Analysis and Detection of Jamming Attacks in an All-Optical Network

by

Poompat Saengudomlert

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1998

Author ....
Department of Electrical Engineering and Computer Science
May 1, 1998

Certified by ....
Dr. Muriel Médard
MIT Lincoln Laboratory Staff
Thesis Supervisor

Certified by ....
Robert G. Gallager
Professor of EECS
Thesis Supervisor

Accepted by ....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Analysis and Detection of Jamming Attacks in an All-Optical Network

by

Poompat Saengudomlert

Submitted to the Department of Electrical Engineering and Computer Science
on May 1, 1998, in partial fulfillment of the
requirements for the degree of
Master of Science

Abstract

Existing supervisory schemes which can be used for detecting jamming attacks in all-optical networks are mostly based on average power detection and the use of pilot signals. We point out that these methods are not adequate to handle all types of jamming attacks. Based on a newly designed attack detection device at a network node, a novel attack detection scheme against jamming attacks is proposed. Preliminary results on the performance of the proposed attack detection scheme are given for both in-band and out-of-band jamming attacks. To analyze the performance of the attack detection scheme in further detail, we find, under certain conditions, the worst case in-band jamming attack scenarios which correspond to the smallest probabilities of being detected for a given value of average degraded bit error rate. We then investigate how our proposed attack detection scheme performs under the worst case in-band jamming attack scenarios. Finally, we make some modification to our attack detection scheme to improve its performance.

Thesis Supervisor: Dr. Muriel Médard
Title: MIT Lincoln Laboratory Staff

Thesis Supervisor: Robert G. Gallager
Title: Professor of EECS

Acknowledgments

I wish to express my utmost gratitude to Dr. Muriel Médard who closely supervised my master thesis from the very beginning to the very end. Her invaluable advice, her patience as well as her kindness helped make this work possible.

I also wish to thank professor Robert Gallager for his insightful comments on various parts of the thesis, professor Amos Lapidoth for the research experience in my first year of the graduate study, and the MIT EECS Department together with the Advanced Networks Group of MIT Lincoln Laboratory who provided me with financial supports.

This master thesis is devoted to my beloved family and my dearest Chaliga.

Contents

List of Figures

List of Tables

Chapter 1

Motivation for a Novel Jamming Attack Detection Scheme

In this chapter, we outline the basic information on all-optical networks (AONs) that is relevant to the development of an attack detection scheme. Throughout this work, we consider only attacks upon the AON infrastructure. We shall explain why the detection of such attacks in an AON is an important problem. In addition, we explain how the problem of attack detection in an AON is different from the similar problems in electro-optic networks, and from the detection of network failures.

We shall introduce two major attack categories, jamming and eavesdropping, together with existing network supervisory mechanisms which can be used as attack detection schemes. This work concentrates on the detection of jamming attacks. In the final section, we argue that existing supervisory mechanisms are not sufficient to handle all types of jamming attacks.

1.1 Introduction to AONs

1.1.1 Basic Concepts of AONs

In an AON, data transmission within the network is entirely in the optical domain. Aside from the user-to-network interfaces (equivalently the access points), there are no optical-to-electronic or electronic-to-optical conversions of transmitted signals at all nodes within the network. However, there is usually a separate supervisory network, which may not be all-optical, for management and control purposes.

One major advantage of AONs with respect to their electro-optic counterparts is their much higher bandwidth. In an electro-optic network, the fiber medium is used in place of copper (or other types of transmission medium) for data transmission between network nodes. When data signals arrive at a network node, they are converted to electrical signals which can then be processed for the purposes of routing, signal regeneration, error correction, and so on. While an electro-optic network such as the Fiber Distributed Data Interface (FDDI) can transmit up to the order of 100 megabits per second [7], recent demands for gigabit applications such as video conferencing and remote access to medical images have indicated the need for an even higher network capacity.

The transmission rate of an electro-optic network is limited by the so-called "electronic bottleneck", which refers to the time consuming processes of optical-to-electronic conversion, electrical signal processing, and electronic-to-optical conversion of data signals at network nodes. In this regard, an AON is a natural solution to the need of higher network capacity since transmitted signals remain optical throughout the network. At present, various testbeds and laboratory experiments have achieved an aggregated throughput of over 1 terabit per second in AONs [10, 8, 22]. Higher network throughputs are expected in the future since much of the 25-terahertz fiber bandwidth remains unused.

Most existing research AONs can be divided into two groups based on the network multiple-access scheme. Simultaneous transmissions on the same fiber can be accomplished by using either different optical wavelengths in wavelength division multiple-access (WDMA) or different time slots in time division multiple-access (TDMA) or a combination of both. Compared to WDMA, TDMA requires synchronization and is therefore more vulnerable to dispersion effects. In addition, the use of TDMA in an AON requires further research and development [6]. This work will therefore concentrate on WDMA networks.

In a WDMA network, the fiber bandwidth is divided into optical "wavebands" each of which can currently support the transmission rate up to 10 gigabits per second (capacity of the standard Optical Carrier 192 (OC-192)) [20]. The process of assigning a waveband which is routed through an AON to establish an end-to-end optical transmission path is called "wavelength routing".

Wavelength routing provides a "transparent" transmission path. Transmitted signals may be amplified but are not processed or regenerated at network nodes. To the network nodes, it makes no difference whether transmitted signals are analog or digital, or whether the modulation scheme is amplitude shift keying (ASK) or phase shift keying (PSK). Consequencely, an AON can simultaneously support different types of data traffic regardless of their formats or modulation schemes. In addition, transparency provides backward compatibility since new types of data traffic can be supported by existing networks.

Its high throughput and its ability to support various types of network traffics suggest that an AON can serve as a backbone for a broadband network designed to support diversity of applications by simultaneously carrying data, voice, images and video.

For convenience in future discussion on AONs, we shall refer to a connection between two adjacent network nodes as a "link", and a connection between any two network nodes as a "light path" between them. In general, a light path can consist of a single link or multilple links.

1.1.2 Basic Components of AONs

Basic physical components of AONs are optical terminals (OTs), transmission lasers, optical receivers, fiber cables, splitters and combiners, switches, multiplexers and demultiplexers, and optical amplifiers. These components are used to construct AONs consisting of user terminals at access points (APs), optical links and network nodes.

OTs are user devices attached to an AON through the AP interface. They are normally electro-optic devices that transmit data to an AON using transmission lasers and/or receive data from an AON using optical receivers. Optical receivers are commonly made of photodetectors which are square-law detectors. Consequently, direct detection is easily implemented and is used in several AON testbeds instead of coherent detection. We shall address the distinction between direct detection and coherent detection in the next section.

Fiber cables are used to transmit signals across the links between network nodes. In the mostly used wavelength vicinity of 1.5 micrometer, signals travelling along the fiber undergo attenuation of about 0.2 decibel per kilometer [14]. Accordingly, there is a need for optical amplifiers to compensate for signal attenuation on fiber links. In addition, signals travelling at different wavelengths on the same fiber can interact with one another. In particular, a signal travelling at one wavelength may cause amplification or attenuation of a signal at another wavelength on the same fiber [20].

A combiner combines optical signals from multiple fibers onto a single fiber. A splitter does the opposite by splitting a signal from a single fiber to multiple fibers. A demultiplexer assigns each of multiple signals (at different wavelengths) on the same fiber to its individual output fiber. A multiplexer combines individual signals (at different wavelengths) from multiple fibers onto a single fiber. A switch is used to route optical signals from one fiber to another. Examples of switches are add-drop multiplexers (ADMs) and wavelength selective switches (WSSs) used in WDMA networks. In general, a switch can be mechanical, opto-electronic, or optical. These components are commonly used at network nodes. One undesirable characteristic is component crosstalk, which allows a portion of a signal travelling through a switch at a network node to be present on another output fiber link belonging to a different light path.

Optical amplifiers are used mainly to compensate for signal attenuation along fiber links. They are also used at network nodes to compensate for losses at switching components. In particular, pre-amplifiers are optical amplifiers used to amplify the input to the network node. Conversely, post-amplifers are optical amplifers used at the output of the network node. At present, the most commonly used optical amplifiers are erbium doped fiber amplifiers (EDFAs) which operate well in the 1.5-micrometer fiber bandwidth.

In addition to naturally occurring noise in passive AON components, EDFAs generate amplified spontaneous emission (ASE) noise which is a dominant noise source in EDFAs. Another important characteristic is the dependency of the EDFA gain on the component noise variance. Moreover, signals in different wavebands on the same fiber can cause "gain competition" at an EDFA by sharing the limited pool of upper-state photons within the same fiber [3].

1.1.3 Modulation and Detection Schemes in AONs

There exist several modulation schemes for the transmission of digital data. In the case of data transmission through telephone lines, the limited bandwidth necessitates the use of higher order symbols, as opposed to binary symbols, to increase the transmission rate. For example, in quatenary phase shift keying (QPSK), a single transmission can represent one of four possible signal values.

In AONs, it is a common practice to represent digital data as streams of binary digits and transmit them using on-off keying (OOK), which is a special case of amplitude shift keying (ASK). In OOK, the light source is turned off during the 0-bit transmission and turned on during the 1-bit transmission. Furthermore, the ON signal of OOK can have either a non-return to zero (NRZ) pulse shape or a return to zero (RZ) pulse shape.

There are several reasons for the choice of OOK in AONs. First of all, modulating lasers is difficult. Because of the simple structure of OOK, it is easier to modulate lasers based on OOK than on other transmission schemes. Moreover, detection of OOK signals can easily be done using photodetectors which are usually square-law detectors. Finally, owing to the enormous bandwidth of the fiber medium, we do not have a restriction on bandwidth; and representing digital data as binary streams is sufficient for practical purposes.

We shall adopt OOK as the transmission scheme throughout this work. Furthermore, we assume a binary data source with equal probabilities of "0" and "1", and the transmission rate of 1 gigabit per second using 1 gigahertz bandwidth.

In AONs, detection of transmitted bits is generally done using square-law magnitude detection of received signals. The decision rule for the receiver can be based on whether the output of a square-law detector at the receiver exceeds a certain threshold. This method is referred to as "direct detection". The right hand side of figure 1-1 shows the decision regions for a decision rule based on direct detection.

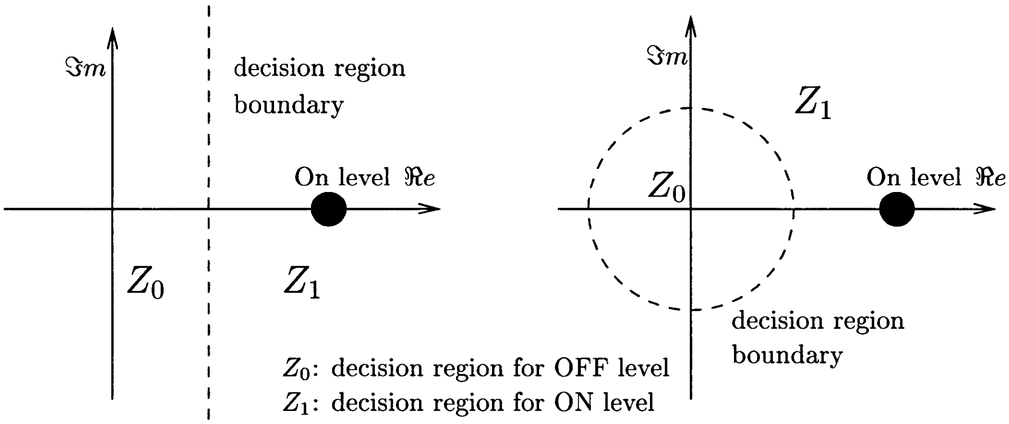


Figure 1-1: Receiver decision regions for coherent detection (left diagram) and for direct detection (right diagram).

Based on the knowledge of both real and imaginary parts of the received signals, the receiver can use a different decision rule whose corresponding decision regions are shown in the left hand side of figure 1-1. This detection method is referred to as "coherent detection". Coherent detection yields a lower bit error rate (BER) than direct detection for the same value of signal-to-noise ratio (SNR). However, the implementation of coherent detection requires additional hardware because photodetectors are square-law detectors.

Typical values of BERs under normal operation in AONs are lower than 10^{-12}; and are much lower than those of non-optical networks. It is important to keep this quantity in mind so that we are aware of the operating regions (in terms of BERs) in the remaining chapters.

1.2 Major Attack Categories

This work concentrates on intentional attacks upon the infrastructure of an AON. By an "attack", we mean an action performed on the infrastructure of an AON with an intention to disrupt network services and/or obtain unauthorized information.

There exist several methods of attacks upon the infrastructure of an AON. However, these attacks can be grouped into two categories for the purpose of attack detection. We consider two categories: eavesdropping and jamming [16].

1.2.1 Eavesdropping

Eavesdropping refers to the attempts of an attacker trying to obtain unauthorized information from an AON. For example, an attacker can obtain the desired information by "tapping", which refers to the method of bending the fiber to change an incidence angle and allowing light to escape from a fiber core [5, 26, 27]. Alternatively, an eavesdropper can obtain the desired information from a "crosstalk channel", which refers to an adjacent output link (leaving from the network node shared by a legitimate light path) that contains portions of legitimate data signals as a result of component crosstalk at the network node.

1.2.2 Jamming

A jamming attack can directly alter data signals if jamming signals have the same frequencies as data signals. We refer to this case as "in-band jamming". Conversely, an "out-of-band jamming" attack occurs when a jamming signal is out of the transmission band but can still degrade the data signals. For example, an out-of-band jamming signal at an EDFA can cause gain competition and attenuate data signals at other frequencies.

This work concentrates on the detection of jamming attacks. Although jamming can occur at fiber links as well as at network nodes, we shall consider only jamming at network nodes since there exist several ways of constructing optical-fiber transmission systems with good protection against line tapping [5, 27].

Note that we do not worry about jamming attacks which do not change the BER significantly. An example of such attacks is in-band jamming with additive low-variance white Gaussian noise at a single network node.

In-Band Jamming Attacks

We concentrate on the case in which an attacker gains legitimate access to a network node (or multiple network nodes) and inserts in-band jamming signals onto a legitimate light path using component crosstalk at a network node. Figure 1-2 illustrates an in-band jamming attack via crosstalk at a network node. Note that jamming can also occur at several network nodes simultaneously.

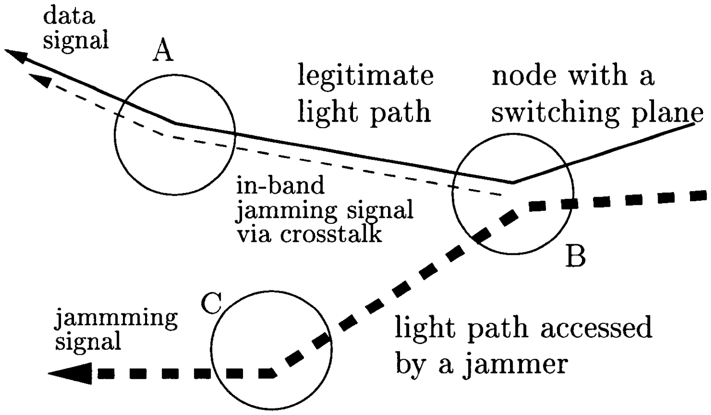


Figure 1-2: Illustration of in-band jamming via component crosstalk at a network node. Node B is the scene of an attack. Node A demonstrates the propagation of the effects of such an attack.

In electronic or electro-optic networks, component crosstalk is significantly lower than 0.03% of the signal level, while the same quantity in an AON is between 0.03% to 1.0% [9]. Since the level of component crosstalk in an AON is higher than in an electro-optic network, an attack using the same set of jamming signals can be more detrimental to the BER in an AON than in its electro-optic counterpart. However, jamming via component crosstalk at network nodes requires high jamming signal power (compared to the power level under normal operation) to affect data signals in a legitimate light path significantly.

Out-of-Band Jamming Attacks Causing Gain Competition at EDFAs

An out-of-band jamming attack can cause gain competition at EDFAs and attenuate data signals. Figure 1-3 illustrates this type of out-of-band jamming attacks. Amplifier gain fluctuation during an attack depends on the type of EDFAs used. Throughout this project, we assume the use of gain-clamped EDFAs in which internal out-of-band oscillation is used to clamp the gain [17].

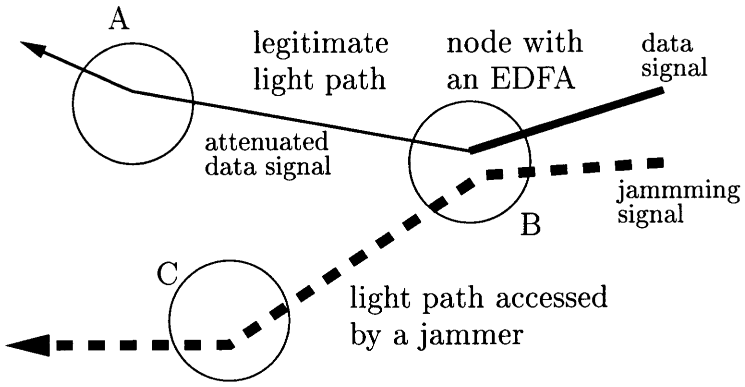


Figure 1-3: Illustration of out-of-band jamming causing gain competition at an EDFA at a network node. Node B is the scene of an attack. Node A demonstrates the propagation of the effects of such an attack.

Figure 1-4 shows gain fluctuation for an EDFA under an out-of-band jamming attack. Figure 1-5 shows similar information for a gain-clamped EDFA.

As in the case of in-band jamming, an attacker can cause gain competition at multiple network nodes simultaneously. We shall discuss the effects of jamming at multiple network nodes in further detail later on.

1.3 Motivation for an Attack Detection Scheme

Before we consider the method of attack detection, it is important to understand why the detection of attacks in an AON is an important problem.

One important characteristic of an AON is its "transparency", which refers to the absence of detection and processing (besides amplification) of transmitted signal at network nodes. Transparency allows an AON to support different types of data traffics simultane-

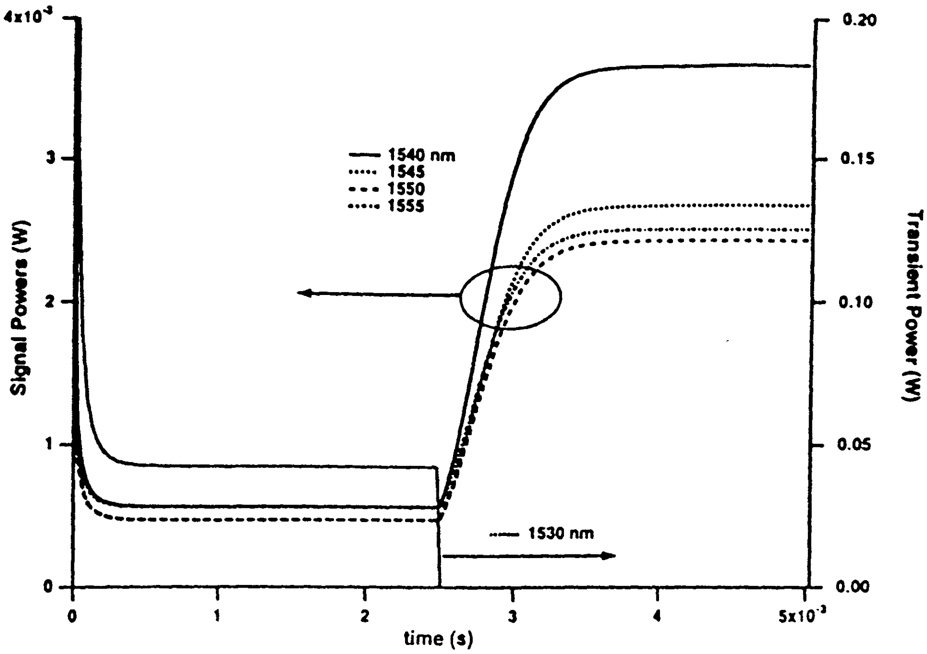


Figure 1-4: EDFA gains of signals at 4 legitimate wavelengths, each with input power 10^{-4} W. There are 16 legitimate wavelengths evenly spaced between 1540 nm and 1555 nm. An attack is a uniform pulse which occurs at 1530 nm (out-of-band) and lasts for 2.5 ms.

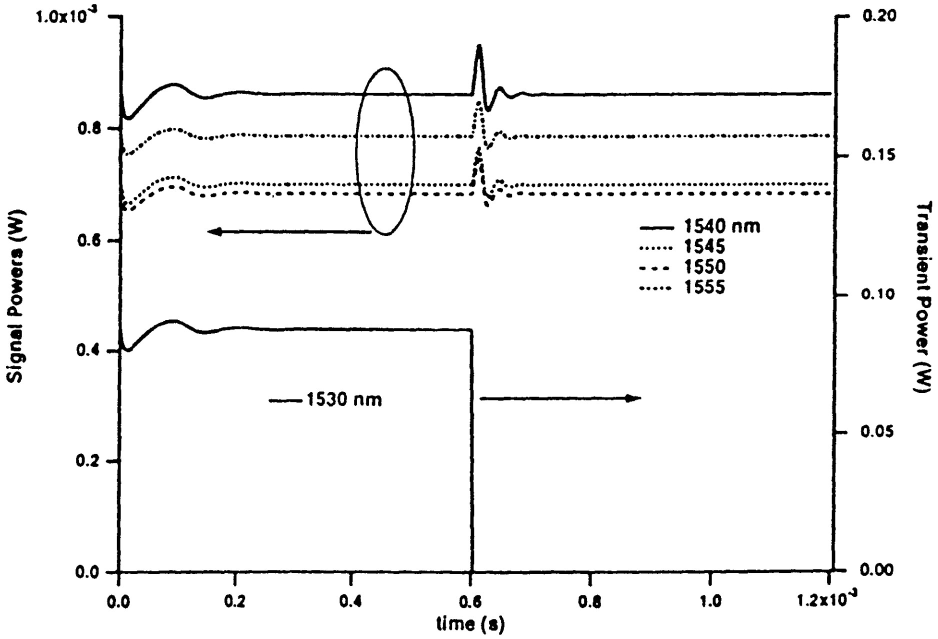


Figure 1-5: Gain-clamped EDFA gain of signals at 4 legitimate wavelengths, each with input power 10^{-4} W. A uniform pulse attack occurs at 1,530 nm (out-of-band) with power 10^{-4} W and lasts for 2.5 ms.

ously regardless of their formats or modulation schemes. In general, the network manager does not know whether a particular data bit stream is coded for error control, or what type of coding scheme is used. Apart from an error control mechanism between the transmitter and the receiver, there is no error detection or error control mechanism for the transmission across individual fiber links. While an error detection mechanism between the transmitter and the receiver may tell us about the presence of an attack, it cannot provide information regarding the location of an attack.

An important consequence of transparency in an AON is "attack propagation", which refers to the propagation of the effects due to an attack at a network node to downstream nodes. In an AON, attenuated or degraded data signals propagate through downstream nodes without being processed. Therefore, the effects due to an attack at a single network node will not be removed by any downstream node.

As a result, simultaneous jamming at multiple network nodes can provide accumulated effects along a light path. For in-band jamming, coherent components of jamming signals at multiple network nodes add up constructively along the light path. In general, intentional coherent in-band jamming at multiple network nodes is hard to perform since an attacker needs to have complete timing information.

For out-of-band jamming that causes gain competition at EDFAs, the effects due to simultaneous jamming at multiple network nodes always add up constructively. In other words, the SNR degradation increases as we move towards downstream nodes.

Attack propagation allows an attacker to insert relatively small in-band or out-of-band jamming signals at multiple network nodes and degrade the overall BER along the light path significantly.

Network transparency and high component crosstalk in AONs raise several security issues which do not exist in electro-optic networks. To provide a secure network connection in an AON, there is a need for the detection of attacks. The detection of attacks which disrupt AON services will serve as an error detection mechanism in the data link layer. On the other hand, the detection of eavesdropping will protect sensitive information. An attack detection system will also provide information regarding the location of an attack. The information is useful for further actions such as rerouting.

Owing to unique characteristics of AONs, the detection of attacks in an AON requires special consideration and a new solution.

1.4 Requirements of an Attack Detection Scheme

In this section, we describe desirable characteristics of an attack detection scheme. In general, an attack detection scheme in an AON should satisfy all of the following requirements.

- Is applicable to all types of data traffic.
- Identifies attacked network nodes.
- Detects an attack in a short period of time.
- Does not require complicated network hardware upgrades.

We shall address each requirement separately.

1.4.1 Applicability to All Types of Data Traffic

Because of transparency in an AON, an attack detection scheme cannot depend on specific characteristics of the data traffic, such as a coding scheme and a transmission rate, since there is no detection of transmitted signals within a network (refer the discussion in section 1.3).

1.4.2 Identification of Attacked Network Nodes

In order to perform appropriate actions after detecting an attack, we may need to know the location of an attack. For example, if we know the locations of network nodes which are under attack, we can reroute the light path to avoid attacked nodes. [18] presents an algorithm for attack localization in an AON provided that we have a detection scheme for detecting whether a particular node is under attack.

1.4.3 Time Requirement of an Attack Detection Scheme

In a high data rate AON, we need to detect an attack quickly for several reasons. The first reason is that the effects of a very short attack (e.g. an attack on a single data bit) become less and less apparent over time as the observation period increases. If the required observation period in our attack detection scheme is too long, a very short attack will not be detected. We shall demonstrate that it is harder to detect an attack in which relatively few number of bits are attacked than to detect an attack in which a large number of bits are attacked.

The second reason is that a large amount of data can be affected by an attack which lasts for a relatively short time. Another reason is the large network "latency", which refers to the number of bits in flight or equivalently the number of bits under transmission within the network at a given time. We provide detailed explanation in what follows.

Consider the consequences of a high data rate for the detection of attacks in an AON. In the case of eavesdropping, we want to find out quickly whether or not an eavesdropper is present since a longer detection time means a larger amount of data being compromised.

For a jamming attack corrupting the BER, we can perform retransmission of data after the detection of an attack. Doing so would require memory buffers at the transmitter and possibly at the receiver. After detecting an attack, we may want to retransmit all data bits under transmission during the observation period for the detection of that particular attack (since we do not know which portion of bits are corrupted). Owing to the large network latency, the number of bits we want to store in memory can be very large. For example, if we take one second to detect a jamming attack in a system with a transmission rate of 1 gigabit per second, we may need to retransmit 1 gigabit of data. Although the time used in retransmitting bits may not be a problem, the amount of memory buffers used to store data bits can be very large; and the associated cost can be very high.

For real time applications in which there is no retransmission of data, we may still want to detect an attack quickly so that the network supervisor can reroute the data traffic onto an alternate light path before sacrificing a large amount of data (provided that such a fast rerouting scheme exists).

1.4.4 Additional Network Upgrades

It is desirable to build an attack detection system in an AON with little network modification. As we shall describe in chapter 2, our proposed attack detection system is constructed from attack detection devices proposed in [16]. The proposed device is a "wrapper", which refers to a wrap-around device to be installed at a network node.

The use of wrap-around devices at network nodes allows a construction of an attack detection system with relatively little network modifications. In addition, we can choose to install these wrappers only in certain parts of an AON which are susceptible to attacks.

Finally, it is important to distinguish between the problem of attack detection and the detection of network failures. The quality of service can be significantly degraded without the presence of network failure. For example, an attacker can perform in-band jamming and increase the BER by several orders of magnitude. A failure detection mechanism will not be able to recognize such an intentional attack since an attack does not cause any malfunction in the network. We shall see in the next section that existing supervisory mechanisms are more apt to detect network failures than to detect intentional attacks.

1.5 Existing Supervisory Mechanisms

Existing supervisory mechanisms which could be used as attack detection schemes are power detection, optical spectral analysis, pilot tone, optical time domain reflectrometry, and bit error rate testing. We shall discuss each method separately. In particular, we shall demonstrate why each method is not sufficient for detecting jamming attacks in an AON. Some comments on the detection of eavesdropping are included for completeness of the discussion. [16] provides the basic ideas outlined in this section.

1.5.1 Power Detection

In power detection, we measure the average signal optical power over a bandwidth of a single waveband or across multiple wavebands, and compare the measured power to the estimated level. If the difference between the measured power and the estimated level is beyond a certain threshold, an attack yielding that amount of change in the power level will be detected.

Power detection has been used to detect network failures such as amplifier failures in an AON [12, 24]. It can be used to detect an eavesdropper who taps off a significant amount of signal power from a legitimate channel [27], or a jammer who introduces in-band jamming signals which significantly alter the average power of transmitted signals. It can also be used to detect significant EDFA gain degradation due to out-of-band jamming attacks.

Power detection techniques rely on the computation of statistical averages which requires an observation of a large number of samples. For the detection scheme to be able to distinguish between the presence and the absence of a jamming attack, an observation of a sufficiently large number of corrupted bits is required. If an attacker performs in-band sporadic jamming to degrade the BER to an unacceptable level without changing the average signal power enough to cause an alarm, the detection scheme based on power detection techniques will not detect such an attack.

As a specific example, an attacker can jam at only one bit in every period of 10 microseconds. At the transmission rate of 1 gigabit per second, jamming signals are present only 0.01% of the time and the average power of data signals together with jamming signals may not be much different from the average power of data signals alone since the power of jamming signals is not significant "on average", i.e. the power of jamming signals is averaged out over a large number of observed bits. As a result, such an attack may not be detected. Note that if such a sporadic attack is out-of-band, the EDFA gains for transmitted signals, which vary as users come on and off, will not be significantly degraded on average. Therefore, the attack will not be detected.

According to [16], an out-of-band jammer can cause gain competition at an EDFA without causing degradation in total power. For example, consider a series of two EDFAs. An out-of-band jammer can cause gain competition at the first EDFA. Consequently, the data signals are attenuated and the output of the first EDFA has a lower SNR. Suppose that the second EDFA has "automatic gain control" (AGC), i.e. the second EDFA maintains its output power at a specified level. As a result, the output of the second EDFA will have the same power level (despite lower SNR) as if there were no attack.

As another example, an attacker can perform an out-of-band jamming attack causing gain competition at an EDFA together with an in-band jamming attack to compensate for the power loss due to gain competition [16].

An eavesdropper may insert in-band jamming signals after tapping data signals from a legitimate fiber and leave the power level unchanged. In this case, an attack will not be detected by power detection techniques. Finally, if power detection is performed only on the legitimate channel, eavesdropping from a crosstalk channel will not be detected.

1.5.2 Optical Spectral Analysis

Optical spectral analyzers (OSAs) measure average optical powers of signals at different frequencies. One obvious benefit of the use of an OSA with respect to power detection is the ability to detect a change in the spectrum shape of transmitted signals, even if the change in the spectrum shape does not alter the average power level over the channel bandwidth.

An OSA can be used to detect in-band jamming attacks which significantly alter the spectrum shape of transmitted signals even if these attacks yield the same average power as before.

Moreover, an OSA can be used to detect out-of-band jamming attacks provided that jamming signal frequencies are still in the range of an OSA. Using an OSA, we are able to detect some sporadic out-of-band jamming attacks which cause gain competition at EDFAs but do not change the average power and cannot be detected by power detection techniques.

An OSA can also be used to detect an eavesdropper who taps off a significant amount of signal power and causes the change in the spectrum shape. In this regard, an OSA does not provide much more information than power detection techniques.

Although we can detect a wider variety of attacks using an OSA, the additional information given by an OSA is still the values of statistical averages. As in the case of power detection, the use of an OSA to detect jamming attacks requires an observation of a large number of corrupted bits in order to distinguish between the presence and the absence of an attack. Therefore, an OSA may not detect in-band sporadic jamming attacks for the same reason as in the case of power detection.

Furthermore, an OSA may not detect sporadic out-of-band jamming attacks when the jamming signal frequencies are outside the OSA spectrum range. Finally, if an OSA is used to detect eavesdropping only on the legitimate channel, eavesdropping from a crosstalk channel will not be detected.

1.5.3 Pilot Tones

Pilot tones are signals which are transmitted along the same paths as the data signals but are distinguishable from the data signals. They are used to detect any transmission disruption that may occur to the data signals.

In general, pilot tones will not detect a jamming attack unless the pilot tones themselves are affected by such an attack. An attack which affects only the transmitted signals may not be detected. For example, if pilot tones are outside the transmission band, an in-band jamming attack will not be detected.

For subcarrier multiplexed (SCM) signals, which are the signals obtained from combining data signals and pilot tones in the same frequency band, an in-band jamming attack may affect the pilot tones and may therefore be detected. However, pilot tones are generally less sensitive to corruption than data signals.

If pilot tones are amplified by the same amplifiers as transmitted signals, an out-of-band jamming attack causing gain competition at EDFAs will attenuate pilot tones and may be detected. However, with AGC in subsequent EDFAs, the power levels of pilot tones may automatically be restored. In this case, an attack will not be detected.

On the other hand, if pilot tones and the transmitted signals are separately amplified, an out-of-band jamming attack which attenuates data signals through gain competition will not be detected by the use of pilot tones.

Note that pilot tones are subject to jamming themselves. An attacker may attempt to modify pilot tones directly after degrading data signals. By doing so, the attacker may restore the power levels of pilot tones to their normal levels and the corresponding attack will not be detected. For example, an attacker can perform out-of-band jamming causing gain competition which attenuates the power levels of pilot tones as well as data signals. At the same time, the attacker introduces in-band jamming signals at the frequencies of pilot tones to compensate for power losses. This combination of attacks will not be detected.

Futhermore, since pilot tones are detected only where we expect to receive the transmitted signals, they do not provide any information about an eavesdropper who taps the desired signals from a crosstalk channel instead of a legitimate channel.

1.5.4 Optical Time Domain Reflectrometry

Optical time domain reflectrometers (OTDRs) analyze echoes of the probe signals at the transmitter end instead of at the receiver end. One unique characteristic of the use of OTDRs which differentiates it from the use of pilot tones is the ability to perform attack detection solely at the transmission side.

OTDRs are mainly used to detect faults, bends and losses in fiber links. By modulating OTDR probe signals, the performance of OTDRs in fault detection can be enhanced [25]. If jamming signals share the same frequencies as OTDR probe signals, the corresponding attacks may be detected. On the other hand, we cannot detect an in-band jamming attack if OTDR probe signals are out of the transmission band.

The use of OTDRs to detect failures across multiple links requires bi-directional EDFAs [13]; otherwise, we cannot detect the reflected OTDR probe signals. In general, we may not have bi-directional EDFAs; and OTDRs are used to detect malfunction on individual sections of fiber links. As in the case of pilot tones, if OTDR probe signals share the same EDFAs as transmitted signals, out-of-band jamming attacks causing gain competition at EDFAs may be detected. However, like pilot tones, OTDR probe signals are generally less sensitive to corruption than data signals. On the other hand, if OTDR probe signals and transmitted signals are separately amplified, out-of-band jamming attacks causing gain competition at EDFAs will not be detected.

OTDR probe signals are subject to jamming themselves. An attacker can perform out-of-band jamming causing gain competition at an EDFA, together with in-band jamming at the frequencies of OTDR probe signals to compensate for power losses. This combination of attacks will not be detected.

Finally, as in the case of pilot tones, if we detect OTDR probe signals only where we expect to receive transmitted signals, eavesdropping from a crosstalk channel will not be detected.

1.5.5 Bit Error Rate Testing

A bit error rate tester (BERT) can be used to monitor the quality of the transmission path of interest in terms of the number of bit errors per time period. Based on the fact that both in-band and out-of-band jamming attacks can degrade transmitted signals and thus make it more likely for the receiver to make decision errors, an unusually high BER is an indicator of the possible presence of an attack on the network. In this regard, a BERT can also be used to detect jamming attacks based on degradation of the BER.

A BERT can also be used to detect eavesdropping provided that the eavesdropper taps off a significant amount of power from transmitted signals and the associated BER deteriorates significantly.

As in power detection and optical spectral analysis, the operation of a BERT is based on computing a statistical average which can consume too much time (see the discussion on why we want to detect an attack quickly in section 1.4.3). Specifically, the time required for a BERT to obtain a good estimate on the BER of a transmission path is in the order of several seconds for a transmission at 1 gigabit per second and the BER in the range of $\approx 10^{-10}$ to 10^{-8} (provided that an attack lasts for several seconds), and in the order of hundreds of seconds for the BER in the range of $\approx 10^{-12}$ to 10^{-10} (provided that an attack lasts for several hundreds of seconds).

Another drawback of the use of BERTs is the need to transmit test patterns, which take up transmission resources. Moreover, the use of a BERT requires information on the modulation scheme. Thus this technique is not appropriate for a transparent network. Finally, BERTs are very expensive.

We shall refer to BERTs again in chapter 2 when we compare the time required for our proposed attack detection scheme to detect some jamming attacks to the time required for a BERT to perform the same task.

1.6 Summary and Outlines of Remaining Chapters

In this chapter, we argued why existing supervisory mechanisms cannot be used to detect all types of jamming attacks in an AON. The insufficiency of existing supervisory mechanisms in detecting jamming attacks serves as the motivation to the construction and the analysis of a novel jamming attack detection scheme which is the subject of the remaining chapters.

In chapter 2, we introduce the attack detection device at a network node as proposed in [16]. Based on this device, we construct a jamming attack detection system. In analyzing its performance, we assume some rather pessimistic jamming attack scenarios. In those scenarios, we consider a light path consisting of 1 and 10 network nodes and OOK transmission at a rate of 1 gigabit per second using 1 gigahertz bandwidth. In addition, coherent components of jamming signals at all network nodes are assumed to be equal and constant throughout an observation period.

We consider separately the cases of in-band jamming and out-of-band jamming. For in-band jamming, we assume that the transmitted signals are always degraded in such a way that the receiver is more likely to make a decision error, i.e. the jamming signals increase the magnitude of the transmitted signal when the OFF signal of OOK is sent and vice versa for the ON signal.

We assume that an out-of-band jamming attack causes gain competition at gain-clamped EDFAs. In particular, we assume an equal percentage of gain attenuation of data signals at all network nodes. With gain attenuation at EDFAs, transmitted signals associated with the OFF level of OOK are not affected while those associated with the ON level are. An important factor is the dependence of the noise variance on the EDFA gain. Therefore, we base our analysis on the simulation data obtained from [1].

The decision rules for the attack detection device at each network node and at the receiver are based on direct detection, which is implemented in several existing AON testbeds. Our goal in chapter 2 is to obtain preliminary results on the performance of the proposed attack detection scheme. Therefore, we adopt direct detection in the analysis. When we investigate the performance of our approach for detecting jamming attacks in general, we shall remove the direct detection constraint and consider coherent detection since it yields a lower BER than direct detection for a given SNR, and since coherent detection can be done using direct detection photodetectors. [19] discusses how we can measure, using square-law detection, the input to a network node and its delayed version as well as the output of a network node and its delayed version in order to find both the magnitude difference and the phase difference between the input and the output of a network node.

In particular, our proposed attack detection scheme employs two levels of alarms. The first-level alarms are generated by attack detection devices at network nodes. The second-level (higher-level) alarm is generated on top of the first-level alarm in each observation period. Decision of the second-level alarm generation is based on the number of alarms generated by attack detection devices in each observation period.

An important parameter is the observation time required for our proposed attack detection scheme to detect an attack with high reliability, i.e. the associated probabilities of decision errors are sufficiently low. We measure the performance of the attack detection scheme in terms of two quantities: the false positive and the false negative probabilities of a second-level alarm in the corresponding observation period. Finally, we compare the observation periods in our attack detection scheme with those of BERTs.

In chapter 3, we assume coherent detection at attack detection devices at network nodes as well as at the receiver. The main goal of this chapter is to find the worst case in-band jamming attack scenario among all scenarios associated with the same average degraded BER. In particular, the worst case in-band jamming attack scenario yields the lowest expected number of first-level alarms in a given observation period. Consequently, it is the attack scenario with the lowest probability of being detected. We want to consider the worst case scenario since we want to be able to provide a guarantee on BER regardless of the policy or the sophistication of an attack.

We choose to describe an attack scenario by specifying coherent components of jamming signals at all network nodes and at each bit time. We formulate the problem of searching for the worst case attack scenario as an optimization problem whose variables are coherent components of jamming signals at all network nodes at each bit time and whose objective is to minimize the expected number of first-level alarms in an observation period.

After deriving the expressions for the average BER and the expected device alarm (first-level alarm) rate functions, we solve the minimization problem for a jamming attack at one network node and at two transmitted bits. We then extend the result to the case of jamming at one network node and at more than two bits.

In parallel, we solve the minimization problem for an attack at one transmitted bit and at two network nodes. We then extend the result to the case of jamming at one transmitted bit and at more than two network nodes.

Finally, we combine the results to establish the worst case in-band jamming attack scenario for jamming at multiple network nodes and at multiple transmitted bits.

In chapter 4, we find out how our proposed attack detection scheme performs under the worst case in-band jamming attack scenarios that we found in chapter 3. As in chapter 2, we measure the performance of the attack detection scheme in terms of the false positive and the false negative probabilities of a second-level alarm in the corresponding observation period. Finally, we suggest the use of hard limiters at the input to network nodes in order to improve the performance of our proposed attack detection scheme. Chapter 5 provides a summary of this work together with directions for future research.

Chapter 2

Novel Jamming Attack Detection Scheme

In this chapter, we construct a jamming attack detection scheme based on the detection device proposed in [19]. The chapter begins with basic information on the attack detection device and related analytical groundwork. Throughout the chapter, we assume the decision rules based on direct detection at the receiver as well as at attack detection devices at all network nodes.

We then introduce a novel attack detection scheme which is a decision system based on the number of alarms generated by attack detection devices at network nodes in an observation period.

To analyze the performance of our attack detection system, we assume some pessimistic in-band jamming attack scenarios. Rather than assuming that jamming signals at different network nodes are additive white Gaussian random variables, we assume that they are coherent and add up constructively along the light path. Moreover, we assume that jamming signals are equally spreaded out at all attacked nodes so that no overt attack is present at any network node.

The problem of determining in-band jamming attack scenarios which are the hardest to detect is relegated to chapter 3. The results obtained in this chapter will indicate that the proposed detection scheme can at least handle some in-band jamming attack scenarios. The performance analysis in the worst case scenarios (which are the hardest to detect) will be postponed to chapter 4.

In the final section, we investigate the detection of out-of-band jamming attacks causing gain competition at EDFAs. The analysis is based on the data obtained from [1] and on some pessimistic out-of-band jamming attack scenarios. The subject of the worst case out-of-band jamming attack scenarios (which are the hardest to detect) is extremely device specific and is outside the scope of this research.

2.1 Attack Detection Device at a Network Node

Our jamming attack detection system is constructed from attack detection devices which wrap around certain network nodes. In general, a network node can consist of switching devices such as ADMs and WSSs, or optical amplifiers such as EDFAs, or a combination of different devices. Each attack detection device contains two taps, one at the input and the other at the output of a network node. Figure 2-1 shows the schematic diagram of the device proposed in [19].

Denote the signal at the first tap by $s(t)$, which after passing through a device delay T_D, becomes $s(t - T_D)$. Denote the sum of the signal and noise at the second tap by $s'(t-T_D)+n(t-T_D)$, or $\gamma s'(t-T_D)+n(t-T_D)$ for a node with an amplifier. Note that we describe the signal component at the second tap by $s'(t - T_D)$ instead of $s(t - T_D)$ to take into account a possible phase shift and polarization due to normal operation of the network node.

Our device takes these two signals and compensates for the phase shift and polarization due to normal operation of the network node, i.e. the device makes the signals from the two taps coherent with each other. The device then performs a signal subtraction and a square-law magnitude detection of the result. In general, the change in phase and polarization of $s(t)$ does not vary rapidly; and we can safely assume, in the absence of any attack, the output of the square law detector is $|n(t - T_D)|^2$, or $|\frac{1}{\gamma}n(t - T_D)|^2$ for a node with an amplifier. For notational simplicity, we shall drop the time notation in what follows.

2.1.1 In-Band Jamming Attack

If there is an in-band jamming signal denoted by J, the output of the square law detector will be $|n + J|^2$ (or $|\frac{1}{\gamma}(n + J)|^2$) , which is likely to be much greater than $|n|^2$ (or $|\frac{1}{\gamma}n|^2$) for an attack with a sufficiently large in-band jamming signal. Based on the output of the

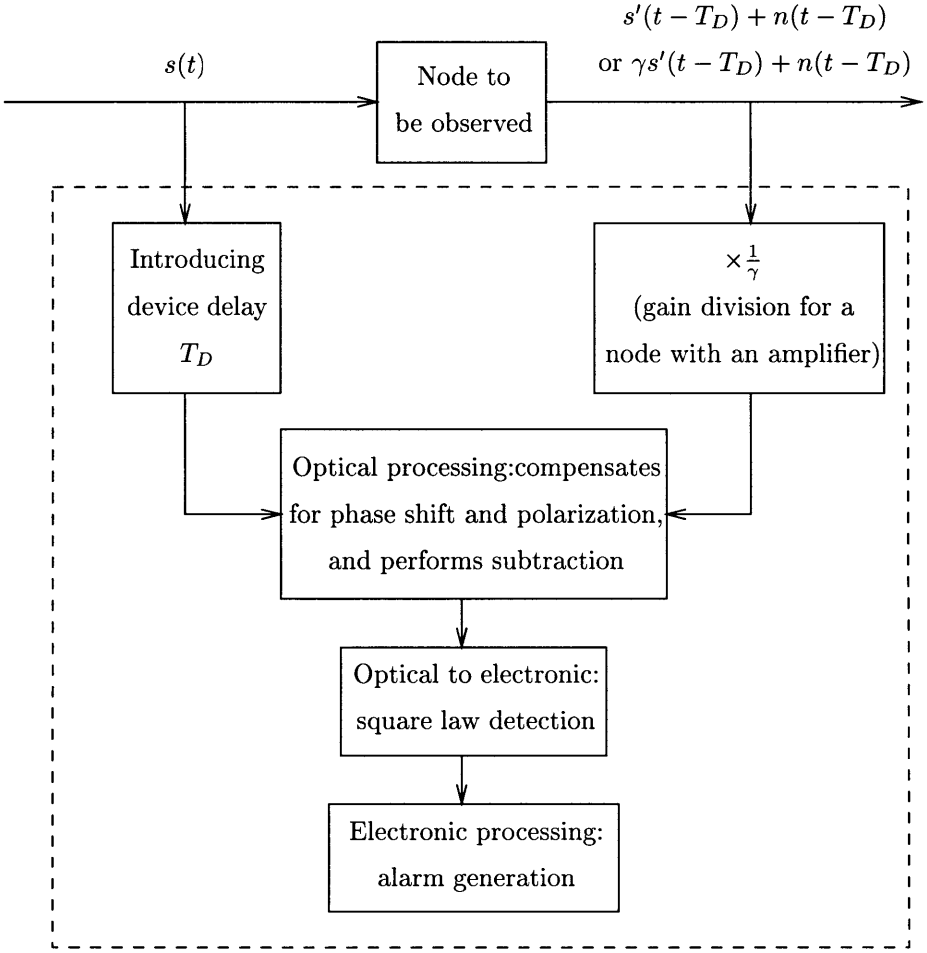

$s(t)$
Node to be observed
$s'(t - T_D) + n(t - T_D)$
or $\gamma s'(t - T_D) + n(t - T_D)$
Introducing device delay T_D
$\times\frac{1}{\gamma}$ (gain division for a node with an amplifier)
Optical processing:compensates for phase shift and polarization, and performs subtraction
Optical to electronic: square law detection
Electronic processing: alarm generation

Figure 2-1: Attack detection device around a node.

square law detector, we set a decision threshold for an alarm generation. As we shall see later on, the value of such a threshold is a design parameter for our attack detection system.

2.1.2 Out-of-Band Jamming Attack

In analyzing gain competition, our analysis assumes that each EDFA at a network node makes up precisely for signal attenuation in the previous link of the light path. Consider an out-of-band jamming attack causing gain competition at an EDFA in a network node. An out-of-band jamming attack which robs the EDFA gain for the legitimate signal by $a\%$ will cause the output of the square law detector to be $\left|\frac{a}{100}s + \frac{n}{\gamma}\right|^2$.

Since the quantities involved in the analysis of the device behaviours are in terms of the probability distribution of the square law detector output, it is necessary to find this distribution before we proceed.

2.2 Probability Distribution of the Output of the Square Law Detector

In what follows, uppercase characters denote random variables while lowercase characters denote deterministic values realized by the random variables.

In general, the output of a square law detector at a given time is of the form $Y = |S + N|^2$, where S and N denote the signal and noise respectively. At the receiver, S refers to the received data signal. At an attack detection device, S may refer to an in-band jamming signal applied at the network node, or the attenuation of the data signal due to gain competition at an EDFA.

Conditioned on $S = s$, we want to find the cumulative probability distribution (c.d.f.) of $Y = |s + N|^2$, i.e. $Pr\{Y \leq A\}$ for a real $A > 0$. [16] suggests the method outlined in this section.

Note that $|s + N|^2 = |s_R + N_R|^2 + |s_I + N_I|^2$, where subscripts R and I denote the real and imaginary parts respectively. We assume N_R and N_I to be two independent Gaussian random variables with the same variance denoted by σ_N^2. We can then consider Y as the sum of two random variables $Y_R = |s_R + N_R|^2$ and $Y_I = |s_I + N_I|^2$.

Let us denote the characteristic function of Y_R by Φ_{Y_R}. We may write that

$$\Phi_{Y_R}(\omega) = \frac{1}{\sigma_N\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{j\omega(n+s_R)^2} e^{-\frac{n^2}{2\sigma_N^2}}\, dn. \tag{2.1}$$

By performing a change of variable $y = (n + s_R)^2$, we can write equation 2.1 as

$$\Phi_{Y_R}(\omega) = \frac{1}{2\sigma_N\sqrt{2\pi}} \int_0^{\infty} e^{j\omega y} \left(e^{-\frac{(\sqrt{y}-s_R)^2}{2\sigma_N^2}} + e^{-\frac{(\sqrt{y}+s_R)^2}{2\sigma_N^2}} \right) \frac{dy}{\sqrt{y}}. \tag{2.2}$$

From equation 2.2, the probability distribution function (p.d.f.) of Y_R is

$$p_{Y_R}(y) = \begin{cases} \frac{1}{2\sigma_N\sqrt{2\pi}} \left(e^{-\frac{(\sqrt{y}-s_R)^2}{2\sigma_N^2}} + e^{-\frac{(\sqrt{y}+s_R)^2}{2\sigma_N^2}} \right) \frac{1}{\sqrt{y}}, & \text{if } y \geq 0 \\ 0 & \text{otherwise} \end{cases}. \tag{2.3}$$

Similarly, the p.d.f. of Y_I is

$$p_{Y_I}(y) \quad = \quad \begin{cases} \frac{1}{2\sigma_N\sqrt{2\pi}} \left(e^{-\frac{(\sqrt{y}-s_I)^2}{2\sigma_N^2}} + e^{-\frac{(\sqrt{y}+s_I)^2}{2\sigma_N^2}} \right) \frac{1}{\sqrt{y}}, & \text{if } y \geq 0 \\ 0 & \text{otherwise} \end{cases} . \tag{2.4}$$

So we can express the c.d.f. $Pr\{Y \leq A\}$ as

$$\begin{aligned} Pr\{Y \leq A\} \quad = \quad & \frac{1}{2\pi\sigma_N^2} \int_0^A \int_0^{A-y_R} \left(e^{\frac{-(\sqrt{y_R}-s_R)^2}{2\sigma_N^2}} + e^{\frac{-(\sqrt{y_R}+s_R)^2}{2\sigma_N^2}} \right) \frac{1}{2\sqrt{y_R}} \\ & \left(e^{\frac{-(\sqrt{y_I}-s_I)^2}{2\sigma_N^2}} + e^{\frac{-(\sqrt{y_I}+s_I)^2}{2\sigma_N^2}} \right) \frac{1}{2\sqrt{y_I}} dy_I dy_R. \end{aligned} \tag{2.5}$$

Making the change of variables $z_R = \sqrt{y_R}$ and $z_I = \sqrt{y_I}$ in equation 2.5, we have

$$\begin{aligned} Pr\{Y \leq A\} \quad = \quad & \frac{1}{2\pi\sigma_N^2} \int_0^{\sqrt{A}} \int_0^{\sqrt{A-z_R^2}} \left(e^{\frac{-(z_R-s_R)^2}{2\sigma_N^2}} + e^{\frac{-(z_R+s_R)^2}{2\sigma_N^2}} \right) \\ & \left(e^{\frac{-(z_I-s_I)^2}{2\sigma_N^2}} + e^{\frac{-(z_I+s_I)^2}{2\sigma_N^2}} \right) dz_I dz_R. \end{aligned} \tag{2.6}$$

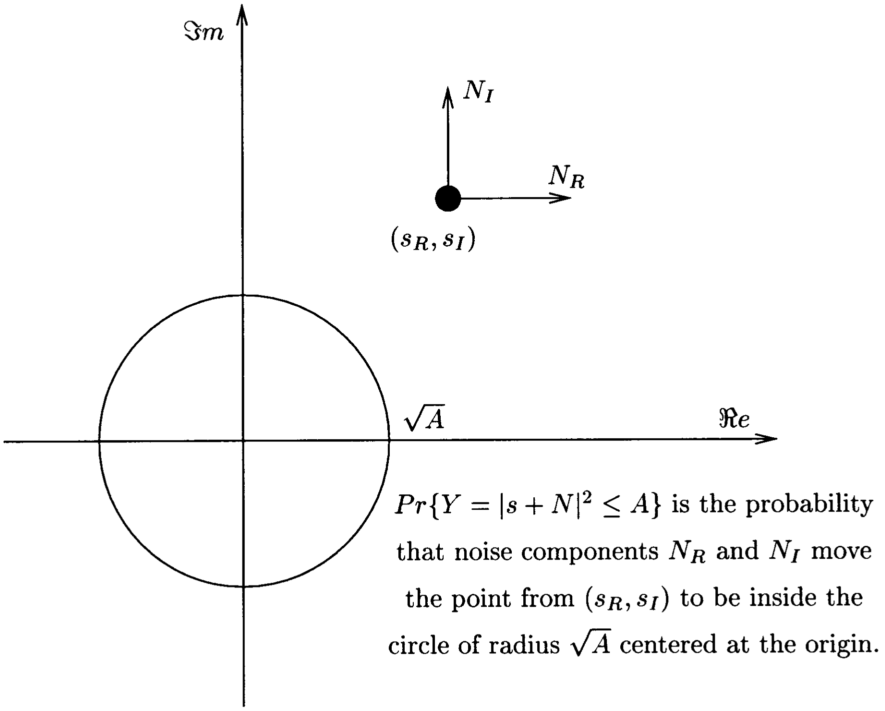
Note that we may simplify the expression on the right hand side of equation 2.6 by considering the integral as the sum of four terms. Using symmetry arguments, or by performing a change of variable, we can see that

$$\begin{aligned} & \frac{1}{2\pi\sigma_N^2} \int_0^{\sqrt{A}} \int_0^{\sqrt{A-z_R^2}} e^{\frac{-(z_R-s_R)^2}{2\sigma_N^2}} e^{\frac{-(z_I+s_I)^2}{2\sigma_N^2}} dz_I dz_R = \\ & \frac{1}{2\pi\sigma_N^2} \int_0^{\sqrt{A}} \int_{-\sqrt{A-z_R^2}}^{0} e^{\frac{-(z_R-s_R)^2}{2\sigma_N^2}} e^{\frac{-(z_I-s_I)^2}{2\sigma_N^2}} dz_I dz_R. \end{aligned} \tag{2.7}$$

We can write all terms on the right hand side of equation 2.6 such that the exponents are functions of $(z_R - s_R)$ and $(z_I - s_I)$ using the method presented in equation 2.7 if necessary. As a result, our expression for $Pr\{Y \leq A\}$ becomes

$$Pr\{Y \leq A\} = \int_{-\sqrt{A}}^{\sqrt{A}} \int_{-\sqrt{A-z_R^2}}^{\sqrt{A-z_R^2}} \frac{1}{2\pi\sigma_N^2} e^{\left(-\frac{(z_R-s_R)^2}{2\sigma_N^2} - \frac{(z_I-s_I)^2}{2\sigma_N^2} \right)} dz_I dz_R. \tag{2.8}$$

Transforming the region of integration from a circular region to a rectangular region



$Pr\{Y = |s+N|^2 \leq A\}$ is the probability that noise components N_R and N_I move the point from (s_R, s_I) to be inside the circle of radius $\sqrt{A}$ centered at the origin.

Figure 2-2: Geometric view of $Pr\{Y \leq A\}$.

helps simplify the method of numerical integration. Using polar coordinates, we can express $Pr\{Y \leq A\}$ in equation 2.8 as

$$Pr\{Y \leq A\} = \int_0^{2\pi} \int_0^{\sqrt{A}} \frac{1}{2\pi\sigma_N^2} r e^{\left(-\frac{(r\cos\theta - s_R)^2}{2\sigma_N^2} - \frac{(r\sin\theta - s_I)^2}{2\sigma_N^2}\right)} dr d\theta. \tag{2.9}$$

Denote this c.d.f. as a function of σ_N^2, A, s_R and s_I by $F_d(\sigma_N^2, A, s_R, s_I)$ as shown below

$$F_d(\sigma_N^2, A, s_R, s_I) = \int_0^{2\pi} \int_0^{\sqrt{A}} \frac{1}{2\pi\sigma_N^2} r e^{\left(-\frac{(r\cos\theta - s_R)^2}{2\sigma_N^2} - \frac{(r\sin\theta - s_I)^2}{2\sigma_N^2}\right)} dr d\theta. \tag{2.10}$$

Figure 2-2 shows a geometric interpretation of $Pr \leq A$. An interesting fact is the dependency of $F_d(\sigma_N^2, A, s_R, s_I)$ on s only through its norm $||s||$. Lemma 2.1 verifies this fact. Using the result of lemma 2.1, we can reexpress $F_d(\sigma_N^2, A, s_R, s_I)$ as

$$F_d(\sigma_N^2, A, ||s||) = \int_0^{2\pi} \int_0^{\sqrt{A}} \frac{1}{2\pi\sigma_N^2} r e^{\left(-\frac{(r\cos\theta - ||s||)^2}{2\sigma_N^2} - \frac{(r\sin\theta)^2}{2\sigma_N^2}\right)} dr d\theta. \tag{2.11}$$

For future reference, we shall denote the c.d.f. $Pr\{Y \leq A\}$ as a function of σ^2, A, and $||s||$ by $F_d(\sigma^2, A, ||s||)$. We finish this section by presenting lemma 2.1.

Lemma 2.1 *Let $Y = |s + N|^2$ where s is a deterministic complex signal and N is a random Gaussian complex signal whose real and imaginary parts are independent zero-mean Gaussian random variables with variance σ_N^2. For $A > 0$, $F_d(\sigma_N^2, A, s_R, s_I)$ as defined in equation 2.10 depends on s only through its norm $||s||$.*

Proof Consider the noise component pointing away from the origin along the θ-axis as shown in figure 2-3. We can express this noise component as the sum of two independent Gaussian random variables

$$N_C = N_R \cos\theta + N_I \sin\theta,$$

where θ is equal to $\arctan(s_I/s_R)$. By statistical independence, the variance of N_C is the sum of the two variances and is equal to

$$\begin{aligned} \mathrm{Var}(N_C) &= \mathrm{Var}(N_R)\cos^2\theta + \mathrm{Var}(N_I)\sin^2\theta \\ &= \sigma_N^2(\cos^2\theta + \sin^2\theta) = \sigma_N^2. \end{aligned}$$

In the similar fashion, consider the other noise component (denoted by N_T) orthogonal to N_C as shown in figure 2-3. We can express N_T together with its variance as.

$$\begin{aligned} N_T &= -N_R \sin\theta + N_I \cos\theta \\ \mathrm{Var}(N_T) &= \mathrm{Var}(N_R)\sin^2\theta + \mathrm{Var}(N_I)\cos^2\theta \\ &= \sigma_N^2(\sin^2\theta + \cos^2\theta) = \sigma_N^2. \end{aligned}$$

We now verify that N_C and N_T are independent Gaussian random variables. Let $E[X]$ denote the expected value of a random variable X. Consider their covariance given below,

$$\begin{aligned} E[N_C N_T] &= E\left[(N_R\cos\theta + N_I\sin\theta)(-N_R\sin\theta + N_I\cos\theta)\right] \\ &= E\left[-N_R^2\cos\theta\sin\theta + N_R N_I\cos^2\theta - N_R N_I\sin^2\theta + N_I^2\sin\theta\cos\theta\right] \\ &= -E[N_R^2]\cos\theta\sin\theta + E[N_R N_I]\cos^2\theta - E[N_R N_I]\sin^2\theta + E[N_I^2]\sin\theta\cos\theta \\ &= -E[N_R^2]\cos\theta\sin\theta + E[N_I^2]\cos\theta\sin\theta \qquad (2.12) \end{aligned}$$

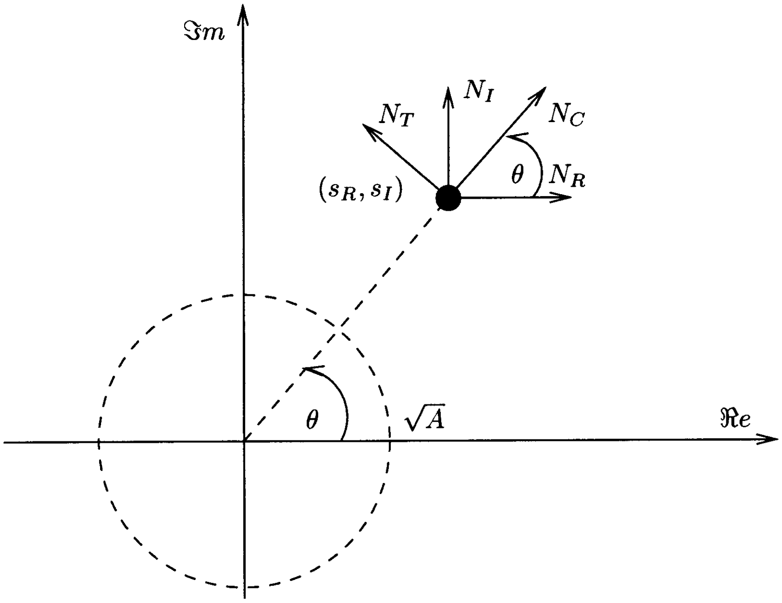


Figure 2-3: Illustration of the fact that $F_d(\sigma_N^2, A, s_R, s_I)$ only depends on s through its norm for the proof of lemma 2.1.

$$= -\sigma_N^2 \cos\theta \sin\theta + \sigma_N^2 \cos\theta \sin\theta$$
$$= 0,$$

where equality 2.12 follows from the fact that N_R and N_I are independent and thus uncorrelated. Note that N_C and N_T are jointly Gaussian since their linear combination is a linear combination of N_R and N_I which is a Gaussian random variable. Since N_C and N_T are jointly Gaussian and uncorrelated, it follows that they are independent.

By rotating the axes counterclockwise by an angle θ, we can treat the calculation of $F_d(\sigma_N^2, A, s_R, s_I)$ as if s were located at $(||s||, 0)$ and there were two independent noise components each with variance σ_N^2 along the horizontal and the vertical axes. Thus $F_d(\sigma_N^2, A, s_R, s_I)$ depends on s only through its norm $||s||$ and we can reexpress $F_d(\sigma_N^2, A, s_R, s_I)$ in equation 2.10 as $F_d(\sigma_N^2, A, ||s||)$ in equation 2.11. □

2.3 Scenario for the Analysis of Jamming Attack Detection

We would like to find out how the detection device behaves and construct a scheme for detecting jamming attacks. In this section, we set up a scenario to base our analysis on.

2.3.1 Network Topology

Assume the transmission of data across M successive network nodes. We shall consider the cases when M is equal to 1 and 10 respectively. For the analysis in this chapter, we assume that all M network nodes are similarly affected by a jamming attack, and that the corresponding effects are constant throughout an observation period.

2.3.2 Noise and SNR

At each network node at any given bit time, we have two independent additive white Gaussian noise components denoted by N_R and N_I. Assume the noise variance $\sigma_{N_R}^2$ and $\sigma_{N_I}^2$ to be equal (denoted by $\frac{1}{M}\sigma_N^2$ so that the total noise variance across the light path at any given bit time is $2\sigma_N^2$). To obtain later numerical results, we assume σ_N^2 to be 0.5.

Assume an end-to-end SNR of 16 dB (approximated from $10\log_{10}40$). In addition, assume that the SNR degradation at network nodes is much more significant than the SNR degradation along the fiber. Furthermore, assume that all network nodes cause the same level of SNR degradation. Based on these assumptions, when $M = 10$, we have an SNR of 16 dB across the light path and an SNR of 26 dB across each network node.

2.3.3 Transmission and Detection Scheme

Assume that we transmit data bits whose values are equally likely to be 0 or 1 using OOK. Assume the transmission rate of 1 Gb/s and the bandwidth of 1 GHz. In addition, we assume direct detection at the receiver.

Denote the square magnitude of the ON level by P, the following equation relates P to a specific value of SNR.

$$SNR = \frac{1}{2}\left(\frac{P}{2\sigma_N^2}\right) + \frac{1}{2}\left(\frac{0}{2\sigma_N^2}\right). \tag{2.13}$$

For 16 dB SNR, the ON level (square magnitude) is $160\sigma_N^2$ since $40 = \frac{1}{2}\left(\frac{160\sigma_N^2}{2\sigma_N^2}\right) + \frac{1}{2}(0)$. This ON level can be achieved by different values of s_R and s_I. For example, (s_R, s_I) can be

$(\sqrt{80}\sigma_N, \sqrt{80}\sigma_N)$ or $(\sqrt{160}\sigma_N, 0)$. For a given decision threshold, it is known that different values of (s_R, s_I), given the constraint that $s_R^2 + s_I^2$ (equivalently $||s||^2$) is constant, lead to the same value of probability of decision error (equivalently the BER). Nevertheless, we provide a proof in proposition 2.1. For the analysis in this chapter, we set (s_R, s_I) to be $(\sqrt{160}\sigma_N$,0).

Proposition 2.1 *Assume that the ON and the OFF signals of OOK are equally likely, and the receiver's decision rule is based on direct detection. Given a decision threshold, all the ON level signals having the same norm result in the same BER.*

Proof Consider the general expression for the BER given below

$$BER = \frac{1}{2}P_{e|\text{OFF}} + \frac{1}{2}P_{e|\text{ON}}, \tag{2.14}$$

where $P_{e|\text{OFF}}$ denotes the probability of error given the OFF signal is sent and vice versa.

Let A denote the decision threshold, we can express $P_{e|\text{OFF}}$ and $P_{e|\text{ON}}$ as

$$\begin{aligned} P_{e|\text{OFF}} &= 1 - F_d(\sigma_N^2, A, 0), & (2.15) \\ P_{e|\text{ON}} &= F_d(\sigma_N^2, A, ||s||). & (2.16) \end{aligned}$$

It is clear from equation 2.15 that $P_{e|\text{OFF}}$ does not depend on (s_R, s_I). Equation 2.16 tells us that $P_{e|\text{ON}}$ depends on (s_R, s_I) only through its norm $||s||$ (consequence of lemma 2.1). Using equations 2.14, 2.15 and 2.16, we conclude that the BER only depends on s through its norm. Thus all the ON level signals having the same norm result in the same BER. □

2.3.4 Calculation of the BER

In the absence of an attack, we can compute the end-to-end BER. For analytical purposes, we shall occasionally express relevant quantities in terms of the function F_d defined in equation 2.11 before carrying out numerical computation. Let A denote the optimal threshold for the decision rule at the receiver, it follows that

$$BER_{\text{no attack}} = \frac{1}{2}(1 - F_d(\sigma_N^2, A, 0)) + \frac{1}{2}F_d(\sigma_N^2, A, \sqrt{160}\sigma_N) \tag{2.17}$$

By trial and error, it turns out that the optimal decision threshold A is approximately 22 (or $44\sigma_N^2$). We then have

$$\begin{aligned} BER_{\text{no attack}} &= \frac{1}{2}(1 - F_d(0.5, 22, 0)) + \frac{1}{2}F_d(0.5, 22, \sqrt{80}) \\ &\approx 4.6 \times 10^{-10}. \end{aligned}$$

From here on, we shall refer to this value as BER_{baseline}.

2.3.5 Guaranteed BER

We would like to construct the detection scheme that generates an alarm whenever the end-to-end BER (denoted by $BER_{\text{end-to-end}}$) is greater than 10^{-8} (in comparison to BER_{baseline} of 4.6×10^{-10}). We shall compare the performance of our detection scheme to that of a BERT in terms of the required observation period.

2.4 Attack Detection Scheme: In-Band Jamming

The goal of this section and the next is to construct a detection scheme which generates an alarm whenever $BER_{\text{end-to-end}}$ is greater than 10^{-8}. In this section, we shall construct a detection scheme to handle in-band jamming attacks.

2.4.1 Degraded BER

We first investigate the effects of in-band jamming attacks. First we find the $BER_{\text{end-to-end}}$ given the degradation of $a\%$ of the ON level (square magnitude) at each network node, or equivalently a possible total degradation of $Ma\%$ across M nodes. Note that we take a rather pessimistic view. In general, it is very hard for an attacker to jam multiple network nodes coherently so that the effects of a jamming attack add up constructively across M network nodes.

In addition, we consider the worst case degradation, i.e. signals are always degraded in such a way that errors are more likely. In other words, the jamming signal decreases the output magnitude of the square law detector when the ON level signal is sent and vice versa. In general, it is very hard for an attacker to cause the worst case degradation since an attacker must assign jamming signals according to transmitted data bits. The following

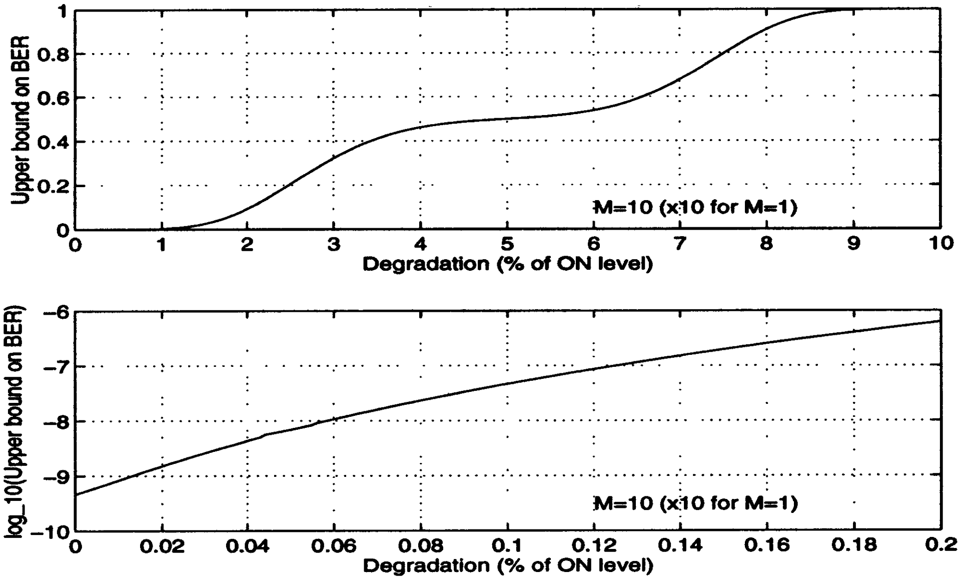


Figure 2-4: Upper bound on BER versus degradation at each of the M (equal to 10) network nodes in percentage of the ON level $(160\sigma_N^2)$. For $M = 1$, multiply the degradation level (horizontal axis) by 10. Note that the bottom plot is the zoomed version of the top one.

expression provides the upper bound on $BER_{\text{end-to-end}}$ with $a\%$ degradation allowed at each of the M network nodes

$$\begin{aligned}\overline{BER}_{end-to-end} &= \frac{1}{2}(1 - F_d(\sigma_N^2, A, \sqrt{(.01Ma)160}\sigma_N)) \\ &+\frac{1}{2}F_d(\sigma_N^2, A, \sqrt{(1-.01Ma)160}\sigma_N),\end{aligned} \tag{2.18}$$

Note that we use the notation $\overline{BER}_{\text{end-to-end}}$ to emphasize that the value is an upperbound on $BER_{\text{end-to-end}}$. Figure 2-4 shows the curve of $\overline{BER}_{end-to-end}$ versus the degradation at each of the M network nodes in percentage of the ON level $(160\sigma_N^2)$.

2.4.2 Probability of Alarm Generation by Attack Detection Devices

An attack detection device detects an attack when the output of a square-law detector exceeds a certain threshold. The corresponding decision regions are the same as the ones in the right hand side of figure 1-1. We shall refer to an alarm generated by an attack detection device at a network node as a "device alarm".

For a given threshold, there is a false positive probability (FP), the probability of detecting an attack when there is none. Moreover, there is a false negative probability (FN), the probability of not detecting an attack when there is one. Note that the value of FP depends solely on the threshold value while the value of FN depends on both the threshold value and the magnitude of a jamming signal.

It is necessary at this point to distinguish between FP and FN at a single network node and FP and FN across multiple network nodes when $M > 1$. We will refer to the former case with FP_{node} and FN_{node}, and to the latter case with $FP_{\text{end-to-end}}$ and $FN_{\text{end-to-end}}$ respectively.

In particular, suppose the threshold is set at $t\%$ of the ON level ($160\sigma_N^2$) and the magnitude of the jamming tone is $a\%$ of the ON level. The corresponding FP_{node} and FN_{node} are

$$\begin{aligned} FP_{\text{node}} &= 1 - F_d\left(\frac{1}{M}\sigma_N^2, (.01t)160\sigma_N^2, 0\right), && (2.19) \\ FN_{\text{node}} &= F_d\left(\frac{1}{M}\sigma_N^2, (.01t)160\sigma_N^2, \sqrt{(.01a)160}\sigma_N\right). && (2.20) \end{aligned}$$

Note that the noise variance is adjusted to $\frac{1}{M}\sigma_N^2$ since we are looking at each individual node with the SNR of $16+10\log_{10} M$ dB instead of the end-to-end SNR of 16 dB (see the discussion in section 2.3.2).

In the case of multiple network nodes ($M > 1$), we shall consider that a device alarm at any bit time has occured if at least one attack detection device at any network node generates a device alarm. This assumption leads to the computation of $FN_{\text{end-to-end}}$ and $FP_{\text{end-to-end}}$ given next.

We can think of $FN_{\text{end-to-end}}$ as the probability that none of the M network nodes generates a device alarm. Given a jamming signal, an event that an alarm is not generated depends on AWGN introduced at a network node. Since our model assumes that noise components introduced at different network nodes are independent, we have that the generation of device alarms at different network nodes are independent. Therefore, we assert that

$$FN_{\text{end-to-end}} = (FN_{\text{node}})^M . \qquad (2.21)$$

$FP_{\text{end-to-end}}$ is the probability that at least one of the M network nodes generates a device alarm. From the independence of device alarm generation at different network nodes, we assert that

$$FP_{\text{end-to-end}} = 1 - (1 - FP_{\text{node}})^M . \tag{2.22}$$

Figures 2-5 and 2-6 show example curves of $FP_{\text{end-to-end}}$ and $FN_{\text{end-to-end}}$ respectively.

2.4.3 Construction of an Attack Detection Scheme

We are now ready to construct a novel jamming attack detection scheme. Our detection scheme is based on the generation of alarms by attack detection devices at network nodes.

Keep in mind that the presence of a device alarm does not mean that a transmitted bit is corrupted. A device alarm simply notifies the user that the bit in transmission has a higher probability of decision error at the receiver. If an attack does not significantly corrupt transmitted bits and does not affect our communication more severely than by occurring natural noise, we do not worry about detecting such an attack.

Therefore, we do not want to conclude there is an attack unless the number of device alarms in an observation period (equivalently the device alarm rate which is the number of device alarms in an observation period divided by the number of bits in an observation period) exceeds a certain threshold.

Based on the given arguments, we propose two levels of alarms which will be referred to as a "device alarm" (first-level) and an "attack alarm" (second-level). We shall consider that a device alarm at any bit time has occured if at least one detection device at any network node generates a device alarm. An attack alarm will turn on when the number of device alarms in a given observation period (in bits) exceeds a certain threshold. Let FP_{attack} and FN_{attack} denote the FP and FN of the attack alarm respectively.

In what follows, we assume a constant jamming attack throughout an observation period. Later on, we shall determine the observation period required for the detection scheme to have sufficiently low $FP_{\text{end-to-end}}$ and $FN_{\text{end-to-end}}$.

Expected Device Alarm Rate Versus $\overline{BER}_{\text{end-to-end}}$ Curves

To find appropriate threshold values for the device alarm and the attack alarm, consider the curves of the expected device alarm rate (the expected number of device alarms in an observation period divided by the number of bits in an observation period) versus $\overline{BER}_{\text{end-to-end}}$. The expected device alarm rate is equal to $FP_{\text{end-to-end}}$ when no attack is present, and is

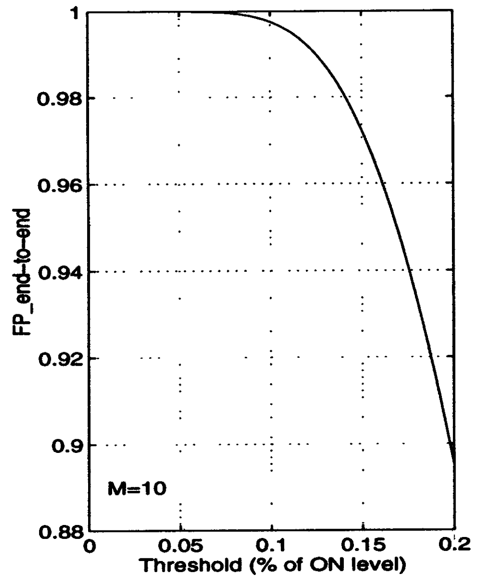


Figure 2-5: $FP_{\text{end-to-end}}$ versus detection device threshold (allowed degradation) at each network node.

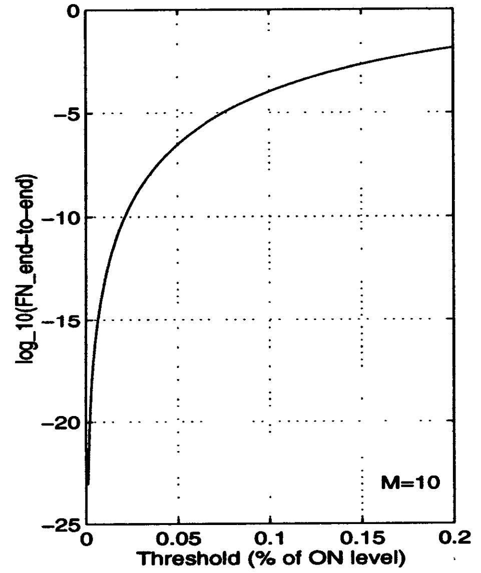


Figure 2-6: $FN_{\text{end-to-end}}$ versus detection device threshold (allowed degradation) at each network node given that the jamming signal magnitude is 0.06% of the ON level for M=10 (0.6% for M=1). With these jamming signal magnitudes, $BER_{\text{end-to-end}}$ can be as high as 10^{-8}.

equal to $1 - FN_{\text{end-to-end}}$ when there is an attack. Figures 2-7 and 2-8 show the curves of the expected device alarm rate versus $\overline{BER}_{\text{end-to-end}}$ for different values of detection device threshold.

For each curve in figures 2-7 and 2-8, the starting point on the left has the expected device alarm rate equal to $FP_{\text{end-to-end}}$ while the corresponding BER is BER_{baseline} (equal to 4.6×10^{-10}). This is the situation with no attack. The points along the curve in the positive direction correspond to in-band jamming attacks with higher and higher jamming signal magnitudes. We would like to have a significant difference between the expected device alarm rate with no attack and the rate with an attack causing $\overline{BER}_{\text{end-to-end}}$ to exceed 10^{-8}. In addition, we would like to be able to detect an attack in a small amount of time.

Finding the Threshold Values

We now have enough information to find an appropriate threshold value for the device alarm as a percentage of the ON level ($160\sigma_N^2$) and an appropriate threshold value for the attack alarm in terms of the number of device alarms in an observation period.

Note that our transmission rate is 1 Gb/s. Using the information in figures 2-7 and 2-8, we can calculate the estimated numbers of device alarms in an observation period. Tables 2.1 and 2.2 show the expected number of device alarms for different detection device thresholds and for different lengths of an observation period.

Since the occurrences of device alarms are discrete events, we cannot distinguish between two events both of which correspond to less than one alarm. Therefore, any combination of an attack detection device threshold and an observation period yielding an expected number of device alarms less than 1 in the presence of an attack is not useful.

From tables 2.1 and 2.2, we can see that a higher threshold yields a lower expected number of device alarms in an observation period but a higher ratio of the expected number of device alarms associated with an attack to the one associated with no attack.

Since we would like to have a high expected number of device alarms when an attack is present so that we can detect it quickly (using a short observation period), a high detection device threshold is not favourable in this viewpoint. At the same time, we would like to have a significant difference in the expected number of alarms with and without an attack. In this viewpoint, a high detection device threshold is favourable.

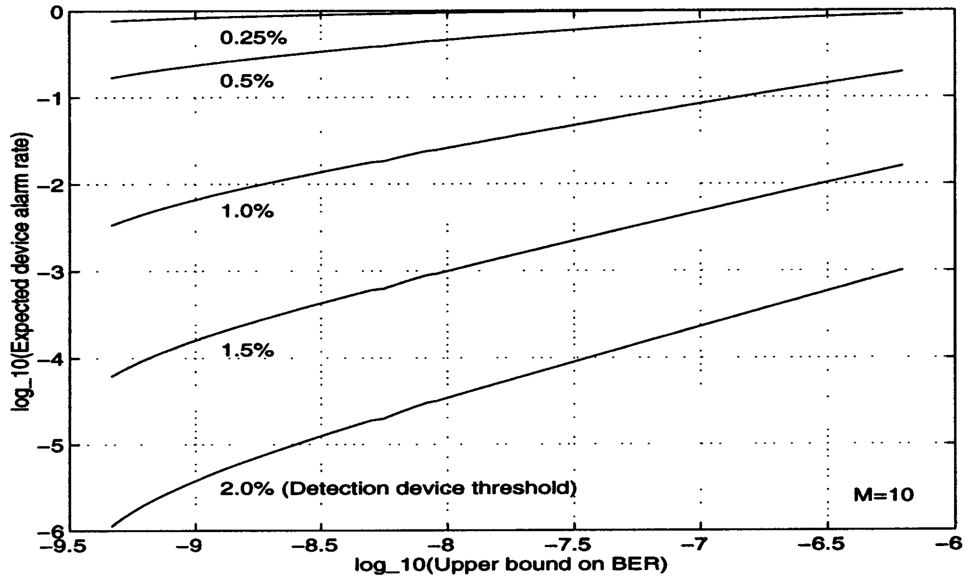


Figure 2-7: Expected device alarm rate versus $\overline{BER}_{\text{end-to-end}}$ for different detection device thresholds (M=10, in-band jamming).

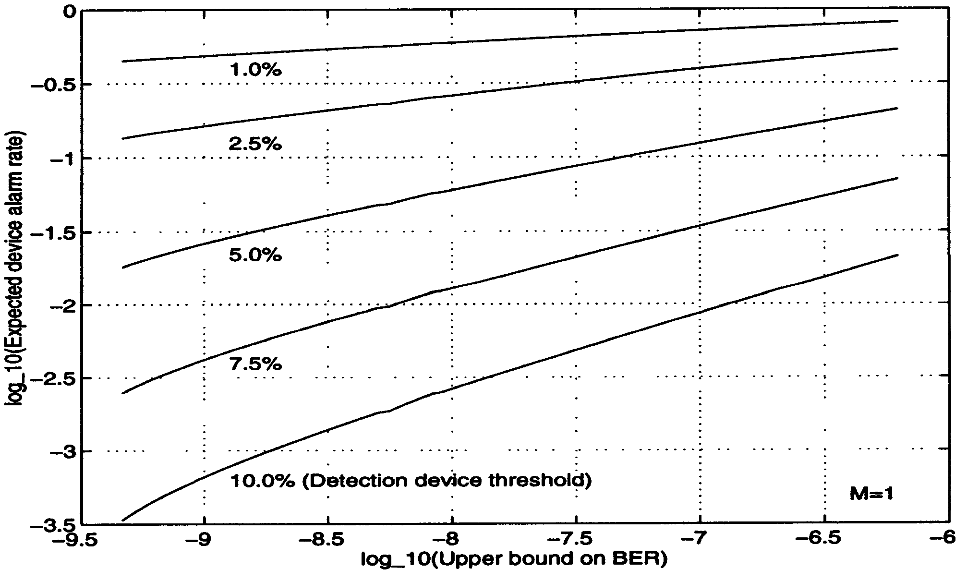


Figure 2-8: Expected device alarm rate versus $\overline{BER}_{\text{end-to-end}}$ for different detection device thresholds (M=1, in-band jamming).

	0.25%	0.5%	1.0%	1.5%	2.0%
10 μs	7,700/9,500	1,700/4,600	34/260	0.7/10	0.01/0.3
1 μs	770/950	170/460	3.4/26	0.07/1	0.001/0.03
0.1 μs	77/95	17/46	0.34/2.6	0.007/0.1	0.0001/0.003

Table 2.1: Expected number of device alarms in an observation period. The top row contains the detection device threshold values in percentage of the ON level ($160\sigma_N^2$). The first column contains the length of an observation period. The first number of each entry in the table is the expected number of device alarms when there is no attack, while the second number is the expected number of device alarms when there is an attack causing $\overline{BER}_{\text{end-to-end}}$ to be approximately 10^{-8}. In this case, $M = 10$.

	1.0%	2.5%	5.0%	7.5%	10.0%
10 μs	4,500/6,000	1,400/2,600	270/600	25/130	6.5/26
1 μs	450/600	140/260	27/60	2.5/13	0.65/2.6
0.1 μs	45/60	14/26	2.7/6	0.25/1.3	0.065/0.26

Table 2.2: Expected number of device alarms in an observation period ($M = 1$). All the entries have the same meanings as in table 2.1.

In general, if the detection device threshold is set too high, we need a long observation period to wait for a sufficient amount of device alarms to occur before we can recognize an attack. On the other hand, if the threshold is set too low, we may not be able to distinguish between events associated with the presence and the absence of an attack. In what follows, we present a few examples to demonstrate how one can assign threshold values for both levels of alarms. Our goal is to have both FP_{attack} and FN_{attack} approximately no greater than 10^{-10}.

Example 2.1

Assume $M = 10$ and the detection device threshold of 0.25%. If an observation period is 1 μs, we have from table 2.1 that the expected number of device alarms in an observation period is 950 and 770 with and without an attack respectively. At the transmission rate of 1 Gb/s, we transmit 1,000 bits in 1 μs. Therefore, the number of device alarms in an observation period can be at most 1,000.

Denote the attack alarm threshold by β. Let X denote the number of device alarms in an observation period. It follows that X is a binomial random variable whose probability of success, or equivalently the probability of device alarm generation, is equal to $FP_{\text{end-to-end}}$ when there is no attack, and is equal to $1 - FN_{\text{end-to-end}}$ when an attack is present. The values of FP_{attack} and FN_{attack} is given by

$$\begin{aligned} FP_{\text{attack}} &= Pr\{X \geq \beta \mid \text{no attack}\} \\ &= \sum_{i=\beta}^{1,000} \binom{1,000}{i} FP^{i}_{\text{end-to-end}}(1 - FP_{\text{end-to-end}})^{1,000-i} \qquad (2.23) \\ FN_{\text{attack}} &= Pr\{X < \beta \mid \text{attack}\} \\ &= \sum_{i=0}^{\beta-1} \binom{1,000}{i} (1 - FN_{\text{end-to-end}})^{i} FN^{1,000-i}_{\text{end-to-end}}. \qquad (2.24) \end{aligned}$$

We shall set β to be such that $FP_{\text{attack}} \approx 10^{-10}$ and find the corresponding value of FN_{attack}. Note that if $\beta > 1,000 \times (1 - FN_{\text{end-to-end}})$, or equivalently the expected number of device alarms in the presence of an attack is lower than β, we cannot construct a detection scheme based on the given detection device threshold and observation period.

In this case, we have that $FP_{\text{end-to-end}} \approx 0.77$, and $FN_{\text{end-to-end}} \approx 0.05$. We find, by trial and error, the value of β to be ≈ 851. The corresponding FN_{attack} is $< 10^{-16}$.

We conclude that with the observation period of 1 μs, the detection device threshold of 0.25%, and the attack alarm threshold of 851, our detection scheme can detect a constant jamming attack (yielding the BER above 10^{-8}) that lasts for longer than 1 μs.

Example 2.2

Assume $M = 1$ and the detection device threshold of 1.0%. For the observation period of 1 μs, we have $\beta \approx 550$. The corresponding FN_{attack} is 7.41×10^{-4}, which is too high for our purpose. If we change the observation period to 10 μs, we can choose β to be 4,960. The corresponding FP_{attack} is $< 10^{-10}$ and the corresponding FN_{attack} is $< 10^{-16}$.

We conclude that with the observation period of 10 μs, the detection device threshold of 1.0%, and the attack alarm threshold of 4,960, our detection scheme can detect a constant jamming attack (yielding the BER above 10^{-8}) that lasts for longer than 10 μs.

In both examples, the required observation period is much shorter than the time required if we are to use a BERT to detect the same jamming attacks. In particular, we detect simple in-band jamming attack scenarios in less than 10 μs in our examples. A similar task of distinguishing between the BERs of 4.6×10^{-10} (the value of BER_{baseline}) and 10^{-8} would require several seconds if we use a BERT. For example, using an observation period of 10 s, a BERT can expect 4.6 bit errors when there is no attack and 100 bit erros when there is an attack yielding the BER of 10^{-8}. A shorter observation period will not work since the expected number of bit errors in the presence of an attack do not differ significantly from the expected number of bit errors in the absence of an attack. Therefore, Our attack detection scheme is 6 orders of magnitude faster than a BERT.

As a final note, to achieve a better estimate when FN_{attack} is extremely small (as denoted by $< 10^{-16}$ in our examples), we can use large deviation theory [2]. The calculation of FN_{attack} in these cases is the subject of future research.

In conclusion, we have demonstrated with examples how one can construct a jamming attack detection system and indentify the corresponding observation period.

2.5 Attack Detection Scheme: Out-of-Band Jamming Causing Gain Competition at EDFAs

In this section, we investigate the effects of out-of-band jamming attacks causing gain competition at EDFAs. We then demonstrate with examples that our proposed detection scheme in section 2.4 can also be applied to detect out-of-band jamming attacks in this section.

2.5.1 EDFA Characteristics

We assume that the gain value γ of an EDFA is real and constant; and we can correctly perform gain estimation at the device. In the case of in-band jamming, we do not specify the source of noise. However, in this section, we are concerned with ASE noise which is dominant in EDFAs. An important property is the dependence of the noise variance on the EDFA gain.

We obtain the data of noise variances together with EDFA gains from [1]. This set of data corresponds to the transmission at 1540 nm where the gain fluctuation is the largest during an attack. Table 2.3 shows the values of noise variances associated with EDFA gains. We shall base our analysis on this set of data.

For future reference, let σ_N^2 denote the noise variance (corresponding to the gain γ)

when there is no attack, and $\sigma_{\tilde{N}}^2$ denote the noise variance when the EDFA gain is degraded due to gain competition.

EDFA Gain	Degradation	Noise Variance	EDFA Gain	Degradation	Noise Variance
86.160	0%	1.7590×10^{-7}	85.585	0.6674%	1.7489×10^{-7}
86.052	0.1253%	1.7571×10^{-7}	85.424	0.8542%	1.7461×10^{-7}
85.908	0.2925%	1.7546×10^{-7}	85.135	1.1896%	1.7410×10^{-7}
85.786	0.4341%	1.7524×10^{-7}	84.580	1.8338%	1.7313×10^{-7}
85.679	0.5583%	1.7506×10^{-7}	84.234	2.2354%	1.7252×10^{-7}

Table 2.3: Noise variances (σ_N^2 and $\sigma_{\tilde{N}}^2$) associated with EDFA square magnitude gains (γ^2) for the transmission at 1540 nm and a bit rate of 1 Gb/s using 1 GHz bandwidth.

2.5.2 Degraded BER

Given that an attack attenuates the EDFA gain at each of the M network nodes by $a\%$, we have the following expression for $BER_{\text{end-to-end}}$,

$$\begin{aligned} BER_{\text{end-to-end}} &= \frac{1}{2}(1 - F_d(\sigma_{\tilde{N}}^2, A, 0)) \\ &\quad + \frac{1}{2}F_d(\sigma_{\tilde{N}}^2, A, \sqrt{(1 - 0.01a)^M 160\sigma_N}). \end{aligned} \tag{2.25}$$

As a reminder, A is the optimal decision threshold at the receiver, which was found is section 2.3.4 to be approximately $44\sigma_{\tilde{N}}^2$.

2.5.3 False Positive and False Negative Probabilities

Suppose the detection device threshold is set at $t\%$ of the ON level ($160\sigma_{\tilde{N}}^2$) and the corresponding EDFA gain is γ. Given the gain degradation of $a\%$ at each EDFA, the expressions for FP_{node} and FN_{node} are

$$FP_{\text{node}} = 1 - F_d\left(\frac{1}{M\gamma^2}\sigma_{\tilde{N}}^2, (.01t)\frac{1}{\gamma^2}160\sigma_{\tilde{N}}^2, 0\right), \tag{2.26}$$

$$FN_{\text{node}} = \frac{1}{2}F_d\left(\frac{1}{M\gamma^2}\sigma_{\tilde{N}}^2, (.01t)\frac{1}{\gamma^2}160\sigma_{\tilde{N}}^2, 0\right) \tag{2.27}$$

$$+\frac{1}{2}F_d\left(\frac{1}{M\gamma^2}\sigma_{\tilde{N}}^2, (.01t)\frac{1}{\gamma^2}160\sigma_{\tilde{N}}^2, \sqrt{(.01a)\frac{1}{\gamma^2}160\sigma_N}\right), \tag{2.28}$$

where the ON level is scaled by $\frac{1}{\gamma^2}$ to reflect the gain loss prior to the network node.

	0.25%	0.5%	1.0%	1.5%	2.0%
1 μs	770/1000	170/997	3.4/937	0.07/623	0.001/250
0.1 μs	77/100	17/99.7	0.34/93.7	0.007/62.3	0.0001/25
0.01 μs	7.7/10	1.7/9.97	0.034/9.37	0.0007/6.23	0.00001/2.5

Table 2.4: Expected number of device alarms in an observation period ($M = 10$, out-of-band jamming). All the entries have the same meanings as in table 2.1.

	0.25%	0.5%	1.0%	1.5%	2.0%
1 μs	900/996	520/983	89/954	13/900	1.7/794
0.1 μs	90/99.6	52/98.3	8.9/95.4	1.3/90	0.17/79.4
0.01 μs	9/9.96	5.2/9.83	0.89/9.54	0.13/9	0.017/7.94

Table 2.5: Expected number of device alarms in an observation period ($M = 5$, out-of-band jamming). All the entries have the same meanings as in table 2.1.

The values of $FP_{\text{end-to-end}}$ and $FN_{\text{end-to-end}}$ can be found from equations 2.22 and 2.21.

2.5.4 Construction of an Attack Detection Scheme

Figures 2-9 and 2-10 show the curves of the expected device alarm rate versus $BER_{\text{end-to-end}}$ for different values of detection device thresholds. Tables 2.4 and 2.5 present the corresponding expected numbers of device alarms in an observation period with and without an attack.

Note that we consider the case when $M = 5$ instead of $M = 1$ since the maximal gain degradation in table 2.3 does not yield a degraded level of BER above 10^{-8} in the case of $M = 1$. We can construct the detection scheme based on the procedures given in example 2.1 of section 2.4. Our goal is to have both FP_{attack} and FN_{attack} approximately no greater than 10^{-10}. We end this section with examples.

Example 2.3

Assume that $M = 10$, the detection device threshold is 0.5%, and the observation period is 0.1 μs. In this case, we find β to be 44. The corresponding FP_{attack} is $< 10^{-10}$, and the corresponding FN_{attack} is $< 10^{-16}$. Therefore, we can detect a constant out-of-band jamming attack (yielding the BER above 10^{-8}) which lasts longer than 0.1 μs.

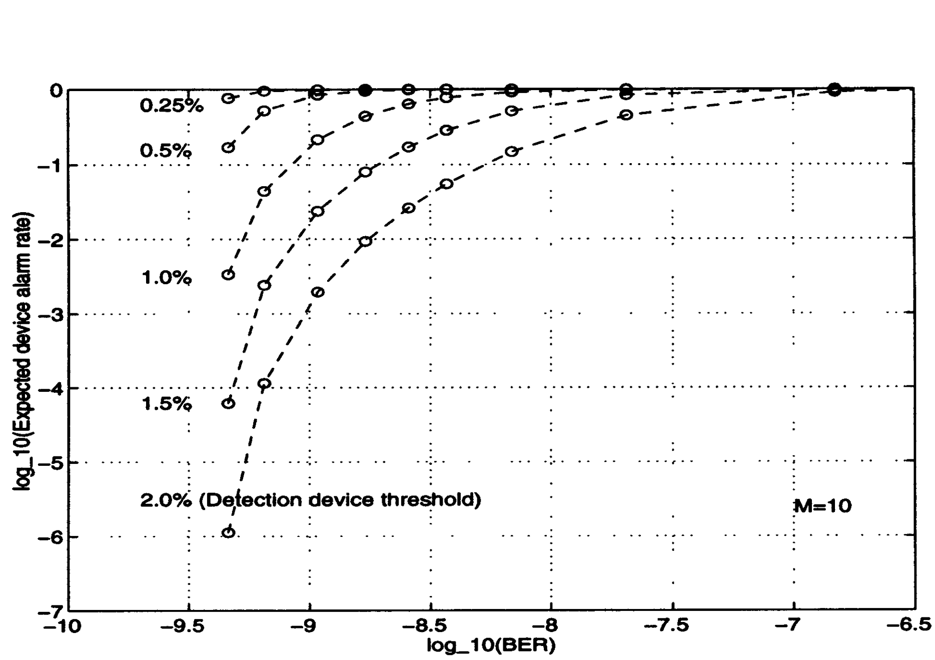


Figure 2-9: Expected device alarm rate versus $BER_{\text{end-to-end}}$ for different detection device thresholds (M=10, out-of-band jamming).

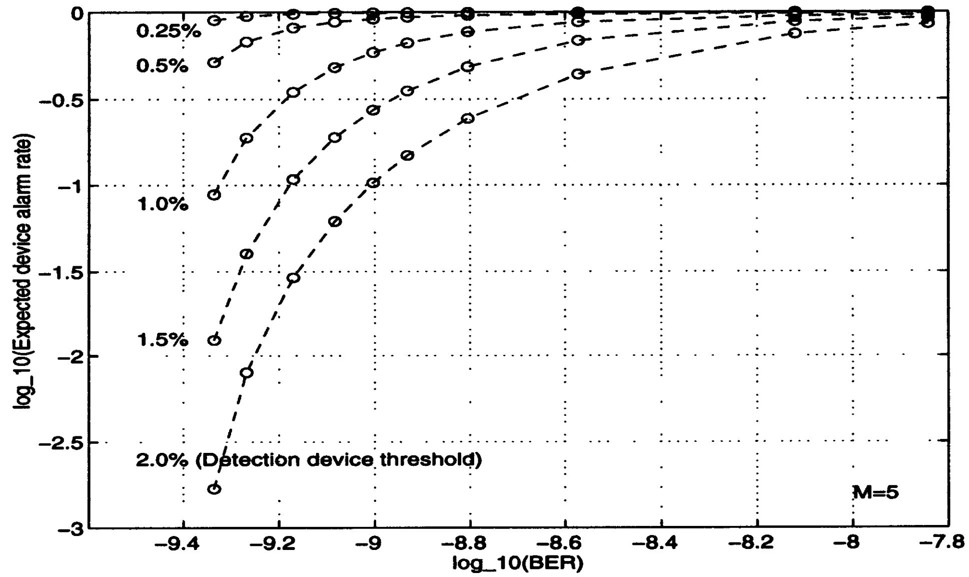


Figure 2-10: Expected device alarm rate versus $BER_{\text{end-to-end}}$ for different detection device thresholds (M=5, out-of-band jamming).

Example 2.4

Assume that $M = 5$, the detection device threshold is 1.0%, and the observation period is 0.1 μs. In this case, we find β to be 31. The corresponding FP_{attack} is $< 10^{-10}$, and the corresponding FN_{attack} is $< 10^{-16}$. Therefore, we can detect a constant out-of-band jamming attack (yielding the BER above 10^{-8}) which lasts longer than 0.1 μs.

We argue in section 2.4 that the time required for a BERT to distinguish between the absence and the presence of an attack yielding the degraded BER of 10^{-8} is in the order of several seconds. Both examples show that the required observation time for our attack detection scheme is 7 orders of magnitude smaller than the time required by a BERT.

2.6 Summary

In this chapter, we have demonstrated how one can construct a detection scheme to detect some in-band and out-of-band jamming attacks associated with the BER above 10^{-8}. The preliminary results obtained in sections 2.4 and 2.5 indicate a satisfactory performance of our attack detection scheme. Having demonstrated that our attack detection scheme can perform well at least in some pessimistic attack scenarios, we are encouraged to investigate the performance limit of this approach for detecting jamming attacks in a general case when we eliminate the direct detection constraint. This investigation is carried out in the next chapter.

Chapter 3

Worst Case In-Band Jamming Attack Scenarios

In this chapter, we search for in-band jamming attack scenarios that yield the lowest probability of getting caught by our proposed attack detection scheme constructed in chapter 2. The results obtained in this chapter will be useful for the performance analysis of our attack detection scheme in chapter 4.

We have pointed out in section 1.1.3 that, with coherent detection, we can improve the BER of the communication path of interest. Since our purpose is to investigate the performance of our approach for detecting jamming attacks in general rather than to construct a detection scheme for a particular implementation, it is appropriate to perform our analysis without the constraint that direct detection is used.

We shall consider only in-band jamming attacks. The subject of the worst case out-of-band jamming attack scenarios is extremely device specific and is outside the scope of this research.

3.1 Setup for the Analysis

3.1.1 OOK Signalling and the BER

Assume in all analyses throughout this chapter that we transmit data using OOK. In addition, assume that the receiver as well as the attack detection devices at all network nodes use coherent detection.

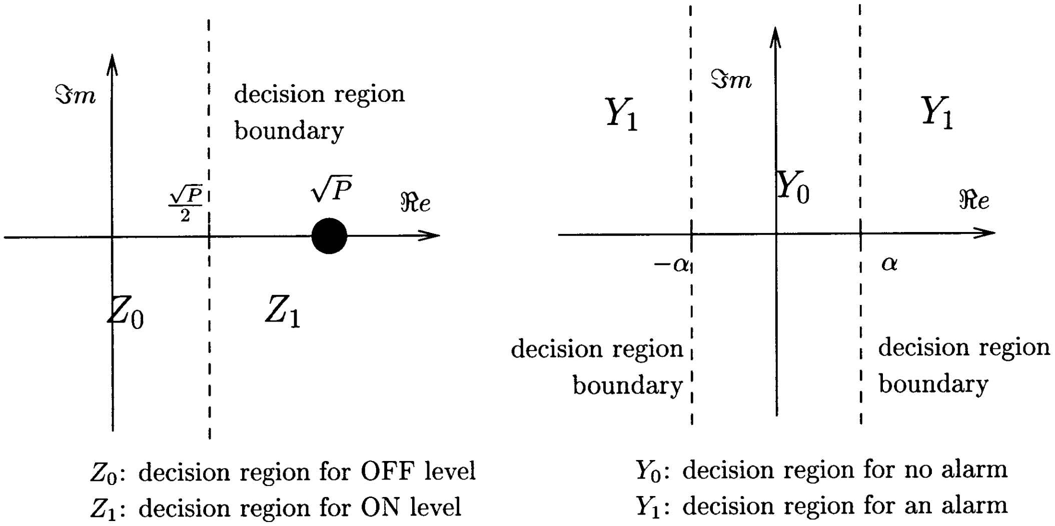


Figure 3-1: The left diagram shows the decision regions for the receiver, while the right diagram shows the decision regions for the attack detection device at each network node.

Let s_R and s_I denote the real and imaginary parts of a signal s. Let s^{OFF} and s^{ON} denote the OFF and the ON signals for OOK signalling. We shall assume that the OFF and the ON signals arequally likely. By expressing s in the form (s_R, s_I), we can write s^{OFF} as (0,0) and s^{ON} as $(s_R^{\text{ON}}, s_I^{\text{ON}})$. Assume the AWGN at each network node at any bit time has variance σ^2 for both real and imaginary parts. Given that there are M network nodes in the light path of interest, it follows that the end-to-end BER can be expressed as $Q\left(\frac{||s^{\text{ON}}||}{2\sqrt{M\sigma^2}}\right)$, where Q is the complementary cumulative distribution function of a zero-mean, unit-variance Gaussian random variable, and $||s||$ is the norm of the signal s. Note that the overall noise is the accumulation of noise components at all M network nodes since there is no processing of signals at network nodes.

Because the BER depends only on the norm of the ON signal, given that the power of the ON signal is P, we can choose s^{ON} to be $(\sqrt{P}, 0)$ for convenience in the analysis that follows. With this choice of s^{ON}, our decision region is shown in the left diagram of figure 3-1. Note that these decision regions are the same as the ones in the left diagram of figure 1-1. In the absence of an attack, we can express the baseline value of the end-to-end BER as $Q\left(\frac{\sqrt{P}}{2\sqrt{M\sigma^2}}\right)$.

Let $J^{(i)}$ be the jamming signal at node i. Using the fact that attacks propagate along

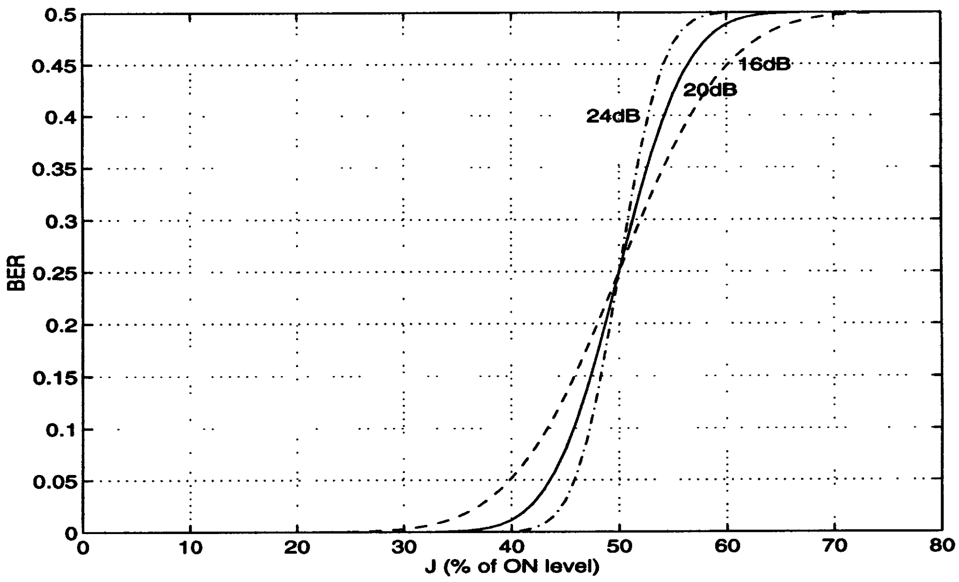


Figure 3-2: The BER curves as a function of accumulated coherent components of jamming signals (equal to $\sum_{i=1}^{T} J_C^{(i)}$) in percentage of $\sqrt{P}$.

the light path, we can express the end-to-end BER for a particular bit as follow

$$BER_{\text{end-to-end}} = \frac{1}{2}Q\left(\frac{\frac{\sqrt{P}}{2} - \sum_{i=1}^{M} J_C^{(i)}}{\sqrt{M\sigma^2}}\right) + \frac{1}{2}Q\left(\frac{\frac{\sqrt{P}}{2} + \sum_{i=1}^{M} J_C^{(i)}}{\sqrt{M\sigma^2}}\right), \tag{3.1}$$

where $J_C^{(i)}$ denotes the component of $J^{(i)}$ which is coherent to the transmitted signal. Coherent components of jamming signals are assumed to add up constructively along the light path.

Note that the expression in equation 3.1 is different from the expression given in equation 2.18 since we use coherent detection instead of direct detection. Figure 3-2 shows the BER curves for different values of the SNR as functions of accumulated coherent components of jamming signals (equal to $\sum_{i=1}^{M} J_C^{(i)}$) in percentage of $\sqrt{P}$. To find the value of P for a particular SNR, we use the relation

$$SNR = \frac{1}{2}\left(\frac{P}{2M\sigma^2}\right) + \frac{1}{2}0, \tag{3.2}$$

where a factor of 2 in $2M\sigma^2$ takes into account the imaginary component as well as the real component of the noise.

3.1.2 Device Alarm Generation and the Device Alarm Rate at a Network Node

Consider the output d of an attack detection device at a particular network node. Assume that coherent detection is used. We would like to generate an alarm at the network node when the jamming signal is likely to degrade the BER significantly, or equivalently when the magnitude of d_C (the component of d which is coherent to the transmitted signal) is large. It is appropriate to set the decision regions for alarm generation in the following fashion: generate a device alarm when $|d_C| > \alpha$ where α is a real positive number generally smaller than $\sqrt{P}/2$ since we are trying to detect the BER degradation rather than the error itself. This decision region is shown in the right diagram of figure 3-1.

For a given detection device threshold α and a jamming signal at the network node $J^{(i)}$, we have the expected device alarm rate or equivalently the probability of alarm generation at the network node equal to

$$\text{Alarm rate} = Q\left(\frac{\alpha - J_C^{(i)}}{\sigma}\right) + Q\left(\frac{\alpha + J_C^{(i)}}{\sigma}\right). \tag{3.3}$$

Figure 3-3 shows the expected device alarm rate curve together with the BER curve for the SNR of 20 dB and α of $0.15\sqrt{P}$. Plot 3-4 shows the expected device alarm rate at a single network node versus the BER curves for different values of α.

3.1.3 Formulation of the Optimization Problem

We now formulate the problem of searching for the "worst case" in-band jamming attack scenario as an optimization problem. We start by defining what we mean by the worst case in-band jamming attack scenario.

Let T denote the length of an observation period in bits, and $\tilde{B}$ denote the average degraded level of BER over T bits. Among all in-band jamming attacks associated with $\tilde{B}$, the attack scenario yielding the smallest number of device alarms in a period of T bits has the smallest probability of getting caught. It is therefore appropriate to consider the "worst case" attack scenario as the one which yields the smallest expected number of device alarms in a period of T bits, or equivalently the smallest expected device alarm rate (the expected number of device alarms in an observation period divided by the number of bits in an observation period).

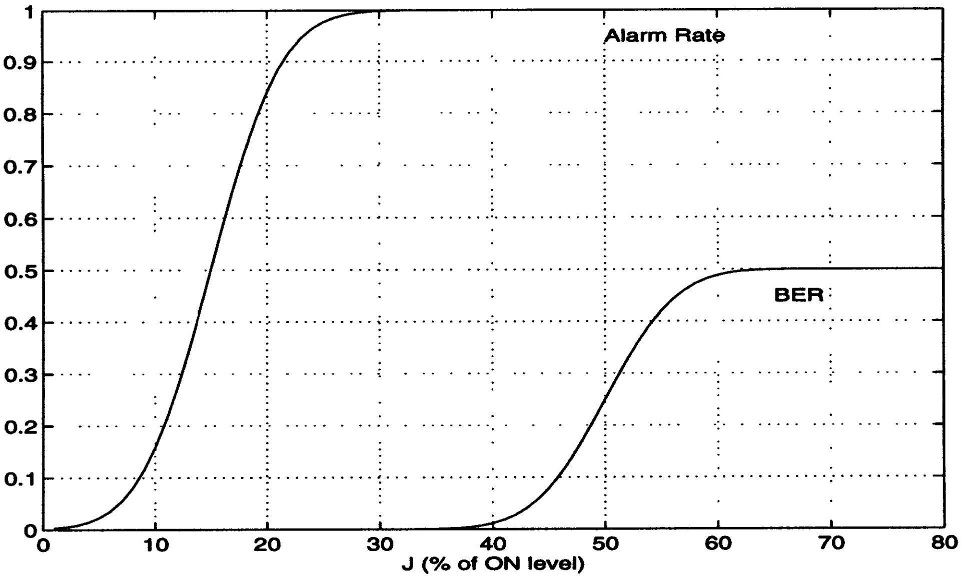


Figure 3-3: The expected device alarm rate curve as a function of the jamming signal (coherent component) at a single network node together with the BER curve as a function of accumulated coherent components of jamming signals along the light path. The jamming signals (coherent components) in both cases are in percentages of $\sqrt{P}$. In this plot, the SNR is 20 dB, the attack detection device threshold α is $0.15\sqrt{P}$.

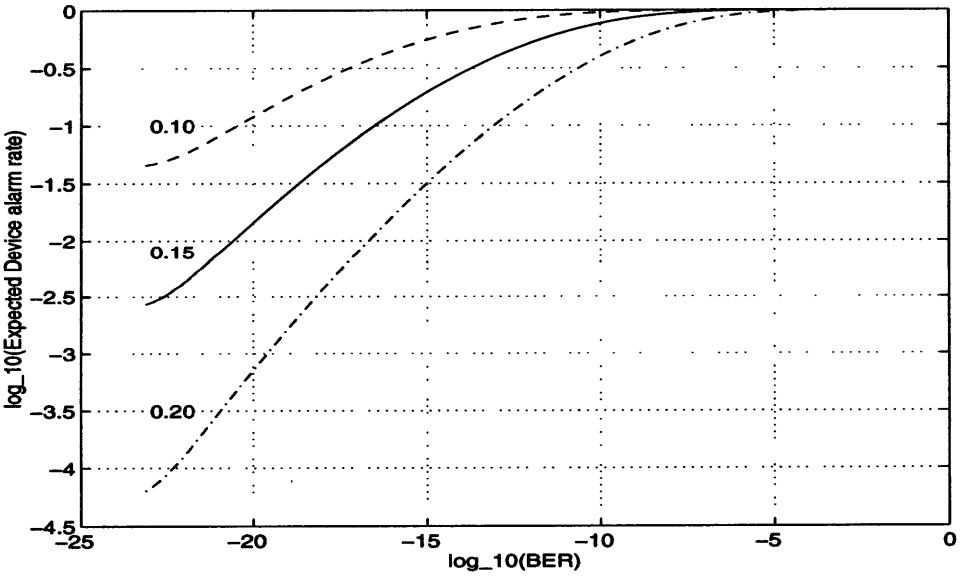


Figure 3-4: Expected device alarm rate versus BER for different attack detection device thresholds (α) in proportion of $\sqrt{P}$. In this plot, the SNR is 20 dB, and there is one network node ($M = 1$).

We shall describe an in-band jamming attack scenario in terms of the coherent components of jamming signals at all network nodes at all bit times. In this regard, searching for the worst case attack scenario is equivalent to finding the values of the corresponding jamming signals (coherent components) which yield the smallest expected device alarm rate. Therefore, we can treat the jamming signals (coherent components) as variables and find their values by solving the following optimization problem.

$$\begin{array}{ll} \text{minimize} & \text{expected device alarm rate} \\ \text{subject to} & \text{average BER in } T \text{ bits} = \tilde{B}. \end{array}$$

For future reference, let $B(J)$ and $A(J)$ denote the BER and the expected device alarm rate functions whose expressions in equations 3.1 and 3.3 are presented again below.

$$B(J) = \frac{1}{2}Q\left(\frac{\frac{\sqrt{P}}{2} - J}{\sqrt{M\sigma^2}}\right) + \frac{1}{2}Q\left(\frac{\frac{\sqrt{P}}{2} + J}{\sqrt{M\sigma^2}}\right) \tag{3.4}$$

$$A(J) = Q\left(\frac{\alpha - J}{\sigma}\right) + Q\left(\frac{\alpha + J}{\sigma}\right). \tag{3.5}$$

Note that the argument of the BER function in equation 3.4 is the accumulated coherent components of jamming signals along the light path. The argument of the expected device alarm rate function in equation 3.5 is the coherent component of a jamming signal at a network node.

3.2 Worst Case Scenario for In-band Jamming at Multiple Bits at a Single Network Node

Assume there is only one network node in the light path of interest (M=1). The goal of this section is to find the worst case scenario for in-band jamming at T ($T > 1$) bits. We shall start with the simplest case with $T = 2$ and extend the result to a general case with $T > 2$.

3.2.1 Worst Case Scenario for In-band Jamming at Two Bits at a Single Network Node

Denote the coherent component of the jamming signal at the first bit by J_1 and at the second bit by J_2. Finding the worst case jamming attack scenario is similar to solving the following minimization problem for J_1 and J_2.

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\left(A(J_1) + A(J_2)\right) \\ \text{subject to} \quad & \tilde{B} = \frac{1}{2}\left(B(J_1) + B(J_2)\right) \end{aligned} \tag{3.6}$$

Using the Lagrange multiplier method, we form the Lagrangean $L(J_1, J_2, \lambda)$ defined by

$$L(J_1, J_2, \lambda) = \frac{1}{2}\left(A(J_1) + A(J_2)\right) - \lambda\left(\frac{1}{2}\left(B(J_1) + B(J_2)\right) - \tilde{B}\right).$$

By differentiating the Lagrangean with respect to each variable and solving the system of simultaneous equations, we get the locations of extrema. In this case the system of equations is

$$\frac{\partial}{\partial J_i}L = 0 \quad = \quad \frac{1}{2}A'(J_i) - \lambda\frac{1}{2}B'(J_i),\ i = 1, 2 \tag{3.7}$$

$$\frac{\partial}{\partial \lambda}L = 0 \quad = \quad \frac{1}{2}\left(B(J_1) + B(J_2)\right) - \tilde{B}. \tag{3.8}$$

We would like to find the values of J_1 and J_2 at the extrema. To find possible solutions for equation 3.7, consider the curve of $A'(J_i)$ superposed on the curve of $\lambda B'(J_i)$. Continuing with our example (SNR=20 dB,α=$0.15\sqrt{P}$), figure 3-5 shows example curves of $A'(J_i)$ and $\lambda B'(J_i)$ for different values of λ. Note that, for the existence of nonzero solutions, λ must be positive.

The points of intersection of $A'(J_i)$ and $\lambda B'(J_i)$ specify the solutions (J_1, J_2) to equation 3.7. Because the functions $A(J)$ and $B(J)$ are even, we shall consider only nonnegative values of J_i.

One can observe from figure 3-5 that the curves of $A'(J_i)$ and $\lambda B'(J_i)$ (with $\lambda > 0$) can intersect at no more than two nonnegative points: zero and a strictly positive value. Since $A'(0) = B'(0) = 0$ (equations 3.4 and 3.5), it is easy to see that zero is indeed a solution to equation 3.7. The fact that at most one positive solution exists is verified in lemma 3.3

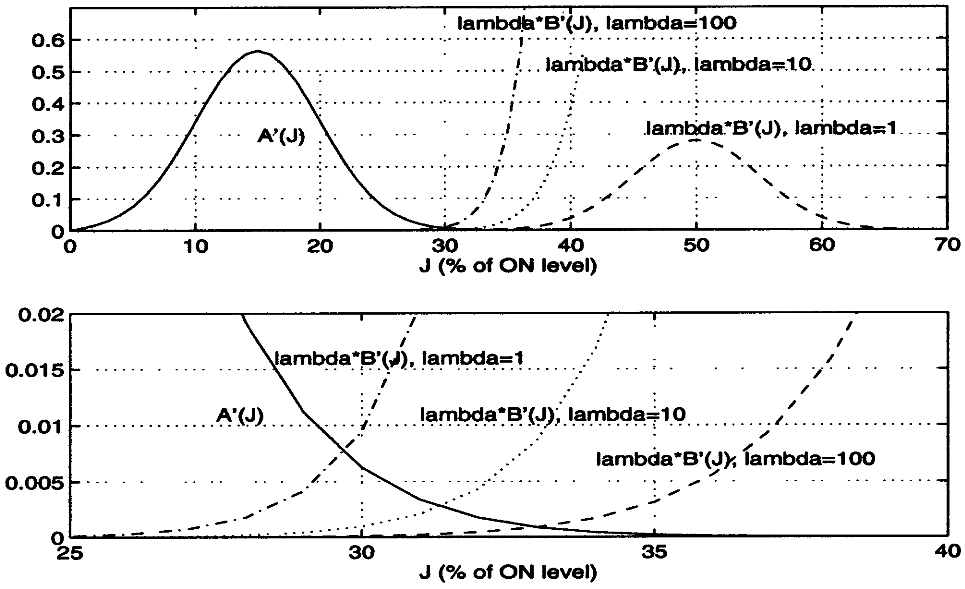


Figure 3-5: The curve of $A'(J_i)$ superposed on the curves of $\lambda B'(J_i)$ for different values of λ ($\lambda > 0$). Notice that for all values of λ, the two curves intersect at zero and a strictly positive value. In this plot, the SNR is 20 dB, and α is $0.15\sqrt{P}$.

with some restrictions given in assumption 3.1.

Assumption 3.1 *Let σ denote the variance of both real and imaginary parts of the AWGN at any network node at any bit time, α denote the attack detection device threshold, and η = $\sqrt{P}/2$. Assume that $\sigma < \alpha < \eta - \sqrt{3}\sigma$.*

Notice that assumption 3.1 generally holds when the SNR is sufficiently high since the value of σ will be small in comparison to the value of η.

Lemma 3.1 *Given that $\lambda > 0$, $\eta > \alpha$, consider the function $f(x)$ of the form*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\alpha)^2}{2\sigma^2}} - \lambda \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\eta)^2}{2\sigma^2}}. \tag{3.9}$$

$f(x)$ has the following characteristics:

i) There is a unique solution to $f(x) = 0$.

ii) $f(x)$ *has a unique local maximum and a unique local minimum.*

iii) Let x^{max} *and* x^{min} *denote the locations of the maximum and the minimum respectively. We have* $x^{max} < \alpha < \eta < x^{min}$.

iv) As x *increases from* $-\infty$ *to* x^{max}, $f(x)$ *strictly increases from the limit value of 0 to* $f(x^{max})$ *and has a single saddle point where* $f(x)$ *changes from convexity* $(\smile)$ *to concavity* $(\frown)$.

v) As x *increases from* x^{max} *to* x^{min}, $f(x)$ *strictly decreases.*

vi) As x *increases from* x^{min} *to* ∞, $f(x)$ *strictly increases from* $f(x^{min})$ *to the limit value of 0 and has a single saddle point at which* $f(x)$ *changes from convexity* $(\smile)$ *to concavity* $(\frown)$.

Proof The proof is provided in appendix A. □

Lemma 3.2 *Under assumption 3.1,* $f(x)$ *as defined in equation 3.9 has a strictly decreasing second derivative in* $(x^s, \alpha]$, *where* x^s *denotes the unique saddle point in* $(-\infty, x^{max})$.

Proof The proof is provided in appendix A. □

Lemma 3.3 *Given* $\lambda > 0$, *and assumption 3.1, there is at most one positive solution to the equation* $A'(J) = \lambda B'(J)$ *in* $(0, \eta - \sigma)$.

Proof The proof is provided in appendix A. □

Given that only zero and one other positive value are solutions to equation 3.7, there are two forms of solutions to equation 3.8, namely $J_1 = J_2$, and J_1 (or J_2) $= 0$ with J_2 (or J_1) > 0. Let $\tilde{J}$ be such that $B(\tilde{J}) = \tilde{B}$ and J^s be such that $\frac{1}{2}B(0) + \frac{1}{2}B(J^s) = \tilde{B}$. It follows that the extrema are $(\tilde{J}, \tilde{J})$, $(0, J^s)$, and $(J^s, 0)$. Therefore, any extremum must correspond to either an attack with equal coherent jamming signal components at the two bits (referred to as constant jamming) or an attack with jamming at only one bit (referred to as sporadic jamming).

We would like to find out when, if ever, constant jamming, $(J_1, J_2)=(\tilde{J}, \tilde{J})$, yields a lower expected device alarm rate than sporadic jamming, $(J_1, J_2)=(0, J^s)$ or $(J^s, 0)$, and vice versa.

According to [15], given the locations of the extrema, we can investigate their properties by performing the second derivative test for constrained extrema. In particular, we evaluate the bordered Hessian $|H|_{(J_1,J_2)}$ of the Lagrangean which in this case is equal to

$$\begin{aligned} |H|_{(J_1,J_2)} &= \begin{vmatrix} 0 & -B'(J_1) & -B'(J_2) \\ -B'(J_1) & L''(J_1) & 0 \\ -B'(J_2) & 0 & L''(J_2) \end{vmatrix} \\ &= -L''(J_1)(B'(J_2))^2 - L''(J_2)(B'(J_1))^2 \\ &= -\frac{1}{2}\left[A''(J_1) - \lambda B''(J_1)\right](B'(J_2))^2 \\ &\quad -\frac{1}{2}\left[A''(J_2) - \lambda B''(J_2)\right](B'(J_1))^2. \end{aligned} \tag{3.10}$$

As a reminder, (J_1, J_2) is a local minimum when $|H|_{(J_1,J_2)} < 0$, a local maximum when $|H|_{(J_1,J_2)} > 0$, and no conclusion can be made when $|H|_{(J_1,J_2)} = 0$. For constant jamming, the bordered Hessian in equation 3.10 can be written as

$$|H|_{(\tilde{J},\tilde{J})} = -\left[A''(\tilde{J}) - \frac{A'(\tilde{J})}{B'(\tilde{J})}B''(\tilde{J})\right](B'(\tilde{J}))^2, \tag{3.11}$$

where λ is found from the equation $A'(\tilde{J}) = \lambda B'(\tilde{J})$. Figure 3-6 shows the curve of $|H|_{(\tilde{J},\tilde{J})}$ as a function of $\tilde{J}$. Notice that, for all values of $\tilde{J}$ (equivalently for all values of $\tilde{B}$), $|H|_{(\tilde{J},\tilde{J})}$ is positive. Lemma 3.5 verifies this fact under assumption 3.1.

Lemma 3.4 *Under assumption 3.1,* $\frac{A'(J)}{B'(J)} < \frac{A''(0)}{B''(0)}$ *for all* $J > 0$.

Proof We use x as a dummy variable. For convenience, the expressions for $A'(x), A''(x), B'(x)$ and $B''(x)$ are given below

$$\begin{aligned} A'(x) &= \frac{1}{\sqrt{2\pi}\sigma}\left[e^{-\frac{(x-\alpha)^2}{2\sigma^2}} - e^{-\frac{(x+\alpha)^2}{2\sigma^2}}\right] && (3.12) \\ A''(x) &= \frac{1}{\sqrt{2\pi}\sigma^3}\left[-(x-\alpha)e^{-\frac{(x-\alpha)^2}{2\sigma^2}} + (x+\alpha)e^{-\frac{(x+\alpha)^2}{2\sigma^2}}\right] && (3.13) \\ B'(x) &= \frac{1}{\sqrt{2\pi}\sigma}\left[e^{-\frac{(x-\eta)^2}{2\sigma^2}} - e^{-\frac{(x+\eta)^2}{2\sigma^2}}\right] && (3.14) \\ B''(x) &= \frac{1}{\sqrt{2\pi}\sigma^3}\left[-(x-\eta)e^{-\frac{(x-\eta)^2}{2\sigma^2}} + (x+\eta)e^{-\frac{(x+\eta)^2}{2\sigma^2}}\right] && (3.15) \end{aligned}$$

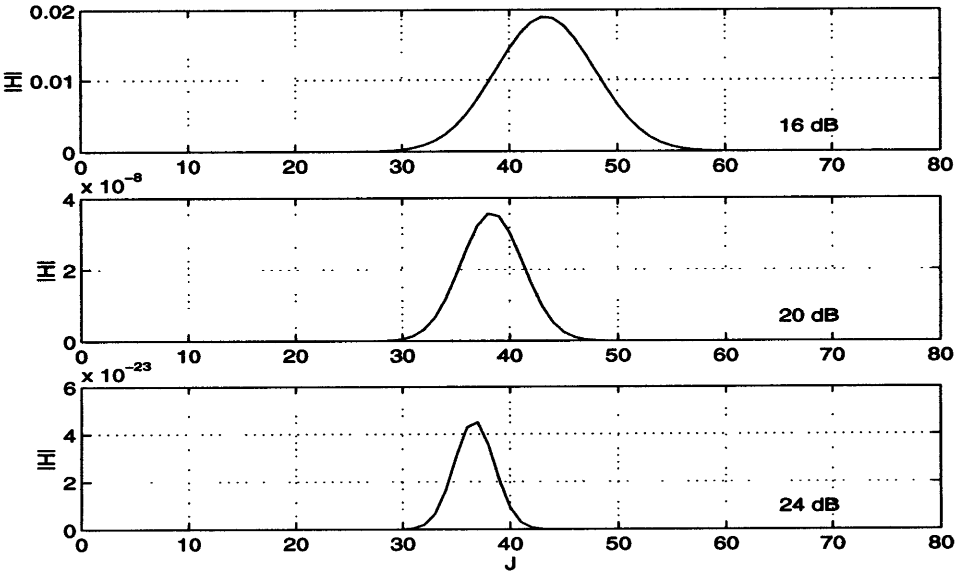


Figure 3-6: $|H|_{(\tilde{J},\tilde{J})}$ as a function of $\tilde{J}$. The values of α for the SNRs of 16, 20, and 24 dB are $0.30\sqrt{P}$, $0.15\sqrt{P}$, and $0.10\sqrt{P}$ respectively.

We now prove the lemma in the following fashion,

$$\begin{aligned}
\frac{A'(x)}{B'(x)} &= \frac{e^{-\frac{(x-\alpha)^2}{2\sigma^2}} - e^{-\frac{(x+\alpha)^2}{2\sigma^2}}}{e^{-\frac{(x-\eta)^2}{2\sigma^2}} - e^{-\frac{(x+\eta)^2}{2\sigma^2}}} \\
&= \frac{e^{-\frac{\alpha^2}{2\sigma^2}}\left[e^{\frac{x\alpha}{\sigma^2}} - e^{-\frac{x\alpha}{\sigma^2}}\right]}{e^{-\frac{\eta^2}{2\sigma^2}}\left[e^{\frac{x\eta}{\sigma^2}} - e^{-\frac{x\eta}{\sigma^2}}\right]} \\
&= \frac{\alpha e^{-\frac{\alpha^2}{2\sigma^2}}\left[\frac{\sigma^2}{\alpha}e^{\frac{x\alpha}{\sigma^2}} - \frac{\sigma^2}{\alpha}e^{-\frac{x\alpha}{\sigma^2}}\right]}{\eta e^{-\frac{\eta^2}{2\sigma^2}}\left[\frac{\sigma^2}{\eta}e^{\frac{x\eta}{\sigma^2}} - \frac{\sigma^2}{\eta}e^{-\frac{x\eta}{\sigma^2}}\right]} \\
&< \frac{\alpha e^{-\frac{\alpha^2}{2\sigma^2}}}{\eta e^{-\frac{\eta^2}{2\sigma^2}}} \\
&= \frac{A''(0)}{B''(0)},
\end{aligned}$$

where the inequality follows from the fact that the function $g(y) = \frac{1}{y}e^{xy} - \frac{1}{y}e^{-xy}$ is increasing in $(0, \infty)$ for all $x > 0$, hence $\frac{\left[\frac{\sigma^2}{\alpha}e^{\frac{x\alpha}{\sigma^2}} - \frac{\sigma^2}{\alpha}e^{-\frac{x\alpha}{\sigma^2}}\right]}{\left[\frac{\sigma^2}{\eta}e^{\frac{x\eta}{\sigma^2}} - \frac{\sigma^2}{\eta}e^{-\frac{x\eta}{\sigma^2}}\right]} < 1$ □

Lemma 3.5 *Under assumption 3.1, $|H|_{(\tilde{J},\tilde{J})}$ as given in equation 3.11 is positive for all $\tilde{J} < \eta - \sigma$ (equivalently $\tilde{B} < B(\eta - \sigma)$).*

Proof We use x as a dummy variable. To show that $|H|_{(\tilde{J},\tilde{J})} > 0$, it is sufficient to show that $G(x) = \frac{A'(x)}{B'(x)}B''(x) - A''(x) < 0$ in $(0, \eta - \sigma)$. Given that the two curves intersect at x^*, we can think of $\frac{A'(x^*)}{B'(x^*)}B''(x^*)$ as the slope of the $\lambda B'(x)$ curve and $A''(x^*)$ as the slope of the $A'(x)$ curve at the intersection point x^*.

Given an intersection at $x^* \in (0, \eta - \sigma)$, we have $\lambda = \frac{A'(x^*)}{B'(x^*)}$. It follows from lemma 3.4 that $\lambda < \frac{A''(0)}{B''(0)}$, or $\lambda B''(0) < A''(0)$. Since $\lambda B'(0) = A'(0) = 0$, $\lambda B''(0) < A''(0)$ yields $\lambda B'(0^+) < A'(0^+)$.

The fact that x^* is the only intersection point in $(0, \eta - \sigma)$ (lemma 3.3) and $\lambda B'(0^+) < A'(0^+)$ imply that, at the intersection point x^*, the slope of $\lambda B'(x)$, which is equal to $\frac{A'(x^*)}{B'(x^*)}B''(x^*)$, is larger than the slope of $A'(x)$, which is equal to $A''(x^*)$. Equivalently, $G(x^*) < 0$ for all $x^* \in (0, \eta - \sigma)$. □

Lemma 3.5 tells us that constant jamming yields a local maximal expected device alarm rate.

For sporadic jamming, the bordered Hessian can be written as

$$\begin{aligned} |H|_{(J^s,0)} &= -\frac{1}{2}\left[A''(0) - \frac{A'(J^s)}{B'(J^s)}B''(0)\right](B'(J^s))^2 \\ &\quad -\frac{1}{2}\left[A''(J^s) - \frac{A'(J^s)}{B'(J^s)}B''(J^s)\right](B'(0))^2 \\ &= -\frac{1}{2}\left[A''(0) - \frac{A'(J^s)}{B'(J^s)}B''(0)\right](B'(J^s))^2, \end{aligned} \tag{3.16}$$

where the last equality follows from the fact that $B'(0) = 0$ (equation 3.14).

Figure 3-7 shows the curve of $|H|_{(J^s,0)}$ as a function of J^s. Notice that for all values of J^s (equivalently for all values of $\tilde{B}$), $|H|_{(J^s,0)}$ is negative. Lemma 3.6 verifies this fact under assumption 3.1.

Lemma 3.6 *Under assumption 3.1, $|H|_{(J^s,0)}$ as given in equation 3.16 is negative for all $J^s < \eta - \sigma$ (equivalently $\tilde{B} < \frac{1}{2}B(0) + \frac{1}{2}B(\eta - \sigma)$).*

Proof It is sufficient to show that $\frac{A''(0)}{B''(0)}B'(J) > A'(J)$ for all x. However, this statement

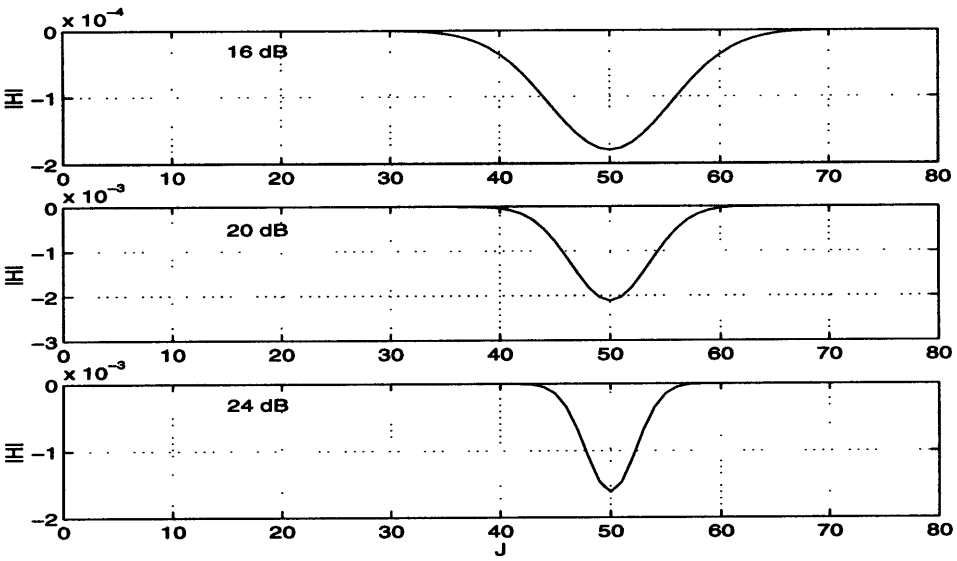


Figure 3-7: $|H|_{(J^s,0)}$ as a function of J^s. The values of α for the SNRs of 16, 20, and 24 dB are $0.30\sqrt{P}$, $0.15\sqrt{P}$, and $0.10\sqrt{P}$ respectively.

is exactly the content of lemma 3.4. □

Lemma 3.6 tells us that sporadic jamming yields a local minimal expected device alarm rate.

We now argue that the constraint set defined in equation 3.8 has no end points, i.e. the constraint set is a continuous region. Notice that the function $\frac{1}{2}B(J_1) + \frac{1}{2}B(J_2)$ is a continuous function of (J_1, J_2). Consider the first quadrant of (J_1, J_2), the constraint set is a continuous path from $(0, J^s)$ to $(\tilde{J}, \tilde{J})$ and finally to $(J^s, 0)$. By symmetry of the function $B(J)$, in the second quadrant, the constraint set is a continuous path from $(0, J^s)$ to $(-\tilde{J}, \tilde{J})$ and finally to $(-J^s, 0)$. The constraint set behaves similarly in the third and the fourth quadrants. Therefore, the constraint set is a continuous path around the origin with no end points.

Since there is no other jamming attack scenario that corresponds to an extremum in the constraint set and the constraint set has no end points (no boundary case to check for a global minimum and a global maximum), we claim that the expected device alarm rates associated with all other jamming scenarios are bounded by the local maximum and the local minimum that we found. In other words, $(J^s, 0)$ and $(0, J^s)$ are indeed global minima.

Note that the maximal value of $\tilde{B}$ such that both J_1 and J_2 are still in $[0, \eta - \alpha)$ is $\frac{1}{2}B(\eta - \alpha)+ \frac{1}{2}B(0)$. To see this, consider when $\tilde{B}$ is greater than $\frac{1}{2}B(\eta - \alpha) + \frac{1}{2}B(0)$. In this case, a jamming attack at only one bit yielding the BER of $\tilde{B}$ must have a jamming signal J^s greater than $\eta - \alpha$; and we no longer have both J_1 and J_2 in $[0, \eta - \sigma)$.

For convenience, we shall use the value $\frac{1}{2}B(\eta - \alpha)$ as the upper bound on $\tilde{B}$ since $B(0)$ is generally very small. In conclusion, we have established the following proposition.

Proposition 3.1 *Under assumption 3.1, given the average degraded BER of $\tilde{B} < \frac{1}{2}B(\eta - \sigma)$, an attack scenario in which an attacker jams at only one out of two bits has the smallest expected device alarm rate and is therefore the worst case attack scenario.*

As a final note, empirical data suggests that proposition 3.1 holds without assumption 3.1. In addition, it also holds regardless of the value of degraded BER ($\tilde{B}$).

3.2.2 Worst Case Scenario for Jamming at T Bits at a Single Network Node

In this section, we extend the result from the worst case scenario for jamming at two bits to the worst case scenario for jamming at T ($T > 2$) bits. Denote the coherent component (to the transmitted signal) of the jamming signal at the ith bit by J_i. We find the worst case jamming attack scenario by solving the following minimization problem,

$$\begin{aligned} \text{minimize} \quad & \frac{1}{T}\sum_{i=1}^{T} A(J_i) \\ \text{subject to} \quad & \tilde{B} = \frac{1}{T}\sum_{i=1}^{T} B(J_i). \end{aligned} \tag{3.17}$$

The Lagrangean, denoted by $L(J_1, ..., J_T, \lambda)$, in this case is $\frac{1}{T}\sum_{i=1}^{T} A(J_i)$ $- \lambda\left(\frac{1}{T}\sum_{i=1}^{T} B(J_i) - \tilde{B}\right)$. Performing partial differentiation with respect to J_i gives us the similar condition for an extremum as in equation 3.7 for the case of jamming at two bits,

$$\frac{\partial}{\partial J_i}L = 0 = \frac{1}{T}A'(J_i) - \lambda\frac{1}{T}B'(J_i),\ i = 1, ..., T. \tag{3.18}$$

We have from lemma 3.3 that under assumption 3.1 there exists at most one positive solution to $A'(J_i) = \lambda B'(J_i)$ in $(0, \eta-\sigma)$. We shall assume that $\tilde{B} < \frac{1}{M}B(\eta-\sigma)+\frac{M-1}{M}B(0)$ to guarantee that $J_i \in [0, \eta-\sigma)$ for all i regardless of the jamming attack scenario. To see why this assumption is necessary, consider an attack scenario when $\tilde{B} > \frac{1}{M}B(\eta-\sigma)+\frac{M-1}{M}B(0)$. In the case of jamming at only one bit, the jamming signal must be greater than $\eta-\sigma$; and we no longer have that $J_i \in [0, \eta-\sigma)$ for all i. For convenience, we use $\frac{1}{M}B(\eta-\sigma)$ instead of $\frac{1}{M}B(\eta-\sigma)+\frac{M-1}{M}B(0)$ as the restriction on $\tilde{B}$ since the value of $B(0)$ is generally very small.

At a given extremum, lemma 3.3 tells us that each J_i can take only two possible values, namely zero and a strictly positive value denoted by J_t^s where t is the number of bits that are jammed. For the remaining discussion, we shall refer to jamming at t out of T bits with jamming signal (coherent component) J_t^s as scenario t. All possible scenarios are 1,2,...,T.

Let A_t denote the average device alarm rate corresponding to scenario t, we now show that proposition 3.1 implies that $A_T > A_{T-1} > ... > A_1$. This conclusion is stated as proposition 3.2.

Lemma 3.7 *Let $\gamma < \frac{1}{2}$. In the problem of jamming at two bits given below*

$$\begin{aligned} \text{minimize} \quad & \gamma A(J_1) + (1-\gamma)A(J_2) \\ \text{subject to} \quad & \tilde{B} = \gamma B(J_1) + (1-\gamma)B(J_2), \end{aligned}$$

given assumption 3.1 and $\tilde{B} < \gamma B(\eta-\sigma)$, jamming at only one out of the two bits yields the smallest expected device alarm rate and thus corresponds to the worst case attack scenario.

Proof Note that the condition $\tilde{B} < \gamma B(\eta-\sigma)$ ensures that both J_1 and J_2 are in $[0, \eta-\sigma)$. By differentiating the Lagrangean with respect to J_i, we have the same conditions for an extremum as in the problem given in expression 3.6, namely $A'(J_i) = \lambda B'(J_i)$. Lemma 3.3 tells us that constant jamming and jamming only at one bit are the only two scenarios that correspond to the extrema. It remains to investigate the bordered Hessian $|H|_{(J_1,J_2)}$.

For constant jamming, the expression and the value of $|H|_{(J_1,J_2)}$ is exactly the same as equation 3.11 in the problem given in expression 3.6 and thus constant jamming yields a local maximal rate of the device alarm.

For sporadic jamming, the value of $|H|_{(J_1,J_2)}$ is $-\left[A''(0) - \frac{A'(J^s)}{B'(J^s)}B''(0)\right](B'(J^s))^2$ multiplied by either γ or $(1-\gamma)$ depending on whether J_1 or J_2 is equal to jamming signal J^s. In either case, $|H|_{(J_1,J_2)}$ is negative as in the problem given in expression 3.6. Thus sporadic jamming yields a local minimal rate of the device alarm.

Similarly to the arguments used to construct proposition 3.1, since the constraint set is bounded and the expected device alarm rate is bounded within the constraint set, we conclude that jamming at one bit yields the global minimal rate of the device alarm and thus corresponds to the worst case scenario. □

Lemma 3.8 *Let J_t^s be the jamming signal (coherent component) associated with scenario t in which jamming occurs at t out of T bits and the degraded BER is equal to $\tilde{B}$. Let A_t be the expected device alarm rate corresponding to scenario t. Given assumption 3.1 and $\tilde{B} < \frac{1}{T}B(\eta - \sigma)$, for jamming at one network node and at T bits, we have that $A_T > A_{T-1} > ... > A_1$.*

Proof The condition $\tilde{B} < \frac{1}{T}B(\eta - \sigma)$ guarantees that $J_i \in [0, \eta - \sigma)$ for all i. Using induction, we first show that $A_T > A_{T-1}$ and then show that $A_{T-j} > A_{T-j-1}$. This will imply that $A_T > A_{T-1} > ... > A_1$.

Let us compare scenarios T and $T-1$ yielding the device alarm rate A_T and A_{T-1} respectively. Notice that comparing these two scenarios is essentially the same as comparing the two extrema in the problem of lemma 3.7 with γ equal to $\frac{T-1}{T}$. The extrema of interest correspond to (J_1, J_2) equal to (J_T^s, J_T^s) for scenario T and equal to $(J_{T-1}^s, 0)$ for scenario $T-1$. Lemma 3.7 tells us that $A_T > A_{T-1}$.

Now compare scenarios $T-j$ to scenario $T-j-1$. Consider the problem of minimizing the average device alarm rate over the $T-j$ bits that are jammed in scenario $T-j$. Note that the device alarm rate contribution from the bits not considered is the same in both cases, namely $\frac{T-j}{T}A(0)$. This problem is again similar to comparing the two extrema in the problem of lemma 3.7 with γ now equal to $\frac{T-j-1}{T-j}$ and the BER constraint equal to $\tilde{B} - \frac{T-j}{T}B(0)$. The extrema of interest are (J_{T-j}^s, J_{T-j}^s) and $(J_{T-j-1}^s, 0)$. Lemma 3.7 tells us that $A_{T-j} > A_{T-j-1}$. □

Proposition 3.2 *Under assumption 3.1, given $\tilde{B} < \frac{1}{T}B(\eta - \sigma)$, an attack scenario in which* *an attacker jams at only one out of T bits has the smallest expected device alarm rate and is therefore the worst case attack scenario.*

Proof The proposition is the direct consequence of lemma 3.8 for $t = 1$. □

We conclude that given assumption 3.1 and $\tilde{B} < \frac{1}{T}B(\eta - \sigma)$, jamming at only one out of T bits corresponds to the worst case attack scenario. Finally, our empirical data suggests that proposition 3.2 holds regardless of assumption 3.1 and the restriction on $\tilde{B}$. Nevertheless, assumption 3.1 will generally be satisfied in practical scenarios; and the value of the BER above $\frac{1}{2}B(\eta - \sigma)$ is generally higher than our guaranteed BER by several orders of magnitude. We shall illustrate these issues with specific examples in chapter 4.

3.3 Worst Case Scenario for Jamming at Multiple Bits at Multiple Network Nodes

In the previous section, we considered scenarios in which a jammer can attack at multiple bits but at a single network node. In this section, we consider scenarios in which a jammer can attack at multiple bits and at multiple network nodes. The final goal is to find the worst case scenario associated with the smallest expected number of device alarms in an observation period.

Let $J_i^{(k)}$ denote the jamming signal magnitude (coherent component) at the ith bit at node k. Since degradation due to the attack propagates along the light path, the overall BER of the ith bit for an attack with $(J_i^{(1)},J_i^{(2)},...,J_i^{(M)})$ is $B(\sum_{k=1}^{M} J_i^{(k)})$, where M is the number of network nodes in the light path. The average BER in an observation period is given by

$$BER = \frac{1}{T}\sum_{i=1}^{T} B\left(\sum_{k=1}^{M} J_i^{(k)}\right). \qquad (3.19)$$

The expected device alarm rate for a given bit depends on how we want to measure the number of alarm generations in an observation period. One reasonable scheme is to consider, for any particular bit, that an alarm is generated if at least one of the network nodes in the light path generates a device alarm. Note that we use this scheme in the analysis in chapter 2. In this case, the expected device alarm rate in an observation period is given by

$$\text{Alarm rate (scheme 1)} = \frac{1}{T}\sum_{i=1}^{T}\left[1 - \prod_{k=1}^{M}\left(1 - A(J_i^{(k)})\right)\right]. \tag{3.20}$$

Another reasonable scheme is to sum up the number of alarm generations at each network node over an observation period, then sum up the number of alarm generations across all network nodes. The corresponding expected device alarm rate is given by

$$\text{Alarm rate (scheme 2)} = \frac{1}{T}\sum_{i=1}^{T}\sum_{k=1}^{M} A(J_i^{(k)}). \tag{3.21}$$

Searching for the worst case jamming attack scenario is therefore equivalent to solving the following minimization problem

$$\begin{aligned} \text{minimize} \quad & \frac{1}{T}\sum_{i=1}^{T}\left[1 - \prod_{k=1}^{M}\left(1 - A(J_i^{(k)})\right)\right] \text{ for scheme 1} \\ \text{or minimize} \quad & \frac{1}{T}\sum_{i=1}^{T}\sum_{k=1}^{M} A(J_i^{(k)}) \text{ for scheme 2} \\ \text{subject to} \quad & \tilde{B} = \frac{1}{T}\sum_{i=1}^{T} B\left(\sum_{k=1}^{M} J_i^{(k)}\right). \end{aligned} \tag{3.22}$$

We shall concentrate our analysis on scheme 1 and provide comments on scheme 2 in appendix B. We start by finding the worst case jamming attack scenario among jamming attacks at a single bit but at multiple network nodes.

Finally, the threshold value α depends on the value of M. In general, the larger M is, the smaller α will be since we made a pessimistic assumption that small coherent components of jamming signals can add up constructively across the light path and affect the BER significantly. We shall assume that the value of α in the case of M network nodes is a $1/M$ times the value of α in the case of one network node. Following our example, for the SNR = 20 dB, we have $\alpha = 0.15\sqrt{P}$ for $M = 1$ and $\alpha = 0.075\sqrt{P}$ for $M = 2$, and so on.

3.3.1 Worst Case Scenario for an Attack at a Single Bit at Two Network Nodes

In this section, we solve the minimization problem 3.22 when $T = 1$ and $M = 2$ for scheme 1. The solution to the similar problem for scheme 2 is given, for a specific example, in appendix B.

Consider the transmission of a single bit over the communication path containing two network nodes ($T = 1$,$M = 2$). Let $J^{(k)}$ denote the jamming signal (coherent component) at node k. We shall solve the following minimization problem

$$\begin{array}{ll} \text{minimize} & 1 - \left[1 - A(J^{(1)})\right]\left[1 - A(J^{(2)})\right] \text{ for scheme 1} \\ \text{subject to} & \hat{J} = J^{(1)} + J^{(2)}, \end{array} \qquad (3.23)$$

where $\hat{J} > 0$ and $B(\hat{J}) = \hat{B}$. Note that the constraint given in the problem is equivalent to the constraint $\hat{B}$=$B(J^{(1)} + J^{(2)})$ since $B(J)$ is a one-to-one function.

Note that we use the constant $\hat{B}$ to distinguish between the degraded BER at a particular bit due to accumulated jamming signals at M network nodes ($\hat{B}$ in this section) and the average degraded BER over an observation period of T bits ($\tilde{B}$ in section 3.2).

For convenience, let $C(J)$ denote the complementary expected device alarm rate function $1 - A(J)$. In terms of $C(J^{(i)})$, the objective function of the minimization problem can be rephrased as minimizing the product $-C(J^{(1)})C(J^{(2)})$. The Lagrangean and its derivatives with respect to $J^{(1)}$ and $J^{(2)}$ in this case are

$$\begin{aligned} L(J^{(1)}, J^{(2)}, \lambda) &= -C(J^{(1)})C(J^{(2)}) - \lambda(J^{(1)} + J^{(2)} - \hat{J}) && (3.24) \\ \frac{\partial}{\partial J^{(1)}}L = 0 &\Rightarrow C'(J^{(1)})C(J^{(2)}) = -\lambda && (3.25) \\ \frac{\partial}{\partial J^{(2)}}L = 0 &\Rightarrow C(J^{(1)})C'(J^{(2)}) = -\lambda && (3.26) \\ \frac{\partial}{\partial \lambda}L = 0 &\Rightarrow J^{(1)} + J^{(2)} = \hat{J}. && (3.27) \end{aligned}$$

Equations 3.25 and 3.26 yield the following relation

$$\frac{C'(J^{(1)})}{C(J^{(1)})} = \frac{C'(J^{(2)})}{C(J^{(2)})} = \text{constant}. \qquad (3.28)$$

Figure 3-8 shows the curve of $C'(J)/C(J)$. Notice that the function is one-to-one. Lemma 3.10 verifies this fact under a restriction on J.

Lemma 3.9 *Consider the function $A(J)$ as given in equation 3.5.*

i) There is a unique positive solution (denoted by J^*) to $A''(J) = 0$. Moreover, $J^* > \alpha$.

ii) $A'(J) > 0$ in $[0, \infty)$.

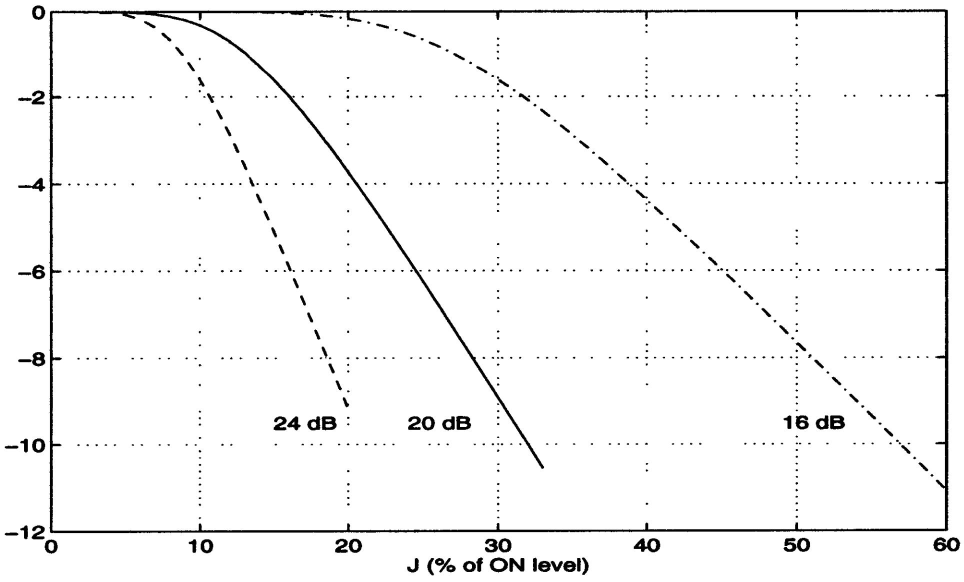


Figure 3-8: (Scheme 1) $C'(J)/C(J)$ curves for different values of the SNR. Note that the function is one-to-one.

Proof We use x as a dummy variable. Consider $-A''(x)$ as the slope of $-A'(x)$. The expression of $-A'(x)$ is given below

$$-A'(x) = \frac{1}{\sqrt{2\pi}\sigma}\left[e^{-\frac{(x+\alpha)^2}{2\sigma^2}} - e^{-\frac{(x-\alpha)^2}{2\sigma^2}}\right]. \tag{3.29}$$

Note that $-A'(x)$ has the same form as the function given in equation 3.9 of lemma 3.1. In particular, $-\alpha$ in equation 3.29 takes the place of α in equation 3.9, while α in equation 3.29 takes the place of η in equation 3.9. Therefore, various properties from lemma 3.1 hold for the function $-A'(x)$.

From lemma 3.1, we know that $-A'(x)$ has only one local maximum located before $-\alpha$ ($< -\alpha$) and one local minimum after α ($> \alpha$). We also know that $-A'(x)$ strictly increases (from the limit value of 0) in the interval from $-\infty$ to the local maximum, strictly decreases in the interval from the local maximum to the local minimum, and strictly increases (to the limit value of 0) in the interval from the local minimum to ∞.

i) It follows that in the interval from 0 to the local minimum, $-A'(x)$ strictly decreses and therefore $-A''(x) < 0$. At the local minimum, we have $-A''(x) = 0$. In the interval from the local minimum to ∞, $-A'(x)$ strictly increases and thus $-A''(x) > 0$. Therefore, we have $-A''(x) = 0$ only at the local minimum of $-A'(x)$. In addition, the location of this local minimum is in the positive direction of α ($> \alpha$).

ii) Note that $A'(0) = 0$ (from equation 3.12). We know that $-A'(x)$ decreases and is thus negative in $[0, J^*)$. In addition, $-A'(x)$ increases to the limit value of 0 in (J^*, ∞). Therefore, $-A'(x)$ is negative throughout the interval $(0, \infty)$, or equivalently $A(x) > 0$ in $[0, \infty)$. □

Lemma 3.10 *The function $C'(J)/C(J)$ strictly decreases in $[0, J^*)$, where J^* is the unique positive solution to $A''(J) = 0$.*

Proof We shall show that $C'(J)/C(J)$ strictly decreases or equivalently its derivative is negative in $[0, J^*)$. Consider the derivative of $C'(J)/C(J)$ given below

$$\begin{aligned}\frac{d}{dJ}\left(\frac{C'(J)}{C(J)}\right) &= \frac{C(J)C''(J) - (C'(J))^2}{(C(J))^2} \\ &= \frac{-(1 - A(J))A''(J) - (A'(J))^2}{(C(J))^2}. \qquad (3.30)\end{aligned}$$

The function on the right hand side of equation 3.30 is continuous since it is the derivative of a continuous function $C'(J)/C(J)$. Note that $1 - A(J) > 0$ for all J since $A(J)$ as given in 3.5 is the probability value and is thus bounded by 1. In addition, lemma 3.9 tells us that $A''(J) > 0$ in $[0, J^*)$. Applying these facts to the right hand side of equation 3.30, one can see that the derivative of $C'(J)/C(J)$ is negative in $[0, J^*)$. □

It follows that for any value of $\hat{J} < 2J^*$, an extremum of the objective function $-C(J^{(1)})C(J^{(2)})$ must have equal components, namely $(\hat{J}/2, \hat{J}/2)$, since there is a unique solution for each $J^{(i)}$ to equations 3.25 and 3.26 which are conditions for an extremum. It remains to investigate the nature of this extremum.

We now argue that $(\hat{J}/2, \hat{J}/2)$ is indeed the global minimum of the constraint set $J^{(1)} + J^{(2)} = \hat{J}$. It is sufficient to show that $-C(J^{(1)})C(J^{(2)})$ at another point in the constraint set is larger than $-(C(\hat{J}/2))^2$. Consider the point $(3\hat{J}/2, -\hat{J}/2)$, we have that

$$\begin{aligned}-C(3\hat{J}/2)C(-\hat{J}/2) &= -C(3\hat{J}/2)C(\hat{J}/2) \\ &> -(C(\hat{J}/2))^2,\end{aligned}$$

where the equality follows the fact that $A(J)$ is symmetric; and so is $C(J) = 1 - A(J)$. The inequality follows from the fact that $C(3\hat{J}/2) < C(\hat{J}/2)$ since $C(J)$ is a strictly decreasing function (its first derivative $-A'(J)$ is negative in $[0, \infty)$ as shown in lemma 3.9).

Note that at either end of the constraint set (when $(J^{(1)}, J^{(2)}) = (-\infty, \infty)$ or $(\infty, -\infty)$), the objective function approaches the maximal possible value, i.e. $-C(J^{(1)})C(J^{(2)})$ approaches zero. Therefore, we do not approach a global minimum in the limit as we approach either of these end points.

We conclude that the only extremum $(\hat{J}/2, \hat{J}/2)$ is indeed the global minimum. Notice that the condition $\hat{J} < 2J^*$ is equivalent to the condition $\hat{B} < B(2J^*)$. In general, the value of J^* $(> \alpha)$ is approximately close to α. For convenience in later analysis, we choose to restrict $\hat{B}$ to be below $B(2\alpha)$. Thus we have established the following proposition.

Proposition 3.3 *For scheme 1, given that $\hat{B} < B(2\alpha)$, an attack scenario in which the coherent components of jamming signals at the two network nodes are equal has the smallest expected device alarm rate and is therefore the worst case attack scenario.*

Finally, empirical data suggests that lemma 3.10 holds in the range $[0, \infty)$. Therefore, we can expect that proposition 3.3 holds regardless of the restriction on $\hat{B}$.

3.3.2 Worst Case Scenario for an Attack at a Single Bit at M Network Nodes

In this section, we extend our result from a scenario in which jamming can occur at two network nodes to the one in which jamming can occur at M $(M > 2)$ network nodes. We still consider jamming attacks which degrade the BER of a single transmitted bit. Moreover, we still adopt scheme 1 as the attack detection scheme. The corresponding minimization problem is of the form

$$\begin{aligned} \text{minimize} \quad & 1 - \prod_{k=1}^{M} \left[1 - A(J^{(k)})\right] \text{ for scheme 1} \\ \text{subject to} \quad & \hat{J} = \sum_{k=1}^{M} J^{(k)}. \end{aligned} \tag{3.31}$$

The solution to the similar problem for scheme 2 is given, for a specific example, in appendix B.

As before, let $C(J)$ denote the complementary expected device alarm rate $1 - A(J)$. We can rephrase the objective of problem 3.31 as to minimize the product $-\prod_{k=1}^{M} C(J^{(k)})$. The Lagrangean $L(J^{(1)}, ...J^{(M)}, \lambda)$ is equal to $-\prod_{k=1}^{M} C(J^{(k)}) - \lambda(\sum_{k=1}^{M} J^{(k)} - \hat{J})$. Differentiating the Lagrangean with respect to each $J^{(k)}$ yields

$$-\lambda = C'(J^{(1)}) \prod_{k \neq 1} C(J^{(k)}) = C'(J^{(2)}) \prod_{k \neq 2} C(J^{(k)}) = ... = C'(J^{(M)}) \prod_{k \neq M} C(J^{(k)}). \quad (3.32)$$

Multiplying through by $\left[\prod_{k=1}^{M} C(J^{(k)})\right]^{-1}$, we get the following relation,

$$\frac{C'(J^{(1)})}{C(J^{(1)})} = \frac{C'(J^{(2)})}{C(J^{(2)})} = ... = \frac{C'(J^{(M)})}{C(J^{(M)})},$$

which is the same condition required for an extremum in the two-node case (equation 3.28). Lemma 3.10 tells us that $C'(J)/C(J)$ is a strictly decreasing one-to-one function in $[0, J^*)$. Given that $\hat{J} < MJ^*$, or equivalently the jamming signal (coherent component) associated with equal jamming at all M network nodes does not exceed J^*, there exists a unique extremum at $(\hat{J}/M, ..., \hat{J}/M)$.

In the same fashion as in the case of two network nodes ($M = 2$), we can establish that $(\hat{J}/M, ..., \hat{J}/M)$ is indeed the global minimum (see the arguments given in section 3.3.1). For convenience in later analysis, we choose to restrict $\hat{B}$ to be below $B(M\alpha)$ instead of $B(MJ^*)$. Thus we have established the following proposition.

Proposition 3.4 *For scheme 1, given $\hat{B} < B(M\alpha)$, an attack scenario in which the coherent components of jamming signals at M network nodes are equal has the smallest expected device alarm rate and is therefore the worst case attack scenario.*

Empirical data suggests that proposition 3.4 holds regardless of the restriction on $\hat{B}$. In appendix B, we show, for a specific example, that the similar conclusion can be made for scheme 2 although the restriction on $\hat{B}$ is tighter in the case of scheme 2.

3.3.3 Worst Case Scenario for an attack at T Bits at M Network Nodes

This section pursues the final objective of the chapter. We find the worst case attack scenario when an attacker can jam at multiple bits and at multiple network nodes. To do so, we solving the previously constructed minimization problem (problem 3.22) which is presented again below,

$$\begin{aligned}
\text{minimize} \quad & -\frac{1}{T}\sum_{i=1}^{T}\prod_{k=1}^{M} C(J_i^{(k)}) \text{ for scheme 1} \\
\text{or minimize} \quad & \frac{1}{T}\sum_{i=1}^{T}\sum_{k=1}^{M} A(J_i^{(k)}) \text{ for scheme 2} \\
\text{subject to} \quad & \tilde{B} = \frac{1}{T}\sum_{i=1}^{T} B\left(\sum_{k=1}^{M} J_i^{(k)}\right)
\end{aligned} \tag{3.33}$$

As before, $C(J)$ is the complementary alarm rate function which is equal to $1 - A(J)$. In any optimal solution for scheme 1, we must have, for each i, the minimum value of $-\prod_{k=1}^{M} C(J_i^{(k)})$ subject the constraint on $\sum_{k=1}^{M} J_i^{(k)}$. Otherwise, we can reassign the values of $J_i^{(k)}$ such that their sum remains the same but $-\prod_{k=1}^{M} C(J_i^{(k)})$ is smaller. The same arguments can be applied for scheme 2.

Let J_i denote $\sum_{k=1}^{M} J_i^{(k)}$. In addition, let $A_M(J)$ denote the minimal expected device alarm rate for a single bit for an attack at M network nodes with jamming signals whose coherent components sum up to J. Similarly, let $C_M(J)$ denote the corresponding maximal complementary expected device alarm rate for a single bit. Using these functions, we can rewrite our minimization problem as

$$\begin{aligned}
\text{minimize} \quad & -\frac{1}{T}\sum_{i=1}^{T} C_M(J_i) \text{ for scheme 1} \\
\text{or minimize} \quad & \frac{1}{T}\sum_{i=1}^{T} A_M(J_i) \text{ for scheme 2} \\
\text{subject to} \quad & \tilde{B} = \frac{1}{T}\sum_{i=1}^{T} B(J_i).
\end{aligned} \tag{3.34}$$

Using the functions $A_M(J)$ and $C_M(J)$, we can treat the problem as if we were to find the worst case attack scenario for jamming at multiple bits but at a single network node. The similar problem is solved in section 3.2. The only difference is the expected device alarm rate functions $A_M(J)$ and $-C_M(J)$ in this section instead of $A(J)$ in section 3.2.

Consider the function $C_M(J)$. For $J < MJ^*$, proposition 3.4 tells us that $C_M(J) = (C(J/M))^M$ since the worst case scenario corresponds to the event in which coherent components of jamming signals at M network nodes are equal. Therefore, we can express $C'_M(J)$ and $C''_M(J)$ when $M > 1$ as

$$\frac{d}{dJ}C_M(J) = (C(J/M))^{M-1}C'(J/M) \tag{3.35}$$

$$\frac{d^2}{dJ^2}C_M(J) = \frac{M-1}{M}(C(J/M))^{M-2}(C'(J/M))^2 + \frac{1}{M}(C(J/M))^{M-1}C''(J/M) \tag{3.36}$$

Note that since $C(J) = 1 - A(J)$, we have $C'(J) = -A'(J)$ and $C''(J) = -A''(J)$.

In what follows, we concentrate on finding the solution to problem 3.33 for scheme 1. The solution to the similar problem for scheme 2 is given, for a specific example, in appendix B.

We shall start solving problem 3.33 in the case when $T = 2$ and then extend the result to a general case when $T > 2$. For $T = 2$, the Lagrangean and its derivatives are

$$L(J_1, J_2, \lambda) = -\frac{1}{2}[C_M(J_1) + C_M(J_2)] - \lambda\left(\frac{1}{2}[B(J_1) + B(J_2)] - \tilde{B}\right) \tag{3.37}$$

$$\frac{\partial}{\partial J_i}L = 0 = C'_M(J_i) - \lambda B'(J_i), i = 1, 2 \tag{3.38}$$

$$\frac{\partial}{\partial \lambda} = 0 = \frac{1}{2}B(J_1) + \frac{1}{2}B(J_1) - \tilde{B} \tag{3.39}$$

Figure 3-9 shows the curves of $-C'_M(J_i)$ for a few values of M superposed on the curve of $B'(J_i)$. Since $C'_M(0) = 0$ (equation 3.35) and $B'(0) = 0$ (equation 3.14), 0 is always a solution to $C'_M(J_i) = \lambda B'(J_i)$. It follows that there exist extrema of the forms $(\tilde{J}, \tilde{J})$ and $(0, J^s)$ (or $(J^s, 0)$). In any case, an extremum must satisfy conditions 3.38 and 3.39.

The bordered Hessian of the extremum $(\tilde{J}, \tilde{J})$ can be expressed as

$$|H|_{(\tilde{J},\tilde{J})} = \left[C''_M(\tilde{J}) - \frac{C'_M(\tilde{J})}{B'(\tilde{J})}B''(\tilde{J})\right](B'(\tilde{J}))^2. \tag{3.40}$$

The bordered Hessian of the extrema $(J^s, 0)$ and $(0, J^s)$ can be expressed as

$$|H|_{(J^s,0)} = \frac{1}{2}\left[C''_M(0) - \frac{C'_M(J^s)}{B'(J^s)}B''(0)\right](B'(J^s))^2. \tag{3.41}$$

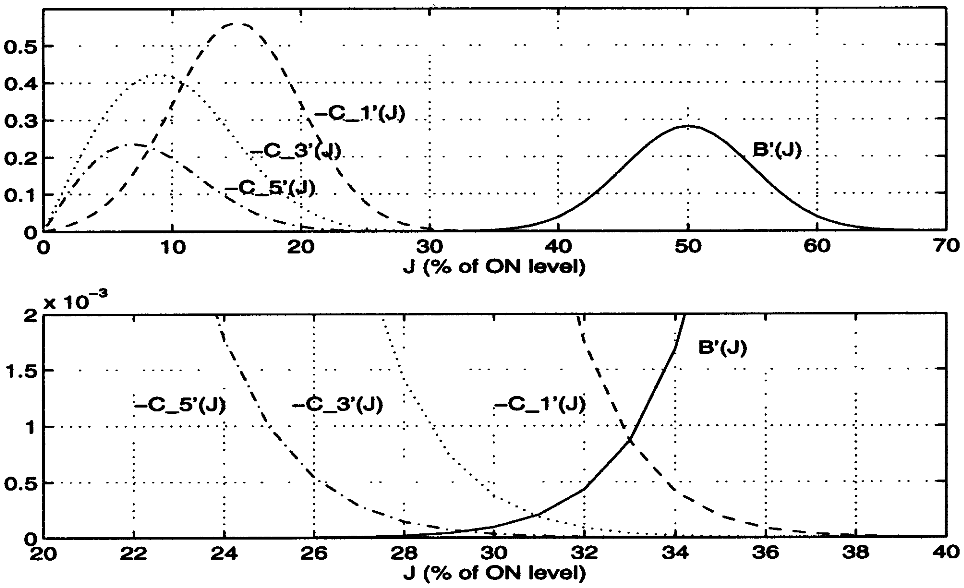


Figure 3-9: (Scheme 1) The curve of $B'(J)$ superposed on the curves of $-C'_M(J)$ for a few values of M. In the range $(0, \eta - \sqrt{M}\sigma)$, there exists at most one positive intersection point ($\eta - \sqrt{M}\sigma$ is $0.45\sqrt{P}$ in this case). The SNR is 20 dB, and α is $0.15\sqrt{P}$ (for $M = 1$).

In what follows, we shall proceed with a specific example. The conclusion in a general case is the subject of future research.

Example 3.1

We shall continue with our example case: SNR = 20 dB, $\alpha = 0.15\sqrt{P}$ ($M = 1$). Figure 3-9 shows the curves of $-C'_M(J_i)$ for a few values of M superposed on the curve of $B'(J_i)$.

We have argued that zero is a solution to equation 3.38; and there exist extrema of the forms $(\tilde{J}, \tilde{J})$ and $(0, J^s)$ (or $(J^s, 0)$).

Consider the range $(0, \eta - \sigma)$, for this example, there exists at most one intersection in $(0, \eta - \sqrt{M}\sigma)$ (we can consider a larger interval but $(0, \eta - \sqrt{M}\sigma)$ is large enough for our purpose and is consistent with the result in section 3.2). Thus, given that $\tilde{B} < \frac{1}{2}B(\eta - \sqrt{M}\sigma) + \frac{1}{2}B(0)$, any extremum must correspond to either constant jamming $(\tilde{J}, \tilde{J})$ or sporadic jamming $(0, J^s)$ (or $(J^s, 0)$).

Figure 3-10 shows the curves of $|H|_{(\tilde{J},\tilde{J})}$ in equation 3.40 for different values of M. Figure 3-11 shows the curves of $|H|_{(J^s,0)}$ in equation 3.41 for different values of M. In all cases, $|H|_{(\tilde{J},\tilde{J})} > 0$ and $|H|_{(J^s,0)} < 0$ in $(0, \eta - \sqrt{M}\sigma)$. Thus constant jamming corresponds

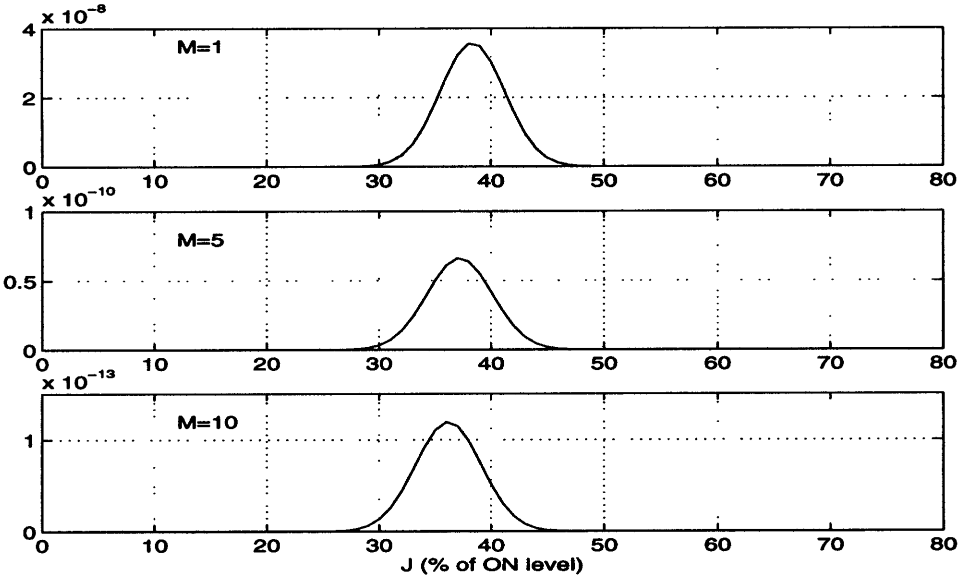


Figure 3-10: (Scheme 1) $|H|_{(\tilde{J},\tilde{J})}$ as a function of $\tilde{J}$. In all cases, $|H|_{(\tilde{J},\tilde{J})} > 0$ in $(0, \eta - \sqrt{M}\sigma)$. The SNR is 20 dB, and α is $0.15\sqrt{P}$ (for $M = 1$).

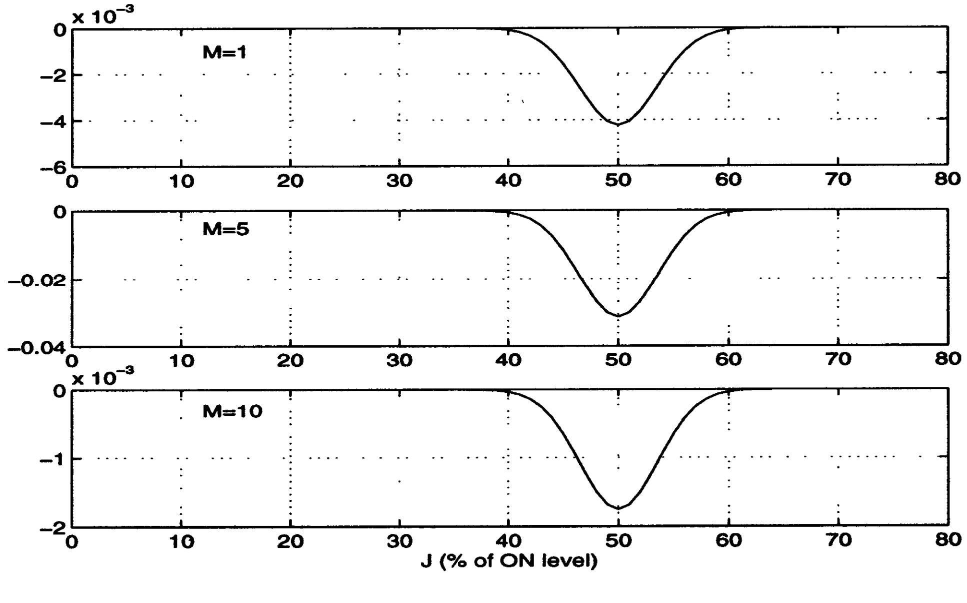


Figure 3-11: (Scheme 1) $|H|_{(J^s,0)}$ as a function of J^s. In all cases, $|H|_{(J^s,0)} < 0$ in $(0, \eta - \sqrt{M}\sigma)$. The SNR is 20 dB, and α is $0.15\sqrt{P}$ (for $M = 1$).

to the local maximum while sporadic jamming corresponds to the local minimum.

We have argued in section 3.2.1 that the constraint set defined by equation 3.39 (which is the same as equation 3.8 in section 3.2.1) has no end points (no boundary). Since there is no other extremum in this constraint set, it follows that sporadic jamming with $(J^s, 0)$ or $(0, J^s)$ yields the smallest expected device alarm rate and thus corresponds to the worst case attack scenario.

We can extend the result to the case when $T > 2$ using precisely the same arguments given in the proof of proposition 3.2 (see section 3.2.2). Therefore, given that $C_M(J_i) = (C(J_i/M))^M$, as long as $\tilde{B} < \frac{1}{T}B(\eta - \sqrt{M}\sigma) + \frac{T-1}{T}B(0)$, the worst case attack scenario corresponds to the event in which a jammer attacks at a single bit, and coherent components of jamming signals at M network nodes at that bit are equal. For convenience, we choose to restrict $\tilde{B}$ to be below $\frac{1}{T}B(\eta - \sqrt{M}\sigma)$ since the value of $B(0)$ is generally very small.

Finally, to guarantee that our complementary device alarm rate function is indeed given by $C_M(J_i) = (C(J_i/M))^M$, we must have $J_i < M\alpha$ (proposition 3.4). We thus arrive at the following conclusion.

In our example, given $\tilde{B} < \frac{1}{T}B(\xi)$, where $\xi = \min(M\alpha, \eta - \sqrt{M}\sigma)$, the worst case attack scenario corresponds to the event in which a jammer attacks at a single bit, and coherent components of jamming signals at M network nodes at that bit are equal.

Although we have found the worst case attack scenario only for a specific example, the procedures involved in example 3.1 are so general that we expect the conclusion to hold in other practical scenarios to be considered in chapter 4.

Finally, in appendix B, we show that the similar conclusion can be obtained, for a specific example, in the case of scheme 2. However, the restriction on the value of $\tilde{B}$ is tighter in the case of scheme 2 than in the case of scheme 1.

3.4 Summary

In this chapter, we searched for the worst case in-band jamming attack scenario for a given value of degraded BER ($\tilde{B}$). In particular, the worst case attack scenario has the smallest expected numbers of device alarms in a given observation period. We summarize the results in what follows.

Restrictions on the Degraded BER

In each scenario in the analysis, we have found the worst case attack scenario given that the degraded BER $\tilde{B}$ does not exceed a certain value. The restrictions on $\tilde{B}$ depend on the following parameters.

- M: the number of network nodes in the light path.
- σ^2: the variance of a real part and an imaginary part of the AWGN at any bit time at each network node.
- $\eta = \sqrt{P}/2$: where P is the ON level signal power of OOK.
- α: threshold of the decision rule for attack detection devices at network nodes.

Assumption on the Attack Detection Device Threshold

We have made assumption 3.1 on the attack detection device threshold α. In particular, we assume $\sigma < \alpha < \eta - \sigma$.

Note that assumption 3.1 is generally satisfied when the SNR is sufficiently high since the value of η and σ will differ significantly.

Worst Case In-band Jamming Attack Scenarios

When a jammer can attack at 1 network node and at T bits, we have shown in a general case under assumption 3.1 that, given $\tilde{B} < \frac{1}{T}\ B(\eta-\sigma)$, the worst case scenario corresponds to the event in which jamming occurs at one out of T bits.

When a jammer can attack at M network nodes and at 1 bit, we have shown (for scheme 1) in a general case under assumption 3.1 that, given $\tilde{B} < B(M\alpha)$, the worst case scenario corresponds to the event in which coherent components of jamming signals at M network nodes are equal.

When a jammer can attack at M network nodes and at T bits, we have shown for example 3.1 under assumption 3.1 that, given $\tilde{B} < \frac{1}{T}B(\xi)$ where $\xi = \min(M\alpha, \eta - \sqrt{M}\sigma)$, the worst case attack scenario corresponds to the event in which jamming occurs at a single bit, and coherent components of jamming signals at M network nodes at that bit are equal.

We expect the conclusion in example 3.1 to hold for other practical scenarios to be considered in chapter 4.

Chapter 4

Example Attack Detection Systems Against In-Band Jamming Attacks

Having found the worst-case in-band jamming attack scenarios in chapter 3, we can investigate the performance of our attack detection scheme under these scenarios.

We start the chapter with a discussion on the calcution of FP_{attack} and FN_{attack}, the false positive and the false negative probabilities of an attack alarm. We then present a few examples of attack detection systems together with a suggestion of how to modify our attack detection scheme to improve its performance.

4.1 Calculation of FP_{attack} and FN_{attack}

In chapter 2, we assume that coherent components of jamming signals at M network nodes are equal and are constant throughout an observation period of T bits. In addition, we assume that an alarm is generated at any bit time if at least one of the network nodes in the light path generates a device alarm (scheme 1). These assumptions enable us to think of an alarm generation in a given bit time as a Bernoulli random variable, and the number of device alarms in a given observation period as a binomial random variable.

Under the worst case in-band jamming attack scenarios found in chapter 3, coherent components of jamming signals at M network nodes are equal but are not constant throughout an observation period. In this case, we can no longer think of the number of device alarms as a binomial random variable. In what follows, we shall provide exact expressions for FP_{attack} and FN_{attack}.

4.1.1 FP_{attack} and FN_{attack}: Scheme 1

As a reminder, in scheme 1, we consider that a device alarm is generated at a given bit time if at least one of the network nodes in the light path generates a device alarm. Therefore, the total number of device alarms in an observation period (denoted by X) is at most T. In addition, X is the sum of T Bernoulli random variables whose probability of device alarm generation p is given by

$$p = \begin{cases} FP_{\text{end-to-end}} = 1 - (1 - A(0))^M, & \text{no attack} \\ 1 - FN_{\text{end-to-end}} = 1 - (1 - A(J))^M, & \text{attack with } J \text{ at each node} \end{cases}, \tag{4.1}$$

where $A(J)$ is defined in equation 3.5. For convenience, let p_o denote the probability of a device alarm generation given no attack, and p_J denote the probability of a device alarm generation given a coherent component of the jamming signal at each of the M network nodes equal to J.

Under no attack, we can express FP_{attack} as

$$\begin{aligned} FP_{\text{attack}} &= Pr\{X \geq \beta \mid \text{no attack}\} \\ &= \sum_{i=\beta}^{T} \binom{T}{i} p_o^i (1 - p_o)^{T-i}, \end{aligned} \tag{4.2}$$

where β is the threshold value for an attack alarm. We shall assume throughout the chapter that β is an integer greater than or equal to 1.

Under the worst case attack scenario, coherent components of jamming signals are equal at M network nodes but jamming occurs at one out of T bits. Let the jamming signal (coherent component) at each network node be J. It follows that X is the sum of two types of Bernoulli random variables: one Bernoulli random variables with probability of device alarm generation p_J, and $(T - 1)$ Bernoulli random variables with probability of device alarm generation p_o.

Let X_J denote the value of the Bernoulli random variable associated with p_J, and X_o

denote the sum of $(T-1)$ Bernoulli random variables associated with p_o. It is easy to see that $X = X_J + X_o$. We can calculate FN_{attack} in the following fashion

$$\begin{aligned} FN_{\text{attack}}^{1\text{ bit}} &= Pr\left\{X \leq \beta - 1 \mid \text{attack}\right\} \\ &= \sum_{i=0}^{1} Pr\left\{X_J = i\right\} Pr\left\{X_o \leq \beta - 1 - i\right\} \\ &= \sum_{i=0}^{1} \binom{1}{i} p_J^i (1-p_J)^{1-i} Pr\left\{X_o \leq \beta - 1 - i\right\} \\ &= \sum_{i=0}^{1} \binom{1}{i} p_J^i (1-p_J)^{1-i} \sum_{j=0}^{\beta-1} \binom{T-1}{j} p_o^j (1-p_o)^{T-1-j}. \qquad (4.3) \end{aligned}$$

Similarly, given that coherent components of jamming signals are equal at M network nodes and are equal in a period of t bits $(t < T)$. we have that

$$FN_{\text{attack}}^{t\text{ bits}} = \sum_{i=0}^{t} \binom{t}{i} p_J^i (1-p_J)^{t-i} \sum_{j=0}^{\beta-1} \binom{T-t}{j} p_o^j (1-p_o)^{T-t-j}. \qquad (4.4)$$

4.1.2 FP_{attack} and FN_{attack}: Scheme 2

As a reminder, in scheme 2, we count the number of device alarms in an observation period by summing the number of device alarm generations at each network node over an observation period of T bits, and then summing the number of device alarm generations across all M network nodes.

In this case, the total number of device alarms in an observation period of T bits (denoted by X) is at most TM. In addition, X is the sum of TM Bernoulli random variables whose probability of device alarm generation p is given by

$$p = \begin{cases} FP_{\text{node}} = A(0), & \text{no attack} \\ 1 - FN_{\text{node}} = A(J), & \text{attack with } J \text{ at each node} \end{cases}. \qquad (4.5)$$

For convenience, let p_o denote the probability of an alarm generation given no attack, and p_J denote the probability of an alarm generation given a coherent component of the jamming signal at each of the M network nodes equal to J.

Using the same procedures as in the case of scheme 1, we can calculate the values of

FP_{attack} and FN_{attack} in the following fashion

$$FP_{\text{attack}} = \sum_{i=\beta}^{TM} \binom{TM}{i} p_o^i (1-p_o)^{TM-i} \tag{4.6}$$

$$FN_{\text{attack}}^{1\text{ bit}} = \sum_{i=0}^{M} \binom{M}{i} p_J^i (1-p_J)^{M-i} \sum_{j=0}^{\beta-1} \binom{M(T-1)}{j} p_o^j (1-p_o)^{M(T-1)-j} \tag{4.7}$$

$$FN_{\text{attack}}^{t\text{ bits}} = \sum_{i=0}^{tM} \binom{tM}{i} p_J^i (1-p_J)^{tM-i} \sum_{j=0}^{\beta-1} \binom{M(T-t)}{j} p_o^j (1-p_o)^{M(T-t)-j}. \tag{4.8}$$

4.2 Requirements on the Attack Detection Scheme Performance

The performance of our proposed attack detection scheme depends on the following parameters.

- SNR: The signal-to-noise ratio of the communication path of interest.
- M: The number of network nodes in the communication path.
- T: The number of bits observed, or an observation period of the attack detection scheme.
- $\tilde{B}$: The maximal tolerable BER.
- α: The threshold value for device alarm generation at an attack detection device at each network node (first-level alarm).
- β: The threshold value for attack alarm generation (second-level alarm).
- $\overline{FP}_{\text{attack}}$: Upper limit on the probability of concluding that there is an attack yielding the BER above $\tilde{B}$ when there is none.
- $\overline{FN}_{\text{attack}}$: Upper limit on the probability of concluding that there is no attack yielding the BER above $\tilde{B}$ when there is one.

The design parameters are dependent on one another. Therefore, we shall specify some parameters and find the corresponding values of the others. In particular, we shall assume that the values of SNR and M are given. Moreover, we are given the requirements on the attack detection scheme performance in terms of $\tilde{B}$, $\overline{FP}_{\text{attack}}$, and $\overline{FN}_{\text{attack}}$. We then find the appropriate values for α, β, and T such that the constructed attack detection system satisfies all the requirements, or conclude that it is not possible to construct such an attack detection scheme.

4.3 Examples of Attack Detection Schemes

We present in this section examples of attack detection schemes against in-band jamming attacks. We concentrate on scheme 1 of the attack detection scheme. Comments on scheme 2 of the attack detection scheme can be found in appendix B.

Example 4.1

Assume the following parameters: SNR = 24 dB, M=1, $\tilde{B} = 10^{-10}$, $\overline{FP}_{\text{attack}} = 10^{-8}$, and $\overline{FN}_{\text{attack}} = 10^{-8}$.

For SNR = 24 dB, assumption 3.1 is satisfied given that α is between $0.0315\sqrt{P}$ and $0.4685\sqrt{P}$ (where P is the ON level signal power of OOK). Proposition 3.2 tells us that in the worst case attack scenario jamming occurs at only one out of T bits in an observation period.

By trial and error, an appropriate value of T is 10^4. The corresponding threshold values are $\alpha = 0.22\sqrt{P}$, and $\beta = 1$.

We calculate FP_{attack} in the following fashion

$$\begin{aligned} FP_{\text{attack}} &= 1 - Pr\{\text{no alarm}\} \\ &= 1 - (1 - p_o)^{10{,}000} \\ &\approx 3.09 \times 10^{-8}, \end{aligned}$$

where p_o is given in equation 4.1.

On the other hand, we can calculate FN_{attack} as follows

$$\begin{aligned} FN_{\text{attack}} &= Pr\{\text{no alarm}\} \\ &= (1 - p_o)^{9{,}999} \times (1 - p_J) \\ &\approx 3.55 \times 10^{-8}, \end{aligned}$$

where p_o and p_J are given in equation 4.1.

Therefore, by setting the parameters $\alpha = 0.22\sqrt{P}$, $\beta = 1$, and $T = 10^4$, we can construct an in-band jamming attack detection system that generates an attack alarm whenever the average BER in an observation period exceeds 10^{-10}.

Example 4.2

Assume the following parameters: SNR = 24 dB, M=10, $\tilde{B} = 10^{-8}$, $\overline{FP}_{\text{attack}} = 10^{-8}$, and $\overline{FN}_{\text{attack}} = 10^{-8}$. Notice that except for M, all the other parameters have the same values as in example 4.1.

It turns out that we cannot construct an attack detection scheme that can handle the worst case attack scenario in this case. Consider the following results under the worst case attack scenarios. Note that FP_{attack} and FN_{attack} are calculated in the similar manner as in example 4.1.

For $T = 1{,}000$ ($\alpha = 0.07\sqrt{P}$, $\beta = 1$), we have $FP_{\text{attack}} \approx 2.27\times10^{-8}$. The corresponding FN_{attack} is ≈ 0.995.

For $T = 10^4$ ($\alpha = 0.075\sqrt{P}$, $\beta = 1$), we have $FP_{\text{attack}} \approx 2.53 \times 10^{-9}$. The corresponding FN_{attack} is ≈ 0.995.

For $T = 10^5$ ($\alpha = 0.075\sqrt{P}$, $\beta = 1$), we have $FP_{\text{attack}} \approx 2.58 \times 10^{-8}$. The corresponding FN_{attack} is ≈ 0.9974.

This example shows that varying the value of T hardly affects the performance of the attack detection system under the worst case scenarios. We conclude that an attack detection system satisfying the given restrictions cannot be constructed.

In the next example, we demonstrate that an attack detection scheme can be constructed in the case of $M = 10$ if we are willing to use a higher SNR and relax the performance requirement on $\tilde{B}$.

Example 4.3

Assume the following parameters: SNR = 28 dB, M=10, $\tilde{B} = 10^{-4}$, $\overline{FP}_{\text{attack}} = 10^{-8}$, and $\overline{FN}_{\text{attack}} = 10^{-8}$. Notice that except for the SNR and $\tilde{B}$, all the other parameters have the same values as in example 4.2.

In this case, an appropriate value of T is 1,000. The corresponding threshold values are $\alpha = 0.043\sqrt{P}$, and $\beta = 1$.

Using the same method as in example 4.1. We have $FP_{\text{attack}} \approx 8.42 \times 10^{-8}$ and $FN_{\text{attack}} \approx 5.18 \times 10^{-8}$.

Therefore, by setting the parameters $\alpha = 0.043\sqrt{P}$, $\beta = 1$, and $T = 1,000$, we can construct an in-band jamming attack detection scheme that generates an attack alarm whenever the BER exceeds 10^{-4}.

As a final note, in this particular example, $\tilde{B}$ is so high that the restriction $\tilde{B} < \frac{1}{T}B(\xi)$ as given in example 3.1 is not satisfied. Nevertheless, we have verified in this case that $C'(J)/C(J)$ is one-to-one beyond α and that the restriction on $\tilde{B}$ (such that lemma 3.10 and proposition 3.4 hold) can be relaxed. Accordingly, the worst case scenario still corresponds to the event in which coherent components of jamming signals are equal at 10 network nodes but jamming occurs at only one bit.

Note that we have to increase both the SNR and $\tilde{B}$ so that our attack detection scheme can handle worst case attack scenarios. However, these modifications are a high price to pay. In the next section, we propose an alternative method of modifying our attack detection scheme to handle worst case in-band jamming attack scenarios using hard limiters at the input to network nodes.

4.4 Modification to the Attack Detection Scheme Using Hard Limiters at the Input to Network Nodes

As a reminder, this work concentrates on in-band jamming attacks which are performed in the following fashion. An attacker gains access to an adjacent input link to a network node shared by a legitimate light path. Using crosstalk at network node components, an attacker can insert in-band jamming signals to a legitimate light path. To insert jamming signals with large magnitudes, an attacker must use signals with very high power as inputs to the network node.

Under the worst case attack scenarios, an attacker performs sporadic jamming by inserting jamming signals at only a fraction of transmitted bits. To degrade the BER by sporadic jamming, the jammer needs to insert jamming signals of large magnitudes, or equivalently input signals with very high instantaneous power to the network node.

A "hard limiter" is a device that limits the power of signals passing through it. In the presence of hard limiters, an attacker cannot degrade the BER by inserting input signals with high instantaneous power to affect a small fraction of bits since the excess power will be cut off.

Note that jamming signals have relatively high power compared to the signal power under normal operation; therefore, the presence of hard limiters will not affect normal operation of our communication system.

Provided that the jammer attacks at a sufficiently large number of bits, the worst case attack scenario corresponds to the event in which coherent components of jamming signals at M network nodes are equal but jamming occurs at only $t(t > 1)$ out of T bits. We shall refer to this worst case attack scenario as scenario t in what follows.

When the value of t is sufficiently high, our attack detection scheme will be able to distinguish between the presence and the absence of an attack. Finally, finding the cut-off power level of a hard limiter is equivalent to finding a value of t^* such that scenario $t > t^*$ will be detected by our attack detection scheme. We shall demonstrate this with an example.

Example 4.4

We shall continue with the scenario in example 4.2, namely SNR $=$ 24 dB, M=10, $\tilde{B} = 10^{-8}$, $\overline{FP}_{\text{attack}} = 10^{-8}$, and $\overline{FN}_{\text{attack}} = 10^{-8}$.

Appropriate values are $T = 1,000$, $\alpha = 0.04\sqrt{P}$, $\beta = 9$, and $t^* = 15$. Under scenario t^*, we can calculate FP_{attack} using equation 4.2 and FN_{attack} using equation 4.4. In this case, $FP_{\text{attack}} \approx 1.76 \times 10^{-8}$, and $FN_{\text{attack}} \approx 4.56 \times 10^{-9}$.

Let J_t^s denote the coherent component of jamming signals at each network node at any bit time in scenario t. We claim that all attack scenarios associated with lower expected device alarm rates must have at a jamming signal (accumulated coherent components across 10 nodes) greater than $J_{t^*}^s$ at least at one bit time.

To justify the claim, imagine trying to decrease the expected device alarm rate by modifying scenario t^*. To decrease the expected device alarm rate, we need an attack scenario in which jamming is more sporadic in nature (a consequence of the results from chapter 3). In order to achieve the same BER by jamming more sporadically, we need both weaker and stronger jamming signals.

We conclude that the attack detection system constructed in this example (with $T = 1,000$, $\alpha = 0.04\sqrt{P}$, and $\beta = 9$) can detect all jamming attacks given that the jamming signal (accumulated coherent components across 10 nodes) at each bit time does not exceed $J^s_{t^*}$, which is $\approx 0.354\sqrt{P}$.

Based on $J^s_{t^*}$ and the knowledge of component crosstalk levels, we can accordingly assign a cut-off power level for hard limiters at network nodes. For example, if the crosstalk level is at -20 dB, in this case, we can set the cut-off power level at the input to each of the 10 network nodes to be $\approx 3.54\sqrt{P}$.

Finally, to make sure that we have the same conclusion for the worst case attack scenarios as in example 3.1, we need to have $\tilde{B}$ less than $\frac{1}{T}B(\xi)$ (see example 3.1). In this case $\xi = M\alpha = 0.40\sqrt{P}$ and $\frac{1}{T}B(\xi) \approx 3.8 \times 10^{-7}$. So the restriction on $\tilde{B}$ is satisfied.

4.5 Summary

In this chapter, we present a few examples of attack detection systems. In addition, we have also given an example in which our proposed attack detection scheme fails to detect the worst case jamming attack scenario. With a modification to the attack detection scheme using hard limiters at the input to network nodes, we are able to construct an attack detection scheme that can handle the worst case jamming attack scenarios, and thus all jamming attack scenarios.

Chapter 5

Summary and Directions for Future Research

5.1 Summary of Results

In this section, we provide an overview of what we accomplished in this work. The detailed explanations are provided in the summaries at the end of all chapters.

We proposed a novel attack detection scheme which is able to detect a wider variety of jamming attacks than any of the existing methods described in chapter 1. We have shown in chapter 2 that under some rather pessimistic jamming attack scenarios, our proposed attack detection scheme performs satisfactorily both in the case of in-band jamming and in the case of out-of-band jamming causing gain competition at EDFAs.

To further investigate the performance of our approach for detecting in-band jamming attacks, we searched among all jamming attack scenarios yielding the same degraded level of BER for the worst case attack scenario, which corresponds to the lowest expected number of device alarms in an observation period, and consequently the smallest probability of being detected. The worst case in-band jamming attack scenarios are the subject of chapter 3. We found the worst case attack scenarios under some restrictions which are generally satisfied in practical scenarios some of which are considered in chapter 4.

In chapter 4, we analyzed the performance of our attack detection scheme under the worst case in-band jamming attack scenarios found in chapter 3. In particular, we presented an example such that the proposed attack detection scheme constructed up to that point could not detect the worst case attack scenarios. We then proposed a modification to the attack detection scheme using hard limiters at the input to network nodes. Finally, we showed that the modified detection scheme can detect the worst case jamming attack scenarios, and thus all jamming attack scenarios.

5.2 Applicability of the Attack Detection Scheme

Notice that our attack detection scheme does not distinguish between the sources of interfering signals which corrupt transmitted signals. In particular, the attack detection device at the network node generates an alarm whenever the output of the detection device is greater than a certain threshold. The cause of an alarm may in fact be the signal inserted to the network node by the jammer, or simply the noise which happens to be at a significantly high level. Therefore, to be more precise, it may be more appropriate to view an "attack alarm" as an alarm notifying an excessive level of degraded BER. However, to keep the motivation clear, and to avoid some ambiguities which may arise in some parts of this work, we decided to keep the notion of an "attack alarm".

As a result, our proposed attack detection scheme can be applied to satisfy a more general purpose of maintaining a certain level of BER, which is a general performance metric, in a communication path.

5.3 Directions for Future Research

While we performed an analysis on the effects of in-band jamming attacks and searched for the worst case in-band jamming attack scenarios, a parallel analysis is yet to be done in the case of out-of-band jamming attacks causing gain competitions at EDFAs at network nodes.

We obtained empirical data which suggests the validity of various theoretical results in chapter 3 without restrictions and assumptions that we adopted for the purpose of the proofs. It will be useful to have alternative proofs of theoretical results in chapter 3 which do not rely on the assumptions that we made.

Appendix A

Proofs of Various Lemmas

Lemma 3.1 *Given that $\lambda > 0$, $\eta > \alpha$, consider the function $f(x)$ of the form*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\alpha)^2}{2\sigma^2}} - \lambda \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\eta)^2}{2\sigma^2}}.$$

$f(x)$ has the following characteristics:

i) There is a unique solution to $f(x) = 0$.

ii) $f(x)$ has a unique local maximum and a unique local minimum.

iii) Let x^{max} and x^{min} denote the locations of the maximum and the minimum respectively. We have $x^{max} < \alpha < \eta < x^{min}$.

iv) As x increases from $-\infty$ to x^{max}, $f(x)$ strictly increases from the limit value of 0 to $f(x^{max})$ and has a single saddle point where $f(x)$ changes from convexity ($\smile$) to concavity ($\frown$).

v) As x increases from x^{max} to x^{min}, $f(x)$ strictly decreases.

vi) As x increases from x^{min} to ∞, $f(x)$ strictly increases from $f(x^{min})$ to the limit value of 0 and has a single saddle point at which $f(x)$ changes from convexity ($\smile$) to concavity ($\frown$).

Proof Figure A-1 shows some example curves of $f(x)$.

i) To show that $f(x) = 0$ has a unique solution, we can simplify the equation $f(x) = 0$ to the following form

$$-(x-\alpha)^2 = -(x-\eta)^2 + 2\sigma^2 \ln \lambda,$$

which we can solve to obtain a unique answer $x^o = \frac{\eta^2 - \alpha^2 - 2\sigma^2 \ln \lambda}{2(\eta - \alpha)}$.

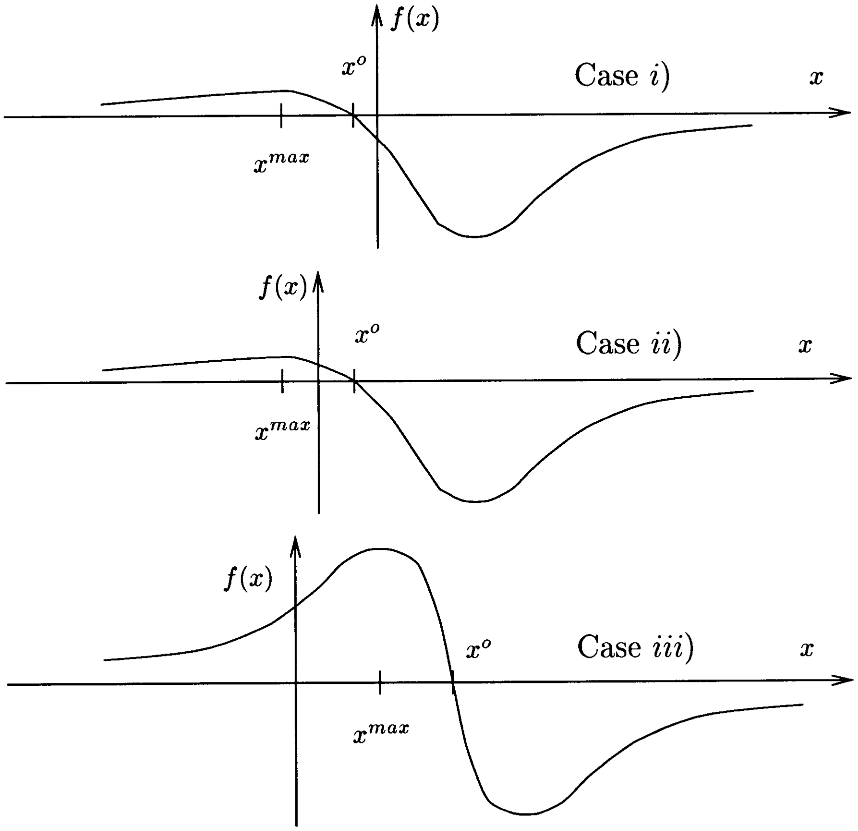


Figure A-1: Illustrations of $f(x)$ in three different cases for the proof of lemma 3.3.

ii) Consider the first derivative of $f(x)$ given below

$$f'(x) = -\frac{x-\alpha}{\sqrt{2\pi}\sigma^3}e^{-\frac{(x-\alpha)^2}{2\sigma^2}} + \lambda\frac{x-\eta}{\sqrt{2\pi}\sigma^3}e^{-\frac{(x-\eta)^2}{2\sigma^2}}. \tag{A.1}$$

When x is in $[\alpha, \eta]$, $f'(x)$ is strictly negative and thus $f(x)$ strictly decreases in $[\alpha, \eta]$. In the limit as $x \to -\infty$, the first term dominates and we have $\lim_{x\to-\infty} f'(x) > 0$. As $x \to \infty$, the second term dominates and we have $\lim_{x\to\infty} f'(x) > 0$.

Based on the fact that $\lim_{x\to-\infty} f(x)$=0 and $\lim_{x\to-\infty} f'(x) > 0$, we know $f(x)$ increases from the limit value of 0 as x increases from $-\infty$. Since $f(x)$ is continuous and has a zero crossing (solution) at x^o, $f(x)$ must have at least one local maximum. Since $f(x)$ decreases in the range $[\alpha, \eta]$, there exists a local maximum x^{max} and $x^{max} < \alpha$. Similar arguments can be made to show that there exist a local minimum x^{min} and $x^{min} > \eta$.

Consider the expression of $f'(x)$ in equation A.1. Denote the term $\frac{x-\alpha}{\sqrt{2\pi}\sigma^3}e^{-\frac{(x-\alpha)^2}{2\sigma^2}}$ by $f_1(x)$ and the term $\lambda\frac{x-\eta}{\sqrt{2\pi}\sigma^3}e^{-\frac{(x-\eta)^2}{2\sigma^2}}$ by $f_2(x)$. Consider the ratio between $f_1(x)$ and $f_2(x)$ given below

$$\frac{f_1(x)}{f_2(x)} = \frac{(x-\alpha)}{\lambda(x-\eta)} \times \frac{e^{-\frac{(x-\alpha)^2}{2\sigma^2}}}{e^{-\frac{(x-\eta)^2}{2\sigma^2}}}. \tag{A.2}$$

We can think of $\frac{f_1(x)}{f_2(x)}$ as the product of two quantities whose derivatives are given by

$$\begin{aligned} \frac{d}{dx}\left(\frac{x-\alpha}{\lambda(x-\eta)}\right) &= -\frac{\eta-\alpha}{\lambda(x-\eta)^2} < 0, x \neq \eta \\ \frac{d}{dx}\left(\frac{e^{-\frac{(x-\alpha)^2}{2\sigma^2}}}{e^{-\frac{(x-\eta)^2}{2\sigma^2}}}\right) &= -\frac{\eta-\alpha}{\sigma^2}e^{-\frac{2x(\eta-\alpha)+\alpha^2-\eta^2}{2\sigma^2}} < 0. \end{aligned}$$

Since $\frac{f_1(x)}{f_2(x)}$ is a product of two strictly decreasing quantities, it must be a strictly decreasing function. Although $\frac{f_1(x)}{f_2(x)}$ has discontinuity at η, it still holds that $\frac{f_1(x)}{f_2(x)}$ is a decreasing function in both intervals $(-\infty, \eta)$ and (η, ∞).

Since $\frac{f_1(x)}{f_2(x)}$ strictly decreases in $(-\infty, \eta)$, x^{max} is the only value of x in $(-\infty, \eta)$ at which $\frac{f_1(x)}{f_2(x)}$ is equal to 1, or equivalently $f'(x)$ is equal to 0. From the fact that $f(x)$ strictly decreases in $[\alpha, \eta]$, we know that there cannot be an extremum in $[\alpha, \eta]$. Therefore, x^{max} is the only extremum (which happens to be a local maximum) in $(-\infty, \alpha)$ $(\subset (-\infty, \eta))$.

By the same arguments, x^{min} is the only value of x in (η, ∞) at which $f'(x) = 0$. Thus x^{min} is the only extremum (which happens to be a local minimum) in (η, ∞).

We conclude that $f(x)$ has only one local maximum and one local minimum.

iii) The proof follows directly from the arguments given in the proof of *ii)*

iv), vi) Since there is no extremum in $(-\infty, x^{max})$ and $\lim_{x\to-\infty} f'(x) > 0$ and $f(x)$ is continuous in $(-\infty, x^{max})$, $f(x)$ strictly increases in $(-\infty, x^{max})$. Similarly, since there is no extremum in (x^{min}, ∞) and $\lim_{x\to\infty} f'(x) > 0$ and $f(x)$ is continuous in (x^{min}, ∞), $f(x)$ strictly increases in (x^{min}, ∞).

Consider the second derivative of $f(x)$ as shown below

$$f''(x) = \frac{1}{\sqrt{2\pi\sigma^5}}\left[(x-\alpha)^2 - \sigma^2\right] e^{-\frac{(x-\alpha)^2}{2\sigma^2}} - \lambda\frac{1}{\sqrt{2\pi\sigma^5}}\left[(x-\eta)^2 - \sigma^2\right] e^{-\frac{(x-\eta)^2}{2\sigma^2}}. \quad \text{(A.3)}$$

Note that in the limit, $\lim_{x\to-\infty} f''(x) > 0$ and $\lim_{x\to\infty} f''(x) < 0$. Let $g_1(x)$ and $g_2(x)$ denote the first and the second terms of $f''(x)$ in equation A.3 as follows

$$\begin{aligned} g_1(x) &= \frac{1}{\sqrt{2\pi\sigma^5}}\left[(x-\alpha)^2 - \sigma^2\right] e^{-\frac{(x-\alpha)^2}{2\sigma^2}} \\ g_2(x) &= \lambda\frac{1}{\sqrt{2\pi\sigma^5}}\left[(x-\eta)^2 - \sigma^2\right] e^{-\frac{(x-\eta)^2}{2\sigma^2}}. \end{aligned}$$

We shall use the similar argument as in the proof of *ii)*. Consider the ratio $\frac{g_1(x)}{g_2(x)}$ as the product of two quantities

$$\frac{g_1(x)}{g_2(x)} = \frac{1}{\lambda}\frac{(x-\alpha)^2 - \sigma^2}{(x-\eta)^2 - \sigma^2} \times \frac{e^{-\frac{(x-\alpha)^2}{2\sigma^2}}}{e^{-\frac{(x-\eta)^2}{2\sigma^2}}}.$$

We find the derivatives of both quantities to be

$$\begin{aligned} \frac{d}{dx}\left(\frac{1}{\lambda}\frac{(x-\alpha)^2 - \sigma^2}{(x-\eta)^2 - \sigma^2}\right) &= \frac{-2(\eta-\alpha)\left[(x-\eta)(x-\alpha) + \sigma^2\right]}{\lambda\left[(x-\eta)^2 - \sigma^2\right]^2} \\ &< 0, x < \alpha \text{ or } x > \eta, x \neq \eta \pm \sigma \\ \frac{d}{dx}\left(\frac{e^{-\frac{(x-\alpha)^2}{2\sigma^2}}}{e^{-\frac{(x-\eta)^2}{2\sigma^2}}}\right) &= -\frac{\eta-\alpha}{\sigma^2} e^{-\frac{2x(\eta-\alpha)+\alpha^2-\eta^2}{2\sigma^2}} < 0, \end{aligned}$$

where we use the fact that $x < \alpha$ or $x > \eta$ to arrive at the first inequality.

Since $\frac{g_1(x)}{g_2(x)}$ is the product of two strictly decreasing quantities, it must also be a strictly decreasing function of x in $(-\infty, \alpha)$ and in (η, ∞), except possibly at discontinuities which occur at $\eta \pm \sigma$.

We now argue that $f''(x^{max}) < 0$ and $f''(x^{min}) > 0$, since a twice differentiable function must be concave $\frown$ at a local maximum and convex $\smile$ at a local minimum. Since $\lim_{x\to-\infty}$ $f''(x) > 0$, or equivalently $f''(x)$ is convex $\smile$ in the limit, there must be at least one saddle point where $f(x)$ changes from convexity ($\smile$) to concavity ($\frown$) in $(-\infty, x^{max})$. Similarly, since $\lim_{x\to\infty} f''(x) < 0$, or equivalently $f''(x)$ in concave $\frown$ in the limit, there must be at least one saddle point where $f(x)$ changes from convexity ($\smile$) to concavity ($\frown$) in (x^{min}, ∞).

If $x^{max} < \eta - \sigma$, we have in the range $(-\infty, x^{max})$ $(\subset (-\infty, \eta - \sigma))$ a unique saddle point x^s at which $\frac{g_1(x^s)}{g_2(x^s)} = 0$ since there must be at least one saddle point in $(-\infty, x^{max})$.

We shall show that even if $x^{max} \geq \eta - \sigma$, we still have a unique saddle point in $(-\infty, x^{max})$. Since $\frac{g_1(x^s)}{g_2(x^s)}$ strictly decreases in $(-\infty, \eta - \sigma)$, $(\eta - \sigma, \eta + \sigma)$, and $(\eta + \sigma, \infty)$, we can have at most one saddle point in each of the three intervals.

At the point of discontinuity $\eta - \sigma$, we have from equation A.3 that

$$f''(\eta - \sigma) = \frac{1}{\sqrt{2\pi}\sigma^5}\left[(\eta - \alpha - \sigma)^2 - \sigma^2\right] e^{-\frac{(\eta-\alpha-\sigma)^2}{2\sigma^2}},$$

which is equal to 0 only if $\eta - \alpha = 2\sigma$. However, for $x^{max} \geq \eta - \sigma$, we must have $\eta - \sigma \leq x^{max} < \alpha$ or $\eta - \alpha < \sigma$. So it cannot happen that $\eta - \alpha = 2\sigma$ and $\eta - \sigma$ is not a saddle point when $x^{max} \geq \eta - \sigma$.

As a result, for $x^{max} > \eta - \sigma$, we can have either one or two saddle points in $(-\infty, x^{max})$. In the case of one saddle point, we must have one saddle point in either $(-\infty, \eta - \sigma)$ or $(\eta - \sigma, x^{max})$ $(\subset (\eta - \sigma, \eta + \sigma))$. In the case of two saddle points, one of the saddle points is in $(-\infty, \eta - \sigma)$ and the other is in $(\eta - \sigma, x^{max})$ $(\subset (\eta - \sigma, \eta + \sigma))$. We do not have to consider saddle points in $[\eta + \sigma, \infty)$ since we know that $x^{max} < \alpha < \eta + \sigma$.

However, since $\lim_{x\to-\infty} f''(x) > 0$ and $f''(x^{max}) < 0$, we cannot have an even number of saddle points in $(-\infty, x^{max})$. Thus we have a unique saddle point in $(-\infty, x^{max})$ for $x^{max} \geq \eta - \sigma$.

We now repeat the same arguments to show that there is a unique saddle point in (x^{min}, ∞). If $x^{min} > \eta + \sigma$, we have in (x^{min}, ∞) $(\subset (\eta + \sigma, \infty))$ a unique saddle point since we have argued that there must be at least one saddle point in (x^{min}, ∞).

At the point of discontinuity $\eta + \sigma$, we have from equation A.3 that

$$f''(\eta + \sigma) = \frac{1}{\sqrt{2\pi\sigma^5}} \left[(\eta - \alpha + \sigma)^2 - \sigma^2 \right] e^{-\frac{(\eta-\alpha+\sigma)^2}{2\sigma^2}},$$

which is never equal to 0 since $\eta > \alpha$. It follows that $\eta + \sigma$ is not a saddle point when $x^{min} \leq \eta + \sigma$.

Therefore, for $x^{min} < \eta + \sigma$, we can have either one or two saddle points in (x^{min}, ∞). In the case of one saddle point, we must have one saddle point in either $(x^{min}, \eta + \sigma)$ $(\subset (\eta - \sigma, \eta + \sigma))$ or $(\eta + \sigma, \infty)$. In the case of two saddle points, one of the saddle points is in $(x^{min}, \eta + \sigma)$ $(\subset (\eta - \sigma, \eta + \sigma))$ and the other is in $(\eta + \sigma, \infty)$. We do not have to consider saddle points in $(-\infty, \eta - \sigma]$ since we know that $\eta - \sigma < \eta < x^{min}$.

However, since $\lim_{x \to \infty} f''(x) < 0$ and $f''(x^{min}) > 0$, we cannot have an even number of saddle points in (x^{min}, ∞). Thus we have a unique saddle point in (x^{min}, ∞) for $x^{min} \leq \eta + \sigma$.

v) Since there is a unique local maximum and a unique local minimum and there is no other extreme point, it follows that the function $f(x)$ must strictly decrease as we move from x^{max} to x^{min}. □

Lemma 3.2 *Under assumption 3.1, $f(x)$ as defined in equation 3.9 has a strictly decreasing second derivative in $(x^s, \alpha]$, where x^s denotes the unique saddle point in $(-\infty, x^{max})$.*

Proof Under assumption 3.1, $f''(x)$ decreases in $(\alpha - \sqrt{3}\sigma, \alpha]$ since in $(\alpha - \sqrt{3}\sigma, \alpha]$, both terms $g_1(x)$ and $-g_2(x)$ in equation A.3 have negative derivatives, and at $x = \alpha$, $g_1(x)$ has a zero derivative while $-g_2(x)$ has a negative derivative.

We are left to show that $f''(x)$ decreases in the region $(x^s, \alpha - \sqrt{3}\sigma]$. Consider the third derivative of $f(x)$ given below

$$\begin{aligned} f'''(x) &= \frac{1}{\sqrt{2\pi\sigma^7}} (\alpha - x) \left[(x - \alpha)^2 - 3\sigma^2 \right] e^{-\frac{(x-\alpha)^2}{2\sigma^2}} \\ & \quad -\lambda \frac{1}{\sqrt{2\pi\sigma^7}} (\eta - x) \left[(x - \eta)^2 - 3\sigma^2 \right] e^{-\frac{(x-\eta)^2}{2\sigma^2}}. \end{aligned} \tag{A.4}$$

Let $h_1(x)$ and $h_2(x)$ denote the first and the second terms of $f'''(x)$ in equation A.4 as follows

$$\begin{aligned}
h_1(x) &= \frac{1}{\sqrt{2\pi\sigma^7}}(\alpha - x)\left[(x-\alpha)^2 - 3\sigma^2\right] e^{-\frac{(x-\alpha)^2}{2\sigma^2}} \\
h_2(x) &= \lambda\frac{1}{\sqrt{2\pi\sigma^7}}(\eta - x)\left[(x-\eta)^2 - 3\sigma^2\right] e^{-\frac{(x-\eta)^2}{2\sigma^2}}.
\end{aligned}$$

We can express $\frac{h_1(x)}{h_2(x)}$ as the product of two quantities

$$\frac{h_1(x)}{h_2(x)} = \frac{1}{\lambda}\frac{(x-\alpha)\left[(x-\alpha)^2 - 3\sigma^2\right]}{(x-\eta)\left[(x-\eta)^2 - 3\sigma^2\right]} \times \frac{e^{-\frac{(x-\alpha)^2}{2\sigma^2}}}{e^{-\frac{(x-\eta)^2}{2\sigma^2}}}.$$

The derivative of the second term in equation A.4 is

$$\frac{d}{dx}\left(\frac{e^{-\frac{(x-\alpha)^2}{2\sigma^2}}}{e^{-\frac{(x-\eta)^2}{2\sigma^2}}}\right) = -\frac{\eta-\alpha}{\sigma^2}e^{-\frac{2x(\eta-\alpha)+\alpha^2-\eta^2}{2\sigma^2}} < 0,$$

while the derivative of the first term in equation A.4 is

$$\begin{aligned}
&\frac{d}{dx}\left(\frac{1}{\lambda}\frac{(x-\alpha)\left[(x-\alpha)^2 - 3\sigma^2\right]}{(x-\eta)\left[(x-\eta)^2 - 3\sigma^2\right]}\right) \\
= &\frac{(x-\eta)\left[(x-\eta)^2 - 3\sigma^2\right]\left[3(x-\alpha)^2 - 3\sigma^2\right]}{\lambda(x-\eta)^2\left[(x-\eta)^2 - 3\sigma^2\right]^2} \\
&-\frac{(x-\alpha)\left[(x-\alpha)^2 - 3\sigma^2\right]\left[3(x-\eta)^2 - 3\sigma^2\right]}{\lambda(x-\eta)^2\left[(x-\eta)^2 - 3\sigma^2\right]^2} &\text{(A.5)} \\
< &\ 0 \text{ (to be justified)}, x \leq \alpha - \sqrt{3}\sigma. &\text{(A.6)}
\end{aligned}$$

To justify inequality A.6, let $h_3(x)$ denote the numerator of the right hand side of equation A.5. We choose to ignore the denominator since it is always positive because $\lambda > 0$ and we only want to know about the sign of the derivative, which is the same as the sign of $h_3(x)$ whose expression is given by

$$\begin{aligned}
h_3(x) &= (x-\eta)\left[(x-\eta)^2 - 3\sigma^2\right]\left[3(x-\alpha)^2 - 3\sigma^2\right] \\
&\quad -(x-\alpha)\left[(x-\alpha)^2 - 3\sigma^2\right]\left[3(x-\eta)^2 - 3\sigma^2\right]. &\text{(A.7)}
\end{aligned}$$

For $x \in (x^s, \alpha - \sqrt{3}\sigma)$, we have $x < \alpha - \sqrt{3}\sigma$ and it is easy to verify that all the quantities in square brackets are positive. Using the fact that $x - \eta < x - \alpha$, we can bound $h_3(x)$ in equation A.5 by

$$\begin{aligned}
h_3(x) &< (x-\alpha)\left[(x-\eta)^2-3\sigma^2\right]\left[3(x-\alpha)^2-3\sigma^2\right] \\
&\quad -(x-\alpha)\left[(x-\alpha)^2-3\sigma^2\right]\left[3(x-\eta)^2-3\sigma^2\right] \\
&= (x-\alpha)\times 6\sigma^2\left[(x-\eta)^2-(x-\alpha)^2\right] \\
&< 0,
\end{aligned}$$

where the last inequality follows from the fact that $x \in (x^s, \alpha-\sqrt{3}\sigma)$, hence $x < \alpha$.

For $x = \alpha - \sqrt{3}\sigma$, we have from equation A.7 that

$$\begin{aligned}
h_3(x) &= (x-\eta)\left[(x-\eta)^2-3\sigma^2\right]\left[3(x-\alpha)^2-3\sigma^2\right] \\
&< 0,
\end{aligned}$$

since the quantities in the square brackets are positive and $x < \eta$ by assumption 3.1.

Thus $\frac{h_1(x)}{h_2(x)}$ is a product of two strictly decreasing quantities and it is therefore a strictly decreasing function of x in $(x^s, \alpha-\sqrt{3}\sigma]$

We have just shown that there can be at most one point $\hat{x} \in (x^s, \alpha-\sqrt{3}\sigma)$ $(\subset (-\infty, \alpha-\sqrt{3}\sigma))$ such that $f'''(\hat{x}) = 0$. We shall show that $f'''(x^s) < 0$, or equivalently $\frac{h_1(x^s)}{h_2(x^s)} < 1$. Since $\frac{h_1(x)}{h_2(x)}$ is decreasing in $(x^s, \alpha-\sqrt{3}\sigma]$, this result will imply that $\frac{h_1(x)}{h_2(x)} < 1$ for all $x \in (x^s, \alpha-\sqrt{3}\sigma]$. We can then conclude that there does not exist such an $\hat{x}$ in $(x^s, \alpha-\sqrt{3}\sigma]$ and $f''(x)$ is decreasing in $(x^s, \alpha-\sqrt{3}\sigma]$.

To show that $f'''(x^s) < 0$, we can rewrite and bound the expression for $f'''(x)$ as follows

$$\begin{aligned}
f'''(x) &= -\frac{1}{\sqrt{2\pi\sigma^7}}(x-\alpha)\left[(x-\alpha)^2-3\sigma^2\right]e^{-\frac{(x-\alpha)^2}{2\sigma^2}} \\
&\quad +\lambda\frac{1}{\sqrt{2\pi\sigma^7}}(x-\eta)\left[(x-\eta)^2-3\sigma^2\right]e^{-\frac{(x-\eta)^2}{2\sigma^2}} \\
&= -\frac{x-\alpha}{\sigma^2}\left[\frac{(x-\alpha)^2-\sigma^2}{\sqrt{2\pi\sigma^5}}\right]e^{-\frac{(x-\alpha)^2}{2\sigma^2}}+\lambda\frac{x-\eta}{\sigma^2}\left[\frac{(x-\eta)^2-\sigma^2}{\sqrt{2\pi\sigma^5}}\right]e^{-\frac{(x-\eta)^2}{2\sigma^2}} \\
&\quad +\frac{2}{\sigma^2}\frac{x-\alpha}{\sqrt{2\pi\sigma^3}}e^{-\frac{(x-\alpha)^2}{2\sigma^2}}-\lambda\frac{2}{\sigma^2}\frac{x-\eta}{\sqrt{2\pi\sigma^3}}e^{-\frac{(x-\eta)^2}{2\sigma^2}} \\
&< -\frac{x-\eta}{\sigma^2}f''(x)-\frac{2}{\sigma^2}f'(x),
\end{aligned}$$

where we have used the fact that $x-\alpha > x-\eta$ to construct the inequality. By definition,

$f''(x^s) = 0$ and since $x^s < x^{max}$ we have from lemma 3.1 that $f'(x^s) > 0$. It then follows from the above expression that $f'''(x^s) < 0$. □

Lemma 3.3 *Given $\lambda > 0$, and assumption 3.1, there is at most one positive solution to the equation $A'(J) = \lambda B'(J)$ in $(0, \eta - \sigma)$.*

Proof In what follows, we use x as a dummy variable for the argument of all relevant functions (x takes the role of J). We can express $A'(x)$ - $\lambda B'(x)$ as

$$\begin{aligned} A'(x) - \lambda B'(x) &= \frac{1}{\sqrt{2\pi\sigma^2}}\left[e^{-\frac{(x-\alpha)^2}{2\sigma^2}} - \lambda e^{-\frac{(x-\eta)^2}{2\sigma^2}}\right] - \frac{1}{\sqrt{2\pi\sigma^2}}\left[e^{-\frac{(x+\alpha)^2}{2\sigma^2}} - \lambda e^{-\frac{(x+\eta)^2}{2\sigma^2}}\right] \\ &= f(x) - f(-x), \end{aligned}$$

where $f(x)$ is defined as in lemma 3.1. Let x^o, x^{max}, x^{min} denote its unique solution (zero crossing), the only local maximum, and the only local minimum respectively. Lemma 3.1 has helped us understand the behaviour of $f(x)$.

Let x^* denote the smallest positive value such that $f(x^*)$ $-f(-x^*) = 0$. Note that x^* may or may not exist. Finding x^* is equivalent to finding a point of intersection between $f(x)$ and $f(-x)$. For convenience, let $f_-(x)$ denote $f(-x)$. We shall provide the proof for three separate cases which cover all the possibilities (see the illustration in figure A-1).

i) $x^o \leq 0$: We shall show that no intersection exists in $(0, x^{min})$. From part *iii)* of lemma 3.1, we know that $x^{min} > \eta$. It then follows that there cannot be an intersection in $(0, \eta - \sigma)$ $(\subset (0, x^{min}))$.

To show that no intersection exists in $(0, x^{min})$, consider two distinct cases: $-x^o > x^{min}$ and $-x^o \leq x^{min}$. When $-x^o > x^{min}$, we have that $f(0) = f_-(0)$ and $f(x)$ is decreasing while $f_-(x)$ is increasing in $(0, x^{min})$. Therefore, there cannot be an intersection in $(0, x^{min})$.

When $-x^o \geq x^{min}$, we have that $f(0) = f_-(0)$ and $f(x)$ is decreasing while $f_-(x)$ is increasing in $(0, x^o)$. Therefore, there cannot be an intersection in $(0, x^o)$. In $[x^o, x^{min})$, $f_-(x)$ is non-negative while $f(x)$ is strictly negative. Therefore, there cannot be an intersection in $[x^o, x^{min})$. In conclusion, there cannot be an intersection in $(0, x^{min})$.

ii) $x^o > 0, x^{max} \leq 0$: We first argue there is no intersection in $(0, -x^{max}]$ since $f(0) =$ $f_-(0)$ and $f(x)$ remains below $f(0)$ while $f_-(x)$ is increasing in $(0, -x^{max}]$.

Consider three separate intervals: $(0, \alpha]$, $(\alpha, \alpha + \sigma]$, and $(\alpha + \sigma, \eta - \sigma)$. We shall argue that there can be no intersection in any of them.

- Consider the interval $(0, \alpha]$. If $-x^{max} \geq \alpha$, there is no intersection in $(0, \alpha]$ since there is no intersection in $(0, -x^{max}]$. So we consider the case when $-x^{max} < \alpha$. In this case, the intersection, if exists, must lie in $(-x^{max}, \alpha]$.

  Let x^s denote the unique saddle point in $(-\infty, x^{max})$. In addition, let $\delta > 0$ and $x^{max} + \delta \leq \alpha$. We have the following inequality.

$$\begin{aligned} |f'(x^{max} + \delta)| &= -\int_{x^{max}}^{x^{max}+\delta} f''(x)dx \\ &> -\int_{x^{max}-\delta}^{x^{max}} f''(x)dx \\ &= |f'(x^{max} - \delta)|, \end{aligned} \tag{A.8}$$

  where the inequality is justified in the case $x^{max} - \delta > x^s$ from the fact that $f''(x)$ is decreasing in $(x^s, \alpha]$ (lemma 3.2); and in the case $x^{max} - \delta < x^s$ from the fact that $f''(x)$ is nonnegative in $(x^{max} - \delta, x^s]$ and is decreasing in $(x^s, \alpha]$.

  Let $\epsilon \in (-x^{max}, \alpha]$ and $\delta = x^{max} + \epsilon$ (the difference of x^{max} and $-\epsilon$). Consider the following sequence of inequalities.

$$\begin{aligned} f(\epsilon) - f(-\epsilon) &= \int_{x^{max}}^{\epsilon} f'(x)dx + \int_{-\epsilon}^{x^{max}} f'(x)dx \\ &= \int_{x^{max}-\delta}^{x^{max}} f'(x)dx + \int_{x^{max}}^{x^{max}+\delta} f'(x)dx + \int_{x^{max}+\delta}^{\epsilon} f'(x)dx \\ &= \int_{x^{max}-\delta}^{x^{max}} |f'(x)|dx - \int_{x^{max}}^{x^{max}+\delta} |f'(x)|dx + \int_{x^{max}+\delta}^{\epsilon} f'(x)dx \\ &< 0 + \int_{x^{max}+\delta}^{\epsilon} f'(x)dx \\ &< 0, \end{aligned} \tag{A.9}$$

  where the first inequality is the direct result from A.8 and the second inequality follows from the fact that $f'(x) < 0$ in $(x^{max}, \alpha]$ ($\supset (x^{max} + \delta, \epsilon)$).

  Inequality A.9 asserts that there can be no intersection in $(-x^{max}, \alpha]$. Furthermore, inequalities A.9 and A.8 tell us that, given $-x^{max} < \alpha$, $f(\alpha) < f_-(\alpha)$ and $f'(\alpha) < f'_-(\alpha)$ (equivalently $f(x)$ decreases faster than $f_-(x)$ at $x = \alpha$).

For convenience, we define the components of $f(x)$ as $f_\alpha(x)$ and $f_\eta(x)$ whose expressions are given below,

$$\begin{aligned} f(x) &= f_\alpha(x) + f_\eta(x) && \text{(A.10)} \\ f_\alpha(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\alpha)^2}{2\sigma^2}} && \text{(A.11)} \\ f_\eta(x) &= -\lambda \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\eta)^2}{2\sigma^2}}. && \text{(A.12)} \end{aligned}$$

Figure A-2 shows example curves of $f_\alpha(x)$ and $f_\eta(x)$.

- Consider the interval $(\alpha, \alpha+\sigma]$. If $-x^{max} \geq \alpha+\sigma$, there is no intersection in $(0, \alpha+\sigma]$ since there is no intersection in $(0, -x^{max}]$. So we consider the case when $-x^{max} < \alpha + \sigma$.

  Under assumption 3.1, $f(x)$ is strictly concave $\frown$ in $(\alpha, \alpha+\sigma]$ since $f_\alpha(x)$ and $f_\eta(x)$ are both decreasing and concave $\frown$ (see figure A-2). Therefore, $f(x)$ is decreasing and concave $\frown$ at every point in $(\alpha, \alpha+\sigma]$. On the other hand, the contribution of $f_\alpha(x)$ to $f_-(x)$ is decreasing but convex $\smile$, while the contribution of $f_\eta(x)$ is increasing. An additional condition that $f'(\alpha) < f'_-(\alpha)$ will lead to the conclusion that $f(x)$ is decreasing faster than $f_-(x)$ at every point in $(\alpha, \alpha+\sigma]$.

  For $-x^{max} < \alpha$, we have shown that $f(\alpha) < f_-(\alpha)$ and $f'(\alpha) < f'_-(\alpha)$. For $-x^{max} > \alpha$ ($f(x)$ is decreasing and $f_-(x)$ is increasing at $x = \alpha$), we have that $f(\alpha) < f_-(\alpha)$ and $f'(\alpha) < 0 < f'_-(\alpha)$. In both cases, we have that $f(\alpha) < f_-(\alpha)$ and $f(x)$ decreases faster than $f_-(x)$ at every point in $(\alpha, \alpha+\sigma]$. Therefore, there cannot be an intersection in $(\alpha, \alpha+\sigma]$.

  Consequently, given that $-x^{max} < \alpha + \sigma$, we have that $f(\alpha+\sigma) < f_-(\alpha+\sigma)$ and $f'(\alpha+\sigma) < f'_-(\alpha+\sigma)$.

- Consider the interval $(\alpha+\sigma, \eta-\sigma]$. If $-x^{max} \geq \eta - \sigma$, there is no intersection in $(0, \eta-\sigma]$ since there is no intersection in $(0, -x^{max}]$. So we consider the case when $-x^{max} < \eta - \sigma$.

  In $(\alpha+\sigma, \eta-\sigma)$, the contributions of $f_\alpha(x)$ to $f(x)$ and $f_-(x)$ are both decreasing and convex $\smile$. However, at any particular point $x \in (\alpha+\sigma, \eta-\sigma)$, the contribution of $f_\alpha(x)$ to $f(x)$ corresponds to a point on the normal curve (with the bell shape)

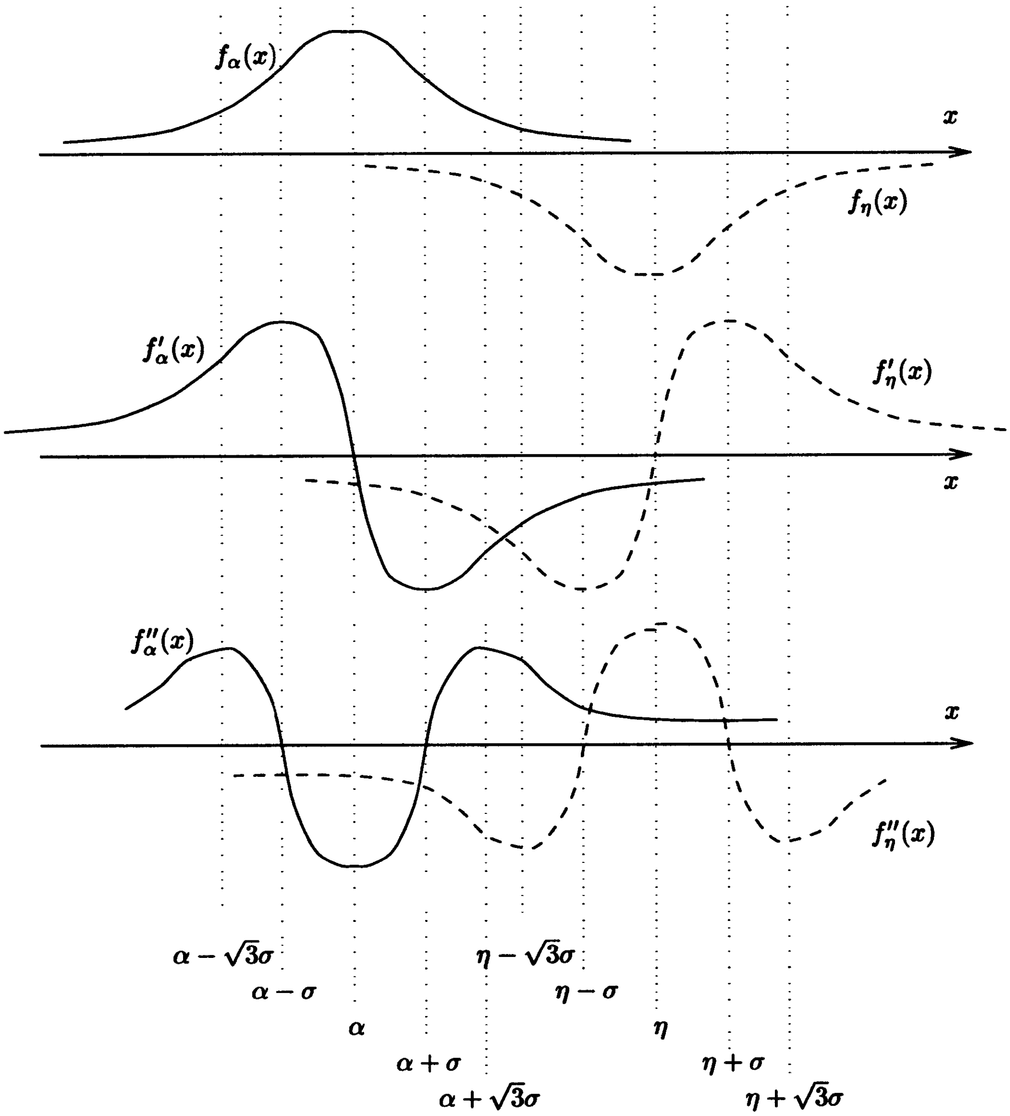


Figure A-2: $f_\alpha(x)$ and $f_\eta(x)$ as defined in equations A.11 and A.12.

whose distance from the center of the curve (equivalently the mean) is smaller than the distance from the mean of the point on the normal curve corresponding to the contribution of $f_{\alpha}(x)$ to $f_{-}(x)$. Therefore, the contribution of $f_{\alpha}(x)$ to $f(x)$ decreases faster than its contribution to $f_{-}(x)$. On the other hand, the contribution of $f_{\eta}(x)$ to $f(x)$ is decreasing and concave $\frown$, while its contribution to $f_{-}(x)$ is increasing. An additional condition that $f'(\alpha+\sigma) < f'_{-}(\alpha+\sigma)$ will lead to the conclusion that $f(x)$ is decreasing faster than $f_{-}(x)$ at every point in $(\alpha+\sigma, \eta-\sigma]$.

For $-x^{max} < \alpha+\sigma$, we have shown that $f(\alpha+\sigma) < f_{-}(\alpha+\sigma)$ and $f'(\alpha+\sigma) < f'_{-}(\alpha+\sigma)$. For $-x^{max} > \alpha+\sigma$ ($f(x)$ is decreasing and $f_{-}(x)$ is increasing at $x = \alpha+\sigma$), we have that $f(\alpha+\sigma) < f_{-}(\alpha+\sigma)$ and $f'(\alpha+\sigma) < 0 < f'_{-}(\alpha+\sigma)$. In both cases, we have that $f(\alpha+\sigma) < f_{-}(\alpha+\sigma)$ and $f(x)$ decreases faster than $f_{-}(x)$ at every point in $(\alpha+\sigma, \eta-\sigma]$. Therefore, there cannot be an intersection in $(\alpha+\sigma, \eta-\sigma]$.

iii) $x^{o} > 0, x^{max} > 0$: We first argue there is no intersection in $(0, x^{max}]$ since $f(0) = f_{-}(0)$ and $f(x)$ is increasing while $f_{-}(x)$ is decreasing in $(0, x^{max}]$. Consider two separate intervals: (x^{max}, α) and $[\alpha, \eta-\sigma)$.

- We now show that there can be at most one intersection in (x^{max}, α). Given that x^{*} is an intersection point and the fact that $f(x^{max}) > f_{-}(x^{max})$, we must have that

$$f'(x^{*}) < f'_{-}(x^{*}), \tag{A.13}$$

since the intersection must occur before x^{o} and both $f(x)$ and $f_{-}(x)$ are decreasing in (x^{max}, x^{o}). For $f(x)$ and $f_{-}(x)$ to intersect, $f(x)$ has to decrease faster than $f_{-}(x)$ at the intersection point.

Consider a point $x \in (x^{*}, \alpha)$ and let $\epsilon = x - x^{*}$. We shall show that $f'(x) < f'_{-}(x)$ using the following sequence of inequalities.

$$\begin{aligned} f'(x) - f'_{-}(x) &= f'(x^{*}) + \int_{x^{*}}^{x^{*}+\epsilon} f''(x)dx \\ &\quad - f'_{-}(x^{*}) - \int_{x^{*}}^{x^{*}+\epsilon} f''_{-}(x)dx \\ &= f'(x^{*}) + \int_{x^{*}}^{x^{*}+\epsilon} f''(x)dx \end{aligned}$$

$$\begin{aligned} & -f'_-(x^*) - \int_{-x^*-\epsilon}^{-x^*} f''(x)dx \\ < \;& 0 + \int_{x^*}^{x^*+\epsilon} f''(x)dx - \int_{-x^*-\epsilon}^{-x^*} f''(x)dx \qquad \text{(A.14)} \\ < \;& 0, \qquad \text{(A.15)} \end{aligned}$$

where the inequality A.14 follows from inequality A.13. We now prove inequality A.15. Let x^s denote the unique saddle point in $(-\infty, x^{max})$ (lemma 3.1, part *iv*)). If $x^s < -x^* - \epsilon$, then inequality A.15 holds as a consequence of lemma 3.2. If $-x^* - \epsilon \le x^s < -x^*$, we have that

$$\int_{-x^*-\epsilon}^{x^s} f''(x)dx > 0 \qquad \text{(A.16)}$$

$$0 > \int_{x^s}^{-x^*} f''(x)dx > \int_{x^*}^{-x^s} f''(x)dx \qquad \text{(A.17)}$$

$$\int_{-x^s}^{x^*+\epsilon} f''(x)dx < 0, \qquad \text{(A.18)}$$

where inequality A.16 follows from the fact that $f(x)$ is convex $\smile$ in $(-\infty, x^s)$. Inequality A.17 follows as a consequence of lemma 3.2. Inequality A.18 follows since $f(x)$ is concave $\frown$ in (x^s, α) ($\supset (-x^s, \alpha)$). Combining inequalities A.16, A.17 and A.18, inequality A.15 holds.

If $-x^* < x^s$, we have that $f(x)$ is convex $\smile$ in $(-x^* - \epsilon, -x^*)$. Note that $x^* > x^s$ since $x^* > x^{max}$ (there is no intersection in $(0, x^{max}]$) and $x^{max} > x^s$ ($f(x)$ must be concave $\frown$ at a local maximum so x^s must lie before x^{max}). It follows that $(x^*, x^* + \epsilon) \subset (x^s, \alpha)$ and thus $f(x)$ is concave $\frown$ in $(x^*, x^* + \epsilon)$. As a result,

$$\int_{-x^*-\epsilon}^{-x^*} f''(x)dx > 0 \qquad \text{(A.19)}$$

$$\int_{x^*}^{x^*+\epsilon} f''(x)dx < 0. \qquad \text{(A.20)}$$

Combining inequalities A.19 and A.20, inequality A.15 holds.

Therefore, after the intersection point x^*, $f(x)$ decreases faster than $f_-(x)$ at all points in (x^*, α). It follows that there can be at most one intersection point in (x^{max}, α). In addition, if there exists an intersection point in (x^{max}, α), we can assert that $f(\alpha) < f_-(\alpha)$ and $f'(\alpha) < f'_-(\alpha)$.

- Consider the interval $[\alpha, \eta - \sigma)$. We shall show that, given an intersection in $(0, \alpha)$, there can be no additional intersection in $[\alpha, \eta - \sigma)$. In addition, given no intersection in $(0, \alpha)$, there is at most one intersection in $[\alpha, \eta - \sigma)$.

  We first show that $-x^s < \alpha$, i.e. $f_-(x)$ is convex $\smile$ in $[\alpha, \infty)$. Since x^s is the only saddle point in $(-\infty, x^{max})$ where $f(x)$ changes from being convex $\smile$ to being concave $\frown$ (lemma 3.1), it is sufficient to show that $f''(-\alpha) > 0$.

  Since we are in case *iii)* where $x^{max} > 0$, we must have that $f'(0) > 0$. From equation A.1, we can find a bound on λ as shown below.

$$\lambda < \frac{\alpha}{\eta} e^{-\frac{\alpha^2}{2\sigma^2} + \frac{\eta^2}{2\sigma^2}} \tag{A.21}$$

  Consider the following expression for $f''(-\alpha)$ obtained from equation A.3,

$$\begin{aligned}
f''(-\alpha) &= \frac{1}{\sqrt{2\pi}\sigma^5} \left[(4\alpha^2 - \sigma^2) e^{-\frac{4\alpha^2}{2\sigma^2}} - \lambda(\alpha^2 + \eta^2 + 2\alpha\eta - \sigma^2) e^{-\frac{\alpha^2+\eta^2+2\alpha\eta}{2\sigma^2}} \right] \\
&> \frac{1}{\sqrt{2\pi}\sigma^5} \left[(4\alpha^2 - \sigma^2) e^{-\frac{4\alpha^2}{2\sigma^2}} - \frac{\alpha}{\eta} e^{-\frac{\alpha^2}{2\sigma^2} + \frac{\eta^2}{2\sigma^2}} (\alpha^2 + \eta^2 + 2\alpha\eta - \sigma^2) e^{-\frac{\alpha^2+\eta^2+2\alpha\eta}{2\sigma^2}} \right] \\
&= \frac{e^{-\frac{2\alpha^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma^5} \left[(4\alpha^2 - \sigma^2) e^{-\frac{2\alpha^2}{2\sigma^2}} - \left(\frac{\alpha}{\eta}(\alpha^2 - \sigma^2) + \alpha\eta + 2\alpha^2 \right) e^{-\frac{2\alpha\eta}{2\sigma^2}} \right] \\
&= \frac{e^{-\frac{2\alpha^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma^5} \left(2\alpha^2 \left[e^{-\frac{2\alpha^2}{2\sigma^2}} - e^{-\frac{2\alpha\eta}{2\sigma^2}} \right] + \left[\alpha^2 e^{-\frac{2\alpha^2}{2\sigma^2}} - \alpha\eta e^{-\frac{2\alpha\eta}{2\sigma^2}} \right] \right. \\
&\quad \left. + (\alpha^2 - \sigma^2) \left[e^{-\frac{2\alpha^2}{2\sigma^2}} - \frac{\alpha}{\eta} e^{-\frac{2\alpha\eta}{2\sigma^2}} \right] \right) \qquad \text{(A.22)} \\
&> 0, \qquad \text{(A.23)}
\end{aligned}$$

  where the first inequality is the direct result of inequality A.21. The second inequality results from the fact that all quantities in square brackets in equation A.22 are positive and from assumption 3.1. It is easy to see that the quantities in the first and the third square brackets are positive. The quantity in the second square bracket is positive since the function $xe^{-\frac{x}{\sigma^2}}$ is decreasing in (σ^2, ∞), equivalently its derivative $\left(1 - \frac{x}{\sigma^2}\right) e^{-\frac{x}{\sigma^2}}$ is negative in (σ^2, ∞).

  In conclusion, we have shown that

$$f''_-(x) > 0, x \in [\alpha, \infty). \tag{A.24}$$

Let $x^{s2} < x^{s3} < x^{s4} < ... < x^{sK}$ denote all the saddle points, if exist, in $[\alpha, \eta - \sigma)$. Note that x^{sK} denotes the last saddle point before $\eta - \sigma$. We shall consider each of the following intervals: $[\alpha, x^{s2})$, $[x^{s2}, x^{s3})$, $[x^{s3}, x^{s4})$, ..., $[x^{sK}, \eta - \sigma)$. We shall show that in each of these intervals, the following two properties hold.

1. If an intersection occurs before the interval, there can be no additional intersection in the interval.
2. If no intersection occurs before the interval, there can be at most one intersection in the interval.

In what follows, we prove the two properties for each interval. Note that both $f(x)$ and $f_-(x)$ are decreasing in all the intervals we consider.

- Consider the interval $[\alpha, x^{s2})$. We now show that, given an intersection before α (in (x^{max}, α)), there is no additional intersection in $[\alpha, x^{s2})$. As a reminder, we have shown that, given an intersection before α, $f(\alpha) < f_-(\alpha)$ and $f'(\alpha) < f'_-(\alpha)$.

  Let $x \in [\alpha, x^{s2})$ and $\epsilon = x - \alpha$. Consider the following sequence of inequalities.

$$\begin{aligned} f'(x) - f'_-(x) &= f'(\alpha) + \int_{\alpha}^{\alpha+\epsilon} f''(x)dx \\ &\quad -f'_-(\alpha) - \int_{\alpha}^{\alpha+\epsilon} f''_-(x)dx \\ &< \int_{\alpha}^{\alpha+\epsilon} f''(x)dx - \int_{\alpha}^{\alpha+\epsilon} f''_-(x)dx \\ &< 0, \end{aligned}$$

  where the first inequality holds since $f'(\alpha) < f'_-(\alpha)$ and the second inequality holds since $f''(\alpha) < 0$ (a consequence of lemma 3.2) and thus $f(x)$ is concave $\frown$ in $[\alpha, x^{s2})$, while $f_-(x)$ is convex $\smile$ in $[\alpha, x^{s2})$ (inequality A.24). Since $f(\alpha) < f_-(\alpha)$ and $f(x)$ decreases faster than $f_-(x)$ at every point in $[\alpha, x^{s2})$, there is no additional intersection in $[\alpha, x^{s2})$. In addition, since $f(\alpha) < f_-(\alpha)$ and $f'(\alpha) < f'_-(\alpha)$, we have that $f(x^{s2}) < f_-(x^{s2})$ and $f'(x^{s2}) < f'_-(x^{s2})$.

  Note that if there is no saddle point in $[\alpha, \eta - \sigma)$ (equivalently x^{s2} does not exist), the fact that $f'(\alpha) < f'_-(\alpha)$ and $f(x)$ is concave $\frown$ while $f_-(x)$ is convex

$\smile$ tells us that $f(x)$ decreases faster than $f_-(x)$ at every point in $[\alpha, \eta - \sigma)$ and no additional intersection is possible in $[\alpha, \eta - \sigma)$.

Given that there is no intersection before α (in (x^{max}, α)), we now show that there is at most one intersection in $[\alpha, x^{s2})$. As a reminder, we have in this case that $f(\alpha) > f_-(\alpha)$ since $f(0) = f_-(0)$, $f'(0) > 0 > f'_-(0)$, and no intersection has occured before α.

Let x^* denote the smallest point of intersection, if exists, in $[\alpha, x^{s2})$. Since $f(\alpha) > f_-(\alpha)$ and both $f(x)$ and $f_-(x)$ are decreasing, we must have that $f'(x^*) < f'_-(x^*)$.

We argue that $f(x)$ decreases faster than $f_-(x)$ at every point in (x^*, x^{s2}) since $f'(x^*) < f'_-(x^*)$ and $f(x)$ is concave $\frown$ while $f_-(x)$ is convex $\smile$ (inequality A.24). Therefore, there can be no additional intersection in (x^*, x^{s2}). Moreover, given that x^* exists in $[\alpha, x^{s2})$, we have that $f(x^{s2}) < f_-(x^{s2})$ and $f'(x^{s2}) < f'_-(x^{s2})$.

Finally, if there is no saddle point in $[\alpha, \eta - \sigma)$ (equivalently x^{s2} does not exist), we can let x^* denote the smallest intersection point in $[\alpha, \eta - \sigma)$. The fact that $f'(x^*) < f'_-(x^*)$ and $f(x)$ is concave $\frown$ while $f_-(x)$ is convex $\smile$ (inequality A.24) tells us that $f(x)$ decreases faster than $f_-(x)$ at every point in $(x^*, \eta - \sigma)$ and no additional intersection is possible in $(x^*, \eta - \sigma)$.

- Consider the interval $[x^{s2}, x^{s3})$. We now show that, given an intersection before x^{s2}, there is no additional intersection in $[x^{s2}, x^{s3})$. As a reminder, we have shown that, given an intersection before x^{s2}, $f(x^{s2}) < f_-(x^{s2})$ and $f'(x^{s2}) < f'_-(x^{s2})$. In $[x^{s2}, x^{s3})$, $f(x)$ is decreasing and convex $\smile$ (by the definition of x^{s2}). Similarly, $f_-(x)$ is decreasing and convex $\smile$ (inequality A.24). However, we claim that $f(x)$ decreases faster than $f_-(x)$ at every point in $[x^{s2}, x^{s3})$. This claim together with the fact that $f(x^{s2}) < f_-(x^{s2})$ imply that there is no additional intersection in $[x^{s2}, x^{s3})$, $f(x^{s3}) < f_-(x^{s3})$, and $f'(x^{s3}) < f'_-(x^{s3})$.

  We now argue why the claim that $f(x)$ decreases faster than $f_-(x)$ is true. In $[x^{s2}, x^{s3})$, the contribution of $f_\alpha(x)$ to $f(x)$ is decreasing faster than its contribution to $f_-(x)$ since the interval $[x^{s2}, x^{s3})$ is closer to α than the interval

$(-x^{s3}, -x^{s2}]$ and $f_\alpha(x)$ is convex $\smile$ in both intervals. At the same time, in $[x^{s2}, x^{s3})$, the contribution of $f_\eta(x)$ to $f(x)$ is decreasing while its contribution to $f_-(x)$ is increasing. An additional fact that $f'(x^{s2}) < f_-'(x^{s2})$ leads to the conclusion that $f(x)$ decreases faster than $f_-(x)$ at every point in $[x^{s2}, x^{s3})$ and our claim is justified.

If x^{s3} does not exist, we can conclude that, given an intersection before x^{s2}, there can be no additional intersection in $[x^{s2}, \eta - \sigma)$.

Given that there is no intersection before x^{s2}, we now show that there is at most one intersection in $[x^{s2}, x^{s3})$. As a reminder, we have in this case that $f(x^{s2}) > f_-(x^{s2})$ since $f(0) = f_-(0)$, $f'(0) > 0 > f_-'(0)$, and no intersection has occured before x^{s2}

Let x^* denote the smallest point of intersection, if exists, in $[x^{s2}, x^{s3})$. Since $f(x^{s2}) > f_-(x^{s2})$ and both $f(x)$ and $f_-(x)$ are decreasing, we must have that $f'(x^*) < f_-'(x^*)$.

We argue that $f(x)$ decreases faster than $f_-(x)$ at every point in (x^*, x^{s3}) since $f'(x^*) < f_-'(x^*)$ and we have argued that the overall contribution of $f_\alpha(x)$ and $f_\eta(x)$ to $f(x)$ is decreasing faster than the overall contribution of $f_\alpha(x)$ and $f_\eta(x)$ to $f_-(x)$ in $[x^{s2}, x^{s3})$ $(\supset (x^*, x^{s3}))$. Therefore, there can be no additional intersection in (x^*, x^{s3}). Moreover, given that x^* exists in $[x^{s2}, x^{s3})$, we have that $f(x^{s3}) < f_-(x^{s3})$ and $f'(x^{s3}) < f_-'(x^{s3})$.

Finally, if x^{s3} does not exist, we can let x^* denote the smallest intersection point in $[x^{s2}, \eta - \sigma)$. In this case, we still have that $f'(x^*) < f_-'(x^*)$ and $f(x)$ decreases faster than $f_-(x)$ at every point in $[x^*, \eta - \sigma)$. Therefore, there can be no additional intersection in $(x^*, \eta - \sigma)$.

- Consider the interval $[x^{s3}, x^{s4})$. Similar to the case of interval $[\alpha, x^{s2})$, $f(x)$ is decreasing and concave $\frown$ while $f_-(x)$ is decreasing and convex $\smile$.

  It follows that the proof that there can be no additional intersection in $[x^{s3}, x^{s4})$ given an intersection before x^{s3}, and that there can be at most one intersection in $[x^{s3}, x^{s4})$ given no intersection before x^{s3}, is similar to the proof for the interval $[\alpha, x^{s2})$ and is thus omitted.

- Consider the interval $[x^{s4}, x^{s5})$. Similar to the case of interval $[x^{s4}, x^{s5})$, both $f(x)$ and $f_{-}(x)$ are decreasing and convex $\smile$. However, the overall contribution of $f_{\alpha}(x)$ and $f_{\eta}(x)$ to $f(x)$ is decreasing faster than the overall contribution of $f_{\alpha}(x)$ and $f_{\eta}(x)$ to $f_{-}(x)$.

  It follows that the proof that there can be no additional intersection in $[x^{s4}, x^{s5})$ given an intersection before x^{s4}, and that there can be at most one intersection in $[x^{s4}, x^{s5})$ given no intersection before x^{s4}, is similar to the proof for the interval $[x^{s2}, x^{s3})$ and is thus omitted.

The proofs for the other intervals can be done in the same fashion as in either $[\alpha, x^{s2})$ or $[x^{s2}, x^{s3})$. We conclude that, given an intersection before α, there can be no additional intersection in $[\alpha, \eta - \sigma)$. On the other hand, given no intersection before α, there can be at most one intersection in $[\alpha, \eta - \sigma)$.

Combining the results for the intervals $(0, x^{max}]$, (x^{max}, α), and $[\alpha, \eta - \sigma)$, we conclude that there can be at most one intersection in $(0, \eta - \sigma)$. □

Appendix B

Comments on Scheme 2 of the Attack Detection System

In this section, we shall find the worst case attack scenario under scheme 2 for a particular example in which the SNR is 20 dB, and the attack detection device threshold α is $0.15\sqrt{P}$ $(M = 1)$. The analysis presented in this section is done in parallel with the work in section 3.3.

However, all the results presented in this appendix are not derived theoretically as in some parts of section 3.3. Our approach in solving the corresponding minimization problem in this appendix is based on various observations of the function curves. Our goal is to compare the results obtained for scheme 2 to the results for scheme 1. Establishing the theoretical results in a general case under scheme 2 is the subject of future research.

In what follows, we rely on the same set of notations as in section 3.3.

B.1 Attack at a Single Bit at Two Network Nodes

Finding the worst case scenario in this case is equivalent to solving for $J^{(1)}$ and $J^{(2)}$ in the following optimization problem.

$$\begin{aligned} &\text{minimize} && A(J^{(1)}) + A(J^{(2)}) \text{ for scheme 2} \\ &\text{subject to} && \hat{J} = J^{(1)} + J^{(2)}, \end{aligned} \tag{B.1}$$

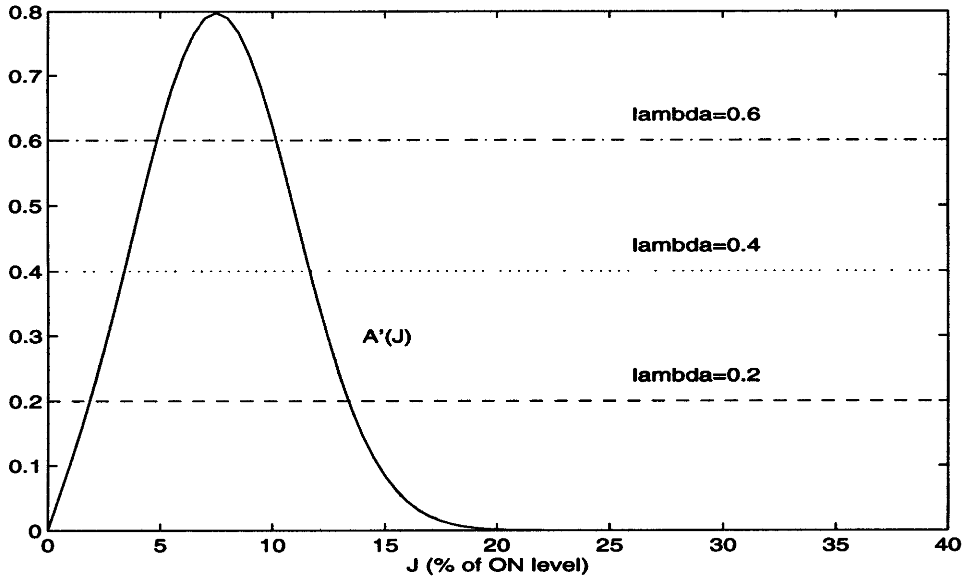


Figure B-1: (Scheme 2) The curve of $A'(J_i)$ superposed on the curves of λ for different values of λ ($\lambda > 0$). Note that there are at most two positive intersection points. In this plot, the SNR is 20 dB, and α is $0.15\sqrt{P}$ (for $M = 1$).

The Lagrangean and its derivatives with respect to $J^{(1)}$ and $J^{(2)}$ are

$$L(J^{(1)}, J^{(2)}, \lambda) = A(J^{(1)}) + A(J^{(2)}) - \lambda(J^{(1)} + J^{(2)} - \hat{J}) \tag{B.2}$$

$$\frac{\partial}{\partial J^{(k)}} L = 0 = A'(J^{(k)}) - \lambda, k = 1, 2, \tag{B.3}$$

Consider the curve of $A'(J)$ superposed on the constant curve λ ($\lambda > 0$) as shown in figure B-1. By observation (without theoretical proof), we can see there are at most two positive intersection points for any positive λ.

As a result, an extremum could correspond to constant jamming with $(J^{(1)}, J^{(2)}) = (\hat{J}/2, \hat{J}/2)$, or sporadic jamming with $(J^{(1)}, J^{(2)}) = (J^a, J^b)$ or (J^b, J^a), where $J^b > J^a$.

We can investigate the nature of these extrema by evaluating the bordered Hessian of the Lagrangean which is

$$|H|_{(J^{(1)},J^{(2)})} = \begin{vmatrix} 0 & -1 & -1 \\ -1 & A''(J^{(1)}) & 0 \\ -1 & 0 & A''(J^{(2)}) \end{vmatrix} = -A''(J^{(1)}) - A''(J^{(2)}). \tag{B.4}$$

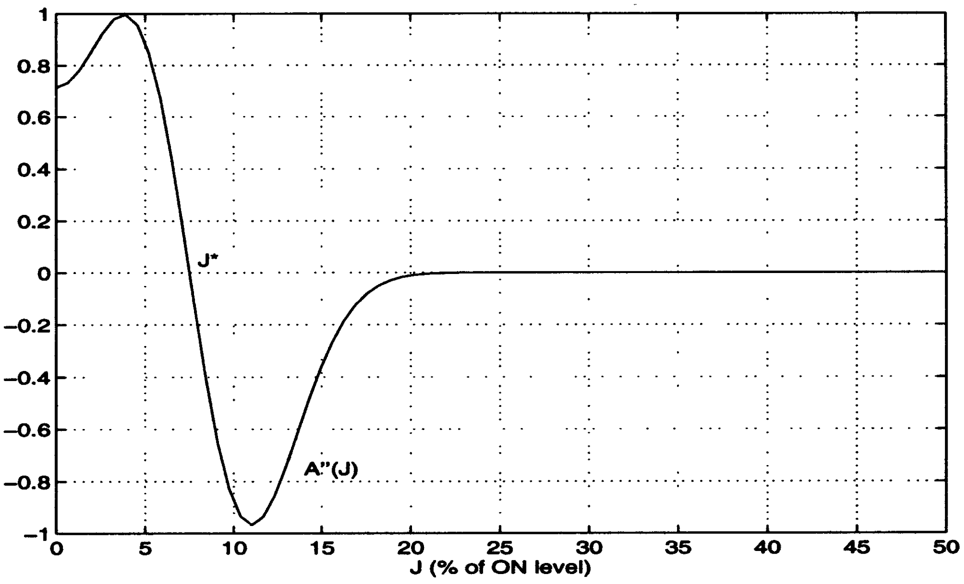


Figure B-2: (Scheme 2) The curve of $A''(J)$. Note that there is a unique positive value J^* such that $A''(J^*) = 0$. In this plot, the SNR is 20 dB, and α is $0.15\sqrt{P}$ (for $M = 1$).

For constant jamming, the bordered Hessian $|H|_{(\hat{J}/2,\hat{J}/2)}$ of the Lagrangean is equal to $-2A''(\hat{J}/2)$. Figure B-2 shows the curve of $A''(J)$. Note that there is a unique value denoted by J^* at which $A''(J^*) = 0$ (lemma 3.9). Observe that $A''(J) > 0$ in $(0, J^*)$ and $A''(J) < 0$ in (J^*, ∞).

It follows that $(\hat{J}/2, \hat{J}/2)$ corresponds to a local minimum for $\hat{J} < 2J^*$ and a local maximum for $\hat{J} > 2J^*$.

We next investigate the extrema of the form (J^a, J^b) and (J^b, J^a). We claim that $J^a + J^b > 2J^*$. To see this, note that the condition $J^a + J^b > 2J^*$ is equivalent to $J^b - J^* > J^* - J^a$. This inequality holds for all values of J^a and J^b if the reflection of $A'(J)$, $J > J^*$ across $J = J^*$ into the region $(0, J^*)$ is above the graph of $A'(J)$ in $(0, J^*)$. Figure B-3 verifies the claim for this example.

Therefore, given that $\hat{J} < 2J^*$, the local minimum $(\hat{J}/2, \hat{J}/2)$ is the only extremum in the constraint set of problem B.1. At either end point of this constraint set, i.e. when $(J^{(1)}, J^{(2)}) = (-\infty, \infty)$ or $(\infty, -\infty)$, we have that the expected alarm rate takes the maximal possible value of 2 since $\lim_{J\to\pm\infty} A(J) = 1$. Therefore, the objective function cannot be minimized in the limit (at the end points). We conclude that $(\hat{J}/2, \hat{J}/2)$ is indeed the global minimum.

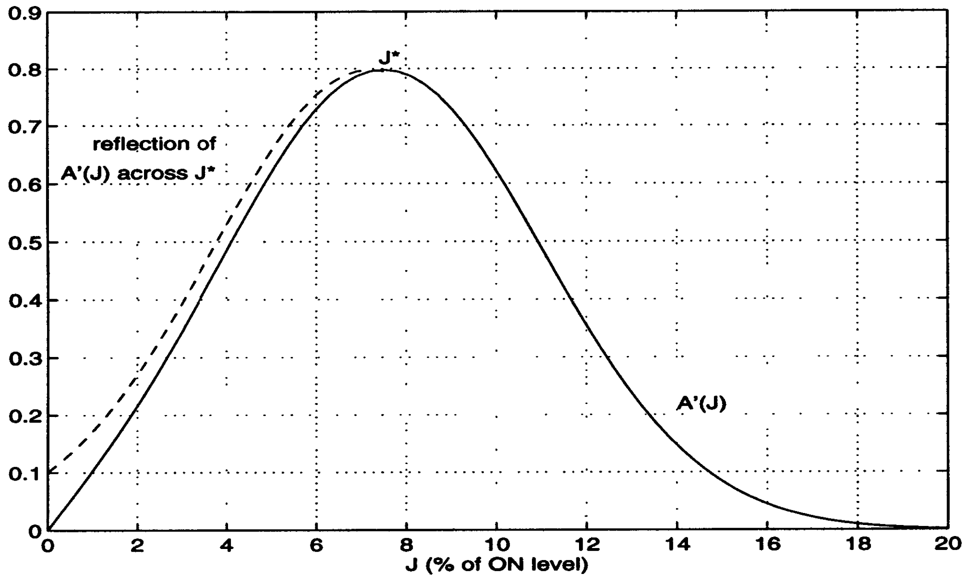


Figure B-3: (Scheme 2) The curve of $A'(J)$ together with its reflection across J^*, where $A''(J^*) = 0$. Note that the reflection of $A'(J)$ lies above $A'(J)$ in the interval $[0, J^*)$.

In summary, given that $\hat{B} < B(2J^*)$, the wost case attack scenario corresponds to the event in which coherent components of jamming signals at two network nodes are equal.

B.2 Attack at a Single Bit at Multiple Network Nodes

In this case, the corresponding optimization problem is

$$\begin{aligned} &\text{minimize} \quad && \sum_{k=1}^{M} A(J^{(k)}) \text{ for scheme 2} \\ &\text{subject to} \quad && \hat{J} = \sum_{k=1}^{M} J^{(k)}. \end{aligned} \tag{B.5}$$

The Lagrangean $L(J^{(1)}, ..., J^{(M)}, \lambda)$ is equal to $\sum_{k=1}^{M} A(J^{(k)}) - \lambda(\sum_{k=1}^{M} J^{(k)} - \hat{J})$. Differentiating the Lagrangean yields the same condition for each $J^{(k)}$ as in the two-node case, namely $A'(J^{(k)}) = \lambda$ (equation B.3). As before, each $J^{(k)}$ can take only one of the two values denoted by J^a and J^b, where $J^b > J^a$.

Using the results from the case of jamming at 1 bit and 2 network nodes (section B.1), we claim that, for $\hat{B} < B(2J^*)$, the worst case scenario corresponds to the event in which coherent components of jamming signals at all M network nodes are equal.

To see that the above claim is true, note that $J^a + J^b > 2J^*$ (see section B.1). It follows that, for $\hat{J} < 2J^*$ (equivalently $B(\hat{J}) < B(2J^*)$), $J^{(k)}$ can take only a single value since the sum of two distinct values of $J^{(k)}$ exceeds $2J^*$. Therefore, there is only one extremum $(\hat{J}/M, ..., \hat{J}/M)$. We now argue that this extremum is indeed the global minimum.

We shall proceed by contradiction. Assume there exists a scenario with jamming signals (coherent components) denoted by $(J^{*(1)}, J^{*(2)}..., J^{*(M)})$ such that not all signals are equal; and this scenario yields the lowest expected device alarm rate. Consider two unequal signals in such a scenario, namely $J^{*(i_1)}$ and $J^{*(i_2)}$. Based on the optimality of equal jamming signals (coherent components) at two network nodes when $J^{*(i_1)} + J^{*(i_2)} < 2J^*$, it follows that there exists a lower expected device alarm rate when both $J^{*(i_1)}$ and $J^{*(i_2)}$ are equal to $\frac{1}{2}J^{*(i_1)} + \frac{1}{2}J^{*(i_2)}$. We thus arrive at a contradiction.

In summary, given that $\hat{B} < B(2J^*)$, the wost case attack scenario corresponds to the event in which coherent components of jamming signals at M network nodes are equal.

In general, the value of J^* $(> \alpha)$ is approximately close to α. For convenience in later analysis, we choose to restrict $\hat{B}$ to be below $B(2\alpha)$ instead of $B(2J^*)$ (as in the analysis for scheme 1 in section 3.3).

B.3 Attack at Multiple Bits at Multiple Network Nodes

Based on the discussion in section 3.3.3, the corresponding optimization problem in this case is given in problem 3.34 which we represent again below for convenience.

$$\begin{aligned} \text{minimize} \quad & \frac{1}{T}\sum_{i=1}^{T} A_M(J_i) \text{ for scheme 2} \\ \text{subject to} \quad & \tilde{B} = \frac{1}{T}\sum_{i=1}^{T} B(J_i) \end{aligned} \tag{B.6}$$

We have argued in section B.2 that, given $J < 2\alpha$, $A_M(J) = MA(J/M)$. In this case, we can express $A'_M(J)$ and $A''_M(J)$ as $A'(J/M)$ and $\frac{1}{M}A''(J/M)$ respectively.

Consider again the optimization problem B.6, the Lagrangean and its derivatives are equal to

$$L(J_1, J_2, \lambda) \quad = \quad \frac{1}{2}[A_M(J_1) + A_M(J_2)] - \lambda\left(\frac{1}{2}[B(J_1) + B(J_2)] - \tilde{B}\right)$$

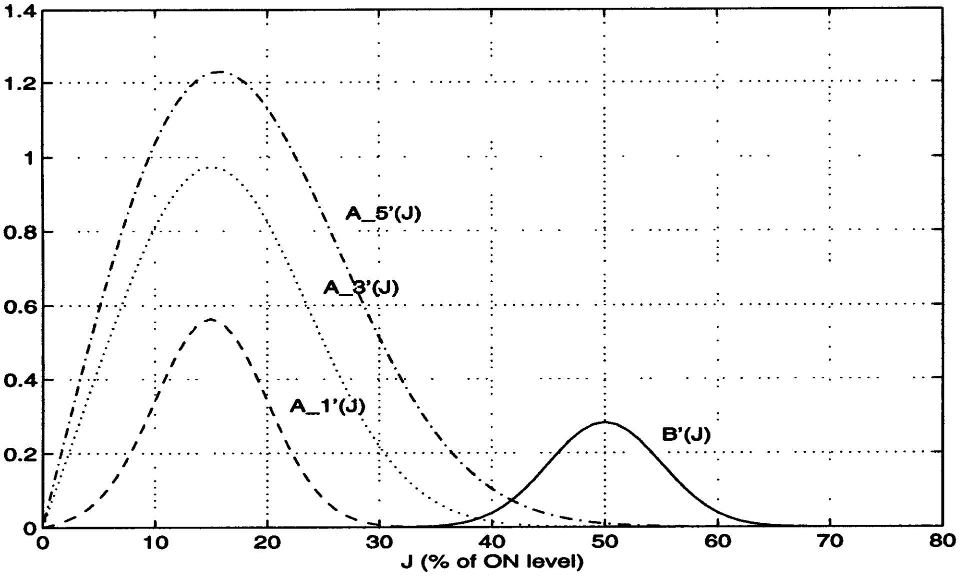


Figure B-4: (Scheme 2) The curve of $B'(J)$ superposed on the curves of $A'_M(J)$ for a few values of M. In the range $(0, \eta - \sqrt{M}\sigma)$, there exists at most one positive intersection point ($\eta - \sqrt{M}\sigma$ is $0.45\sqrt{P}$ in this case). The SNR is 20 dB, and α is $0.15\sqrt{P}$ (for $M = 1$).

$$\frac{\partial}{\partial J_i} L = 0 = A'_M(J_i) - \lambda B'(J_i). \tag{B.7}$$

Figure B-4 shows the curves of $A'_M(J_i)$ for a few values of M superposed on the curve of $B'(J_i)$. Note that 0 is always a solution to $A'_M(J_i) = \lambda B'(J_i)$ since $A'_M(0) = A'(0) = 0$ (equation 3.12) and $B'(0) = 0$ (equation 3.14).

Consider the range $(0, \eta - \sqrt{M}\sigma)$, there exists at most one intersection in $(0, \eta - \sqrt{M}\sigma)$. Thus, given that $\tilde{B} < \frac{1}{2}B(\eta - \sqrt{M}\sigma)$, any extremum must correspond to either constant jamming $(\tilde{J}, \tilde{J})$ or sporadic jamming $(0, J^s)$ (or $(J^s, 0)$).

The bordered Hessian of the extremum of the form $(\tilde{J}, \tilde{J})$ can be expressed as

$$|H|_{(\tilde{J},\tilde{J})} = -\left[\frac{1}{M}A''(\tilde{J}/M) - \frac{A'(\tilde{J}/M)}{B'(\tilde{J})}B''(\tilde{J})\right](B'(\tilde{J}))^2. \tag{B.8}$$

Figure B-5 shows the curves of $|H|_{(\tilde{J},\tilde{J})}$ as given in equation B.8 as a function of $\tilde{J}$ for different values of M. Observe that in all cases, $|H|_{(\tilde{J},\tilde{J})} > 0$ in $(0, \eta - \sqrt{M}\sigma)$ and thus constant jamming corresponds to a local maximal expected device alarm rate.

For the extrema of the form $(J^s, 0)$ or $(0, J^s)$. Their bordered Hessian can be expressed

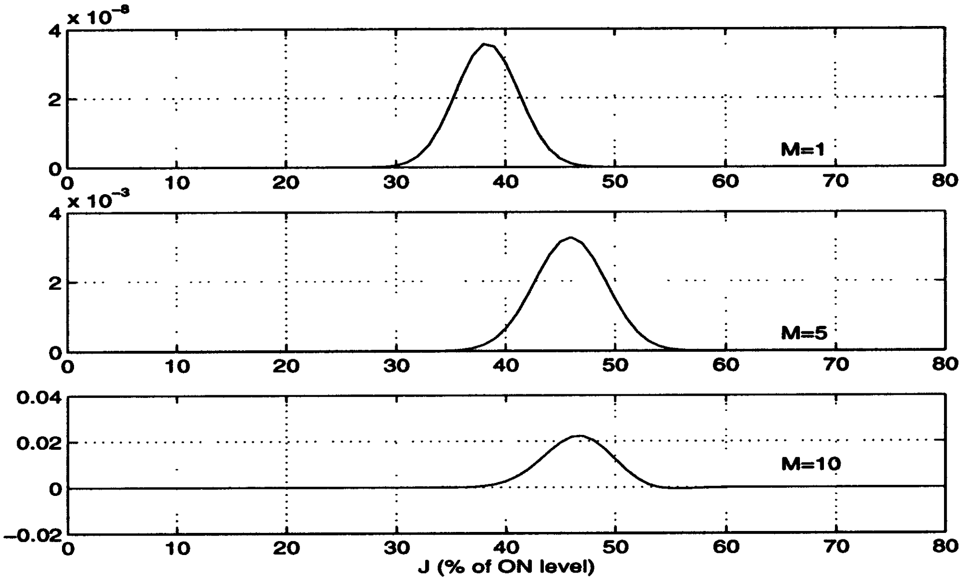


Figure B-5: (Scheme 2) $|H|_{(\tilde{J},\tilde{J})}$ as a function of $\tilde{J}$. In all cases, $|H|_{(\tilde{J},\tilde{J})} > 0$ in $(0, \eta - \sqrt{M}\sigma)$. The SNR is 20 dB, and α is $0.15\sqrt{P}$ (for $M = 1$).

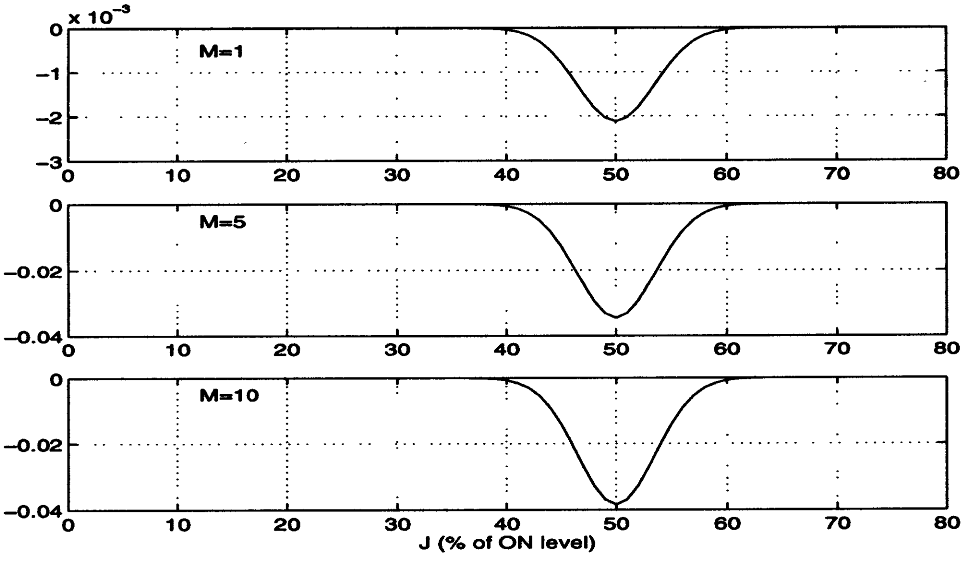


Figure B-6: (Scheme 2) $|H|_{(J^s,0)}$ as a function of J^s. In all cases, $|H|_{(J^s,0)} < 0$ in $(0, \eta - \sqrt{M}\sigma)$. The SNR is 20 dB, and α is $0.15\sqrt{P}$ (for $M = 1$).

as

$$|H|_{(J^s,0)} = -\frac{1}{2}\left[\frac{1}{M}A''(0) - \frac{A'(J^s/M)}{B'(J^s)}B''(0)\right](B'(J^s))^2. \tag{B.9}$$

Figure B-6 shows the curves of $|H|_{(J^s,0)}$ as given in equation B.9 for different values of M. Observe that $|H|_{(J^s,0)} < 0$ in $(0, \eta - \sqrt{M}\sigma)$ and thus sporadic jamming corresponds to a local minimal expected device alarm rate.

Since there is no other extremum and since the constraint set in problem B.6 has no boundary (note that this constraint set is the same as the one defined by equation 3.39 and we have argued in section 3.2.1 that it has no boundary), sporadic jamming with $(J^s, 0)$ or $(0, J^s)$ yields the smallest expected device alarm rate and thus corresponds to the worst case attack scenario.

We can extend the result to the case when $T > 2$ using the same arguments given in the proof of proposition 3.2 in section 3.2.2. To do so, we must make sure that all the arguments given in proof of proposition 3.2 are still valid. Notice that as long as the desired average BER is less than $\frac{1}{T}B(\eta - \sqrt{M}\sigma)$, we are in the region that all the arguments in the proof of proposition 3.2 can still be applied.

Therefore, given the device alarm rate function $A_M(J_i) = MA(J_i/M)$, for jamming at M network nodes and at T bits, as long as $\tilde{B} < \frac{1}{T}B(\eta - \sqrt{M}\sigma)$, the worst case attack scenario corresponds to the event in which jamming occurs at a single bit.

To make sure that our alarm rate function is indeed $A_M(J_i) = MA(J_i/M)$, we must guarantee that $J_i < 2\alpha$ (section B.2). Therefore, given that $\tilde{B} < \frac{1}{T}B(\xi)$, where $\xi = \min(2\alpha, \eta - \sqrt{M}\sigma)$, the worst case attack scenario corresponds to the event in which jamming occurs at a single bit, and coherent components of jamming signals at M network nodes at that particular bit are equal.

In comparison to the results for scheme 1, we can see that the restriction on the value of degraded BER (denoted by $\tilde{B}$) is tighter for scheme 2 than for scheme 1. In both cases, the restriction on $\tilde{B}$ can be expressed as $\tilde{B} < \frac{1}{T}B(\xi)$. The values of ξ are given below.

$$\xi = \begin{cases} \min(M\alpha, \eta - \sqrt{M}\sigma) & \text{for scheme 1} \\ \min(2\alpha, \eta - \sqrt{M}\sigma) & \text{for scheme 2} \end{cases} \tag{B.10}$$

Bibliography

[1] S.R. Chinn, EDFA simulation results, Lincoln Laboratory, 1997.

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

[3] E. Desurvire, *Erbium-Doped Fiber Amplifiers*, New York: Wiley & Sons, 1994.

[4] R.H. Eng, A.A. Lazar, W. Wang, "A Probabilistic Approach to Fault Diagnosis in Linear Lightwave Networks," *IEEE Journal on Selected Areas in Communications*, vol. 11, no. 9, December 1993, pp. 1438-1448.

[5] R. Erkander, "Optical Fibre Security System ZAT 4," Ericsson Review, Vol. 67, no. 1, 1986, pp. 35-41.

[6] S.G. Finn and R.A. Barry, "Optical Servies in Future Broadband Networks," *IEEE Network*, November/December 1996, vol. 10, no. 6, pp. 7-13.

[7] R.G. Gallager and D.P. Bertsekas, *Data Networks*, New Jersey: Prentice Hall, 1992.

[8] A.H. Gnauck et al., "One Terabit/s Transmission Experiment," OFC Postdeadline Paper PD 20, February 1996.

[9] P.E. Green, *Fiber Optic Networks*, Englewood Cliffs, New Jersey: Prentice Hall, 1993.

[10] B.R. Hemenway et al., "Demonstration of a Reconfigurable Wavelength Routed Network at 1.14 Tbits/s", OFC 97 Postdeadline Paper PD 26, February 1997.

[11] I.P. Kaminow et al., "A Precompetitive Consortium on Wide-band All Optical Networks," *IEEE JSAC*, vol. 14, no. 5, June 1996, pp. 780-99.

[12] I. Katzela, G. Ellinas, T.E. Stern, "Fault Diagnosis in the Linear Lightwave Network," LEOS Summer Topical Meetings, August 7-11, 1995, pp. 41-42.

[13] Y.W. Lai, Y.K. Chen, W.I. Way, "Novel Supervisory Technique Using Wavelength-Division-Multiplexed OTDR in EDFA Repeatered Transmission Systems," *IEEE Photonics Technology Letters*, vol. 6, no. 3, March 1994, pp. 446-451.

[14] E.A. Lee and D.G. Messerschmitt, *Digital Communication*, Massachusetts: Kluwer, 1994.

[15] J.E. Marsden and A.J. Tromba, *Vector Calculus*, New York: Freeman and Company, 1988.

[16] M. Médard, D. Marquis, S.R. Chinn, "Attack Detection Methods for All-Optical Networks," the Internet Society's Symposium on Network and Distributed System Security (NDSS), March 1998.

[17] M. Médard, S.R. Chinn, P. Saengudomlert, "Attack Detection in All-Optical Networks," OFC 98 Technical Digest paper ThD4, February 1998.

[18] M. Médard, R. Bergman and S. Chan, "Distributed Algorithms for Attack Localization in All-Optical Networks," the Internet Society's Symposium on Network and Distributed System Security (NDSS), March 1998.

[19] M. Médard and S.R. Chinn, "Proposed Methods of Detection for Security Attacks," Lincoln Laboratory, 1997.

[20] M. Médard, D. Marquis, R.A. Barry, S.G. Finn , "Security Issues in All-Opical Networks," *IEEE Network*, May/June 1997, pp. 42-48.

[21] M. Médard and D. Marquis, "A Taxonomy of Vulnerabilities of All-Optical Networks to Nefarious Attack," Lincoln Laboratory, 1997.

[22] H. Onaka et al., "1.1 Tb/s WDM Transmission over a 150 km 1.3 mm Zero-Dispersion Single-Mode Fiber," OFC 96 Postdeadline paper PD 19, February 1996.

[23] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, NY: McGraw-Hill, 1991.

[24] N. Schroff, M. Schwartz, *Fault Detection/Identification in the Linear Lightwave Network*, CU/CTR/TR 243-91-24, Columbia University, 1991.

[25] M. Sumida, "OTDR Performance Enhancement Using a Quaternary FSK Modulated Probe and Coherent Detection," *IEEE Photonics Technology Letters*, vol. 7, no. 3, March 1995, pp. 336-338.

[26] F. Tosco, *Fiber Optic Communications Handbook*, New York: McGraw Hill, 1990, pp. 237-241.

[27] A.V. Yakovlev, "An Optical-Fiber System for Transmitting Confidential Information," *Telecommunication and Radio Engineering*, vol. 10, no. 4, 1995.

[28] //www.ll.mit.edu/aon/