Network Reliability Analysis under the Multivariate Normal Model

by

Jerrod Wigmore

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of

Master of Science in Aeronautics and Astronautics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2021

Author . . . .
Department of Aeronautics and Astronautics
August 17, 2021

Certified by. . . . .
Eytan Modiano
Professor of Aeronautics and Astronautics
Thesis Supervisor

Certified by. . . . .
Brooke Shrader
Senior Technical Staff at MIT Lincoln Laboratory
Thesis Supervisor

Accepted by . . . .
Jonathan How
Chair, Graduate Program Committee

Network Reliability Analysis under the Multivariate Normal Model

by

Jerrod Wigmore

Abstract

High Frequency (HF) radios have been used since the early 20th century for long-distance communication. HF communication systems primarily utilize skywave propagation in which the ionosphere is used to reflect radiowaves back to Earth. The performance of HF communication links is directly tied to the ionospheric propagation medium. The ionosphere is a highly variable and irregular environment that creates many challenges in the design of robust HF communication networks. Ionospheric characteristics vary temporally and will be spatially correlated. As a result, link failures within an HF network may also be correlated. In this thesis we develop a novel model for probabilistic link failures that captures the correlation expected in HF communication networks. We focus on two problems related to the design of robust HF networks- 1.) The Network Reliability Problem, which seeks to compute the probability a network is operational in the presence of random link failures, and 2.) the Most Reliable Path Problem, which seeks to identify the most reliable path between two nodes in a network.

Thesis Supervisor: Eytan Modiano
Title: Professor of Aeronautics and Astronautics

Thesis Supervisor: Brooke Shrader
Title: Senior Technical Staff at MIT Lincoln Laboratory

Acknowledgments

This research would not be possible without the combined efforts and guidance from my thesis advisors, Professor Eytan Modiano and Dr. Brooke Shrader. I am extremely grateful for their encouragement throughout the research process as well as their dedication to my development as a researcher. I would also like to sincerely thank Lincoln Labs for funding my research and providing me the opportunity to intern with them this past summer.

Next, I would like to thank my mentors during my undergraduate studies at the University of South Carolina, Dr. Roger Dougal and Dr. David Matolak, for their advice and insistence on pursuing a graduate degree. I would also like to thank Beth Marois for making sure I actually completed the proper steps to receive my graduate degree. The past two years would not have been the same without the following people: Francisco, Soumya, Chirag, Alejandro, Akshay, George, Sam, Tank, John, Sasha, Liza, Artur, Skylar, Katie, and Gavin. You all played a major part in keeping me sane throughout graduate school and I sincerely thank every single one of you for that.

Last but not least, I would like to thank my family for all of their love and encouragement. I know my parents worked relentlessly to make sure I had the opportunities that were never afforded to them, and for that, I am beyond grateful.

THIS PAGE INTENTIONALLY LEFT BLANK

Contents

List of Figures

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1

Introduction

High Frequency (HF) communication is a wireless communication paradigm that allows users to communicate over long-distances without without the use of satellites or terrestrial relays. The dominant mode of operation is *Skywave* propagation in which radio signals in the HF frequency band (3-30 MHz) are directed at an angle towards the sky. Upon reaching the ionosphere, the signals undergo a process of refraction and are eventually reflected back to Earth. The ionosphere is an ionized region of Earth's atmosphere and acts a natural reflector for HF radio wave signals. This ionospheric reflection capability coupled with HF signals' low free-space attenuation allows for HF radio signals to be received up to 10,000km from their transmitting location. This ionospheric propagation mechanism provides a unique opportunity for global connectivity using very minimal infrastructure.

Ever since the advent of satellites, the use of HF communications has declined due to to its limited data rate and frequency management problems. Nonetheless, it offers advantages over satellite communication (SATCOM) which make HF a worthy alternative or backup to SATCOM for beyond line-of-sight communications. From a national security perspective, satellites are more vulnerable to destruction whereas the ionosphere is permanent fixture of Earth's atmosphere. For example, an electromagnetic pulse (EMP) may disrupt ionospheric propagation in the short-term, however the same EMP may permanently destroy a SATCOM system [21, Chapter 1]. Another major benefit of HF over SATCOM is the relatively low costs and com-

plexity of deployment of HF radios compared to satellites. Although HF antennas may be large, they are relatively simple and more cost effective compared to deploying a satellite with the same coverage. This relative simplicity and low costs makes its valuable for civilian use in applications where large coverage is needed without the use of intermediate relays. Examples include disaster relief operations where existing communication infrastructure may be compromised, communicating with ships at sea or large aircraft where line-of-sight of sight methods are unavailable, or even communicating with remote areas of the globe [25].

There are however challenges to HF radio communications that are not encountered in traditional line-of-sight wireless communications. The performance of HF communication systems are heavily tied to the behavior of the ionosphere which is a dynamic and inhomogeneous propagation medium. The structure and characteristics of the ionosphere results from a complex interaction of solar radiation with Earth's atmosphere. Ionospheric characteristics that influence HF signal propagation such as free-electron density and layer height vary in time and space. These variations have been well-studied and modeled on a global scale, however the exact channel characteristics for an HF links can only be known via real-time sounding. Additionally, regions of the ionosphere may be in a "disturbed" state due to naturally occurring impulsive events such as solar flares and geomagnetic storms which may have adverse effects on communication links in the affected areas. Networked HF systems have to be able to account for these natural and sometimes impulsive variations of the propagation medium.

The non-uniformity of the ionosphere creates a unique quality of radio HF networks compared to most line-of-sight wireless networks. In traditional line-of-sight wireless networks, the channel properties of different links are assumed to be stationary as well as statistically independent. Due to the regional correlation of ionospheric variability, we know this independence assumption is not valid for all links within an HF network. The ionosphere may be constantly changing with time, however these variations are not purely stochastic in nature and may show high degrees of spatial correlation. For example, Figure 1-1 displays the predicted received power all over the
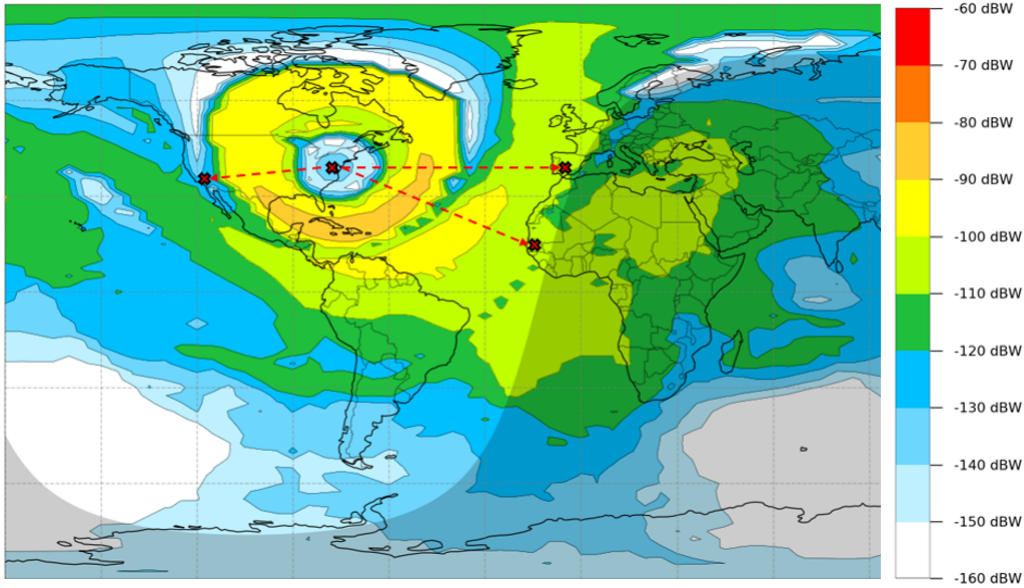
Figure 1-1: Map of received power for omnidirectional 14.1 MHz transmission from Washington, DC at 19:00 UTC in the month of August. Produced using VOACAP online propagation predictor: https://www.voacap.com/hf/

globe from an omnidirectional 14.1 MHz transmission originating Washington, DC. Overlayed on the receive power map are three possible receiver locations in Spain, West Africa, and California. Since the received power between the Spanish and West African link are similar, we can assume the ionospheric characteristics about the regions in which their respective signals propagate are also similar. In contrast, the ionospheric characteristics for the link to California would not be as statistically correlated as the two other links. From a network design perspective, the correlation of link performance in an HF network must be taken into account. Links with a high degree of correlation are prone to simultaneous failures which may have serious implications for a large networks that rely on multiple HF radio links to establish full network connectivity.

1.1   Network Reliability

Network Reliability is well-studied field that aims to quantify the robustness of a system in the presence of random component failures. These systems are modeled as
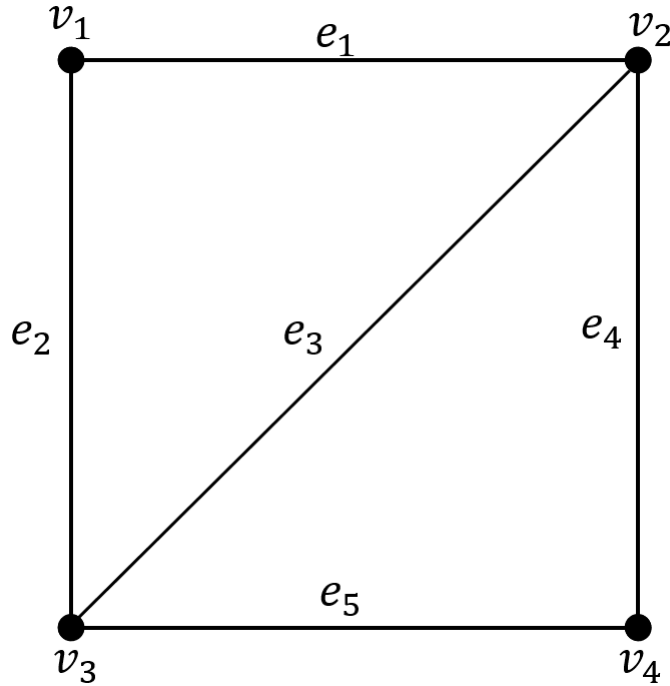
Figure 1-2: Bridge graph with four vertices and five edges. For reliability modeling, each edge would be assigned a probability of failure

graphs $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, and when applied to communication networks, the vertices/nodes $\mathbf{V} = [v_1, ..., v_n]$ represent the communication centers/radios while the edges $\mathbf{E} = [e_1, ..., e_m]$ represent the communication links between nodes. Associated with each link $e_i \in \mathbf{E}$ is some probability of operation $p_i \in \mathbf{p}$. The traditional network reliability model typically uses all or most of the following set of simplifying assumptions to make the analysis more tractable:

1. All nodes do not fail, thus the network reliability is only depends on edge failures

2. Edges can only be in one of two possible states: operating or failed

3. All edge failure probabilities are statistically independent

The last assumption of statistical independence is central to most works in the network reliability literature, yet the validity of this assumption is dependent on the system being modeled. First lets describe what statistical independence means in the context of communication networks. In general, two random events X and Y are

independent if and only if their joint probability of occurrence is the product of their marginal probabilities [19]:

$$\mathbb{P}(X \text{ occurs AND } Y \text{ occurs}) = \mathbb{P}(X \text{ occurs})\mathbb{P}(Y \text{ occurs}) \tag{1.1}$$

As a result, the occurrence of X does not alter the probability of the occurrence of Y and vice-versa:

$$\mathbb{P}(Y \text{ occurs}|X \text{ occurs}) = \mathbb{P}(Y \text{ occurs}) \tag{1.2}$$

From a network reliability perspective, this means the failure of link e_i does not affect the probability of failure of any other link within the network. This assumption may be true when modeling line-of-sight wireless networks, however in HF Networks we know that link failures may be spatially correlated.

1.2 Research Aims and Objectives

Understanding the reliability of multi-hop HF Networks inspired the development of our Multivariate Normal (MVN) Model: a novel probabilistic model for analyzing the reliability of networks with correlated failures. In this model we assume the set of all signal strengths for the links within the network can be modeled as a correlated multivariate normal random variable. There are two major benefits to this model compared to other approaches to capture correlated failures:

1. For a network with m links, it only requires the specification of $\binom{m+1}{2}$ correlation coefficients to model all $2^m - 1$ possible statistical dependencies between sets of links.

2. Its a general model that does not require the modeling of the external process that cause the correlated link failures

The aim of this work was to develop methodologies to solve two commonly tackled problems in the network reliability literature under the MVN Model:

1. Network Reliability Problem: Compute the probability a network is operational

17

2. Most Reliable Path Problem: Identify the most reliable sequence of edges connecting two nodes

In this thesis, we describe these problems and how they differ between the traditional network reliability model and the MVN model. We also demonstrate how most methodologies previously developed for solving these problems when link failures are independent are non-generalizable to the case of correlated failures.

1.3 Prior Correlated Failure Models

Here we present a brief literature review on other approaches to incorporate correlated failures into network reliability analysis.

Early approaches to capture correlation required the specification of conditional probabilities and utilized the probabilistic chain rule to compute the overall reliability. An obvious issue with this approach is that there are $2^m - 1$ conditional probabilities that must be specified for a network of m links. The q-Ψ model utilizes a Markovian model where only conditional probabilities of adjacent links are considered [45], [4]. As a result of the Markovian property, the order in which the chain rule expansion is used changes the joint probability thus this approach lacked a "consistency" property. The ϵ-model was introduced to overcome the issue of inconsistency in the q-ψ model and introduces and utilizes a set of inputs ϵ, that details the perturbations from the independent model caused by the correlation between link failures [39]. Like other conditional probability based models, the inputs of the ϵ-model is exponential in the number of links.

Lam and Li developed the Event-Based Network Model (EBRM) which is one of the first works to capture correlation without the specification of conditional probabilities as inputs [29]. The EBRM models each failure causing mechanism using independent "event elements" which are assigned fixed probabilities and cause failures of their associated links. This model may also suffer from the exponential growth of the input parameter space as to capture all possible dependencies for m links, it would require the specification of $2^m - 1$ event elements. First introduced in [26],

Shared Risk Link Groups framework captures correlation by assigning links into sets such that a failure of one link in the set causes the failures of all other links within the same set. This model was inspired by optical WDM Networks where different logical links may share a common physical conduit. Probabilistic Shared Risk Groups is an extension of the SRLG framework that allows for non-deterministic dependent failures [30].

Multiple geographic failure models have been proposed which model external processes disrupting the physical infrastructure of the network creating spatially correlated failures. These events are modeled as geometric cuts or disks [34] [35], spatial point processes [41], or even based on physical models such as rain fading in cellular networks [49]. Geographic failure-based models typically focus on identifying regional vulnerabilities in the network based on its geographic topology and the specific model of the external disturbances.

There have been few works that use non-Bernoulli distributions to incorporate correlation into Network reliability analysis. Nguyen et all extend the idea of probabilistic graphs to uncertain graphs in which edge reliabilities are described using a boolean expressions of indicator random variables. Multiple link probabilities may share the same indicator random variable in their associated boolean expression, thus can be correlated. This approach is a generative model where non-Bernoulli distributions are sampled to eventually generate probabilities of failure for each links [36] [37]. Botev et al develops a static network reliability model based on a Marshall-Olkin copula, which models networks of components with time-to-failure distributions. This approach specifies multivariate lifetime or repair distributions for links in the graph and the equivalent static reliability formulation is made by assessing whether the network is operational by time 1 [10]. Botev's modeling is slightly different than our MVN model as its underlying multivariate distribution is artificial in the sense that it doesn't represent an underlying physical quantity that dictates the reliability. Instead the latent distributions are generated by the user as a mechanism for importance sampling in Network Reliability estimations.

1.4 Chapter Outline

The remainder of this thesis is organized as follows:

Chapter 2 provides background information on HF communication systems, the inspiration for the MVN Model. It covers the ionosphere and its effects on HF radiowaves, as well as the various HF Channel models. We then describe how the correlation of ionospheric characteristics and the effect of the ionosphere on the propagation of radio waves leads to the assumption that the received signal strengths on links in a network may be modeled as correlated normal random variables.

Chapter 3 formulates the MVN Model. First we detail the independent Bernoulli random failure model used in most of the Network Reliability literature. Then we contrast this model with the MVN model which is able to capture correlated failures in networks. We then discuss the reliability problem and describe exact methods of computing the reliability and the limitations of the exact methods.

Chapter 4 discusses methods to estimate the reliability under the MVN model. First we provide an abstract representation of the estimation problem. It then describes how approaches previously used in the literature for the Bernoulli model are not valid when link failures are correlated. Two algorithms are then provided to estimate the reliability under the MVN model and their performance is compared.

Chapter 5 explores the most reliable path problem under the MVN model. It first describes how the problem under the MVN model lacks the property of optimal substructure which is essential for the previous techniques used to solve this problem when link failures are independent. It then describes a method creating optimal substructure that permits a greedy algorithm to enumerate the most reliable path. An upper and lower bound on the most reliable path is given based on assuming all links are fully correlated or independent respectively. Finally, it describes an approach where we assume links are independent and solve the equivalent most reliable path problem may be a valid approach for solving the most reliable path problem under the MVN Model.

Chapter 2

HF Communication

The following chapter provides background information on HF communication. First, we discuss the composition and variability of the ionosphere which HF communication is reliant upon. Next, we cover HF skywave propagation and its relationship to the ionosphere. Finally, we discuss HF channel modeling and the characteristics that lead to the MVN model.

2.1   The Ionosphere

The ionosphere is an outer region of Earth's atmosphere that consists of partially ionized gases and free electrons. Ionization occurs in sunlit regions of the ionosphere when ultra-violet and x-ray solar radiation collides with neutral gas molecules. During this interaction, these neutral gas molecules release an electron and become positively charged. Recombination occurs when free electrons combine with the positively charged ions to form inert molecules. During the day the rate of ionization is generally greater than the rate of ionization. At night, recombination continues while ionization effectively ceases. The ionosphere can be broken up into multiple regions which are distinct in their electron density and height. The D-region is the lowest (70-90km above the Earth) region, and primarily acts as an attenuator to HF radiowaves. D-region attenuation is greatest during the day when ionization peaks, however this region completely dissipates shortly after sunset through the recombina-
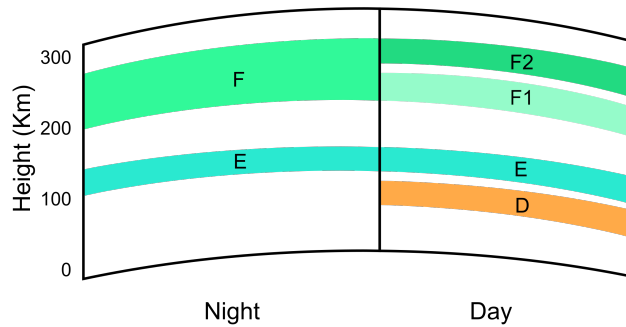
Figure 2-1: Ionospheric Regions and their heights [40]

tion process. Attenuation by the D-region is also frequency dependent, with higher frequency signals attenuated more than lower frequency signals. The E-region's altitude is approximately 90km to 130km above Earth's surface, and unlike the D-region, refracts HF radiowaves and can be used for medium range communications on the lower end of the HF spectrum (less than 10 MHz). The E-region also dissipates at night but does not completely disappear. The F-region is the dominant region used for long-range HF communications due to its height (130-1000km) and its high free-electron density. At night, the F-region is a single layer but during the day this region splits into two distinct layers, the lower F1 (130-310km) and and upper F2 layer (200-1000). Peak ionization is generally achieved between mid-day to mid-afternoon, and due to its height allows HF signals to propagate up to 4,000km on a single reflection off of ionosphere. [21, Chapter 2].

The ionosphere is a highly dynamic and inhomogeneous propagation medium. Its properties vary temporally and spatially, and both variations primarily coincide with the amount of solar radiation absorbed by the particular region. Solar activity varies on a nine to fourteen year solar cycle which corresponds to how much radiation is being emitted by the sun. [5]. A well-known index of solar activity, the sunspot number (SSN), roughly characterizes the number of spots visible on the solar disk. It is used in predicting HF propagation as the SSN is proportional to the component of solar activity which most severely influences HF systems [21, Chapter 2]. During

solar minimum, fewer free electrons are produced in the ionosphere and only lower frequency HF signals will be reflected. Meanwhile, during solar maximum solar radiation is at its peak thus more free electrons are available in the ionosphere thus higher frequencies in the HF bands may be used. Also during solar maximum there is a greater likelihood of large solar flares occurring. Solar flares increase the ionization of and absorption by the D-layer and can cause brief loss of propagation for HF signals [5].

Seasonal variation in the ionosphere is due to changes in the distance from the sun as well as the tilt of Earth's axis. For the hemisphere experiencing summer, the apparent elevation of the sun is greater and it receives more direct illumination for a longer period each day compared to the other hemisphere experiencing winter. This means at noon in the summer, ionization of all layers is typically greater than the winter months and higher frequency propagation is supported. However, there are "anomalous features" such as the so called "seasonal anomaly" which refers to greater support at noon for higher frequencies in the Winter than those in the Summer during solar maximum [31]. Hourly variation in the ionosphere stems from the rotation of the Earth about its axis. Supported frequencies for a given region are higher during the day due to direct exposure to the sun.

Generally free electron density increases during periods that correspond to increased solar radiation on all timescales (i.e. daily: noon; seasonally: summer; solar cyclically: solar maximum). Thus higher frequency propagation is supported within the HF Bands during the periods of increased solar radiation. However, this trend is primarily for a non-disturbed ionosphere in the mid-latitude regions of the Northern and Southern Hemispheres. Around Earth's poles and the equator we see many irregularities that cause deviations away from this overall trend. In particular the High Latitude Region is the portion of the ionosphere whose properties are largely reliant on the magnetospheric fluctuations and corpuscular radiation as opposed to solar radiation. In this region we see a more frequent occurrence of ionospheric irregularities and disturbances, which provide special challenges for HF propagation that don't exist in mid-latitude or equatorial regions. For an in depth look into the iono-

spheric differences of the High Latitude and Equatorial regions, readers are directed to [21, Chapter 3] and the sources within. There are a wide range of irregularities that may exist in certain regions in the ionosphere and their impact on HF propagation is important for a complete understanding of the dynamic HF propagation medium. Table 2.1 contains a brief description of the most widely studied irregularities and ionospheric disturbances as well as their effect on HF communication Links.

Name	Cause	Effect	Locations	Duration
Ionospheric Region Irregularities				
Sporadic E (Es)	Anomalous form of ionization with little direct impact relationship to solar ionizing radiation. Size, duration, and effect is dependent on geographic location and thus is often separated into three different types (Mid-latitude Es, Equatorial Es, and Auroral Es)	Increased Ionization of E-Region	Can occur anywhere globally and its region of effect can range from a few km to hundreds of km	Minutes to hours
Spread F	Generally geomagnetic disturbances Associated with ionospheric storms	Scattering of radiowaves	Equatorial and high latitudes with area effected up to 10km by a single event	Hours
Sudden Ionospheric Disturbances				
Short-wave fadeouts (SWF)	Large solar flares causing increased ionization of the D-layer	Increased absorption of mostly the lower frequencies. May render the entire HF spectrum unusable if the solar flare is large enough	Sunlit regions	Few-minutes to a few hours (depending on magnitude of flare)
Polar Cap Absorption (PCA)	High energy protons from Solar flare ionizing D-region along magnetic field lines to the polar regions	Increased HF Absorption for Circuits along polar regions	Polar Regions	Minutes to days
Ionospheric Storm	Solar event causing large disturbances in Earth's geomagnetic field	Brief increase then prolonged decrease in electron density leading to only lower frequencies being supported	Generally higher latitudes are more effected	Up to a couple of days

Table 2.1: Table of ionospheric irregularities and disturbances. Adopted from [2] and [5]

2.2 HF Communication

The inhomogenous and dynamic ionosphere creates a unique propagation environment for HF radiowaves. The following section describes the mechanisms behind ionospheric propagation and the multitude of effects the ionosphere has on radiowaves in the HF Frequency band.

2.2.1 Ionospheric Propagation

When an electromagnetic wave passes through an ionized region, the upper part of the wave continually slows causing the wave to refract or bend as it propagates. The amount of refraction depends on the ionization density of the medium, the frequency of the electromagnetic wave, and the angle of incidence [21, Chapter 4]. Thus for HF communication, the signal path between a transmitter and receiver is determined by the electron density along the portion of the path that passes through the ionosphere, the frequency of the signal, and the transmitting antenna's elevation angle.
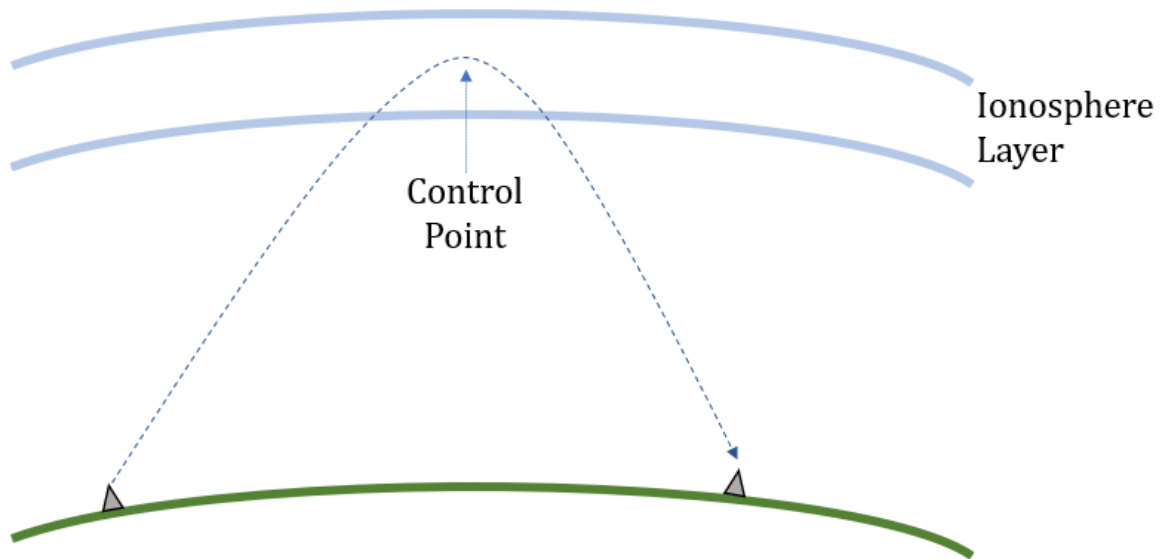
Figure 2-2: Basic single-hop path of skywave propagation

For a given layer, its critical frequency is the highest frequency at which a vertically incident radio wave refracts 180 degrees [2]. As a function of free electron density N (electrons per m^3), it is approximately:

$$f_0 \approx 9\sqrt{N} \tag{2.1}$$

Similarly, for oblique transmissions, the maximum usable frequency (MUF) for a given layer is given by:

$$f_M \approx f_0 \sec(\phi) \tag{2.2}$$

where ϕ is the angle of incidence on the reflecting refracting [25]. Carrier frequency also determines the amount of energy lost to absorption for a particular path. The D-region is the primary attenuator along an ionospheric path, thus its electron density and layer thickness determines the lowest usable frequency often referred to as the Absorption Limiting Frequency (ALF) [5].
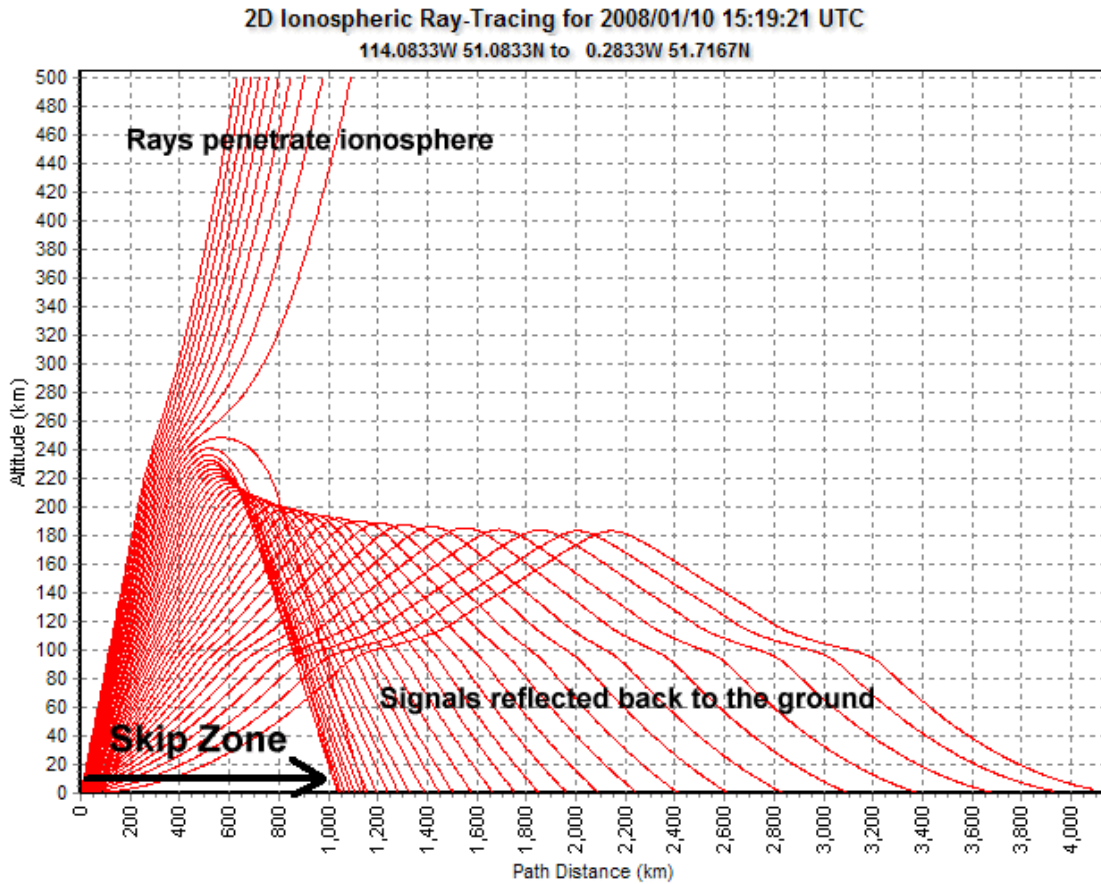
Figure 2-3: Signal paths as a function of incident angle using a fixed carrier frequency. Originally from [38]

The potential paths of a transmission for a fixed frequency is shown in Figure 2-3. Rays corresponding with high elevation angles penetrate through the ionosphere and are not reflected, while low elevation angles are able to propagate a considerable distance before their return to earth. As the elevation angle increases, the ground distance of the corresponding ray decreases until reaching the point where the ele-

vation angle exceeds the maximum angle for the given frequency. The area on the ground where no signal is able to be received is called the *skip-zone.* For antennas with omnidirectional radiation patterns, this leads to an annular coverage pattern that is unique to HF. At ground distances just beyond the skip-zone, there is a collection of rays that have the same end point. The *high-rays* penetrate further into the ionosphere thus have a longer total path length, while the *low-rays* are more immediately reflected, however both rays will combine at the same endpoint. Since the propagation path of the high-ray is longer, the signal will face more attenuation and be more delayed than the lower ray. The bifurcation into high and low rays is not unique to the area just beyond the skip-zone, and may occur at greater ground distances when operating below the maximum useable frequency for a link.
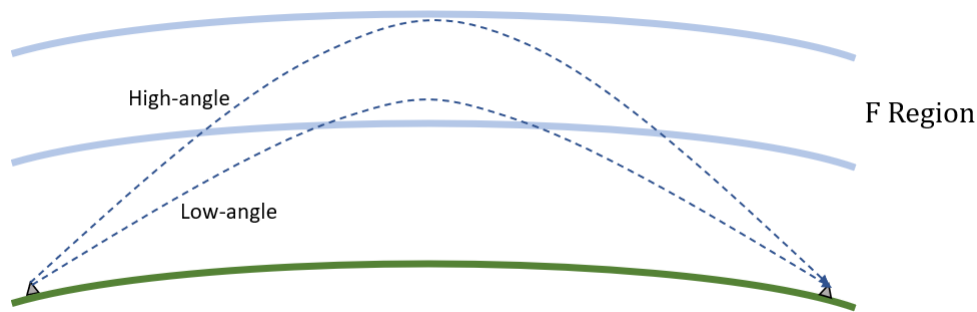
Figure 2-4: Splitting into High and Low Rays

The ray-tracing depictions given so far do not tell the complete story of ionospheric propagation. Earth's magnetic field causes electromagnetic waves to split into two separately polarized components once entering the ionosphere: the ordinary (O) and extraordinary (X) waves. The refractive index of the ionosphere is different for the O and X waves. The O waves takes a slightly longer path through the ionosphere and are subjected its effects for longer. Upon recombination of the O and X waves, the polarization of the combined wave will appear to have been rotated due to different amounts of phase change of the O and X waves through their respective trajectories. For a more in depth coverage of this magneto-ionic splitting and Faraday rotation, readers are referred to [21, Chapter 4]

27

Over long distances a signal may take multiple "hops" off the ground and the ionosphere to reach its destination. The *mode* of propagation refers to the specific layer(s) and bounces of the received signal as shown in 2-5 (modes: 1F, 2F, 2E, etc). Ground distances greater than 4000km, such as across oceans, generally rely on multiple hops off the F1 or F2-region [1].
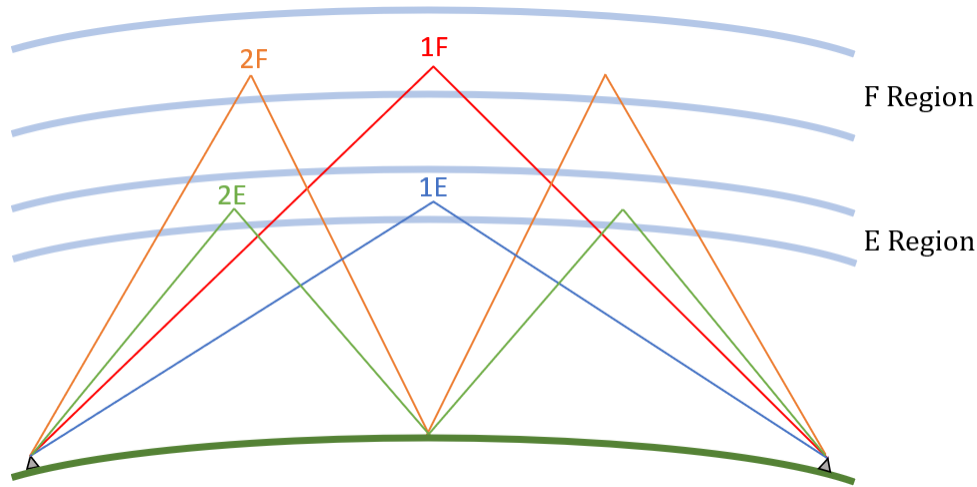
Figure 2-5: Possible single/multi-hop trajectories

Due to the dynamic nature of the ionosphere, propagation prediction for frequency planning/selection has been imperative for the operation of HF links. If a selected transmission frequency is too high given the current conditions (i.e. exceeds the oblique critical frequency) the signal will penetrate through all layers of the ionosphere and not be refracted. If a selected frequency is too low given the current conditions (i.e. is below the ALF) either the D-region attenuates the signal too much to be received or the signal is unable to penetrate the E-region and is refracted back at shorter distances than what is required. Due to this frequency dependent attenuation, it is desirable to operate as close to the oblique critical frequency as possible.

So far all discussions of HF propagation have centered on oblique incidence sky-waves, where the elevation angles are relatively low and link distances are generally long. An alternative method of propagation revolves around transmitting at lower frequencies and very high (near vertical) elevation angles. This method is called

Near-Vertical Incidence Skywaves (NVIS) and has a range of 250km with a small or non-existent skip zone [24]. For this to occur, the transmission frequency must be below the critical frequency of the ionosphere at the site of transmission. NVIS can be used to overcome local terrain barriers (such as mountains) that would interfere with low-angled signals or for broadcasts in a local region.
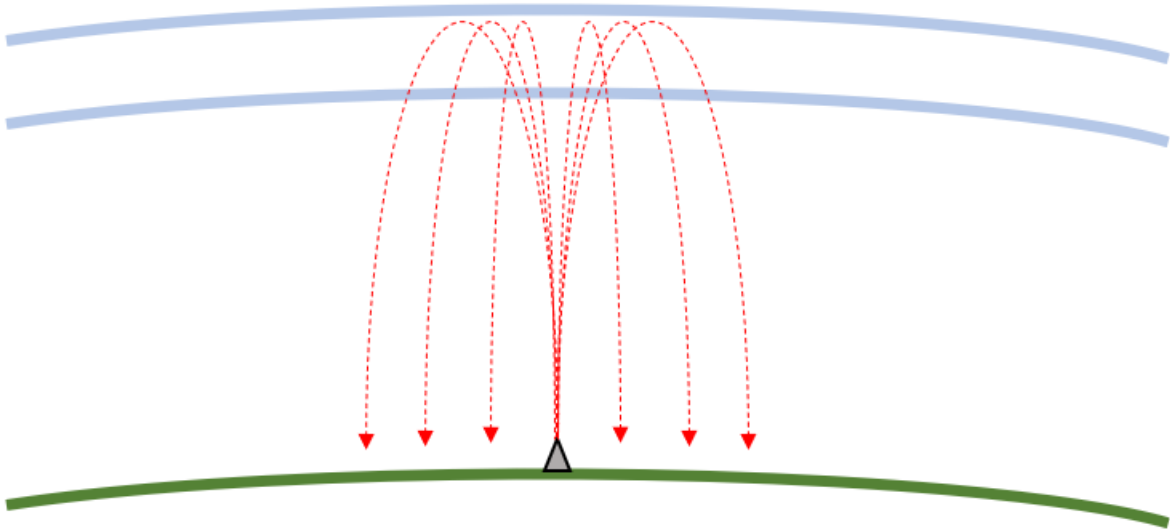
Figure 2-6: NVIS Propagation

2.2.2 Signal Attenuation

For a given path-geometry, there are several attenuation factors whose effects will be compounded on the received signal [18]:

1. Free-Space Path-Loss: Caused by the spatial spread of the signal and can be determined via the following:

$$L_{FS} = (4\pi df/c)^2 \tag{2.3}$$

where d is the total path-distance, f is the carrier frequency, and c is the speed of light.

2. Polarization Loss: Due to the mismatch of the antenna polarization and wave

polarization, the receive antenna may not be able to absorb all of the incoming signal's energy

3. Ionospheric Absorption: Primarily from the D Region, energy will absorbed by the ionosphere. This attenuation is inversely proportional to the carrier frequency.

4. Ground Absorption/Scattering: For multi-hop signal paths, some energy will be lost due to absorption or scattering by the ground

These factors lead to slow variations in received signal strength and are generally well-understood and can be quantified.

2.2.3 Multipath

A single pulse sent from a transmitter over the HF Channel may take multiple trajectories before being received by the receiver. This phenomenon, called *multipath* propagation, may be introduced by many different propagation mechanisms in the HF Channel. Each of these path lengths will differ so the amplitude and phase of the received components will differ. At any one time, the received signal is a superposition of the individual multipath components and for digital communications will cause inter-symbol interference (ISI). Multi-mode multipath occurs when a signal is received from two or more ionospheric layers simultaneously, such as simultaneous 1F and 1E propagation. This also encapsulates multi-hop multipath when a the received signal is a superposition of multiple reflections off ionospheric layers (ex. simultaneous 2E and 1F propagation). A receiver may also see multipath arising from the same propagation mode. Within the same mode, propagation may be comprised of four separate components due to high and low rays as well as magneto-ionic splitting [2]. It is quite possible that multiple of these multipath mechanisms occur simultaneously thus many time delayed versions of a transmission may be received up to milliseconds apart. Without diversity techniques, severe multipath spread can limit the symbol rate in digital communications to several hundred symbols/second to prevent ISI [21, Chapter 4].

2.2.4 Fading

Fading refers to the fluctuation of received signal energy with time. The main causes of fading over the HF channel and their fading period are as follows:

Fading Category	Fade Period
Motion of ionospheric inhomogeneities	< 1 second
Ionospheric motion changing multipath spread	1-10 second(s)
Faraday Effect	0.1-2 minutes
Changes in D-layer absorption (diurnal and disturbance induced)	5-30 minutes
Focusing/defocusing caused by change in layer curvature	10-60 minutes
Variation in ionospheric support for a particular frequency	Aperiodic

Table 2.2: Fading types in the HF Channel, adopted from [18] and [2]

The magnitude of fading may be dependent (selective) or independent (non-selective) of the carrier frequency. Generally fading caused by the reception of multiple modes is frequency selective, while single mode fading is non-selective [18]. The total variation in the received signal over the HF Channel will be due to the combination of fading effects over all timescales superimposed upon the effects of multipath.

2.2.5 HF Channel Modeling

Variation in signal strength is a random process and in the HF Channel has temporal effects over a variety of time-scales. The following models capture variation in the long, short, and intermediate time-scales respectively.

Propagation Predictors

Propagation prediction programs such as ITURHFPROP and VOACAP can be used predict changes in received power over the timescales of an hour and longer. These are based on experimentally derived ionospheric coefficient maps along with propagation models, and provide predictions of monthly median statistics for a given point-to-point HF communication link. The following inputs are generally needed for propagation prediction programs:

31

1. Latitude and longitude of the transmitter and receiver

2. Month and time of day (typically at the granularity of an hour)

3. Smoothed sunspot number or other ionospheric activity index

4. Frequencies of interest

5. System parameters (antenna model, man-made noise at the receiver, transmit power, data rate, etc)

From these inputs the programs can predict metrics at the receiver such as signal strength, SNR, and link reliability as well path-centric metrics such as the maximum useable frequency and viable modes of propagation for each frequency of interest. These metrics however are monthly median predictions meaning at the specific input hour, SSN, and frequencies of interest, the predictions provide the median, lower-, and upper-deciles in which the metric of interest will exceed for 50%, 10%, and 90% of the days of the month respectively [1]. These can be powerful tools for long-term planning, however, they do not provide the granularity required to understand the real-time performance of a communication link. Figures 2-7 through 2-10 demonstrate VOACAP's coverage prediction capabilities. The input parameters were as follows:

1. Location: Washington, DC (38.8771, -77.0409)

2. Time Parameters: August over 00, 06, 12, and 18 UTC

3. SSN: 24

4. Frequency: 10.1 MHz (30m Band)

5. System Parameters: Isotropic Antenna (0 dBi) transmitting at 80W

TX: FM18LV (38.88N, 77.04W) ● Aug, 00 UTC, SSN:24, 10.1 MHz ● 80 W, Mode: CW
TX Ant: V14GD.ANT, -1.0˚, RX Ants: D10M.ANT. Noise: -153 dBW
Made in www.voacap.com, 2021-08-03
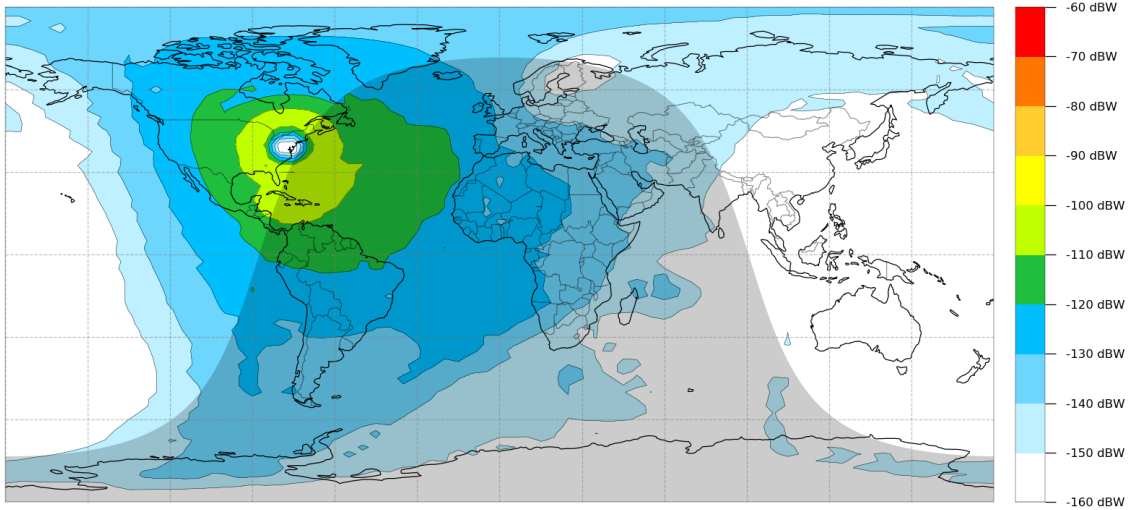
Figure 2-7: VOACAP received power coverage prediction for 00 UTC

TX: FM18LV (38.88N, 77.04W) ● Aug, 06 UTC, SSN:24, 10.1 MHz ● 80 W, Mode: CW
TX Ant: V14GD.ANT, -1.0˚, RX Ants: D10M.ANT. Noise: -153 dBW
Made in www.voacap.com, 2021-08-03
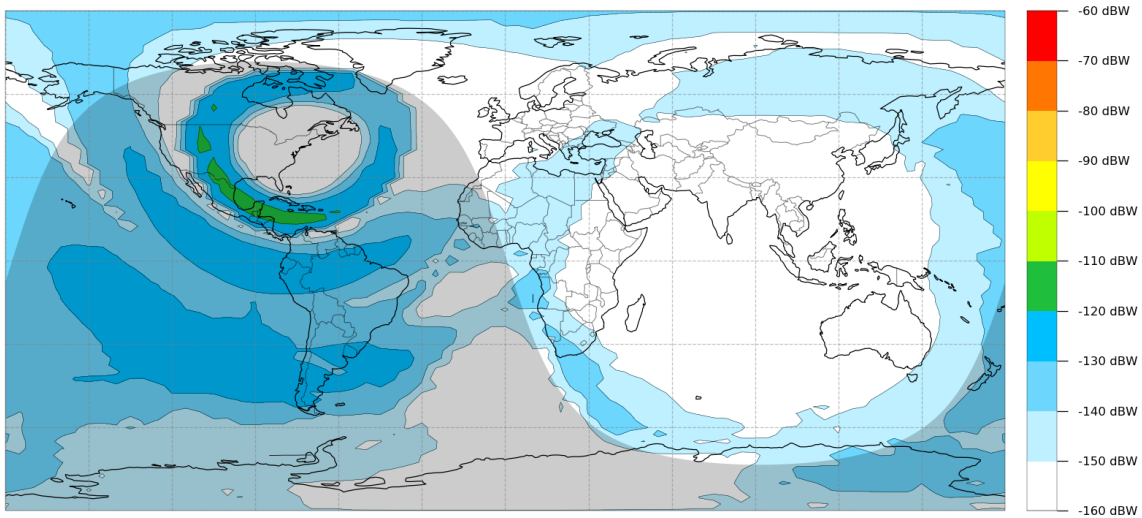
Figure 2-8: VOACAP received power coverage prediction for 06 UTC

33

TX: FM18LV (38.88N, 77.04W) ● Aug, 12 UTC, SSN:24, 10.1 MHz ● 80 W, Mode: CW
TX Ant: V14GD.ANT, -1.0˚, RX Ants: D10M.ANT. Noise: -153 dBW
Made in www.voacap.com, 2021-08-03
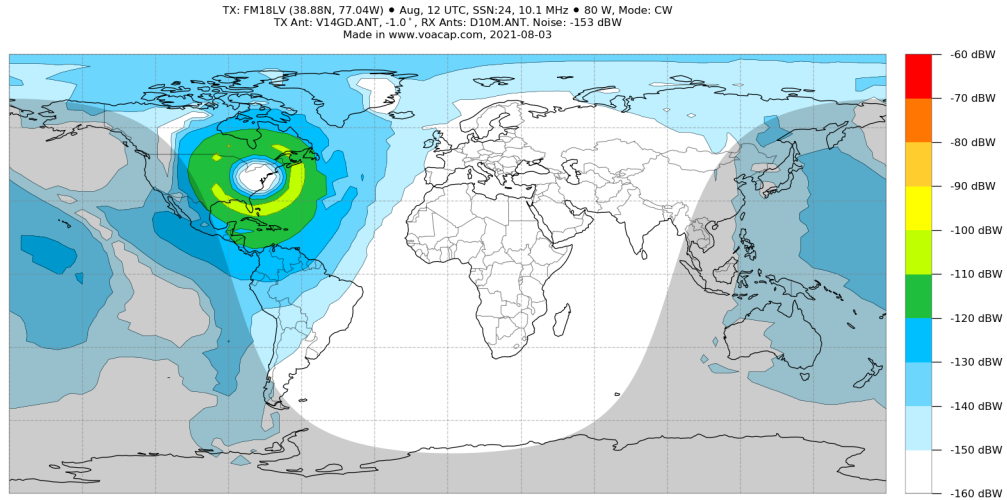
Figure 2-9: VOACAP received power coverage prediction for 12 UTC

TX: FM18LV (38.88N, 77.04W) ● Aug, 18 UTC, SSN:24, 10.1 MHz ● 80 W, Mode: CW
TX Ant: V14GD.ANT, -1.0˚, RX Ants: D10M.ANT. Noise: -153 dBW
Made in www.voacap.com, 2021-08-03
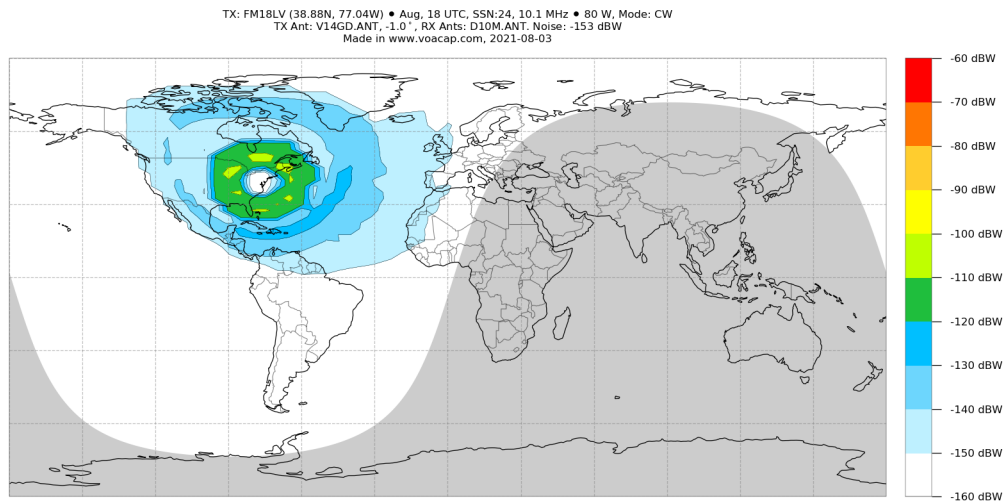
Figure 2-10: VOACAP received power coverage prediction for 18 UTC

In each of the figures, the shaded and unshaded regions correspond to regions of night and day respectively. These figures capture the dynamic nature of the iono-spheric channel as there are no locations with consistent received signal strengths throughout the day at this single frequency.

Watterson Model

In contrast to the long-term statistical behavior captured by propagation prediction programs, the Watterson Model captures the fine-grained behavior of ionospheric paths on the scale of micro- to milliseconds. The Watterson model assumes the ionosphere can be modeled as stationary over a short-interval ($\sim$ 10 minutes) for a band-limited ($\sim$ 10kHz) channel. The base model represents the HF Channel using an ideal tap delay line, where each tap corresponds to a resolvable propagation path. Each tap gain function models the received signal for the magneto-ionic components of the particular path as:

$$G_i(t) = G_{X,i} \exp\left(j2\pi\theta_{X,i}t\right) + G_{O,i} \exp\left(j2\pi\theta_{O,i}t\right) \tag{2.4}$$

where the $G_{\alpha,i}$ terms are sample functions of two complex independent Gaussian processes, and the exponential allow the addition of Doppler shifts for each component [47].

Midterm Variation Model

Variation on the order of seconds to minutes is not captured by either the Watterson Model nor propagation prediction programs. This time-scale is important in the design, analysis, and performance evaluation of an HF Network as networking protocols are sensitive to the variation in the ionosphere at this scale [25]. Mid-term variations arise in the HF Channel due to ionospheric motion, Faraday fading, the focusing effect. Frequency domain analysis of midterm fading measurements have found that variation over this timescale can be subdivided into two distinct normal processes in the decibel scale. The component with the longest period called the *Long-Term Variation* (LTV) has a fade interval of approximately 3 minutes while the *Intermediate-Term Variation* (ITV) component has a fade interval of approximately 5 seconds. The standard deviation of each component ranges from 1-4dB throughout the day [6].

2.3 Assumptions

In our MVN model, we model the received signal strengths in an HF network as normally distributed and correlated. The following section describes the justification for these two assumptions.

2.3.1 Log-normal Shadowing

A widely used model in the wireless communication literature is the log-path loss model [42]. This model captures distance dependent attenuation in environments with a high density of obstructions such as inside buildings or within cities . In this model the average received signal strength in dB ($P_{Rx}[dB]$) is:

$$P_{Rx}[dB] = P_{Tx}[dB] + PL_0 + 10\gamma \log_{10}(\frac{d}{d_0}) + L_G \qquad (2.5)$$

where $P_{Tx}[dB]$ is the transmit power in dB, d_0 is a short reference distance from the transmitter, PL_0 is the path-loss at reference distance d_0, γ is an experimentally derived path-loss exponent, d is the distance between the transmitter and the receiver, and L_g is zero-mean Gaussian random variable. The Gaussian log-fading component was originally based on empirical measurements in cluttered environments [46], but has statistical foundations as well: in a cluttered environment there will be multiple rays between transmitter and receiver, and if the attenuation on each ray is modeled as a random variable, then by the Central Limit Theorem, the sum of the power in each ray at the receiver will converge to the normal distribution [16].

A similar justification can be said for using log-normal fading to model large scale fading in the HF channel. In the HF channel a radiowave will take multiple trajectories through the inhomogenous ionosphere before reaching the receiver. The overall attenuation on each ray will be a combination of losses due to diffraction, scattering, absorption, etc. all of which can be modeled as random variables. Thus by the Central Limit Theorem, the mean received power between a transmitter and receiver can be modeled as normally distributed about a distance dependent mean.

2.3.2   Spatial Correlation of the HF Channels

As stated in Chapter 1, we are interested in modeling and analyzing HF networks comprised of multiple point-to-point HF communication links. In traditional line-of-sight wireless networks it can usually be assumed that the channel parameters for each link are static and independent from one another. There are scenarios in traditional wireless networks where these assumptions do not hold such as in the event of rain [50], [13], [23], or shadowing by environmental obstructions [3]. However, in these examples spatial correlation is induced by the presence of a disturbance in the propagation medium. Analogously, channel parameters of links in an HF Network would exhibit spatial correlation if links crossed through regions effected by one of the events listed in Table 2.1.

However, even under benign ionospheric conditions HF channel parameters would exhibit correlation induced by the spatial correlation of the propagation medium itself. Numerous experimental studies have demonstrated the spatial correlation of ionospheric characteristics such as the F2 critical frequency [32], [44] and the total electron content [51]. The spatial relationship of signal strengths may also be seen in Figures 2-7 through 2-10, however it can be observed that this relationship is also time dependent. The correlation between communication links in an HF Network will depend on the spatial separation of links' ionospheric paths, thus correlation would coincide with the spatial separation of their control points [22]. For a one-hop mode, this is assumed to lie directly above the midpoint between the transmitter and receiver [1]. Channel sounding experiments conducted by Goodman and colleagues have shown that correlation in the HF propagation environment may decrease the reliability of the overall network [20].

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 3

The MVN Reliability Model

The following chapter introduces the Network Reliability Problem and the Multivariate Normal (MVN) Model. The traditional Bernoulli failure model that is often used in Network Reliability literature is described in Section 3.2 and our novel MVN Model formulation is given Section 3.3. Section 3.4 describes the various Network Reliability problems and definitions and Section 3.5 presents methods to solve them under the MVN Model.

3.1 Bernoulli Model

Traditionally, probabilistic graphs $\mathbf{G} = (\mathbf{V}, \mathbf{E}, \mathbf{p})$ have been used to model communication networks with random link failures. The nodes $\mathbf{V} = [v_1, \ldots, v_n]$ represent the n communication centers or relays and it is typically assumed they do not fail. The m undirected edges $E = [e_1, \ldots, e_m]$ of the graph represent the links between pairs of nodes $e_i \in (v_j, v_k)$ and can be in one of two states: either operational or failed. The state of link e_i is denoted by the random variable X_i, where $\{X_i = 1\}$ denotes the event e_i is in the operational state, while $\{X_i = 0\}$ denotes the event e_i has failed. For more concise probabilistic expressions, the operational link state may also be denoted by x_i and the failed state x'_i in the following sections. In the Bernoulli Model, the probability link e_i is operational is p_i while the probability link e_i has failed is $1 - p_i = q_i$

$$\mathbb{P}(\{X_i = 1\}) = \mathbb{P}(x_i) = p_i \tag{3.1}$$

$$\mathbb{P}(\{X_i = 0\}) = \mathbb{P}(x_i') = 1 - p_i = q_i \tag{3.2}$$

Links may also be modeled using directed edges or arcs $\mathbf{A} = [a_{ij}]$ where their connectivity is not assumed to be bidirectional, the probability link a_{ij} is operational p_{ij} may not equivalent to the probability link a_{ji} is operational p_{ji}.

In the Bernoulli model it is assumed that link failures are independent, meaning the state of link e_i does not affect the state probability of any other link e_j:

$$\mathbb{P}(X_i | X_j) = \mathbb{P}(X_i) \ \forall \ i, j \tag{3.3}$$

As a result, the probability of the event that both link e_i and e_j are both operational, or their *joint reliability*, is given as the product of their marginal reliabilities:

$$\mathbb{P}(x_i \cap x_j) = \mathbb{P}(x_i)\mathbb{P}(x_j) = p_i p_j \ \forall \ i, j \tag{3.4}$$

3.2 MVN Model Formulation

Inspired by the spatial correlation and the normally distributed signal distribution expected in an HF Network we developed the following model.

Let the network be modeled as a graph $\mathbf{G} = (\mathbf{V}, \mathbf{E}, \mathbf{s}, \boldsymbol{\tau})$. Associated with each link $e_i \in \mathbf{E}$ is a normally distributed random variable $s_i \sim N(\mu_i, \sigma_i^2)$ and a fixed threshold τ_i. The random variable s_i represents the signal strength (or similar metric such as SNR) for link e_i, while τ_i is a fixed value that represents the minimum signal strength required to meet some performance criteria. If the signal is below this threshold then the link is considered to be in a failed state. It's also assumed that the vector $\mathbf{s} = [s_1, \ldots, s_m]$ is jointly normally distributed with mean vector $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_m]$ and positive-definite covariance matrix Σ, where:

$$\Sigma(i,j) = \begin{cases} \sigma_i^2 & \text{if } i = j \\ Cov(s_i, s_j) & \text{if } i \neq j \end{cases} \tag{3.5}$$

The marginal and joint probability density functions are given as:

$$f_{s_i}(s; \mu_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(s-\mu_i)^2}{2\sigma^2}\right) \tag{3.6}$$

$$f_{\mathbf{s}}(\mathbf{s}; \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{\frac{m}{2}}|\Sigma|} exp\left(-\frac{1}{2}(\mathbf{s}-\boldsymbol{\mu})^\mathsf{T}\Sigma(\mathbf{s}-\boldsymbol{\mu})\right) \tag{3.7}$$

which correspond to a univariate and multivariate normal distribution respectively.

Like the Bernoulli model, marginal link reliabilities can be given as a fixed value, however, they are based on evaluating the corresponding univariate normal integral for that link:

$$\mathbb{P}(x_i) = \mathbb{P}(s_i \geq \tau_i) = \int_{\tau_i}^{\infty} f_{s_i}(s; \mu_i, \sigma_i^2) ds = p_i \tag{3.8}$$

The key difference between the Bernoulli and MVN model lies in the joint probabilities of link states for two or more links. For example, for two links e_i and e_j with their signal covariance $\sigma_{i,j} \neq 0$:

$$\mathbb{P}(\{s_i \geq \tau_i\} \cap \{s_j \geq \tau_j\}) = \int_{\tau_i}^{\infty} \int_{\tau_j}^{\infty} \frac{1}{2\pi|\Sigma|} exp\left(-\frac{1}{2}(\mathbf{s}-\boldsymbol{\mu})^\mathsf{T}\Sigma(\mathbf{s}-\boldsymbol{\mu})\right) ds_j ds_i$$

where:

$$\boldsymbol{\mu} = [\mu_i, \mu_j]^\mathsf{T}$$

$$\Sigma = \begin{bmatrix} \sigma_i^2 & \sigma_{i,j} \\ \sigma_{i,j} & \sigma_j^2 \end{bmatrix}$$

Similarly, the probability link e_i has failed and e_j is operation is given as:

$$\mathbb{P}(\{s_i < \tau_i\} \cap \{s_j \geq \tau_j\}) = \int_{-\infty}^{\tau_i} \int_{\tau_j}^{\infty} \frac{1}{2\pi|\Sigma|} exp\left(-\frac{1}{2}(\mathbf{s}-\boldsymbol{\mu})^\mathsf{T}\Sigma(\mathbf{s}-\boldsymbol{\mu})\right) ds_j ds_i$$

where the mean and covariance parameters remain unchanged. If their covariance was zero, $\sigma_{i,j} = 0$, then s_i and s_j would be statistically independent and their joint reliability could be evaluated as the product of their marginals as in the Bernoulli model.

This difference in the joint probabilities also changes the expression for conditional reliabilities as:

$$\mathbb{P}(X_i|X_j) = \frac{\mathbb{P}(X_i \cap X_j)}{\mathbb{P}(X_j)} \tag{3.9}$$

which does not simplify to the univariate probability $\mathbb{P}(X_i)$ unless their correlation $\rho_{ij} = 0$.

Since there are still only two possible states for each link, operational (x_i) or failed (x_i'), the state of $\mathbf{G}$ can described by the random state vector $\mathbf{X} \in \mathcal{X}$, where $|\mathcal{X}| = 2^m$. Any graph state $\mathbf{X}$ can be partitioned into two sets: the operational set $\mathbf{x} = [X_i : X_i = x_i]$ and the failed set $\mathbf{x}' = [X_i : X_i = x_i']$.

The probability of the event $\mathbf{G}$ is in state $\mathbf{X} = [X_1, ..., X_m]$ is:

$$\mathbb{P}(\mathbf{X}) = \mathbb{P}\left(\bigcap_{i=1}^{m} X_i)\right) \tag{3.10}$$

which under the MVN model becomes:

$$\mathbb{P}(\mathbf{X}) = \mathbb{P}\left(\bigcap_{i=1}^{m}\{s_i \underset{x_i'}{\overset{x_i}{\gtrless}} \tau_i\})\right) \tag{3.11}$$

In the Bernoulli model (or the uncorrelated MVN case) this simplifies to:

$$\mathbb{P}(\mathbf{X}) = \prod_{i=1}^{m} \mathbb{P}(X_i) \tag{3.12}$$

3.3 Network Reliability

Network reliability is defined as the probability of the event the network is in an "operational" state. Whether or not a graph state $\mathbf{X}$ is operational or not will depend on the network's intended function, but we will focus on three of the most common

definitions: two-terminal, k-terminal, and all-terminal reliability. These definitions and examples as applied to the graph depicted in Figure 3-1 are given below along with the definitions of important graph substructures for reliability.
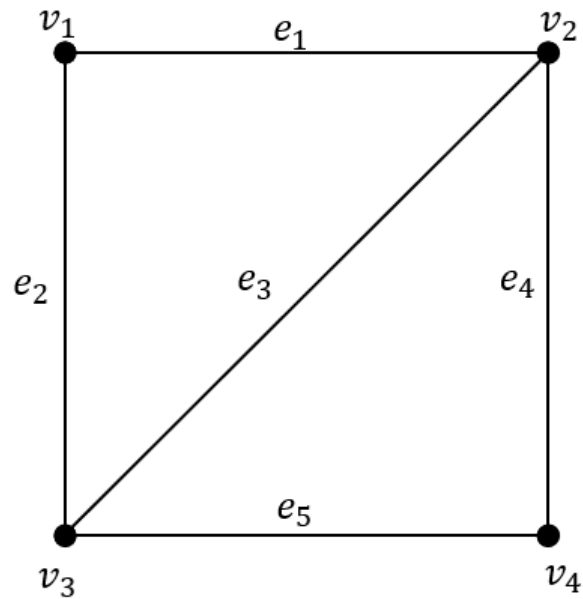
Figure 3-1: Bridge Graph

Definitions:

Pathset (or Path) - PS: A sequence of edges which join a sequence of distinct vertices

Ex. (e_2, e_5) and (e_1, e_3, e_5) are paths between nodes v_1 and v_4

Cutset(or Cut) - CS: A subset of edges such that their removal from the graph partitions it into two (or more) separate subgraphs.

Ex. (e_1, e_2), (e_2, e_3, e_5), and (e_1, e_3, e_4, e_5) are all cuts

Minimal Cut(or mincut) - C: A cut set such that the removal of any link from the set would make the remaining links no longer a cut set. Every cutset contains one or more minimal cuts.

Ex. (e_1, e_2) and (e_2, e_3, e_5) are both mincuts, while (e_1, e_3, e_4, e_5) is not but contains the mincut (e_1, e_3, e_4)

Two-terminal Reliability - $\mathbf{Rel_{st}(G)}$: The probability of the event that for a given

43

node pair $(v_s, v_t) \in \mathbf{V}$ there is at least one path connecting the two nodes.

K-terminal Reliability - $\mathbf{Rel_K(G)}$: The probability of the event that there is at least one path between k specified nodes

All-terminal Reliability - $\mathbf{Rel_A(G)}$: The probability the event that for every node pair $(v_i, v_j) \in \mathbf{V}$ there is at least one path connecting the two nodes.

When applied to directed graphs, these definitions may be referred to as all-terminal/two-terminal/k-terminal *connectivity* instead of reliability.

For all definitions of reliability, we can define a structure function $\Psi(\cdot)$ which indicates when a graph's state $\mathbf{X} = [X_1, \dots, X_m]$ is operational based on the specified network operation definition. For example, for two-terminal reliability:

$$\Psi_{st}(\mathbf{X}) = \begin{cases} 1 & \text{if a path exists between nodes } v_s \text{ and } v_t \\ 0 & \text{otherwise} \end{cases} \tag{3.13}$$

3.4   Evaluating Network Reliability

All definitions of reliability may be expressed as the probability its structure function evaluates to one:

$$Rel(\mathbf{G}) = \mathbb{P}(\{\Psi(\mathbf{X}) = 1\}) = \sum_{\mathbf{X} \in \mathcal{X}} \mathbb{P}(\mathbf{X})\Psi(\mathbf{X}) \tag{3.14}$$

where each term of the summation is non-zero only if $\Psi(\mathbf{X})$ for that particular state evaluates to one. The set of all possible graph states $\mathcal{X}$ contains the $|\mathcal{X}| = 2^m$ possible values of the random Bernoulli vector $\mathbf{X} = [X_i]_{i=1}^m$. The naive evaluation of the above expression requires the enumeration of all 2^m states and thus quickly becomes prohibitive for large networks.

All reliability expressions can also be evaluated in terms of their minpaths or mincuts. Letting $\{P_i\}$ denotes the event that all links the ith minpath are operational and $\{C_i\}$ denotes the event that all links the ith mincut have failed:

$$Rel(\mathbf{G}) = \mathbb{P}(\{\mathbf{X} \text{ contains a minpath}\}) = \mathbb{P}(\bigcup_{P_i \in \mathcal{P}} \{P_i\}) \tag{3.15}$$

$$= 1 - \mathbb{P}(\{\mathbf{X} \text{ contains a mincut}\}) = 1 - \mathbb{P}(\bigcup_{C_i \in \mathcal{C}} \{C_i\}) \tag{3.16}$$

Where the probability of a minpath $\mathbb{P}(\{P_i\})$ is the joint probability of all links in the minpath:

$$\mathbb{P}(\{P_i\}) = \mathbb{P}(\bigcap_{j:e_j \in P_i} x_j) \tag{3.17}$$

Both 5.1 and 3.16 are expressed in terms of unions of non-disjoint events. This contrasts Equation 3.14 as only one graph state may exist at a single time thus there is no overlap in their event space and the individual graph states are mutually exclusive. Meanwhile it is possible (and often common) for multiple overlapping minpaths or mincuts to exists in a single graph state thus they are not mutually exclusive sets. To evaluate reliability using minpaths, the inclusion-exclusion principal must be used:

$$\mathbb{P}\left(\bigcup_{i=1}^{K}\{P_i\}\right) = \sum_{i=1}^{K}\mathbb{P}\{P_i\} - \sum_{i<j}^{K}\mathbb{P}(\{P_i\} \cap \{P_j\}) + \sum_{i<j<k}^{K}\mathbb{P}(\{P_i\} \cap \{P_j\} \cap \{P_k\}) +$$
$$... + (-1)^{K-1}\sum_{i<...<K}\mathbb{P}\left(\bigcap_{k=1}^{K}\{P_k\}\right)$$

$$\tag{3.18}$$

Even when links are independent, there are two main issues with the naive implementation of this approach:

1. Generating all paths/mincuts requires exponential time since the number of paths/mincuts may be exponential in $|V|$.

2. Generating all subsets of the minpaths/mincuts can take 2^h time where h is the number of paths/mincuts.

Thus the naive implementation is a doubly exponential algorithm [14].
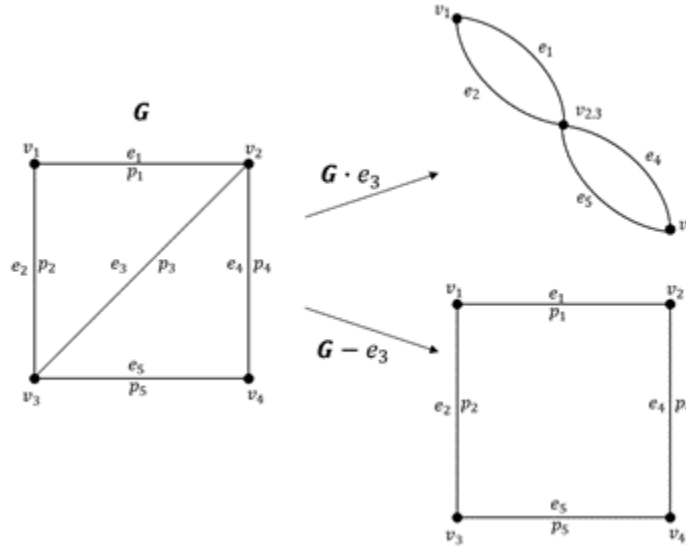
Figure 3-2: Edge Contraction $\mathbf{G} \cdot e_3$ and deletion $\mathbf{G} - e_3$ of e_3

3.4.1 Factoring Method

As will be explained in the following chapter, there are many methods of evaluating/approximating reliability under the Bernoulli Model that do not work under the MVN due to correlation. One method that does have applicability to both the Bernoulli and the MVN model is the Factoring Method. First described by [33], this approach relies on repeatedly applying a pair of operations called edge contractions and edge deletions. Contracting e_i from a graph $(\mathbf{G} \cdot e_i)$ merges the two endnodes of e_i into a single node and removes e_i along with any parallel edges. Deleting e_i from a graph $(\mathbf{G} - e_i)$ simply removes e_i from the graph without effecting any other nodes or edges. These operations are demonstrated in Figure 3-2

The reliability of a graph can be expressed as the weighted sum of the reliabilities of the subgraphs created via the contraction/deletion process.

Theorem 3.4.1 (Factoring Theorem [14]). *For a given network* $\mathbf{G}$ *with independent link failures:*

$$Rel(\mathbf{G}) = Rel(\mathbf{G} \cdot e_i)\mathbb{P}(x_i) + Rel(\mathbf{G} - e_i)\mathbb{P}(x_i') \tag{3.19}$$

This theorem can be reapplied to subsequent subgraphs $G \cdot e_i$ and $G - e_i$, leading to a recursive algorithm to compute the reliability of any graph. The factoring

method can equivalently be thought of as a conditioning process, where edge contraction is equivalent to conditioning on a link being in the operational state, and edge deletion is equivalent to conditioning on the event a link is in the failed state. In the Bernoulli Model, this has no effect on the reliability of subgraphs as all links are independent thus the probability link e_j fails in subgraph $\mathbf{G} \cdot e_i$ is q_j. However, when links are correlated the conditional distributions must be considered when computing the reliabilities of subgraphs. This leads to the following Correlated Factoring Theorem:

Theorem 3.4.2 (Correlated Factoring Theorem). *For a given network* $\mathbf{G}$*, let the set* $\mathbf{x}$ *correspond to the links conditioned to be operational/contracted in the original graph* $\mathbf{G}$ *and the set* $\mathbf{x}'$ *correspond to the links conditioned to be failed/deleted in the original graph* $\mathbf{G}$*.*

$$Rel(\mathbf{G}|\mathbf{x} \cap \mathbf{x}') = Rel(\mathbf{G}|x_i \cap \mathbf{x} \cap \mathbf{x}')\mathbb{P}(x_i|\mathbf{x} \cap \mathbf{x}') + Rel(\mathbf{G}|x_i' \cap \mathbf{x} \cap \mathbf{x}')\mathbb{P}(x_i'|\mathbf{x} \cap \mathbf{x}') \quad (3.20)$$

When the sets $\mathbf{x}$ and $\mathbf{x}'$ are empty this simplifies to:

$$Rel(\mathbf{G}) = Rel(\mathbf{G}|x_i)\mathbb{P}(x_i) + Rel(\mathbf{G}|x_i')\mathbb{P}(x_i') \quad (3.21)$$

which is identical to Theorem 3.4.1

For a simple proof of the validity of these theorems we use the fact that reliability is simply the probability of the event $\mathbf{G}$ is operational and the fact that the event space of link e_i, $\mathcal{X}_i$, is confined to either an operating x_i or failed x_i' state which are disjoint event spaces. Thus by the total law of probability:

$$\mathbb{P}(\mathbf{G} \text{ is operational}) = \sum_{X_i \in \mathcal{X}_i} \mathbb{P}(\mathbf{G} \text{ is operational}|X_i)\mathbb{P}(X_i) \quad (3.22)$$

47

which is equivalently:

$$Rel(\mathbf{G}) = Rel(\mathbf{G}|x_i)\mathbb{P}(x_i) + Rel(\mathbf{G}|x_i')\mathbb{P}(x_i') \tag{3.23}$$

$$= Rel(\mathbf{G} \cdot e_i)\mathbb{P}(x_i) + Rel(\mathbf{G} - e_i)\mathbb{P}(x_i') \tag{3.24}$$

$$\tag{3.25}$$

Note that Theorem 3.4.2 is not dependent on the underlying signal distribution being a multivariate Gaussian and would still valid true under a generally probabilistic framework for reliability that incorporates correlation. When applied to the MVN model, the link signal distribution $\mathbf{s}$ becomes a truncated multivariate normal distribution after any contraction or deletion. For example, after the contraction operation $\mathbf{G} \cdot e_i$, the signal distribution for the rest of the links would be given by:

$$f_{\mathbf{s}|x_i}(\mathbf{s}|x_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{f_{\mathbf{s}}(\mathbf{s})\mathbb{I}(s_i \geq \tau_i)}{\mathbb{P}(s_i \geq \tau_i)} \tag{3.26}$$

where $f_{\mathbf{s}}$ is the original signal distribution prior to conditioning. Like the MVN distribution, probabilities associated with the truncated MVN distribution do not have closed forms but can be evaluated numerically.

Chapter 4

Approximating MVN Reliability

Due to the NP-Completeness of Network Reliability problems, most of the literature concerning the Bernoulli model revolves around estimating the reliability instead of exact computations. Simulation based methods such as Monte Carlo methods are often used in the Network Reliability literature when the size of the network makes computing the exact reliability infeasible. Even for the Bernoulli Model there is not a universal algorithm that's optimal for all graphs and class of reliability problems. In this chapter, two different simulation methods are presented to estimate the unreliability.

4.1 Problem Abstraction

All Network reliability problems may be abstracted to the following: We have a continuous sample space $\mathcal{S} \in \mathbb{R}^m$ that contains all possible realizations of our signal vector $\mathbf{s}$. We have a function $I : \mathcal{S} \to \mathcal{X}$, where $\mathcal{X}$ is subdivided 2^m disjoint subspaces $(\mathcal{X}_1, ..., \mathcal{X}_{2^m})$ corresponding to all the possible link states of $\mathbf{G}$. The function I is the evaluation of $\mathbb{I}(s_i > \tau_i) \ \forall \ i \in 1, .., m$. The weight of any $\mathcal{X}_i$ can only be computed by integrating the region of $\mathcal{S}$ that maps to $\mathcal{X}_i$ under function I

$$\mathcal{W}(\mathcal{X}_i) = \int\limits_{s:I(s)\to\mathcal{X}_i} f_{\mathbf{s}}(s)d\mathbf{s} \tag{4.1}$$

where $f_{\mathbf{s}}$ is the probability distribution of $\mathbf{s}$. All $\mathcal{X}_i$ are disjoint therefore the sum of their weights is equivalent to the sum of all weights in $\mathcal{S}$:

$$\sum_i \mathcal{W}(\mathcal{X}_i) = \mathcal{W}(\mathcal{S}) = 1 \tag{4.2}$$

We also have another function $\Psi : \mathcal{X} \to \Gamma$, where Γ is comprised of two disjoint subsets $\mathcal{C}$ and $\mathcal{C}'$ corresponding to connected and disconnected states under the structure function Ψ.The weight of $\mathcal{C}$ is the sum of weights of $\mathcal{X}_i$ such that $\mathcal{X}_i$ maps to $\mathcal{C}$ under Ψ:

$$\mathcal{W}(\mathcal{C}) = \sum_{i:\Psi(\mathcal{X}_i)\to\mathcal{C}} \mathcal{W}(\mathcal{X}_i) \tag{4.3}$$

Similarly for $\mathcal{C}'$

$$\mathcal{W}(\mathcal{C}') = \sum_{i:\Psi(\mathcal{X}_i)\to\mathcal{C}'} \mathcal{W}(\mathcal{X}_i) \tag{4.4}$$

And because Γ is the union of the two disjoint subsets $\mathcal{C}$ and $\mathcal{C}'$:

$$\mathcal{W}(\mathcal{C}') = 1 - \mathcal{W}(\mathcal{C}) \tag{4.5}$$

The goal of Network Reliability approximations is to estimate the weight of $\mathcal{C}$ (the reliability) or $\mathcal{C}'$ (the unreliability), which can also be defined by estimating the weight of the region of $\mathcal{S}$ that maps to $\mathcal{C}$ or $\mathcal{C}'$, referred to as $\mathcal{S}_{\mathcal{C}}$ and $\mathcal{S}_{\mathcal{C}'}$ respectively.

The following properties make this a challenging problem:

1. The functions I and Ψ are non-invertible so the relationship between $\mathcal{S}$ and Γ may only be computed by mapping $\mathcal{S} \to \mathcal{X} \to \Gamma$

2. The space of $\mathcal{S}$ scales with m thus may be very large thus it may be infeasible to map $\mathcal{S} \to \mathcal{X} \to \Gamma$ for the entire space of $\mathcal{S}$

3. The number of subspaces $\mathcal{X}_i$ grows with 2^m

4. The weight of $\mathcal{C}'$ might be very small, meaning it is challenging to sample s that will map to $\mathcal{C}'$
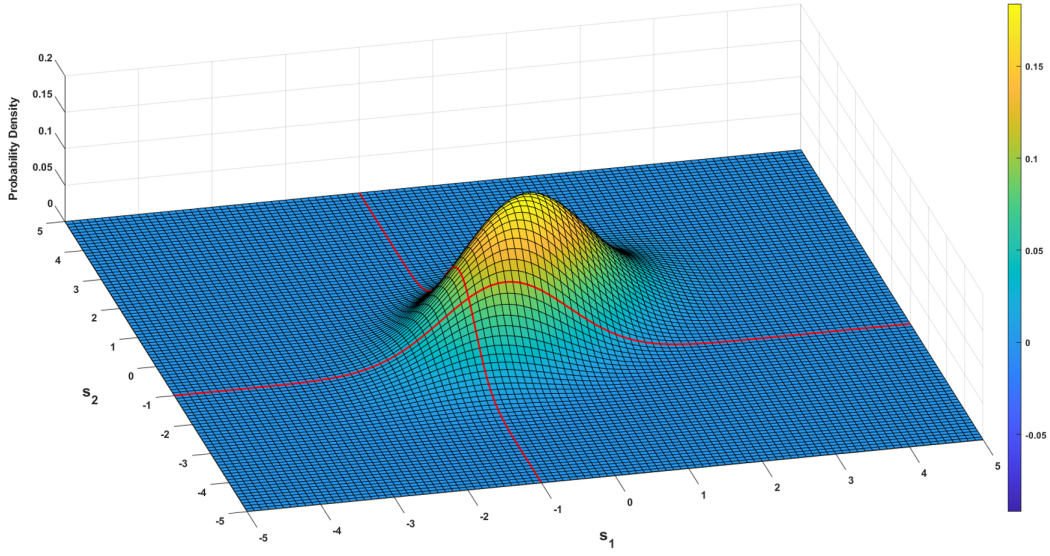
Figure 4-1: Probability density function of $f_{\mathbf{s}}(\mathbf{s}; \mu, \boldsymbol{\Sigma})$ with red lines depicting τ_1 and τ_2

Item four corresponds to estimating the unreliability in a reliable network, and the inverse problem of estimating reliability in an unreliable network is just as hard for the same reason. The problems of estimating $\mathcal{W}(\mathcal{C}')$ as $\mathcal{W}(\mathcal{C}') \to 0$ is referred to as the rare-event unreliability problem and is often a problem of interest.

Say we are given a small network with the signal distribution of $\mathbf{s} = [s_1, s_2] \sim \mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with the following parameters:

$$\boldsymbol{\mu} = [0, 0]^{\mathsf{T}}$$

$$\Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

and the thresholds $\boldsymbol{\tau} = [-1, -1]^{\mathsf{T}}$.

Figure 4-1 depicts the probability density function in three-dimensions while Figure 4-2 depicts its contour plot. The corners of Figure 4-2 also label the regions of $\mathcal{S}$
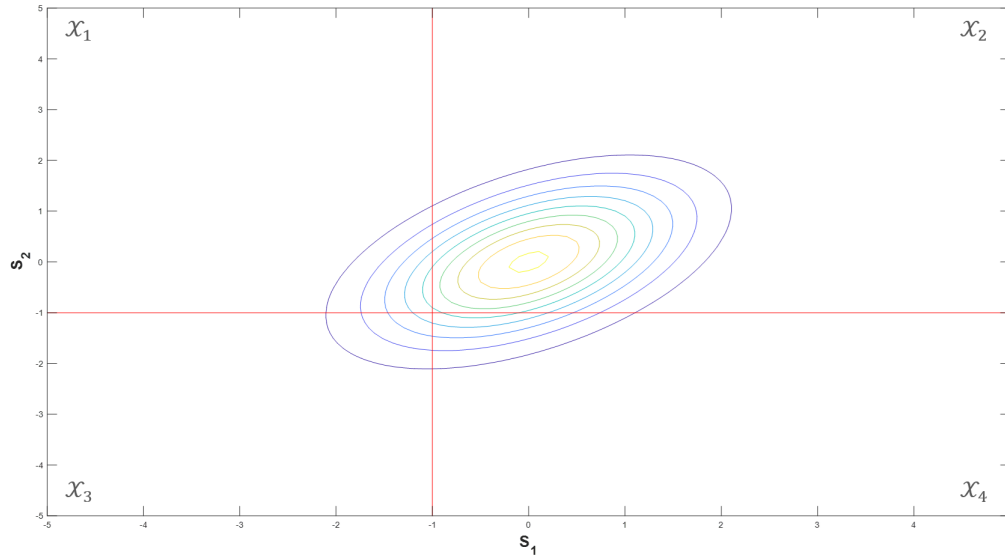
Figure 4-2: Contour plot for Figure 4-1

that correspond to each $\mathcal{X}_i \in \mathcal{X}$ based on the function I.

$$\mathcal{X}_1 = [x_1', x_2]$$

$$\mathcal{X}_2 = [x_1, x_2]$$

$$\mathcal{X}_3 = [x_1', x_2']$$

$$\mathcal{X}_4 = [x_1, x_2']$$

Now the mapping of $\mathcal{X}$ to Γ depends on the network topology. There are two possible configurations given that $\mathcal{S} \in \mathbb{R}^2$: $\mathbf{G}_1$ where e_1 and e_2 are in series, and $\mathbf{G}_2$ where e_1 and e_2 are in parallel both of which are shown in 4-3

If we are concerned with estimating the all-terminal unreliability then Ψ maps $\mathcal{X}_1 \to \mathcal{C}$ and $\{\mathcal{X}_2, \ \mathcal{X}_3 \ \mathcal{X}_4\} \to \mathcal{C}'$ for $\mathbf{G}_1$, and $\{\mathcal{X}_1, \ \mathcal{X}_2 \ \mathcal{X}_4\} \to \mathcal{C}$ and $\mathcal{X}_3 \to \mathcal{C}'$ for $\mathbf{G}_2$. To compute the unreliability for either of these topologies, we would need to compute the weights of all $\mathcal{X}_i$ mapped to $\mathcal{C}'$ using equation 4.1, and then sum these weights. In this example, the number of link states is small thus the exact computation of these weights is feasible. However, due to the number of states increasing exponentially with the number of links, approximation methods will be more efficient exact computation in situations outside of trivial examples. The following chapter provides the basic
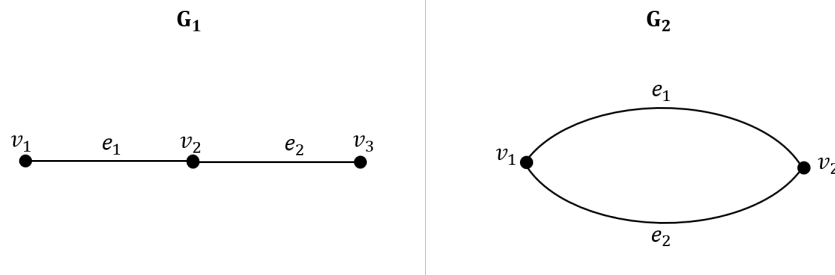
Figure 4-3: Possible graphs for the bivariate normal distribution - $\mathbf{G}_1$: Series graph; $\mathbf{G}_2$: Parallel graph

theory on estimators needed for understanding the efficiency of the approximation algorithms.

4.2 Estimation

Suppose $\hat{\theta}$ is an estimation of some quantity of interest θ. If $Z_1,..., Z_K$ are independent samples of a random variable Z with expected value $E[Z] = \theta$, then an unbiased estimator of θ is:

$$\hat{\theta} = \frac{1}{K} \sum_{i=1}^{K} Z_i \tag{4.6}$$

An efficient estimator $\hat{\theta}_{eff}$ is one that estimates θ in the "best" possible manner, where "best" depends on an assigned loss function. The Mean-Squared Error $E[(\theta - \hat{\theta})^2]$ is most often used in which case:

$$E[(\hat{\theta} - \theta)^2] = Var(\hat{\theta}) + (E[\hat{\theta} - \theta])^2 \tag{4.7}$$

where on the right hand side, the first term $Var(\hat{\theta})$ is the variance of our estimator and the second term $(E[\hat{\theta} - \theta])^2$ refers to the bias of the estimator. Generally to minimize the MSE we want an unbiased estimator $(E[\hat{\theta} - \theta])^2$=0 and seek to minimize the estimator's variance. The accuracy of the our estimator can be measured in terms of

53

its relative error:

$$RE(\hat{\theta}) = \frac{\sqrt{Var(\hat{\theta})}}{\theta} \tag{4.8a}$$

$$= \frac{\sqrt{Var(Z)}}{\theta\sqrt{K}} \tag{4.8b}$$

In the rare-event context, θ is typically dependent on some parameter γ, such that $\theta(\gamma) \to 0$ as $\gamma \to \infty$. For analyzing the efficiency of estimators under the rare-events we care about the following measures (arranged from the strongest to weakest condition)[17, Chapter 10]:

1. $\hat{\theta}$ has **Asymptotically Vanishing Relative Error** if:

$$\limsup_{\gamma\to\infty} \frac{\sqrt{Var(Z(\gamma))}}{\theta(\gamma)^2} = 0 \tag{4.9}$$

2. $\hat{\theta}$ has **Bounded Relative Error** and is **Strongly Efficient** if:

$$\limsup_{\gamma\to\infty} \frac{\sqrt{Var(Z(\gamma))}}{\theta(\gamma)^2} \leq Q < \infty \tag{4.10}$$

3. $\hat{\theta}$ is **Logarithmically Efficient** if:

$$\limsup_{\gamma\to\infty} \frac{\sqrt{Var(Z(\gamma))}}{\theta(\gamma)^{2-\epsilon}} = 0 \ \forall \ \epsilon > 0 \tag{4.11}$$

or equivalently

$$\liminf_{\gamma\to\infty} \left| \frac{\ln \sqrt{Var(Z(\gamma))}}{\ln \theta(\gamma)^2} \right| \geq 1 \tag{4.12}$$

An estimator with bounded relevant error implies the number of samples needed to estimate θ to a given relative accuracy is bounded in γ. For a logarithmically efficient estimator, the number of trials needed to estimate θ within a given relative error grows at a rate of $\mathbf{o}(\log \theta(\gamma))$[9].

4.3 Direct Monte Carlo

The following section describes a Direct Monte Carlo method for estimating the unreliability of a network under the MVN Model. It can be applied to two-terminal/k-terminal/all-terminal unreliability and can easily be modified to estimate the reliability. The objective of this estimator is to approximate the weight of $\mathcal{C}'$ which is equivalent to the probability the given network has failed. Given graph $\mathbf{G} = (\mathbf{V}, \mathbf{E}, \mathbf{s}, \boldsymbol{\tau})$, and structure function Ψ, we do the following:

Algorithm 1: Direct Monte Carlo for Unreliability

   input : $\mathbf{G} = (\mathbf{V}, \mathbf{E}, \mathbf{s}, \boldsymbol{\tau})$

   output: $\hat{U}$ - Estimate of the unreliability

   for $k \leftarrow 1$ **to** K **do**
       | Sample $\mathbf{s}^{(k)} = [s_1^{(k)}, ..., s_m^{(k)}]$
       | Evaluate $X_i^{(k)} = \mathbb{I}(s_i^{(k)} \geq \tau_i) \; \forall \; i$
       | Evaluate $Z^{(k)} = 1 - \Psi(\mathbf{X}^{(k)})$

   $\hat{U} \leftarrow \frac{1}{K} \sum_{k=1}^{K} Z^{(k)}$

Under the MVN Model, to create a link-state sample $\mathbf{X}^{(k)}$ we must first sample $\mathbf{s}^{(k)}$ from $\mathbf{s} \sim \mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. This can be done using the combination of the Box-Muller method for generating univariate standard normal random variables and linear transformations using of the mean vector $\boldsymbol{\mu}$ and Cholesky-decomposition of the covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^{\intercal}$. Once $\mathbf{s}^{(k)}$ has been sampled, the equivalent graph state is generated by assessing $\mathbb{I}(s_i^{(k)} \geq \tau_i)$ for all $i \in [1, m]$. The DMC algorithm for reliability would be almost identical, except $Z^{(k)} = \Psi(\mathbf{X}^{(\mathbf{k})})$. The structure function $\Psi(\mathbf{X}^{(K)})$ can be evaluated using a depth first search (DFS) which has worst-case runtime of $\mathbf{O}(|\mathbf{V}| + |\mathbf{E}|)$. Given the Cholesky-decomposition $\mathbf{L}$ is precomputed, the generation of $\mathbf{s}^{(k)}$ is on the order of $\mathbf{O}(|\mathbf{E}|)$ assuming univariate Gaussian's can be generated in constant time. As a result, the runtime of the Direct Monte Carlo algorithm is dominated by the required number of trials needed to achieve some specified relative error. Consider our estimator $Z = Z(\tau) = \Psi(\mathbf{X}(\mathbf{s}, \tau))$. Z evaluates to either zero or one on each trial thus is a Bernoulli random variable with expectation and variance

given by following:

$$E[Z] = E[Z(\tau)] = Unrel(\mathbf{G}(\tau)) = U(\tau) \tag{4.13}$$

$$Var(Z) = U(\tau)(1 - U(\tau)) \tag{4.14}$$

As $\tau \to \infty$, $Var(Z(\tau)) \to U(\tau)$, thus:

$$\liminf_{\tau \to \infty} \left| \frac{\ln \sqrt{Var(Z(\tau))}}{\ln U(\tau)^2} \right| = \frac{\ln U(\tau)}{\ln U(\tau)^2} = 1/2 < 1 \tag{4.15}$$

thus the DMC estimor is not logarithmically efficient meaning it will taking a prohibitively large number of trials to estimate the unreliability as $U \to \infty$. For example, the relative error the DMC estimator is:

$$RE(\hat{U}) = \frac{\sqrt{Var(\hat{U})}}{U} = \sqrt{\frac{1 - U}{K \times U}} \approx \sqrt{\frac{1}{K \times U}} \tag{4.16}$$

where K is the number of trials. Thus to achieve a relative error of 1% when the unreliability is 10^{-6} the number of trials needed is approximately:

$$K \approx \frac{1}{RE(\hat{U})^2 \times U} = 10^{10} \tag{4.17}$$

4.4 Challenges Introduced by Correlation

Estimator performance can be improved by utilizing known information about the model. This gives us the ability to reduce the variance of our estimator. The variance reduction techniques most frequently seen in literature for the estimation of Network Reliability revolve around the ability to *perfectly sample* either minpaths or mincuts in $\mathbf{G}$. An algorithm that perfectly samples minpaths will generate a minpath with probability proportional to its occurrence probability:

$$\mathbb{P}(P_i \text{ is generated}) = \frac{\mathbb{P}(P_i)}{\sum_j \mathbb{P}(P_j)} \tag{4.18}$$
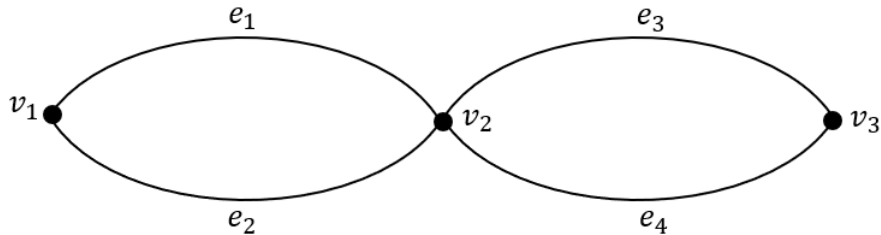
Figure 4-4: Parallel Graph

If one can perfectly sample either minpaths or mincuts, the reliability as expressed in Equations 5.1 or 3.16 may be approximated using the DNF probability FPRAS developed by Karp, Luby, and Madras [28]. The perfect sampling algorithms considered were based on two main concepts: iterative graph transformations or random walks. The follow two sections describe how these operations don't produce perfectly sampled minpaths or mincuts under the correlation introduced by the MVN model.

4.4.1 Random Walks

It has been well documented in the network reliability literature that a random walk on $\mathbb{G}$ produces a perfectly sampled spanning tree [48] [12]. A random walk is performed on $\mathbb{G}$ by starting at some initial vertex v_0, and transitioning to a neighboring vertex with probability proportional to the probability (or weight) of their connecting edge. Recording the edge traversed for the first time when each vertex is encountered produces a spanning tree. A more effective approach is to used a loop-erased random walk where cycles are erased from the random walk if the walk revisits a node. These are sequential algorithms, thus their transition probabilities can be described as using the conditional probabilities of prior events. The total probability of the sequential events is the product of their conditional probabilities. When these sampled events are independent of one another, the product of the conditional events simplifies to the product of the marginal probabilities of each event. In the case of random walks on $\mathbb{G}$ this is needed to produce a spanning tree with the correct perfect sampling probability. This is best exemplified via an example:

Say we want to sample a rooted spanning tree with root node v_1 in graph depicted in Figure 4-4. For the spanning tree $P_1 = [e_1, e_3]$ to be perfectly sampled, the algorithm would have to produce P_1 with probability:

$$\frac{\mathbb{P}(x_1 \cap x_3)}{\mathbb{P}(x_1 \cap x_3) + \mathbb{P}(x_1 \cap x_4) + \mathbb{P}(x_2 \cap x_3) + \mathbb{P}(x_2 \cap x_4)} \tag{4.19}$$

to satisfy Equation 4.18.

Starting at v_1, and letting $\mathbb{P}(\mathbb{S}_j[e_i])$ denote the probability e_i is traversed in step j, the transition probabilities for the first iteration of the random walk would be:

$$\mathbb{P}(\mathbb{S}_1[e_i]) = \begin{cases} \frac{\mathbb{P}(x_1)}{\mathbb{P}(x_1) + \mathbb{P}(x_2)} & \text{if } i = 1 \\ \frac{\mathbb{P}(x_2)}{\mathbb{P}(x_1) + \mathbb{P}(x_2)} & \text{if } i = 2 \end{cases} \tag{4.20}$$

For the next step, one must consider the conditional probabilities given the prior traversals. The following transitional probabilities for step 2 would be applied if e_1 was traversed in step 1 of the random walk:

$$\mathbb{P}(\mathbb{S}_2[e_i]) = \begin{cases} \frac{\mathbb{P}(x_3|x_1)}{\mathbb{P}(x_3|x_1) + \mathbb{P}(x_4|x_1)} & \text{if } i = 3 \\ \frac{\mathbb{P}(x_4|x_1)}{\mathbb{P}(x_3|x_1) + \mathbb{P}(x_4|x_1)} & \text{if } i = 4 \end{cases} \tag{4.21}$$

Thus the probability of the random walk outputting the spanning tree given by $P_1 = [e_1, e_3]$ is:

$$\mathbb{P}(P_1 \text{ is generated}) = \frac{\mathbb{P}(x_1)}{\mathbb{P}(x_1) + \mathbb{P}(x_2)} \times \frac{\mathbb{P}(x_3|x_1)}{\mathbb{P}(x_3|x_1) + \mathbb{P}(x_4|x_1)} \tag{4.22}$$

When link failures are independent, this simplifies to:

$$\mathbb{P}(P_1 \text{ is generated}) = \frac{\mathbb{P}(x_1)}{\mathbb{P}(x_1) + \mathbb{P}(x_2)} \times \frac{\mathbb{P}(x_3}{\mathbb{P}(x_3) + \mathbb{P}(x_4)} \tag{4.23}$$

$$= \frac{\mathbb{P}(x_1)\mathbb{P}(x_3)}{\mathbb{P}(x_1)\mathbb{P}(x_3) + \mathbb{P}(x_1)\mathbb{P}(x_4) + \mathbb{P}(x_2)\mathbb{P}(x_3) + \mathbb{P}(x_2)\mathbb{P}(x_4)} \tag{4.24}$$

Which is equivalent to Equation 4.19 under the independent link assumption. The

denominator in the above expression is the sum of the probabilities of all spanning trees with root node v_1, thus Equation 4.18 is satisfied and the random walk perfectly sampled P_1. However, if correlation cannot be disregarded Equation 4.22 simplifies to:

$$\mathbb{P}(P_1 \text{ is generated}) =$$

$$= \frac{\mathbb{P}(x_1)\mathbb{P}(x_3|x_1)}{\mathbb{P}(x_1)\mathbb{P}(x_3|x_1) + \mathbb{P}(x_2)\mathbb{P}(x_3|x_1) + \mathbb{P}(x_1)\mathbb{P}(x_4|x_1) + \mathbb{P}(x_2)\mathbb{P}(x_4|x_1)}$$

$$= \frac{\mathbb{P}(x_1 \cap x_3)}{\mathbb{P}(x_1 \cap x_3)) + \mathbb{P}(x_2)\mathbb{P}(x_3|x_1) + \mathbb{P}(x_1 \cap x_4) + \mathbb{P}(x_2)\mathbb{P}(x_4|x_1)}$$

which is not equivalent to Equation 4.19. This demonstrates how random walks fail to perfectly sample spanning trees when links are correlated.

4.4.2   Graph Transformations

In the Bernoulli model one can usually divide a graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ into multiple subgraphs $\mathbf{G}^{(1)} = (\mathbf{V}^{(1)}, \mathbf{E}^{(1)}), \dots, \mathbf{G}^{(k)} = (\mathbf{V}^{(k)}, \mathbf{E}^{(k)})$ where $\mathbf{V}^{(i)}$ shares a single node with $\mathbf{V}^{(i+1)}$ for all $i = 1, ..., k-1$. In this scenario, these subgraphs are in series with one another. Since link failures are independent, the overall reliability is the same as the product of the subgraph reliabilities:

$$Rel(\mathbf{G}) = \prod_{i=1}^{k} Rel(\mathbf{G}^{(i)}) \tag{4.25}$$

Another outcome of this ability to subdivide the graph is that one can transform a set of n parallel links into an equivalent single link and vice-versa without effecting the reliability of the overall graph.
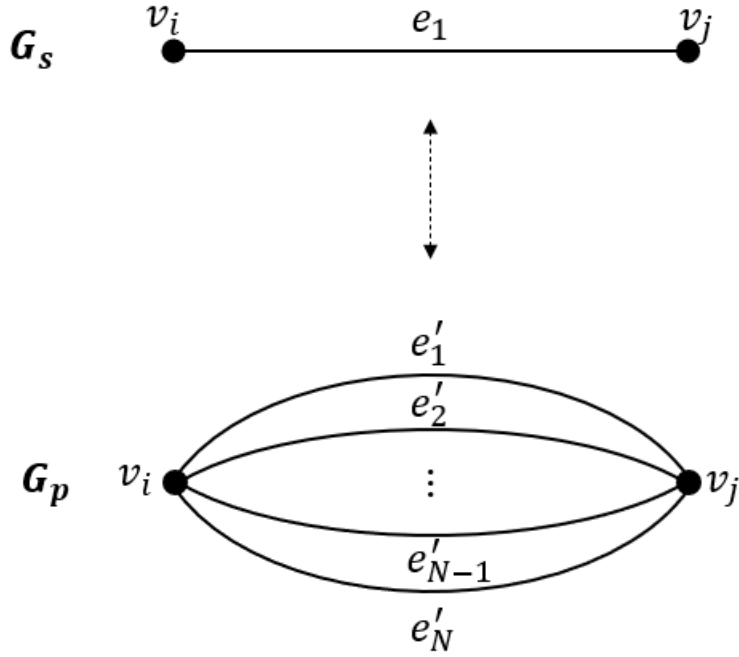
Figure 4-5: Series to Parallel Conversion

Where one can solve for the reliability of the subgraph defined as the two-terminal reliability from v_i to v_j as:

$$p_p = 1 - (1 - p_s)^{1/N} \; (\textit{Series to Parallel}) \tag{4.26}$$

$$p_s = 1 - (1 - p_p)^N \; (\textit{Parallel to Series}) \tag{4.27}$$

These transformations are an effective method for reducing the size of **G** for easier computation, and in some simple graphs may be used to solved for the overall graph reliability. This divisibility of graphs under the Bernoulli model is also used Recursive Contraction Algorithm which is a method of sampling mincuts in the Bernoulli model [27]. In this algorithm, every link e_i with failure probability q_i is replaced with a "bundle" of k_e parallel links with the same endpoint as e_i but with failure probability $1 - \theta$ where $k_i = \lceil -\frac{\ln(q_i)}{\theta} \rceil$, such that the failure probability of the equivalent bundle is now $(1 - \theta)^{k_i}$. As $\theta \to 0$, the failure probability converges to q_i. In general, graph transformations that preserve the reliability of the graph are called *reductions* [14].

60

The Factoring Method is an example of reductions that can be applied in both the Bernoulli and MVN Models. However, the other reductions listed in this section are not possible under the MVN Model as they would require adding or deleting random variables from our multivariate distribution **s** which would result in changes to the reliability of the original graph.

4.5   General Splitting Method

The following method first introduced by [11] allows for the estimation of unreliability in reliability graphs with correlated failures and does not require enumerating mincuts or minpaths. This approach uses a dynamic representation of the Network Reliability that is given below:

Assume link e_i is operational for s_i units of time before failing, where s_i is random. We can specify some deterministic time τ and the variable $X_i(\tau) = \mathbb{I}(s_i \geq \tau)$ signifies if the link e_i is still operational at time τ. $\mathbf{G}(\mathbf{X}(\tau))$ where $\mathbf{X}(\tau) = [X_1(\tau), ..., X_m(\tau)]$ is the subgraph of $\mathbf{G}$ that only contains the edges that for which $X_i(\tau) = 1$ meaning they are still operational at time τ. The network is operational at τ if the graph remains connected after τ time units i.e. $\Psi(\mathbf{X}(\tau)) = 1$. Unreliability of the network can then be expressed as the probability the network is disconnected at time τ. Even though this model is based around a dynamic process of time-dependent network failures, if we let the failure times follow a multivariate normal distribution $\mathbf{s} = [s_1, ..., s_m] \sim \mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ the formulation is mathematically identical ours in the MVN model.

For the following explanation of the Generalized Splitting method we will use the time-based description of the model as the methods are more intuitive when explained in terms of time instead of signal strengths. Given a realization of the random vector **s**, the score of **s**, $W(\mathbf{s})$, is the last time the network is operational:

$$W(\mathbf{s}) = \sup\{\gamma : \Psi(\mathbf{X}(\gamma)) = 1\} \tag{4.28}$$

This score is independent of originally specified threshold τ, however if $W(\mathbf{s}) > \tau$

then we know $\Psi(\mathbf{X}(\tau)) = 1$. To evaluate $W(\mathbf{s})$ we use the following algorithm:

Algorithm 2: Score Function

   input : Sampled signal vector $\mathbf{s} = [s_1, ..., s_m]$
   output: Score $W(\mathbf{s})$
      Apply permutation $\boldsymbol{\pi}$ to $\mathbf{s}$ such that $s_{\pi_1} < s_{\pi_2} < ... < s_{\pi_m}$
      for $b \leftarrow 1$ **to** m **do**
         $\mathbf{X}_b \leftarrow \mathbf{X}(\mathbf{s}_{\pi_\mathbf{b}})$
         if $\Psi(\mathbf{X}_b) = 1$ **then**
            break
      return $W(\mathbf{s}) \leftarrow s_{\pi_{b-1}}$

Based on this score function, the unreliability of a network can equivalently be expressed as:

$$Rel(\mathbf{G}) = \mathbb{P}(W(\mathbf{s}) < \tau) \tag{4.29}$$

A Direct Monte Carlo approach can be implemented using this score function in place of using the structure function on sampled states, however the relative error remains the same as in Equation 4.16. This means when a network failure is a rare-event, it suffers from the same limitations as the DMC implementation in Section 4.2, and is also more costly per trial. Returning to the abstract formulation of the Network Reliability problem, the region $\mathcal{S}_{\mathcal{C}'}$ is equivalent to the region of $\mathcal{S}$ where $W(s) < \tau$ denoted as $\mathcal{S}_\tau$. If we defined a set of intermediate thresholds $\tau_0 < \tau_1 < ..., \tau_T = \tau$, then we can further decompose the subspace $\mathcal{S}$ into nested subspaces $\mathcal{S}_{\tau_0}, \mathcal{S}_{\tau_1}, ..., \mathcal{S}_\tau$.
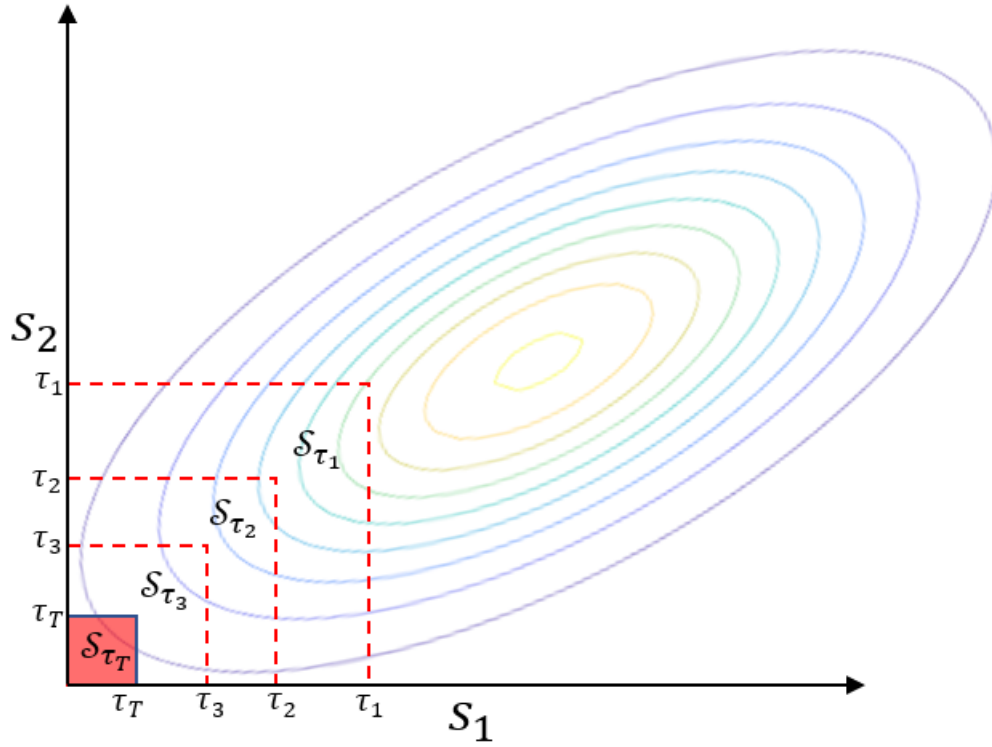
Figure 4-6: Nested subspaces for parallel graph with small τ

The subspace $\mathcal{S}_{\tau_i}$ will contain all subspaces $\mathcal{S}_{\tau_j}$ where $j > i$, thus these are not disjoint subspaces. Now we can represent the probability that a sample $\mathbf{s}$ lies in $\mathcal{S}_\tau$ as

$$\mathbb{P}(\mathbf{s} \in \mathcal{S}_{\tau_T}) = \mathbb{P}(\mathbf{s} \in \mathcal{S}_{\tau_0})\mathbb{P}(\mathbf{s} \in \mathcal{S}_{\tau_1}|\mathbf{s} \in \mathcal{S}_{\tau_0}) \cdots \mathbb{P}(\mathbf{s} \in \mathcal{S}_{\tau_T}|\mathbf{s} \in \mathcal{S}_{\tau_{T-1}}) \qquad (4.30)$$

Simulation-based methods need to generate sufficiently many samples of the rare-event in order to provide an accurate approximation. However if we can decompose the state space into nested subsets, then we can represent the rare event as the intersection of a nested sequence of events. For for example, if we are given a set of intermediate thresholds $\tau_0 > \tau_1 > ..., > \tau_T = \tau$ and we generate sample $\mathbf{s}$ we know that if $W(\mathbf{s}) > \tau_i$, then $W(\mathbf{s}) > \tau_j$ for all $j > i$. This nested structure of the state space allows us to represent the probability of generating a sample such that

$W(\mathbf{s}) > \tau$ as:

$$\mathbb{P}(W(\mathbf{s}) > \tau) = \mathbb{P}(W(\mathbf{s}) > \tau_0)\mathbb{P}(W(\mathbf{s}) > \tau_1 | W(\mathbf{s}) > \tau_0)...$$
$$...\mathbb{P}(W(\mathbf{s}) > \tau_T | W(\mathbf{s}) > \tau_{T-1}) \tag{4.31}$$

These conditional probabilities are much larger than the probability we actually seek, thus are easier to generate samples from. Markov Chain Monte Carlo (MCMC) is a method of approximately sampling from an arbitrary distributed. The main idea is to generate a Markov Chain whose limiting distribution is equal to the desired distribution [17, Chapter 6]. The General Splitting algorithm uses MCMC to sample from the conditional probability distributions given above. The main idea of the GS approach is to define a Markov chain on the state space $\mathcal{S}$ that evolves under a conditional distribution that pushes the chain from the region $\mathcal{S}_t$ to $\mathcal{S}_{t+1}$. We begin with a partition of levels $\infty = \tau_0 > \tau_1 > ..., > \tau_T = \tau$ and a pre-selected splitting factor $F \geq 2$. The levels are chosen such that:

$$\varrho_t = \mathbb{P}(W(\mathbf{s}) > \tau_t | W(\mathbf{s}) > \tau_{t-1}) \approx 1/F \tag{4.32}$$

for all t=1,...,T-1 and are estimated through an independent pilot algorithm. For each level we construct a Markov Chain $\mathbf{S}_{t,j}$ having stationary density equal to the density of $\mathbf{S}$ conditional on $W(\mathbf{S}) > \tau_t$:

$$f_t(\mathbf{s}) = f_{\mathbf{s}}(\mathbf{s}) \frac{\mathbb{I}(W(\mathbf{s} > \tau_t))}{\mathbb{P}(W(\mathbf{s} > \tau_t)} \tag{4.33}$$

where $f_{\mathbf{s}}(\mathbf{s})$ is our unconditional distribution ($f_{\mathbf{s}}(\mathbf{s}) \sim \mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$). For the MVN Model $f_t(\mathbf{s})$ with corresponds to a truncated normal distribution where the region of truncation restricts the density to be non-zero only where $W(\mathbf{s} > \tau_t)$. The transitional kernel density $\kappa_t(\mathbf{S}_{t,j+1} | \mathbf{S}_{t,j})$ controls the evolution of the chain and is constructed via a hit-and-run sampler which is described below. At the t-th stage, if the current state of the Markov Chain is generated from the density f_t and evolves according to $\kappa_t(\mathbf{S}_{t,j+1} | \mathbf{S}_{t,j})$ then each subsequent state also has density f_t. This means the Markov

chain will never go below the level τ_t that has already been reached. Each chain is halted if the number of transitions exceeds the splitting factor F without reaching the region defined by the next threshold $\mathcal{S}_{\tau_{t+1}}$. The following algorithm is for generating a single starting chain and provides an unbiased estimate of the unreliability.

Algorithm 3: GS Algorithm for Network Unreliability

---

input : Splitting Factor F, number of levels T, and thresholds $\tau_1,,,,\tau_T$

output: Unbiased estimate of the unreliability $Z^{(k)}$

Sample $\mathbf{s}$ from unconditional density $f_{\mathbf{s}}(\mathbf{s})$

if $W(\mathbf{s}) < \tau_1$ **then**

   | $\quad Y_1 \leftarrow \mathbf{s}$

else

   | $\quad$ **return** $Z^{(k)} \leftarrow 0$

for $t \leftarrow 2$ **to** T **do**

   | $\quad Y_t \leftarrow 0$

   | $\quad$ **for** *all* $B_1 \in Y_{t-1}$ **do**

   | $\quad\quad$ **for** $i = 1$ **to** F **do**

   | $\quad\quad\quad$ Sample B_i from $\kappa_{t-1}(\mathbf{B}_{i+1}|\mathbf{B}_i)$

   | $\quad\quad\quad$ **if** $W(B_i) < \tau_t$ **then**

   | $\quad\quad\quad\quad$ add B_i to Y_t

return $Z^{(k)} \leftarrow |Y_\tau|/F^{T-1}$

This algorithm would be run for K trials to obtain $Z^{(1)}, \cdots, Z^{(K)}$. One can then estimate the unreliability using Equation 4.6. The variance of our estimator σ_Z^2 can be estimated via the empirical variance:

$$\hat{\sigma}^2 = \frac{1}{K-1} \sum_{k=1}^{K} (Z^{(k)} - \hat{U}_k)^2 \qquad (4.34)$$

where $\hat{U}_k$ is the estimate of the unreliability after the kth trial.

4.5.1 Hit-and-Run Sampler

The hit-and-run sampler is an MCMC method of generating random variables from a target distribution:

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{p_{\mathbf{x(x)}}}{\mathcal{Z}} \tag{4.35}$$

Under the MVN Model, the density $\kappa_{t-1}(\mathbf{B}_{i+1}|\mathbf{B}_i)$ corresponds with sampling from a truncated normal distribution. The following hit-and-run sampler can be used to generate the next states.

Algorithm 4: Truncated MVN Hit-and-Run Sampler

input : A sample $\mathbf{s_{i-1}}$ such that $W(\mathbf{s_{i-1}}) > \tau_t$, positive integer b, $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$

output: $\mathbf{s_{i+1}}$ drawn from $\kappa_t(\mathbf{s}_{i+1}|\mathbf{s}_i)$

$\mathbf{s} \leftarrow \mathbf{s_i}$

for $i = 1$ **to** b **do**

$\quad$ Sample $\boldsymbol{\delta}$ from $\mathbf{N(0, I)}$

$\quad$ $\mathbf{d} \leftarrow \left(\frac{\delta_1}{||\boldsymbol{\delta}||}, ..., \left(\frac{\delta_m}{||\boldsymbol{\delta}||} \right) \right)$

$\quad$ Sample λ from

$$\Lambda \sim N \left(-\frac{(\mathbf{s} - \boldsymbol{\mu})^{\intercal}\boldsymbol{\Sigma}^{-1}\mathbf{d}}{\mathbf{d}^{\intercal}\boldsymbol{\Sigma}^{-1}\mathbf{d}}, \frac{1}{\mathbf{d}^{\intercal}\boldsymbol{\Sigma}^{-1}\mathbf{d}} \right)$$

$\quad$ **if** $W(\mathbf{s} + \lambda\mathbf{d}) < \tau_t$ **then**

$\quad\quad |\quad \mathbf{s} \leftarrow \mathbf{s} + \lambda\mathbf{d}$

$\quad$ **else**

$\quad\quad |\quad \mathbf{s} \leftarrow \mathbf{s}$

$\mathbf{s}_{i+1} \leftarrow \mathbf{s}$

4.5.2 Pilot Algorithm

To determine the appropriate thresholds $\tau_1, ..., \tau_T$ such that Equation 4.32 is satisfied the following Adaptive Multilevel (ADAM) splitting algorithm is used.

Algorithm 5: ADAM Algorithm

input : Splitting factor F, sample size N_0 (a large integer multiple of F),

and final threshold τ

output: Number of levels T, thresholds $\tau_1, ..., \tau_T$, conditional probability

estimates $\varrho_1, ..., \varrho_T$

$q \leftarrow N_0/F$

Generate N_0 samples $\mathbf{s_1}, ..., \mathbf{s_N}$ from unconditional density

Sort the samples in increasing order of $W(\mathbf{s})$, say $\mathbf{s_{(1)}}, ..., \mathbf{s_{(N)}}$

$\tau_1 \leftarrow [W(\mathbf{s}_q) + W(\mathbf{s}_{q+1})]/F$

$t \leftarrow 1$

while $\tau_t > \tau$ **do**

$\quad t \leftarrow t + 1$

$\quad \mathbf{Y}_{t-1} \leftarrow \{\mathbf{s}_{(q)}, ..., \mathbf{s}_{(N)}\}$

$\quad \mathbf{Y}_t \leftarrow \emptyset$

$\quad$ **for** *all* $\mathbf{s}_0 \in \mathbf{Y}_{t-1}$ **do**

$\quad\quad$ **for** $j = 1$ *to* F **do**

$\quad\quad\quad$ Sample $\mathbf{s}_j$ from the density $\kappa_{t-1}(\cdot|\mathbf{s}_{j-1})$ and add it to $\mathbf{Y}_t$

$\quad$ Sort the elements of $\mathbf{Y}_t$ be increasing order of $W(\mathbf{s})$, say $\mathbf{s_{(1)}}, ..., \mathbf{s_{(N)}}$

$\quad \tau_t \leftarrow \max\left([W(\mathbf{s}_q) + W(\mathbf{s}_{q+1})]/F, \tau\right)$

$\quad$ **if** $\tau_t > \tau$ **then**

$\quad\quad \varrho_t = 1/F$

$\quad$ **else**

$\quad\quad$ Let $Y_\tau = [\mathbf{s}_{(1)}, ..., \mathbf{s}_{(N_t)}]$ be the largest subset of elements in $\mathbf{Y}_t$ such

$\quad\quad$ that $W(\mathbf{s}) \geq \tau$

$\quad\quad \varrho_T = N_T/N_0$

$T \leftarrow t$

This algorithm is very similar to the main GS algorithm, except it determines where the thresholds are based on the splitting factors. We start by generating N_o samples from the unconditional distribution, and determine a threshold τ_1 such that q of them have $W(\mathbf{s}) < \tau_1$. Then in subsequent runs we generate N_o samples from conditional distribution given by Equation 4.33, and once again determine a threshold τ_t such that q of the sample have $W(\mathbf{s}) < \tau_t$. This process halts when the

threshold determined exceeds our intended final threshold τ. In addition to producing the thresholds that will be used in Algorithm 2, $\tau_1, ..., \tau_T$, the pilot algorithm also produces estimates of conditional probabilities $\varrho_1, ..., \varrho_T$. These can be used to form an initial estimate of the unreliability:

$$\hat{U}_{pilot} = \prod_{t=1}^{T} \varrho_t \qquad (4.36)$$

Unlike Algorithm 2, each stage performs a fixed number of simulations and the thresholds are determined online using random populations $\{\mathbf{Y}_t\}$. As a result, $\hat{U}_{pilot}$ is not an unbiased estimate of the unreliability[11].

4.6 Simulation

The following section gives results from implementations of the approximation algorithms described in Sections 4.3 and 4.5. First, we describe a method of generating MVN parameters based on the spatial relationships of edges in a graph. This MVN parameter generation method is an abstraction of the distant dependent means and control point dependent correlation coefficients expected in an HF Network.

4.6.1 MVN Parameter Generation

The MVN Model requires an of a graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ as well as a specification of the signal distribution $\mathbf{s} \sim \mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and thresholds $\boldsymbol{\tau}$. To simulate network reliability with parameters that would model spatial relationships of an HF Network the following geometric model was used: Let $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ be given on a two-dimensional plane, with node coordinates given by the tuple vector $\mathbf{L} = [(x_1, y_1), ..., (x_n, y_n)]$ for nodes $\mathbf{V} = [V_1, ..., V_N]$. Three user-defined inputs are also needed:

1. Transmit Power vector $\mathbf{T} = [T_1, ..., T_n]$

2. Variance vector $\boldsymbol{\sigma} = [\sigma_1, ..., \sigma_m]$

3. Threshold vector $\boldsymbol{\tau} = [\tau_1, ..., \tau_m]$

where n and m are the number of nodes and edges in $\mathbf{G}$ respectively. In order to use an undirected graphical model, we will assume that all input parameters are the same for each node and link, thus the common transmit power, variance, and thresholds will be refered to as T, σ, and τ respectively

To define the means signal strength for each link and correlations between the signal strengths of links, we require the link lengths:

$$l(e_i) = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2} \tag{4.37}$$

for $e_i \in (v_a, v_b)$, as well as the distance between link midpoints:

$$M(e_i, e_j) = \sqrt{\left(\frac{(x_a + x_b) - (x_c + x_d)}{2}\right)^2 + \left(\frac{(y_a + y_b) - (y_c + y_d)}{2}\right)^2} \tag{4.38}$$

for $e_j \in (V_c, V_d)$. The mean signal strength over link e_i is:

$$\mu_i = T - l(e_i) \tag{4.39}$$

meaning the path-loss is modeled as linear which is unrealistic for an HF Network (or any other wireless network), but once again this geometric model captures overall spatial trends seen in an HF network and is not intended to model the physical propagation environment. The correlation between links e_i and e_j is given by the squared-exponential kernel:

$$Corr(s_i, s_j) = \rho_{i,j} = exp\left(-\frac{M_{i,j}^2}{2l^2}\right) \tag{4.40}$$

where l is the *characteristic length* which is user defined and allows one to modify the magnitude of the relationship between spacial separation and correlation. Kernels, also called covariance functions, define nearness or similarity (usually in space or time) between inputs x_i and x_j. For a function $k(x_i, x_j)$ to be a valid covariance function,

it must be positive semi-definite meaning:

$$\int k(x_i, x_j) f(x_i) f(x_j) d\mu(x_i) d\mu(x_j) \geq 0 \tag{4.41}$$

for all $f \in L_2(\mathcal{X}, \omega)$ where ω denotes a measure. For more information on kernels, readings are directed to [43].
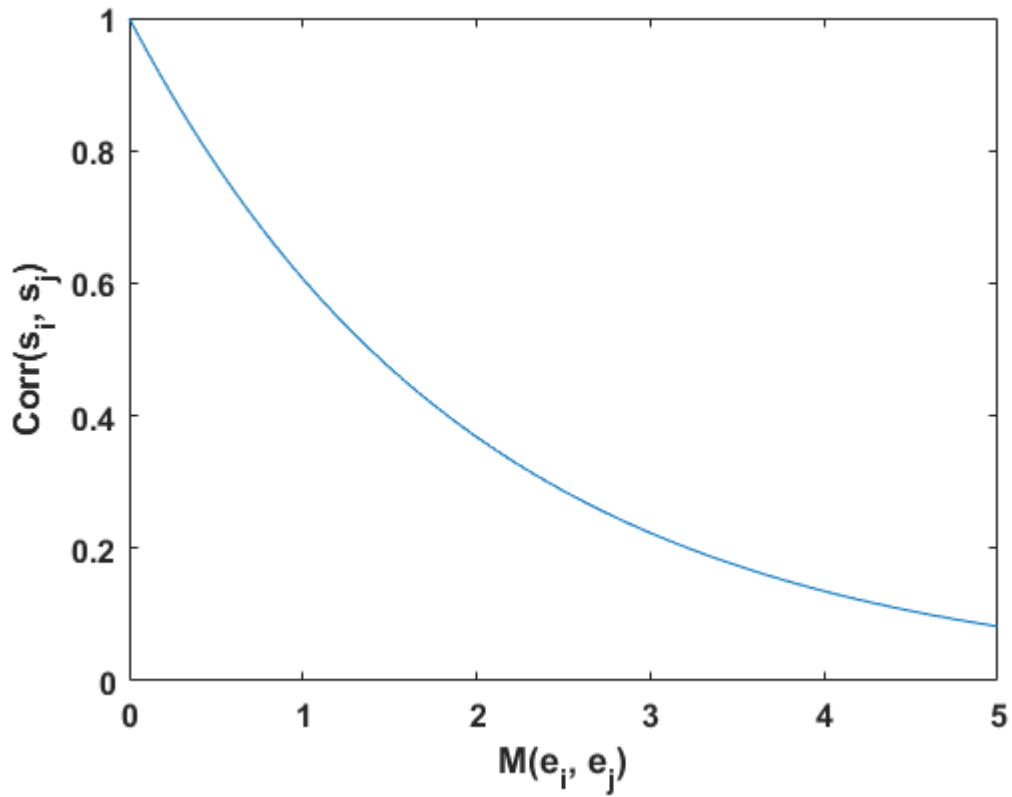
Figure 4-7: Correlation as a function of midpoint separation $M(e_i, e_j)$ under the squared exponential kernel with characteristic length $l = 1$
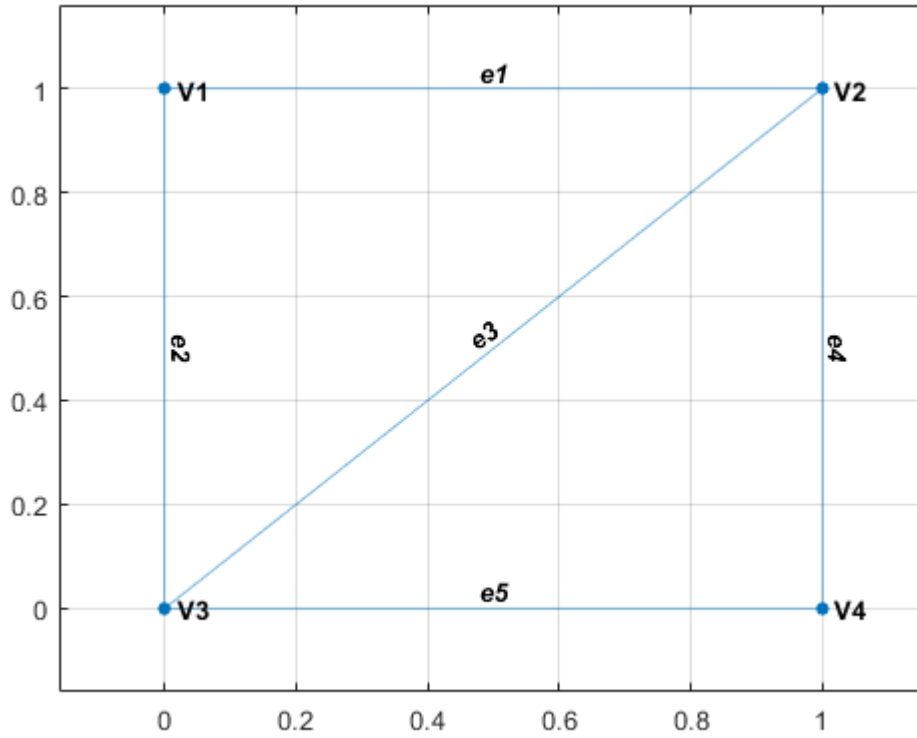
Figure 4-8: Bridge Graph on 2-dimensional plane

As an example, the bridge graph given in 4-8 the following input parameters were used:

Input Parameters	
T	4
σ	1
τ	0
l	1

which resulted in the mean vector $\boldsymbol{\mu} = [3, 3, 2.56, 3, 3]^{\mathsf{T}}$ and the covariance matrix given in Table 4.1

	s_1	s_2	s_3	s_4	s_5
s_1	1.00	0.61	0.78	0.61	0.37
s_2	0.61	1.00	0.78	0.37	0.61
s_3	0.78	0.78	1.00	0.78	0.78
s_4	0.61	0.37	0.78	1.00	0.61
s_5	0.37	0.61	0.78	0.61	1.00

Table 4.1: $\Sigma = \mathrm{Cov}(s_i, s_j)$ for Bridge Graph example

4.6.2 Algorithm Results

The following section reports numerical experiments for the DMC and GS methods. Each algorithm was tested on both the bridge graph shown in 4-8 and the larger graph shown in 4-10. The MVN parameters were generated using the methods described in 4.6.1. For all experiments the following parameters were held constant: $\sigma = 1$, $l = 1$, and $\tau = 0$. As a result of σ and l being constant, the covariance matrix Σ was also constant between all experiments on the same graph. To demonstrate the relationship between the unreliability and the run-time/accuracy of these approximation methods, the unreliability was reduced by increasing the transmit power of each node, thus increasing all μ_i, for each subsequent experiment. The variance of the estimator, σ_Z^2, was estimated between each iteration of the algorithms using Equation 4.34 which was then used to estimate the relative error:

$$RE(\hat{U}) = \frac{\hat{\sigma}_Z}{\hat{U}_k \sqrt{k}} \tag{4.42}$$

where $\hat{U}_k$ was the estimate of the unreliability after trial k. Both the DMC and GS algorithm would output their estimate $\hat{U}_k$ when $RE(\hat{U})$ fell below some small ϵ. For DMC algorithm, the number of trials had a limit of 5×10^7 put into place for the rare failure event experiments. The pilot algorithm of the GS method used an initial sample size of 3000 for all trials. An explanation off all metrics in the results tables can be found in Table 4.2.

Experiment #N	
Reliability Parameters	
T	Transmit power for all nodes
U	Unreliability computed via full state enumeration
DMC Results	
$\widehat{U}$	Estimated unreliability by the DMC Algorithm
$\text{RE}(\widehat{U})$	Estimated Relative Error of the DMC Estimate
Trials	Total number of trials conducted by the DMC Algorithm
Approximation Error	Normalized error $\lvert U - \widehat{U}\rvert / U$
Total Runtime [sec]	Total Runtime of the DMC Algorithm
GS Results	
$\widehat{U}_{pilot}$	Estimated unreliability by the GS Pilot Algorithm
Pilot Runtime [sec]	Runtime of the GS Pilot Algorithm
$\widehat{U}$	Final estimated unreliability by the GS Algorithm
$\text{RE}(\widehat{U})$	Estimated Relative Error of the GS Estimate
Trials	Total number of trials conducted by the main GS Algorithm
Approximation Error	Normalized error $\lvert U - \widehat{U}\rvert / U$
Total Runtime [sec]	Total Runtime of the GS Algorithm (including Pilot Runtime)

Table 4.2: Description of all metrics for the results

For the first three experiments, T was chosen such that the DMC algorithm would converge in under 5×10^7 trials. These results are given in Table 4.3.

Experiments 1,2, and 3 correspond to the scenarios of high, moderate, and low unreliability respectively. The unreliability for each trial U was computed by enumerating each state in $\mathbf{G}$, computing its probability, and then summing the probabilities of each state. For all experiments both the DMC algorithm and GS algorithm were able to achieve a relative error of 0.0300 prior to halting and produced estimates with a relative error $<6\%$. The difference in runtimes is most significant in the high and moderate reliability scenarios, where the ratio of the DMC Runtime to GS Runtime is 0.0177 and 0.032 respectively. The initial estimates output by the GS Pilot algo-

Experiment #1		Experiment #2		Experiment #3	
Reliability Parameters		**Reliability Parameters**		**Reliability Parameters**	
T	2	T	3	T	4
U	0.1721	U	0.0169	U	5.058e-04
DMC Results		**DMC Results**		**DMC Results**	
$\widehat{U}$	0.1706	$\widehat{U}$	0.0164	$\widehat{U}$	4.78e-04
RE($\widehat{U}$)	0.0300	RE($\widehat{U}$)	0.0300	RE($\widehat{U}$)	0.0300
Trials	9824	Trials	71129	Trials	2.33e06
Approximation Error	0.0082	Approximation Error	0.0329	Approximation Error	0.0556
Total Runtime [sec]	1.0392	Total Runtime [sec]	4.70	Total Runtime [sec]	160.99
GS Results		**GS Results**		**GS Results**	
$\widehat{U}_{pilot}$	0.1724	$\widehat{U}_{pilot}$	0.0165	$\widehat{U}_{pilot}$	4.87e-04
Pilot Runtime [sec]	17.1930	Pilot Runtime [sec]	29.21	Pilot Runtime [sec]	70.40
$\widehat{U}$	0.1743	$\widehat{U}$	0.0171	$\widehat{U}$	5.16e-04
RE($\widehat{U}$)	0.0300	RE($\widehat{U}$)	0.0300	RE($\widehat{U}$)	0.0300
Trials	7248	Trials	10731	Trials	15671
Total Runtime [sec]	58.65	Total Runtime [sec]	143.63	Total Runtime [sec]	448.08
Approximation Error	0.0132	Approximation Error	0.0086	Approximation Error	0.0196

Table 4.3: Results of Experiments 1, 2 and 3 for Bridge Graph

rithm were accurate in these cases and for all three experiments actually provided better estimates than the DMC algorithm and sometimes even the main GS Algorithm. However, the Pilot Algorithm uses a fixed number of trials thus its accuracy is not guaranteed. One thing to note is the difference in the increase of the number of trials between subsequent experiments. The number of trials for the GS algorithm increases by less than 50% between each consecutive experiment, while the number of trials increases exponentially for the DMC algorithm. The time per trial is fixed for DMC algorithm thus we also see an exponential increase in overall runtime. The DMC algorithm however proved to be much more effective in for these unreliabilities as was expected.

For Experiment #4, the unreliability was reduced to 4.20×10^{-6} making the faiure of **G** a rare-event.

Experiment #4	
Reliability Parameters	
T	5
U	4.20e-06
DMC Results	
$\widehat{U}$	3.90e-06
RE($\widehat{U}$)	0.0716
Trials	5e07
Approximation Error	0.0721
Total Runtime [sec]	3.31e03
GS Results	
$\widehat{U}_{pilot}$	4.36e-06
Pilot Runtime [sec]	97.65
$\widehat{U}$	4.37e-06
RE($\widehat{U}$)	0.0300
Trials	24241
Total Runtime [sec]	916.47
Approximation Error	0.0400

Table 4.4: Experiment #4 results for Bridge Graph

For the rare failure event scenario examined under Experiment #4, the DMC algorithm was halted after running for over 3000 seconds after reaching its limit of 5×10^7 trials. The GS Algorithm did achieve convergence of its relative error to 0.03 and was able to provide an estimate with low approximation error in less than 1000 seconds. The relative errors of the DMC method and GS method are compared in Figure 4-9. Only first 1.38×10^7 trials are plotted for the DMC method to equate the runtimes between the two plots. It can be seen that the relative error decreases far more rapidly before leveling off at around 0.1 and then gradually decreases to its stopping point of 0.03. Meanwhile the relative error of the DMC method takes longer to approach a leveling point, and the leveling point is about 0.2 thus the gradual approach of 0.03 takes significantly longer.
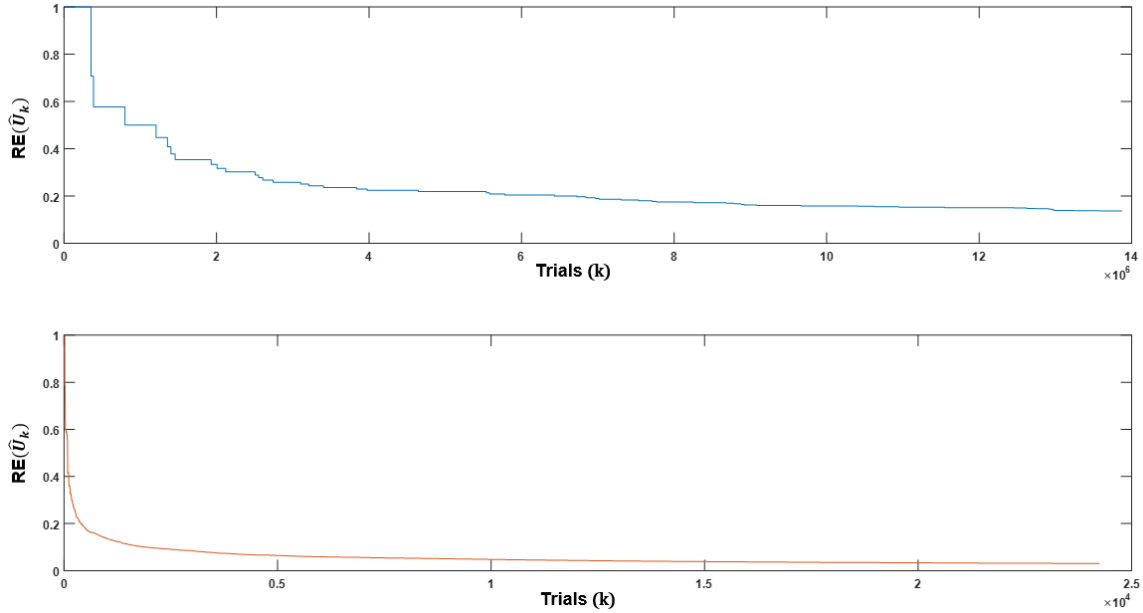
Figure 4-9: Relative error for the DMC Method (Top) and the GS Method (Bottom) for Experiment #4

For Experiment #5 the graph shown in 4-10 was used and the results are shown below in Table 4.5. Comparing the results of Experiment #1 to #5 we see see that the number of trials does not effect the number of trials needed for convergence either the DNC nor GS methods. The difference in the runtime of the GS method between Experiments #1 and #5 can be explained by the longer average computation time of $W(\mathbf{s})$ via Algorithm 2 for the larger graph. The computing of U via full state enumeration took 489 seconds for Experiment #5 while the same computation only took 0.98 seconds for Experiment #1. This is because the Bridge graph only had $2^5 = 32$ states while the graph in 4-10 had $2^{14} = 16384$ states. This demonstrates how unreliability computation becomes unfeasible quickly for larger graphs.

These results demonstrate how the DMC method is an efficient and accurate method for computing the unreliability. Even for an unreliability on the order of 10^{-4} the DMC method outperformed the GS algorithm. Only when the unreliability becomes extremely low does the GS method become more effective.

These results also illustrate how both methods are invariant to the size of $\mathbf{G}$ thus

76

Experiment #5	
Reliability Parameters	
T	3
U	0.1671
DMC Results	
$\widehat{U}$	0.1591
$\text{RE}(\widehat{U})$	0.0300
Trials	10317
Approximation Error	0.0479
Total Runtime [sec]	1.0877
GS Results	
$\widehat{U}_{pilot}$	0.1752
Pilot Runtime [sec]	23.38
$\widehat{U}$	0.1644
$\text{RE}(\widehat{U})$	0.0300
Trials	7330
Total Runtime [sec]	79.84
Approximation Error	0.0161

Table 4.5: Experiment #5 results

are not restricted to small graphs like full state enumeration.

To use these algorithms in practical setting an initial run of the DMC method could be performed with a more modest limit on the maximum number of trials; for example $\sim 10^7$. If the the relative error of the DMC method does not converge within this limit, then it is known that we are dealing with a rare-failure event unreliability, and the GS method should be used instead.
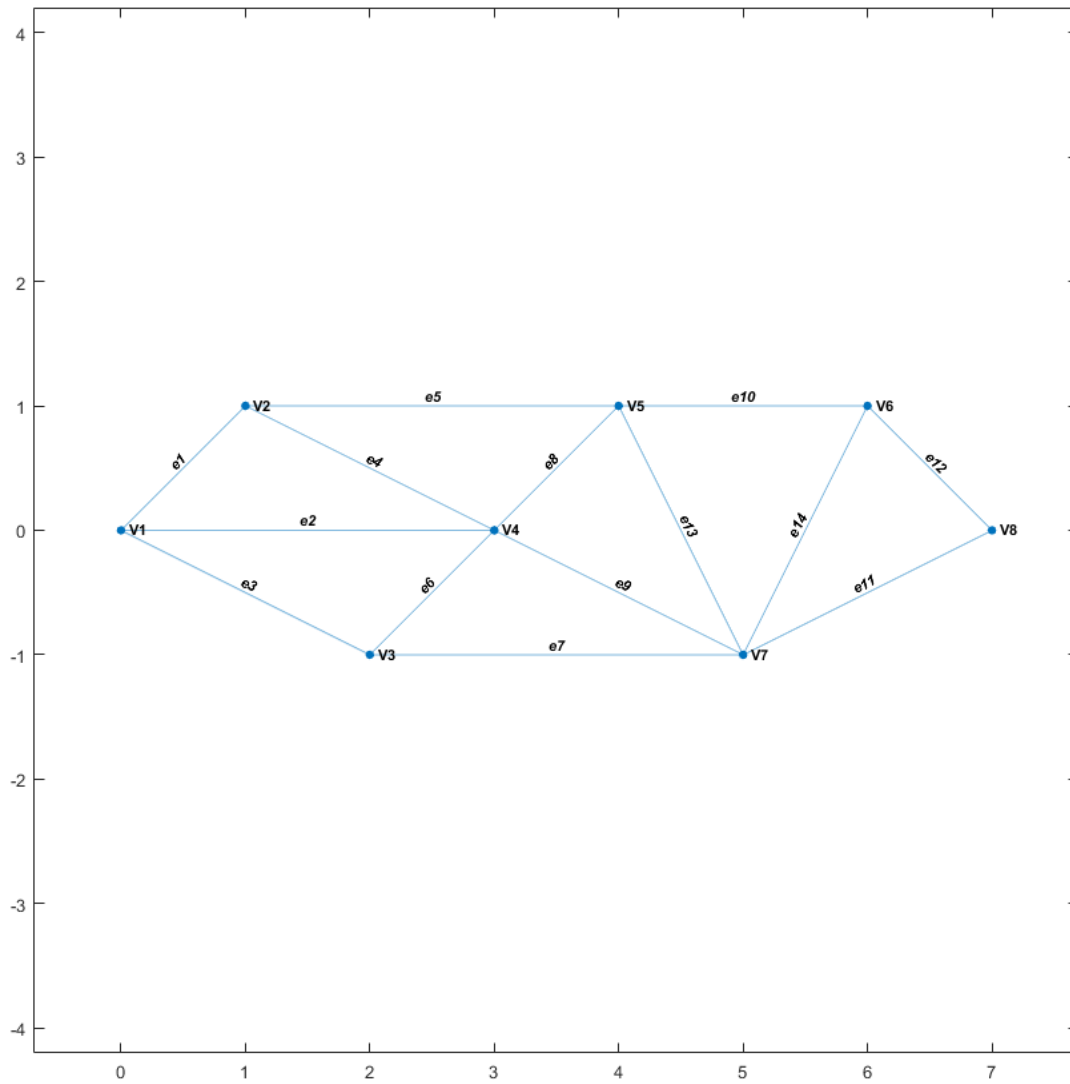
Figure 4-10: Fourteen link graph used for Experiment #5

Chapter 5

Most Reliable Path Problem

In this chapter we consider the problem of finding the most reliable path between two nodes. Section 5.1 will provide an overview of this problem and the difference between solving this problem under the Bernoulli and MVN Models. Section 5.2 provides a method to solve the most reliable path problem under the MVN Model while Section 5.3 provides upper and lower bounds on the reliability of the most reliable path.

5.1 Shortest Path Problem Formulation

Given $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, and the set of all paths $\mathcal{P}_{s,t}$ from nodes v_s to v_t, the most reliable path $P_{s,t}^* \in \mathcal{P}_{s,t}$ has the maximum joint reliability or equivalently the minimum joint failure probability of all ST-paths:

$$P_{st}^* = arg \max_{P \in \mathcal{P}_{st}} \mathbb{P}\left(\bigcap_{e_i \in P} x_i\right) = arg \min_{P \in \mathcal{P}_{st}} \mathbb{P}\left(\bigcup_{e_i \in P} x_i'\right) \tag{5.1}$$

First we consider the case where link reliabilities are independent as in the Bernoulli model. We can rewrite equation 5.1 under the independence assumption as the product of marginal reliabilities:

$$P_{st}^* = \arg \max_{P \in \mathcal{P}_{st}} \mathbb{P}\left(\bigcap_{e_i \in P} x_i\right) = \arg \max_{P \in \mathcal{P}_{st}} \prod_{e_i \in P} p_i \tag{5.2}$$

79

Using the fact that the both the maximum and logarithm functions are non-decreasing, we can equivalently express the objective function as:

$$
\begin{aligned}
P_{st}^* &= \arg\max_{P \in \mathcal{P}_{st}} \prod_{e_i \in P} p_i \\
&= \arg\max_{P \in \mathcal{P}_{st}} \log\left(\prod_{e_i \in P} p_i\right) \\
&= \arg\max_{P \in \mathcal{P}_{st}} \sum_{e_i \in P} \log(p_i) \\
&= \arg\min_{P \in \mathcal{P}_{st}} -\sum_{e_i \in P} \log(p_i) \quad\quad (5.3)
\end{aligned}
$$

This formulation is equivalent to the single pair shortest path problem with edge lengths given by $d(e_i) = -\log p_i$ [7].

5.1.1   Shortest Path Problem

Shortest path problems are applicable to many applications outside of communication networks such as transportation networks, optimal control, and as sub-problems to other complex graph theoretic problems [7]. Consider a graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ in which each edge e_i is assigned a distance d_i. Given any path $P = [e_l, e_m, ..., e_z]$, the length of the path is $d_l + d_m + \cdots + d_z$. There are multiple variants of the shortest path problem, but the most well known is the problem of finding the shortest path from every node to some terminal node designated as v_1. The shortest path problem exhibits the property of *optimal substructure*: if the shortest path from v_i to v_k goes through an intermediate node v_j, then the portion of the path from v_i to v_j is the shortest path between v_i to v_j and the portion of the path from v_j to v_k is the shortest path between v_j and v_k [8]. This optimal substructure lends itself to greedy algorithms, dynamic programming, or linear programming formulations of the problem. For more information on these approaches to solving the shortest-path problem, readers are directed to [15, Chapter 24].
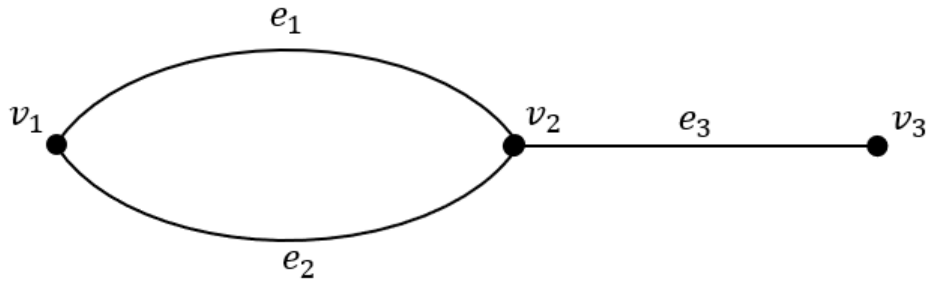
Figure 5-1: Example for non-optimal substructure

5.1.2 Lack of Optimal Substructure

When link failures are independent, the optimal substructure property of the most reliable path problem is maintained. Under correlated link failures, the most reliable path problems loses optimal substructure and this can be illustrated via the following example:

Consider the graph shown in Figure 5-1. Optimal substructure would imply if $\mathbb{P}(x_1 \cap x_3) \geq \mathbb{P}(x_2 \cap x_3)$, then $\mathbb{P}(x_1) \geq \mathbb{P}(x_2)$. If all links were independent $\mathbb{P}(x_1 \cap x_3) = \mathbb{P}(x_1)\mathbb{P}(x_3)$ and $\mathbb{P}(x_2 \cap x_3) = \mathbb{P}(x_2)\mathbb{P}(x_3)$, thus it would require $\mathbb{P}(x_1) \geq \mathbb{P}(x_2)$ and the problem would have optimal substructure as the most reliable path from v_1 to v_3 would be $[e_1, e_3]$ which contains the most reliable path from v_1 to v_2, e_1. To demonstrate how this could change under correlation let $1 > \mathbb{P}(x_2) > \mathbb{P}(x_1) > \mathbb{P}(x_3)$. Let links e_1 and e_3 be fully correlated: $\mathbb{P}(x_1 \cap x_3) = min(\mathbb{P}(x_1), \mathbb{P}(x_3)) = \mathbb{P}(x_3)$, while links e_2 and e_3 are independent: $\mathbb{P}(x_2 \cap x_3) = \mathbb{P}(x_2)\mathbb{P}(x_3)$. Then the most reliable path from v_1 to v_2 is $[e_2]$, while the most reliable path from v_1 to v_3 is $[e_1, e_3]$ which doesn't contain the most reliable path from v_1 to v_2.

5.2 Graph Transformation Technique

The following technique transforms the graph with correlated link failures into a form that has optimal substructure for the most reliable path problem. The approach uses a modified depth first traversal to enumerate all paths from v_s to v_t.

Algorithm 6: Path Enumeration Algorithm

`Path_DFS`$(G, v_i, v_t, \mathcal{P}_{s,t},\ stack)$

    if $v_i = v_t$ **then**
        add stack to $\mathcal{P}_{s,t}$

        return
    foreach v_j *in neighbors(v_j)* **do**

        if v_j *not in stack* **then**
            push v_j to stack

            `Path_DFS`$(G, v_j, v_t, \mathcal{P}_{s,t},\ stack)$

            pop v_j from stack
    return $\mathcal{P}_{s,t}$

A traditional DFS algorithm would exclude the line "pop v_j from the stack", but this allows the algorithm to backtrack and revisit previously visited nodes for the Path Enumeration Algorithm. The output of the algorithm $\mathcal{P}_{s,t}$ would be a set of all paths from v_s to v_t in terms of the nodes in order of their traversal. Using $\mathcal{P}_{s,t}$, we create an equivalent *path graph* by connecting v_s and v_t by parallel branches composed of the paths in $\mathcal{P}_{s,t}$. The distribution associated with each edge within the same branch will be conditioned on all preceding edges in the branch being operational. This idea is illustrated in Figure 5-2.
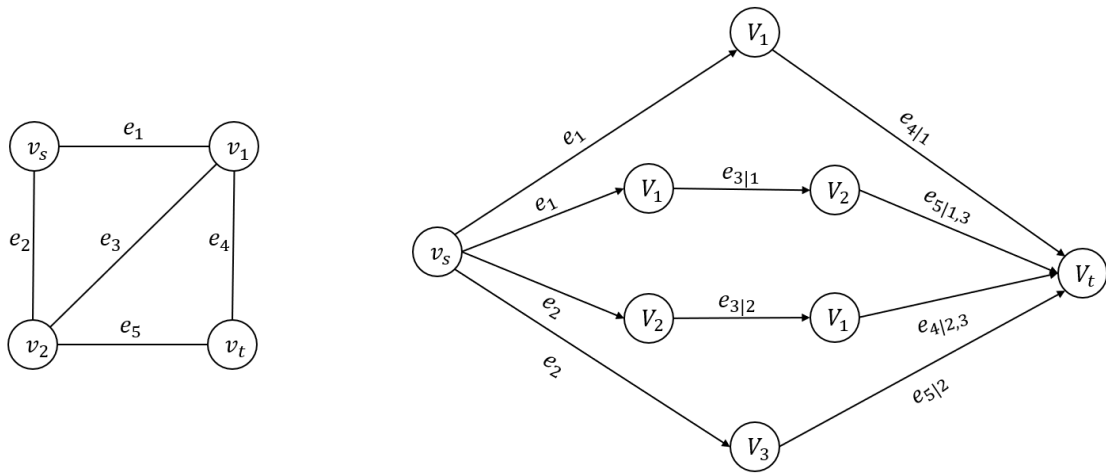

Figure 5-2: Original graph (left) and corresponding path graph (right)

As seen from above, each branch between v_s and v_t is a path in the original graph. All neighbors of v_s will be connected by an edge whose distribution is unchanged from the original graph. The distribution of subsequent edges along a branch are conditional on preceding edges being operational. Letting the branches in the path graph be labeled as $B_1, ..., B_4$ from the top branch to the bottom, the MVN model distribution of $e_{4|1}$ along the B_1 would be the joint distribution of s_1 and s_4 conditioned on $\{s_1 > \tau_1\}$:

$$f_{4|1}(\mathbf{s}) = \frac{f_{1,4}(\mathbf{s})\mathbb{I}(s_1 > \tau_1)}{\mathbb{P}(s_1 > \tau_1)} \tag{5.4}$$

where $f_{1,4}(\mathbf{s})$ is the joint distribution of s_1 and s_4. The equivalent probability $e_{4|1}$ being operational would be $\mathbb{P}(s_4 > \tau_4 | s_1 > \tau_1)$ which can be found via the integration of $f_{4|1}(\mathbf{s})$:

$$\mathbb{P}(x_4|x_1) = \int_{\tau_4}^{\infty} \int_{\tau_1}^{\infty} f_{4|1}(\mathbf{s}) ds_1 ds_4 \tag{5.5}$$

The resulting probability of each path may be evaluated by taking the product of all conditional probabilities along each branch. For example, for B_1:

$$\mathbb{P}(B_1) = \mathbb{P}(s_1 > \tau_1)\mathbb{P}(s_4 > \tau_4 | s_1 > \tau_1) \tag{5.6}$$

However, this method would be computationally expensive. To improve upon this exhaustive enumeration of all probabilities, a greedy algorithm may now be used to traverse the path graph from v_s to v_t. Prior to the start of this algorithm a stack L_i is initialized for each branch B_i of the path graph and the starting node v_s is added to each stack. Associated with each L_i is a cost c_i which is equivalent to the probability of the edge being on that connects the head of L_i to the next possible edge in B_i. For example, at the start of the algorithm:

$$L_1 = [v_s] \rightarrow c_1 = \mathbb{P}(x_1)$$
$$L_2 = [v_s] \rightarrow c_2 = \mathbb{P}(x_1)$$
$$L_3 = [v_s] \rightarrow c_3 = \mathbb{P}(x_2)$$
$$L_4 = [v_s] \rightarrow c_4 = \mathbb{P}(x_2)$$

At each iteration of the algorithm, the branch (or branches) with the greatest c_i will be chosen to add the next node in the branch to the stack, and their costs would be updated. For example, if $\mathbb{P}(x_1) > \mathbb{P}(x_2)$ then after the first iteration our stacks and costs would be:

$$L_1 = [v_s, v_1] \rightarrow c_1 = \mathbb{P}(x_4|x_1)$$
$$L_2 = [v_s, v_1] \rightarrow c_2 = \mathbb{P}(x_3|x_1)$$
$$L_3 = [v_s] \rightarrow c_3 = \mathbb{P}(x_2)$$
$$L_4 = [v_s] \rightarrow c_4 = \mathbb{P}(x_2)$$

This process would continue until a branch adds v_t to its stack meaning that branch corresponds to the most reliable path.

This method is computation expensive, as just the modified DFS algorithm has worst case time-complexity of O(V!). However, it is guaranteed to identify the most reliable path and it offers a possible speedup from complete path enumeration and computation as the greedy algorithm ensures not all path probabilities are fully computed.

5.3   Bounding Path Reliabilities

Considering the case where correlation between all links is restricted to $(0, 1)$, such as in our geometric MVN parameter generation model described in Section 4.6.1, the following bounds on path reliabilities may be applied.

5.3.1   Upper Bound

The reliability of a path is improved as the correlation between links within the path is increased. Given path $P = [e_1, e_2, ...e_k]$, a simple upper bound on the path reliability

corresponds to the case where all links are fully correlated:

$$Rel(P) = \mathbb{P}\left(\bigcap_{i=1}^{k} x_i\right) \leq min(\mathbb{P}(x_1), \mathbb{P}(x_2), ..., \mathbb{P}(x_k)) \qquad (5.7)$$

Identifying the path from v_s to v_t with the greatest upper bound can be done by computing the marginal reliabilities of all links in $\mathbf{G}$ and then running a widest-path algorithm on $\mathbf{G}$ using the marginal reliabilities as weights. Whereas a shortest path algorithm identifies the minimum weighted path from v_s to v_t, a widest path algorithm identifies the st-path(s) with the maximum weight of the minimum edge weights within the path. The widest path problem may be solved via the following modified Dijkstra's algorithm:

Algorithm 7: Widest Path Algorithm

input : Graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, Marginal Reliabilities $\mathbf{p} = [p_1, ..., p_m]$, Source Node v_s, and Terminal Node v_t

output: Weight of widest path W and vector of visited nodes **prev**

$Q \leftarrow \mathbf{V}$

foreach v_j in $1 : |\mathbf{V}|$ **do**
    $\mathbf{W}[v_j] \leftarrow \infty$
    $\mathbf{prev}[v_j] \leftarrow \emptyset$
$\mathbf{W}[v_s] = 0$

while Q is ***not*** $empty$ **do**
    $v_u \leftarrow$ vertex in Q with min $W[v_i]$
    Delete v_u from Q
    if $v_u = v_t$ **then**
        break
    foreach $neighbor$ v_n of v_u **do**
        $w \leftarrow min(W[v_u], p(v_u, v_n))$
        if $w > W[v_n]$ **then**
            $W[v_n] \leftarrow w$
            $prev[v_n] \leftarrow v_u$
return W, $prev$

The upper bound corresponding to Equation 5.10 will be $\mathbf{W}[v_t]$, while the path from v_s to v_t can be found via by backtracking from v_t to v_s in **prev**. Like Dijkstra's

shortest path algorithm, the runtime of this algorithm is $O((n + m)\log(n))$ where $n = |\mathbf{V}|$ and $m = |\mathbf{E}|$. This upper bound may be loose but is easily computed with the widest path algorithm. Another upper bound that may be applied is:

$$Rel(P) = \mathbb{P}\left(\bigcap_{i=1}^{h} x_i\right) \leq \min\left(\mathbb{P}\left(\bigcap_{J \subset \{1,...,h\}} x_i\right)\right) \tag{5.8}$$

This means that a path containing h links can be upper bounded by the joint reliability of any subset of the h links along the path. This bound does not lend itself to an algorithm as it is a minimization over a probability of intersection of different subsets. Similar the the most reliable path problem, this problem lacks optimal substructure unless all elements in the path are uncorrelated.

5.3.2 Lower Bound

Given path $P = [e_1, e_2, ...e_h]$, a lower bound the reliability of the path corresponds the scenario where all links are independent of one another. In this case:

$$\prod_{i=1}^{h} \mathbb{P}(x_i) \leq \mathbb{P}\left(\bigcap_{i=1}^{h} x_i\right) \tag{5.9}$$

Considering all links to be independent simplifies the MVN Model to the Bernoulli Model with $p_i = \mathbb{P}(s_i > \tau)$. As described by Section 5.1.1, we can reformulate the most reliable path problem as the shortest path problem with edge lengths given by $d[e_i] = -log(p_i)$ when all links are independent. Thus we can solve for the path with the greatest lower bound via a shortest path algorithm.

5.3.3 Tightness of Bounds

While simulating graphs via the geometric MVN parameter generation method described in Section 4.6.1, it was observed that correlation had a negligible effect on the most reliable path and could usually be ignored. Ignoring correlation in this case means considering all links to be independent with marginal reliabilities given by

86

$\mathbb{P}(s_i > \tau)$ and joint reliabilities computed as a product of marginal reliabilities rather than the probability of their intersection. This is the same strategy as identifying the path with the greatest lower bound from Section 5.3.2.

For this section, we will assume correlation between two links is dependant on a non-increasing function of the distance between their midpoints, such as in the geometric MVN parameter generation method described in Section 4.6.1. However, these results may generalize to other spatial correlation models a well, as long as correlation is inversely related to link separation. Under the midpoint correlation model, we can specify a decorrelation distance d' such that if the midpoint separation $M_{i,j}$ between links e_i and e_j is greater that d' the correlation between the two links is negligible. Assume the minimum number of links of any st-path in $\mathbf{G}$ is $h \geq 2$ for a particular node pair v_s and v_t. Given the distance between v_s and v_t is $d_{s,t}$, the midpoint separation distance between the first and last link of any path is $M_{1,h} \geq d_{s,t}/2$. Assuming $d_{s,t}/2 > d'$, then we know the correlation between the first e_1 and last link e_h in the path will be negligible and we can approximate their joint reliability as $\mathbb{P}(x_i \cap x_h) \approx \mathbb{P}(x_1)\mathbb{P}(x_h)$. This is significant because by the upper bound given in Equation 5.10, we know the reliability of the path would be upper bounded by $\mathbb{P}(x_1)\mathbb{P}(x_h)$.

This idea can also be extended to larger subsets of links. If subset of links $Q \in P$ within a path are all uncorrelated with one another, we can upper bound the reliability of the path as the product of the marginal reliabilities of the uncorrelated subset:

$$Rel(P) = \mathbb{P}\left(\bigcap_{i=1}^{h} x_i\right) \leq \prod_{i:e_i \in Q} \mathbb{P}(x_i) \qquad (5.10)$$

For large graphs with $d_{s,t} \gg d'$, this decorrelation of links along a path leads the lower bound, where correlation is disregarded, to be a much better approximation of the most reliable path compared to the upper bound. The following example demonstrates how disregarding correlation by computing path reliabilities as the product of marginal link reliabilities may be an appropriate strategy for enumerating the most reliable path when links are correlated under a distance dependent correlation model.
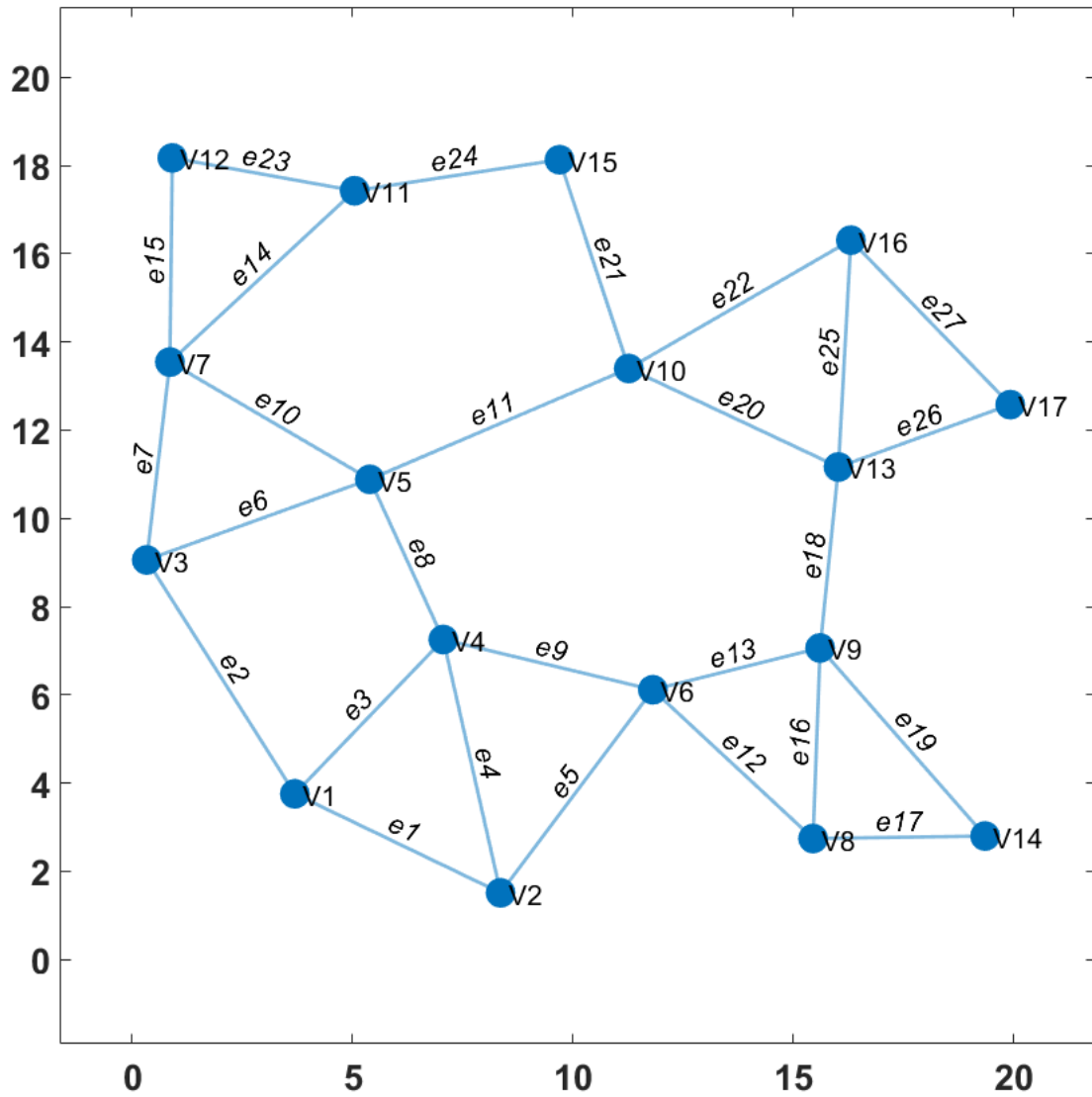
Figure 5-3: Large graph used for most reliable path analysis

Input Parameters	
T	10
σ	8
τ	0
l	5

Table 5.1: User inputs to generate MVN parameters via the the methods described in Section 4.6

The geometric MVN parameter generation method with inputs shown in Table 5.1 was used for the graph shown in Figure 5-3. The resulting univariate parameters and the resulting marginal reliabilities for each link are shown in Table 5.2. To give insight into the relationship between midpoint separation and the resulting link correlation, the midpoint separation distances for the first eight links are shown in Table 5.3 and the resulting correlation matrix is shown in Table 5.4.

All paths from v_1 to v_{17} were enumerated, and their reliabilities under the MVN model was computed. The product of link marginal reliabilities (lower bound) and the minimum marginal reliability (upper bound) for each path was also computed. Figure 5-4 provides the three most reliable paths in descending order of their reliability, as well as their bounds, marginal link reliabilities, and correlation.

For all paths, the correlation between links is only significant for neighboring links and this is true for paths even beyond the top three most reliable path. As a result the lower bounds are tighter than the upper bounds for the three most reliable links. Its also evident that the most reliable path also has the greatest lower bound and the greatest upper bound. The paths are presented in descending order of their reliabilities and this order would be preserved for the top three paths if they were ranked by their lower bounds. The same is not true for their upper bounds as there are paths beyond the top three whose upper bound is greater than the upper bound of the third most reliable path. These patterns are not unique to this particular example, and the fact that the greatest lower bound corresponded to the most reliable path was prevalent in all other simulations conducted for this thesis.

| Edge	End Nodes		Distance	μ_i	σ_i^2	τ_i	Marginal Reliability
e1	V1	V2	5.181	4.819	8	0	0.956
e2	V1	V3	6.279	3.721	8	0	0.906
e3	V1	V4	4.858	5.142	8	0	0.965
e4	V2	V4	5.890	4.110	8	0	0.927
e5	V2	V6	5.757	4.243	8	0	0.933
e6	V3	V5	5.383	4.617	8	0	0.949
e7	V3	V7	4.518	5.482	8	0	0.974
e8	V4	V5	3.998	6.002	8	0	0.983
e9	V4	V6	4.889	5.111	8	0	0.965
e10	V5	V7	5.258	4.742	8	0	0.953
e11	V5	V10	6.391	3.609	8	0	0.899
e12	V6	V8	4.960	5.040	8	0	0.963
e13	V6	V9	3.913	6.087	8	0	0.984
e14	V7	V11	5.714	4.286	8	0	0.935
e15	V7	V12	4.629	5.371	8	0	0.971
e16	V8	V9	4.324	5.676	8	0	0.978
e17	V8	V14	3.904	6.096	8	0	0.984
e18	V9	V13	4.125	5.875	8	0	0.981
e19	V9	V14	5.675	4.325	8	0	0.937
e20	V10	V13	5.257	4.743	8	0	0.953
e21	V10	V15	4.987	5.013	8	0	0.962
e22	V10	V16	5.818	4.182	8	0	0.930
e23	V11	V12	4.204	5.796	8	0	0.980
e24	V11	V15	4.702	5.298	8	0	0.969
e25	V13	V16	5.151	4.849	8	0	0.957
e26	V13	V17	4.147	5.853	8	0	0.981
e27	V16	V17	5.195	4.805	8	0	0.955

Table 5.2: Univariate distribution parameters and marginal reliabilities for the graph shown in Figure 5-3 using input parameters from Table 5.1

Dist	e1	e2	e3	e4	e5	e6	e7	e8
e1	0.00	5.51	2.94	2.43	4.23	7.99	10.23	6.44
e2	5.51	0.00	3.48	6.05	8.48	3.67	5.10	4.99
e3	2.94	3.48	0.00	2.59	5.01	5.13	7.52	3.67
e4	2.43	6.05	2.59	0.00	2.44	7.40	9.93	4.92
e5	4.23	8.48	5.01	2.44	0.00	9.50	12.09	6.52
e6	7.99	3.67	5.13	7.40	9.50	0.00	2.63	3.48
e7	10.23	5.10	7.52	9.93	12.09	2.63	0.00	6.06
e8	6.44	4.99	3.67	4.92	6.52	3.48	6.06	0.00

Table 5.3: Distances between the midpoints for links e_1 through e_8

Corr	e1	e2	e3	e4	e5	e6	e7	e8
e1	1.000	0.545	0.841	0.889	0.699	0.279	0.123	0.436
e2	0.545	1.000	0.784	0.481	0.237	0.764	0.595	0.608
e3	0.841	0.784	1.000	0.874	0.606	0.591	0.323	0.764
e4	0.889	0.481	0.874	1.000	0.887	0.334	0.139	0.616
e5	0.699	0.237	0.606	0.887	1.000	0.165	0.054	0.427
e6	0.279	0.764	0.591	0.334	0.165	1.000	0.871	0.784
e7	0.123	0.595	0.323	0.139	0.054	0.871	1.000	0.480
e8	0.436	0.608	0.764	0.616	0.427	0.784	0.480	1.000

Table 5.4: Correlation matrix for e_1 through e_8

Path #1
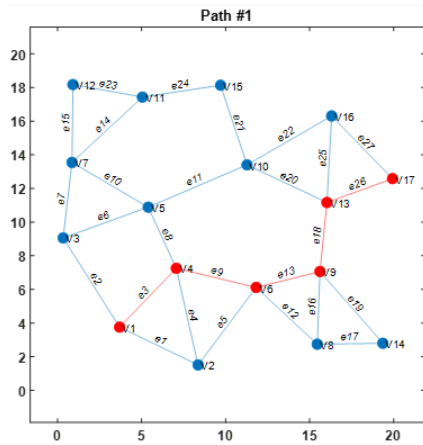
Reliability		Lower Bound		Upper Bound	
0.912		0.882		0.965	

Link	e3	e9	e13	e18	e26
Marginal Reliability	0.965	0.965	0.984	0.981	0.981

Corr	e3	e9	e13	e18	e26
e3	1.00	0.70	0.24	0.09	0.02
e9	0.70	1.00	0.69	0.39	0.14
e13	0.24	0.69	1.00	0.81	0.40
e18	0.09	0.39	0.81	1.00	0.78
e26	0.02	0.14	0.40	0.78	1.00

Path #2
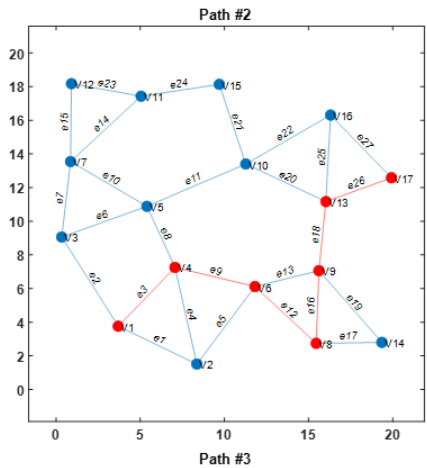
Reliability		Lower Bound		Upper Bound	
0.892		0.846		0.963	

Link	e3	e9	e12	e16	e18	e26
Marginal Reliability	0.965	0.965	0.963	0.978	0.981	0.981

Corr	e3	e9	e12	e16	e18	e26
e3	1.00	0.70	0.25	0.13	0.09	0.02
e9	0.70	1.00	0.64	0.45	0.39	0.14
e12	0.25	0.64	1.00	0.93	0.59	0.23
e16	0.13	0.45	0.93	1.00	0.70	0.34
e18	0.09	0.39	0.59	0.70	1.00	0.78
e26	0.02	0.14	0.23	0.34	0.78	1.00

Path #3
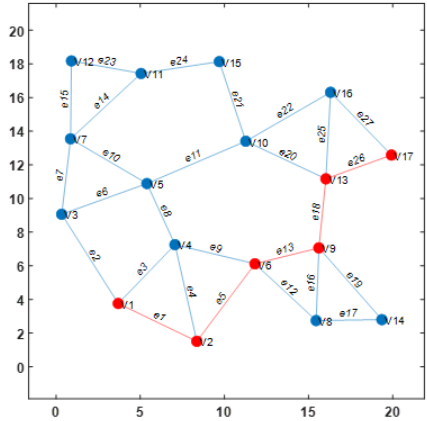
Reliability		Lower Bound		Upper Bound	
0.882		0.845		0.933	

Link	e1	e5	e13	e18	e26
Marginal Reliability	0.956	0.933	0.984	0.981	0.981

Corr	e1	e5	e13	e18	e26
e1	1.00	0.70	0.22	0.06	0.01
e5	0.70	1.00	0.66	0.30	0.08
e13	0.22	0.66	1.00	0.81	0.40
e18	0.06	0.30	0.81	1.00	0.78
e26	0.01	0.08	0.40	0.78	1.00

Figure 5-4: Three most reliable paths with their reliabilities, lower/upper bounds, link marginal reliabilties, and correlation between all links within the path

5.4   Conclusions and Future Work

The results in this chapter characterize the Most Reliable Path Problem under the MVN model and provides a method to enumerate its solution. However, this method relies on the explicit enumeration of all paths thus may take exponential time in large graphs. The bounding methods offer insight into the problem and may offer opportunities to prune the set of all possible paths needed to be considered.

The lack of optimal substructure introduced by correlation makes this problem inherently different than the problem under the Bernoulli model. Much like the Network Reliability Problem, we believe simulation based methods are the best direction for providing computationally efficient solutions to this problem.

THIS PAGE INTENTIONALLY LEFT BLANK

Bibliography

[1] ITU-R Study Group 3. *RECOMMENDATION ITU-R P.533-14 - Method for the prediction of the performance of HF circuits**.

[2] ITU-R Study Group 3. *Ionosphere and its Effects on Radiowave Propagation.* 1998.

[3] P. Agrawal and N. Patwari. Correlated link shadow fading in multi-hop wireless networks. *IEEE Transactions on Wireless Communications*, 8(8):4024–4036, Aug 2009.

[4] J. A. Assiri. *Development of Dependent Failure Reliability Models for Distributed Communication Networks.* PhD thesis, Oregon State University, 1980.

[5] Bureau of Meteorology Australian Government. Introduction to hf radio propagation. https://www.sws.bom.gov.au/Category/Educational/Other20Topics/Radio20Communication/Intro20to20HF20Radio.pdf.

[6] W. Batts, W. Furman, and E. Koski. Empirically characterizing channel quality variation on hf ionospheric channels. In *HF 07, The 8th Nordic Conference on HF Communications*, Aug 2007.

[7] D. Bertsekas and R. Gallager. *Data Networks (2nd Ed.).* Prentice-Hall, Inc., USA, 1992.

[8] D. Bertsimas and J.N. Tsitsiklis. *Introduction to Linear Optimization.* Athena Scientific series in optimization and neural computation. Athena Scientific, 1997.

[9] J. Blanchet and P. Glynn. Efficient rare-event simulation for the maximum of heavy-tailed random walks. *The Annals of Applied Probability*, 18(4):1351–1378, 2008.

[10] Z. I. Botev, P. L'Ecuyer, G. Rubino, R. Simard, and B. Tuffin. Static network reliability estimation via generalized splitting. *INFORMS Journal on Computing*, 25(1):56–71, Feb 2013.

[11] Z. I. Botev, P. L'Ecuyer, and B. Tuffin. Dependent failures in highly reliable static networks. In *Proceedings of the 2012 Winter Simulation Conference (WSC)*, page 1–12, Dec 2012.

[12] A. Broder. Generating random spanning trees. In *Proceedings of the 30th Annual Symposium on Foundations of Computer Science*, SFCS '89, page 442–447, USA, 1989. IEEE Computer Society.

[13] M. Cheffena, L. E. Braten, and T. Ekman. On the space-time variations of rain attenuation. *IEEE Transactions on Antennas and Propagation*, 57(6):1771–1782, Jun 2009.

[14] C. J. Colbourn. *The Combinatorics of Network Reliability*. Oxford University Press, Inc., USA, 1987.

[15] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to algorithms*. MIT Press, 3rd ed edition, 2009.

[16] A.J. Coulson, A.G. Williamson, and R.G. Vaughan. A statistical basis for log-normal shadowing effects in multipath fading channels. *IEEE Transactions on Communications*, 46(4):494–502, Apr 1998.

[17] Z. I. Botev D. P. Kroise, T. Taimre. *Handbook of Monte Carlo Methods*. John Wiley & Sons, Ltd, 2011.

[18] N. C. Davies. *Digital Radio and Its Application in the HF (2-30MHz) Band*. PhD thesis, The University of Leeds, May 2004.

[19] R. G. Gallager. *Stochastic Processes: Theory for Applications*. Cambridge University Press, Dec 2013.

[20] J. Goodman, J. Ballard, and E. Sharp. A long-term investigation of the hf communication channel over middle- and high-latitude paths. *Radio Science*, 32(4):1705–1715, Aug 1997.

[21] J. M. Goodman. *HF Communications: Science and Technology*. Nostrand Reinhold, 1992.

[22] J. M. Goodman. Availability correlation distances for hf communications in a global network. In *Seventh International Conference on HF Radio Systems and Techniques*, page 140–144, Jul 1997.

[23] G. Hendrantoro, R.J.C. Bultitude, and D.D. Falconer. Use of cell-site diversity in millimeter-wave fixed cellular systems to combat the effects of rain attenuation. *IEEE Journal on Selected Areas in Communications*, 20(3):602–614, Apr 2002.

[24] ITU. *Technical and operational characteristics of land mobile MF/HF systems*. Number M. 1795. Mar 2007.

[25] E. E. Johnson, E. Koski, W. N. Furman, M. Jorgenson, and J. Nieto. *Third-generation and Wideband HF radio communications*. Artech House mobile communications series. Artech House, 2013.

[26] I. P. Kaminow and T. L. Koch. *Optical Fiber Telecommunications III*, volume 3A. Academic Press, 1st edition.

[27] D. R. Karger and C. Stein. A new approach to the minimum cut problem. *Journal of the ACM (JACM)*, 43(4):601–640, Jul 1996.

[28] R. M Karp, M. Luby, and N. Madras. Monte-carlo approximation algorithms for enumeration problems. *Journal of Algorithms*, 10(3):429–448, Sep 1989.

[29] Y. F. Lam and V. K. Li. Reliability modeling and analysis of communication networks with dependent failures. *IEEE Transactions on Communications*, 34(1):82–84, Jan 1986.

[30] H. Lee, E. Modiano, and K. Lee. Diverse routing in networks with probabilistic failures. *IEEE/ACM Transactions on Networking*, 18(6):1895–1907, Dec 2010.

[31] L. Liu, B. Zhao, W. Wan, B. Ning, M. Zhang, and M. He. Seasonal variations of the ionospheric electron densities retrieved from constellation observing system for meteorology, ionosphere, and climate mission radio occultation measurements. *Journal of Geophysical Research: Space Physics*, 114(A2), 2009.

[32] L. F. McNamara. Spatial correlations of fof2 deviations and their implications for global ionospheric models: 2. digisondes in the united states, europe, and south africa. *Radio Science*, 44(2), 2009.

[33] F. Moskowitz. The analysis of redundancy networks. *Transactions of the American Institute of Electrical Engineers, Part I: Communication and Electronics*, 77(5):627–632, Nov 1958.

[34] S. Neumayer and E. Modiano. Network reliability with geographically correlated failures. In *2010 Proceedings IEEE INFOCOM*, page 1–9. IEEE, Mar 2010.

[35] S. Neumayer, G. Zussman, R. Cohen, and E. Modiano. Assessing the vulnerability of the fiber infrastructure to disasters. *IEEE/ACM Transactions on Networking*, 19(6):1610–1623, Dec 2011.

[36] H. H. Nguyen, K. Palani, and D. M. Nicol. An approach to incorporating uncertainty in network security analysis. In *Proceedings of the Hot Topics in Science of Security: Symposium and Bootcamp*, HoTSoS, page 74–84. Association for Computing Machinery, Apr 2017.

[37] H. H. Nguyen, K. Palani, and D. M. Nicol. Extensions of network reliability analysis. In *2019 49th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, page 88–99, Jun 2019.

[38] Cowpip (C. Oler). Signal ray-tracing using proplab-pro 3 showing the formation of a skip zone. `https://commons.wikimedia.org/wiki/File:Skip_Zone_Example.gif/`, 2008.

[39] S. N. Pan and J. D. Spragins. Dependent failure reliability models for tactical communications networks. In *Proceedings of International Conference on Communications*, page 765–771, 1983.

[40] Phirosiberia. Ionospheric layers. `https://commons.wikimedia.org/wiki/File:Ionosphere_Layers_en.svg`, Jun 2009.

[41] M. Rahnamay-Naeini, J. E. Pezoa, G. Azar, N. Ghani, and M. M. Hayat. Modeling stochastic correlated failures and their effects on network reliability. In *2011 Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN)*, page 1–6. IEEE, Jul 2011.

[42] Theodore Rappaport. *Wireless Communications: Principles and Practice*. Prentice Hall PTR, USA, 2nd edition, 2001.

[43] C. E. Rasmussen and C. K. I. Williams. *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, 2006.

[44] C. M. Rush. An ionospheric observation network for use in short-term propagation predictions. *ITU Telecommunication Journal*, 43:544–549, Aug 1976.

[45] J. D. Spraggins and J. A. Assiri. Communication network reliability calculations with dependent failures. In *Proceedings of National Telecommunications Conference*, pages 25.2.1–25.2.5, 1980.

[46] G.L. Turin, F.D. Clapp, T.L. Johnston, S.B. Fine, and D. Lavry. A statistical model of urban multipath propagation. *IEEE Transactions on Vehicular Technology*, 21(1):1–9, Feb 1972.

[47] C. Watterson, J. Juroshek, and W. Bensema. Experimental confirmation of an hf channel model. *IEEE Transactions on Communication Technology*, 18(6):792–803, Dec 1970.

[48] D. B. Wilson. Generating random spanning trees more quickly than the cover time. In *Proceedings of the Twenty-Eighth Annual Acm Symposium on the Theory of Computing*, page 296–303. ACM, 1996.

[49] F. Yaghoubi, M. Furdek, A. Rostami, P. Oehlen, and L. Wosinska. Reliable topology design of wireless networks under correlated failures. In *2018 IEEE International Conference on Communications (ICC)*, page 1–6, May 2018.

[50] F. Yaghoubi, M. Furdek, A. Rostami, P. Oehlen, and L. Wosinska. Reliable topology design of wireless networks under correlated failures. In *2018 IEEE International Conference on Communications (ICC)*, page 1–6, May 2018.

[51] X. Yue, W. Wan, L. Liu, and T. Mao. Statistical analysis on spatial correlation of ionospheric day-to-day variability by using gps and incoherent scatter radar observations. *Annales Geophysicae*, 25(8):1815–1825, Aug 2007.