the nature of our problem necessarily involves geometric structures in networks, the technical analysis requires substantial efforts to deal with scaling of the size of geometric structures and analyze the impact of geometry on consistency of tests, which are not captured in the two sample problem.

1.1.2 Anomalous Data Stream Detection

For the anomalous data stream detection problem, there are totally n sequences out of which s anomalous sequences need to be detected. Each *typical* sequence consists of m independent and identically distributed (i.i.d.) samples drawn from a distribution p, whereas each *anomalous* sequence contains i.i.d. samples drawn from a distribution q that is distinct from p. The distributions p and q are assumed to be unknown. The goal is to build distribution-free tests to detect the s anomalous data sequences generated by q out of all data sequences.

Solutions to such a problem is very useful in many applications. For example, in cognitive wireless networks, signals follow different distributions either p or q depending on whether the channel is busy or vacant. A major issue in such a network is to identify vacant channels out of a large number of busy channels based on their corresponding signals in order to utilize vacant channels for improving spectral efficiency. This problem was studied in [12] and [13] under the assumption that p and q are known, whereas here, we study the problem with unknown p and q. Other applications include detecting anomalous DNA sequences out of typical sequences, detecting virus infected computers from other virus free computers, and detecting slightly modified images from other untouched images.

The parametric model of the problem has been well studied, e.g., [12], in which it is assumed that the distributions p and q are known in advance and can be exploited for detection. However, the nonparametric model is less explored, in which it is assumed that the distributions p and q are unknown and can be arbitrary. Recently, Li, Nitinawarat and Veeravalli proposed the divergence-based generalized likelihood tests in [14], and characterized the error decay exponents of these tests. However, [14] studied only the case when the distributions p and q are discrete with finite

alphabets, and their tests utilize empirical probability mass functions of p and q.

For this problem, we study the nonparametric model, in which distributions p and q can be continuous and arbitrary. The major challenges to solve this problem (compared to the discrete case studied in [14]) lie in: (1) it is difficult to accurately estimate continuous distributions with limited samples for further anomalous data stream detection; (2) it is difficult to design low complexity tests for continuous distributions; and (3) building distribution-free consistent tests (and further guaranteeing exponential error decay) is challenging for arbitrary distributions.

For this problem, we apply MMD as a metric to construct our tests for detecting anomalous data sequences. In contrast to consistency analysis in classical theory as in [14], which assumes that the problem dimension (i.e., the number n of sequences and the number s of anomalous sequences) is fixed and the sample size m increases, our focus is on the regime in which the problem dimension (i.e., n and s) increases. This is motivated by applications, in which anomalous sequences are required to be detected out of a large number of typical data sequences. It is clear that as n (and possibly s) becomes large, it is increasingly challenging to consistently detect all anomalous sequences. It then requires that the sample size m correspondingly increases in order to guarantee more accurate detection. Hence, we are interested in characterizing how the sample size m should scale with n and s in order to guarantee the consistency of our tests.

Our main contributions for the problem of anomalous data stream detection are listed as follows.

(1) We construct MMD-based distribution-free tests, which enjoy low computational complexity and are proven to be powerful for nonparametric detection.

(2) We analyze the performance guarantee for the proposed MMD-based test. We bound the probability of error as a function of the sample size m, the number s of anomalous sequences, and the total number n of sequences. We then show that with s known, the constructed test is exponentially consistent if m scales at the order $\Omega(\log n)$ for any p and q, whereas with s unknown, m should scale at the order $\omega(\log n)$ (i.e., strictly larger than $\Omega(\log n)$). Thus, lack of the information about s results in order level increase in sample size m needed for consistent detection. We further

develop low complexity consistent tests by exploiting the asymptotic behavior of s and n.

(3) We further derive a necessary condition which states that no test can be consistent for arbitrary p and q if m scales at the order $O(\log n)$, thus establishing the order level optimality of the MMD-based test.

(4) We also provide an interesting example study, in which the distribution q is the mixture of the distribution p and the anomalous distribution $\tilde{q}$. In such a case, the anomalous sequence contains only sparse samples from the anomalous distribution. Our results for such a model quantitatively characterize the impact of the sparsity level of anomalous samples on the scaling behavior of the sample size m, in order to guarantee consistency of the proposed tests.

We further provide numerical results to demonstrate our theoretical assertions and compare our tests with other competitive approaches. Our numerical results demonstrate that the MMD-based test has a better performance than the divergence-based generalized likelihood test proposed in [14] when the sample size m is not very large. We also demonstrate that the MMD-based test outperforms (or performs as well as) other competitive tests via a real data set.

1.2 Secure Communications over Broadcast Networks

In wireless networks, communication signals are transmitted via the open medium of the free space, and hence can be easily eavesdropped upon by any receiver within transmission ranges. This broadcast nature of radio channels is one of the major challenges to the design of secure wireless communications. Some commonly used security approaches employed in current wireless systems may encounter potential problems as wireless networks incorporate more communication patterns and flexible structures. For example, a popular approach to secure wireless communications is to pre-deploy a secret certificate into mobile devices, based on which devices can establish keys. However, for device-to-device (D2D) communications recently proposed for LTE networks, such an approach cannot adapt easily for a mobile device to directly communicate with a large set of devices in a unicast fashion. Furthermore, public-key based encryption is also not applicable in

many cases, as mobile devices may not be equipped with sufficiently high computational resources for implementing public-key algorithms.

In the seminal work by Wyner [15], a physical layer approach to secrecy was proposed, which exploits randomness in statistical communication channels as resources to achieve secure communications. Without inherently employing secret keys, such a new security approach, if applied to wireless networks, can significantly reduce requirements on the infrastructure and improve communication flexibility and dynamics. Wyner's result was further generalized to the case in which the transmitter further sends one common message to both the legitimate receiver and the eavesdropper by Csiszár and Körner in [16].

Following the initial studies in [15] and [16], broadcast channels with various decoding and secrecy constraints have been studied intensively (see [17] and [18] for more references). Wyner's wiretap model was further studied when the legitimate and eavesdropping channels take specific forms. As some key examples, the Gaussian wiretap channel was studied in [19]; the multiple-input multiple-output (MIMO) wiretap channel with the transmitter, the legitimate receiver, and/or the eavesdropper equipped with multiple antennas was studied in [20–25]; and the compound wiretap channel, in which there are multiple legitimate receivers and single/multiple eavesdroppers, was studied in [26–30].

Csiszár and Körner's broadcast model was further studied for the Gaussian fading channel in [31], and for the MIMO channel in [32]. This model was generalized in [33] to two compound scenarios, in which the legitimate receiver (i.e., receiver 1) and the eavesdropper (i.e., receiver 2) are respectively replaced by two receivers with the same decoding and secrecy requirements. Furthermore, Csiszár and Körner's model was also generalized in [34] to the compound scenario, in which each receiver is replaced by multiple users. And such a model was further generalized in [35] to the case with one legitimate receiver and multiple eavesdroppers, in which the legitimate receiver is required to decode all messages, and the eavesdroppers are required to satisfy the layered secrecy requirements, i.e., one more message is required to be secured as channel quality gets one level worse.

As further generalizations of the Wyner and Csiszár-Körner models, a class of broadcast channels with an additional eavesdropper were intensively studied. In the model considered in [36] and [37], a transmitter has two independent messages intended for two legitimate receivers, respectively, and wishes to keep the two messages confidential from an (additional) eavesdropper. Such a model was further studied in [38], when the channel is corrupted by additive Gaussian noise. The multiple antenna version of the above model was studied in [39] and [40]. Furthermore, the multi-antenna channel was generalized in [41] to the compound scenario with each receiver and the eavesdropper being replaced by a group of co-located users. The model was also generalized and studied in [42] for the case with an arbitrary number of legitimate receivers (and hence with an arbitrary number of independent messages respectively for each receiver), and the fading channel of such a model was studied in [29].

Apart from the above class of broadcast channels, another class of models consisting of receivers that are expected to not only receive certain information from the transmitter but also be kept ignorant of certain other information have also been studied. In the model studied in [43], a transmitter has two independent messages with each intended for one receiver and required to be kept secure from the other receiver. The MIMO version of such a model was studied in [44–46]. Furthermore, such a model was generalized in [47] to the case in which the transmitter has one more common message for both receivers, and users are equipped with multiple antennas. The compound scenario of the preceding model with each receiver being replaced by a group of co-located users was studied in [47].

In fact, these models can be unified under a more general framework, in which a transmitter sends a number of messages to a set of receivers over a broadcast channel, and the receivers' channel quality can be ordered in a certain way. Each receiver can possibly serve as a legitimate user expecting a certain subset of messages, and/or as an eavesdropper that should be kept ignorant of a certain subset of messages.

In the second part of the thesis, we study two broadcast models: the degraded broadcast channels with layered decoding and layered secrecy and the degraded broadcast channel with secrecy

outside a bounded range. Layered decoding requires that as the channel quality gets one level better, one more message is decoded, and layered secrecy requires that as the channel quality gets one level worse, one more message is kept secure. Here, we focus on degraded channels for two reasons: 1) degraded channels often arise naturally in practical applications such as in the context of Gaussian fading channels that model wireless communication channels; 2) the performance for degraded channels can often be characterized in simpler forms that can facilitate the illustration of central ideas. However, all achievable schemes designed for degraded channels are applicable to non-degraded channels except that the optimality of the schemes are not easy to prove (due to the difficulty in developing outer bounds that match achievable regions).

1.2.1 Degraded Broadcast Channel with Layered Decoding and Layered Secrecy

In the degraded broadcast channel with layered decoding and layered secrecy (see Fig. 4.1), a transmitter wishes to transmit K messages to K receivers. The channel outputs at receivers naturally satisfy the degradedness condition, i.e., from receiver K to receiver 1, the quality of their channels gets worse gradually. It is required that receiver k decode one more message W_k than receiver $k-1$ for $k = 2, \ldots, K$, and this additional message W_k should be kept secure from all receivers with worse channel outputs, i.e., with lower indices.

Such a model captures practical scenarios in which users are ranked to receive files with different security levels. For example, a WiFi network in a company consists of a number of legitimate users. Users with certain ranks are allowed to receive files up to certain security levels, and should be kept ignorant of files with higher security levels. Hence users with higher ranks are able to see more files. It is also possible to set the channel quality based on users' ranks by assigning more communication resources to higher ranked users. Another example is in social networks in which one user wishes to share more resources with close friends and fewer resources with other friends. As we show in Chapter 4, this model is equivalent to a secret sharing problem.

For the problem of degraded broadcast channel with layered decoding and layered secrecy, we

list our main contributions as follows.

(1) We establish the secrecy capacity regions for both the discrete memoryless channel (DMC) and Gaussian MIMO channel.

(2) We construct the achievable scheme based on superposition coding and random binning. More specifically, for each message, W_k, one layer is designed and superposed on the layer designed for W_{k-1}. The codewords within each layer are further divided into a number of bins, and the corresponding message is encoded as the bin number, which the index inside the bin serves as a random source to protect the message. Thus, the receivers that are required to decode this message can tell which bin the codeword is in and hence decode the message, while those receivers with worse channel quality are kept ignorant of the message.

(3) We provide a more involved analysis of leakage rates than the cases with two or three receivers, which generalize the current techniques for two or three receivers. For the DMC, we develop a novel generation of the analysis of the leakage rate provided in [48] for one legitimate receiver to multiple receivers. This approach carries complementary insights for analyzing the leakage rate for scenarios with layered decoding and secrecy constraints.

(4) We provide converse proofs for the DMC and Gaussian MIMO channel which require careful constructions of auxiliary random variables (covariance matrices) in a recursive form, which also generalizes the existing techniques for two or three receivers.

1.2.2 Degraded Broadcast Channel with Secrecy Outside a Bounded Range

For the model with layered decoding and secrecy described in the previous section, the additional message decoded by a better receiver needs to be kept secure from the receiver with only one level worse channel quality (layered secrecy, zero secrecy range). Although such a model is feasible for broadcast channels with discrete states (i.e., quality of receivers can be captured by discrete channel states), it cannot capture the scenarios with receivers' channel quality varying continuously. For such a case, it is more reasonable to require the message to be secured from the receivers with

a certain amount of worse channel quality, instead of being secured from the receiver with one level worse channel quality, which is not even well defined for continuous channel quality. To be more explicit, we use an example to illustrate the motivation of such a model. Consider a degraded broadcast channel with infinite number of receivers, in which h denotes the amplitude of the channel gain (the larger h, the better the channel). In this case, it is impossible to require that the message intended for receivers with $h \geq h_0$ to be secured from receivers with $h < h_0$, because no positive secrecy rate can be achieved. Instead, it is more nature to require that the messages intended for receivers with $h \geq h_0$ to be secured from receivers with $h \leq h_0 - \Delta$, where $\Delta > 0$. We refer to such a secrecy requirement as *secrecy outside of a bounded range*.

In this thesis, we study the four-receiver case of the above model (see Fig. 5.1), in which a transmitter sends information to four receivers over a discrete memoryless channel. The channel quality is assumed to gradually degrade from receiver 4 to receiver 1. There are in total four messages W_1, W_2, W_3, W_4 intended for four receivers with the following decoding and secrecy requirements. Receiver k is required to decode messages $W_1, \ldots, W_k$, for $k = 1, 2, 3, 4$. Furthermore, message W_3 needs to be kept secure from receiver 1, and message W_4 need to be kept secure from receiver 1 and receiver 2. Thus, each message is required to be kept secure from receivers with two-level worse channel quality.

For the problem of degraded broadcast channel with secrecy outside a bounded range, we list our main contributions as follows.

(1) We establish the secrecy capacity region for the four-receiver model.

(2) We propose a novel achievable scheme which is based on the following techniques: (1) superposition coding, which encodes each message into one layer in order to satisfy the layered decoding requirements at the four receivers; (2) embedded coding, which exploits the secrecy requirement outside a bounded range to use lower-layer messages as random sources to secure higher-layer messages; (3) random binning, which provides further randomness to secure each message at its corresponding layer; and (4) rate splitting and sharing, which turns out to be critical for this model to further enlarge the achievable region.

(3) We develop a converse proof which exploits the insight obtained from the achievable scheme.

We mainly illustrate why rate splitting and sharing is useful here. Consider the case where layer 3 is sufficient to secure layer 4. Random binning is then not necessary in layer 4. Hence, simply using techniques superposition coding and embedded coding yields the rate of W_4 to be bounded by the decoding capability of receiver 4 given decoding of the three other messages. It turns out to be very difficult to develop the converse proof for the resulting achievable region, which suggests that such an achievable region may not be large enough. Indeed, the previous achievable scheme ignores the fact that under assumption of this case, part of layer 3 (say W_{31}) is good enough to secure the remaining part of layer 3 (say W_{32}) and layer 4 from receiver 2. Hence, W_{32} can be counted towards either the rate R_3 or the rate R_4, which provides the flexibility to enlarge R_4 and correspondingly the achievable region. Such an idea motivates our development of splitting W_3 into two parts W_{31} and W_{32} and sharing W_{32} between R_3 and R_4. The converse for this resulting achievable region can be developed, suggesting that rate splitting and sharing are important for establishing the secrecy capacity region.

We further note that during the preparation of this thesis, we were able to establish the secrecy capacity region for the case with arbitrary k-receiver case. The results will be drafted for conference and journal submissions in the near future.

1.3 Summary of Contributions and Thesis Organization

The rest of the thesis is organized as follows. In Chapter 2, we present our results for the problem of anomalous geometric structure detection over networks. In Chapter 3, we present our results for the problem of anomalous data stream detection. In Chapter 4, we present our results for the problem of degraded broadcast channel with layered decoding and layered secrecy, and its application to multi-secret sharing problem. In Chapter 5, we present our results for the problem of degraded broadcast channel with secrecy outside a bounded range. In Chapter 6, we summarize

our results and discuss about future works.

As a summary, works reported in this thesis has led to two journal publications, two journal submissions and one journal in preparation, together with eleven conference publications [49–61]. We provide a list of these publications as follows.

Journal papers:

1. S. Zou, Y. Liang, L. Lai, and S. Shamai (Shitz), “An information theoretical approach to secret sharing," IEEE Trans. Inform. Theory, vol. 61, no. 6, pp. 3121-3136, 2015.

2. S. Zou, Y. Liang, L. Lai, H. V. Poor, and S. Shamai (Shitz), “Broadcast networks with layered decoding and layered secrecy: Theory and applications," Proceedings of the IEEE, vol. 103, no. 10, pp. 1841-1856, 2015.

3. S. Zou, Y. Liang, H. V. Poor. “Nonparametric Detection of Geometric Structures over Networks", submitted to IEEE Trans. Inform. Theory, 2015.

4. S. Zou, Y. Liang, H. V. Poor, X. Shi. “Nonparametric Detection of Anomalous Data Streams via Kernel Mean Embedding", submitted to IEEE Trans. Inform. Theory, 2015.

5. S. Zou, Y. Liang, L. Lai, H. V. Poor, S. Shamai (Shitz). “Degraded Broadcast Channel with Secrecy Outside a Bounded Range", in preparation.

Conference papers:

1. S. Zou, Y. Liang, S. Shamai (Shitz). “Multiple Access Channel with State Uncertainty at Transmitters", in Proc. IEEE International Symposium on Information Theory (ISIT), 2013.

2. S. Zou, Y. Liang, L. Lai, S. Shamai (Shitz). “Layered Decoding and Secrecy over Degraded Broadcast Channels", in Proc. IEEE 14th Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2013.

3. S. Zou, Y. Liang, L. Lai, S. Shamai (Shitz). “Layered Decoding and Secrecy over Degraded Gaussian MIMO Broadcast Channels and Application in Secret Sharing", in Proc. IEEE International Symposium on Information Theory (ISIT), 2014.

4. S. Zou, Y. Liang, H. V. Poor, X. Shi. “Kernel-Based Nonparametric Anomaly Detection", in Proc. IEEE 15th Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2014.

5. S. Zou, Y. Liang, H. V. Poor. “A Kernel-Based Nonparametric Test For Anomaly Detection Over Line Network", in Proc. IEEE International Workshop on Machine Learning for Signal Processing (MLSP), 2014.

6. S. Zou, Y. Liang, H. V. Poor, X. Shi. “Unsupervised Nonparametric Anomaly Detection: A Kernel Method", in Proc. of 52th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, Oct. 2014.

7. S. Zou, Y. Liang, L. Lai, S. Shamai (Shitz). “Degraded Broadcast Channel: Secrecy Outside of a Bounded Range", in Proc. 2015 IEEE Information Theory Workshop (ITW), Jerusalem, 2015.

8. S. Zou, Y. Liang, L. Lai, S. Shamai (Shitz). “Rate Splitting and Sharing for Degraded Broadcast Channel with Secrecy Outside a Bounded Range", in Proc. IEEE International Symposium on Information Theory (ISIT), Hongkong, 2015.

9. Y. Bu, S. Zou, Y. Liang, and V. V. Veeravalli, “Universal outlying sequence detection for continuous observations", to appear in Proc. IEEE International Conference on Acoustics,Speech and Signal Processing (ICASSP), Shanghai, 2016.

10. S. Zou, Y. Liang, and H. V. Poor, “Nonparametric detection of an anomalous disk over a two-dimensional lattice network", to appear in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, 2016.

11. Y. Bu, S. Zou, Y. Liang, and V. V. Veeravalli, “Estimation of KL Divergence Between Large-Alphabet Distributions", to appear in Proc. IEEE International Symposium on Information Theory (ISIT), 2016.

CHAPTER 2
ANOMALOUS GEOMETRIC STRUCTURE DETECTION

In this chapter, we study the problem of anomalous geometric structure detection. In Section 2.1, we introduce our problem model and the performance metric in the context of detecting existence of anomalous interval over line networks. In Section 2.2, we introduce the metric MMD. In Section 2.3, we present our results for the line network. In Section 2.4, we generalize our approaches to higher dimensional networks. In Section 2.5, we present the numerical results.

2.1 Problem Statement

In this section, we introduce our problem formulation in the context of line network that we study in Section 2.3. We describe generalization of this problem to other networks in Section 2.4 when we present the corresponding results for these networks.

We consider a line network, which consists of nodes $1, \ldots, n$, as shown in Figure 2.1. We use I to denote a subset of consecutive indices of nodes, which is referred to as an *interval*. Here, the length of an interval I refers to the cardinality of I, and is denoted by $|I|$. We assume that any interval with the length between $I_{\min}$ and $I_{\max}$ can be a candidate anomalous interval, and collect

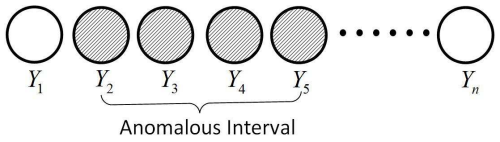

Fig. 2.1: A line network with an anomalous interval.

all candidate anomalous intervals into the following set $\mathcal{I}_n^{(a)}$

$$\mathcal{I}_n^{(a)} = \{I : I_{\min} \leq |I| \leq I_{\max}\}. \tag{2.1}$$

As we explain towards the end of this section, the two problem parameters $I_{\min}$ and $I_{\max}$ play an important role in determining whether the problem is well posed.

We assume that each node, say node i, is associated with a random variable, denoted by Y_i, for $i = 1, \ldots, n$. We consider two hypotheses about the distributions of the line network. For the *null hypothesis* H_0, Y_i for $i = 1, \ldots, n$ are independently and identically distributed (i.i.d.) random variables, and are generated from a distribution p. For the *alternative hypothesis* H_1, there exists an interval $I \in \mathcal{I}_n^{(a)}$ over which Y_i (with $i \in I$) are i.i.d. and are generated from a distribution $q \neq p$, and otherwise, Y_i are i.i.d. and generated from the distribution p. Thus, the alternative hypothesis is composite due to the fact that $\mathcal{I}_n^{(a)}$ contains multiple candidate anomalous intervals, and these intervals differentiate from each other by their length and location in the network. We further assume that under both hypotheses, each node generates only one sample. Putting the problem into a context, H_0 models the scenario when the observations Y_i are background noise, and H_1 models the scenario when some Y_i (for $i \in I$) are observations activated by an anomalous intrusion.

In contrast to previous work, we assume that the distributions p and q are *arbitrary* and *unknown a priori*. For this problem, we are interested in the asymptotic scenario, in which the number of nodes goes to infinity, i.e., $n \to \infty$. The performance of a test for such a system is captured by two types of errors. The *type I error* refers to the event that samples are generated from the null hypothesis, but the detector determines an anomalous event occurs. We denote the

probability of such an event as $P(H_1|H_0)$, or $P_{H_0}(error)$. The *type II error* refers to the case that an anomalous event occurs but the detector claims that samples are generated from the null hypothesis. We denote the probability of such an event as $P(H_0|H_1)$, or $P_{H_1}(error)$. We define the following minimax risk to measure the performance of a test:

$$R_m^{(n)} = P(H_1|H_0) + \max_{I \in \mathcal{I}_n^{(a)}} P(H_0|H_{1,I}). \tag{2.2}$$

Definition 2.1. *A test is said to be consistent if the minimax risk* $R_m^{(n)} \to 0$*, as* $n \to \infty$.

It can be seen from the definition of $\mathcal{I}_n^{(a)}$ that $I_{\min}$ and $I_{\max}$ determine the number of candidate anomalous intervals. Furthermore, if there exists an anomalous interval, $I_{\min}$ determines the least number of samples generated by q and $n - I_{\max}$ determines the least number of samples generated by p. As $n \to \infty$, to guarantee asymptotically small probability of error, both $I_{\min}$ and $I_{\max}$ must scale with n to provide sufficient information about p and q in order to yield accurate distinction between the two hypotheses. This suggests that as the network becomes larger, only a large enough but not too large anomalous object can be detected. Therefore, our goal in this problem is to characterize how $I_{\min}$ and $I_{\max}$ should scale with the network size in order for a test to successfully distinguish between the two hypotheses. Such conditions on $I_{\min}$ and $I_{\max}$ can thus be interpreted as the resolution of the corresponding test.

2.2 Introduction to MMD

We adopts the emerging technique based on mean embedding of distributions into a reproducing kernel Hilbert space (RKHS) [62, 63]. The idea is to map probability distributions into a RKHS with an associated kernel such that distinguishing the two probabilities can be carried out by distinguishing their corresponding embeddings in the RKHS. Since RKHS naturally carries a distance metric, mean embeddings of distributions can be compared easily based on their distances in the RKHS. Such a metric is referred to as the *maximum mean discrepancy (MMD)* [11, 64]. In the

following, we briefly introduce the idea of mean embedding of distributions into RKHS and the metric of MMD.

Suppose $\mathcal{P}$ is a set of probability distributions, and suppose $\mathcal{H}$ is the RKHS with an associated kernel $k(\cdot,\cdot)$ (see [65, 66] for an introduction of RKHS). We define a mapping from $\mathcal{P}$ to $\mathcal{H}$ such that each distribution $p \in \mathcal{P}$ is mapped to an element in $\mathcal{H}$ as follows

$$\mu_p(\cdot) = \mathbb{E}_p[k(\cdot, x)] = \int k(\cdot, x) dp(x).$$

Here, $\mu_p(\cdot)$ is referred to as the *mean embedding* of the distribution p into the Hilbert space $\mathcal{H}$. Due to the reproducing property of $\mathcal{H}$, it is clear that $\mathbb{E}_p[f] = \langle \mu_p, f \rangle_{\mathcal{H}}$ for all $f \in \mathcal{H}$.

It is desirable that the embedding is *injective* such that each $p \in \mathcal{P}$ is mapped to a unique element $\mu_p \in \mathcal{H}$. It has been shown in [63, 67–69] that for many RKHSs such as those associated with Gaussian and Laplace kernels, the mean embedding is injective. In order to distinguish between two distributions p and q, [11] introduced the following quantity of MMD based on the mean embeddings μ_p and μ_q of p and q in RKHS:

$$\text{MMD}[p, q] := \|\mu_p - \mu_q\|_{\mathcal{H}}. \tag{2.3}$$

It is clear that the MMD equals the distance between the mean embeddings μ_p and μ_q of the two distributions p and q, and can be used to distinguish between p and q due to the injectivity of mean embedding.

It can also be shown that

$$\text{MMD}[p, q] = \sup_{\|f\|_{\mathcal{H}} \leq 1} \mathbb{E}_p[f(x)] - \mathbb{E}_q[f(x)].$$

Namely, $\text{MMD}[p, q]$ achieves the maximum of the mean difference of a function between the two distributions over all unit-norm functions in the RKHS $\mathcal{H}$.

Due to the reproducing property of kernel, it can be easily shown that

$$\text{MMD}^2[p,q] = \mathbb{E}_{x,x'}[k(x,x')] - 2\mathbb{E}_{x,y}[k(x,y)] + \mathbb{E}_{y,y'}[k(y,y')], \tag{2.4}$$

where x and x' have independent but the same distribution p, and y and y' have independent but the same distribution q. An unbiased estimator of $\text{MMD}^2[p,q]$ based on l_1 samples of X and l_2 samples of Y is given as follows,

$$\begin{aligned}\text{MMD}_u^2[X,Y] =& \frac{1}{l_1(l_1-1)}\sum_{i=1}^{l_1}\sum_{j\neq i}^{l_1} k(x_i,x_j) + \frac{1}{l_2(l_2-1)}\sum_{i=1}^{l_2}\sum_{j\neq i}^{l_2} k(y_i,y_j) \\ & - \frac{2}{l_1 l_2}\sum_{i=1}^{l_1}\sum_{j=1}^{l_2} k(x_i,y_j).\end{aligned} \tag{2.5}$$

We note that other estimators of the $\text{MMD}^2[p,q]$ are also available, which can be used for our problem. We focus on the unbiased estimator given above to convey the central idea.

2.3 Detection of Anomalous Interval in Line Network

2.3.1 Test and Performance

We construct a nonparametric test using the unbiased estimator in (2.5) and the scan statistics. For each interval I, let Y_I denote the samples in the interval I, and $Y_{\bar{I}}$ denote the samples outside the interval I. We compute $\text{MMD}_{u,I}^2(Y_I, Y_{\bar{I}})$ for all intervals $I \in \mathcal{I}_n^{(a)}$. Under the null hypothesis H_0, all samples are generated from the distribution p. Hence, for each $I \in \mathcal{I}_n^{(a)}$, $\text{MMD}_{u,I}^2(Y_I, Y_{\bar{I}})$ yields an estimate of $\text{MMD}^2[p,p]$, which is zero. Under the alternative hypothesis H_1, there exists an anomalous interval I^* in which the samples are generated from distribution q. Hence, $\text{MMD}_{u,I^*}^2(Y_{I^*}, Y_{\bar{I^*}})$ yields an estimate of $\text{MMD}^2[p,q]$, which is bounded away from zero due to the fact that $p \neq q$.

Based on the above understanding, we build the following test:

$$\max_{I:I\in\mathcal{I}_n^{(a)}} \text{MMD}^2_{u,I}(Y_I, Y_{\bar{I}}) \begin{cases} \geq t, & \text{determine } H_1 \\ < t, & \text{determine } H_0 \end{cases} \tag{2.6}$$

where t is a threshold parameter. It is anticipated that with sufficiently accurate estimate of MMD and an appropriate choice of the threshold t, the test in (2.6) should provide desired performance. The following theorem characterizes the performance of the proposed test.

Theorem 2.1. *Suppose the test in* (2.6) *is applied to the nonparametric problem described in Section 2.1. Further assume that the kernel in the test satisfies* $0 \leq k(x,y) \leq K$ *for all* (x,y)*. Then, the type I and type II errors are upper bounded respectively as follows:*

$$\begin{aligned} P(H_1|H_0) &\leq \sum_{I:I_{\min}\leq|I|\leq I_{\max}} \exp\left(-\frac{t^2|I|(n-|I|)}{8K^2n}\right) \\ &= \sum_{I_{\min}\leq i\leq I_{\max}} (n-i+1)\exp\left(-\frac{t^2 i(n-i)}{8K^2n}\right), \end{aligned} \tag{2.7}$$

$$P(H_0|H_{1,I}) \leq \exp\left(-\frac{(\textit{MMD}^2[p,q]-t)^2|I|(n-|I|)}{8nK^2}\right), \textit{ for } I\in\mathcal{I}_n^{(a)} \tag{2.8}$$

where t *is the threshold of the test that satisfies* $t < \textit{MMD}^2[p,q]$*.*

Furthermore, the test (2.6) *is consistent if*

$$I_{\min} \geq \frac{16K^2(1+\eta)}{t^2}\log n, \tag{2.9}$$

$$I_{\max} \leq n - \frac{16K^2(1+\eta)}{t^2}\underbrace{\log\cdots\cdots\log}_{\textit{arbitrary k number of } \log} n\ , \tag{2.10}$$

where η *is any positive constant.*

Proof. See Section 2.6. □

We note that many kernels satisfy the boundedness condition required in Theorem 2.1, such as Gaussian kernel and Laplacian Kernel.

The above theorem implies that to guarantee consistency of the proposed test, the minimum length $I_{\min}$ should scale at the order $I_{\min} = \Omega(\log n)$. Furthermore, $n - I_{\max}$ should scale at the order $\Omega(\underbrace{\log \cdots \cdots \log}_{\text{arbitrary k number of } \log} n)$ which can goes to infinity arbitrarily slow. Hence, the number of candidate anomalous intervals in the set $\mathcal{I}_n^{(a)}$ is $\Theta(n^2)$, which is at the same order as the number of all intervals. Hence, at the order sense, not many intervals are excluded from being anomalous.

It can be seen that the conditions (2.9) and (2.10) on $I_{\min}$ and $I_{\max}$ are asymmetric. This can be understood by the upper bound (2.7) on the type I error, which is a sum over all candidate anomalous intervals with length between $I_{\min}$ and $I_{\max}$. Due to the specific geometric structure of the line network, as the length $|I|$ increases, the number of candidate anomalous intervals with length $|I|$ equals $n - |I| + 1$ and decreases as $|I|$ increases. Although the term $\exp\left(-\frac{t^2 i(n-i)}{8K^2 n}\right)$ in (2.7) is symmetric over i with respect to $\frac{n}{2}$, the entire term $(n - i + 1)\exp\left(-\frac{t^2 i(n-i)}{8K^2 n}\right)$ is not symmetric, which consequently yields the asymmetric conditions on $I_{\min}$ and $I_{\max}$.

Theorem 2.1 requires that the threshold t in the test (2.6) to be less than $\text{MMD}^2[p, q]$. In practice, the information of $\text{MMD}^2[p, q]$ may or may not be available depending on specific applications. If it is known, then the threshold t can be set as a constant smaller than $\text{MMD}^2[p, q]$. If it is unknown, then the threshold t needs to scale to zero as n gets large in order to be asymptotically smaller than $\text{MMD}^2[p, q]$. We summarize these two cases in the following two corollaries.

Corollary 2.1. *If the value $MMD^2[p, q]$ is known a priori, we set the threshold $t = (1-\delta)MMD^2[p, q]$ for any $0 < \delta < 1$. The test in* (2.6) *is consistent, if $I_{\min}$ and $I_{\max}$ satisfy the following conditions,*

$$\begin{aligned} I_{\min} &\geq \frac{16K^2(1+\eta')}{MMD^4[p,q]} \log n \\ I_{\max} &\leq n - \frac{16K^2(1+\eta')}{MMD^4[p,q]} \underbrace{\log \cdots \cdots \log}_{\textit{arbitrary k number of } \log} n \ , \end{aligned} \tag{2.11}$$

where η' is any positive constant.

Corollary 2.1 follows directly from Theorem 2.1 by setting $\eta' = \frac{1+\eta}{(1-\delta)^2} - 1$.

Corollary 2.2. *If the value MMD*$^2[p, q]$ *is unknown, we set the threshold t to scale with* n*, such that* $\lim_{n\to\infty} t_n = 0$*. The test in* (2.6) *is consistent, if* $I_{\min}$ *and* $I_{\max}$ *satisfy the following conditions,*

$$I_{\min} \geq \frac{16K^2(1+\eta)}{t_n^2} \log n$$
$$I_{\max} \leq n - \frac{16K^2(1+\eta)}{t_n^2} \underbrace{\log \cdots\cdots \log}_{\textit{arbitrary k number of } \log} n \ , \tag{2.12}$$

where η *is any positive constant.*

Corollary 2.2 follows directly from Theorem 2.1 by noting that $t_n < \text{MMD}^2[p, q]$ for n large enough.

We note that Corollary 2.2 holds for any t_n that satisfies $\lim_{n\to\infty} t_n = 0$. It is clear from Corollary 2.2 that for the case when $\text{MMD}^2[p, q]$ is unknown, $I_{\min}$ should scale at the order $\omega(\log n)$, and $n - I_{\max}$ should scale at the order $\omega(\underbrace{\log \cdots\cdots \log}_{\text{arbitrary k number of } \log} n)$. Hence, comparison of the above two corollaries implies that the prior knowledge about $\text{MMD}^2[p, q]$ is very important for network ability to identifying anomalous events. If $\text{MMD}^2[p, q]$ is known, then the network can resolve an anomalous object with the size $\Omega(\log n)$. However, if such knowledge is unknown, the network can resolve only bigger anomalous objects with the size $\omega(\log n)$.

We note that Theorem 2.1 and Corollaries 2.1 and 2.2 characterize the conditions to guarantee test consistency for a pair of fixed but unknown distributions p and q. Hence, the conditions (2.9), (2.10), (2.11) and (2.12) depend on the underlying distributions p and q. In fact, these conditions further yield the following condition that guarantees the proposed test to be universally consistent for any arbitrary p and q.

Proposition 2.1 (Universal Consistency)**.** *Consider the nonparametric problem given in Section 2.1. Further assume the test in* (2.6) *applies a bounded kernel with* $0 \leq k(x, y) \leq K$ *for any*

(x, y). *Then the test* (2.6) *is universally consistent for any arbitrary pair of* p *and* q*, if*

$$I_{\min} = \omega(\log n)$$
$$I_{\max} = n - \Omega(\underbrace{\log \cdots\cdots \log}_{\text{arbitrary } k \text{ number of } \log} n) \tag{2.13}$$

Proof. This result follows from (2.9), (2.10), (2.11) and (2.12) and the fact that MMD$[p, q]$ is a constant for any given p and q. □

2.3.2 Necessary Conditions

In Section 2.3.1, Proposition 2.1 suggests the sufficient conditions on $I_{\min}$ and $I_{\max}$ to guarantee the proposed nonparametric test to be universally consistent for arbitrary p and q. In the following theorem, we characterize the necessary conditions on $I_{\min}$ and $I_{\max}$ that any test must satisfy in order to be universally consistent for arbitrary p and q.

Theorem 2.2. *For the nonparametric detection problem described in Section 2.1 over a line network, any test must satisfy the following conditions on* $I_{\min}$ *and* $I_{\max}$ *in order to be universally consistent for arbitrary* p *and* q*:*

$$I_{\min} = \omega(\log n)$$
$$\text{and } n - I_{\max} \to \infty, \text{ as } n \to \infty. \tag{2.14}$$

Proof. See Section 2.7. The idea of the proof is to first lower bound the minimax risk by the Bayes risk of a simpler problem. Then for such a problem, we show that there exist p and q (in fact Gaussian p and q) such that even the optimal parametric test is not consistent under the conditions given in the theorem. This thus implies that under the same condition, no nonparametric test is universally consistent for arbitrary p and q. □

It can be seen that the necessary condition on $I_{\min}$ in (2.14) matches the sufficient condition in (2.13) at the order level which implies that the proposed test is order level optimal in $I_{\min}$.

Furthermore, the sufficient condition on $I_{\max}$ can arbitrarily slowly converge to n, which is also very close to the necessary condition on $I_{\max}$. Thus we have the following theorem.

Theorem 2.3 (Optimality)**.** *Consider the nonparametric detection problem described in Section 2.1. The MMD-based test* (2.6) *is order level optimal in terms of* $I_{\min}$ *and nearly order level optimal in terms of* $I_{\max}$ *required to guarantee universal test consistency for arbitrary* p *and* q*.*

2.4 Generalization to Other Networks

In this section, we generalize our study to three other networks in order to demonstrate more generality of our approach. For each network, our study further demonstrates how the geometric property of the network affects the conditions required to guarantee the test consistency.

2.4.1 Detection of Anomalous Interval in Ring Network

In this subsection, we consider a ring network (see Figure 2.2), in which n nodes are located over a ring. We define an interval I to be a subset of consecutive nodes over the ring. We consider the following set of candidate anomalous intervals,

$$\mathcal{I}_n^{(a)} = \{I : I_{\min} \leq |I| \leq I_{\max}\}, \tag{2.15}$$

where $I_{\min}$ and $I_{\max}$ are minimal and maximal lengths of all candidate anomalous intervals. Despite similarities that the ring network shares with the line network, its major difference lies in that the number of candidate anomalous intervals with size k is n (which remains the same as k increases) as opposed to $n - k + 1$ in the line network (which decreases as k increases). Consequently, the number of sub-hypotheses in H_1 is different. Such difference is reflected in the results that we present next.

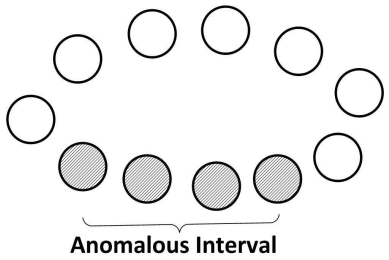

Fig. 2.2: A ring network with an anomalous interval

We construct the test as follows:

$$\max_{I:I\in\mathcal{I}_n^{(a)}} \text{MMD}^2_{u,I}(Y_I, Y_{\bar{I}}) \begin{cases} \geq t, & \text{determine } H_1 \\ < t, & \text{determine } H_0 \end{cases} \tag{2.16}$$

where Y_I denotes the samples in the interval I, $Y_{\bar{I}}$ denotes the samples outside the interval I, and t is a threshold parameter.

If we apply the test (2.16) with a bounded kernel, then it can be shown (see Section 2.8) that the type I and type II errors are bounded as follows:

$$P(H_1|H_0) \leq \exp\left(2\log n - \frac{2t^2 \min\{I_{\min}(n-I_{\min}), I_{\max}(n-I_{\max})\}}{16nK^2}\right), \tag{2.17}$$

$$P(H_0|H_{1,I}) \leq \exp\left(-\frac{(\text{MMD}^2[p,q]-t)^2|I|(n-|I|)}{8nK^2}\right), \text{ for } I \in \mathcal{I}_n^{(a)} \tag{2.18}$$

where t is the threshold of the test that satisfies $t < \text{MMD}^2[p,q]$. Furthermore, the test in (2.16) is consistent, if

$$I_{\min} \geq \frac{16K^2(1+\eta)}{t^2}\log n, \tag{2.19}$$

$$I_{\max} \leq n - \frac{16K^2(1+\eta)}{t^2}\log n, \tag{2.20}$$

where η is any positive constant. The detailed proof can be found in Section 2.8. Comparing the above conditions with Theorem 2.1 suggests that although the sufficient conditions on $I_{\min}$

are the same, the conditions on $I_{\max}$ reflect order level difference in line and ring network. For line network, an anomalous interval can be close to the entire network with only a gap of length $\Omega(\underbrace{\log \cdots \cdots \log}_{\text{arbitrary k number of log}} n)$. However, for ring network, the gap can be as large as $\Omega(\log n)$. Such difference in tests' resolution of anomalous intervals is mainly due to the difference in network geometry that further affects the error probability of tests. By carefully comparing the two types of errors, in fact, the type II error converges to zero as the network size goes to infinity as long as the number of anomalous samples (i.e., length of anomalous intervals) and the number of typical samples (i.e., the gap between anomalous intervals and the entire network) both scale with n to infinity. Thus, the conditions for the type II error being asymptotically small are the same for the two types of networks. The situation is different for the type I error. The key observation is that the number of candidate anomalous intervals with size k is $n-k+1$ in a line network (which decreases as k increases), but is n in a ring network (which remains the same as k increases). Such difference can be as significant as the order level if k is close to n, say $n-\Omega(\log n)$. Consequently, the type I error for a line network can be much smaller than that for a ring network, resulting more relaxed condition on $I_{\max}$ to guarantee consistency.

Similarly to the line network, setting the threshold t for the test (2.16) can be considered in two cases with and without the information of $\text{MMD}[p,q]$. If $\text{MMD}[p,q]$ is known, set $t=(1-\delta)\text{MMD}^2[p,q]$. Otherwise, t can be chosen to scale to zero as n goes to infinity. Similar results as in Corollary 2.1 and Corollary 2.2 can then be derived for a ring network.

Furthermore, (2.19) and (2.20) imply that the test (2.16) is *universally consistent* for any arbitrary p and q, if

$$I_{\min}=\omega(\log n), \quad \text{and} \quad n-I_{\max}=\omega(\log n). \tag{2.21}$$

Following the arguments similar to those for the line network, it can be shown that any test must satisfy the following necessary conditions required on $I_{\min}$ and $I_{\max}$ in order to be *universally*

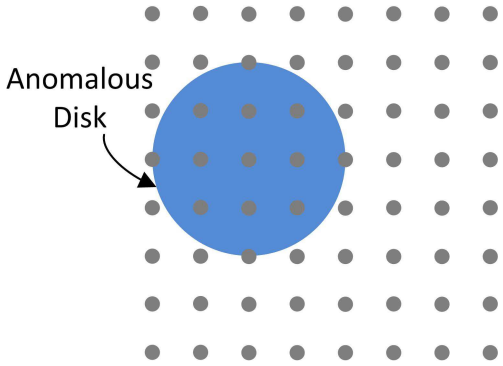

Fig. 2.3: Two-dimensional lattice network with an anomalous disk.

consistent for arbitrary p and q:

$$I_{\min} = \omega(\log n), \quad \text{and} \quad n - I_{\max} \to \infty, \text{as } n \to \infty. \tag{2.22}$$

The detailed proof can be found in Section 2.9.

Therefore, combining the above sufficient and necessary conditions, we conclude the following optimality property for the proposed test.

Theorem 2.4 (Optimality)**.** *Consider the problem of nonparametric detection of an interval over a ring network. The MMD-based test* (2.16) *is order level optimal in terms of* $I_{\min}$ *required to guarantee universal test consistency for arbitrary* p *and* q*.*

2.4.2 Detection of Anomalous Disk in Two-Dimensional Lattice Network

We consider a two-dimensional lattice network (see Figure 2.3) consisting of n^2 nodes placed at the corner points of a lattice. Consider the following set of candidate anomalous disks with each disk centered at a certain node with integer radius:

$$\mathcal{D}_n^{(a)} = \{D : D_{\min} \leq |D| \leq D_{\max}\}, \tag{2.23}$$

where $|D|$ denotes the number of nodes within the disk D, $D_{\min} := \min_{D\in\mathcal{D}_n^{(a)}} |D|$ and $D_{\max} := \max_{D\in\mathcal{D}_n^{(a)}} |D|$. The goal is to detect the existence of an anomalous disk over the lattice network. Towards this end, we build the following test:

$$\max_{D:D\in\mathcal{D}_n^{(a)}} \text{MMD}^2_{u,D}(Y_D, Y_{\overline{D}}) \begin{cases} \geq t, & \text{determine } H_1 \\ < t, & \text{determine } H_0 \end{cases} \tag{2.24}$$

where Y_D contains samples within the disk D, and $Y_{\overline{D}}$ contains samples outside the disk D. If we apply this test with a bounded kernel, then the type I error can be bounded as follows:

$$P(H_1|H_0) \leq \exp\left(3\log n - \frac{2t^2 \min\{D_{\min}(n^2 - D_{\min}), D_{\max}(n^2 - D_{\max})\}}{16n^2K^2}\right), \tag{2.25}$$

and the type II error can be bounded as follows:

$$P(H_0|H_{1,D}) \leq \exp\left(-\frac{(\text{MMD}^2[p,q] - t)^2|D|(n^2 - |D|)}{8n^2K^2}\right) \text{ for } D \in \mathcal{D}^{(a)}, \tag{2.26}$$

where t is the threshold of the test that satisfies $t < \text{MMD}^2[p,q]$.

It can be further shown that if $D_{\min}$ and $D_{\max}$ satisfy the following conditions:

$$D_{\min} \geq \frac{24K^2(1+\eta)}{t^2}\log n, \tag{2.27}$$

$$D_{\max} \leq n^2 - \frac{24K^2(1+\eta)}{t^2}\log n, \tag{2.28}$$

where η is any positive constant, then the test (2.24) is consistent. Interestingly, the largest disk within a two-dimensional lattice network has radius to be $\frac{n}{2}$ and areas to be $\frac{\pi n^2}{4} \approx 0.79n^2$, which contains at most cn^2 nodes with $c < 1$ for large n. This implies that the bound on $D_{\max}$ in (2.28) is satisfied automatically when n is large.

Hence, for large n, (2.27) implies that the test (2.24) is *universally consistent* for any arbitrary

p and q, if

$$D_{\min} = \omega(\log n). \tag{2.29}$$

Furthermore, following the arguments similar to those for the line network, it can be shown that any test must satisfy the following necessary condition required on $D_{\min}$ in order to be *universally consistent* for arbitrary p and q:

$$D_{\min} = \omega(\log n). \tag{2.30}$$

Therefore, combining the above sufficient and necessary conditions, we conclude the following optimality property for the proposed test.

Theorem 2.5 (Optimality)**.** *Consider the problem of nonparametric detection of a disk over two-dimensional lattice network. The MMD-based test* (2.24) *is order level optimal in the size of disks required to guarantee universal test consistency for arbitrary p and q.*

2.4.3 Detection of Anomalous Rectangle in Lattice Network

We consider a r-dimensional lattice network consisting of n^r nodes placed at the corner points of a lattice network. Consider the following set of candidate anomalous rectangles:

$$S_n^{(a)} := \{S = [I_1 \times I_2 \times \ldots \times I_r] : S_{\min} \leq |S| \leq S_{\max}\},$$

where I_i for $1 \leq i \leq r$ denotes an interval contained in $[1, n]$ with consecutive indices, $|S|$ denotes the number of nodes in the rectangle S, $S_{\min} := \min_{S \in \mathcal{S}_n^{(a)}} |S|$, and $S_{\max} := \max_{S \in \mathcal{S}_n^{(a)}} |S|$. The goal is to detect existence of an anomalous r-dimensional rectangle. Towards this end, we build

the following test,

$$\max_{S:S\in\mathcal{S}_n^{(a)}} \text{MMD}^2_{u,S}(Y_S, Y_{\overline{S}}) \begin{cases} \geq t, & \text{determine } H_1 \\ < t, & \text{determine } H_0 \end{cases} \tag{2.31}$$

where Y_S contains samples within the rectangular S, and $Y_{\overline{S}}$ contains samples outside the rectangular S. If we apply this test with a bounded kernel, then the type I error is bounded as follows:

$$P(H_1|H_0) \leq \exp\left(2r\log n - \frac{2t^2\min\{S_{\min}(n^r - S_{\min}), S_{\max}(n^r - S_{\max})\}}{16n^rK^2}\right), \tag{2.32}$$

and the type II error is bounded as follows:

$$P(H_0|H_{1,S}) \leq \exp\left(-\frac{(\text{MMD}^2[p,q] - t)^2|S|(n^r - |S|)}{8n^rK^2}\right), \text{ for } S \in \mathcal{S}^{(a)} \tag{2.33}$$

where t is the threshold of the test that satisfies $t < \text{MMD}^2[p,q]$.

It can be further shown that if $S_{\min}$ and $S_{\max}$ satisfy the following conditions:

$$S_{\min} \geq \frac{16rK^2(1+\eta)}{t^2}\log n \tag{2.34}$$

$$S_{\max} \leq n^r - \frac{16rK^2(1+\eta)}{t^2}\log n, \tag{2.35}$$

where η is any positive constant, then the test in (2.31) is consistent.

We here note an important fact that as long as the largest anomalous rectangle is not the entire lattice network, it can at most contain $n^r - n^{r-1}$ nodes, which satisfies the condition (2.35) for large n as well as the following condition

$$n^r - S_{\max} \to \infty \text{ as } n \to \infty. \tag{2.36}$$

Consequently, (2.35) and (2.36) are equivalent, both requiring the largest anomalous rectangle not to be the entire network. Thus, the conditions (2.34) and (2.36) imply that the test (2.31) is

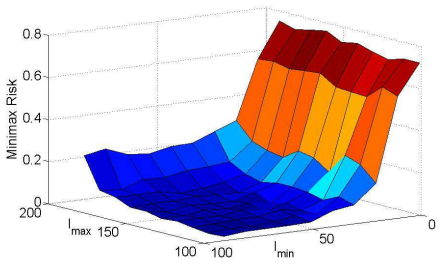

Fig. 2.4: Minimax risk for a line network.

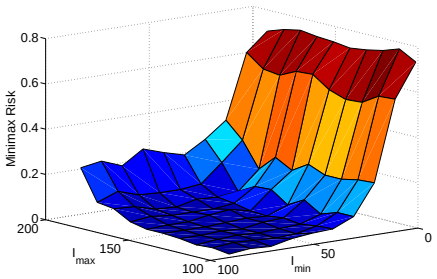

Fig. 2.5: Minimax risk for a ring network.

universally consistent for any arbitrary p and q, if

$$S_{\min} = \omega(\log n), \quad \text{and} \quad n^r - S_{\max} \to \infty \text{ as } n \to \infty. \tag{2.37}$$

Furthermore, following the arguments similar to those for the line network, it can be shown that any test must satisfy the following necessary conditions required on $S_{\min}$ and $S_{\max}$ in order to be *universally consistent* for arbitrary p and q:

$$S_{\min} = \omega(\log n), \quad \text{and} \quad n^r - S_{\max} \to \infty \text{ as } n \to \infty. \tag{2.38}$$

Therefore, combining the above sufficient and necessary conditions, we conclude the following optimality property for the proposed test.

Theorem 2.6 (Optimality)**.** *Consider the problem of nonparametric detection of a rectangle over a lattice network. The MMD-based test* (2.31) *is order level optimal to guarantee universal test consistency for arbitrary* p *and* q*.*

2.5 Numerical Results

In this section, we provide numerical results to demonstrate the performance of our tests and compare our approach with other competitive approaches.

In the first experiment, we apply our test to the line and ring network. We set the network size

Table 2.1: Minimax risk for a line network

$I_{\min} \setminus I_{\max}$	100	110	130	160	190
1	0.73	0.73	0.72	0.76	0.76
11	0.58	0.61	0.55	0.59	0.59
31	0.09	0.11	0.10	0.09	0.27
61	0.03	0.03	0.05	0.06	0.21
91	0.02	0.02	0.04	0.05	0.24

Table 2.2: Minimax risk for a ring network.

$I_{\min} \setminus I_{\max}$	100	110	130	160	190
1	0.73	0.78	0.74	0.74	0.71
11	0.58	0.61	0.53	0.60	0.63
31	0.09	0.10	0.12	0.12	0.27
61	0.03	0.03	0.03	0.05	0.27
91	0.02	0.03	0.03	0.05	0.24

$n = 200$, the distribution p to be Gaussian with mean zero and variance one, and the anomalous distribution q to be Gaussian with mean one and variance one. We use Gaussian kernel with $\sigma = 1$. In Figures 2.4 and 2.5, we plot the minimax risk (normalized by 2) for line and ring network as functions of $I_{\min}$ and $I_{\max}$. For further illustration, we also list some values of the two risk functions in Tables 2.1 and 2.2. It can be seen from Tables 2.1 and 2.2, and Figures 2.4 and 2.5 that the risk functions decrease as $I_{\min}$ increases and as $I_{\max}$ decreases. This is reasonable because as $I_{\min}$ increases and as $I_{\max}$ decreases, the number of candidate anomalous intervals decreases, which reduces the difficulty of detection. The minimum numbers of samples inside and outside the anomalous interval also increase, respectively, which provide more accurate information about the distributions.

In the next experiment, we compare the performance of our test with other competitive tests including the student t-test, the Smirnov test [70], the Wolf test [70], Hall test [71], kernel-based KFDA test [72] and kernel-based KDR test [73]. We consider a line network with the network size $n = 100$. We set the distribution p to be Gaussian with zero mean and variance 2, and set the anomalous distribution q to be a mixture of Gaussian distributions with equal probability taking $\mathcal{N}(-1, 1)$ and $\mathcal{N}(1, 1)$. Hence, distributions p and q have the same mean and variance.

Table 2.3: Comparison of nonparametric approaches over a line network.

$(I_{\min}, I_{\max})$	t-test	Smirnov	KFDA	KDR	MMD
(10,95)	0.90	0.92	0.70	0.66	**0.66**
(10,50)	0.88	0.90	0.51	0.56	**0.55**
(45,95)	0.89	0.93	0.54	0.43	**0.43**
(45,50)	0.83	0.62	0.06	0.06	**0.05**

In Table 2.3, we list some values of the risk function of our MMD-based test and other nonparametric tests with respect to various values of $I_{\min}$ and $I_{\max}$. It can be seen that the student t-test fails, because the test relies on difference in mean and variance to distinguish two distributions, which are the same in our experiment. The Smirnov test estimates the cumulative distribution function (CDF) first and then takes the maximum difference of the two cumulative distribution functions as the test statistics. For continuous distributions, accurately estimating the CDF from samples requires a large amount of data, which is not feasible in our experiment. For the three kernel-based tests KFDA, KDR and MMD, the performance are very close. In particular, for large enough $I_{\min}$ and small enough $I_{\max}$, the kernel-based tests yield small risk. Among these three kernel-based tests, MMD has a slightly better performance. In terms of the computational complexity, KFDA is much higher than KDR and MMD-based tests.

2.6 Proof of Theorem 2.1: Performance Guarantee

We first introduce the McDiarmid's inequality which is useful in bounding the probability of error in our proof.

Lemma 2.1 (McDiarmid's Inequality)**.** *Let $f : \mathcal{X}^m \to \mathbb{R}$ be a function such that for all $i \in \{1, \ldots, m\}$, there exist $c_i < \infty$ for which*

$$\sup_{X \in \mathcal{X}^m, \tilde{x} \in \mathcal{X}} |f(x_1, \ldots, x_m) - f(x_1, \ldots x_{i-1}, \tilde{x}, x_{i+1}, \ldots, x_m)| \leq c_i. \tag{2.39}$$

Then for any probability measure P_X over m independent random variables $X := (X_1 \ldots, X_m)$,

and every $\epsilon > 0$*,*

$$P_X\left(f(X) - E_X(f(X)) > \epsilon\right) < \exp\left(-\frac{2\epsilon^2}{\sum_{i=1}^m c_i^2}\right), \tag{2.40}$$

where E_X *denotes the expectation over* P_X*.*

We now derive bounds on $P(H_1|H_0)$ and $P(H_0|H_{1,I})$ for the test (2.6). We first have

$$\begin{aligned}\text{MMD}^2_{u,I}(Y_I, Y_{\bar{I}}) =& \frac{1}{|I|(|I|-1)}\sum_{i\in I}\sum_{\substack{j\neq i\\ j\in I}} k(y_i, y_j) + \frac{1}{(n-|I|)(n-|I|-1)}\sum_{i\notin I}\sum_{\substack{j\neq i\\ j\notin I}} k(y_i, y_j)\\ &- \frac{2}{|I|(n-|I|)}\sum_{i\in I}\sum_{j\notin I} k(y_i, y_j).\end{aligned} \tag{2.41}$$

Under H_0, all samples are generated from distribution p. Hence, $\mathbb{E}[\text{MMD}^2_{u,I}(Y_I, Y_{\bar{I}})] = 0$.

In order to apply the McDiarmid's inequality to bound the error probabilities $P(H_1|H_0)$ and $P(H_0|H_{1,I})$, we evaluate the following quantities. There are n variables that affects the value of $\text{MMD}^2_{u,I}(Y_I, Y_{\bar{I}})$. We study the influence of these n variables on $\text{MMD}^2_{u,I}(Y_I, Y_{\bar{I}})$ in the following two cases. For $i \in I$, change of y_i affects $\text{MMD}^2_{u,I}(Y_I, Y_{\bar{I}})$ through the following terms,

$$\frac{2}{|I|(|I|-1)}\sum_{\substack{j\neq i\\ j\in I}} k(y_i, y_j) - \frac{2}{|I|(n-|I|)}\sum_{j\notin I} k(y_i, y_j). \tag{2.42}$$

For $i \notin I$, change of y_i affects $\text{MMD}^2_{u,I}(Y_I, Y_{\bar{I}})$ through the following terms,

$$\frac{2}{(n-|I|)(n-|I|-1)}\sum_{\substack{j\neq i\\ j\notin I}} k(y_i, y_j) - \frac{2}{|I|(n-|I|)}\sum_{j\in I} k(y_i, y_j). \tag{2.43}$$

Since the kernel we use is bounded, i.e., $0 \leq k(x, y) \leq K$ for any x, y, we have that for $i \in I$, $c_i = \frac{4K}{|I|}$, and for $i \notin I$, $c_i = \frac{4K}{n-|I|}$, where c_i serves the role as in (2.39).

Therefore, by applying McDiarmid's inequality, we obtain

$$P_{H_0}(\text{MMD}^2_{u,I}(Y_I, Y_{\bar{I}}) > t) \leq \exp\left(-\frac{2t^2|I|(n-|I|)}{16nK^2}\right). \tag{2.44}$$

Hence,

$$
\begin{aligned}
P(H_1|H_0) &= P_{H_0}\left(\max_{I:I\in\mathcal{I}_n^{(a)}} \mathrm{MMD}^2_{u,I}(Y_I, Y_{\bar{I}}) > t\right) \\
&\overset{(a)}{\leq} \sum_{I:I\in\mathcal{I}_n^{(a)}} P_{H_0}(\mathrm{MMD}^2_{u,I}(Y_I, Y_{\bar{I}}) > t) \\
&\leq \sum_{I:I\in\mathcal{I}_n^{(a)}} \exp\left(-\frac{2t^2|I|(n-|I|)}{16nK^2}\right) \\
&= \sum_{I:I_{\min}\leq|I|\leq I_{\max}} \exp\left(-\frac{2t^2|I|(n-|I|)}{16nK^2}\right) \\
&\overset{(b)}{=} \sum_{I:I_{\min}\leq|I|\leq n-\frac{16K^2(1+\eta)}{t^2}\log n} \exp\left(-\frac{2t^2|I|(n-|I|)}{16nK^2}\right) \\
&\quad + \sum_{I:n-\frac{16K^2(1+\eta)}{t^2}\log n+1\leq|I|\leq I_{\max}} \exp\left(-\frac{2t^2|I|(n-|I|)}{16nK^2}\right)
\end{aligned} \tag{2.45}
$$

where (a) is due to Boole's inequality, η in (b) is a positive constant, and the second term in (b) is equal to zero if $n - \frac{16K^2(1+\eta)}{t^2}\log n + 1 \geq I_{\max}$.

It can be shown that if $I_{\min} \geq \frac{16K^2(1+\eta)}{t^2}\log n$, then the first term in (2.45) can be bounded as follows,

$$
\begin{aligned}
&\sum_{I:I_{\min}\leq|I|\leq n-\frac{16K^2(1+\eta)}{t^2}\log n} \exp\left(-\frac{2t^2|I|(n-|I|)}{16nK^2}\right) \\
&\qquad \overset{(a)}{\leq} n^2 \exp\left(-\frac{2t^2|I|(n-|I|)}{16nK^2}\right)\Big|_{|I|=\frac{16K^2(1+\eta)}{t^2}\log n} \\
&\qquad = \exp\left(-2\eta\log n + o(n)\right) \to 0, \text{ as } n\to\infty,
\end{aligned} \tag{2.46}
$$

where (a) is due to the fact that there are at most n^2 number of candidate anomalous intervals contributing to the sum, and $|I|(n-|I|)$ is minimized by the value $|I| = \frac{16K^2(1+\eta)}{t^2}\log n$ within the range of $|I|$.

We next bound the second term in (2.45).

$$\sum_{n-\frac{16K^2(1+\eta)}{t^2}\log n+1\le|I|\le I_{\max}} \exp\left(-\frac{2t^2|I|(n-|I|)}{16nK^2}\right) \tag{2.47}$$

$$= \sum_{n-\frac{16K^2(1+\eta)}{t^2}\log n+1\le|I|\le n-\frac{16K^2(1+\eta)}{t^2}\log\log n} \exp\left(-\frac{2t^2|I|(n-|I|)}{16nK^2}\right) \tag{2.48}$$

$$+ \sum_{n-\frac{16K^2(1+\eta)}{t^2}\log\log n+1\le|I|\le I_{\max}} \exp\left(-\frac{2t^2|I|(n-|I|)}{16nK^2}\right)$$

$$\le \left(\frac{16K^2(1+\eta)}{t^2}\log n\right)^2 \exp\left(-\frac{2t^2(n-\frac{16K^2(1+\eta)}{t^2}\log\log n)\frac{16K^2(1+\eta)}{t^2}\log\log n}{16nK^2}\right)$$

$$+ \sum_{n-\frac{16K^2(1+\eta)}{t^2}\log\log n+1\le|I|\le I_{\max}} \exp\left(-\frac{2t^2|I|(n-|I|)}{16nK^2}\right), \tag{2.49}$$

where the first term in (2.49) converges to zero as n goes to infinity. The second term in (2.49) can be bounded as

$$\sum_{n-\frac{16K^2(1+\eta)}{t^2}\log\log n+1\le|I|\le I_{\max}} \exp\left(-\frac{2t^2|I|(n-|I|)}{16nK^2}\right)$$

$$\le \left(\frac{16K^2(1+\eta)}{t^2}\log\log n\right)^2 \exp\left(-\frac{2t^2 I_{\max}(n-I_{\max})}{16nK^2}\right) \tag{2.50}$$

which converges to zero as $n\to\infty$ if

$$I_{\max} \le n - \frac{16K^2(1+\eta)}{t^2}\log\log\log n. \tag{2.51}$$

In fact, the condition (2.51) can be further relaxed by decomposing the second term in (2.49) following the steps similar to (2.48) and (2.49). Such a procedure can be repeated for arbitrary

finite times, say $k-2$ times, and it can be shown that (2.47) converges to zero as $n \to \infty$ if

$$I_{\max} \leq n - \frac{16K^2(1+\eta)}{t^2} \underbrace{\log \cdots \log \log}_{\text{arbitrary } k \text{ number of } \log} n \ . \tag{2.52}$$

Therefore, we conclude that the type I error, i.e., $P(H_1|H_0)$, converges to zero as $n \to \infty$ if the following conditions are satisfied:

$$\begin{aligned} I_{\min} &\geq \frac{16K^2(1+\eta)}{t^2} \log n \\ I_{\max} &\leq n - \frac{16K^2(1+\eta)}{t^2} \underbrace{\log \cdots \log \log}_{\text{arbitrary } k \text{ number of } \log} n \end{aligned} \tag{2.53}$$

for any positive integer k.

We next continue to bound the type II error $\max\limits_{I \in \mathcal{I}_n^{(a)}} P(H_0|H_{1,I})$ as follows.

$$\begin{aligned} &\max_{I \in \mathcal{I}_n^{(a)}} P(H_0|H_{1,I}) \\ &\quad = \max_{I \in \mathcal{I}_n^{(a)}} P_{H_{1,I}} \left(\max_{I' \in \mathcal{I}_n^{(a)}} \mathbf{MMD}^2_{u,I'}(Y_{I'}, Y_{\bar{I}'}) < t \right) \\ &\quad \overset{(a)}{\leq} \max_{I \in \mathcal{I}_n^{(a)}} P_{H_{1,I}} (\mathbf{MMD}^2_{u,I}(Y_I, Y_{\bar{I}}) < t) \\ &\quad = \max_{I \in \mathcal{I}_n^{(a)}} P_{H_{1,I}} (\mathbf{MMD}^2[p,q] - \mathbf{MMD}^2_{u,I}(Y_I, Y_{\bar{I}}) > \mathbf{MMD}^2[p,q] - t) \\ &\quad \overset{(b)}{\leq} \max_{I \in \mathcal{I}_n^{(a)}} \exp \left(-\frac{(\mathbf{MMD}^2[p,q] - t)^2 |I|(n-|I|)}{8K^2 n} \right) \end{aligned} \tag{2.54}$$

where (a) holds by taking $I' = I$, and (b) holds by applying McDiarmid's inequality. It can be

easily checked that under the condition (2.53),

$$\begin{aligned}
&\max_{I\in\mathcal{I}_n^{(a)}} P(H_0|H_{1,I}) \\
&\qquad \leq \exp\left(-\frac{(\text{MMD}^2[p,q]-t)^2|I|(n-|I|)}{8K^2n}\right)\Big|_{|I|=n-\frac{16K^2(1+\eta)}{t^2}\underbrace{\log\cdots\log\log}_{\text{arbitrary } k \text{ number of } \log} n} \\
&\qquad \to 0, \text{ as } n\to\infty. \qquad (2.55)
\end{aligned}$$

Therefore, we conclude that the condition (2.53) guarantees that $R_m^{(n)} \to 0$ as $n \to \infty$, and thus guarantees the consistency of the test (2.6).

2.7 Proof of Theorem 2.2: Necessary Conditions

The idea is to consider the following problem which has lower risk than our original problem, and show that there exist distributions (in fact for Gaussian p and q), under which such a risk is bounded away from zero for all tests if the necessary conditions are not satisfied.

First consider the following problem, in which all candidate anomalous intervals have the same length k, and hence there are in total $n-k+1$ candidate anomalous intervals. Furthermore, suppose the distribution p is Gaussian with mean zero and variance one, and the distribution q is Gaussian with mean $\mu > 0$ and variance one. We define the minimax risk of a test for such a problem as follows:

$$R_m(k) = P(H_1|H_0) + \max_{|I|=k} P(H_0|H_{1,I}), \qquad (2.56)$$

and we denote the minimum minimax risk as $R_m^*(k)$. We further assign uniform distribution over all candidate anomalous intervals under the alternative hypothesis H_1, i.e., each candidate

anomalous interval has the same probability $\frac{1}{n-k+1}$ to occur. Thus the Bayes risk is given by

$$R_b = P(H_1|H_0) + \frac{1}{n-k+1}\sum_{|I|=k} P(H_0|H_{1,I}), \tag{2.57}$$

and we use R_b^* to denote the minimum Bayes risk over all possible tests. It is clear that

$$R_m^*(k) \geq R_b^*.$$

It is justified in [4] that the optimal Bayes risk R_b^* can be lower bounded as follows.

$$R_b^* \geq 1 - \frac{1}{2}\sqrt{\mathbb{E}e^{\mu^2 Z} - 1}, \tag{2.58}$$

where $Z = |S \bigcap S'|$ with S and S' being independently and uniformly drawn at random from all candidate anomalous intervals.

We next characterize the distribution of the random variable Z in order to evaluate the lower bound. We are interested only in the case with $k < \frac{n}{2}$. It can be shown that

$$\begin{aligned} P(Z=i) &= \frac{2(n-2k+1+i)}{(n-k+1)^2}, \text{ for } 1 \leq i \leq k-1 \\ P(Z=k) &= \frac{1}{n-k+1} \\ P(Z=0) &= 1 - \sum_{i=1}^{k-1} \frac{2(n-2k+1+i)}{(n-k+1)^2} - \frac{1}{n-k+1}. \end{aligned} \tag{2.59}$$

Based on the above distribution of Z, we obtain

$$\begin{aligned}
&\mathbb{E}e^{\mu^2 Z} - 1 \\
&= \sum_{j=1}^{k-1} \frac{2(n-2k+1+j)e^{\mu^2 j}}{(n-k+1)^2} + \frac{e^{\mu^2 k}}{n-k+1} + 1 - \frac{2(k-1)(n-\frac{3}{2}k+1)+n-k+1}{(n-k+1)^2} - 1 \\
&= \frac{2(n-2k+1)}{(n-k+1)^2} \sum_{j=1}^{k-1} e^{\mu^2 j} + \frac{2}{(n-k+1)^2} \sum_{j=1}^{k-1} j e^{\mu^2 j} + \frac{e^{\mu^2 k}}{n-k+1} \\
&\quad - \frac{2(k-1)(n-\frac{3}{2}k+1)+n-k+1}{(n-k+1)^2} \\
&\overset{(a)}{\leq} \frac{2(n-2k+1)}{(n-k+1)^2} \int_1^k e^{\mu^2 x} dx + \frac{2}{(n-k+1)^2} \int_1^k x e^{\mu^2 x} dx + \frac{e^{\mu^2 k}}{n-k+1} \\
&\quad - \frac{2(k-1)(n-\frac{3}{2}k+1)+n-k+1}{(n-k+1)^2} \\
&= \frac{2(n-2k+1)}{(n-k+1)^2}(e^{\mu^2 k} - e^{\mu^2}) + \frac{2}{(n-k+1)^2}(\frac{1}{\mu^2} k e^{\mu^2 k} - \frac{1}{\mu^4} e^{\mu^2 k} - \frac{1}{\mu^2} e^{\mu^2} + \frac{1}{\mu^4} e^{\mu^2}) \\
&\quad + \frac{e^{\mu^2 k}}{n-k+1} - \frac{2(k-1)(n-\frac{3}{2}k+1)+n-k+1}{(n-k+1)^2}.
\end{aligned} \tag{2.60}$$

It can be checked that if $k \leq \frac{1}{2\mu^2} \log n$, (2.60) converges to zero as n goes to infinity. Hence, $R_b^* \geq 1$ as n goes to infinity, which further implies that $R_m^*(k) > 1$, as $n \to \infty$, and thus any test is no better than random guess. Since μ can be any constant, there always exists Gaussian p and q such that no test can be consistent as long as $k \leq c \log n$, where c is any constant.

Now consider the original problem with the minimax risk

$$R_m = P(H_1|H_0) + \max_{I_{\min} \leq |I| \leq I_{\max}} P(H_0|H_{1,I}). \tag{2.61}$$

It can be shown that

$$R_m^* \geq R^*(k), \text{ if } k = I_{\min}$$

where R_m^* denotes the minimum risk over all possible tests. Based on the above argument on $R^*(k)$, it is clear that if $I_{\min} \leq c \log n$, there exists no test such that R_m^* converges to zero as n goes to infinity for arbitrary distributions p and q.

Furthermore, consider the case with only one candidate anomalous interval I with length k. The risk in this case is

$$R(k) = P(H_1|H_0) + P(H_0|H_{1,I}) \tag{2.62}$$

where $|I| = k$. It is also clear that $R_m^* \geq R^*(k)$ where $k = I_{\max}$. For such a simple case, the problem reduces to the two-sample problem, detecting whether the samples in the interval I and the samples outside of the interval I are generated from the same distribution. In order to guarantee $R^*(k) \to 0$ as $n \to \infty$, k and $n-k$ should both scale with n to infinity. Thus, in order to guarantee $R_m^* \to 0$, as $n \to \infty$, $n - I_{\max} \to \infty$ is necessary for any test to be universally consistent. This concludes the proof.

2.8 Proof of Sufficient Conditions for Ring Networks

Following steps similar to those in Section 2.6, we derive the following bound

$$\begin{aligned}
P(H_1|H_0) &\leq \sum_{I \in \mathcal{I}_n^{(a)}} \exp\left(-\frac{2t^2|I|(n-|I|)}{16nK^2}\right) \\
&\overset{(a)}{=} \sum_{i=I_{\min}}^{I_{\max}} n \exp\left(-\frac{2t^2 i(n-i)}{16nK^2}\right) \\
&\overset{(b)}{\leq} \sum_{i=I_{\min}}^{I_{\max}} n \exp\left(-\frac{2t^2 \min\{I_{\min}(n-I_{\min}), I_{\max}(n-I_{\max})\}}{16nK^2}\right) \\
&\overset{(c)}{\leq} n^2 \exp\left(-\frac{2t^2 \min\{I_{\min}(n-I_{\min}), I_{\max}(n-I_{\max})\}}{16nK^2}\right) \\
&= \exp\left(2\log n - \frac{2t^2 \min\{I_{\min}(n-I_{\min}), I_{\max}(n-I_{\max})\}}{16nK^2}\right)
\end{aligned} \tag{2.63}$$

where (a) is due to the fact in the ring network, there are n candidate anomalous intervals with size i, (b) is due to the fact that $i(n-i)$ is lowered bounded by $\min\{I_{\min}(n-I_{\min}), I_{\max}(n-I_{\max})\}$, and (c) is due to the fact that $I_{\max} - I_{\min} \leq n$.

It can be checked that $P(H_1|H_0) \to 0$ as $n \to \infty$ if

$$\frac{16K^2(1+\eta)}{t^2}\log n \leq I_{\min} \leq I_{\max} \leq n - \frac{16K^2(1+\eta)}{t^2}\log n. \tag{2.64}$$

Furthermore, following steps similar to those in Section 2.6, we can derive an upper bound on the type II error and show that it converges to zero as $n \to \infty$ if

$$I_{\min} \to \infty, \quad n - I_{\max} \to \infty, \tag{2.65}$$

Combining the two conditions completes the proof.

2.9 Proof of Necessary Conditions for Ring Networks

The proof follows the idea in Section 2.7 for the line network. Here, we consider a problem in which all the candidate anomalous intervals have the same length k, i.e., there are in total $n-k+1$ candidate anomalous intervals. Furthermore, suppose the distribution p is Gaussian with mean zero and variance one, and the distribution q is Gaussian with mean $\mu > 0$ and variance one. We define the minimax risk of a test for such a problem as follows:

$$R_m(k) = P(H_1|H_0) + \max_{|I|=k} P(H_0|H_{1,I}), \tag{2.66}$$

and we denote the minimum minimax risk as $R^*(k)$. We further assign uniform distribution over all candidate anomalous intervals under the alternative hypothesis H_1, i.e., each candidate anomalous interval has the same probability $\frac{1}{n-k+1}$ to be anomalous. Hence, the Bayes risk is given by

$$R_b = P(H_1|H_0) + \frac{1}{n}\sum_{|I|=k} P(H_0|H_{1,I}), \tag{2.67}$$

and we use R_b^* to denote the minimum Bayes risk over all possible tests. It is clear that

$$R_k^* \geq R_b^*.$$

In order to apply the same property in (2.58), we characterize the distribution of the random variable Z as follows.

$$\begin{aligned} P(Z=i) &= \frac{2}{n}, \text{ for } i = 1, \ldots, k-1, \\ P(Z=k) &= \frac{1}{n}, \\ P(Z=0) &= \frac{n-2k+1}{n}. \end{aligned} \tag{2.68}$$

Then we have

$$\begin{aligned} &\mathbb{E}e^{\mu^2 Z} - 1 \\ &= \sum_{i=1}^{k-1} \frac{2}{n} e^{\mu^2 i} + \frac{1}{2} e^{\mu^2 k} + \frac{n-2k+1}{n} - 1 \\ &\leq \frac{2}{n\mu^2} e^{\mu^2 k} - \frac{2}{n} e^{\mu^2} + \frac{1}{n} e^{\mu^2 k} - \frac{2k-1}{n} \\ &\to 0, \text{ if } k \leq \mathcal{O}(\log n). \end{aligned} \tag{2.69}$$

By arguments similar to those for the line network, we conclude that $I_{\min} > c \log n$ for any constant c and $n - I_{\max} \to \infty$ as $n \to \infty$ are necessary to guarantee any test to be universally consistent for arbitrary distributions p and q.

CHAPTER 3
ANOMALOUS DATA STREAM DETECTION

In this chapter, we study the problem of anomalous data stream detection. In Section 3.1, we introduce our problem model and performance metric. In Section 3.2, we present our tests and results on the sufficient condition on the sample complexity for consistency of MMD-based tests. In Section 3.3, we present our results on the necessary condition for any universally consistent test. In Section 3.4, we present our numerical results.

3.1 Problem Statement

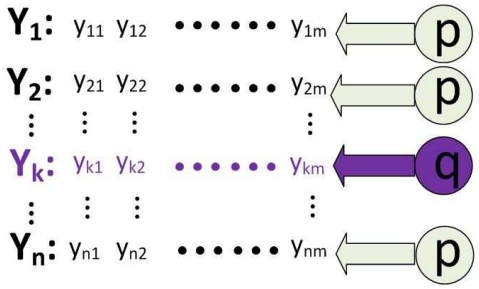

Fig. 3.1: An anomalous data stream detection problem.

In this chapter, we study an anomalous data stream detection problem (see Figure 3.1), in which there are in total n data sequences denoted by Y_k for $1 \leq k \leq n$. Each data sequence Y_k consists of

m i.i.d. samples $y_{k1}, \ldots, y_{km}$ drawn from either a typical distribution p or an anomalous distribution q, where $p \neq q$. In the sequel, we use the notation $Y_k := (y_{k1}, \ldots, y_{km})$. We assume that the distributions p and q are arbitrary and unknown in advance. Our goal is to build distribution-free tests to detect data sequences generated by the anomalous distribution q.

We assume that s out of n data sequences are anomalous, i.e., are generated by the anomalous distribution q. We study both cases with s known and unknown, respectively. We are interested in the asymptotical regime, in which the number n of data sequences goes to infinity. We assume that the number s of anomalous sequences satisfies $\frac{s}{n} \to \alpha$ as $n \to \infty$, where $0 \leq \alpha \leq 1$. This includes the following three cases: (1) s is fixed, and nonzero as $n \to \infty$; (2) $s \to \infty$, but $\frac{s}{n} \to 0$ as $n \to \infty$; and (3) $\frac{s}{n}$ approaches to a positive constant, which is less than or equal to 1. Some of our results are also applicable to the case with $s = 0$, i.e., the null hypothesis in which there is no anomalous sequence. We will comment on such a case when the corresponding results are presented.

We next define the probability of detection error as the performance measure of tests. We let $\mathcal{I}$ denote the set that contains indices of all anomalous data sequences. Hence, the cardinality $|\mathcal{I}| = s$. We let $\hat{\mathcal{I}}^n$ denote a sequence of index sets that contain indices of all anomalous data sequences claimed by a corresponding sequence of tests.

Definition 3.1. *A sequence of tests are said to be consistent if*

$$\lim_{n\to\infty} P_e = \lim_{n\to\infty} P\{\hat{\mathcal{I}}^n \neq \mathcal{I}^n\} = 0. \tag{3.1}$$

We note that the above definition of consistency is with respect to the number n of sequences instead of the number m of samples. However, as n becomes large (and possibly as s becomes large), it is increasingly challenging to consistently detect all anomalous data sequences. It then requires that the number m of samples becomes large enough in order to more accurately detect anomalous sequences. Therefore, the limit in the above definition in fact refers to the asymptotic regime, in which m scales fast enough as n goes to infinity in order to guarantee asymptotically

small probability of error.

Furthermore, for a consistent test, it is also desirable that the error probability decays exponentially fast with respect to the number m of samples.

Definition 3.2. *A sequence of tests are said to be exponentially consistent if*

$$\liminf_{m\to\infty} -\frac{1}{m}\log P_e = \liminf_{m\to\infty} -\frac{1}{m}\log P\{\hat{\mathcal{I}}^n \neq \mathcal{I}^n\} > 0. \tag{3.2}$$

In this chapter, our goal is to construct distribution-free tests for detecting anomalous sequences, and characterize the scaling behavior of m with n (and possibly s) so that the developed tests are consistent (and possibly exponentially consistent).

An example with sparse anomalous samples. In this chapter, we also study an interesting example, in which the distribution q is a mixture of the distribution p with probability $1-\epsilon$ and an anomalous distribution $\tilde{q}$ with probability ϵ, where $0 < \epsilon \leq 1$, i.e., $q = (1-\epsilon)p + \epsilon\tilde{q}$. It can be seen that if ϵ is small, the majority of samples in an anomalous sequence are drawn from the distribution p, and only sparse samples are drawn from the anomalous distribution $\tilde{q}$. The value of ϵ captures the sparsity level of anomalous samples. Here, ϵ can scale as n increases, and is hence denoted by ϵ_n. We study how ϵ_n affects the number of samples needed for consistent detection.

3.2 Test and Performance Guarantee

In this section, we study both cases with s known and unknown, respectively. We then study the example with sparse anomalous samples.

3.2.1 Known s

In this subsection, we consider the case with s known. We start with a simple case with $s = 1$, and then study the more general case, in which $\frac{s}{n} \to \alpha$ as $n \to \infty$, where $0 \leq \alpha \leq 1$.

Consider the case with $s = 1$. For each sequence Y_k, we use $\overline{Y}_k$ to denote the $(n-1)m$

dimensional sequence that stacks all other sequences together, as given by

$$\overline{Y}_k = \{Y_1, \ldots, Y_{k-1}, Y_{k+1}, \ldots, Y_n\}.$$

We then compute $\text{MMD}_u^2[Y_k, \overline{Y}_k]$ for $1 \leq k \leq n$. It is clear that if Y_k is the anomalous sequence, then $\overline{Y}_k$ is fully composed of typical sequences. Hence, $\text{MMD}_u^2[Y_k, \overline{Y}_k]$ is a good estimator of $\text{MMD}^2[p, q]$, which is a positive constant. On the other hand, if Y_k is a typical sequence, $\overline{Y}_k$ is composed of $n-2$ sequences generated by p and only one sequence generated by q. As n increases, the impact of the anomalous sequence on $\overline{Y}_k$ is negligible, and $\text{MMD}_u^2[Y_k, \overline{Y}_k]$ should be asymptotically close to zero. Based on the above understanding, we construct the following test when $s = 1$. The sequence k^* is claimed to be anomalous if

$$k^* = \arg \max_{1 \leq k \leq n} \text{MMD}_u^2[Y_k, \overline{Y}_k]. \tag{3.3}$$

The following proposition characterizes the condition under which the above test is consistent.

Proposition 3.1. *Consider the anomalous data stream detection problem with one anomalous sequence, i.e., $s = 1$. Suppose the test* (3.3) *applies a bounded kernel with $0 \leq k(x, y) \leq K$ for any (x, y). Then, the probability of error is upper bounded as follows,*

$$P_e \leq \exp\Big(\log n - \frac{m(\textit{MMD}^2[p, q] - \xi)^2}{16K^2(1 + \Theta(\frac{1}{n}))} \Big), \tag{3.4}$$

where ξ is a constant which can be picked arbitrarily close to zero. Furthermore, the test (3.3) *is exponentially consistent if*

$$m \geq \frac{16K^2(1+\eta)}{\textit{MMD}^4[p, q]} \log n, \tag{3.5}$$

where η is any positive constant.

Proof. See Section 3.5. □

Proposition 3.1 implies that for the scenario with one anomalous sequence, $\Omega(\log n)$ samples are sufficient to guarantee consistent detection.

We next consider the case with $s \geq 1$. More specifically, we consider the case with $\frac{s}{n} \to \alpha$ as $n \to \infty$, where $0 \leq \alpha < \frac{1}{2}$. Although we focus on the case with $\alpha < \frac{1}{2}$, the case with $\alpha > \frac{1}{2}$ is similar, with the roles of p and q being exchanged. We first study the case with s known. Our test is a natural generalization of the test (3.3) except now the test picks the sequences with the largest s values of $\text{MMD}_u^2[Y_k, \overline{Y}_k]$, which is given by

$$\begin{aligned}\hat{\mathcal{I}} =&\{k : \text{MMD}_u^2[Y_k, \overline{Y}_k] \text{ is among the } s \text{ largest} \\ &\text{values of } \text{MMD}_u^2[Y_i, \overline{Y}_i] \text{ for } i = 1, \ldots, n\}. \end{aligned} \tag{3.6}$$

The following theorem characterizes the condition under which the above test is consistent.

Theorem 3.1. *Consider the anomalous data stream detection problem with s anomalous sequences, where $\frac{s}{n} \to \alpha$ as $n \to \infty$ and $0 \leq \alpha < \frac{1}{2}$. Assume the value of s is known. Further assume that the test* (3.6) *applies a bounded kernel with $0 \leq k(x, y) \leq K$ for any (x, y). Then the probability of error is upper bounded as follows,*

$$P_e \leq \exp\Big(\log((n-s)s) - \frac{m((1-2\alpha)\textit{MMD}^2[p,q] - \xi)^2}{16K^2(1+\Theta(\frac{1}{n}))}\Big), \tag{3.7}$$

where ξ is a constant which can be picked arbitrarily close to zero. Furthermore, the test (3.6) *is exponentially consistent for any p and q if*

$$m \geq \frac{16K^2(1+\eta)}{(1-2\alpha)^2\textit{MMD}^4[p,q]} \log(s(n-s)), \tag{3.8}$$

where η is any positive constant.

Proof. We analyze the performance of the test (3.6). Without loss of generality, we assume that the first s sequences are anomalous and are generated from distribution q. Hence, the probability

of error can be bounded as,

$$\begin{aligned} P_e =& P\bigg(\exists k > s : \text{MMD}_u^2[Y_k, \overline{Y}_k] > \min_{1\le l\le s} \text{MMD}_u^2[Y_l, \overline{Y}_l]\bigg) \\ &\le \sum_{k=s+1}^{n}\sum_{l=1}^{s} P\bigg(\text{MMD}_u^2[Y_k, \overline{Y}_k] > \text{MMD}_u^2[Y_l, \overline{Y}_l]\bigg). \end{aligned} \tag{3.9}$$

Using the fact that $\frac{s}{n} \to \alpha$, where $0 \le \alpha < \frac{1}{2}$, and using (3.51) and (3.52), we can show that

$$\mathbb{E}\big[\text{MMD}_u^2[Y_l, \overline{Y}_l]\big] \to (1-\alpha)^2\text{MMD}^2[p, q], \tag{3.10}$$

as $n \to \infty$ for $1 \le l \le s$, and

$$\mathbb{E}\big[\text{MMD}_u^2[Y_k, \overline{Y}_k]\big] \to \alpha^2\text{MMD}^2[p, q], \tag{3.11}$$

as $n \to \infty$ for $s+1 \le k \le n$. Hence, there exists a constant ξ such that

$$0 < \xi < (1-\alpha)^2\text{MMD}^2[p, q] - \alpha^2\text{MMD}^2[p, q]$$

and

$$\mathbb{E}\big[\text{MMD}_u^2[Y_k, \overline{Y}_k] - \text{MMD}_u^2[Y_l, \overline{Y}_l]\big] < \alpha^2\text{MMD}^2[p, q] - (1-\alpha)^2\text{MMD}^2[p, q] + \xi, \tag{3.12}$$

for large enough n.

Therefore, we obtain,

$$
\begin{aligned}
&P\bigg(\text{MMD}_u^2[Y_k,\overline{Y}_k]-\text{MMD}_u^2[Y_l,\overline{Y}_l]>0\bigg)\\
&=P\bigg(\text{MMD}_u^2[Y_k,\overline{Y}_k]-\text{MMD}_u^2[Y_l,\overline{Y}_l]-\mathbb{E}\big[\text{MMD}_u^2[Y_k,\overline{Y}_k]-\text{MMD}_u^2[Y_l,\overline{Y}_l]\big]\\
&\qquad>-\mathbb{E}\big[\text{MMD}_u^2[Y_k,\overline{Y}_k]-\text{MMD}_u^2[Y_l,\overline{Y}_l]\big]\bigg)\\
&\leq P\bigg(\text{MMD}_u^2[Y_k,\overline{Y}_k]-\text{MMD}_u^2[Y_l,\overline{Y}_l]-\mathbb{E}\big[\text{MMD}_u^2[Y_k,\overline{Y}_k]-\text{MMD}_u^2[Y_l,\overline{Y}_l]\big]\\
&\qquad>((1-\alpha)^2-\alpha^2)\text{MMD}^2[p,q])-\xi\bigg),
\end{aligned}
\tag{3.13}
$$

for large enough n.

Applying McDiarmid's inequality, we obtain,

$$
P_e\leq\exp\left(\log((n-s)s)-\frac{m((1-2\alpha)\text{MMD}^2[p,q]-\xi)^2}{16K^2(1+\Theta(\frac{1}{n}))}\right).
\tag{3.14}
$$

Since ξ can be arbitrarily small, we conclude that if,

$$
m\geq\frac{16K^2(1+\eta)}{(1-2\alpha)^2\text{MMD}^4[p,q]}\log(s(n-s)),
\tag{3.15}
$$

where η is any positive constant, then $P_e\to 0$, as $n\to\infty$. It is also clear that if the above condition is satisfied, P_e converges to zero exponentially fast with respect to m. □

We note that $\log((n-s)s)=\Theta(\log n)$, for $1\leq s<n$. Hence, Theorem 3.1 implies that even with s anomalous sequence, the test (3.6) requires only $\Omega(\log n)$ samples in each data sequence in order to guarantee consistency of the test. Hence, the increase of s does not affect the order level requirement on the sample size m. We further note that Theorem 3.1 is also applicable to the case in which $\alpha>\frac{1}{2}$ simply with the roles of p and q exchanged.

Remark 3.1. *For the case with $\frac{s}{n}\to 0$, as $n\to\infty$, we can also build a test with reduced computational complexity as follows. For each Y_k, instead of using $n-1$ sequences to build $\overline{Y}_k$ as in*

the test (3.6)*, we take any* l *sequences out of the remaining* $n-1$ *sequences to build a sequence* $\widetilde{Y}_k$*, such that* $\frac{l}{n} \to 0$ *and* $\frac{s}{l} \to 0$ *as* $n \to \infty$*. Such an* l *exists for any* s *and* n *satisfying* $\frac{s}{n} \to 0$ *(e.g.,* $l = \sqrt{sn}$*). It can be shown that using* $\widetilde{Y}_k$ *to replace* $\overline{Y}_k$ *in the test* (3.6) *still leads to consistent detection under the same condition given in Theorem 3.1. Since* l *is much smaller than* n*, computational complexity is substantially reduced.*

We note that Theorem 3.1 (which includes Proposition 3.1 as a special case) characterizes the conditions to guarantee test consistency for a pair of fixed but unknown distributions p and q. Hence, the condition (3.8) depends on the underlying distributions p and q. In fact, such a condition further yields the following condition that guarantees the test to be universally consistent for arbitrary p and q.

Proposition 3.2 (Universal Consistency)**.** *Consider the anomalous data stream detection problem, where* $\frac{s}{n} \to \alpha$ *as* $n \to \infty$ *and* $0 \leq \alpha < \frac{1}{2}$*. Assume* s *is known. Further assume that the test* (3.6) *applies a bounded kernel with* $0 \leq k(x, y) \leq K$ *for any* (x, y)*. Then the test* (3.6) *is universally consistent for any arbitrary pair of* p *and* q*, if*

$$m = \omega(\log n). \tag{3.16}$$

Proof. This result follows from (3.8) and the facts that $\log((n-s)s) = \Theta(\log n)$ and MMD$[p, q]$ is constant for any given p and q. □

3.2.2 Unknown s

In this subsection, we consider the case, in which the value of s is unknown. And we focus on the scenario that $\frac{s}{n} \to 0$, as $n \to \infty$. This includes two cases: (1) s is fixed and (2) $s \to \infty$ and $\frac{s}{n} \to 0$ as $n \to \infty$. Without knowledge of s, the test in (3.6) is not applicable anymore, because it depends on the value of s.

In order to build a test now, we first observe that for each k, although $\overline{Y}_k$ contains mixed samples from p and q, it is dominated by samples from p due to the above assumption on s. Thus,

for large enough m and n, $\text{MMD}_u^2[Y_k, \overline{Y}_k]$ should be close to zero if Y_k is drawn from p, and should be far away enough from zero (in fact, close to $\text{MMD}^2[p, q]$) if Y_k is drawn from q. Based on this understanding, we construct the following test:

$$\widehat{\mathcal{I}} = \{k : \text{MMD}_u^2[Y_k, \overline{Y}_k] > \delta_n\} \tag{3.17}$$

where $\delta_n \to 0$ and $\frac{s^2}{n^2\delta_n} \to 0$ as $n \to \infty$. The reason for the condition $\frac{s^2}{n^2\delta_n} \to 0$ is to guarantee that δ_n converges to 0 more slowly than $\text{MMD}_u^2[Y_k, \overline{Y}_k]$ with Y_k drawn from p so that as n goes to infinity, δ_n asymptotically falls between $\text{MMD}_u^2[Y_k, \overline{Y}_k]$ with Y_k drawn from p and $\text{MMD}_u^2[Y_k, \overline{Y}_k]$ with Y_k drawn from q. We note that the scaling behavior of s as n increases needs to be known in order to pick δ_n for the test. This is reasonable to assume because mostly in practice the scale of anomalous data sequences can be estimated based on domain knowledge.

The following theorem characterizes the condition under which the test (3.17) is consistent.

Theorem 3.2. *Consider the anomalous data stream detection problem with s anomalous sequences, where $\frac{s}{n} \to 0$, as $n \to \infty$. Assume that s is unknown in advance. Further assume that the test* (3.17) *adopts a threshold δ_n such that $\delta_n \to 0$ and $\frac{s^2}{n^2\delta_n} \to 0$, as $n \to \infty$, and the test applies a bounded kernel with $0 \le k(x, y) \le K$ for any (x, y). Then the probability of error is upper bounded as follows:*

$$\begin{aligned} P_e \le & \exp\left(\log s - \frac{m(\textit{MMD}^2[p, q] - \delta_n)^2}{16K^2(1 + \Theta(\frac{1}{n}))}\right) \\ & + \exp\left(\log(n - s) - \frac{m(\delta_n - \mathbb{E}\big[\textit{MMD}_u^2[Y_k, \overline{Y}_k]\big])^2}{16K^2(1 + \Theta(\frac{1}{n}))}\right). \end{aligned} \tag{3.18}$$

Furthermore, the test (3.17) *is consistent if*

$$m \ge 16(1 + \eta)K^2 \max\left\{\frac{\log(\max\{1, s\})}{(\textit{MMD}^2[p, q] - \delta_n)^2}, \frac{\log(n - s)}{(\delta_n - \mathbb{E}\big[\textit{MMD}_u^2[Y, \overline{Y}]\big])^2}\right\}, \tag{3.19}$$

where η is any positive constant. In the above equation, $\mathbb{E}[\textit{MMD}_u^2[Y, \overline{Y}]]$ is a constant, where Y is

a sequence generated by p and $\overline{Y}$ is a stack of $(n-1)$ sequences with s sequences generated by q and the remaining sequences generated by p.

Proof. We analyze the performance of the test (3.17). Without loss of generality, we assume that the first s sequences are the anomalous sequences. Hence,

$$\begin{aligned} P_e &= P\bigg(\big(\exists 1 \leq l \leq s : \text{MMD}_u^2[Y_l, \overline{Y}_l] \leq \delta_n\big) \text{or } \big(\exists s+1 \leq k \leq n : \text{MMD}_u^2[Y_k, \overline{Y}_k] > \delta_n\big)\bigg) \\ &\leq \sum_{l=1}^{s} P\bigg(\text{MMD}_u^2[Y_l, \overline{Y}_l] \leq \delta_n\bigg) + \sum_{k=s+1}^{n} P\bigg(\text{MMD}_u^2[Y_k, \overline{Y}_k] > \delta_n\bigg). \end{aligned} \tag{3.20}$$

Using the fact that $\frac{s}{n} \to 0$ as $n \to \infty$, and using (3.51) and (3.52) we obtain,

$$\mathbb{E}\big[\text{MMD}_u^2[Y_l, \overline{Y}_l]\big] \to \text{MMD}^2[p, q], \tag{3.21}$$

$$\mathbb{E}\big[\text{MMD}_u^2[Y_k, \overline{Y}_k]\big] \to 0, \tag{3.22}$$

as $n \to \infty$, for $1 \leq l \leq s$ and $s+1 \leq k \leq n$.

Due to (3.21), for any constant ϵ, $-\mathbb{E}\big[\text{MMD}_u^2[Y_l, \overline{Y}_l]\big] < -\text{MMD}^2[p, q] + \epsilon$ for large enough n.

For $1 \leq l \leq s$, we drive,

$$\begin{aligned} &P\bigg(\text{MMD}_u^2[Y_l, \overline{Y}_l] \leq \delta_n\bigg) \\ &\quad = P\bigg(\text{MMD}_u^2[Y_l, \overline{Y}_l] - \mathbb{E}\big[\text{MMD}_u^2[Y_l, \overline{Y}_l]\big] \leq -\mathbb{E}\big[\text{MMD}_u^2[Y_l, \overline{Y}_l]\big] + \delta_n\bigg) \\ &\quad \leq P\bigg(\text{MMD}_u^2[Y_l, \overline{Y}_l] - \mathbb{E}\big[\text{MMD}_u^2[Y_l, \overline{Y}_l] \leq -(\text{MMD}^2[p, q] - \epsilon - \delta_n)\bigg), \end{aligned} \tag{3.23}$$

for large enough n. Therefore, by applying McDiarmid's inequality, we obtain,

$$\begin{aligned} & P\bigg(\text{MMD}_u^2[Y_l, \overline{Y}_l] \leq \delta_n\bigg) \\ & \leq \exp\bigg(-\frac{2(\text{MMD}^2[p,q] - \epsilon - \delta_n)^2}{\frac{16K^2}{m}(1+\Theta(\frac{1}{n})) + \frac{16K^2}{m}(1+\Theta(\frac{1}{n}))}\bigg) \\ & = \exp\bigg(-\frac{m(\text{MMD}^2[p,q] - \epsilon - \delta_n)^2}{16K^2(1+\Theta(\frac{1}{n}))}\bigg), \end{aligned} \tag{3.24}$$

for large enough n.

For $s+1 \leq k \leq n$,

$$\begin{aligned} & P\bigg(\text{MMD}_u^2[Y_k, \overline{Y}_k] > \delta_n\bigg) \\ & = P\bigg(\text{MMD}_u^2[Y_k, \overline{Y}_k] - \mathbb{E}\big[\text{MMD}_u^2[Y_k, \overline{Y}_k]\big] > \delta_n - \mathbb{E}\big[\text{MMD}_u^2[Y_k, \overline{Y}_k]\big]\bigg). \end{aligned} \tag{3.25}$$

Using the fact that $\frac{s^2}{n^2\delta_n} \to 0$ as $n \to \infty$, we can show that

$$\frac{\mathbb{E}\big[\text{MMD}_u^2[Y_k, \overline{Y}_k]\big]}{\delta_n} \to 0,$$

as $n \to \infty$. Hence, for large enough n, $\delta_n > \mathbb{E}\big[\text{MMD}_u^2[Y_k, \overline{Y}_k]\big]$. Therefore, using McDiarmid's inequality, we have

$$\begin{aligned} & P\bigg(\text{MMD}_u^2[Y_k, \overline{Y}_k] > \delta_n\bigg) \\ & \leq \exp\bigg(-\frac{2(\delta_n - \mathbb{E}\big[\text{MMD}_u^2[Y_k, \overline{Y}_k]\big])^2}{\frac{16K^2}{m}(1+\Theta(\frac{1}{n})) + \frac{16K^2}{m}(1+\Theta(\frac{1}{n}))}\bigg) \\ & = \exp\bigg(-\frac{m(\delta_n - \mathbb{E}\big[\text{MMD}_u^2[Y_k, \overline{Y}_k]\big])^2}{16K^2(1+\Theta(\frac{1}{n}))}\bigg). \end{aligned} \tag{3.26}$$

Therefore,

$$\begin{aligned} P_e &\leq s \exp\left(- \frac{m(\mathbf{MMD}^2[p,q] - \epsilon - \delta_n)^2}{16K^2(1+\Theta(\frac{1}{n}))} \right) \\ &\quad + (n-s)\exp\left(- \frac{m(\delta_n - \mathbb{E}\big[\mathbf{MMD}_u^2[Y_k, \overline{Y}_k]\big])^2}{16K^2(1+\Theta(\frac{1}{n}))} \right) \\ &= \exp\left(\log s - \frac{m(\mathbf{MMD}^2[p,q] - \epsilon - \delta_n)^2}{16K^2(1+\Theta(\frac{1}{n}))} \right) \\ &\quad + \exp\left(\log(n-s) - \frac{m(\delta_n - \mathbb{E}\big[\mathbf{MMD}_u^2[Y_k, \overline{Y}_k]\big])^2}{16K^2(1+\Theta(\frac{1}{n}))} \right), \end{aligned} \tag{3.27}$$

for large enough n. Hence, we conclude that if

$$m \geq \frac{16(1+\eta)K^2}{(\mathbf{MMD}^2[p,q] - \delta_n)^2} \log s, \tag{3.28}$$

and

$$m \geq \frac{16(1+\eta)K^2}{(\delta_n - \mathbb{E}\big[\mathbf{MMD}_u^2[Y_k, \overline{Y}_k]\big])^2} \log(n-s), \tag{3.29}$$

where η is any positive constant, then $P_e \to 0$, as $n \to \infty$.

When $s = 0$, $P_e = \sum_{k=1}^{n} P(\mathbf{MMD}_u^2[Y_k, \overline{Y_k}] > \delta_n)$. Then applying (3.26), we have if

$$m \geq \frac{16(1+\eta)K^2}{(\delta_n - \mathbb{E}\big[\mathbf{MMD}_u^2[Y_k, \overline{Y}_k]\big])^2} \log n, \tag{3.30}$$

where η is any positive constant, then $P_e \to 0$, as $n \to \infty$. □

We note that Theorem 3.2 is also applicable to the case with $s = 0$, i.e., the null hypothesis when there is no anomalous sequence. We further note that the test (3.17) is not exponentially consistent. In fact, when there is no null hypothesis (i.e., $s > 1$), an exponentially consistent test can be built as follows. For each subsect $\mathcal{S}$ of $1, \ldots, n$, we compute $\mathbf{MMD}_u^2[Y_{\mathcal{S}}, \overline{Y}_{\mathcal{S}}]$, and the test finds the set of indices corresponding to the largest average value. However, for such a test to be consistent, m needs to scale linearly with n, which is not desirable.

Theorem 3.2 implies that m should be in the order $\omega(\log n)$ to guarantee test consistency, because $\frac{s}{n} \to 0$ and $\delta_n \to 0$ as $n \to \infty$. Compared to the case with s known (for which it is sufficient for m to scale at the order $\Theta(\log n)$), the threshold on m has order level increase due to lack of the knowledge of s. Furthermore, the above understanding on the order level condition on m also yields the following sufficient condition for the test to be universally consistent.

Proposition 3.3 (Universal Consistency)**.** *Consider the anomalous data stream detection problem, where $\frac{s}{n} \to 0$, as $n \to \infty$. We assume that s is unknown in advance. Further assume that the test* (3.17) *adopts a threshold δ_n such that $\delta_n \to 0$ and $\frac{s^2}{n^2\delta_n} \to 0$, as $n \to \infty$, and the test applies a bounded kernel with $0 \leq k(x, y) \leq K$ for any (x, y). Then the test* (3.17) *is universally consistent for any arbitrary pair of p and q, if*

$$m = \omega(\log n). \tag{3.31}$$

Comparison between Proposition 3.3 with Proposition 3.2 implies that the knowledge of s does not affect the order level sample complexity to guarantee a test to be universally consistent.

3.2.3 Example with Sparse Anomalous Samples

We study the example with the anomalous distribution $q = (1 - \epsilon_n)p + \epsilon_n\tilde{q}$ as we introduce in Section 3.1. The following result characterizes the impact of sparsity level ϵ_n on the scaling behavior of m to guarantee consistent detection.

Corollary 3.1. *Consider the model with the typical distribution p and the anomalous distribution $q = (1 - \epsilon_n)p + \epsilon_n\tilde{q}$, where $0 < \epsilon_n \leq 1$. If s is known, then the test* (3.6) *is consistent if*

$$m \geq \frac{16K^2(1 + \eta)}{(1 - 2\alpha)^2\epsilon_n^4 \mathit{MMD}^4[p, \tilde{q}]} \log(s(n - s)), \tag{3.32}$$

where η is any positive constant.

If s is unknown, then the test (3.17) *is consistent if*

$$m \geq 16(1+\eta)K^2 \max\Big\{ \frac{\log(\max\{1,s\})}{(\epsilon_n^2 \mathit{MMD}^2[p,\tilde{q}] - \delta_n)^2} \, , \frac{\log(n-s)}{(\delta_n - \mathbb{E}\big[\mathit{MMD}_u^2[Y,\overline{Y}]\big])^2} \Big\}, \tag{3.33}$$

where η is any positive constant, $\frac{s^2\epsilon_n^2}{n^2\delta_n} \to 0$ and $\frac{\delta_n}{\epsilon_n^2} \to 0$ as $n \to \infty$, Y is a sequence generated by p, and $\overline{Y}$ is a stack of $(n-1)$ sequences with s sequences generated by $\tilde{q}$ and the remaining sequences generated by p.

Proof. The proof follows from Theorems 3.1 and 3.2 by substituting:

$$\begin{aligned}
&\mathrm{MMD}^2[p,q] \\
&= \mathbb{E}_{x,x'}[k(x,x')] - 2\mathbb{E}_{x,y}[k(x,y)] + \mathbb{E}_{y,y'}[k(y,y')] \\
&= \mathbb{E}_{x,x'}[k(x,x')] - 2(1-\epsilon_n)\mathbb{E}_{x,x'}[k(x,x')] - 2\epsilon_n\mathbb{E}_{x,\tilde{y}}[k(x,\tilde{y})] + (1-\epsilon_n)^2\mathbb{E}_{x,x'}[k(x,x')] \\
&\quad + 2\epsilon_n(1-\epsilon_n)\mathbb{E}_{x,\tilde{y}}[k(x,\tilde{y})] + \epsilon_n^2\mathbb{E}_{\tilde{y},\tilde{y}'}[k(\tilde{y},\tilde{y}')] \\
&= \epsilon_n^2 \mathrm{MMD}^2[p,\tilde{q}],
\end{aligned} \tag{3.34}$$

where x and x' are independent but have the same distribution p, y and y' are independent but have the same distribution q, and $\tilde{y}$ and $\tilde{y}'$ are independent but have the same distribution $\tilde{q}$. □

Corollary 3.1 implies that if ϵ_n is a constant, then the scaling behavior of m needed for consistent detection does not change. However, if $\epsilon_n \to 0$ as $n \to \infty$, i.e., anomalous sequences contain more sparse anomalous samples, then m needs to scale faster with n in order to guarantee consistent detection. This is reasonable because the sample size m should have a higher order to cancel out the impact of the increasingly sparse anomalous samples in each anomalous sequence. Corollary 3.1 explicitly captures such tradeoff between the sample size m and the sparsity level ϵ_n of anomalous samples in addition to n and s.

3.3 Necessary Condition and Optimality

In Section 3.2, we characterize sufficient conditions on the sample size m under which the MMD-based test is guaranteed to be consistent for any distribution pair p and q. In this section, we characterize conditions under which no test is universally consistent for arbitrary p and q. We first study the case with $s = 1$ for which we develop our key idea of the proof. We then generalize our study to the case with $s \geq 1$.

Proposition 3.4. *Consider the anomalous data stream detection problem with one anomalous sequence. If the sample size m satisfies*

$$m = O(\log n), \tag{3.35}$$

then there exists no test that is universally consistent for any arbitrary distribution pair p and q.

Proof. The idea of the proof is to show that for a certain distribution pair p and q, even the optimal parametric test (with known p and q) is not consistent under the condition given in the theorem. This thus implies that under the same condition, no nonparametric test is universally consistent for arbitrary p and q.

We first introduce an interesting property of Gaussian distribution, which is useful for bounding the probability of error for our problem.

Lemma 3.1. *[74] For the standard Gaussian distribution with mean zero and variance one, there exists positive constants c_1 and c_2 such that the cumulative distribution function (CDF) $\Phi(x)$ of the standard Gaussian distribution satisfies the following inequalities:*

$$\frac{c_1}{\log n} < \sup_{-\infty<x<\infty} |\Phi^n(a_n x + b_n) - G(x)| < \frac{c_2}{\log n} \tag{3.36}$$

for all positive integer n, where $G(x) = e^{e^{-x}}$ (i.e., the CDF of the Gumbel distribution), $a_n b_n = 1$.

In particular, b_n can be approximated as

$$b_n = \sqrt{2\log n} - \frac{\frac{1}{2}\log(4\pi\log n)}{\sqrt{2\log n}} + O\left(\frac{1}{\log n}\right). \tag{3.37}$$

Our main idea of the proof is to show that under a certain distribution pair p and q, even the optimal parametric test is not consistent under the condition given in the theorem. This thus implies that under the same condition, no nonparametric test is universally consistent for arbitrary p and q. Towards this end, we consider the case, in which p and q are Gaussian with the same variance but mean shift, i.e., $p = \mathcal{N}(0,1)$ and $q = \mathcal{N}(1,1)$. The optimal test with known p and q is the following maximum likelihood (ML) test.

$$\hat{i} = \arg\max_{1\le i\le n}\{P_i(Y^{nm})\}, \tag{3.38}$$

where $P_i(Y^{nm})$ denotes the probability of Y^{nm} if the i-th sequence is anomalous. The probability of error under the ML test is given by:

$$P_e = \frac{1}{n}\sum_{i=1}^{n}\mathcal{P}_i\Big(P_i(Y^{nm}) \le \max_{k\neq i} P_k(Y^{nm})\Big), \tag{3.39}$$

where $\mathcal{P}_i$ denotes the probability evaluated when i-th sequence is anomalous. By the symmetry of the problem,

$$\mathcal{P}_i\Big(P_i(Y^{nm}) \le \max_{k\neq i} P_k(Y^{nm})\Big) = \mathcal{P}_j\Big(P_j(Y^{nm}) \le \max_{k\neq j} P_k(Y^{nm})\Big), \tag{3.40}$$

for any $1 \le i, j \le n$. Hence, we have

$$\begin{aligned} P_e &= \mathcal{P}_1\Big(P_1(Y^{nm}) \le \max_{k\neq 1} P_k(Y^{nm})\Big) \\ &= \mathcal{P}_1\Big(\frac{1}{\sqrt{m}}\sum_{i=1}^{m} Y_{1i} \le \max_{2\le k\le n}\frac{1}{\sqrt{m}}\sum_{i=1}^{m} Y_{ki}\Big). \end{aligned} \tag{3.41}$$

For convenience, we define $B_1 := \frac{1}{\sqrt{m}} \sum_{i=1}^{m} Y_{1i}$, and $B_k := \frac{1}{\sqrt{m}} \sum_{i=1}^{m} Y_{ki}$, for $2 \leq k \leq n$. Hence, $B_1 \sim \mathcal{N}(\sqrt{m}, 1)$, and $B_k \sim \mathcal{N}(0, 1)$, and they are independent from each other. With the above definitions, the probability of error can be written as

$$\begin{aligned} P_e &= \mathcal{P}\Big(B_1 \leq \max_{2 \leq k \leq n} B_k\Big) \\ &= 1 - \mathcal{P}\Big(\max_{2 \leq k \leq n} B_k < B_1\Big) \\ &= 1 - \mathbb{E}_B\Big\{\Phi^{n-1}(B_1)\Big\} \end{aligned} \tag{3.42}$$

where Φ is the CDF of B_k.

By Lemma 3.1, there exists a constant c independent of n, such that for all positive integer n, and for all real values x,

$$G\Big(\frac{x - b_n}{a_n}\Big) - \frac{c}{\log n} \leq \Phi^n(x) \leq G\Big(\frac{x - b_n}{a_n}\Big) + \frac{c}{\log n}, \tag{3.43}$$

where a_n, b_n are optimal normalizing constants, and $G(x) = e^{-e^{-x}}$ is the CDF of the Gumbel distribution.

Hence,

$$\begin{aligned} P_e &= 1 - \mathbb{E}_B \Phi^{n-1}(B_1) \\ &\geq 1 - \frac{c}{\log(n-1)} - \mathbb{E}_B\Big\{G\Big(\frac{B_1 - b_{n-1}}{a_{n-1}}\Big)\Big\} \\ &= 1 - \frac{c}{\log(n-1)} - \mathbb{E}_T\Big\{G(T)\Big\}, \end{aligned} \tag{3.44}$$

where $T = \frac{B_1 - b_{n-1}}{a_{n-1}}$, and $T \sim \mathcal{N}(\frac{\sqrt{m} - b_{n-1}}{a_{n-1}}, \frac{1}{a_{n-1}^2})$. The second term in (3.44) can be further

bounded as

$$\begin{aligned}
&\mathbb{E}_T\Big\{G(T)\Big\} \\
&\quad = \int_{-\infty}^{0} e^{-e^{-t}} p(t)dt + \int_{0}^{+\infty} e^{-e^{-t}} p(t)dt \\
&\quad \leq e^{-1} + P(T \geq 0)
\end{aligned} \tag{3.45}$$

where

$$P(T \geq 0) = Q\left(\frac{0 - \frac{\sqrt{m} - b_{n-1}}{a_{n-1}}}{\frac{1}{a_{n-1}}}\right) = Q(b_{n-1} - \sqrt{m}). \tag{3.46}$$

In the above equations, $Q(\cdot)$ denotes the tail probability of the standard Gaussian distribution. If $m \leq 2(1-\eta)\log n$, where η is any positive constant, $b_{n-1} - \sqrt{m} \to \infty$, $Q(b_{n-1} - \sqrt{m}) \to 0$. Hence,

$$\lim_{n\to\infty} \mathbb{E}_T[G(T)] \leq e^{-1}. \tag{3.47}$$

Thus, with $\frac{c}{\log n} \to 0$

$$\lim_{n\to\infty} P_e \geq 1 - e^{-1} \approx 0.6321 > 0 \tag{3.48}$$

as $n \to \infty$. Therefore, if $m = O(\log n)$, where η is any positive constant, there exists no consistent test for any arbitrary distributions p and q. □

We now generalize our result to the case with $s \geq 1$, and provide the following proposition.

Proposition 3.5. *Consider the anomalous data stream detection problem with s anomalous sequences. If the sample size m satisfies*

$$m = O\left(\frac{\log \frac{n}{s}}{s}\right), \tag{3.49}$$

then there exists no test that is universally consistent for arbitrary distribution pair p and q.

Proof. It can be shown that the probability of error of this problem is lower bounded by a special scenario, in which anomalous sequences can only be a group of s sequences with consecutive indices, i.e., one of the following possibilities: the $(is+1)$-th to $(i+1)s$-th sequences, for $i = 0, \ldots, \lfloor \frac{n}{s} \rfloor - 1$. Hence, there are $\lfloor \frac{n}{s} \rfloor$ candidates. Such a specific scenario can be viewed as the problem of detecting one anomalous sequence with length ms out of $\lfloor \frac{n}{s} \rfloor$ sequences. The proposition then follows from arguments similar to those used to prove Proposition 3.4. □

The sufficient and necessary conditions on sample complexity that we derive so far establish the following performance optimality for the MMD-based test.

Theorem 3.3 (Optimality)**.** *Consider the nonparametric anomalous data stream detection problem with $s \geq 1$. For s being known and unknown, the MMD-based test* (3.6) *(under the conditions in Propositions 3.2) and the test* (3.17) *(under the conditions in Proposition 3.3) are respectively order level optimal in sample complexity required to guarantee universal consistency for arbitrary p and q.*

Proof. The proof follows by comparing Propositions 3.2 and 3.3 with Proposition 3.5 and observing the fact that $m = O(\log n)$ in Proposition 3.5 for finite s. □

3.4 Numerical Results

In this section, we provide numerical results to demonstrate our theoretical assertions, and compare our MMD-based tests with a number of other tests. We also apply our MMD based test to a real data set.

We first demonstrate our theorem on sample complexity. We note that although the following experiment is performed for chosen distributions p and q, our tests are nonparametric and do not exploit the information about p and q. We choose the distribution p to be Gaussian with mean zero and variance one, i.e., $\mathcal{N}(0,1)$, and choose the anomalous distribution q to be Laplace distribution

with mean one and variance one. We use the Gaussian kernel $k(x, x') = \exp(-\frac{|x-x'|^2}{2\sigma^2})$ with $\sigma = 1$. We set $s = 1$. We run the test for cases with $n = 40$ and 100, respectively. In Figure 3.2, we plot how the probability of error changes with m. For illustrational convenience, we normalize m by $\log n$, i.e., the horizontal axis represents $\frac{m}{\log n}$. It is clear from the figure that when $\frac{m}{\log n}$ is above a certain threshold, the probability of error converges to zero, which is consistent with our theoretical results. Furthermore, for different values of n, the two curves drop to zero almost at the same threshold. This observation confirms Proposition 3.1, which states that the threshold on $\frac{m}{\log n}$ depends only on the bound K of the kernel and MMD of the two distributions. Both quantities are constant for the two values of n.

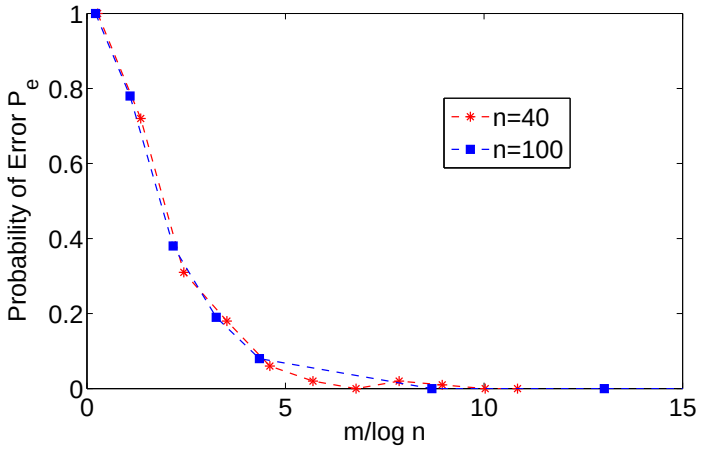

Fig. 3.2: The performance of the MMD-based test.

We next compare the MMD-based test with the divergence-based generalized likelihood test developed in [14]. Since the test in [14] is applicable only when the distributions p and q are discrete and have finite alphabets, we set the distributions p and q to be binary with p having probability 0.3 to take “0" (and probability 0.7 to take “1"), and q having probability 0.7 to take “0" (and probability 0.3 to take “1"). We let $s = 1$ and assume that s is known. We let $n = 50$. In Figure 3.3, we plot the probability of error as a function of the sample size m. It can be seen that the MMD-based test outperforms the divergence-based generalized likelihood test when the sample size m is small. We note that it has been shown in [14] that the generalized likelihood test has optimal convergence rate in the limiting case when n is infinite. Our numerical comparison,

on the other hand, demonstrates that the MMD-based test performs as well as or even better than the generalized likelihood test for moderate n.

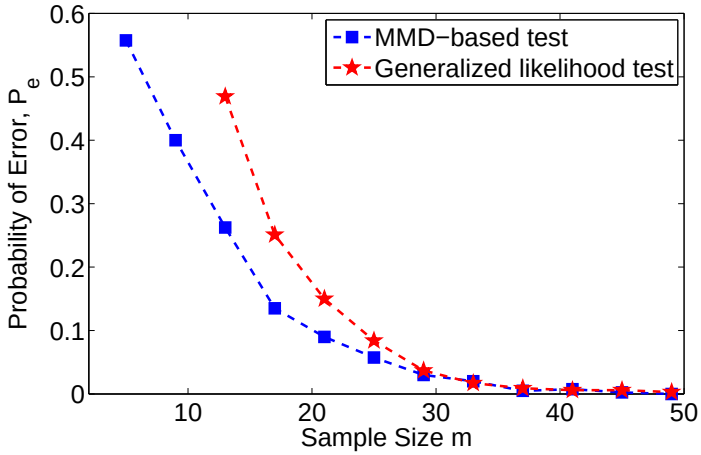

Fig. 3.3: Comparison of the MMD-based test with divergence-based generalized likelihood test.

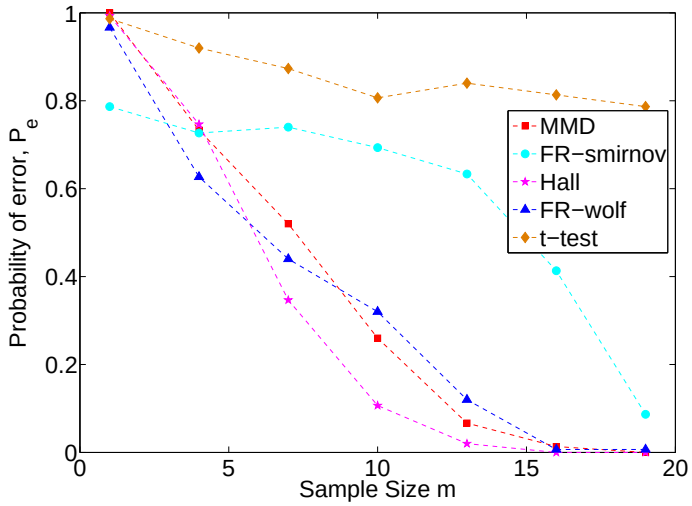

Fig. 3.4: Comparison of the MMD-based test with four other tests on a real data set.

We finally compare the performance of the MMD-based test with a few other competitive tests on a real data set. We choose the collection of daily maximum temperature of Syracuse (New York, USA) in July from 1993 to 2012 as the typical data sequences, and the collection of daily maximum temperature of Makapulapai (Hawaii, USA) in May from 1993 to 2012 as anomalous sequences. Here, each data sequence contains daily maximum temperatures of a certain day across

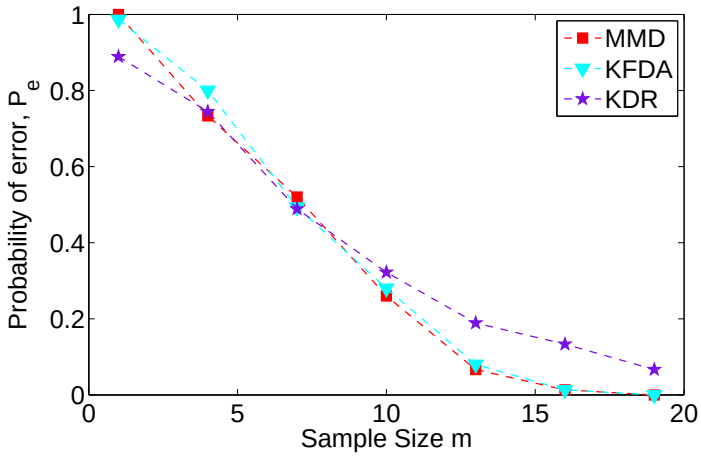

Fig. 3.5: Comparison of the MMD-based test with two other kernel-based tests on a real data set.

twenty years from 1993 to 2012. In our experiment, the data set contains 32 sequences in total, including one temperature sequence of Hawaii and 31 sequences of Syracuse. The probability of error is averaged over all cases with each using one sequence of Hawaii as the anomalous sequence. Although it seems easy to detect the sequence of Hawaii out of the sequences of Syracuse, the temperatures we compare for the two places are in May for Hawaii and July for Syracuse, during which the two places have approximately the same mean in temperature. In this way, it may not be easy to detect the anomalous sequence (in fact, some tests do not perform well as shown in Figure 3.4).

We first compare the performance of the MMD-based test with t-test, FR-Wolf test, FR-Smirnov test, and Hall test on the above data set. For the MMD-based test, we use the Gaussian kernel with $\sigma = 1$. In Figure 3.4, we plot the probability of error as a function of the length of sequence m for all tests. It can be seen that the MMD-based test, Hall test, and FR-wolf test have the best performances, and all of the three tests are consistent with the probability of error converging to zero as m goes to infinity. Furthermore, comparing to Hall and FR-wolf tests, the MMD-based test has the lowest computational complexity.

We further compare the performance of MMD-based test with the kernel-based tests KFDA and KDR for the same data set. For all three tests, we use Gaussian kernel with $\sigma = 1$. In Figure 3.5,

we plot the probability of error as a function of the length of sequence for all tests. It can be seen that all tests are consistent with the probability of error converging to zero as m increases, and the MMD-based test has the best performance among the three tests.

3.5 Proof of Proposition 3.1: Performance Guarantee

In order to analyze the probability of error for the test (3.3), without loss of generality, we assume that the first sequence is the anomalous sequence generated by the anomalous distribution q. Hence,

$$\begin{aligned} P_e = P(k^* \neq 1) &= P\bigg(\exists k \neq 1 : \text{MMD}_u^2[Y_k, \overline{Y}_k] > \text{MMD}_u^2[Y_1, \overline{Y}_1]\bigg) \\ &\leq \sum_{k=2}^{n} P\bigg(\text{MMD}_u^2[Y_k, \overline{Y}_k] > \text{MMD}_u^2[Y_1, \overline{Y}_1]\bigg). \end{aligned} \tag{3.50}$$

For notational convenience, we stack $Y_1, \ldots, Y_n$ into a nm dimensional row vector $Y = \{y_i, 1 \leq i \leq nm\}$, where $Y_k = \{y_{(k-1)m+1}, \ldots, y_{km}\}$. And we define $n' = (n-1)m$. We then have,

$$\text{MMD}_u^2[Y_1, \overline{Y}_1] = \frac{1}{m(m-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^{m,m} k(y_i, y_j) + \frac{1}{n'(n'-1)} \sum_{\substack{i,j=m+1 \\ i \neq j}}^{nm} k(y_i, y_j) - \frac{2}{mn'} \sum_{\substack{i=1 \\ j=m+1}}^{m,nm} k(y_i, y_j). \tag{3.51}$$

For $2 \leq k \leq n$, we have,

$$\begin{aligned} \text{MMD}_u^2[Y_k, \overline{Y}_k] = \frac{1}{m(m-1)} &\sum_{\substack{i,j=(k-1)m+1 \\ i \neq j}}^{km,km} k(y_i, y_j) + \frac{1}{n'(n'-1)} \bigg(\sum_{\substack{i,j=1 \\ i \neq j}}^{m,m} k(y_i, y_j) + 2 \sum_{\substack{i=1 \\ j=m+1}}^{m,(k-1)m} k(y_i, y_j) \\ &+ 2 \sum_{\substack{i=1 \\ j=km+1}}^{m,nm} k(y_i, y_j) + \sum_{\substack{i,j=m+1 \\ i \neq j}}^{(k-1)m,(k-1)m} k(y_i, y_j) + \sum_{\substack{i,j=km+1 \\ i \neq j}}^{nm,nm} k(y_i, y_j) + 2 \sum_{\substack{i=m+1 \\ j=km+1}}^{(k-1)m,nm} k(y_i, y_j) \bigg) \\ &- \frac{2}{mn'} \bigg(\sum_{\substack{i=1 \\ j=(k-1)m+1}}^{m,km} k(y_i, y_j) + \sum_{\substack{i=m+1 \\ j=(k-1)m+1}}^{(k-1)m,km} k(y_i, y_j) + \sum_{\substack{i=(k-1)m+1 \\ j=km+1}}^{km,nm} k(y_i, y_j) \bigg). \end{aligned} \tag{3.52}$$

We define $\Delta_k = \text{MMD}_u^2[Y_k, \overline{Y}_k] - \text{MMD}_u^2[Y_1, \overline{Y}_1]$. It can be shown that,

$$\mathbb{E}[\text{MMD}_u^2[Y_1, \overline{Y}_1]] = \text{MMD}^2[p, q],$$

and

$$\begin{aligned}\mathbb{E}[\text{MMD}_u^2[Y_k, \overline{Y}_k]] = \mathbb{E}_{x,x'}k(x, x') + \frac{1}{(n-1)m((n-1)m-1)}\Big(m(m-1)\mathbb{E}_{y,y'}k(y, y') \\ + 2m^2(n-2)\mathbb{E}_{x,y}k(x, y) + ((n-2)m-1)(n-2)m\mathbb{E}_{x,x'}k(x, x')\Big) \\ - \frac{2}{(n-1)m^2}\Big(m^2\mathbb{E}_{x,y}k(x, y) + (n-2)m^2\mathbb{E}_{x,x'}k(x, x')\Big) \\ \to 0, \text{ as } n \to \infty, \end{aligned} \tag{3.53}$$

where x and x' are independent but have the same distribution p, y and y' are independent but have the same distribution q. Hence, there exists a constant ξ that satisfies

$$\mathbb{E}[\text{MMD}_u^2[Y_k, \overline{Y}_k]] < \xi < \text{MMD}^2[p, q], \tag{3.54}$$

for large enough n. Here, ξ can be arbitrarily close to zero as $n \to \infty$.

We next divide the entries in $\{y_1, \ldots, y_{nm}\}$ into three groups: $Y_1 = \{y_1, \ldots, y_m\}$, $Y_k = \{y_{(k-1)m+1} \ldots, y_{km}\}$, and $\widehat{Y}_k$ that contains the remaining entries. We define Y_{-a} as Y with the a-th component y_a being removed.

For $1 \leq a \leq m$, y_a affects Δ_k through the following terms

$$\begin{aligned}&\frac{1}{n'(n'-1)}\Big(2\sum_{\substack{j=1\\j\neq a}}^{m} k(y_a, y_j) + 2\sum_{j=m+1}^{(k-1)m} k(y_a, y_j) + 2\sum_{j=km+1}^{nm} k(y_a, y_j)\Big) \\ &- \frac{2}{mn'}\sum_{j=(k-1)m+1}^{km} k(y_a, y_j) - \frac{2}{m(m-1)}\sum_{\substack{j=1\\k\neq a}}^{m} k(y_a, y_j) + \frac{2}{mn'}\sum_{j=m+1}^{nm} k(y_a, y_j). \end{aligned} \tag{3.55}$$

Hence, for $1 \le a \le m$, we have

$$|\Delta_k(Y_{-a}, y_a) - \Delta_k(Y_{-a}, y'_a)| \le \frac{4K}{m}\left(1 + \Theta\left(\frac{1}{n}\right)\right). \tag{3.56}$$

For $(k-1)m + 1 \le a \le km$, y_a affects Δ_k through

$$\frac{2}{m(m-1)} \sum_{\substack{j=(k-1)m+1 \\ j \ne a}}^{km} k(y_a, y_j) - \frac{2}{mn'}\left(\sum_{i=1}^{m} k(y_i, y_a) + \sum_{i=m+1}^{(k-1)m} k(y_i, y_a) + \sum_{j=km+1}^{nm} k(y_a, y_j)\right)$$
$$- \frac{2}{n'(n'-1)} \sum_{\substack{j=m+1 \\ j \ne a}}^{nm} k(y_a, y_j) + \frac{2}{mn'} \sum_{i=1}^{m} k(y_a, y_i). \tag{3.57}$$

Hence, for $(k-1)m + 1 \le a \le km$, we have

$$|\Delta_k(Y_{-a}, y_a) - \Delta_k(Y_{-a}, y'_a)| \le \frac{4K}{m}\left(1 + \Theta\left(\frac{1}{n}\right)\right). \tag{3.58}$$

For $m + 1 \le a \le (k-1)m$ and $km + 1 \le a \le nm$, y_a affects Δ_k through

$$\frac{2}{n'(n'-1)}\left(\sum_{i=1}^{m} k(y_i, y_a) + \sum_{\substack{i=m+1 \\ i \ne a}}^{(k-1)m} k(y_i, y_a) + \sum_{j=km+1}^{nm} k(y_a, y_j)\right) - \frac{2}{mn'} \sum_{j=(k-1)m+1}^{km} k(y_a, y_j)$$
$$- \frac{2}{n'(n'-1)} \sum_{\substack{j=m+1 \\ j \ne a}}^{nm} k(y_a, y_j) + \frac{2}{mn'} \sum_{i=(k-1)m+1}^{km} k(y_i, y_a). \tag{3.59}$$

Hence, for $m + 1 \le a \le (k-1)m$ or $km + 1 \le a \le nm$, we have

$$|\Delta_k(Y_{-a}, y_a) - \Delta_k(Y_{-a}, y'_a)| \le \frac{1}{m}\Theta\left(\frac{1}{n}\right). \tag{3.60}$$

We further derive the following probability,

$$
\begin{aligned}
&P\bigg(\text{MMD}_u^2[Y_k, \overline{Y}_k] > \text{MMD}_u^2[Y_1, \overline{Y}_1]\bigg) \\
&= P\bigg(\text{MMD}_u^2[Y_k, \overline{Y}_k] - \text{MMD}_u^2[Y_1, \overline{Y}_1] + \text{MMD}^2[p,q] > \text{MMD}^2[p,q]\bigg) \\
&\overset{(a)}{\leq} P\bigg(\text{MMD}_u^2[Y_k, \overline{Y}_k] - \text{MMD}_u^2[Y_1, \overline{Y}_1] + \text{MMD}^2[p,q] - \mathbb{E}[\text{MMD}_u^2[Y_k, \overline{Y}_k]] > \text{MMD}^2[p,q] - \xi\bigg),
\end{aligned}
\tag{3.61}
$$

where (a) follows from (3.54).

Combining (3.56), (3.58), (3.60), and applying McDiarmid's inequality, we have,

$$
\begin{aligned}
&P\bigg(\text{MMD}_u^2[Y_k, \overline{Y}_k] > \text{MMD}_u^2[Y_1, \overline{Y}_1]\bigg) \\
&\quad \leq \exp\bigg(-\frac{2(\text{MMD}^2[p,q] - \xi)^2}{2m\frac{16K^2}{m^2}(1+\Theta(\frac{1}{n})) + \frac{1}{m}\Theta(\frac{1}{n})}\bigg) \\
&\quad = \exp\bigg(-\frac{m(\text{MMD}^2[p,q] - \xi)^2}{16K^2(1+\Theta(\frac{1}{n}))}\bigg)
\end{aligned}
\tag{3.62}
$$

Hence,

$$
P_e \leq \exp\bigg(\log n - \frac{m(\text{MMD}^2[p,q] - \xi)^2}{16K^2(1+\Theta(\frac{1}{n}))}\bigg). \tag{3.63}
$$

Since ξ can be picked arbitrarily close to zero, we conclude that if

$$
m \geq \frac{16K^2(1+\eta)}{\text{MMD}^4[p,q]} \log n, \tag{3.64}
$$

where η is any positive constant, then $P_e \to 0$ as $n \to \infty$. It is also clear that if the above condition is satisfied, P_e converges to zero exponentially fast with respect to m. This completes the proof.

CHAPTER 4
DEGRADED BROADCAST CHANNEL WITH LAYERED DECODING AND LAYERED SECRECY

In this chapter, we study the model of the degraded broadcast channel with layered decoding and layered secrecy. In Section 4.1, we introduce the model of the degraded broadcast channel with layered decoding and layered secrecy constraints. In Section 4.2, we present our results on the secrecy capacity region. In Section 4.3, we present our results on the multi-secret sharing problem.

4.1 Channel Model

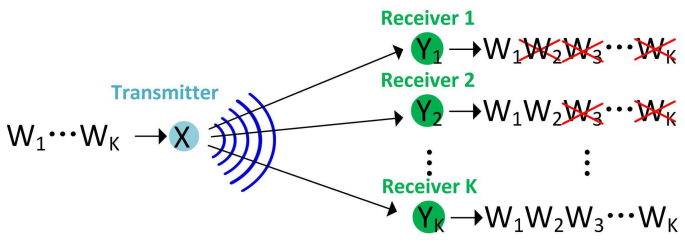

Fig. 4.1: Degraded broadcast channel with layered decoding and layered secrecy.

In this section, we introduce the model of the degraded broadcast channel with layered decoding and secrecy constraints (see Fig. 4.1), in which a transmitter transmits to K receivers. The channel transition probability function is given by $P_{Y_1 \cdots Y_K|X}$, in which $X \in \mathcal{X}$ is the channel input and $Y_k \in \mathcal{Y}_k$ is the channel output of receiver k for $k = 1, \ldots, K$. It is assumed that the receivers have degraded outputs, i.e., $Y_1, \cdots, Y_K$ satisfy the following Markov chain condition:

$$X \to Y_K \to Y_{K-1} \to \ldots \to Y_2 \to Y_1. \tag{4.1}$$

Hence, the quality of channels gradually degrades from receiver K to receiver 1. The transmitter has K messages $W_1, \cdots, W_K$ intended for the K receivers. The system is required to satisfy the following layered decoding and secrecy constraints. For $k = 1, \ldots, K$, receiver k needs to decode the messages $W_1, \cdots, W_k$, and to be kept ignorant of messages $W_{k+1}, \cdots, W_K$ (see Fig. 4.1 for an illustration).

A $(2^{nR_1}, \cdots, 2^{nR_K}, n)$ code for the channel consists of

- K message sets: $W_k \in \mathcal{W}_k = \{1, \cdots, 2^{nR_k}\}$ for $k = 1, \cdots, K$, which are independent from each other and each message is uniformly distributed over the corresponding message set;
- An (possibly stochastic) encoder f^n: $\mathcal{W}_1 \times \cdots \times \mathcal{W}_K \to \mathcal{X}^n$;
- K decoders $g_k^n : \mathcal{Y}_k^n \to (\mathcal{W}_1, \cdots, \mathcal{W}_k)$ for $k = 1, \cdots, K$.

Hence, a secrecy rate tuple $(R_1, \cdots, R_K)$ is said to be *achievable*, if there exists a sequence of $(2^{nR_1}, \cdots, 2^{nR_K}, n)$ codes such that both the average error probability

$$P_e^n = \Pr\left(\cup_{k=1}^K \{(W_1, \cdots, W_k) \neq g_k^n(Y_k^n)\}\right) \tag{4.2}$$

and the leakage rate at each receiver k for $k = 1, \ldots, K$

$$\frac{1}{n} I(W_{k+1}, \cdots, W_K; Y_k^n | W_1, \cdots, W_k) \tag{4.3}$$

approach zero as n goes to infinity.

Here, condition (4.2) implies that each receiver k is able to decode messages $W_1, \ldots, W_k$, while (4.3) implies that receiver k is kept ignorant of messages $W_{k+1}, \ldots, W_K$. The secrecy capacity region is defined as the set of all achievable rate tuples.

The degraded Gaussian MIMO broadcast channel with layered decoding and secrecy constraints is further studied. In this model, the received signal at receiver k for one channel use is given by

$$\mathbf{Y}_k = \mathbf{X} + \mathbf{N}_k, \text{ for } k = 1, \ldots, K, \tag{4.4}$$

where the channel input $\mathbf{X}$, the channel output $\mathbf{Y}_k$ and the noise $\mathbf{N}_k$ are r-dimensional vectors. Furthermore, the noise variables $\mathbf{N}_k$ are zero-mean Gaussian random vectors with covariance matrices $\mathbf{\Sigma}_k$ for $k = 1, \ldots, K$ that satisfy the following order:

$$\mathbf{0} \prec \mathbf{\Sigma}_K \preceq \mathbf{\Sigma}_{K-1} \preceq \cdots \preceq \mathbf{\Sigma}_1. \tag{4.5}$$

The channel input $\mathbf{X}$ is subject to a covariance constraint

$$E[\mathbf{X}\mathbf{X}^\top] \preceq \mathbf{S} \tag{4.6}$$

where $\mathbf{S} \succ \mathbf{0}$. Since the secrecy capacity region does not depend on the correlation across the channel outputs, we can adjust the correlation between the noise vectors such that the channel inputs and channel outputs satisfy the following Markov chain:

$$\mathbf{X} \to \mathbf{Y}_K \to \mathbf{Y}_{K-1} \to \ldots \to \mathbf{Y}_2 \to \mathbf{Y}_1. \tag{4.7}$$

Hence, the quality of channels gradually degrades from receiver K to receiver 1.

4.2 Characterization of Secrecy Capacity Region

For the discrete memoryless degraded broadcast channel with layered decoding and secrecy constraints, we characterize the secrecy region in the following theorem.

Theorem 4.1. *The secrecy capacity region of the degraded broadcast channel with layered decoding and secrecy constraints as described in Section 4.1 contains rate tuples* $(R_1, \cdots, R_K)$ *satisfying*

$$\begin{aligned} R_1 &\leq I(U_1; Y_1), \\ R_k &\leq I(U_k; Y_k|U_{k-1}) - I(U_k; Y_{k-1}|U_{k-1}), \\ &\qquad\qquad \textit{for } k = 2, \ldots, K-1, \\ R_K &\leq I(X; Y_K|U_{K-1}) - I(X; Y_{K-1}|U_{K-1}), \end{aligned} \tag{4.8}$$

for some $P_{U_1 U_2 \ldots U_{K-1} X}$ *such that the following Markov chain holds*

$$U_1 \to U_2 \to \ldots \to U_{K-1} \to X \to Y_K \to \ldots \to Y_1. \tag{4.9}$$

Proof. The proof of achievability and converse are provided in Section 4.4 and Section 4.6, respectively. □

We here briefly introduce the idea of the achievable scheme, which is based on the stochastic encoding (i.e., random binning) and superposition coding. For each message, we design one layer of codebook. This codebook contains codewords that are divided into a number of bins, where the bin number contains the information of the corresponding message. The receivers that are required to decode the message can tell which bin the codeword is in with a small probability of error, while other receivers (i.e., those with worse channel quality) are kept ignorant of this message. These layers of codebooks are superposed together via superposition coding. The major challenge of the achievability proof arises in the analysis of leakage rates, which is much more involved than the cases with two secure messages studied in [39, 41]. In our proof, we develop novel generalization

of the analysis provided in [48] for the case with one secure message to multiple secure messages. The details can be referred to Section 4.4.

Furthermore, we characterize the secrecy capacity region for the degraded Gaussian MIMO channel with layered decoding and secrecy constraints in the following theorem.

Theorem 4.2. *The secrecy capacity region of the degraded Gaussian MIMO broadcast channel with layered decoding and secrecy constraints as described in Section 4.1 contains all rate tuples* $(R_1, \ldots, R_K)$ *that satisfy the following inequalities:*

$$\begin{aligned} R_1 &\leq \frac{1}{2}\log\frac{|\mathbf{\Sigma}_1+\mathbf{S}|}{|\mathbf{\Sigma}_1+\mathbf{S}_1|} \\ R_k &\leq \frac{1}{2}\log\frac{|\mathbf{\Sigma}_k+\mathbf{S}_{k-1}|}{|\mathbf{\Sigma}_k+\mathbf{S}_k|} - \frac{1}{2}\log\frac{|\mathbf{\Sigma}_{k-1}+\mathbf{S}_{k-1}|}{|\mathbf{\Sigma}_{k-1}+\mathbf{S}_k|}, \\ &\qquad\qquad \textit{for } 2 \leq k \leq K-1, \\ R_K &\leq \frac{1}{2}\log\frac{|\mathbf{\Sigma}_K+\mathbf{S}_{K-1}|}{|\mathbf{\Sigma}_K|} - \frac{1}{2}\log\frac{|\mathbf{\Sigma}_{K-1}+\mathbf{S}_{K-1}|}{|\mathbf{\Sigma}_{K-1}|}, \end{aligned} \tag{4.10}$$

for some $\mathbf{0} \preceq \mathbf{S}_{K-1} \preceq \mathbf{S}_{K-2} \preceq \ldots \preceq \mathbf{S}_2 \preceq \mathbf{S}_1 \preceq \mathbf{S}$.

Proof of Achievability. The achievability of region (4.10) follows by choosing the auxiliary random variables $\mathbf{U}_1, \ldots, \mathbf{U}_{K-1}, \mathbf{X}$ to be jointly Gaussian distributed and satisfy the following Markov chain condition:

$$\mathbf{U}_1 \rightarrow \mathbf{U}_2 \rightarrow \ldots \rightarrow \mathbf{U}_{K-1} \rightarrow \mathbf{X}, \tag{4.11}$$

where the covariance of $\mathbf{U}_k$ is set to be $\mathbf{S} - \mathbf{S}_k$ for $k = 1, \ldots, K-1$, and the covariance of $\mathbf{X}$ is set to be $\mathbf{S}$. □

Proof of Converse. See Section 4.7. □

We note that due to the layered secrecy constraints, the major challenge in the converse proof for the secrecy capacity region lies in development of upper bounds in certain recursive structures for three or more consecutive layers of receivers. Our contribution here lies in the construction of a series of covariance matrices representing input resources for layered messages such that the

secrecy rates can be upper bounded as the desired recursive forms in terms of these covariance matrices. The details can be referred to Section 4.7.

4.3 Application to Secret Sharing

In this section, we apply our result in Theorem 4.2 to study a secret sharing problem, in which a dealer wishes to share K secrets $W_1, W_2, \ldots, W_K$ with K participants via a broadcast channel (see Fig. 4.2). The channel input sent by the dealer is denoted by $\mathbf{X}$ and the channel output received at participant k is denoted by Y_k for $k = 1, \ldots, K$. It is required that participant 1 decodes W_1, and participant 1 and 2 decode W_1 and W_2 by sharing their outputs (Y_1, Y_2), but W_2 should be kept secure from participant 1. Such requirements extend to k participants for $k = 1, \ldots, K$ in the sense that participants 1 to k can recover the first k messages $W_1, \ldots, W_k$ by sharing their outputs $(Y_1, \ldots, Y_k)$, but the new message W_k should be secure from the first $k-1$ participants. Hence, as one more participant joins the group, one more secret can be recovered, and this new secret is secure from (and hence cannot be recovered by) a smaller group. The goal is to characterize the secret sharing capacity region, which contains all possible achievable rate tuples $(R_1, R_2, \ldots, R_K)$ for K secrets.

This secret sharing problem involves sharing multiple secrets in a layered fashion, and is challenging to solve using the classical approach based on number theory. Here, we solve this problem by constructing an equivalent broadcast model described in Section 4.1.

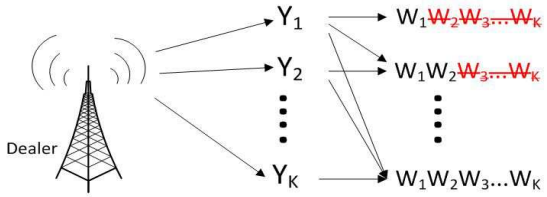

Fig. 4.2: The model of secret sharing via a broadcast channel.

We assume that the dealer communicates to the participants via a Gaussian multiple input single

output (MISO) broadcast channel corrupted by additive Gaussian noise variables. The dealer has K antennas and each receiver has one antenna. The relationship of the channel input from the dealer and the channel outputs at all participants is characterized as

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_K \end{pmatrix} = \mathbf{H} \begin{pmatrix} X_1 \\ \vdots \\ X_K \end{pmatrix} + \begin{pmatrix} Z_1 \\ \vdots \\ Z_K \end{pmatrix} \tag{4.12}$$

where $\mathbf{H}$ is the $K \times K$ channel matrix, which is assumed to be invertible, $(Y_1, \ldots, Y_K)$ are channel outputs at the K participants, $(X_1, \ldots X_K)$ are the channel inputs from the K antennas of the dealer, and $(Z_1, \ldots, Z_K)$ is a random Gaussian vector with the covariance matrix $\mathbf{\Sigma}$ with each entry $\mathbf{\Sigma}_{ij} = E[Z_i Z_j] = \sigma_{ij}^2$. We assume that the dealer's input is subject to a resource constraint, $E[\mathbf{X}\mathbf{X}^T] \preceq \mathbf{S}$.

We note that it is reasonable to assume that $\mathbf{H}$ is invertible in order to guarantee that each participant's output contains new information compared to other participants so that new secret can be recovered when this participant joins a group.

We reformulate the above secret sharing model into a degraded MIMO broadcast communication system by designing a virtual receiver for each sharing group of participants. More specifically, we design a virtual receiver $\mathbf{V}_k$ for the group of the first k participants, i.e., $\mathbf{V}_k = (Y_1, \ldots, Y_k)$, for $1 \leq k \leq K$. For technical convenience, we add $K - k$ outputs $\widetilde{Y}_{k+1}, \ldots, \widetilde{Y}_K$ to $\mathbf{V}_k$ so that it contains K components, i.e., the virtual receiver $\mathbf{V}_k$ has K antennas. The channel

outputs at those K antennas are given by,

$$\mathbf{V}_k = \begin{pmatrix} Y_1 \\ \vdots \\ Y_k \\ \widetilde{Y}_{k+1} \\ \vdots \\ \widetilde{Y}_K \end{pmatrix} = \mathbf{H} \begin{pmatrix} X_1 \\ \vdots \\ X_K \end{pmatrix} + \begin{pmatrix} Z_1 \\ \vdots \\ Z_k \\ Z_{k+1} + t\widetilde{Z}_{k+1} \\ \vdots \\ Z_K + t\widetilde{Z}_K \end{pmatrix} \tag{4.13}$$

where $\widetilde{Z}_k$, $2 \leq k \leq K$, is random Gaussian noise variables with mean zero and variance $\tilde{\sigma}_{kk}^2 > 0$, and $\widetilde{Z}_k$ is independent from all other random variables. Here, t is a large enough constant (i.e., $t \to \infty$), so that $\widetilde{Y}_{k+1}, \ldots, \widetilde{Y}_K$ are fully corrupted by the noise. We define a new random Gaussian vector $\mathbf{Z}_V(k) = (Z_1, \ldots, Z_k, Z_{k+1} + t\widetilde{Z}_{k+1}, \ldots, Z_K + t\widetilde{Z}_K)^T$ and rewrite (4.13) as

$$\mathbf{V}_k = \mathbf{H}\mathbf{X} + \mathbf{Z}_V(k), \text{ for } k = 1, \ldots, K. \tag{4.14}$$

Since the channel matrix $\mathbf{H}$ is invertible, we have

$$\mathbf{H}^{-1}\mathbf{V}_k = \mathbf{X} + \mathbf{H}^{-1}\mathbf{Z}_V(k). \tag{4.15}$$

By treating $\mathbf{H}^{-1}\mathbf{V}_k$ as the new channel output $\mathbf{V}'_k$ at virtual receiver $\mathbf{V}_k$, and define a new random Gaussian noise vector $\mathbf{Z}'_V(k) = \mathbf{H}^{-1}\mathbf{Z}_V(k)$, we have

$$\mathbf{V}'_k = \mathbf{X} + \mathbf{Z}'_V(k), \tag{4.16}$$

which is equivalent to the model in (4.14).

We now state a lemma that provides the order of the covariance matrices of $\mathbf{Z}'_V(k)$, denoted by $\mathbf{\Sigma}'_V(k)$, for $1 \leq k \leq K$.

Lemma 4.1. *Let $\mathbf{Z}'_V(k)$, $1 \leq k \leq K$, be random Gaussian vectors defined as above. The covari-*

ance matrices of $\mathbf{Z}'_V(k)$ *satisfy the following ordering property:*

$$\mathbf{\Sigma}'_V(1) \succeq \mathbf{\Sigma}'_V(2) \succeq \ldots \succeq \mathbf{\Sigma}'_V(K). \tag{4.17}$$

Proof. For any $1 \leq k \leq K-1$,

$$Z'_{V_k} = Z'_{V_{k+1}} + \mathbf{H}^{-1}(0, \cdots, 0, t\widetilde{Z}_{k+1}, 0, \cdots, 0)^T, \tag{4.18}$$

where $\mathbf{H}^{-1}(0, \cdots, 0, t\widetilde{Z}_{k+1}, 0, \cdots, 0)^T$ is a random Gaussian vector, hence the covariance matrices of $\mathbf{Z}'_{V_k}$ satisfy such an order in (4.17). □

Therefore, by designing virtual receivers, we reformulate the problem of secret sharing via the MISO broadcast channel into the problem of secure communication over the degraded MIMO broadcast channel described in Section 4.1. It can also be seen that the requirements of the secret sharing problem is equivalent to the layered decoding and secrecy requirements for the communication problem. Thus, the secret sharing capacity region equals the secrecy capacity region of the degraded MIMO broadcast channel. Thus applying Theorem 4.2 we obtain the following secret sharing capacity region.

Corollary 4.1. *The capacity region for the secret sharing problem described above contains rate tuples* $(R_1, R_2, \ldots, R_K)$ *satisfying*

$$\begin{aligned}
R_1 &\leq \frac{1}{2}\log\frac{|\mathbf{\Sigma}'_V(1)+\mathbf{S}|}{|\mathbf{\Sigma}'_V(1)+\mathbf{S}_1|} \\
R_k &\leq \lim_{t\to\infty}\frac{1}{2}\log\frac{|\mathbf{\Sigma}'_V(k)+\mathbf{S}_{k-1}|}{|\mathbf{\Sigma}'_V(k)+\mathbf{S}_k|} - \frac{1}{2}\log\frac{|\mathbf{\Sigma}'_V(k-1)+\mathbf{S}_{k-1}|}{|\mathbf{\Sigma}'_V(k-1)+\mathbf{S}_k|}, \\
&\qquad\qquad \textit{for } 2 \leq k \leq K-1, \\
R_K &\leq \lim_{t\to\infty}\frac{1}{2}\log\frac{|\mathbf{\Sigma}'_V(K)+\mathbf{S}_{K-1}|}{|\mathbf{\Sigma}'_V(K)|} - \frac{1}{2}\log\frac{|\mathbf{\Sigma}'_V(K-1)+\mathbf{S}_{K-1}|}{|\mathbf{\Sigma}'_V(K-1)|},
\end{aligned} \tag{4.19}$$

for some $\mathbf{0} \preceq \mathbf{S}_{K-1} \preceq \mathbf{S}_{K-2} \preceq \ldots \preceq \mathbf{S}_2 \preceq \mathbf{S}_1 \preceq \mathbf{S}$.

4.4 Achievability Proof of Theorem 4.1

The achievability proof is based on stochastic encoding and superposition coding. We use random codes and fix a joint probability distribution $P_{U_1 \cdots U_{K-1} X}$ satisfying the Markov chain condition given in (4.9). Let $T_\epsilon^n(P_{U_1 \ldots U_{K-1} X Y_1 \ldots Y_K})$ denote the strongly jointly ϵ-typical set based on the fixed distribution.

Random codebook generation: In the following achievability proof, for notational convenience, we write X as U_K, i.e., $P_{U_1 \cdots U_{K-1} X} = P_{U_1 \cdots U_K}$.

- Generate 2^{nR_1} independent and identically distributed (i.i.d.) u_1^n with distribution $\prod_{i=1}^n p(u_{1,i})$. Index these codewords as $u_1^n(w_1)$, $w_1 \in [1, 2^{nR_1}]$.
- For each $u_{k-1}^n(w_1, w_2, l_2, \cdots, w_{k-1}, l_{k-1})$, $k = 2, \cdots, K$, generate $2^{n\widetilde{R}_k}$ i.i.d. sequences u_k^n with distribution $\prod_{i=1}^n p(u_{k,i}|u_{k-1,i})$. Partition these sequences into 2^{nR_k} bins, each with $2^{n(\widetilde{R}_k - R_k)}$ sequences. We use $w_k \in [1 : 2^{nR_k}]$ to denote the bin index, and $l_k \in [1 : 2^{n(\widetilde{R}_k - R_k)}]$ to denote the index within each bin. Hence each u_k^n is indexed by $(w_1, w_2, l_2, \cdots, w_k, l_k)$.

The chosen codebook is revealed to the transmitter and all receivers.

Encoding: To send a message tuple $(w_1, w_2, \ldots, w_K)$, for each $2 \le k \le K$, the encoder randomly generate $l_k \in [1 : 2^{n(\widetilde{R}_k - R_k)}]$ based on a uniform distribution. The transmitter then sends $u_K^n(w_1, w_2, l_2, \cdots, w_K, l_K)$.

Decoding: For $k = 1, \ldots, K$, receiver k claims that $(\hat{w}_1, \cdots, \hat{w}_k)$ is sent, if there exists a unique tuple $(\hat{w}_1, \hat{w}_2, \hat{l}_2, \cdots, \hat{w}_k, \hat{l}_k)$ such that $(u_1^n(\hat{w}_1), u_2^n(\hat{w}_1, \hat{w}_2, \hat{l}_2), \ldots, u_k^n(\hat{w}_1, \hat{w}_2, \hat{l}_2, \cdots, \hat{w}_k, \hat{l}_k), y_k^n) \in T_\epsilon^n(P_{U_1 \ldots U_k Y_k})$. Otherwise, it declares an error.

Analysis of error probability: By the law of large numbers and the packing lemma [48], it can be shown that if the following inequalities are satisfied, receiver k (for $k = 1, \ldots, K$) can decode messages $w_1, w_2, \ldots, w_k$ with a vanishing error probability:

$$\begin{aligned} R_1 &\le I(U_1; Y_1), \\ \widetilde{R}_k &\le I(U_k; Y_k | U_{k-1}), \quad \text{for } 2 \le k \le K. \end{aligned} \tag{4.20}$$

Analysis of leakage rate: We first compute an average of the leakage rate over the random codebook ensemble as follows. For convenience, we let $W^k = (W_1, \ldots, W_k)$, $W_{k+1}^K = (W_{k+1}, \ldots, W_K)$, and $L^K = (L_1, \ldots, L_K)$. We note that l_k appeared above is realization of the random variable L_k here.

$$
\begin{aligned}
& I(W_{k+1}^K; Y_k^n|\mathcal{C}) \\
& \overset{(a)}{=} I(W^K, L^K; Y_k^n|\mathcal{C}) - I(W^k, L^K; Y_k^n|W_{k+1}^K, \mathcal{C}) \\
& \overset{(b)}{\leq} I(W^K, L^K; Y_k^n|\mathcal{C}) - I(W^k, L^K; Y_k^n|W_{k+1}^K, \mathcal{C}) \\
& \overset{(c)}{\leq} I(U_K^n; Y_k^n|\mathcal{C}) - I(W^k, L^K; Y_k^n|W_{k+1}^K, \mathcal{C}) \\
& = I(U_K^n; Y_k^n|\mathcal{C}) - H(W^k, L^K|W_{k+1}^K, \mathcal{C}) \\
& \quad + H(W^k, L^K|Y_k^n, W_{k+1}^K, \mathcal{C}),
\end{aligned} \tag{4.21}
$$

where step (a) is due to the independence of W^k and W_{k+1}^K, (b) follows from Fano's inequality, step (c) follows from the Markov chain $(W^K, L^K) \to (U_K^n, \mathcal{C}) \to Y_k^n$.

We bound the above three terms one by one. For the first term, we have

$$
\begin{aligned}
& I(U_K^n; Y_k^n|\mathcal{C}) \overset{(a)}{=} I(U_k^n, U_K^n; Y_k^n|\mathcal{C}) \\
& = I(U_k^n; Y_k^n|\mathcal{C}) + I(U_K^n; Y_k^n|U_k^n, \mathcal{C}) \\
& \leq H(U_k^n|\mathcal{C}) + I(U_K^n; Y_k^n|U_k^n, \mathcal{C}) \\
& \leq \sum_{j=1}^{k} \widetilde{R}_j + H(Y_k^n|U_k^n, \mathcal{C}) - H(Y_k^n|U_K^n, U_k^n, \mathcal{C}) \\
& = n\sum_{j=1}^{k} \widetilde{R}_j + \sum_{j=1}^{n} H(Y_{k,j}|U_k^n, Y_k^{j-1}, \mathcal{C}) - \sum_{j=1}^{n} H(Y_{k,j}|U_K^n, U_k^n, Y_k^{j-1}, \mathcal{C}) \\
& \overset{(b)}{\leq} n\sum_{j=1}^{k} \widetilde{R}_j + \sum_{j=1}^{n} H(Y_{k,j}|U_{k,j}) - \sum_{j=1}^{n} H(Y_{k,j}|U_{K,j}) \\
& = n\sum_{j=1}^{k} \widetilde{R}_j + nH(Y_k|U_k) - nH(Y_k|U_K) = n\sum_{j=1}^{k} \widetilde{R}_j + nI(U_K; Y_k|U_k),
\end{aligned} \tag{4.22}
$$

where (a) follows from the Markov chain $U_k^n \to U_K^n \to Y_k^n$, (b) follows from the fact that $H(Y_{k,j}|U_k^n, Y_k^{j-1}, \mathcal{C}) \le H(Y_{k,j}|U_{k,j})$ and from the Markov chain $(U_k^n, U_K^{j-1}, U_{K,j+1}^n, Y_k^{j-1}, \mathcal{C}) \to U_{K,j} \to Y_{k,j}$.

For the second term, due to the independence of $W_1, \cdots, W_K$ and $L_1, \cdots, L_K$, we have

$$H(W^k, L^K|W_{k+1}^K, \mathcal{C}) = \sum_{j=1}^{k} n\widetilde{R}_j + \sum_{j=k+1}^{K} n(\widetilde{R}_j - R_j). \tag{4.23}$$

We now bound the last term as follows.

$$\begin{aligned}
&H(W^k, L^K|Y_k^n, W_{k+1}^K, \mathcal{C}) \\
&= H(W^k|Y_k^n, W_{k+1}^K, \mathcal{C}) + H(L^K|Y_k^n, W^K, \mathcal{C}) \\
&\overset{(a)}{\le} H(L_{k+1}^K|Y_k^n, W^K, L^k, \mathcal{C}) + 2n\epsilon_n \\
&= \sum_{j=k+1}^{K} H(L_j|Y_k^n, W^K, L^{j-1}, \mathcal{C}) + 2n\epsilon_n \\
&\overset{(b)}{=} \sum_{j=k+1}^{K} H(L_j|Y_k^n, W^K, L^{j-1}, U_{j-1}^n, \mathcal{C}) + 2n\epsilon_n \\
&\le \sum_{j=k+1}^{K} H(L_j|Y_k^n, U_{j-1}^n, W_j) + 2n\epsilon_n \\
&\overset{(c)}{\le} \sum_{j=k+1}^{K} n(\tilde{R}_j - R_j - I(U_j; Y_k|U_{j-1})) + n\epsilon_n' \\
&\overset{(d)}{=} \sum_{j=k+1}^{K} n(\tilde{R}_j - R_j) - I(U_K; Y_k|U_k) + n\epsilon_n',
\end{aligned} \tag{4.24}$$

where (a) follows from the chain rule and Fano's inequality, (b) follows from the fact that U_{j-1}^n is a function of $(\mathcal{C}, W^{j-1}, L^{j-1})$, and (c) follows due to Lemma 4.2 with the condition that $\tilde{R}_j - R_j \ge I(U_j; Y_k|U_{j-1})$, and (d) follows from the Markov chain $U_1 \to U_2 \to \cdots \to U_K \to Y_k$.

Lemma 4.2. *If $\tilde{R}_j - R_j \ge I(U_j; Y_k|U_{j-1})$ for $k+1 \le j \le K$, then*

$$\frac{1}{n} H(L_j|Y_k^n, U_{j-1}^n, W_j) \le \widetilde{R}_j - R_j - I(U_j; Y_k|U_{j-1}) + \epsilon_n''.$$

Proof. See Section 4.5. □

Combining the analysis of the three terms together, we conclude that as $n \to \infty$ for $1 \leq k \leq K-1$, $\frac{1}{n} I(W_{k+1}^K; Y_k^n | \mathcal{C}) \to 0$, if

$$\widetilde{R}_k - R_k \geq I(U_k; Y_{k-1} | U_{k-1}), \qquad \text{for } 2 \leq k \leq K. \tag{4.25}$$

It is also clear that the sum of the error probability and the leakage rates averaged over the codebook ensemble converges to zero as $n \to \infty$. Hence, there exists one codebook such that the error probability and the leakage rate converge to zero as $n \to \infty$.

Combining the bounds in (4.20) and (4.25), we obtain that the rate tuple $(R_1, \cdots, R_K)$ is achievable if

$$\begin{aligned} &R_1 \leq I(U_1; Y_1), \\ &R_k \leq I(U_k; Y_k | U_{k-1}) - I(U_k; Y_{k-1} | U_{k-1}), \text{ for } 2 \leq k \leq K. \end{aligned} \tag{4.26}$$

4.5 Proof of Lemma 4.2

We first bound $\frac{1}{n} H(L_j | Y_k^n, U_{j-1}^n, w_j)$ for any w_j, and hence, $\frac{1}{n} H(L_j | Y_k^n, U_{j-1}^n, W_j)$ is bounded.

Fix $L_j = l_j$ and a joint typical sequence $(u_{j-1}^n, y_k^n) \in T_\epsilon^{(n)}(U_{j-1}, Y_k)$. We define

$$N(w_j, l_j, u_{j-1}^n, y_k^n) := |\{\tilde{l}_j \neq l_j : (U_j^n(w_j, \tilde{l}_j), u_{j-1}^n, y_k^n) \in T_\epsilon^{(n)}\}|. \tag{4.27}$$

In fact, $N(w_j, l_j, u_{j-1}^n, y_k^n)$ can be viewed as a binomial distributed random variable, with $2^{n(\widetilde{R}_j - R_j)} - 1$ Bernoulli distributed random variables, each taking the value 1 with probability

$$2^{-nI(U_j; Y_k | U_{j-1}) - n\delta_n(\epsilon)} \leq p \leq 2^{-nI(U_j; Y_k | U_{j-1}) + n\delta_n(\epsilon)} \tag{4.28}$$

It can be shown that the expectation and variance of N satisfy the following inequalities:

$$2^{n(\tilde{R}_j - R_j) - nI(U_j;Y_k|U_{j-1}) - n\delta_n(\epsilon) - n\epsilon_n} \leq E(N(w_j, l_j, u_{j-1}^n, y_k^n)) \leq 2^{n(\tilde{R}_j - R_j) - nI(U_j;Y_k|U_{j-1}) + n\delta_n(\epsilon) - n\epsilon_n}, \tag{4.29}$$

$$Var(N(w_j, l_j, u_{j-1}^n, y_k^n)) \leq 2^{n(\tilde{R}_j - R_j) - nI(U_j;Y_k|U_{j-1}) + n\delta_n(\epsilon) - n\epsilon_n}, \tag{4.30}$$

where $\delta_n(\epsilon), \epsilon_n \to 0$ as $n \to \infty$.

We next define the random event,

$$\varepsilon(w_j, l_j, u_{j-1}^n, y_k^n) := \{N(w_j, l_j, u_{j-1}^n, y_k^n) \geq 2^{n(\tilde{R}_j - R_j - I(U_j;Y_k|U_{j-1}) + \delta_n(\epsilon) - \epsilon_n/2) + 1}\}. \tag{4.31}$$

Using Chebyshev's inequality, we obtain

$$\begin{aligned}
&P\left(\varepsilon(w_j, l_j, u_{j-1}^n, y_k^n)\right) \\
&= P\big(N(w_j, l_j, u_{j-1}^n, y_k^n) \geq 2^{n(\tilde{R}_j - R_j - I(U_j;Y_k|U_{j-1}) + \delta_n(\epsilon) - \epsilon_n/2) + 1}\big) \\
&\leq P\big(N(w_j, l_j, u_{j-1}^n, y_k^n) \geq E(N(w_j, l_j, u_{j-1}^n, y_k^n)) + 2^{n(\tilde{R}_j - R_j - I(U_j;Y_k|U_{j-1}) + \delta_n(\epsilon) - \epsilon_n/2)}\big) \\
&\leq P\big(|N(w_j, l_j, u_{j-1}^n, y_k^n) - E(N(w_j, l_j, u_{j-1}^n, y_k^n))| \geq 2^{n(\tilde{R}_j - R_j - I(U_j;Y_k|U_{j-1}) + \delta_n(\epsilon) - \epsilon_n/2)}\big) \\
&\leq \frac{Var(N(w_j, l_j, u_{j-1}^n, y_k^n))}{2^{2n(\tilde{R}_j - R_j - I(U_j;Y_k|U_{j-1}) + \delta_n(\epsilon) - \epsilon_n/2)}} \\
&\leq \frac{1}{2^{n(\tilde{R}_j - R_j - I(U_j;Y_k|U_{j-1}) + \delta_n(\epsilon))}}
\end{aligned} \tag{4.32}$$

which goes to zero as $n \to \infty$ if $\tilde{R}_j - R_j \geq I(U_j; Y_k|U_{j-1})$. This implies that

$$P\left(\varepsilon(w_j, l_j, u_{j-1}^n, y_k^n)\right) \to 0$$

as $n \to \infty$.

For each message w_j, we define the following random variable and event:

$$N(w_j) := |\{\tilde{l}_j : (U_j^n(w_j, \tilde{l}_j), Y_k^n, U_{j-1}^n) \in T_\epsilon^{(n)}, \tilde{l}_j \neq L_j\}|$$
$$\varepsilon(w_j) := \{N(w_j) \geq 2^{n(\tilde{R}_j - R_j - I(U_j;Y_k|U_{j-1}) + \delta_n(\epsilon) - \epsilon_n/2)+1}\}$$

Finally, define the indicator random variable $E(w_j) := 0$ if $(U_j^n(w_j, L_j), Y_k^n, U_{j-1}^n) \in T_\epsilon^{(n)}$ and $\varepsilon(w_j)^c$ occurs; and $E(w_j) := 1$, otherwise. Therefore, we have

$$P\left(E(w_j) = 1\right) \leq P\left((U_j^n(w_j, L_j), U_{j-1}^n, Y_k^n) \notin T_\epsilon^{(n)}\right) + P\left(\varepsilon(w_j) | (U_{j-1}^n, Y_k^n) \in T_\epsilon^{(n)}\right). \quad (4.33)$$

It is clear that the first term in (4.33) goes to zero as $n \to \infty$. For the second term in (4.33), we have

$$\begin{aligned}
&P\left(\varepsilon(w_j) | (U_{j-1}^n, Y_k^n) \in T_\epsilon^{(n)}\right) \\
&\leq \sum_{(u_{j-1}^n, y_k^n) \in T_\epsilon^{(n)}} P\left(u_{j-1}^n, y_k^n\right) P\left(\varepsilon(w_j) | u_{j-1}^n, y_k^n\right) \\
&= \sum_{(u_{j-1}^n, y_k^n) \in T_\epsilon^{(n)}} \sum_{l_j} \left(P\left(u_{j-1}^n, y_k^n\right) P\left(l_j | u_{j-1}^n, y_k^n\right) P\left(\varepsilon(w_j) | u_{j-1}^n, y_k^n, l_j\right) \right) \\
&\to 0, \quad \text{if } \tilde{R}_j - R_j \geq I(U_j; Y_k | U_{j-1}). \qquad (4.34)
\end{aligned}$$

Therefore,

$$\begin{aligned}
&H(L_j | w_j, U_{j-1}^n, Y_k^n) \\
&\leq H(L_j, E(w_j) | w_j, U_{j-1}^n, Y_k^n) \\
&\leq H(E(w_j)) + H(L_j | w_j, U_{j-1}^n, Y_k^n, E(w_j)) \\
&\leq 1 + P\left(E(w_j) = 1\right) H(L_j | w_j, Y_k^n, U_{j-1}^n, E(w_j) = 1) + H(L_j | w_j, Y_k^n, U_{j-1}^n, E(w_j) = 0) \\
&\leq 1 + P\left(E(w_j) = 1\right) n(\tilde{R}_j - R_j) + \log 2^{n(\tilde{R}_j - R_j - I(U_j;Y_k|U_{j-1}) + \delta(\epsilon) - \epsilon/2)+1} \\
&= 1 + n(\tilde{R}_j - R_j) P\left(E(w_j) = 1\right) + n(\tilde{R}_j - R_j - I(U_j; Y_k | U_{j-1}) + \delta(\epsilon) - \epsilon/2) + 1 \qquad (4.35)
\end{aligned}$$

Following from (4.35), we obtain,

$$\lim_{n\to\infty}\frac{1}{n}H(L_j|w_j,U_{j-1}^n,Y_k^n)\leq\tilde{R}_j-R_j-I(U_j;Y_k|U_{j-1})+\delta'(\epsilon), \tag{4.36}$$

where $\delta'(\epsilon)\to 0$ as $n\to\infty$. This concludes the proof.

4.6 Converse Proof of Theorem 4.1

By Fano's inequality and the secrecy requirements, we have the following inequalities

$$\begin{aligned} H(W_k|Y_k^n)&\leq n\epsilon_n, \quad \text{for } 1\leq k\leq K,\\ \frac{1}{n}I(W_{k+1},\ldots,W_K;Y_k^n|W_1,\ldots,W_k)&\leq\epsilon_n, \text{ for } 1\leq k\leq K-1. \end{aligned} \tag{4.37}$$

We let $Y_k^{i-1}:=(Y_{k,1},\ldots,Y_{k,i-1})$, and $Y_{k,i+1}^n:=(Y_{k,i+1},\ldots,Y_{k,n})$. We set $U_{k,i}:=\{W_1,\ldots,W_k,Y_k^{i-1},Y_{k-1,i+1}^n\}$ for $k=1,\ldots,K$ where $Y_0^n=\phi$. Due to the degradedness condition (4.1), it can be verified that $(U_{1,i},\ldots,U_{K-1,i},X_i)$ satisfy the following Markov chain condition:

$$U_{1,i}\to U_{2,i}\to\ldots\to U_{K-1,i}\to X_i\to Y_{K,i}\to\ldots\to Y_{1,i}, \text{for } 1\leq i\leq n. \tag{4.38}$$

We first bound the rate R_1. Since there is no secrecy constraint for W_1, following the standard steps, we obtain the following bound:

$$\begin{aligned} nR_1&=H(W_1)=I(W_1;Y_1^n)+H(W_1|Y_1^n)\leq I(W_1;Y_1^n)+n\epsilon_n\\ &=\sum_{i=1}^n I(W_1;Y_{1i}|Y_1^{i-1})+n\epsilon_n\leq\sum_{i=1}^n I(W_1,Y_1^{i-1};Y_{1i})+n\epsilon_n\\ &=\sum_{i=1}^n I(U_{1,i};Y_{1,i})+n\epsilon_n. \end{aligned} \tag{4.39}$$

For the message W_k, $2 \leq k \leq K$, we derive the bound as shown in (4.40),

$$
\begin{aligned}
nR_k =& H(W_k) = H(W_k|W^{k-1}) \\
=& I(W_k; Y_k^n|W^{k-1}) + H(W_k|W^{k-1}, Y_k^n) \\
\overset{(a)}{\leq}& I(W_k; Y_k^n|W^{k-1}) + n\epsilon_n \\
\overset{(b)}{\leq}& I(W_k; Y_k^n|W^{k-1}) + 2n\epsilon_n - I(W_k; Y_{k-1}^n|W^{k-1}) \\
=& \sum_{i=1}^{n} I(W_k; Y_{k,i}|W^{k-1}, Y_k^{i-1}) + 2n\epsilon_n - \sum_{i=1}^{n} I(W_k; Y_{k-1,i}|W^{k-1}, Y_{k-1,i+1}^n) \\
=& \sum_{i=1}^{n} \Bigg[I(W_k, Y_{k-1,i+1}^n; Y_{k,i}|W^{k-1}, Y_k^{i-1}) - I(W_k, Y_k^{i-1}; Y_{k-1,i}|W^{k-1}, Y_{k-1,i+1}^n) \\
& - I(Y_{k-1,i+1}^n; Y_{k,i}|W^k, Y_k^{i-1}) + I(Y_k^{i-1}; Y_{k-1,i}|W^k, Y_{k-1,i+1}^n) \Bigg] + 2n\epsilon_n \\
\overset{(c)}{=}& \sum_{i=1}^{n} \Bigg[I(Y_{k-1,i+1}^n; Y_{k,i}|W^{k-1}, Y_k^{i-1}) + I(W_k; Y_{k,i}|W^{k-1}, Y_k^{i-1}, Y_{k-1,i+1}^n) \\
& - I(Y_k^{i-1}; Y_{k-1,i}|W^{k-1}, Y_{k-1,i+1}^n) - I(W_k; Y_{k-1,i}|W^{k-1}, Y_k^{i-1}, Y_{k-1,i+1}^n) \Bigg] + 2n\epsilon_n \\
\overset{(d)}{=}& \sum_{i=1}^{n} \Bigg[I(W_k; Y_{k,i}|W^{k-1}, Y_k^{i-1}, Y_{k-1,i+1}^n) - I(W_k; Y_{k-1,i}|W^{k-1}, Y_k^{i-1}, Y_{k-1,i+1}^n) \Bigg] + 2n\epsilon_n \\
=& \sum_{i=1}^{n} \Bigg[I(W_k, Y_k^{i-1}, Y_{k-1,i+1}^n; Y_{k,i}|W^{k-1}) - I(Y_k^{i-1}, Y_{k-1,i+1}^n; Y_{k,i}|W^{k-1}) \\
& - I(W_k, Y_k^{i-1}, Y_{k-1,i+1}^n; Y_{k-1,i}|W^{k-1}) + I(Y_k^{i-1}, Y_{k-1,i+1}^n; Y_{k-1,i}|W^{k-1}) \Bigg] + 2n\epsilon_n \\
=& \sum_{i=1}^{n} \Bigg[I(W_k, Y_k^{i-1}, Y_{k-1,i+1}^n, Y_{k-1}^{i-1}, Y_{k-2,i+1}^n; Y_{k,i}|W^{k-1}) \\
& - I(W_k, Y_k^{i-1}, Y_{k-1,i+1}^n, Y_{k-1}^{i-1}, Y_{k-2,i+1}^n; Y_{k-1,i}|W^{k-1}) - I(Y_k^{i-1}, Y_{k-1,i+1}^n; Y_{k,i}|W^{k-1}) \\
& + I(Y_k^{i-1}, Y_{k-1,i+1}^n; Y_{k-1,i}|W^{k-1}) \Bigg] + 2n\epsilon_n
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \Big[I(W_k, Y_k^{i-1}, Y_{k-1,i+1}^{n}; Y_{k,i} | W^{k-1}, Y_{k-1}^{i-1}, Y_{k-2,i+1}^{n}) \\
&\quad - I(W_k, Y_k^{i-1}, Y_{k-1,i+1}^{n}; Y_{k-1,i} | W^{k-1}, Y_{k-1}^{i-1}, Y_{k-2,i+1}^{n}) + I(Y_{k-1}^{i-1}, Y_{k-2,i+1}^{n}; Y_{k,i} | W^{k-1}) \\
&\quad - I(Y_{k-1}^{i-1}, Y_{k-2,i+1}^{n}; Y_{k-1,i} | W^{k-1}) - I(Y_k^{i-1}, Y_{k-1,i+1}^{n}; Y_{k,i} | W^{k-1}) \\
&\quad + I(Y_k^{i-1}, Y_{k-1,i+1}^{n}; Y_{k-1,i} | W^{k-1}) \Big] + 2n\epsilon_n \\
&\overset{(e)}{=} \sum_{i=1}^{n} \Big[I(W_k, Y_k^{i-1}, Y_{k-1,i+1}^{n}; Y_{k,i} | W^{k-1}, Y_{k-1}^{i-1}, Y_{k-2,i+1}^{n}) \\
&\quad - I(W_k, Y_k^{i-1}, Y_{k-1,i+1}^{n}; Y_{k-1,i} | W^{k-1}, Y_{k-1}^{i-1}, Y_{k-2,i+1}^{n}) \\
&\quad - I(Y_k^{i-1}, Y_{k-1,i+1}^{n}; Y_{k,i} | W^{k-1}, Y_{k-1}^{i-1}, Y_{k-2,i+1}^{n}) \\
&\quad + I(Y_k^{i-1}, Y_{k-1,i+1}^{n}; Y_{k-1,i} | W^{k-1}, Y_{k-1}^{i-1}, Y_{k-2,i+1}^{n}) \Big] + 2n\epsilon_n \\
&\overset{(f)}{\leq} \sum_{i=1}^{n} \Big[I(W_k, Y_k^{i-1}, Y_{k-1,i+1}^{n}; Y_{k,i} | W^{k-1}, Y_{k-1}^{i-1}, Y_{k-2,i+1}^{n}) \\
&\quad - I(W_k, Y_k^{i-1}, Y_{k-1,i+1}^{n}; Y_{k-1,i} | W^{k-1}, Y_{k-1}^{i-1}, Y_{k-2,i+1}^{n}) \Big] + 2n\epsilon_n \\
&= \sum_{i=1}^{n} \Big[I(U_{k,i}; Y_{k,i} | U_{k-1,i}) - I(U_{k,i}; Y_{k-1,i} | U_{k-1,i}) \Big] + 2n\epsilon_n. \qquad (4.40)
\end{aligned}
$$

where (a) follows from Fano's inequality, (b) follows from (4.37), i.e., the secrecy constraint, (c) and (d) follow from the sum identity property in [16, Lemma 7], and (e) and (f) follows from the degradedness condition (4.1).

For $k = K$, based on (4.40), we obtain

$$
\begin{aligned}
nR_K &\leq \sum_{i=1}^{n} \Big[I(U_{K,i}; Y_{K,i} | U_{K-1,i}) - I(U_{K,i}; Y_{K-1,i} | U_{K-1,i}) \Big] + 2n\epsilon_n \\
&= \sum_{i=1}^{n} \Big[I(U_{K,i}, X_i; Y_{K,i} | U_{K-1,i}) - I(U_{K,i}, X_i; Y_{K-1,i} | U_{K-1,i}) - I(X_i; Y_{K,i} | U_{K,i}) \\
&\qquad + I(X_i; Y_{K-1,i} | U_{K,i}) \Big] + 2n\epsilon_n \\
&\leq \sum_{i=1}^{n} \Big[I(X_i; Y_{K,i} | U_{K-1,i}) - I(X_i; Y_{K-1,i} | U_{K-1,i}) \Big] + 2n\epsilon_n, \qquad (4.41)
\end{aligned}
$$

where the last step follows from (4.38) and (4.1). The proof of the converse is completed by defining a uniformly distributed random variable $Q \in \{1, \cdots, n\}$, and setting $U_k \triangleq (Q, U_{k,Q})$, $Y_k \triangleq Y_{k,Q}$, for $k \in [1:K]$, and $X \triangleq (Q, X_Q)$.

4.7 Converse Proof of Theorem 4.2

In this proof, we first introduce some necessary definitions and useful lemmas in the previous studies [41, 75]. We then present our main proof.

4.7.1 Preliminaries

Definition 4.1. *[41] Let* $(\mathbf{U}, \mathbf{X})$ *be an arbitrarily correlated length-*n *random vector pair with well defined densities. The conditional Fisher information matrix of* $\mathbf{X}$ *given* $\mathbf{U}$ *is defined as*

$$\mathbf{J}(\mathbf{X}|\mathbf{U}) = E[\rho(\mathbf{X}|\mathbf{U})\rho(\mathbf{X}|\mathbf{U})^T] \tag{4.42}$$

where the expectation is taken over the joint density $f(\mathbf{u}, \mathbf{x})$*, and the conditional score function* $\rho(\mathbf{x}|\mathbf{u})$ *is given by*

$$\begin{aligned} \rho(\mathbf{x}|\mathbf{u}) &= \nabla \log f(\mathbf{x}|\mathbf{u}) \\ &= \left[\frac{\partial \log f_U(\mathbf{x}|\mathbf{u})}{\partial x_1} \cdots \frac{\partial \log f_U(\mathbf{x}|\mathbf{u})}{\partial x_n} \right]^T. \end{aligned} \tag{4.43}$$

Lemma 4.3. *[41, Theorem 11] Let* $(\mathbf{Z_1}, \mathbf{Z_2}, \mathbf{Z_3}, \mathbf{Z_4})$ *be Gaussian random vectors with covariance matrices* $\mathbf{\Sigma}_1$*,* $\mathbf{\Sigma}_2$*,* $\mathbf{\Sigma}_3$*,* $\mathbf{\Sigma}_4$*, respectively, where*

$$\mathbf{\Sigma}_4 \preceq \mathbf{\Sigma}_3 \preceq \mathbf{\Sigma}_2 \preceq \mathbf{\Sigma}_1. \tag{4.44}$$

Let $(\mathbf{U}, \mathbf{X})$ *be an arbitrarily dependent random vector pair, which is independent of the Gaussian random vectors* $(\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3, \mathbf{Z}_4)$*, and the second moment of* $\mathbf{X}$ *be constrained as* $E[\mathbf{X}\mathbf{X}^T] \preceq \mathbf{S}$*.*

Then, for any feasible $(\mathbf{U}, \mathbf{X})$*, for any* $\mathbf{\Sigma_1}, \mathbf{\Sigma_2}, \mathbf{\Sigma_3}, \mathbf{\Sigma_4}$ *satisfying the order in* (4.44)*, there exists a positive semidefinite matrix* $\mathbf{K}^*$ *such that* $\mathbf{K}^* \preceq \mathbf{S}$*, and*

$$h(\mathbf{X}+\mathbf{Z}_2|\mathbf{U}) - h(\mathbf{X}+\mathbf{Z}_3|\mathbf{U}) = \frac{1}{2}\log\frac{|\mathbf{K}^*+\mathbf{\Sigma}_2|}{|\mathbf{K}^*+\mathbf{\Sigma}_3|}, \tag{4.45}$$

and

$$h(\mathbf{X}+\mathbf{Z}_1|\mathbf{U}) - h(\mathbf{X}+\mathbf{Z}_3|\mathbf{U}) \leq \frac{1}{2}\log\frac{|\mathbf{K}^*+\mathbf{\Sigma}_1|}{|\mathbf{K}^*+\mathbf{\Sigma}_3|}, \tag{4.46}$$

$$h(\mathbf{X}+\mathbf{Z}_3|\mathbf{U}) - h(\mathbf{X}+\mathbf{Z}_4|\mathbf{U}) \geq \frac{1}{2}\log\frac{|\mathbf{K}^*+\mathbf{\Sigma}_3|}{|\mathbf{K}^*+\mathbf{\Sigma}_4|}. \tag{4.47}$$

Lemma 4.4. *[75] Let* $(\mathbf{U}, \mathbf{X})$ *be an arbitrarily correlated random vector pair, and the second moment of* X *is constrained as* $E[\mathbf{X}\mathbf{X}^T] \preceq S$*. Let* $\mathbf{Z}_1$*,* $\mathbf{Z}_2$ *be Gaussian random vectors that are independent from* $(\mathbf{U}, \mathbf{X})$*, and have mean zero and covariance matrices* $\mathbf{\Sigma}_1$*,* $\mathbf{\Sigma}_2$*, respectively, where* $\mathbf{\Sigma}_1 \succeq \mathbf{\Sigma}_2$*. Then* $h(\mathbf{X}+\mathbf{Z}_1|\mathbf{U}) - h(\mathbf{X}+\mathbf{Z}_2|\mathbf{U})$ *is upper and lower bounded as follows,*

$$\begin{aligned}
&\frac{1}{2}\log\frac{|\mathbf{J}(\mathbf{X}+\mathbf{Z}_1|\mathbf{U})^{-1}|}{|\mathbf{J}(\mathbf{X}+\mathbf{Z}_1|\mathbf{U})^{-1}+\mathbf{\Sigma}_2-\mathbf{\Sigma}_1|} \\
&\qquad \leq h(\mathbf{X}+\mathbf{Z}_1|\mathbf{U}) - h(\mathbf{X}+\mathbf{Z}_2|\mathbf{U}) \\
&\qquad\qquad \leq \frac{1}{2}\log\frac{|\mathbf{J}(\mathbf{X}+\mathbf{Z}_2|\mathbf{U})^{-1}+\mathbf{\Sigma}_1-\mathbf{\Sigma}_2|}{|\mathbf{J}(\mathbf{X}+\mathbf{Z}_2|\mathbf{U})^{-1}|}.
\end{aligned} \tag{4.48}$$

Proof. The proof of the unconditioned version of Lemma 4.4 is given in [75] in part B of Section V. The proof can be generalized to the conditioned version by applying mathematical tools given in part D of Section V of [75]. □

Lemma 4.5. *[75, Lemma 17] Let* $(\mathbf{V}, \mathbf{U}, \mathbf{X})$ *be* $n-$*dimentional random vectors with well-defined densities. Moreover, assume that the partial derivatives of* $f(\mathbf{u}|\mathbf{v}, \mathbf{x})$ *with respect to* x_i*,* $i = 1, \ldots, n$*, exist and satisfy*

$$\max_{1\leq i\leq n}\left|\frac{\partial f(\mathbf{u}|\mathbf{x},\mathbf{v})}{\partial x_i}\right| \leq g(\mathbf{u}), \tag{4.49}$$

for some integrable function $g(\mathbf{u})$*. If* $(\mathbf{V}, \mathbf{U}, \mathbf{X})$ *satisfy the Markov chain* $\mathbf{V} \to \mathbf{U} \to \mathbf{X}$*, then*

$$\mathbf{J}(\mathbf{X}|\mathbf{U}) \succeq \mathbf{J}(\mathbf{X}|\mathbf{V}). \tag{4.50}$$

Lemma 4.6. *[75, Lemma 10] Consider the function*

$$r(t) = \frac{1}{2} \log \frac{|\mathbf{A} + \mathbf{B} + t\mathbf{\Delta}|}{|\mathbf{A} + t\mathbf{\Delta}|},\ 0 \leq t \leq 1. \tag{4.51}$$

where $\mathbf{A}$*,* $\mathbf{B}$*,* $\mathbf{\Delta}$ *are real symmetric matrices, and* $\mathbf{A} \succ \mathbf{0}$*,* $\mathbf{B} \succeq \mathbf{0}$*,* $\mathbf{\Delta} \succeq \mathbf{0}$*. Then* $r(t)$ *is continuous and monotonically decreasing with respect to* t*.*

Lemma 4.7. *[75] Suppose* $(\mathbf{U}, \mathbf{X})$ *is a random vector pair with arbitrary joint distribution and the second order moment of* $\mathbf{X}$ *satisfies* $E(\mathbf{X}\mathbf{X}^T) \preceq S$*. Let* $\mathbf{Z}$ *be a random Gaussian vector that is independent from* $\mathbf{U}$ *and* $\mathbf{X}$ *and has mean zero and covariance* $\mathbf{\Sigma}$*. Then we have*

$$0 \preceq \mathbf{J}(\mathbf{X} + \mathbf{Z}|\mathbf{U})^{-1} - \mathbf{\Sigma} \preceq \mathbf{S}. \tag{4.52}$$

Lemma 4.8. *[75] Suppose* $(\mathbf{U}, \mathbf{X})$ *is a random vector pair with arbitrary joint distribution and the second order moment of* $\mathbf{X}$ *satisfies* $E(\mathbf{X}\mathbf{X}^T) \preceq S$*. Let* $\mathbf{Z_1}, \mathbf{Z_2}$ *be random Gaussian vectors that are independent from* $\mathbf{U}$ *and* $\mathbf{X}$ *and have mean zero and covariance matrices* $\mathbf{\Sigma_1} \succeq \mathbf{\Sigma_2}$*. Then we have*

$$\mathbf{J}(\mathbf{X} + \mathbf{Z}_1|\mathbf{U})^{-1} + \mathbf{\Sigma}_2 - \mathbf{\Sigma}_1 - \mathbf{J}(\mathbf{X} + \mathbf{Z}_2|\mathbf{U})^{-1} \succeq \mathbf{0}. \tag{4.53}$$

The proof of Lemma 4.8 follows the arguments in the proof of Lemma 6 in [75] for the unconditional case, but using Corollary 4 in [75] for the conditional case.

4.7.2 Main Proof

Following the converse proof of Theorem 4.1 in Section 4.6, we have the inequalities as follows:

$$R_1 \leq I(\mathbf{U}_1; \mathbf{Y}_1), \tag{4.54}$$

$$R_k \leq I(\mathbf{U}_k; \mathbf{Y}_k|\mathbf{U}_{k-1}) - I(\mathbf{U}_k; \mathbf{Y}_{k-1}|\mathbf{U}_{k-1}), \text{ for } 2 \leq k \leq K, \tag{4.55}$$

where the random variables satisfy the Markov chain condition in (4.9).

We first derive the bounds on R_2 and R_3 in order to show that the bounding techniques can be extended to prove the bounds on $R_4, \ldots, R_K$. We then derive the bound on R_1.

To bound R_2, we start with (4.54), and have

$$\begin{aligned} R_2 &\leq I(\mathbf{U}_2; \mathbf{Y}_2|\mathbf{U}_1) - I(\mathbf{U}_2; \mathbf{Y}_1|\mathbf{U}_1) \\ &\overset{(a)}{=} h(\mathbf{Y}_2|\mathbf{U}_1) - h(\mathbf{Y}_2|\mathbf{U}_2) - (h(\mathbf{Y}_1|\mathbf{U}_1) - h(\mathbf{Y}_1|\mathbf{U}_2)) \\ &= (h(\mathbf{Y}_2|\mathbf{U}_1) - h(\mathbf{Y}_1|\mathbf{U}_1)) - (h(\mathbf{Y}_2|\mathbf{U}_2) - h(\mathbf{Y}_1|\mathbf{U}_2)), \end{aligned} \tag{4.56}$$

where (a) follows from the Markov chain condition in (4.9).

Following from Lemma 4.4, we obtain the following upper and lower bounds on $h(\mathbf{Y}_2|\mathbf{U}_1) - h(\mathbf{Y}_1|\mathbf{U}_1)$.

$$\begin{aligned} &\frac{1}{2}\log\frac{|\mathbf{J}(\mathbf{X}+\mathbf{Z}_2|\mathbf{U}_1)^{-1}|}{|\mathbf{J}(\mathbf{X}+\mathbf{Z}_2|\mathbf{U}_1)^{-1}+\mathbf{\Sigma}_1-\mathbf{\Sigma}_2|} \\ &\qquad \leq h(\mathbf{Y}_2|\mathbf{U}_1) - h(\mathbf{Y}_1|\mathbf{U}_1) \\ &\qquad\qquad \leq \frac{1}{2}\log\frac{|\mathbf{J}(\mathbf{X}+\mathbf{Z}_1|\mathbf{U}_1)^{-1}+\mathbf{\Sigma}_2-\mathbf{\Sigma}_1|}{|\mathbf{J}(\mathbf{X}+\mathbf{Z}_1|\mathbf{U}_1)^{-1}|}. \end{aligned} \tag{4.57}$$

Define

$$\mathbf{A} = \mathbf{J}(\mathbf{X} + \mathbf{Z}_2|\mathbf{U}_1)^{-1},$$
$$\mathbf{B} = \mathbf{\Sigma}_1 - \mathbf{\Sigma}_2,$$
$$\mathbf{\Delta} = \mathbf{J}(\mathbf{X} + \mathbf{Z}_1|\mathbf{U}_1)^{-1} + \mathbf{\Sigma}_2 - \mathbf{\Sigma}_1 - \mathbf{J}(\mathbf{X} + \mathbf{Z}_2|\mathbf{U}_1)^{-1},$$

and

$$r(t) = \frac{1}{2}\log\frac{|\mathbf{A} + \mathbf{B} + t\mathbf{\Delta}|}{|\mathbf{A} + t\mathbf{\Delta}|}.$$

Therefore, (4.57) can be rewritten into,

$$-r(0) \leq h(\mathbf{Y}_2|\mathbf{U}_1) - h(\mathbf{Y}_1|\mathbf{U}_1) \leq -r(1). \tag{4.58}$$

It can be verified that $\mathbf{A} \succ 0$, $\mathbf{B} \succeq 0$, and $\mathbf{\Delta} \succeq 0$. In particular, $\mathbf{\Delta} \succeq 0$ is due to Lemma 4.8.

Following from Lemma 4.6, $r(t)$ is a continuous and monotonically decreasing function in t. Hence, (4.58) implies that there must exist a constant t_1 with $0 \leq t_1 \leq 1$, such that

$$h(\mathbf{Y}_2|\mathbf{U}_1) - h(\mathbf{Y}_1|\mathbf{U}_1) = -r(t_1). \tag{4.59}$$

We define

$$\begin{aligned}\mathbf{S}_1 :&= \mathbf{A} + t_1\mathbf{\Delta} - \mathbf{\Sigma}_2 \\ &= \mathbf{J}(\mathbf{X} + \mathbf{Z}_2|U_1)^{-1} + t_1(\mathbf{J}(\mathbf{X} + \mathbf{Z}_1|\mathbf{U}_1)^{-1} \\ &\quad + \mathbf{\Sigma}_2 - \mathbf{\Sigma}_1 - \mathbf{J}(\mathbf{X} + \mathbf{Z}_2|\mathbf{U}_1)^{-1}) - \mathbf{\Sigma}_2.\end{aligned} \tag{4.60}$$

Therefore,

$$h(\mathbf{Y}_2|\mathbf{U}_1) - h(\mathbf{Y}_1|\mathbf{U}_1) = -r(t_1) = \frac{1}{2}\log\frac{|\mathbf{S}_1 + \mathbf{\Sigma}_2|}{|\mathbf{S}_1 + \mathbf{\Sigma}_1|}. \tag{4.61}$$

It can be seen that $\mathbf{S}_1$ satisfies $\mathbf{A}-\mathbf{\Sigma}_2 \preceq \mathbf{S}_1 \preceq \mathbf{A}+\mathbf{\Delta}-\mathbf{\Sigma}_2$. Following Lemma 4.7, we have

$$\begin{aligned}\mathbf{0} &\preceq \mathbf{J}(\mathbf{X}+\mathbf{Z}_2|\mathbf{U}_1)^{-1}-\mathbf{\Sigma}_2 = \mathbf{A}-\mathbf{\Sigma}_2 \preceq \mathbf{S}_1 \\ &\preceq \mathbf{A}+\mathbf{\Delta}-\mathbf{\Sigma}_2 = \mathbf{J}(\mathbf{X}+\mathbf{Z}_1|\mathbf{U}_1)^{-1}-\mathbf{\Sigma}_1 \preceq \mathbf{S}, \end{aligned} \tag{4.62}$$

which implies

$$\mathbf{0} \preceq \mathbf{S}_1 \preceq \mathbf{S}. \tag{4.63}$$

We next study the term $h(\mathbf{Y}_2|\mathbf{U}_2)-h(\mathbf{Y}_1|\mathbf{U}_2)$. Due to the Markov chain condition (4.9), it is clear that $-I(U_2;Y_2|U_1)+I(U_2;Y_1|U_1)\leq 0$, which implies that

$$\begin{aligned} &h(\mathbf{Y}_2|\mathbf{U}_2)-h(\mathbf{Y}_1|\mathbf{U}_2) \\ &\quad \leq h(\mathbf{Y}_2|\mathbf{U}_1)-h(\mathbf{Y}_1|\mathbf{U}_1) \\ &\quad = \frac{1}{2}\log\frac{|\mathbf{S}_1+\mathbf{\Sigma}_2|}{|\mathbf{S}_1+\mathbf{\Sigma}_1|}. \end{aligned} \tag{4.64}$$

Applying Lemma 4.4, we obtain

$$\begin{aligned} &\frac{1}{2}\log\frac{|\mathbf{J}(\mathbf{X}+\mathbf{Z}_2|\mathbf{U}_2)^{-1}|}{|\mathbf{J}(\mathbf{X}+\mathbf{Z}_2|\mathbf{U}_2)^{-1}+\mathbf{\Sigma}_1-\mathbf{\Sigma}_2|} \\ &\quad \leq h(\mathbf{Y}_2|\mathbf{U}_2)-h(Y_1|\mathbf{U}_2) \\ &\quad\quad \leq \frac{1}{2}\log\frac{|\mathbf{J}(\mathbf{X}+\mathbf{Z}_2|\mathbf{U}_2)^{-1}+\mathbf{\Sigma}_2-\mathbf{\Sigma}_1|}{|\mathbf{J}(\mathbf{X}+\mathbf{Z}_1|\mathbf{U}_2)^{-1}|}. \end{aligned} \tag{4.65}$$

Combining (4.64) and(4.65), we have

$$\begin{aligned} &\frac{1}{2}\log\frac{|\mathbf{J}(\mathbf{X}+\mathbf{Z}_2|\mathbf{U}_2)^{-1}|}{|\mathbf{J}(\mathbf{X}+\mathbf{Z}_2|\mathbf{U}_2)^{-1}+\mathbf{\Sigma}_1-\mathbf{\Sigma}_2|} \\ &\quad \leq h(\mathbf{Y}_2|\mathbf{U}_2)-h(Y_1|\mathbf{U}_2) \leq \frac{1}{2}\log\frac{|\mathbf{S}_1+\mathbf{\Sigma}_2|}{|\mathbf{S}_1+\mathbf{\Sigma}_1|}. \end{aligned} \tag{4.66}$$

We now consider the function

$$r(t) = \frac{1}{2}\log\frac{|\mathbf{A}+\mathbf{B}+t\mathbf{\Delta}|}{|\mathbf{A}+t\mathbf{\Delta}|} \tag{4.67}$$

with $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{\Delta}$ being redefined as,

$$\begin{aligned}\mathbf{A} &= \mathbf{J}(\mathbf{X}+\mathbf{Z}_2|\mathbf{U}_2)^{-1}\\ \mathbf{B} &= \mathbf{\Sigma}_1 - \mathbf{\Sigma}_2\\ \mathbf{\Delta} &= \mathbf{S}_1 + \mathbf{\Sigma}_2 - \mathbf{J}(\mathbf{X}+\mathbf{Z}_2|\mathbf{U}_2)^{-1},\end{aligned} \tag{4.68}$$

where $\mathbf{A} \succ 0$, $\mathbf{B} \succeq 0$, and $\mathbf{\Delta} \succeq 0$. In order to show $\mathbf{\Delta} \succeq \mathbf{0}$, we show that $\mathbf{S}_1 \succeq \mathbf{J}(\mathbf{X}+\mathbf{Z}_2|\mathbf{U}_2)^{-1} - \mathbf{\Sigma}_2$. Using Lemma 4.5, we have

$$\mathbf{J}(\mathbf{X}+\mathbf{Z}_2|\mathbf{U}_2) \succeq \mathbf{J}(\mathbf{X}+\mathbf{Z}_2|\mathbf{U}_1). \tag{4.69}$$

Hence,

$$\mathbf{J}(\mathbf{X}+\mathbf{Z}_2|\mathbf{U}_1)^{-1} \succeq \mathbf{J}(\mathbf{X}+\mathbf{Z}_2|\mathbf{U}_2)^{-1}. \tag{4.70}$$

Since $\mathbf{S}_1 \succeq \mathbf{J}(\mathbf{X}+\mathbf{Z}_2|\mathbf{U}_1)^{-1} - \mathbf{\Sigma}_2$, we have

$$\mathbf{S}_1 \succeq \mathbf{J}(\mathbf{X}+\mathbf{Z}_2|\mathbf{U}_2)^{-1} - \mathbf{\Sigma}_2. \tag{4.71}$$

Thus, (4.66) can be rewritten as

$$-r(0) \leq h(\mathbf{Y}_2|\mathbf{U}_2) - h(\mathbf{Y}_1|\mathbf{U}_2) \leq -r(1). \tag{4.72}$$

Since the function $r(t)$ is monotone and continuous, there exists a constant t_2 with $0 \leq t_2 \leq 1$

such that $h(\mathbf{Y}_2|\mathbf{U}_2) - h(\mathbf{Y}_1|\mathbf{U}_2) = -r(t_2)$. Let $\mathbf{S}_2 = \mathbf{A} + t_2\mathbf{\Delta} - \mathbf{\Sigma}_2$, and obtain

$$h(\mathbf{Y}_2|\mathbf{U}_2) - h(\mathbf{Y}_1|\mathbf{U}_2) = -r(t_2) = \frac{1}{2}\log\frac{|\mathbf{S}_2 + \mathbf{\Sigma}_2|}{|\mathbf{S}_2 + \mathbf{\Sigma}_1|} \tag{4.73}$$

It can be seen that

$$\mathbf{J}(\mathbf{X} + \mathbf{Z}_2|\mathbf{U}_2)^{-1} - \mathbf{\Sigma}_2 = \mathbf{A} - \mathbf{\Sigma}_2 \preceq \mathbf{S}_2 \preceq \mathbf{A} + \mathbf{\Delta} - \mathbf{\Sigma}_2 = \mathbf{S}_1. \tag{4.74}$$

Therefore, combining (4.61) and (4.73), we obtain

$$\begin{aligned} R_2 &\leq \frac{1}{2}\log\frac{|\mathbf{S}_1 + \mathbf{\Sigma}_2|}{|\mathbf{S}_1 + \mathbf{\Sigma}_1|} - \frac{1}{2}\log\frac{|\mathbf{S}_2 + \mathbf{\Sigma}_2|}{|\mathbf{S}_2 + \mathbf{\Sigma}_1|} \\ &= \frac{1}{2}\log\frac{|\mathbf{S}_1 + \mathbf{\Sigma}_2|}{|\mathbf{S}_2 + \mathbf{\Sigma}_2|} - \frac{1}{2}\log\frac{|\mathbf{S}_1 + \mathbf{\Sigma}_1|}{|\mathbf{S}_2 + \mathbf{\Sigma}_1|}. \end{aligned} \tag{4.75}$$

We next derive an upper bound on R_3, which is a necessary step to show that the proof techniques can be iteratively extended to bound $R_4, \ldots, R_K$. Following from (4.54), we have

$$R_3 \leq h(\mathbf{Y}_3|\mathbf{U}_2) - h(\mathbf{Y}_2|\mathbf{U}_2) - (h(\mathbf{Y}_3|\mathbf{U}_3) - h(\mathbf{Y}_2|\mathbf{U}_3)). \tag{4.76}$$

Using Lemma 4.3 and (4.73), we obtain

$$h(\mathbf{Y}_3|\mathbf{U}_2) - h(\mathbf{Y}_2|\mathbf{U}_2) \leq \frac{1}{2}\log\frac{|\mathbf{S}_2 + \mathbf{\Sigma}_3|}{|\mathbf{S}_2 + \mathbf{\Sigma}_2|}. \tag{4.77}$$

Similarly to (4.64), due to the Markov chain condition (4.9), we have

$$h(\mathbf{Y}_3|\mathbf{U}_3) - h(\mathbf{Y}_2|\mathbf{U}_3) \leq h(\mathbf{Y}_3|\mathbf{U}_2) - h(\mathbf{Y}_2|\mathbf{U}_2). \tag{4.78}$$

Using Lemma 4.4 and (4.77) and (4.78), we have,

$$\frac{1}{2}\log\frac{|\mathbf{J}(\mathbf{X}+\mathbf{Z}_3|\mathbf{U}_3)^{-1}|}{|\mathbf{J}(\mathbf{X}+\mathbf{Z}_3|\mathbf{U}_3)^{-1}+\mathbf{\Sigma}_2-\mathbf{\Sigma}_3|} \leq h(\mathbf{Y}_3|\mathbf{U}_3)-h(\mathbf{Y}_2|\mathbf{U}_3) \leq \frac{1}{2}\log\frac{|\mathbf{S}_2+\mathbf{\Sigma}_3|}{|\mathbf{S}_2+\mathbf{\Sigma}_2|}. \tag{4.79}$$

It can be shown that $\mathbf{S}_2 \succeq \mathbf{J}(\mathbf{X}+\mathbf{Z}_3|\mathbf{U}_3)^{-1}-\mathbf{\Sigma}_3$ by using Lemma 4.5 and Lemma 4.8. Then following the similar arguments that yield (4.73), we can show that there exists an $\mathbf{S}_3$, such that $\mathbf{0} \preceq \mathbf{S}_3 \preceq \mathbf{S}_2 \preceq \mathbf{S}_1 \preceq \mathbf{S}$ and

$$h(\mathbf{Y}_3|\mathbf{U}_3)-h(\mathbf{Y}_2|\mathbf{U}_3)=\frac{1}{2}\log\frac{|\mathbf{S}_3+\mathbf{\Sigma}_3|}{|\mathbf{S}_3+\mathbf{\Sigma}_2|}. \tag{4.80}$$

Therefore, substituting (4.77) and (4.80) into (4.76), we obtain

$$\begin{aligned} R_3 &\leq \frac{1}{2}\log\frac{|\mathbf{S}_2+\mathbf{\Sigma}_3|}{|\mathbf{S}_2+\mathbf{\Sigma}_2|}-\frac{1}{2}\log\frac{|\mathbf{S}_3+\mathbf{\Sigma}_3|}{|\mathbf{S}_3+\mathbf{\Sigma}_2|} \\ &= \frac{1}{2}\log\frac{|\mathbf{S}_2+\mathbf{\Sigma}_3|}{|\mathbf{S}_3+\mathbf{\Sigma}_3|}-\frac{1}{2}\log\frac{|\mathbf{S}_2+\mathbf{\Sigma}_2|}{|\mathbf{S}_3+\mathbf{\Sigma}_2|}. \end{aligned} \tag{4.81}$$

Using techniques similar to those for bounding R_2 and R_3, we can derive the desired bounds on $R_4, \ldots, R_K$ iteratively.

Finally, we bound the rate R_1. We introduce a virtual receiver $\mathbf{Y}_0 = \mathbf{X}+\mathbf{Z}_0$, where $\mathbf{Z}_0$ is a Gaussian vector with the covariance matrix of $\mathbf{\Sigma}_0 = t\mathbf{\Sigma}_1$ with $t \geq 1$. Hence, $\mathbf{\Sigma}_0 \succeq \mathbf{\Sigma}_1$. Following from (4.61) and Lemma 4.3, we have,

$$h(\mathbf{Y}_0|\mathbf{U}_1)-h(\mathbf{Y}_1|U_1) \leq \frac{1}{2}\log\frac{|\mathbf{S}_1+\mathbf{\Sigma}_0|}{|\mathbf{S}_1+\mathbf{\Sigma}_1|}, \tag{4.82}$$

for any $t \geq 1$. On the other hand, we have

$$\frac{1}{2}\log(2\pi e)^r|\mathbf{\Sigma}_0| = h(\mathbf{Z}_0) \leq h(\mathbf{Y}_0|\mathbf{U}_1) \leq h(\mathbf{Y}_0) \leq \frac{1}{2}\log(2\pi e)^r|\mathbf{S}+\mathbf{\Sigma}_0|, \tag{4.83}$$

which implies that

$$\frac{1}{2}\log\frac{|\mathbf{\Sigma}_0|}{|\mathbf{S}_1+\mathbf{\Sigma}_0|} \leq h(\mathbf{Y}_0|\mathbf{U}_1) - \frac{1}{2}\log(2\pi e)^r|\mathbf{S}_1+\mathbf{\Sigma}_0| \leq \frac{1}{2}\log\frac{|\mathbf{S}+\mathbf{\Sigma}_0|}{|\mathbf{S}_1+\mathbf{\Sigma}_0|}. \tag{4.84}$$

As $t\to\infty$, $\frac{1}{2}\log\frac{|\mathbf{\Sigma}_0|}{|\mathbf{S}_1+\mathbf{\Sigma}_0|}\to 0$ and $\frac{1}{2}\log\frac{|\mathbf{S}+\mathbf{\Sigma}_0|}{|\mathbf{S}_1+\mathbf{\Sigma}_0|}\to 0$. Hence, $h(\mathbf{Y}_0|\mathbf{U}_1)-\frac{1}{2}\log(2\pi e)^r|\mathbf{S}_1+\mathbf{\Sigma}_0|\to 0$ as $t\to\infty$. Since (4.82) holds for any $t\geq 1$, we have $h(\mathbf{Y}_1|\mathbf{U}_1)\geq\frac{1}{2}\log(2\pi e)^r|\mathbf{S}_1+\mathbf{\Sigma}_1|$.

Following from (4.54),

$$\begin{aligned} R_1 &\leq I(\mathbf{U}_1;\mathbf{Y}_1) \\ &= h(\mathbf{Y}_1) - h(\mathbf{Y}_1|\mathbf{U}_1) \\ &\leq \frac{1}{2}\log(2\pi e)^r|\mathbf{S}+\mathbf{\Sigma}_1| - \frac{1}{2}\log(2\pi e)^r|\mathbf{S}_1+\mathbf{\Sigma}_1| \\ &= \frac{1}{2}\log\frac{|\mathbf{S}+\mathbf{\Sigma}_1|}{|\mathbf{S}_1+\mathbf{\Sigma}_1|}, \end{aligned} \tag{4.85}$$

which completes the proof.

CHAPTER 5
DEGRADED BROADCAST CHANNEL WITH SECRECY OUTSIDE A BOUNDED RANGE

In this chapter, we study the problem of degraded broadcast channel with secrecy outside a bounded range. We focus on the four-receiver degraded broadcast channel with secrecy outside a bounded range, in which each message should be secured from the receivers with two level worse channel quality. In Section 5.1, we introduce our problem model. In Section 5.2, we present our main results on the secrecy capacity region.

5.1 Channel Model

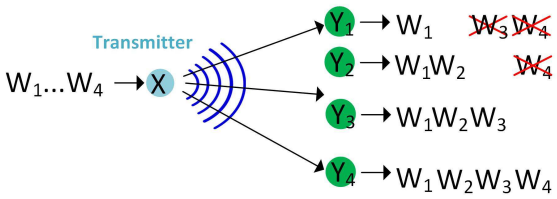

Fig. 5.1: The four-receiver degraded broadcast channel with secrecy outside a bounded range.

In this chapter, we consider a four-receiver degraded broadcast channel with secrecy outside of a bounded range (see Figure 5.1). Here, a transmitter sends information to four receivers over

a discrete memoryless channel with the channel transition probability given by $P_{Y_1Y_2Y_3Y_4|X}$, in which $X \in \mathcal{X}$ denotes the channel input, and $Y_k \in \mathcal{Y}_k$ denotes the channel output at receiver k, for $k = 1, 2, 3, 4$. The channel is assumed to satisfy the degraded condition, i.e., the following Markov chain holds:

$$X \to Y_4 \to Y_3 \to Y_2 \to Y_1. \tag{5.1}$$

Hence, the channel quality gradually degrades from receiver 4 to receiver 1.

The transmitter has four messages W_1, W_2, W_3, W_4 intended for the four receivers with the following decoding and secrecy requirements. For $k = 1, 2, 3, 4$, receiver k is required to decode the messages $W_1, \ldots, W_k$. Furthermore, the message W_3 needs to be kept secure from receiver 1, and the message W_4 needs to be kept secure from receivers 1 and 2 (see Figure 5.1).

A $(2^{nR_1}, 2^{nR_2}, 2^{nR_3}, 2^{nR_4}, n)$ code for the channel consists of

- Four message sets: $W_k \in \mathcal{W}_k = \{1, \cdots, 2^{nR_k}\}$ for $k = 1, 2, 3, 4$, which are independent from each other and each message is uniformly distributed over the corresponding message set;
- A (possibly stochastic) encoder f^n: $\mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{W}_3 \times \mathcal{W}_4 \to \mathcal{X}^n$;
- Four decoders $g_k^n : \mathcal{Y}_k^n \to (\mathcal{W}_1, \cdots, \mathcal{W}_k)$ for $k = 1, 2, 3, 4$.

A secrecy rate tuple (R_1, R_2, R_3, R_4) is *achievable* if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, 2^{nR_3}, 2^{nR_4}, n)$ code such that both the average error probability

$$P_e^n = \Pr\left(\cup_{k=1}^4 \{(W_1, \cdots, W_k) \neq g_k^n(Y_k^n)\}\right) \tag{5.2}$$

and the leakage rate at receivers 1 and 2

$$\frac{1}{n} I(W_3, W_4; Y_1^n | W_1) \quad \text{and} \quad \frac{1}{n} I(W_4; Y_2^n | W_1, W_2) \tag{5.3}$$

go to zero as n goes to infinity.

Our goal is to characterize the *secrecy capacity region* that contains all achievable rate tuples.

5.2 Characterization of Secrecy Capacity Region

Our main result in this chapter is the following characterization of the secrecy capacity region for the model of interest.

Theorem 5.1. *Consider the four-receiver degraded broadcast channel with secrecy outside a bounded range as described in Section 5.1. The secrecy capacity region consists of rate tuples* (R_1, R_2, R_3, R_4) *satisfying*

$$\begin{aligned}
R_1 &\leq I(U_1; Y_1), \\
R_2 &\leq I(U_2; Y_2|U_1), \\
R_3 &\leq I(U_3; Y_3|U_2) + \min\Big(0, I(U_2; Y_2|U_1) - I(U_3; Y_1|U_1)\Big), \\
R_4 &\leq I(X; Y_4|U_3) + I(U_3; Y_3|U_2) - I(X; Y_2|U_2), \\
R_3 + R_4 &\leq I(U_3; Y_3|U_2) + I(X; Y_4|U_3) + \min\Big(0, I(U_2; Y_2|U_1) - I(X; Y_1|U_1)\Big),
\end{aligned} \tag{5.4}$$

for some $P_{U_1U_2U_3X}$ *such that the following Markov chain holds*

$$U_1 \to U_2 \to U_3 \to X \to Y_4 \to Y_3 \to Y_2 \to Y_1. \tag{5.5}$$

The major technical challenge for establishing the above secrecy capacity region lies in providing an achievable region good enough to enable the proof of converse. Here, we briefly introduce our idea of the achievable scheme, which highlights the technical novelty of our design. We provide more detailed proofs in Section 5.3 and Section 5.4.

Our achievable scheme includes the following ingredients:

1. Superposition coding: Due to the requirement of layered decoding, the messages are encoded using superposition coding with each layer corresponding to one message, i.e., layer k corresponds to W_k for $k = 1, 2, 3, 4$.

2. Joint embedded coding and binning: Since the messages do not need to be kept secure from its immediate downstream receiver, such a receiver's message can serve as a random source for

securing the higher layer message in addition to stochastic binning. In fact, if such random source is sufficient for securing the message, binning is not necessary. More specifically, W_3 serves as a random source to secure W_4 from receiver 2 jointly with random binning designed at layer 4 (if necessary). Similarly, W_2 at layer 2 serves as a random source to secure W_3 and W_4 from receiver 1 jointly with binning at layers 3 and 4 (if necessary).

3. Rate splitting and sharing: We split W_3 into two parts, i.e., W_{31} and W_{32}. Such splitting exploits the opportunity (see case 1 in the proof of achievability), that W_{31} is sufficient to secure both W_{32} and W_4 from receiver 2, and thus the rate of W_{32} can be counted towards the rate of either W_3 or W_4. In this way, the rate region may be enlarged.

We note that joint embedded coding and binning is necessary here to exploit the secrecy requirements only outside the bounded range (i.e., the secrecy is not imposed for the immediate downstream receiver). Thus, messages intended for receivers inside the bounded range can serve as random sources for secrecy purpose. Such a scheme cannot be used for the model in Chapter 4 where the secrecy is imposed for the immediate downstream receiver. We further note that the embedded coding here uses messages across superposition layers as random sources for secrecy, which is different from the original embedded coding [39] where the messages serving as random sources are at the same layers as the messages being protected.

In fact, using only the superposition and joint embedded coding and binning is shown to be optimal (i.e., achieve the secrecy capacity region) for the three-receiver model in [60]. However, for the four-receiver model, such an achievable scheme is not in a sufficiently good form for which the machinery of a converse proof is difficult to develop. The major novelty of our scheme lies in developing rating splitting and sharing, which helps to potentially enlarge the achievable region (at least enlarge the region for a given distribution of auxiliary random variables). Consequently, the proof of converse can be developed for such an achievable region, and thus the secrecy capacity region is established.

More specifically, without rate splitting and sharing, superposition and joint embedded coding

and binning yields an achievable region with rates satisfying

$$\begin{aligned}
R_1 &\le I(U_1;Y_1),\\
R_2 &\le I(U_2;Y_2;U_1),\\
R_3 &\le I(U_3;Y_3|U_2) + \min\Big(0, I(U_2;Y_2|U_1) - I(U_3;Y_1|U_1)\Big),\\
R_4 &\le I(X;Y_4|U_3),\\
R_4 &\le I(X;Y_4|U_3) + I(U_3;Y_3|U_2) - I(X;Y_2|U_2),\\
R_3 + R_4 &\le I(U_3;Y_3|U_2) + I(X;Y_4|U_3) + I(U_2;Y_2|U_1) - I(X;Y_1|U_1) \qquad (5.6)
\end{aligned}$$

It is very difficult to develop the converse proof for the bound $R_4 \le I(X;Y_4|U_3)$ in the above region. However, by using rate splitting and sharing, this bound is replaced by the bound $R_3+R_4 \le I(U_3;Y_3|U_2)+I(X;Y_4|U_3)$, and the resulting region (5.4) is larger than the above region (5.6) (for a given distribution of auxiliary random variables). Furthermore, the converse proof for the new bound on $R_3 + R_4$ in (5.4) can be derived, and thus establishes the region (5.4) as the secrecy capacity region.

5.3 Achievability Proof of Theorem 5.1

Fix a distribution $P_{U_1}P_{U_2|U_1}P_{U_3|U_2}P_{X|U_3}P_{Y_1,Y_2,Y_3,Y_4|X}$. We design the achievable schemes for two cases.

Case 1: $I(U_3;Y_3|U_2) > I(X;Y_2|U_2)$

Random codebook generation: Randomly generate the codebook as follows:

- Generate 2^{nR_1} independent and identically distributed (i.i.d.) u_1^n with distribution $\prod_{i=1}^n p(u_{1,i})$. Index these codewords as $u_1^n(w_1)$, $w_1 \in [1, 2^{nR_1}]$.
- For each $u_1^n(w_1)$, generate 2^{nR_2} i.i.d. u_2^n with distribution $\prod_{i=1}^n p(u_{2,i}|u_{1,i})$. Index these codewords as $u_2^n(w_1, w_2)$, $w_2 \in [1, 2^{nR_2}]$.
- For each $u_2^n(w_1, w_2)$, generate $2^{n\widetilde{R}_3}$ i.i.d. u_3^n with distribution $\prod_{i=1}^n p(u_{3,i}|u_{2,i})$. Partition these

codewords into $2^{nR_{31}}$ bins. We further partition each bin into $2^{nR_{32}}$ sub-bins. Hence, there are $2^{n(\widetilde{R}_3-R_{31}-R_{32})}$ u_3^n in each sub-bin. We use $w_{31} \in [1 : 2^{nR_{31}}]$ to denote the bin number, $w_{32} \in [1 : 2^{nR_{32}}]$ to denote the sub-bin number, and $l_3 \in [1 : 2^{n(\widetilde{R}_3-R_{31}-R_{32})}]$ to denote the index within the bin. Hence, each u_3^n is indexed by $(w_1, w_2, w_{31}, w_{32}, l_3)$.

- For each $u_3^n(w_1, w_2, w_{31}, w_{32}, l_3)$, generate $2^{n\widetilde{R}_4}$ i.i.d. x^n with distribution $\prod_{i=1}^n p(x_i|u_{3,i})$. Partition these codewords into $2^{n\overline{R}_4}$ bins. We use $w_4 \in [1 : 2^{n\overline{R}_4}]$ to denote the bin number, $l_4 \in [1 : 2^{n(\widetilde{R}_4-\overline{R}_4)}]$ to denote the index inside the sub-bin. Index those codewords as $x^n(w_1, w_2, w_{31}, w_{32}, l_3, w_4, l_4)$, $w_4 \in [1, 2^{n\widetilde{R}_4}]$.

The chosen codebook is revealed to both the transmitter and receivers.

Encoding: To send a message tuple $(w_1, w_2, w_{31}, w_{32}, w_4)$, the transmitter randomly and uniformly generates $l_3 \in [1 : 2^{n(\widetilde{R}_3-R_{31}-R_{32})}]$ and $l_4 \in [1 : 2^{n(\widetilde{R}_4-\overline{R}_4)}]$, and sends $x^n(w_1, w_2, w_{31}, w_{32}, l_3, w_4, l_4)$.

Decoding:

- Receiver 1 claims that $\widehat{w}_1$ is sent, if there exists a unique $\widehat{w}_1$ such that $\Big(u_1^n(\widehat{w}_1), y_1^n) \in T_\epsilon^n(P_{U_1Y_1}\Big)$. Otherwise, it declares an error.
- Receiver 2 claims that $(\widehat{w}_1, \widehat{w}_2)$ is sent, if there exists a unique pair $(\widehat{w}_1, \widehat{w}_2)$ such that $\Big(u_1^n(\widehat{w}_1)$, $u_2^n(\widehat{w}_1, \widehat{w}_2), y_2^n) \in T_\epsilon^n(P_{U_1U_2Y_2}\Big)$. Otherwise, it declares an error.
- Receiver 3 claims that $(\widehat{w}_1, \widehat{w}_2, \widehat{w}_{31}, \widehat{w}_{32})$ is sent, if there exists a unique tuple $(\widehat{w}_1, \widehat{w}_2, \widehat{w}_{31}, \widehat{w}_{32}$, $\widehat{l_3})$ such that $\Big(u_1^n(\widehat{w}_1), u_2^n(\widehat{w}_1, \widehat{w}_2), u_3^n(\widehat{w}_1, \widehat{w}_2, \widehat{w}_{31}, \widehat{w}_{32}, \widehat{l_3}), y_3^n) \in T_\epsilon^n(P_{U_1U_2U_3Y_3}\Big)$. Otherwise, it declares an error.
- Receiver 4 claims that $\widehat{w}_1, \widehat{w}_2, \widehat{w}_{31}, \widehat{w}_{32}, \widehat{w}_4$ is sent, if there exists a unique tuple $(\widehat{w}_1, \widehat{w}_2, \widehat{w}_{31}, \widehat{w}_{32}$, $\hat{l}_3, \widehat{w}_4, \widehat{l_4})$ such that $\Big(u_1^n(\widehat{w}_1), u_2^n(\widehat{w}_1, \widehat{w}_2), u_3^n(\widehat{w}_1, \widehat{w}_2, \widehat{w}_{31}, \widehat{w}_{32}, \widehat{l_3}), x^n(\widehat{w}_1, \widehat{w}_2, \widehat{w}_{31}, \widehat{w}_{32}, \widehat{l_3}, \widehat{w}_4, \widehat{l_4})$, $y_4^n) \in T_\epsilon^n(P_{U_1U_2U_3XY_4}\Big)$. Otherwise, it declares an error.

Analysis of error probability: It can be shown by the law of large number and packing lemma that receiver k decodes the messages $(w_1, \ldots, w_k)$ with asymptotically small probability of error

for $k = 1, \ldots, 4$ if the following inequalities are satisfied.

$$\begin{aligned} R_1 &\leq I(U_1; Y_1), \\ R_2 &\leq I(U_2; Y_2|U_1), \\ \widetilde{R}_3 &\leq I(U_3; Y_3|U_2), \\ \widetilde{R}_4 &\leq I(X; Y_4|U_3). \end{aligned} \tag{5.7}$$

Analysis of leakage rate: In this model, W_{31}, W_{32}, W_4 are required to be kept secure from receiver 1, and W_4 is required to be kept secure from receiver 2. We note that under the assumption of case 1, i.e., $I(U_3; Y_3|U_2) > I(X; Y_2|U_2)$, part of W_3 (i.e., W_{31}) is sufficient to secure the remaining part of W_3 (i.e., W_{32}) and W_4 from receiver 2 without the necessity of random binning in layer 4[1]. Thus, we strengthen the secrecy requirements as follows: W_{31}, W_{32}, W_4 are kept secure from receiver 1, and W_{32}, W_4 are kept secure from receiver 2. Therefore, it is sufficient to show

$$\frac{1}{n} I(W_{31}, W_{32}, W_4; Y_1^n|W_1, \mathcal{C}) \to 0, \tag{5.8}$$

$$\frac{1}{n} I(W_{32}, W_4; Y_2^n|W_1, W_2, \mathcal{C}) \to 0, \tag{5.9}$$

as $n \to \infty$.

We first bound (5.8) which is the leakage rate of W_{31}, W_{32}, W_4 at receiver Y_1 as follows:

$$\begin{aligned} &I(W_{31}, W_{32}, W_4; Y_1^n|W_1, \mathcal{C}) \\ &= I(W_1, W_2, W_{31}, W_{32}, L_3, W_4, L_4; Y_1^n|\mathcal{C}) - I(W_1, W_2, L_3, L_4; Y_1^n|W_{31}, W_{32}, W_4, \mathcal{C}) \\ &\quad - H(W_{31}, W_{32}, W_4|Y_1^n, W_1, \mathcal{C}) + H(W_{31}, W_{32}, W_4|Y_1^n, \mathcal{C}) \end{aligned}$$

[1]This is only true for securing W_4 from receiver 2. Random binning may still be needed for securing W_3 and W_4 from receiver 1.

$$\begin{aligned}
&\overset{(a)}{\leq} I(W_1, W_2, W_{31}, W_{32}, L_3, W_4, L_4; Y_1^n|\mathcal{C}) - I(W_1, W_2, L_3, L_4; Y_1^n|W_{31}, W_{32}, W_4, \mathcal{C}) + n\epsilon_n \\
&\overset{(b)}{\leq} I(X^n; Y_1^n|\mathcal{C}) - H(W_1, W_2, L_3, L_4|W_{31}, W_{32}, W_4, \mathcal{C}) \\
&+ H(W_1, W_2, L_3, L_4|Y_1^n, W_{31}, W_{32}, W_4, \mathcal{C}), \qquad (5.10)
\end{aligned}$$

where (a) is due to Fano's inequality, (b) follows from the following Markov chain:

$$(W_1, W_2, W_{31}, W_{32}, L_3, W_4, L_4) \to X^n \to Y_1^n.$$

Next, we bound the three terms on the right hand side of (5.10) one by one. The first term is bounded as follows,

$$\begin{aligned}
&I(X^n; Y_1^n|\mathcal{C}) \\
&\overset{(a)}{=} I(U_1^n, X^n; Y_1^n|\mathcal{C}) \\
&= I(U_1^n; Y_1^n|\mathcal{C}) + I(X^n; Y_1^n|U_1^n, \mathcal{C}) \\
&\leq H(U_1^n|\mathcal{C}) + I(X^n; Y_1^n|U_1^n, \mathcal{C})
\end{aligned}$$

$$\begin{aligned}
&= nR_1 + H(Y_1^n|U_1^n, \mathcal{C}) - H(Y_1^n|U_1^n, X^n, \mathcal{C}) \\
&= nR_1 + \sum_{i=1}^{n} H(Y_{1,i}|U_1^n, Y_1^{i-1}, \mathcal{C}) - \sum_{i=1}^{n} H(Y_{1,i}|U_1^n, X^n, Y_1^{i-1}, \mathcal{C}) \\
&\overset{(b)}{\leq} nR_1 + \sum_{i=1}^{n} H(Y_{1,i}|U_{1,i}) - \sum_{i=1}^{n} H(Y_{1,i}|X_i) \\
&= nR_1 + nH(Y_1|U_1) - nH(Y_1|X) \\
&= nR_1 + nI(X; Y_1|U_1), \qquad (5.11)
\end{aligned}$$

where (a) is due to the Markov chain $U_1^n \to X^n \to Y_1^n$, (b) is due to the fact that $H(Y_{1,i}|U_1^n, Y_1^{i-1}, \mathcal{C})$ $\leq H(Y_{1,i}|U_{1,i})$ and the Markov chain $(U_1^n, X^{i-1}, X_{i+1}^n, Y_1^{i-1}, \mathcal{C}) \to X_i \to Y_{1,i}$.

We bound the second term as follows.

$$-H(W_1, W_2, L_3, L_4|W_{31}, W_{32}, W_4, \mathcal{C}) = -nR_1 - nR_2 - n(\widetilde{R}_3 - R_{31} - R_{32}) - n(\widetilde{R}_4 - \overline{R}_4). \tag{5.12}$$

Next, we bound the third term as follows.

$$\begin{aligned}
&H(W_1, W_2, L_3, L_4|Y_1^n, W_{31}, W_{32}, W_4, \mathcal{C}) \\
&= H(W_1|Y_1^n, W_{31}, W_{32}, W_4, \mathcal{C}) + H(W_2, L_3, L_4|Y_1^n, W_1, W_{31}, W_{32}, W_4, \mathcal{C}) \\
&\overset{(a)}{\leq} n\epsilon_n + H(W_2, L_3, L_4|Y_1^n, W_1, W_{31}, W_{32}, W_4, \mathcal{C}) \\
&\overset{(b)}{=} H(W_2, L_3, L_4|Y_1^n, U_1^n, W_{31}, W_{32}, W_4, \mathcal{C}) + n\epsilon_n \\
&\leq H(W_2, L_3, L_4|Y_1^n, U_1^n, W_{31}, W_{32}, W_4) + n\epsilon_n \\
&\overset{(c)}{\leq} nR_2 + n(\widetilde{R}_3 - R_{31} - R_{32}) + n(\widetilde{R}_4 - \overline{R}_4) - nI(X; Y_1|U_1) + 2n\epsilon_n,
\end{aligned} \tag{5.13}$$

where (a) is by Fano's inequality, (b) is due to the fact that U_1^n is a function of W_1 and $\mathcal{C}$, (c) is due to Lemma 5.1 if the conditions in (5.14) are satisfied.

Lemma 5.1. *If the following conditions are satisfied,*

$$\begin{aligned}
R_2 + \widetilde{R}_3 - R_{31} - R_{32} &\geq I(U_3; Y_1|U_1), \\
R_2 &\geq I(U_2; Y_1|U_1), \\
R_2 + \widetilde{R}_3 - R_{31} - R_{32} + \widetilde{R}_4 - \overline{R}_4 &\geq I(X; Y_1|U_1),
\end{aligned} \tag{5.14}$$

we have

$$\begin{aligned}
\frac{1}{n}H(W_2, L_3, L_4|Y_1^n, U_1^n, W_{31}, W_{32}, W_4) \leq &R_2 + (\widetilde{R}_3 - R_{31} - R_{32}) + (\widetilde{R}_4 - \overline{R}_4) \\
&- I(X; Y_1|U_1) + \epsilon_n.
\end{aligned} \tag{5.15}$$

Proof. See Section 5.5. □

Hence, combining (5.11), (5.12) and (5.13), we have as $n \to 0$,

$$\frac{1}{n} I(W_{31}, W_{32}, W_4; Y_1^n | W_1, \mathcal{C}) \to 0, \tag{5.16}$$

if the conditions (5.14) are satisfied.

Secondly, we bound the (5.9) as follows:

$$\begin{aligned}
&\frac{1}{n} I(W_{32}, W_4; Y_2^n | W_1, W_2, \mathcal{C}) \\
&= I(W_1, W_2, W_{31}, W_{32}, W_4, L_3, L_4; Y_2^n | \mathcal{C}) - I(W_1, W_2, W_{31}, L_3, L_4; Y_2^n | W_{32}, W_4, \mathcal{C}) \\
&\quad - H(W_{32}, W_4 | Y_2^n, W_1, W_2, \mathcal{C}) + H(W_{32}, W_4 | Y_2^n, \mathcal{C}) \\
&\leq I(X^n; Y_2^n | \mathcal{C}) - H(W_1, W_2, W_{31}, L_3, L_4 | W_{32}, W_4, \mathcal{C}) \\
&\quad + H(W_1, W_2, W_{31}, L_3, L_4 | Y_2^n, W_{32}, W_4, \mathcal{C}).
\end{aligned} \tag{5.17}$$

Following similar steps for bounding (5.8), we obtain the following bounds on the three terms on the right hand side of (5.17). The first term is bounded as follows,

$$\begin{aligned}
&I(X^n; Y_2^n | \mathcal{C}) \\
&\quad \leq nR_1 + nR_2 + nI(X; Y_2 | U_2).
\end{aligned} \tag{5.18}$$

The second term is bounded as follows,

$$\begin{aligned}
&- H(W_1, W_2, W_{31}, L_3, L_4 | W_{32}, W_4, \mathcal{C}) \\
&\quad = -nR_2 - nR_2 - n(\widetilde{R}_3 - R_{32}) - n(\widetilde{R}_4 - \overline{R}_4).
\end{aligned} \tag{5.19}$$

And the third term is bounded as follows:

$$\begin{aligned}
&H(W_1, W_2, W_{31}, L_3, L_4 | Y_2^n, W_{32}, W_4, \mathcal{C}) \\
&= H(W_1, W_2 | Y_2^n, W_{32}, W_4, \mathcal{C}) + H(W_{31}, L_3, L_4 | Y_2^n, W_1, W_2, W_{32}, W_4, \mathcal{C}) \\
&\overset{(a)}{\leq} H(W_{31}, L_3, L_4 | Y_2^n, W_1, W_2, W_{32}, W_4, \mathcal{C}) + n\epsilon_n \\
&\overset{(b)}{=} H(W_{31}, L_3, L_4 | Y_2^n, U_2^n, W_1, W_2, W_{32}, W_4, \mathcal{C}) + n\epsilon_n \\
&\leq H(W_{31}, L_3, L_4 | Y_2^n, U_2^n, W_{32}, W_4) + n\epsilon_n \\
&\overset{(c)}{\leq} n(\widetilde{R}_3 - R_{32} + \widetilde{R}_4 - \overline{R}_4 - I(X; Y_2 | U_2)) + 2n\epsilon_n,
\end{aligned} \tag{5.20}$$

where (a) is due to Fano's inequality, (b) is due to the fact that U_2^n is a function of $(W_1, W_2, \mathcal{C})$, (c) is by Lemma 5.2, if the conditions in (5.21) are satisfied.

Lemma 5.2. *If the following conditions are satisfied,*

$$\begin{aligned}
\widetilde{R}_3 - R_{32} &\geq I(U_3; Y_2 | U_2), \\
\widetilde{R}_3 - R_{32} + \widetilde{R}_4 - \overline{R}_4 &\geq I(U_4; Y_2 | U_2),
\end{aligned} \tag{5.21}$$

we have

$$\frac{1}{n} H(W_1, W_2, W_{31}, L_3, L_4 | Y_2^n, W_{32}, W_4, \mathcal{C}) \leq \widetilde{R}_3 - R_{32} + \widetilde{R}_4 - \overline{R}_4 - I(X; Y_2 | U_2) + \epsilon_n. \tag{5.22}$$

Proof. See Section 5.6. □

Hence, combining (5.18), (5.19) and (5.20), we have that as $n \to \infty$,

$$\frac{1}{n} I(W_{32}, W_4; Y_2^n | W_1, W_2, \mathcal{C}) \to 0, \tag{5.23}$$

if the conditions in (5.21) are satisfied.

By now, we have obtained constraints as in (5.7), (5.14) and (5.21), such that the decoding and

secrecy requirements are fulfilled.

Note that the lower bound on R_2 is not necessary because if a larger rate $R_2 = I(U_2; Y_2|U_1)$ can be achieved, any rate below can also be achieved by sending independent extra information. Furthermore, in order to maximize our achievable region, we generate $2^{nI(U_k;Y_k|U_{k-1})}$ of u_k^n for each u_{k-1}^n for $k = 2, 3, 4$, i.e., set $R_2 = I(U_2; Y_2|U_1)$, $\widetilde{R}_3 = I(U_3; Y_3|U_2)$ and $\widetilde{R}_4 = I(X; Y_4|U_3)$. Hence, we have the following achievable region:

$$
\begin{aligned}
R_1 &\leq I(U_1; Y_1), \\
R_2 &\leq I(U_2; Y_2|U_1), \\
R_{31} + R_{32} &\leq I(U_3; Y_3|U_2), \\
\overline{R}_4 &\leq I(X; Y_4|U_3), \\
R_{31} + R_{32} &\leq I(U_2; Y_2|U_1) + I(U_3; Y_3|U_2) - I(U_3; Y_1|U_1) \\
R_{31} + R_{32} + \overline{R}_4 &\leq I(U_2; Y_2|U_1) + I(U_3; Y_3|U_2) + I(X; Y_4|U_3) - I(X; Y_1|U_1), \\
R_{32} &\leq I(U_3; Y_3|U_2) - I(U_3; Y_2|U_2), \\
R_{32} + \overline{R}_4 &\leq I(U_3; Y_3|U_2) + I(X; Y_4|U_3) - I(X; Y_2|U_2).
\end{aligned} \tag{5.24}
$$

Rate sharing: It can be observed that W_{32} satisfies the same decoding and secrecy requirements as W_4, and hence its rate can be counted towards R_4 by subtracting the same rate from R_3. Thus, we define $R_3 = R_{31}$, and $R_4 = R_{32} + \overline{R}_4$. By adding these two rates to the above achievable region, and performing the Fourier-Motzkin elimination to remove R_{31}, R_{32}, and $\bar{R}_4$, we obtain the achievable region given in Theorem 1.

Case 2: $I(U_3; Y_3|U_2) \leq I(X; Y_2|U_2)$

Randomly generate the codebook as in case 1 but set $R_{32} = 0$, $R_{31} = R_3$ and $\overline{R}_4 = R_4$. The encoding and decoding procedures are similar to those of case 1. Following steps similar to those for case 1 to analyze the decoding error probability and the leakage rate, we obtain the achievable

region characterized by the following bounds:

$$\begin{aligned}
R_1 &\leq I(U_1;Y_1),\\
R_2 &\leq I(U_2;Y_2|U_1),\\
R_3 &\leq I(U_3;Y_3|U_2),\\
R_4 &\leq I(X;Y_4|U_3),\\
R_3 &\leq I(U_2;Y_2|U_1)+I(U_3;Y_3|U_2)-I(U_3;Y_1|U_1),\\
R_3+R_4 &\leq I(U_2;Y_2|U_1)+I(U_3;Y_3|U_2)+I(X;Y_4|U_3)-I(X;Y_1|U_1),\\
R_4 &\leq I(U_3;Y_3|U_2)+I(X;Y_4|U_3)-I(U_2;Y_2|U_2). \qquad (5.25)
\end{aligned}$$

It can be easily shown that under the assumption of case 2, which is $I(U_3;Y_3|U_2) \leq I(X;Y_2|U_2)$, the achievable region characterized by (5.25) is equivalent to the capacity region characterized in Theorem 1.

5.4 Converse Proof of Theorem 5.1

By Fano's inequality and the secrecy constraints, we have the following inequalities:

$$\begin{aligned}
H(W_k|Y_k^n) &\leq n\epsilon_n, \text{ for } 1 \leq k \leq 4,\\
I(W_3,W_4;Y_1^n|W_1) &\leq n\epsilon_n,\\
I(W_4;Y_2^n|W_1,W_2) &\leq n\epsilon_n. \qquad (5.26)
\end{aligned}$$

To prove the converse, a natural construction of auxiliary random variables is as follows:

$$
\begin{aligned}
U_{1,i} &= (W_1, Y_1^{i-1}), \\
U_{2,i} &= (W_1, W_2, Y_2^{i-1}), \\
U_{3,i} &= (W_1, W_2, W_3, Y_3^{i-1}, Y_{1,i+1}^n), \\
U_{4,i} &= (W_1, \ldots, W_4, Y_4^{i-1}, Y_{2,i+1}^n).
\end{aligned}
$$

It can be shown that the following Markov chain is satisfied:

$$
U_{1,i} \to U_{2,i} \to U_{3,i} \to U_{4,i} \to X_i \to Y_{4,i} \to Y_{3,i} \to Y_{2,i} \to Y_{1,i}, \tag{5.27}
$$

for $i = 1, \ldots, n$.

We show the first three bounds on R_1, R_2, R_3 from the decoding capability. The rate R_1 is bounded as follows:

$$
\begin{aligned}
nR_1 &= H(W_1) \\
&= I(W_1; Y_1^n) + H(W_1|Y_1^n) \\
&\leq I(W_1; Y_1^n) + n\epsilon_n \\
&= \sum_{i=1}^{n} I(W_1; Y_{1i}|Y_1^{i-1}) + n\epsilon_n \\
&\leq \sum_{i=1}^{n} I(W_1, Y_1^{i-1}; Y_{1i}) + n\epsilon_n \\
&= \sum_{i=1}^{n} I(U_{1,i}; Y_{1,i}) + n\epsilon_n.
\end{aligned} \tag{5.28}
$$

The rate R_2 is bounded as follows:

$$\begin{aligned}
nR_2 = H(W_2) = H(W_2|W_1) &= I(W_2; Y_2^n|W_1) + H(W_2|Y_2^n, W_1) \\
&\leq I(W_2; Y_2^n|W_1) + n\epsilon_n \\
&= \sum_{i=1}^{n} I(W_2; Y_{2,i}|W_1, Y_2^{i-1}) + n\epsilon_n \\
&= \sum_{i=1}^{n} I(W_2; Y_{2,i}|W_1, Y_1^{i-1}, Y_2^{i-1}) + n\epsilon_n \\
&\leq \sum_{i=1}^{n} I(W_2, Y_2^{i-1}; Y_{2,i}|W_1, Y_1^{i-1}) + n\epsilon_n \\
&= \sum_{i=1}^{n} I(U_{2,i}; Y_{2,i}|U_{1,i}) + n\epsilon_n. \qquad (5.29)
\end{aligned}$$

The rate R_3 is bounded as follows:

$$\begin{aligned}
nR_3 = H(W_3) &= H(W_3|W_1, W_2) \\
&= I(W_3; Y_3^n|W_1, W_2) + H(W_3|Y_3^n, W_1, W_2) \\
&\leq I(W_3; Y_3^n|W_1, W_2) + n\epsilon_n \\
&= \sum_{i=1}^{n} I(W_3; Y_{3,i}|W_1, W_2, Y_3^{i-1}) + n\epsilon_n \\
&= \sum_{i=1}^{n} I(W_3; Y_{3,i}|W_1, W_2, Y_2^{i-1}, Y_3^{i-1}) + n\epsilon_n \\
&\leq \sum_{i=1}^{n} I(W_3, Y_3^{i-1}, Y_{1,i+1}^n; Y_{3,i}|W_1, W_2, Y_2^{i-1}) + n\epsilon_n \\
&= \sum_{i=1}^{n} I(U_{3,i}; Y_{3,i}|U_{2,i}) + n\epsilon_n. \qquad (5.30)
\end{aligned}$$

Next, we show the bounds from the secrecy constraints. Since W_3 is secured from receiver 1, we

have the following bound:

$$
\begin{aligned}
nR_3 &= H(W_3) - H(W_2) + H(W_2) \\
&\leq H(W_3|W_1, W_2) + n\epsilon_n - H(W_3|Y_3^n, W_1, W_2) - H(W_2) + H(W_2) + n\epsilon_n - I(W_3; Y_1^n|W_1) \\
&= I(W_3; Y_3^n|W_1, W_2) - H(W_2) - H(W_3|W_1) + H(W_2) + H(W_3|Y_1^n, W_1) \\
&= I(W_3; Y_3^n|W_1, W_2) - H(W_2, W_3|W_1) + H(W_2) + H(W_2, W_3|Y_1^n, W_1) \\
&\quad - H(W_2|Y_1^n, W_1, W_3) + 2n\epsilon_n \\
&\overset{(a)}{\leq} I(W_3; Y_3^n|W_1, W_2) - I(W_2, W_3; Y_1^n|W_1) + H(W_2) + 2n\epsilon_n, \qquad (5.31)
\end{aligned}
$$

where (a) is due to the fact that the entropy $H(W_2|Y_1^n, W_1, W_3)$ is nonnegative. Here we note that discarding such an entropy term does not result in a looser bound. This is because if $H(W_2|Y_1^n, W_1, W_3)$ is not a vanishing term (which implies that Y_1 cannot decode W_2 given W_1 and W_3), then W_2 provides enough randomness for protecting W_3, and hence $R_3 \leq I(U_3; Y_3|U_2)$ (which is bounded by the decoding capability of receiver Y_3) should already provide a tighter bound on R_3.

We further bound the three terms in (5.31) one by one. The first term in (5.31) is bounded as follows:

$$
\begin{aligned}
I(W_3; Y_3^n|W_1, W_2) &= \sum_{i=1}^{n} I(W_3; Y_{3,i}|W_1, W_2, Y_3^{i-1}) \\
&= \sum_{i=1}^{n} I(W_3, Y_{1,i+1}^n; Y_{3,i}|W_1, W_2, Y_3^{i-1}) - I(Y_{1,i+1}^n; Y_{3,i}|W_1, W_2, W_3, Y_3^{i-1}) \\
&= \sum_{i=1}^{n} I(W_3, Y_{1,i+1}^n; Y_{3,i}|W_1, W_2, Y_3^{i-1}) - I(Y_3^{i-1}; Y_{1,i}|W_1, W_2, W_3, Y_{1,i+1}^n) \\
&\leq \sum_{i=1}^{n} I(W_3, Y_3^{i-1}, Y_{1,i+1}^n; Y_{3,i}|W_1, W_2, Y_2^{i-1}) - I(Y_3^{i-1}; Y_{1,i}|W_1, W_2, W_3, Y_{1,i+1}^n). \qquad (5.32)
\end{aligned}
$$

The second term in (5.31) is bounded as follows:

$$
\begin{aligned}
& - I(W_2, W_3; Y_1^n | W_1) \\
&= \sum_{i=1}^{n} -I(W_2, W_3; Y_{1,i} | W_1, Y_{1,i+1}^n) \\
&= \sum_{i=1}^{n} -I(W_2, W_3, Y_3^{i-1}; Y_{1,i} | W_1, Y_{1,i+1}^n) + I(Y_3^{i-1}; Y_{1,i} | W_1, W_2, W_3, Y_{1,i+1}^n) \\
&= \sum_{i=1}^{n} -I(Y_3^{i-1}; Y_{1,i} | W_1, Y_{1,i+1}^n) - I(W_2, W_3; Y_{1,i} | W_1, Y_{1,i+1}^n, Y_3^{i-1}) \\
&\quad + I(Y_3^{i-1}; Y_{1,i} | W_1, W_2, W_3, Y_{1,i+1}^n) \\
&= \sum_{i=1}^{n} -I(Y_3^{i-1}; Y_{1,i} | W_1, Y_{1,i+1}^n) - I(W_2, W_3, Y_3^{i-1}, Y_{1,i+1}^n; Y_{1,i} | W_1, Y_1^{i-1}) \\
&\quad + I(Y_3^{i-1}, Y_{1,i+1}^n; Y_{1,i} | W_1, Y_1^{i-1}) + I(Y_3^{i-1}; Y_{1,i} | W_1, W_2, W_3, Y_{1,i+1}^n) \\
&\overset{(a)}{=} \sum_{i=1}^{n} -I(W_2, W_3, Y_3^{i-1}, Y_{1,i+1}^n; Y_{1,i} | W_1, Y_1^{i-1}) + I(Y_3^{i-1}; Y_{1,i} | W_1, W_2, W_3, Y_{1,i+1}^n), \qquad (5.33)
\end{aligned}
$$

where (a) is due to the following fact:

$$
\begin{aligned}
& \sum_{i=1}^{n} [-I(Y_3^{i-1}; Y_{1,i} | W_1, Y_{1,i+1}^n) + I(Y_3^{i-1}, Y_{1,i+1}^n; Y_{1,i} | W_1, Y_1^{i-1})] \\
&= \sum_{i=1}^{n} [-I(Y_1^{i-1}; Y_{1,i} | W_1, Y_{1,i+1}^n) - I(Y_3^{i-1}; Y_{1,i} | W_1, Y_1^{i-1}, Y_{1,i+1}^n) \\
&\quad + I(Y_{1,i+1}^n; Y_{1,i} | W_1, Y_1^{i-1}) + I(Y_3^{i-1}; Y_{1,i} | W_1, Y_1^{i-1}, Y_{1,i+1}^n)] \\
&= 0 \, .
\end{aligned}
$$

Combining (5.32), (5.33) and (5.29), we have

$$
nR_3 \leq \sum_{i=1}^{n} I(U_{3,i}; Y_{3,i} | U_{2,i}) - I(U_{3,i}; Y_{1,i} | U_{1,i}) + I(U_{2,i}; Y_{2,i} | U_{1,i}) + 3n\epsilon_n. \qquad (5.34)
$$

Similarly, we derive another bound on R_4 as follows:

$$\begin{aligned}
nR_4 &= H(W_4) + H(W_3) - H(W_3) \\
&\overset{(a)}{\leq} I(W_4; Y_4^n|W_1, W_2, W_3) + I(W_3; Y_3^n|W_1, W_2) - H(W_3) - I(W_4; Y_2^n|W_1, W_2) + n\epsilon_n \\
&= I(W_4; Y_4^n|W_1, W_2, W_3) + I(W_3; Y_3^n|W_1, W_2) - I(W_3, W_4; Y_2^n|W_1, W_2) \\
&\quad - H(W_3; Y_2^n|W_1, W_2, W_4) + n\epsilon_n \\
&\overset{(b)}{\leq} I(W_4; Y_4^n|W_1, W_2, W_3) + I(W_3; Y_3^n|W_1, W_2) - I(W_3, W_4; Y_2^n|W_1, W_2) + n\epsilon_n, \quad (5.35)
\end{aligned}$$

where (a) is due to Fano's inequality and the secrecy constraint, (b) is due to the non-negativity of entropy.

Next,we bound the three terms on the right hand side of (5.35) one by one. We bound the first term as follows,

$$\begin{aligned}
&I(W_4; Y_4^n|W_1, W_2, W_3) = \sum_{i=1}^{n} I(W_4; Y_{4,i}|W_1, W_2, W_3, Y_4^{i-1}) \\
&= \sum_{i=1}^{n} I(W_4, Y_{2,i+1}^n; Y_{4,i}|W_1, W_2, W_3, Y_4^{i-1}) - I(Y_{2,i+1}^n; Y_{4,i}|W_1, W_2, W_3, W_4, Y_4^{i-1}) \\
&= \sum_{i=1}^{n} I(Y_{1,i+1}^n; Y_{4,i}|W_1, W_2, W_3, Y_4^{i-1}) + I(W_4, Y_{2,i+1}^n; Y_{4,i}|W_1, W_2, W_3, Y_4^{i-1}, Y_{1,i+1}^n) \\
&\quad - I(Y_{2,i+1}^n; Y_{4,i}|W_1, W_2, W_3, W_4, Y_4^{i-1}) \\
&= \sum_{i=1}^{n} I(W_4, Y_4^{i-1}, Y_{2,i+1}^n; Y_{4,i}|W_1, W_2, W_3, Y_3^{i-1}, Y_{1,i+1}^n) \\
&\quad - I(Y_4^{i-1}; Y_{4,i}|W_1, W_2, W_3, Y_3^{i-1}, Y_{1,i+1}^n) \\
&\quad + I(Y_{1,i+1}^n; Y_{4,i}|W_1, W_2, W_3, Y_4^{i-1}) - I(Y_{2,i+1}^n; Y_{4,i}|W_1, W_2, W_3, W_4, Y_4^{i-1}). \quad (5.36)
\end{aligned}$$

We bound the second term as follows,

$$
\begin{aligned}
& I(W_3; Y_3^n | W_1, W_2) \\
&= \sum_{i=1}^{n} I(W_3; Y_{3,i} | W_1, W_2, Y_3^{i-1}) \\
&= \sum_{i=1}^{n} I(W_3, Y_{1,i+1}^n; Y_{3,i} | W_1, W_2, Y_3^{i-1}) - I(Y_{1,i+1}^n; Y_{3,i} | W_1, W_2, W_3, Y_3^{i-1}) \\
&= \sum_{i=1}^{n} I(W_3, Y_{1,i+1}^n, Y_3^{i-1}; Y_{3,i} | W_1, W_2, Y_2^{i-1}) - I(Y_3^{i-1}; Y_{3,i} | W_1, W_2, Y_2^{i-1}) \\
&\quad - I(Y_{1,i+1}^n; Y_{3,i} | W_1, W_2, W_3, Y_3^{i-1}).
\end{aligned} \tag{5.37}
$$

The third term is bounded as follows,

$$
\begin{aligned}
& - I(W_3, W_4; Y_2^n | W_1, W_2) \\
&= \sum_{i=1}^{n} -I(W_3, W_4; Y_{2,i} | W_1, W_2, Y_{2,i+1}^n) \\
&= \sum_{i=1}^{n} -I(W_3, W_4, Y_4^{i-1}; Y_{2,i} | W_1, W_2, Y_{2,i+1}^n) + I(Y_4^{i-1}; Y_{2,i} | W_1, W_2, W_3, W_4, Y_{2,i+1}^n) \\
&= \sum_{i=1}^{n} -I(Y_4^{i-1}; Y_{2,i} | W_1, W_2, Y_{2,i+1}^n) - I(W_3, W_4; Y_{2,i} | W_1, W_2, Y_4^{i_1}, Y_{2,i+1}^n) \\
&\quad + I(Y_4^{i-1}; Y_{2,i} | W_1, W_2, W_3, W_4, Y_{2,i+1}^n) \\
&= \sum_{i=1}^{n} -I(Y_4^{i-1}; Y_{2,i} | W_1, W_2, Y_{2,i+1}^n) - I(W_3, W_4, Y_4^{i-1}, Y_{2,i+1}^n; Y_{2,i} | W_1, W_2, Y_2^{i-1}) \\
&\quad + I(Y_4^{i-1}, Y_{2,i+1}^n; Y_{2,i} | W_1, W_2, Y_2^{i-1}) + I(Y_4^{i-1}; Y_{2,i} | W_1, W_2, W_3, W_4, Y_{2,i+1}^n).
\end{aligned} \tag{5.38}
$$

Combining (5.36), (5.37) and (5.38), and apply Csiszár's sum identity, we have R_4 is bounded

as follows,

$$\begin{aligned}
nR_4 \leq & \sum_{i=1}^{n} I(W_4, Y_4^{i-1}, Y_{2,i+1}^n; Y_{4,i}|W_1, W_2, W_3, Y_3^{i-1}, Y_{1,i+1}^n) \\
& + I(W_3, Y_{1,i+1}^n, Y_3^{i-1}; Y_{3,i}|W_1, W_2, Y_2^{i-1}) \\
& - I(W_3, W_4, Y_4^{i-1}, Y_{2,i+1}^n; Y_{2,i}|W_1, W_2, Y_2^{i-1}) + n\epsilon_n \\
= & \sum_{i=1}^{n} I(U_{4,i}; Y_{4,i}|U_{3,i}) + I(U_{3,i}; Y_{3,i}|U_{2,i}) - I(U_{4,i}; Y_{2,i}|U_{2,i}) + n\epsilon_n \\
\leq & \sum_{i=1}^{n} I(X_i; Y_{4,i}|U_{3,i}) + I(U_{3,i}; Y_{3,i}|U_{2,i}) - I(X_i; Y_{2,i}|U_{2,i}) + n\epsilon_n. \qquad (5.39)
\end{aligned}$$

We next show the sum rate bound of $R_3 + R_4$ as follows,

$$\begin{aligned}
& n(R_3 + R_4) \\
& = H(W_3, W_4) \\
& = H(W_3) + H(W_4) + H(W_2) - H(W_2) \\
& \overset{(a)}{\leq} I(W_2; Y_2^n|W_1) + I(W_3; Y_3^n|W_1, W_2) + I(W_4; Y_4^n|W_1, W_2, W_3) - I(W_2, W_3, W_4; Y_1^n|W_1) \\
& \quad - H(W_2|Y_1^n, W_1, W_3, W_4) + n\epsilon_n \\
& \overset{(b)}{\leq} I(W_2; Y_2^n|W_1) + I(W_3; Y_3^n|W_1, W_2) + I(W_4; Y_4^n|W_1, W_2, W_3) \\
& \quad - I(W_2, W_3, W_4; Y_1^n|W_1) + n\epsilon_n, \qquad (5.40)
\end{aligned}$$

where (a) is due to Fano's inequality and the secrecy constraint, (b) is due to the non-negativity of entropy. We first bound the first term following similar term for showing the bound $R_2 \leq \sum_{i=1}^{n} I(U_{2,i}; Y_{2,i}|U_{1,i})$, we have

$$I(W_2; Y_2^n|W_1) \leq \sum_{i=1}^{n} I(U_{2,i}; Y_{2,i}|U_{1,i}). \qquad (5.41)$$

Secondly, we bound $I(W_3; Y_3^n|W_1, W_2) + I(W_4; Y_4^n|W_1, W_2, W_3)$ as follows,

$$
\begin{aligned}
&I(W_3; Y_3^n|W_1, W_2) + I(W_4; Y_4^n|W_1, W_2, W_3) \\
&= \sum_{i=1}^{n} I(W_3; Y_{3,i}|W_1, W_2, Y_3^{i-1}) + I(W_3; Y_{4,i}|W_1, W_2, W_3, Y_4^{i-1}) \\
&= \sum_{i=1}^{n} I(W_3, Y_{1,i+1}^n; Y_{3,i}|W_1, W_2, Y_3^{i-1}) - I(Y_{1,i+1}^n; Y_{3,i}|W_1, W_2, W_3, Y_3^{i-1}) \\
&\quad + I(W_4, Y_{2,i+1}^n; Y_{4,i}|W_1, W_2, W_3, Y_4^{i-1}) - I(Y_{2,i+1}^n; Y_{4,i}|W_1, W_2, W_3, W_4, Y_4^{i-1}) \\
&= \sum_{i=1}^{n} I(W_3, Y_{1,i+1}^n, Y_3^{i-1}; Y_{3,i}|W_1, W_2, Y_2^{i-1}) - I(Y_3^{i-1}; Y_{3,i}|W_1, W_2, Y_2^{i-1}) \\
&\quad - I(Y_{1,i+1}^n; Y_{3,i}|W_1, W_2, W_3, Y_3^{i-1}) + I(W_4, Y_{2,i+1}^n; Y_{4,i}|W_1, W_2, W_3, Y_4^{i-1}, Y_{1,i+1}^n) \\
&\quad + I(Y_{1,i+1}^n; Y_{4,i}|W_1, W_2, W_3, Y_4^{i-1}) - I(Y_{2,i+1}^n; Y_{4,i}|W_1, W_2, W_3, W_4, Y_4^{i-1}) \\
&= \sum_{i=1}^{n} I(W_3, Y_{1,i+1}^n, Y_3^{i-1}; Y_{3,i}|W_1, W_2, Y_2^{i-1}) \\
&\quad + I(W_4, Y_{2,i+1}^n, Y_4^{i-1}; Y_{4,i}|W_1, W_2, W_3, Y_3^{i-1}, Y_{1,i+1}^n) \\
&\quad - I(Y_4^{i-1}; Y_{4,i}|W_1, W_2, W_3, Y_3^{i-1}, Y_{1,i+1}^n) - I(Y_{1,i+1}^n; Y_{3,i}|W_1, W_2, W_3, Y_3^{i-1}) \\
&\quad + I(Y_{1,i+1}^n; Y_{4,i}|W_1, W_2, W_3, Y_4^{i-1}) - I(Y_{2,i+1}^n; Y_{4,i}|W_1, W_2, W_3, W_4, Y_4^{i-1}) \\
&\quad - I(Y_3^{i-1}; Y_{3,i}|W_1, W_2, Y_2^{i-1}).
\end{aligned} \tag{5.42}
$$

Thirdly, we bound $-I(W_2, W_3, W_4; Y_1^n|W_1)$ as follows,

$$
\begin{aligned}
&- I(W_2, W_3, W_4; Y_1^n|W_1) \\
&= \sum_{i=1}^{n} -I(W_2, W_3, W_4; Y_{1,i}|W_1, Y_{1,i+1}^n) \\
&= \sum_{i=1}^{n} -I(W_2, W_3, W_4, Y_4^{i-1}; Y_{1,i}|W_1, Y_{1,i+1}^n) + I(Y_4^{i-1}; Y_{1,i}|W_1, W_2, W_3, W_4, Y_{1,i+1}^n) \\
&= \sum_{i=1}^{n} -I(Y_4^{i-1}; Y_{1,i}|W_1, Y_{1,i+1}^n) - I(W_2, W_3, W_4; Y_{1,i}|W_1, Y_{1,i+1}^n, Y_4^{i-1}) \\
&\quad + I(Y_4^{i-1}; Y_{1,i}|W_1, W_2, W_3, W_4, Y_{1,i+1}^n)
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} -I(W_2, W_3, W_4, Y_4^{i-1}, Y_{1,i+1}^n; Y_{1,i}|W_1, Y_1^{i-1}) \\
&\quad + I(Y_4^{i-1}, Y_{1,i+1}^n; Y_{1,i}|W_1, Y_1^{i-1}) - I(Y_4^{i-1}; Y_{1,i}|W_1, Y_{1,i+1}^n) \\
&\quad + I(Y_4^{i-1}; Y_{1,i}|W_1, W_2, W_3, W_4, Y_{1,i+1}^n) \\
&= \sum_{i=1}^{n} -I(W_2, W_3, W_4, Y_4^{i-1}, Y_{2,i+1}^n; Y_{1,i}|W_1, Y_1^{i-1}) \\
&\quad + I(Y_{2,i+1}^n; Y_{1,i}|W_1, W_2, W_3, W_4, Y_4^{i-1}, Y_{1,i+1}^n) + I(Y_4^{i-1}, Y_{1,i+1}^n; Y_{1,i}|W_1, Y_1^{i-1}) \\
&\quad - I(Y_4^{i-1}; Y_{1,i}|W_1, Y_{1,i+1}^n) + I(Y_4^{i-1}; Y_{1,i}|W_1, W_2, W_3, W_4, Y_{1,i+1}^n) \\
&\overset{(a)}{=} \sum_{i=1}^{n} -I(W_2, W_3, W_4, Y_4^{i-1}, Y_{2,i+1}^n; Y_{1,i}|W_1, Y_1^{i-1}) \\
&\quad + I(Y_{2,i+1}^n; Y_{1,i}|W_1, W_2, W_3, W_4, Y_4^{i-1}, Y_{1,i+1}^n) \\
&\quad + I(Y_4^{i-1}; Y_{1,i}|W_1, W_2, W_3, W_4, Y_{1,i+1}^n), \qquad (5.43)
\end{aligned}
$$

where (a) is due to the following fact

$$
\begin{aligned}
&I(Y_4^{i-1}, Y_{1,i+1}^n; Y_{1,i}|W_1, Y_1^{i-1}) - I(Y_4^{i-1}; Y_{1,i}|W_1, Y_{1,i+1}^n) \\
&= I(Y_{1,i+1}^n; Y_{1,i}|W_1, Y_1^{i-1}) + I(Y_4^{i-1}; Y_{1,i}|W_1, Y_1^{i-1}, Y_{1,i+1}^n) \\
&\quad - I(Y_1^{i-1}; Y_{1,i}|W_1, Y_{1,i+1}^n) - I(Y_4^{i-1}; Y_{1,i}|W_1, Y_{1,i+1}^n, Y_1^{i-1}) \\
&= 0, \qquad (5.44)
\end{aligned}
$$

by Csiszár's sum identity.

Hence, combining (5.41), (5.42) and (5.43), we have,

$$
\begin{aligned}
& n(R_3+R_4) \\
& \leq \sum_{i=1}^{n} I(U_{2,i};Y_{2,i}|U_{1,i}) + I(U_{3,i};Y_{3,i}|U_{2,i}) + I(U_{4,i};Y_{4,i}|U_{3,i}) - I(U_{4,i};Y_{1,i}|U_{1,i}) \\
& \quad - I(Y_{1,i+1}^n;Y_{3,i}|W_1,W_2,W_3,Y_3^{i-1}) + I(Y_{1,i+1}^n;Y_{4,i}|W_1,W_2,W_3,Y_4^{i-1}) \\
& \quad - I(Y_{2,i+1}^n;Y_{4,i}|W_1,W_2,W_3,W_4,Y_4^{i-1}) + I(Y_{2,i+1}^n;Y_{1,i}|W_1,W_2,W_3,W_4,Y_4^{i-1},Y_{1,i+1}^n) \\
& \quad + I(Y_4^{i-1};Y_{1,i}|W_1,W_2,W_3,W_4,Y_{1,i+1}^n) - I(Y_4^{i-1};Y_{4,i}|W_1,W_2,W_3,Y_3^{i-1},Y_{1,i+1}^n) \\
& \overset{(a)}{\leq} \sum_{i=1}^{n} I(U_{2,i};Y_{2,i}|U_{1,i}) + I(U_{3,i};Y_{3,i}|U_{2,i}) + I(U_{4,i};Y_{4,i}|U_{3,i}) - I(U_{4,i};Y_{1,i}|U_{1,i}) \\
& \leq \sum_{i=1}^{n} I(U_{2,i};Y_{2,i}|U_{1,i}) + I(U_{3,i};Y_{3,i}|U_{2,i}) + I(X_i;Y_{4,i}|U_{3,i}) - I(X_i;Y_{1,i}|U_{1,i}), \qquad (5.45)
\end{aligned}
$$

where (a) is due to the following two facts. The first fact is as follows:

$$
\begin{aligned}
& - I(Y_{1,i+1}^n;Y_{3,i}|W_1,W_2,W_3,Y_3^{i-1}) + I(Y_{1,i+1}^n;Y_{4,i}|W_1,W_2,W_3,Y_4^{i-1}) \\
& - I(Y_4^{i-1};Y_{4,i}|W_1,W_2,W_3,Y_3^{i-1},Y_{1,i+1}^n) \\
& \overset{(b)}{=} -I(Y_3^{i-1};Y_{1,i}|W_1,W_2,W_3,Y_{1,i+1}^n) + I(Y_4^{i-1};Y_{1,i}|W_1,W_2,W_3,Y_{1,i+1}^n) \\
& \quad - I(Y_4^{i-1};Y_{4,i}|W_1,W_2,W_3,Y_3^{i-1},Y_{1,i+1}^n) \\
& = -I(Y_4^{i-1};Y_{1,i}|W_1,W_2,W_3,Y_{1,i+1}^n,Y_3^{i-1}) - I(Y_4^{i-1};Y_{4,i}|W_1,W_2,W_3,Y_3^{i-1},Y_{1,i+1}^n) \\
& = -I(Y_4^{i-1};Y_{4,i}|W_1,W_2,W_3,Y_{1,i+1}^n,Y_3^{i-1},Y_{1,i}) \\
& \leq 0, \qquad (5.46)
\end{aligned}
$$

where step (b) is due to Csiszár's sum identity.

The second fact is as follows,

$$
\begin{aligned}
&- I(Y_{2,i+1}^n; Y_{4,i}|W_1, W_2, W_3, W_4, Y_4^{i-1}) + I(Y_{2,i+1}^n; Y_{1,i}|W_1, W_2, W_3, W_4, Y_4^{i-1}, Y_{1,i+1}^n) \\
&\quad + I(Y_4^{i-1}; Y_{1,i}|W_1, W_2, W_3, W_4, Y_{1,i+1}^n) \\
&\overset{(c)}{=} -I(Y_{1,i+1}^n; Y_{4,i}|W_1, W_2, W_3, W_4, Y_4^{i-1}) - I(Y_{2,i+1}^n; Y_{4,i}|W_1, W_2, W_3, W_4, Y_4^{i-1}, Y_{1,i+1}^n) \\
&\quad + I(Y_{2,i+1}^n; Y_{1,i}|W_1, W_2, W_3, W_4, Y_4^{i-1}, Y_{1,i+1}^n) + I(Y_{1,i+1}^n; Y_{4,i}|W_1, W_2, W_3, W_4, Y_4^{i-1}) \\
&= -I(Y_{2,i+1}^n; Y_{4,i}|W_1, W_2, W_3, W_4, Y_4^{i-1}, Y_{1,i+1}^n, Y_{1,i}) \\
&\leq 0, \qquad (5.47)
\end{aligned}
$$

where step (c) is by chain rule and Csiszár's sum identity.

Furthermore, we derive the second bound for $R_3 + R_4$ as follows.

$$
\begin{aligned}
&n(R_3 + R_4) = H(W_3, W_4) \leq I(W_3; Y_3^n|W_1, W_2) + I(W_4; Y_4^n|W_1, W_2, W_3) + 2n\epsilon_n \\
&= \sum_{i=1}^n I(W_3; Y_{3,i}|W_1, W_2, Y_3^{i-1}) + I(W_4; Y_{4,i}|W_1, W_2, W_3, Y_4^{i-1}) + 2n\epsilon_n \\
&= \sum_{i=1}^n I(W_3, Y_3^{i-1}, Y_{1,i+1}^n; Y_{3,i}|W_1, W_2, Y_2^{i-1}) \\
&\quad - I(Y_3^{i-1}; Y_{3,i}|W_1, W_2, Y_2^{i-1}) - I(Y_{1,i+1}^n; Y_{3,i}|W_1, W_2, W_3, Y_3^{i-1}) \\
&\quad + I(W_4, Y_4^{i-1}, Y_{2,i+1}^n; Y_{4,i}|W_1, W_2, W_3, Y_3^{i-1}, Y_{1,i+1}^n) + I(Y_{1,i+1}^n; Y_{4,i}|W_1, W_2, W_3, Y_4^{i-1}) \\
&\quad - I(Y_4^{i-1}; Y_{4,i}|W_1, W_2, W_3, Y_{1,i+1}^n, Y_3^{i-1}) - I(Y_{2,i+1}^n; Y_{4,i}|W_1, W_2, W_3, Y_4^{i-1}) + 2n\epsilon_n \\
&\overset{(a)}{\leq} \sum_{i=1}^n I(W_3, Y_3^{i-1}, Y_{1,i+1}^n; Y_{3,i}|W_1, W_2, Y_2^{i-1}) \\
&\quad + I(W_4, Y_4^{i-1}, Y_2^{i-1}; Y_{4,i}|W_1, W_2, W_3, Y_3^{i-1}, Y_{1,i+1}^n) \\
&= \sum_{i=1}^n I(U_{3,i}; Y_{3,i}|U_{2,i}) + I(U_{4,i}; Y_{4,i}|U_{3,i}) \\
&\leq \sum_{i=1}^n I(U_{3,i}; Y_{3,i}|U_{2,i}) + I(X_i; Y_{4,i}|U_{3,i}), \qquad (5.48)
\end{aligned}
$$

where (a) can be shown as follows. First of all, we have,

$$I(Y_{1,i+1}^n; Y_{3,i}|W_1, W_2, W_3, Y_3^{i-1}) = I(Y_3^{i-1}; Y_{1,i}|W_1, W_2, W_3, Y_{1,i+1}^n),$$

by Csiszár's sum identity. Further, by (5.49) and (5.50), and the nonnegativity of mutual information, we can show how step (a) is derived as follow,

$$\begin{aligned} &I(Y_3^{i-1}; Y_{1,i}|W_1, W_2, W_3, Y_{1,i+1}^n) + I(Y_{1,i+1}^n; Y_{4,i}|W_1, W_2, W_3, Y_4^{i-1}) \\ &= I(Y_4^{i-1}; Y_{1,i}|W_1, W_2, W_3, Y_3^{i-1}, Y_{1,i+1}^n) \end{aligned} \tag{5.49}$$

and

$$\begin{aligned} &I(Y_4^{i-1}; Y_{1,i}|W_1, W_2, W_3, Y_3^{i-1}, Y_{1,i+1}^n) - I(Y_4^{i-1}; Y_{4,i}|W_1, W_2, W_3, Y_3^{i-1}, Y_{1,i+1}^n) \\ &= -I(Y_4^{i-1}; Y_{4,i}|W_1, W_2, W_3, Y_3^{i-1}, Y_{1,i+1}^n, Y_{1,i}) \\ &\leq 0. \end{aligned} \tag{5.50}$$

Finally, we define a random variable $Q \in \{1, \ldots, n\}$ which is uniformly distributed, and set $U_k \triangleq (Q, U_{k,Q})$, $Y_k \triangleq (Q, Y_{k,Q})$, for $k = 1, 2, 3$, and $X \triangleq (Q, X_Q)$. Then we have the desired bounds by the standard single letter characterization, which concludes the proof.

5.5 Proof of Lemma 5.1

We bound $H(W_2, L_3, L_4|Y_1^n, U_1^n, w_{31}, w_{32}, w_4)$ for each (w_{31}, w_{32}, w_4). Hence, $H(W_2, L_3, L_4|Y_1^n, U_1^n, W_{31}, W_{32}, W_4)$ is bounded.

Firstly, we fix $(W_2, L_3, L_4) = (w_2, l_3, l_4)$ and a joint typical sequence $(u_1^n, y_1^n) \in T_\epsilon^{(n)}(U_1, Y_1)$. We define the following random variable:

$$\begin{aligned} N(w_2, l_3, l_4, u_1^n, y_1^n) := |\{(\widetilde{w}_2, \widetilde{l}_3, \widetilde{l}_4) \neq (w_2, l_3, l_4) : (x^n(\widetilde{w}_2, w_{31}, w_{32}, \widetilde{l}_3, w_4, \widetilde{l}_4), u_1^n, y_1^n) \\ \in T_\epsilon^{(n)}(X^n, U_1^n, Y_1^n)\}| \end{aligned} \tag{5.51}$$

We can show that the expectation of N satisfy the following inequalities:

$$
\begin{aligned}
&2^{n(R_2+\widetilde{R}_3-R_{31}-R_{32}+\widetilde{R}_4-\overline{R}_4-I(X;Y_1|U_1))-n\delta_n(\epsilon)-n\epsilon_n}+2^{n(\widetilde{R}_3-R_{31}-R_{32}+\widetilde{R}_4-R_4-I(X;Y_1|U_2))-n\delta_n(\epsilon)-n\epsilon_n}\\
&+2^{n(\widetilde{R}_4-\overline{R}_4-I(X;Y_1|U_3))-n\delta_n(\epsilon)-n\epsilon_n}\leq N(w_2,l_3,l_4,u_1^n,y_1^n)\\
&\leq 2^{n(R_2+\widetilde{R}_3-R_{31}-R_{32}+\widetilde{R}_4-\overline{R}_4-I(X;Y_1|U_1))+n\delta_n(\epsilon)-n\epsilon_n}+2^{n(\widetilde{R}_3-R_{31}-R_{32}+\widetilde{R}_4-R_4-I(X;Y_1|U_2))+n\delta_n(\epsilon)-n\epsilon_n}\\
&+2^{n(\widetilde{R}_4-\overline{R}_4-I(X;Y_1|U_3))+n\delta_n(\epsilon)-n\epsilon_n},
\end{aligned}
\tag{5.52}
$$

which can be further written as

$$
\begin{aligned}
&2^{n(R_2+\widetilde{R}_3-R_{31}-R_{32}+\widetilde{R}_4-\overline{R}_4-I(X;Y_1|U_1))-n\delta_n(\epsilon)-n\epsilon_n}\big(1+2^{-n(R_2-I(U_2;Y_1|U_1))}\\
&+2^{-n(R_2+\widetilde{R}_3-R_{31}-R_{32}-I(U_3;Y_1|U_1))}\big)\leq N(w_2,l_3,l_4,u_1^n,y_1^n)\\
&\leq 2^{n(R_2+\widetilde{R}_3-R_{31}-R_{32}+\widetilde{R}_4-\overline{R}_4-I(X;Y_1|U_1))+n\delta_n(\epsilon)-n\epsilon_n}\big(1+2^{-n(R_2-I(U_2;Y_1|U_1))}\\
&\quad+2^{-n(R_2+\widetilde{R}_3-R_{31}-R_{32}-I(U_3;Y_1|U_1))}\big).
\end{aligned}
\tag{5.53}
$$

Furthermore, if the following inequalities are satisfied,

$$
\begin{aligned}
R_2&\geq I(U_2;Y_1|U_1),\\
R_2+\widetilde{R}_3-R_3&\geq I(U_3;Y_1|U_1),
\end{aligned}
\tag{5.54}
$$

we can define $\big(1+2^{-n(R_2-I(U_2;Y_1|U_1))}+2^{-n(R_2+\widetilde{R}_3-R_{31}-R_{32}-I(U_3;Y_1|U_1))}\big)=2^{n\epsilon'_n}$, such that $\epsilon'_n\to 0$ as $n\to\infty$.

Then the expectation of $N(w_2,l_3,l_4,u_1^n,y_1^n)$ can be bounded as follows,

$$
\begin{aligned}
2^{n(R_2+\widetilde{R}_3-R_{31}-R_{32}+\widetilde{R}_4-\overline{R}_4-I(X;Y_1|U_1))-n\delta_n(\epsilon)-n\epsilon_n+n\epsilon'_n}&\leq N(w_2,l_3,l_4,u_1^n,y_1^n)\\
&\leq 2^{n(R_2+\widetilde{R}_3-R_{31}-R_{32}+\widetilde{R}_4-\overline{R}_4-I(X;Y_1|U_1))+n\delta_n(\epsilon)-n\epsilon_n+n\epsilon'_n}.
\end{aligned}
\tag{5.55}
$$

Due to the fact that N is binomial distributed, $Var(N) \leq E(N)$. Thus, we have,

$$Var(N(w_2, l_3, l_4, u_1^n, y_1^n)) \leq 2^{n(R_2+\widetilde{R}_3-R_{31}-R_{32}+\widetilde{R}_4-\overline{R}_4-I(X;Y_1|U_1))+n\delta_n(\epsilon)-n\epsilon_n+n\epsilon'_n}. \tag{5.56}$$

Next, we define the random event as follows:

$$\begin{aligned}\varepsilon(w_2, l_3, l_4, u_1^n, y_1^n) :=&\{N(w_2, l_3, l_4, u_1^n, y_1^n) \geq \\ &2^{n(R_2+\widetilde{R}_3-R_{31}-R_{32}+\widetilde{R}_4-\overline{R}_4-I(X;Y_1|U_1))+n\delta_n(\epsilon)-n\epsilon_n/2+n\epsilon'_n+1}\}.\end{aligned} \tag{5.57}$$

The probability of such a random event is bounded as follows:

$$\begin{aligned}&P(\varepsilon(w_2, l_3, l_4, u_1^n, y_1^n)) \\ &= P(N(w_2, l_3, l_4, u_1^n, y_1^n) \geq 2^{n(R_2+\widetilde{R}_3-R_{31}-R_{32}+\widetilde{R}_4-\overline{R}_4)-nI(X;Y_1|U_1)+n\delta_n(\epsilon)-n\epsilon_n/2+n\epsilon'_n+1}) \\ &\leq P(N(w_2, l_3, l_4, u_1^n, y_1^n) \geq E(N(w_2, l_3, l_4, u_1^n, y_1^n)) \\ &\quad + 2^{n(R_2+\widetilde{R}_3-R_{31}-R_{32}+\widetilde{R}_4-\overline{R}_4)-nI(X;Y_1|U_1)+n\delta_n(\epsilon)-n\epsilon_n/2+n\epsilon'_n}) \\ &\leq P(|N(w_2, l_3, l_4, u_1^n, y_1^n) - E(N(w_2, l_3, l_4, u_1^n, y_1^n))| \\ &\quad \geq 2^{n(R_2+\widetilde{R}_3-R_{31}-R_{32}+\widetilde{R}_4-\overline{R}_4)-nI(X;Y_1|U_1)+n\delta_n(\epsilon)-n\epsilon_n/2+n\epsilon'_n}) \\ &\leq \frac{Var(N(w_2, l_3, l_4, u_1^n, y_1^n))}{2^{2n(R_2+\widetilde{R}_3-R_{31}-R_{32}+\widetilde{R}_4-\overline{R}_4)-2nI(X;Y_1|U_1)+2n\delta_n(\epsilon)-n\epsilon_n+2n\epsilon'_n}} \\ &\leq \frac{1}{2^{n(R_2+\widetilde{R}_3-R_{31}-R_{32}+\widetilde{R}_4-\overline{R}_4)-nI(X;Y_1|U_1)+n\delta_n(\epsilon)+n\epsilon'_n}},\end{aligned} \tag{5.58}$$

which goes to zero as $n \to \infty$, if $R_2 + \widetilde{R}_3 - R_{31} - R_{32} + \widetilde{R}_4 - \overline{R}_4 \geq I(X; Y_1|U_1)$. Therefore, we have $P(\varepsilon(w_2, l_3, l_4, u_1^n, y_1^n)) \to 0,$ as $n \to \infty$.

For each (w_{31}, w_{32}, w_4), we define the following random variable and event:

$$\begin{aligned}N(w_{31}, w_{32}, w_4) :=&|\{(\widetilde{w}_2, \widetilde{l}_3, \widetilde{l}_4) : (X^n(W_{31}, W_{32}, W_4, \widetilde{w}_2, \widetilde{l}_3, \widetilde{l}_4), Y_1^n, U_1^n) \in T_\epsilon^{(n)}, \\ &(\widetilde{w}_2, \widetilde{l}_3, \widetilde{l}_4) \neq (W_2, L_3, L_4)\}|\end{aligned}$$

$$\varepsilon(w_{31}, w_{32}, w_4) :=\{N(w_{31}, w_{32}, w_4) \geq 2^{n(\widetilde{R}_3-R_{31}-R_{32}+\widetilde{R}_4-R_4)-nI(X;Y_1|U_1)+n\delta_n(\epsilon)-n\epsilon_n/2+n\epsilon'_n+1}\}.$$

Furthermore, we define the indicator random variable $E(w_{31}, w_{32}, w_4) := 0$ if $(X^n(w_3, w_4, W_2, L_3, L_4), Y_1^n, U_1^n) \in T_\epsilon^{(n)}$ and $\varepsilon(w_{31}, w_{32}, w_4)^c$ occurs; and $E(w_{31}, w_{32}, w_4) := 1$, otherwise. By such a definition, we have the following probability,

$$P(E(w_{31}, w_{32}, w_4) := 1) \leq P((X^n(w_{31}, w_{32}, w_4, W_2, L_3, L_4), Y_1^n, U_1^n) \notin T_\epsilon^{(n)}) + P(\varepsilon(w_{31}, w_{32}, w_4)). \tag{5.59}$$

The first term on the right hand side of the above inequality goes to zero as $n \to \infty$. We now bound the second term as follows,

$$\begin{aligned} P(\varepsilon(w_{31}, w_{32}, w_4)) &\leq \sum_{(u_1^n, y_1^n) \in T_\epsilon^{(n)}} P(u_1^n, y_1^n) P(\varepsilon(w_{31}, w_{32}, w_4) | u_1^n, y_1^n) + P((U_1^n, Y_1^n) \notin T_\epsilon^{(n)}) \\ &= \sum_{(u_1^n, y_1^n) \in T_\epsilon^{(n)}} \sum_{w_2, l_3, l_4} P(u_1^n, y_1^n) P(w_2, l_3, l_4 | u_1^n, y_1^n) P(\varepsilon(w_{31}, w_{32}, w_4) | u_1^n, y_1^n, w_2, l_3, l_4) \\ &\quad + P((U_1^n, Y_1^n) \notin T_\epsilon^{(n)}) \\ &\to 0, \text{ if } R_2 + \widetilde{R}_3 - R_{31} - R_{32} + \widetilde{R}_4 - \overline{R}_4 > I(X; Y_1 | U_1). \end{aligned} \tag{5.60}$$

Therefore,

$$\begin{aligned} &H(W_2, L_3, L_4 | Y_1^n, U_1^n, w_{31}, w_{32}, w_4) \\ &\leq H(W_2, L_3, L_4, E(w_{31}, w_{32}, w_4) | Y_1^n, U_1^n, w_{31}, w_{32}, w_4) \\ &\leq H(E(w_{31}, w_{32}, w_4)) + H(W_2, L_3, L_4 | E(w_{31}, w_{32}, w_4), Y_1^n, U_1^n, w_{31}, w_{32}, w_4) \\ &\leq 1 + P(E(w_{31}, w_{32}, w_4) = 1) H(W_2, L_3, L_4 | E(w_{31}, w_{32}, w_4) = 1, Y_1^n, U_1^n, w_{31}, w_{32}, w_4) \\ &\quad + H(W_2, L_3, L_4 | E(w_{31}, w_{32}, w_4) = 0, Y_1^n, U_1^n, w_{31}, w_{32}, w_4) \\ &\leq 1 + P(E(w_{31}, w_{32}, w_4) = 1) n(R_2 + \widetilde{R}_3 - R_{31} - R_{32} + \widetilde{R}_4 - \overline{R}_4) \\ &\quad + n(R_2 + \widetilde{R}_3 - R_{32} + \widetilde{R}_4 - \overline{R}_4) - nI(X; Y_1 | U_1) + n\delta_n(\epsilon) - n\epsilon_n/2 + n\epsilon_n' + 1. \end{aligned} \tag{5.61}$$

Hence, we have

$$\lim_{n\to\infty}\frac{1}{n}H(W_2,L_3,L_4|Y_1^n,U_1^n,w_{31},w_{32},w_4)\leq R_2+\widetilde{R}_3-R_{31}-R_{32}+\widetilde{R}_4-\overline{R}_4 \\ -I(X;Y_1|U_1)+\delta_n(\epsilon)+\epsilon'_n. \tag{5.62}$$

5.6 Proof of Lemma 5.2

We now bound $H(W_{31},L_3,L_4|Y_2^n,U_2^n,w_{32},w_4)$ for any pair of (w_{32},w_4), and hence, $H(W_{31},L_3,L_4|Y_2^n,U_2^n,W_{32},W_4)$ can be bounded.

We fix $(W_{31},L_3,L_4)=(w_{31},l_3,l_4)$ and a joint typical sequence $(u_2^n,y_2^n)\in T_\epsilon^{(n)}(U_2,Y_2)$. We define a random variable as follows:

$$N(w_{31},l_3,l_4,u_2^n,y_2^n):=|\{(\widetilde{w}_{31},\widetilde{l}_3,\widetilde{l}_4)\neq(w_{31},l_3,l_4):(x^n(\widetilde{w}_{31},w_{32},\widetilde{l}_3,w_4,\widetilde{l}_4),u_2^n,y_2^n) \\ \in T_\epsilon^{(n)}(X^n,U_2^n,Y_2^n)\}|. \tag{5.63}$$

We can show that the expectation of N satisfy the following inequalities:

$$\begin{aligned}&2^{n(\widetilde{R}_3-R_{32}+\widetilde{R}_4-R_4)-nI(X;Y_2|U_2)-n\delta_n(\epsilon)-n\epsilon_n}+2^{n(\widetilde{R}_4-R_4)-nI(X;Y_2|U_2)-n\delta_n(\epsilon)-n\epsilon_n}\\&\leq E(N(w_{31},l_3,l_4,u_2^n,y_2^n))\\&\leq 2^{n(\widetilde{R}_3-R_{32}+\widetilde{R}_4-R_4)-nI(X;Y_2|U_2)+n\delta_n(\epsilon)-n\epsilon_n}+2^{n(\widetilde{R}_4-R_4)-nI(X;Y_2|U_2)+n\delta_n(\epsilon)-n\epsilon_n},\end{aligned} \tag{5.64}$$

where $\delta_n(\epsilon)$, $\epsilon_n\to 0$ as $n\to\infty$. And this can be further written as

$$\begin{aligned}&2^{n(\widetilde{R}_3-R_{32}+\widetilde{R}_4-R_4)-nI(X;Y_2|U_2)-n\delta_n(\epsilon)-n\epsilon_n}(1+2^{-n(\widetilde{R}_3-R_{32}-I(U_3;Y_2|U_2))})\\&\leq E(N(w_{31},l_3,l_4,u_2^n,y_2^n))\\&\leq 2^{n(\widetilde{R}_3-R_{32}+\widetilde{R}_4-R_4)-nI(X;Y_2|U_2)+n\delta_n(\epsilon)-n\epsilon_n}(1+2^{-n(\widetilde{R}_3-R_{32}-I(U_3;Y_2|U_2))}).\end{aligned} \tag{5.65}$$

Furthermore, if $\widetilde{R}_3-R_{32}\geq I(U_3;Y_2|U_2)$, we can define ϵ'_n such that, $2^{n\epsilon'_n}=1+2^{-n(\widetilde{R}_3-R_{32}-I(U_3;Y_2|U_2))}$,

where, $\epsilon'_n \to 0$ as $n \to \infty$.

Then, we can write (5.65) as follows,

$$2^{n(\widetilde{R}_3 - R_{32} + \widetilde{R}_4 - R_4) - nI(X;Y_2|U_2) - n\delta_n(\epsilon) - n\epsilon_n + n\epsilon'_n} \leq E(N(w_{31}, l_3, l_4, u_2^n, y_2^n)) \leq 2^{n(\widetilde{R}_3 - R_{32} + \widetilde{R}_4 - R_4) - nI(X;Y_2|U_2) + n\delta_n(\epsilon) - n\epsilon_n + n\epsilon'_n}. \tag{5.66}$$

Due to the fact that $Var(N) \leq E(N)$, we have the following inequality,

$$Var(N(w_{31}, l_3, l_4, u_2^n, y_2^n)) \leq 2^{n(\widetilde{R}_3 - R_{32} + \widetilde{R}_4 - R_4) - nI(X;Y_2|U_2) + n\delta_n(\epsilon) - n\epsilon_n + n\epsilon'_n}. \tag{5.67}$$

We define the random event:

$$\varepsilon(w_{31}, l_3, l_4, u_2^n, y_2^n) := \{N(w_{31}, l_3, l_4, u_2^n, y_2^n) \geq 2^{n(\widetilde{R}_3 - R_{32} + \widetilde{R}_4 - R_4) - nI(X;Y_2|U_2) + n\delta_n(\epsilon) - n\epsilon_n/2 + n\epsilon'_n + 1}\}.$$

By Chebyshev's inequality, we obtain that

$$\begin{aligned}
&P(\varepsilon(w_{31}, l_3, l_4, u_2^n, y_2^n)) \\
&= P(N(w_{31}, l_3, l_4, u_2^n, y_2^n) \geq 2^{n(\widetilde{R}_3 - R_{32} + \widetilde{R}_4 - R_4) - nI(X;Y_2|U_2) + n\delta_n(\epsilon) - n\epsilon_n/2 + n\epsilon'_n + 1}) \\
&\leq P(N(w_{31}, l_3, l_4, u_2^n, y_2^n) \geq E(N(w_{31}, l_3, l_4, u_2^n, y_2^n)) \\
&\qquad + 2^{n(\widetilde{R}_3 - R_{32} + \widetilde{R}_4 - R_4) - nI(X;Y_2|U_2) + n\delta_n(\epsilon) - n\epsilon_n/2 + n\epsilon'_n}) \\
&\leq P\Bigg(|N(w_{31}, l_3, l_4, u_2^n, y_2^n) - E(N(w_{31}, l_3, l_4, u_2^n, y_2^n))| \\
&\qquad \geq 2^{n(\widetilde{R}_3 - R_{32} + \widetilde{R}_4 - R_4) - nI(X;Y_2|U_2) + n\delta_n(\epsilon) - n\epsilon_n/2 + n\epsilon'_n})\Bigg) \\
&\leq \frac{Var(N(w_{31}, l_3, l_4, u_2^n, y_2^n))}{2^{2n(\widetilde{R}_3 - R_{32} + \widetilde{R}_4 - R_4) - 2nI(X;Y_2|U_2) + 2n\delta_n(\epsilon) - n\epsilon_n + 2n\epsilon'_n}} \\
&\leq \frac{1}{2^{n(\widetilde{R}_3 - R_{32} + \widetilde{R}_4 - R_4) - nI(X;Y_2|U_2) + n\delta_n(\epsilon) + n\epsilon'_n}},
\end{aligned} \tag{5.68}$$

which goes to zero as $n \to \infty$ if $\widetilde{R}_3 - R_{32} + \widetilde{R}_4 - R_4 > I(X; Y_2|U_2)$. This implies that as $n \to \infty$,

$$P(\varepsilon(w_{31}, l_3, l_4, u_2^n, y_2^n)) \to 0.$$

For each pair of (w_{32}, w_4), we define the following random variable and event:

$$\begin{aligned} N(w_{32}, w_4) :=& |\{(\widetilde{w}_{31}, \widetilde{l}_3, \widetilde{l}_4) : (X^n(W_{32}, W_4, \widetilde{w}_{31}, \widetilde{l}_3, \widetilde{l}_4), Y_2^n, U_2^n) \in T_\epsilon^{(n)}, (\widetilde{w}_{31}, \widetilde{l}_3, \widetilde{l}_4) \\ & \neq (W_{31}, L_3, L_4)\}|, \\ \varepsilon(w_{32}, w_4) :=& \{N(w_{32}, w_4) \geq 2^{n(\widetilde{R}_3 - R_{32} + \widetilde{R}_4 - R_4) - nI(X;Y_2|U_2) + n\delta_n(\epsilon) - n\epsilon_n/2 + n\epsilon'_n + 1}\}. \end{aligned} \tag{5.69}$$

And we further define the indicator random variable $E(w_{32}, w_4) := 0$ if $(X^n(w_{32}, w_4, W_{31}, L_3, L_4), Y_2^n, U_2^n) \in T_\epsilon^{(n)}$ and $\varepsilon(w_{32}, w_4)^c$ occurs; and $E(w_{32}, w_4) := 1$, otherwise. By such a definition, we have the following probability,

$$P(E(w_{32}, w_4) := 1) \leq P((X^n(w_{32}, w_4, W_{31}, L_3, L_4), Y_2^n, U_2^n) \notin T_\epsilon^{(n)}) + P(\varepsilon(w_{32}, w_4)). \tag{5.70}$$

The first term on the right hand side of the above inequality goes to zero as $n \to \infty$. We now bound the second term as follows,

$$\begin{aligned} P(\varepsilon(w_{32}, w_4)) \leq & \sum_{(u_2^n, y_2^n) \in T_\epsilon^{(n)}} P(u_2^n, y_2^n) P(\varepsilon(w_{32}, w_4)|u_2^n, y_2^n) + P((U_2^n, Y_2^n) \notin T_\epsilon^{(n)}) \\ = & \sum_{(u_2^n, y_2^n) \in T_\epsilon^{(n)}} \sum_{w_{31}, l_3, l_4} P(u_2^n, y_2^n) P(w_{31}, l_3, l_4|u_2^n, y_2^n) P(\varepsilon(w_{32}, w_4)|u_2^n, y_2^n, w_{31}, l_3, l_4) \\ & + P((U_2^n, Y_2^n) \notin T_\epsilon^{(n)}) \\ \to & 0, \text{ if } \widetilde{R}_3 - R_{32} + \widetilde{R}_4 - R_4 > I(X; Y_2|U_2). \end{aligned} \tag{5.71}$$

Therefore,

$$
\begin{aligned}
&H(W_{31}, L_3, L_4|Y_2^n, U_2^n, w_{32}, w_4) \\
&\leq H(W_{31}, L_3, L_4, E(w_{32}, w_4)|Y_2^n, U_2^n, w_{32}, w_4) \\
&\leq H(E(w_{32}, w_4)) + H(W_{31}, L_3, L_4|E(w_{32}, w_4), Y_2^n, U_2^n, w_{32}, w_4) \\
&\leq 1 + P(E(w_{32}, w_4) = 1)H(W_{31}, L_3, L_4|E(w_{32}, w_4) = 1, Y_2^n, U_2^n, w_{32}, w_4) \\
&\quad + H(W_{31}, L_3, L_4|E(w_{32}, w_4) = 0, Y_2^n, U_2^n, w_{32}, w_4) \\
&\leq 1 + P(E(w_{32}, w_4) = 1)n(\widetilde{R}_3 - R_{32} + \widetilde{R}_4 - R_4) + n(\widetilde{R}_3 - R_{32} + \widetilde{R}_4 - R_4) \\
&\quad - nI(X; Y_2|U_2) + n\delta_n(\epsilon) - n\epsilon_n/2 + n\epsilon'_n + 1. \qquad (5.72)
\end{aligned}
$$

Hence, we have

$$
\lim_{n\to\infty} \frac{1}{n} H(W_{31}, L_3, L_4|Y_2^n, U_2^n, w_{32}, w_4) \leq \widetilde{R}_3 - R_{32} + \widetilde{R}_4 - R_4 - I(X; Y_2|U_2) + \delta_n(\epsilon) + \epsilon'_n. \qquad (5.73)
$$

CHAPTER 6
SUMMARY AND FUTURE WORK

In this section, we first summarize the results presented in this thesis, and then describe a few future directions.

6.1 Summary of the Work

In this thesis, we investigated two security issues in information systems: detection of anomalous data patterns that reflect malicious intrusions into data storage systems and protection of data from malicious attacks during data transmissions. We have applied information theoretic and statistical tools to develop solutions to these problems with guaranteed security performance.

More specifically, we have studied two classes of anomaly detection problems: anomalous geometric structure detection and anomalous data stream detection. For the problem of anomalous geometric structure detection over large networks, We have developed nonparametric tests using the MMD to measure the distance between the mean embeddings of distributions into an RKHS. We have analyzed the performance guarantee of our tests, and characterized the sufficient conditions on the minimum and maximum sizes of candidate anomalous structures to guarantee the consistency of our tests. We have further derived the necessary conditions and showed that our tests are order level optimal and nearly order level optimal respectively in terms of the minimum

and maximum sizes of candidate structures. For the problem of anomalous data stream detection, we have built MMD-based distribution-free tests to detect anomalous data streams. We have characterized the scaling behavior of the sample size m as the total number n of data streams goes to infinity in order to guarantee consistency of our tests. We have further characterized the conditions under which no test is universally consistent for arbitrary p and q, and thus established that our proposed tests are order level optimal.

For the secure communication problem, we have studied two degraded broadcast channel models. For the first model of the degraded broadcast channel with layered decoding and layered secrecy, we have studied the K-receiver degraded DMC and Gaussian MIMO cases. We have fully characterized the secrecy capacity regions of these two channels. We have further proposed an application of such an information theoretic model to the problem of multi-secret sharing, which is difficult to solve using number theoretic tools. We have characterized the secret sharing capacity region by reformulating the secret sharing problem into the problem of the degraded Gaussian MIMO broadcast channel. For the second model of the degraded broadcast channel with secrecy outside a bounded range, we have studied a four-receiver DMC case. We have characterized the secrecy capacity region of this model. We have designed an achievable scheme based on superposition, joint embedded coding and binning, and rate splitting and sharing. Among the techniques, rate splitting and sharing is critical for deriving a larger achievable region, for which the converse can be established.

We note that during the preparation of this thesis, we have characterized the secrecy capacity region for the K-user scenario with secrecy outside a bounded range. The idea of the achievable scheme is to designate one superposition layer to each message with random binning employed for each layer for protecting all upper-layer messages from lower-layer receivers. Such a scheme allows adjacent layers to share rates so that part of the rate of each message can potentially be shared with its immediate upper-layer message to enlarge the rate region. More importantly, we have developed an induction approach to perform Fourier-Motzkin elimination over $2K$ variables among $\Theta(K^2)$ bounds to obtain a close-form achievable rate region.

Furthermore, during the preparation of this thesis, we have studied the problem of estimating KL divergence between large alphabet distributions, in which the alphabet size k of the distributions can scale to infinity. Such a problem is interesting because the KL divergence is an important metric in information theory that measures the distance between two distributions. Such a metric can be further applied to solve the nonparametric anomaly detection problem. The estimation is based on m and n samples from two distributions respectively. We have shown that there does not exist any consistent estimator to guarantee asymptotically small worst case quadratic risk over the set of all pairs of distributions. We further consider a more practical set that contains pairs of distributions with bounded ratio. We have proposed an augmented plug-in estimator and characterized the sufficient and necessary conditions for this test to be consistent, which match in terms of the order of sample complexity. Furthermore, we have shown the necessary conditions for any estimator to be consistent.

6.2 Future Work

The exploration of the problems presented in this thesis further opens a number of research directions in the future.

We have studied the problem of detecting anomalous geometric structure over line, ring, and lattice networks. For future work, we are interested in detecting an event with a certain graph-based connectivity structure. For example, events of interests can be paths, subgraphs, cliques, or trees in networks. Such events can arise naturally in practice. For example, a trail event can model the object detection/tracking problem studied in signal processing, and a subgraph event can model an infected group of nodes in epidemic detection studied in social science settings. Previous studies of this problem mainly focused on parametric models, while our exploration will investigate nonparametric problems and design distribution-free tests.

We will also investigate scenarios where data samples arrive in an online fashion. We will design tests for detecting the change of a structure or a data stream from being typical to anomalous

as early as possible. We will then further characterize the performance of the designed tests in terms of the tradeoff between the detection accuracy, the expected sample size and delay, and the computational complexity of the algorithms.

We have shown the sufficient and necessary sample complexity for the augmented plug-in estimator for consistent estimation of KL divergence. We have further characterized the necessary conditions on the sample complexity for any consistent estimator. By comparing these two conditions, the augmented plug-in estimator is not optimal in terms of sample complexity and has a $\log k$ gap compared to the necessary sample complexity for any consistent estimator. In the future, we will further design consistent estimators that match the developed necessary conditions on the sample complexity.

We have characterized the secrecy capacity region for the discrete memoryless broadcast channel with secrecy outside a bounded range. Extension of such a model can be applied to study more practical fading wiretap channels with continuous channel states, in which messages decoded at a certain receiver are required to be kept secure from receivers that are outside a bounded range (i.e., with a certain level of worse channel quality). It is also of interest to study the models with arbitrary number of receivers in this class in the context of compound scenarios, in which each receiver and/or eavesdropper can represent a group of nodes. Such type of scenarios are more flexible to model practical networks with clusters of receivers.

REFERENCES

[1] E. Arias-Castro, D. L. Donoho, and X. Huo, "Near-optimal detection of geometric objects by fast multiscale methods," *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2402–2425, July 2005. 3

[2] G. Walther, "Optimal and fast detection of spatial clusters with scan statistics," *Ann. Statist.*, vol. 38, no. 2, pp. 1010–1033, 2010. 3

[3] P. M. Pacifico, C. Genovese, I. Verdinelli, and L. Wasserman, "False discovery control for random fields," *J. Amer. Stat. Assoc.*, vol. 99, pp. 1002–1014, 2004. 3

[4] E. Arias-Castro, E. J. Candes, H. Helgason, and O. Zeitouni, "Searching for a trail of evidence in a maze," *Ann. Statist.*, vol. 36, no. 4, pp. 1726–1757, 2008. 3, 41

[5] L. Addario-Berry, N. Broutin, L. Devroye, and G. Lugosi, "On combinatorial testing problems," *Ann. Statist.*, vol. 38, no. 5, pp. 3063–3092, 2010. 3

[6] E. Arias-Castro, E. J. Candes, and A. Durand, "Detection of an anomalous cluster in a network," *Ann. Statist.*, vol. 39, no. 1, pp. 278–304, 2011. 3

[7] J. Sharpnack, A. Rinaldo, and A. Singh, "Changepoint detection over graphs with the spectral scan statistic," in *Proc. International Conference on Artifical Intelligence and Statistics (AISTATS)*, Scottsdale, AZ, May 2013. 3

[8] J. Sharpnack, A. Rinaldo, and A. Singh, "Detecting activations over graphs using spanning tree wavelet bases," in *Artificial Intelligence and Statistics (AISTATS)*, Scottsdale, AZ, May 2013. 3

[9] J. Qian, V. Saligrama, and Y. Chen, "Connected sub-graph detection," in *Proc. International Conference on Artifical Intelligence and Statistics (AISTATS)*, 2014, pp. 796–804. 3

[10] J. Qian and V. Saligrama, "Efficient minimax signal detection on graphs," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2014, pp. 2708–2716. 3

[11] A. Gretton, K. Borgwardt, M. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *J. Mach. Learn. Res.*, vol. 13, pp. 723–773, 2012. 4, 6, 19, 20

[12] L. Lai, H. V. Poor, Y. Xin, and G. Georgiadis, "Quickest search over multiple sequences," *IEEE Trans. Inform. Theory*, vol. 57, no. 8, pp. 5375–5386, Aug. 2011. 6

[13] A. Tajer and H. V. Poor, "Quick search for rare events," *IEEE Trans. Inform. Theory*, vol. 59, no. 7, pp. 4462–4481, July 2013. 6

[14] Y. Li, S. Nitinawarat, and V. V. Veeravalli, "Universal outlier hypothesis testing," *IEEE Trans. Inform. Theory*, vol. 60, no. 7, pp. 4066–4082, July 2014. 6, 7, 8, 65

[15] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975. 9

[16] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inform. Theory*, vol. 24, no. 3, pp. 339–348, May 1978. 9, 89

[17] M. Bloch and J. Barros, *Physical-Layer Security: From Information Theory to Security Engineering*, Cambridge University Press, 2011. 9

[18] Y. Liang, H. V. Poor, and S. Shamai (Shitz), "Information theoretic security," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 4-5, pp. 355–580, Now Publishers, Hanover, MA, USA, 2008. 9

[19] S. Leung-Yan-Cheong and M. E. Hellman, "The Gaussian wire-tap channel," *IEEE Trans. Inform. Theory*, vol. 24, no. 4, pp. 451–456, Jul 1978. 9

[20] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas I: The MISOME wiretap channel," *IEEE Trans. Inform. Theory*, vol. 56, no. 7, pp. 3088–3104, July 2010. 9

[21] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas part II: The MIMOME wiretap channel," *IEEE Trans. Inform. Theory*, vol. 56, no. 11, pp. 5515–5532, Nov 2010. 9

[22] S. Gerbracht, C. Scheunert, and E. A. Jorswieck, "Secrecy outage in MISO systems with partial channel information," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 704–716, April 2012. 9

[23] S. Shafiee and S. Ulukus, "Achievable rates in Gaussian MISO channels with secrecy constraints," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, June 2007, pp. 2466–2470. 9

[24] T. Liu and S. Shamai, "A note on the secrecy capacity of the multiple-antenna wiretap channel," *IEEE Trans. Inform. Theory*, vol. 55, no. 6, pp. 2547–2553, June 2009. 9

[25] F. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," *IEEE Trans. Inform. Theory*, vol. 57, no. 8, pp. 4961–4972, Aug 2011. 9

[26] Y. Liang, G. Kramer, H. V. Poor, and S. Shamai (Shitz), "Compound wire-tap channels," *EURASIP Journal on Wireless Communications and Networking, Special Issue on Wireless Physical Layer Security*, 2009. 9

[27] I. Bjelakovic, H. Boche, and J. Sommerfeld, "Secrecy results for compound wiretap channels," *Problems of Information Transmission*, vol. 49, no. 1, pp. 73–98, 2013. 9

[28] A. Khisti, "On the MISO compound wiretap channel," in *Proc. Information Theory and Applications Workshop (ITA)*, Jan 2010, pp. 1–7. 9

[29] A. Khisti, A. Tchamkerten, and G. W. Wornell, "Secure broadcasting over fading channels," *IEEE Trans. Inform. Theory*, vol. 54, no. 6, pp. 2453–2469, June 2008. 9, 10

[30] R.F. Schaefer and A. Khisti, "Secure broadcasting of a common message with independent secret keys," in *48th Annual Conference on Information Sciences and Systems (CISS)*, March 2014, pp. 1–6. 9

[31] Y. Liang, H. V. Poor, and S. Shamai (Shitz), "Secure communication over fading channels," *IEEE Transactions on Information Theory, Special Issue on Information Theoretic Security*, vol. 54, no. 6, pp. 2470–2492, June 2008. 9

[32] H.D. Ly, T. Liu, and Y. Liang, "Multiple-input multiple-output Gaussian broadcast channels with common and confidential messages," *IEEE Trans. Inform. Theory*, vol. 56, no. 11, pp. 5477–5487, Nov 2010. 9

[33] Y.-K. Chia and A. El Gamal, "Three-receiver broadcast channels with common and confidential messages," *IEEE Trans. Inform. Theory*, vol. 58, no. 5, pp. 2748–2765, May 2012. 9

[34] R.F. Schaefer and H. Boche, "Robust broadcasting of common and confidential messages over compound channels: Strong secrecy and decoding performance," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 10, pp. 1720–1732, Oct 2014. 9

[35] Y. Liang, L. Lai, H. V. Poor, and S. Shamai (Shitz), "A broadcast approach for fading wiretap channels," *IEEE Trans. Inform. Theory*, vol. 60, no. 2, pp. 842–858, Feb 2014. 9

[36] G. Bagherikaram, A. S. Motahari, and A. K. Khandani, "Secrecy rate region of the broadcast channel with an eavesdropper," *available at http://arxiv.org/abs/0910.3658*, 2009. 10

[37] M. Benammar and P. Piantanida, "On the secrecy capacity region of the wiretap broadcast channel," in *Proc. IEEE Information Theory Workshop (ITW)*, Nov 2014, pp. 421–425. 10

[38] G. Bagherikaram, A. S. Motahari, and A.K. Khandani, "Secrecy capacity region of Gaussian broadcast channel," in *43rd Annual Conference on Information Sciences and Systems (CISS)*, March 2009, pp. 152–157. 10

[39] R. Liu, T. Liu, H. V. Poor, and S. Shamai (Shitz), "A vector generalization of Costa's entropy-power inequality with applications," *IEEE Trans. Inform. Theory*, vol. 56, no. 4, pp. 1865 –1879, April 2010. 10, 75, 103

[40] R. Tandon, P. Piantanida, and S. Shamai, "On multi-user MISO wiretap channels with delayed CSIT," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, June 2014, pp. 211–215. 10

[41] E. Ekrem and S. Ulukus, "Degraded compound multi-receiver wiretap channels," *IEEE Trans. Inform. Theory*, vol. 58, no. 9, pp. 5681–5698, 2012. 10, 75, 90

[42] E. Ekrem and S. Ulukus, "Secrecy capacity of a class of broadcast channels with an eavesdropper," *EURASIP J. Wirel. Commun. Netw.*, vol. 2009, pp. 1:1–1:29, Mar. 2009. 10

[43] R. Liu, I. Maric, P. Spasojević, and R.D. Yates, "Discrete memoryless interference and broadcast channels with confidential messages: Secrecy rate regions," *IEEE Trans. Inform. Theory*, vol. 54, no. 6, pp. 2493–2507, June 2008. 10

[44] R. Liu and H. V. Poor, "Secrecy capacity region of a multiple-antenna Gaussian broadcast channel with confidential messages," *IEEE Trans. Inform. Theory*, vol. 55, no. 3, pp. 1235–1249, March 2009. 10

[45] R. Liu, T. Liu, H. V. Poor, and S. Shamai, "Multiple-input multiple-output Gaussian broadcast channels with confidential messages," *IEEE Trans. Inform. Theory*, vol. 56, no. 9, pp. 4215–4227, Sept 2010. 10

[46] S. Yang, M. Kobayashi, P. Piantanida, and S. Shamai, "Secrecy degrees of freedom of MIMO broadcast channels with delayed CSIT," *IEEE Trans. Inform. Theory*, vol. 59, no. 9, pp. 5244–5256, Sept 2013. 10

[47] M. Kobayashi, Y. Liang, S. Shamai, and M. Debbah, "On the compound MIMO broadcast channels with confidential messages," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, June 2009, pp. 1283–1287. 10

[48] A. El Gamal and Y.-H. Kim, *Network Information Theory*, Cambridge University Press, New York, 2012. 12, 76, 81

[49] S. Zou, Y. Liang, L. Lai, and S. Shamai (Shitz), "An information theoretical approach to secrecy sharing," *IEEE Trans. Inform. Theory*, vol. 61, no. 6, pp. 3121–3136, 2015. 15

[50] S. Zou, Y. Liang, L. Lai, H. V.. Poor, and S. Shamai, "Broadcast networks with layered decoding and layered secrecy: Theory and applications," *Proceedings of the IEEE*, vol. 103, no. 10, pp. 1841–1856, 2015. 15

[51] S. Zou, Y. Liang, and H. V. Poor, "Nonparametric detection of geometric structures over networks," *submitted to IEEE Transactions on Information Theory*, 2015. 15

[52] S. Zou, Y. Liang, H. V. Poor, and X. Shi, "Nonparametric detection of anomalous data streams via kernel mean embedding," *submitted to IEEE Transactions on Information Theory*, 2015. 15

[53] S. Zou, Y. Liang, H. V. Poor, and X. Shi, "Kernel-based nonparametric anomaly detection," in *Proc. IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2014, pp. 224–228. 15

[54] S. Zou, Y. Liang, and H. V. Poor, "A kernel-based nonparametric test for anomaly detection over line networks," in *Proc. IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2014. 15

[55] S. Zou, Y. Liang, H. V. Poor, and X. Shi, "Unsupervised nonparametric anomaly detection: A kernel method," in *Proc. Allerton Conference on Communication, Control, and Computing (Allerton)*, 2014, pp. 836–841. 15

[56] S. Zou, Y. Liang, and H. V. Poor, "Nonparametric detection of an anomalous disk over a two-dimensional lattice network," *to appear in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2016. 15

[57] Y. Bu, S. Zou, Y. Liang, and V. V. Veeravalli, "Universal outlying sequence detection for continuous observations," *to appear in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2016. 15

[58] S. Zou, Y. Liang, L. Lai, and S. Shamai (Shitz), "Layered decoding and secrecy over degraded broadcast channels," in *Proc. 2013 IEEE 14th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Darmstadt, Germany, June 2013, pp. 679–683. 15

[59] S. Zou, Y. Liang, L. Lai, and S. Shamai (Shitz), "Layered secure broadcasting over MIMO channels and application in secret sharing," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Hawaii, USA, June 2014, pp. 396–400. 15

[60] S. Zou, Y. Liang, L. Lai, and S. Shamai (Shitz), "Degraded broadcast channel: Secrecy outside of a bounded range," in *Proc. IEEE Information Theory Workshop (ITW)*, 2015. 15, 103

[61] S. Zou, Y. Liang, L. Lai, and S. Shamai, "Rate splitting and sharing for degraded broadcast channel with secrecy outside a bounded range," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, June 2015. 15

[62] A. Berlinet and C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Springer, 2004. 19

[63] B. Sriperumbudur, A. Gretton, K. Fukumizu, G. Lanckriet, and B. Schölkopf, "Hilbert space embeddings and metrics on probability measures," *J. Mach. Learn. Res.*, vol. 11, pp. 1517–1561, 2010. 19, 20

[64] A. Gretton, K. Borgwardt, M. Rasch, B. Schökopf, and A. Smola, "A kernel method for the two-sample-problem," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2007. 19

[65] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, The MIT Press, Cambridge, MA, USA, 2001. 20

[66] T. Hofmann, B. Schölkopf, and A. J. Smola, "Kernel methods in machine learning," *Ann. Statist.*, vol. 36, no. 3, pp. 1171–1220, 2008. 20

[67] K. Fukumizu, A. Gretton, X. Sun, and B. Schölkopf, "Kernel measures of conditional dependence," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2008. 20

[68] B. Sriperumbudur, A. Gretton, K. Fukumizu, G. Lanckriet, and B. Schölkopf, "Injective Hilbert space embeddings of probability measures," in *Proc. Annual Conference on Learning Theory (COLT)*, 2008. 20

[69] K. Fukumizu, B. Sriperumbudur, A. Gretton, and B. Schölkopf, "Characteristic kernels on groups and semigroups," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2009. 20

[70] J. H. Friedman and L. C. Rafsky, "Multivariate generalizations of the wald-wolfowitz and smirnov two-sample tests," *Ann. Statist.*, vol. 7, no. 4, pp. pp. 697–717, 1979. 34

[71] P. Hall and N. Tajvidi, "Permutation tests for equality of distributions in high-dimensional settings," *Biometrika*, vol. 89, no. 2, pp. pp. 359–374, 2002. 34

[72] Z. Harchaoui, F. Bach, and E. Moulines, "Testing for homogeneity with kernel fisher discriminant analysis," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2008. 34

[73] M. Sugiyama, T. Suzuki, and T. Kanamori, *Density Ratio Estimation in Machine Learning*, Cambridge University Press, New York, NY, USA, 2012. 34

[74] P. Hall, "On the rate of convergence of normal extremes," *Journal of Applied Probability*, vol. 16, no. 2, pp. 433–439, 1979. 60

[75] E. Ekrem and S. Ulukus, "The secrecy capacity region of the Gaussian MIMO multi-receiver wiretap channel," *IEEE Trans. Inform. Theory*, vol. 57, no. 4, pp. 2083–2114, 2011. 90, 91, 92

VITA

NAME OF AUTHOR: Shaofeng Zou

PLACE OF BIRTH: Yantai, Shandong, China

DATE OF BIRTH: Oct. 18, 1989

UNDERGRADUATE SCHOOLS ATTENDED:

Shanghai Jiao Tong University, Shanghai, China

DEGREES AWARDED:

B. E., 2011, Shanghai Jiao Tong University, Shanghai, China