Information Warfare and Electronic Warfare Systems

RICHARD A. POISEL

Information Warfare and Electronic Warfare Systems

For a complete listing of titles in the
Artech House Electronic Warfare Library,
turn to the back of this book.

Information Warfare and Electronic Warfare Systems

Richard A. Poisel

ARTECH
HOUSE
BOSTON | LONDON
artechhouse.com

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the U.S. Library of Congress.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library.

Cover design by Vicki Kane

ISBN 13: 978-1-60807-705-2

685 Canton Street
Norwood, MA 02062

10 9 8 7 6 5 4 3 2 1

To Debbie

Contents

Preface

Information warfare/information operations (IW/IO) has evolved as an approach to bring information technologies (IT) to the battlefield. It has been suggested that it is the next evolutionary step in warfighting, breaking away from the industrial age notions surrounding the massing of power; instead, the massing of effects is used. Information superiority is the underlying tenant in these thought processes. The force that can dominate information can win on the modern battlefield.

Electronic warfare (EW) is one of the five legs of IO. The others are: computer network operations (CNO), psychological operations (PSYOPS), military deception (MILDEC), and operations security (OPSEC). CNO approaches IO as attacks (passive and active) against computer networks. PSYOPS addresses changing the attitudes and views of people, both military and civilian. Deception also targets the minds of people, primarily adversarial commanders. It attempts to create believable, but untrue, situations. OPSEC employs methods to keep an adversary unaware of the actual friendly situation. Here we focus on EW and how EW systems and principles can be employed in the IW/IO discipline.

EW includes all aspects of electronic systems that radiate electromagnetic (EM) waves in some sense. As such, it is a fairly broad area. We focus here on communications EW, however. We consider how well EW systems operate against an adversary's attempts to exchange information.

The taxonomy of EW includes three separate but related areas: (1) electronic support (ES), (2) electronic attack (EA), and (3) electronic protection (EP). ES is the domain of noncooperatively intercepting communication signals. ES provides combat information; that is, information that a commander can use immediately to make decisions, as well as information for intelligence generation. The distinction here is to what use is the collected information put and what is the amount of processing requried to generate the desired product.

EA is the domain of actively attacking a communication network by inserting EM radiation into an adversary's networks with the intent of denying the successful exchange of information across that network. ES provides support to EA by providing steerage to targets in some cases.

EP is protecting friendly communication systems from the ES and EA activities of an adversary targeting those friendly communications. We will not focus much on EP here, except to the extent that friendly ES and EA activities can provide EP.

The targeted audience for the material presented here includes technical engineering personnel who are either new to the EW field or are practicing professionals who would like a different view of evaluating EW system performance. A bachelor's degree in engineering is generally required with a working knowledge of linear system theory. The latter includes linear mathematical systems that are usually considered within the domain of matrix theory. Some of the material, in particular that in Chapter 2, is useful for understanding the characteristics of information as a discipline and is relatively nontechnical in nature. A working knowledge of probability theory is also useful; however, the necessary fundamental material is introduced when needed.

This book is structured as follows. We begin with an introduction to IW and EW systems in Chapter 1. The purpose of this chapter is to lay a common foundational understanding of the military service's views of IW.

In Chapter 2 we explore the nature of information by examining some of the basic characteristics. The observe, orient, decide, act (OODA) loop developed by Boyd is introduced as a fundamental model of the decision-making process. The three domains of conflict are discussed, including the fundamental primitives within the three domains.

Tenets of information theory are examined in Chapter 3. The foundations of probability theory that are necessary for the remainder of the material are introduced. Information entropy is presented as the fundamental measure of the amount of information in messages. Common channel models that are widely used in information theory are discussed, as well as the capacity of some common modern communication systems. In particular, the broadcast channel is introduced, which forms the basis for later evaluation of EW system performance in Chapter 9.

An evolving model of IW is discussed in Chapter 4. This model, developed by Kopp and Borden and based on Shannon's theory of information, is used to develop a taxonomy of the attributes of IW. Four canonical forms of IW approaches are developed using this model. We propose adding ES to these canonical forms for a complete analysis of EW systems, as it is generally necessary to use for collecting information so that the four can be applied.

In Chapter 5 we explore how EW systems interrelate with network-centric operations (NCO). The contributions of EW systems to the modern battlefield are presented, with a particular focus on what they bring to situation awareness. The flow diagrams of the fundamentals of EW system operation are discussed, to include ES, EA, and EW target analysis and EW intelligence analysis. The information provided by EW systems is considered along with the information required by EW systems to perform their mission. Data and information fusion are reviewed.

One of the major contributors to effective NCO is the ability to communicate among the several battlespace facilities. We discuss the fundamentals of networking in Chapter 6 and introduce mobile ad hoc networks (MANETs).

Examples of modern MANET protocols are presented and some of their basic characteristics are examined. Security has arisen as one of the principal challenges of MANET protocols, so considerable discussion is included on this important topic. The chapter concludes with a review of EW attacks on MANETs.

One of the major contributions of EW systems to NCO is the information provided to situation assessment, the latter of which leads to situation awareness. We discuss these principles in Chapter 7 and show how the data fusion levels fit into the picture for situation assessment.

An introduction to EW system architectures is provided in Chapter 8. This chapter is a review of ES and EA configurations and how EW systems are designed to provide the information discussed in other chapters.

The theoretical performance of several EW system configurations are presented in Chapter 9. The bases for the analysis presented are Shannon's information theory and the modern concepts of multiple-input multiple-output (MIMO) antenna systems. Analysis of wiretap channels, the basis of ES system evaluation, is presented. Likewise, the arbitrarily varying channels (AVC) are evaluated as the basis for EA system performance. Combinations of wire-tap channels and AVCs are evaluated and the performance compared. Considerations are provided for cooperative and noncooperative ES and EA activities.

Finally, we present the results of computer simulation of some EW system architectures in Chapter 10. An engineering (technical) simulation was performed and conclusions were reached for several architectures. That was followed by an operational simulation to evaluate some representative EW system configurations. We consider sophisticated EW system configurations (thick systems) as well simpler configurations (thin systems). Two different scenarios are considered. One is a scenario in Northeast Asia and the other is in urban terrain, such as in a large city. Both thin and thick system performance are examined for each scenario. The advantages of the various EW system configurations are discussed and recommendations are provided.

The material in the book is presented largely from the point of view of land-mobile forces (Army and Marines). That is a bias of the author, as that is his background. The material is applicable to other scenarios, however. The principal tenets apply to all cases when EW systems are employed.

As always, errors tend to creep into works such as this. Although considerable effort was expended to minimize such mistakes, no claim is made as to their non-existence. In any case, the author has assumed full responsibility when they occur. Constructive feedback is welcome whenever an error is found or recommendations for positive changes are provided.

Chapter 1

Introduction to Information Warfare and Electronic Warfare Systems

1.1 Introduction

Communication systems are usually designed to ensure reliable transmission of information despite physical impairments of the channel (e.g., thermal noise in a receiver or fading and interference in a wireless media). However, many non-military and almost all tactical military communication scenarios require that information be protected against *electronic warfare* (EW) actions.

However, EW activities, especially at the tactical edge, can be used effectively to reduce the effectiveness of an adversary. We will discuss several of these below.

Adversarial parties in secure communications problems can essentially be categorized into two main groups: passive adversaries that intercept and overhear the transmission without any corrupting effects, and active adversaries that manipulate the message or the transmission media. We call the former group eavesdroppers or wiretappers and the latter group jammers.

We consider a class of information protection problems that simultaneously require confidentiality from eavesdropping and integrity from jamming, without focusing on distinguishing identities. For purposes of EW analysis, we determine how vulnerable these functions are to specific EW approaches. As illustrated in Figure 1.1, we investigate the fundamental limits of secure communications subject to these requirements for problems in which an eavesdropper may be wiretapping the channel and a jammer also may be tampering with the channel. Scenarios are included in which the two adversaries take actions independently, that is, they are not cooperating with each other.

We use an information theoretic approach to the problem that facilitates the consideration of computationally unlimited adversaries based on computational

complexity. The *arbitrarily varying wire-tap channel* (AVWTC) is used to model this scenario. The AVWTC combines elements from the wiretap channel model and the arbitrarily varying channel model, both of which are reviewed in Chapter 9. This model consists of a family of wiretap channels dictated by some state that is selected by the jammer in an arbitrary and time-varying manner (as far as the communication link is concerned) and that is unknown to the transmitter and the target receiver. The goal is to determine the privacy capacity of this channel, that is, the maximum rate at which the legitimate users can communicate guaranteeing the reliability of information in the presence of the jammer and its privacy with respect to the eavesdropper. In so doing, we ascertain the effectiveness of the EW approaches.

We concentrate heavily on EW for land-forces concepts here. This notwithstanding, the underlying tenets apply to EW for all military forces and nonmilitary networking and information exchange/processing.

1.2 Global Information Grid

Since mid-1990s, the U.S. military has centered its vision of future forces on the promises of building lighter, leaner, and more lethal forces using the tenets of *network centric warfare* (NCW) (*network-enabled warfare* in the United Kingdom and elsewhere), also known as *network-centric operations* (NCO). By using *information technology* (IT) to connect sensors, shooters, and *decision makers* (DMs) together in a common framework, a military force can achieve rapid, concurrent discovery of enemy activities and dispositions [1]. This information superiority allows U.S. forces to achieve full spectrum dominance over any opposing force [2].

The information superiority that fuels today's lighter, more lethal military forces is increasingly vulnerable to the innovative use of commercially available IT by network-oriented (versus hierarchically structured) adversaries, however.

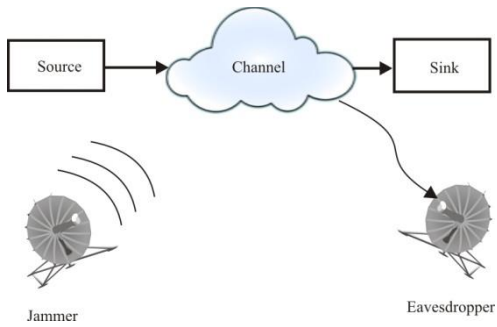


Figure 1.1 General block diagram of the problem of information protection in the presence of a non-cooperative jammer and an eavesdropper.

The Army's *Brigade Combat Team* (BCT) utilizes [3]: "an advanced network architecture that will enable levels of joint connectivity, situational awareness and understanding, and synchronized operations heretofore unachievable."

The DoD is responsible for the largest NCO-related project, the core enabling network of networks itself, the *global information grid* (GIG). The GIG is designed to provide the so called "entry fee" for NCO, the densely interconnected, ultra-high bandwidth, highly reliable information infrastructure, or "infostructure" into which the NCO systems will tie [4]. However, the GIG is primarily focused on providing a long-haul, fixed, high bandwidth, secure backbone for military networking and communications. In order to truly achieve the goals of NCO, all the individual, generally mobile, warfighting entities—tanks, aircraft, *unattended aerial systems* (UASs), soldiers, unattended sensors, indirect fires systems, and *command and control* (C2) assets—must be integrated into the grid.

The Army' backbone architecture is LandWarNet [5].

> LandWarNet is the Army's portion of the global information grid and consists of all globally interconnected, end-to-end Army information capabilities supporting warfighters, policy makers, and support personnel. As the Army's enterprise system of systems, LandWarNet moves information through a seamless network that facilitates information-enabled joint warfighting and supporting operations from the operational base to the edge of tactical formations, down to the individual Soldier.
>
> LandWarNet provides the construct for the Army's transition to the future and is a key contributor for information and decision superiority. LandWarNet will enable voice, data, and video to the edge of tactical formations—ultimately pushing these capabilities lower to our modular U.S. Army's brigades, battalions, and Soldiers. The Future Combat Systems will have a wide array of new information capabilities to achieve conceptual objectives; however, it must be able to pass that information to a variety of organizations with dissimilar levels of capability. LandWarNet is the means to provide linkages between sensors, shooters, and leaders; seamless and secure interoperability, network services; and, end-to-end connectivity throughout the enterprise.

For the Navy, FORCEnet is the path to force integration for the future [6].

> FORCEnet is the operational construct and architectural framework for Naval Warfare in the Information Age which integrates warriors, sensors, networks, command and control, platforms and weapons into a networked, distributed combat force, scalable across the spectrum of conflict from seabed to space and sea to land.
>
> The goal of FORCEnet is to arm U.S. forces with superior knowledge, leading to increased combat power. In pursuit of this goal, FORCEnet will provide a comprehensive network of sensors, analysis tools, and decision aids to support the full array of naval activities, from combat operations to logistics and personnel development. The focused, timely, and accurate data delivered by FORCEnet will help leaders at every level by allowing them to draw on vast amounts of information and share the resultant understanding. This will increase the joint force's ability to synchronize activities throughout the battle space to achieve the greatest impact.

The Air Force [7]:

> The Air Force's contribution to the overarching concept for warfighting operations is the C2 Constellation — the Air Force's components to the GIG. The C2 Constellation is a family of C4ISR systems sharing horizontally and vertically integrated information through machine-to-machine conversations enabled by a peer-based network of sensors, command centers and shooters. Both an operational construct and an architectural framework, it guides our development of people, processes and technology toward network-centric operations.
>
> Key network-centric operation elements of the C2 Constellation include the various platforms and sensors the Air Force provides to the Joint Force Commander and key programs that support command centers such as the Air and Space Operations Center and the Distributed Common Ground Segment. Underpinning programs within the AOC, such as the Theater Battle Management Core System already serve as the joint standard for air operations planning and execution, and we are continuing to migrate these systems to a more modern, web-enabled architecture.

The GIG is not a superficial integration that aims to merely pass simple voice communications and a smattering of digital data. Truly NCO-compliant integration will feature a densely linked network of networks with high bandwidth and sufficient *quality of service* (QoS) to provide a common view of the battlespace to all network nodes, especially the tactical edge nodes of all the services.

We can thus see that the U.S. military is heavily committed to constructing future forces that rely significantly on high bandwidth communications and information to conduct future operations.

1.3 Networks

1.3.1 Operational and Strategic

At the operational and strategic echelons, the GIG network consists mostly of satellite and high bandwidth terrestrial (optical fiber) elements. These capabilities can be put into place and remain relatively static. At the tactical level, networks are less so, and many must be mobile in order to interconnect moving and movable battlespace nodes.

1.3.2 Tactical

If the technology challenges can be worked out, interconnecting the edge nodes will rely upon *mobile adhoc networking* (MANET) technologies. It is necessary to understand the capabilities required to develop and field NCO-enabled forces in order to determine the specific technological requirements for developing the

MANETs that will connect the tactical edge of the military enterprise: the sensors, enablers, and shooters that will perform the military mission.[1] Because of its underpinning importance for NCO, we provide a brief discussion of MANET technology in Chapter 6.

1.4 Information and Information Theory

In Shannon's taxonomy of communications, "content" is separated from the means of communication. The three levels of the communications problem [8] can be expressed as:

- Level A: How accurately can the symbols of communication be transmitted? (The technical problem.)
- Level B: How precisely do the transmitted symbols convey the desired meaning? (The semantic problem.)
- Level C: How effectively does the received meaning affect conduct in the desired way? (The effectiveness problem.)

The technical field of information theory addresses Level A. Levels B and C are considered totally outside this field. EW concerns address both levels A and B, but are only concerned about Level C through effects on the first two levels. We discuss EW effectiveness on Levels A and B throughout this book, but in general we do not combine them into the same conversation—they remain separate topics in the analyses.

We present in Chapter 2 a discussion of the characteristics of information—that is, Level C in the taxonomy. Beginning in Chapter 3, and encompassing the remainder of the book, we delve into some detail on Level A. The effectiveness of EW systems is evaluated based on the premises inherent in Level A of information theory.

1.4.1 Network-Centric Operations Background and Characteristics

The concept of NCO has been known by many different names—network-centric warfare (NCW), cyber war, *command and control* (C2) warfare, and so forth—each with slightly different, but highly overlapping definitions. The many names reflect the struggle of military and strategic thinkers to fully understand the significance of the shift in warfighting doctrine that this philosophy represents.

The concept and origins of NCO are generally credited to Arthur Cebrowski (Vadm, USN, dec.) and John Garstka in the U.S. DoD [9]. They promoted the

[1] The *edge* is meant to refer to power to the edge where the soldiers/sailors/airmen are located and certainly where EW assets are located.

military's evolution from platform-centric to network-centric forces as an inevitable outgrowth of the western countrie's economic and societal evolution of the shift from industrial age to information age philosophies, processes, and tools. IT has altered the business and economic environment by providing ubiquitous international communications, low-cost, high-power computer processing, cheap, high-volume data storage, a proliferation of sensors, and advanced software capabilities that collectively provide precise, readily available information on the operating environment.

The next step is away from stand-alone, powerful, thick computing nodes into what carries the appellation "cloud computing." The user interface in this environment is a simple, thin interface tool (interface to the Internet) with the powerful computers and databases centralized in few locations.

Of course, just having IT tools is not enough. To fully utilize the advantages that IT can provide, an enterprise must also possess the appropriate culture, organizational structure, and set of processes to effectively use these tools to obtain a competitive advantage. With precise information on market demand, inventory levels, commodity prices and availability, and visibility into manufacturing capacity, businesses are now utilizing IT to rapidly adapt to changes in their ecosystems in order to obtain an advantage in their markets [10]. That is, a business with the right organization, processes, and IT-fueled tools can achieve the ultimate competitive advantage: agility.

Agility can be defined as the ability to adapt quickly, but in a sure-footed manner [11]. For the military, agility of forces, organization, resources, and C2 are the fundamental attributes that information age forces must strive to achieve. The NCO cognoscenti propose that the most effective and efficient means to enable agility is the establishment of shared awareness and full collaboration among all the entities in an organization [12]. SA and collaboration require robust communications and rapid exchange of data via one or more networks. The complete networking of battlespace entities is the key enabler to achieving these effects and is the cornerstone of NCO—thus the work network in the title.

These entities will exhibit the NCO characteristics of speed of command, massing of effects, cooperative engagement, high tempo and responsiveness, and self-synchronization to a degree that cannot be matched by any non-NCO capable opponent [13]. Although effectively implementing IT is a central tenet of NCO, clearly it is not solely a technologically driven phenomenon. The processes (doctrine), organizational structure, and culture of an organization are critical enablers to the proper utilization of the tremendous tools presented by the on-going trends in IT of increased processing power, smaller form factors, and lower costs. To proceed with the implementation of NCO by simply pursuing the technology without major changes in the other aspects in the *doctrine, organization, training, materiel, leadership and education, personnel, and facilities* (DOTMLPF) would be futile—a waste of time. However, in the end it is the network that defines NCO, and therefore understanding how to best design,

implement, and protect these networks is the critical materiel element of the NCO transformation.

1.5 Electronic Warfare and NCO

We will cover this topic in greater depth in Chapter 5. Here we provide an outline of the association of EW with NCO.

The U.S. Army defines EW as follows [14]:

> *Electronic warfare* is defined as military action involving the use of electromagnetic and directed energy to control the electromagnetic spectrum or to attack the enemy. Electronic warfare consists of three divisions: electronic attack, electronic protection, and electronic warfare support (JP 3-13.1).
>
> **Electronic Attack**
>
> - *Electronic attack* is a division of electronic warfare involving the use of electromagnetic energy, directed energy, or antiradiation weapons to attack personnel, facilities, or equipment with the intent of degrading, neutralizing, or destroying enemy combat capability and is considered a form of fires (JP 3-13.1). Electronic attack includes—
> - Actions taken to prevent or reduce an enemy's effective use of the electromagnetic spectrum, such as jamming and electromagnetic deception.
> - Employment of weapons that use either electromagnetic or directed energy as their primary destructive mechanism (lasers, radio frequency weapons, particle beams).
> - Offensive and defensive activities including countermeasures.
> - Common types of electronic attack include spot, barrage, and sweep electromagnetic jamming. Electronic attack actions also include various electromagnetic deception techniques such as false target or duplicate target generation.
> - *Directed energy* is an umbrella term covering technologies that relate to the production of a beam of concentrated electromagnetic energy or atomic or subatomic particles (JP 1-02). A directed-energy weapon uses directed energy primarily as a direct means to damage or destroy an enemy's equipment, facilities, and personnel. In addition to destructive effects, directed-energy weapon systems support area denial and crowd control.
> - Examples of offensive electronic attack include—
> - Jamming enemy radar or electronic command and control systems.
> - Using antiradiation missiles to suppress enemy air defenses (antiradiation weapons use radiated energy emitted from the target as their mechanism for guidance onto targeted emitters).
> - Using electronic deception techniques to confuse enemy intelligence, surveillance, and reconnaissance systems.
> - Using directed-energy weapons to disable an enemy's equipment or capability.
> - Defensive electronic attack uses the electromagnetic spectrum to protect personnel, facilities, capabilities, and equipment. Examples include self-protection and other protection measures such as use of expendables (flares and active decoys), jammers, towed decoys, directed-energy infrared countermeasure

systems, and counter-radio-controlled improvised-explosive-device systems. (See JP 3-13.1 for more discussion of electronic attack.)

Electronic Protection

- *Electronic protection* is a division of electronic warfare involving actions taken to protect personnel, facilities, and equipment from any effects of friendly or enemy use of the electromagnetic spectrum that degrade, neutralize, or destroy friendly combat capability (JP 3-13.1). For example, electronic protection includes actions taken to ensure friendly use of the electromagnetic spectrum, such as frequency agility in a radio, or variable pulse repetition frequency in radar. Electronic protection should not be confused with self-protection. Both defensive electronic attack and electronic protection protect personnel, facilities, capabilities, and equipment. However, electronic protection protects from the effects of electronic attack (friendly and enemy), while defensive electronic attack primarily protects against lethal attacks by denying enemy use of the electromagnetic spectrum to guide or trigger weapons.
- During operations, electronic protection includes, but is not limited to, the application of training and procedures for countering enemy electronic attack. Army commanders and forces understand the threat and vulnerability of friendly electronic equipment to enemy electronic attack and take appropriate actions to safeguard friendly combat capability from exploitation and attack. Electronic protection measures minimize the enemy's ability to conduct electronic warfare support (discussed below) and electronic attack operations successfully against friendly forces. To protect friendly combat capabilities, units—
    - Regularly brief force personnel on the EW threat.
    - Ensure that electronic system capabilities are safeguarded during exercises, workups, and predeployment training.
    - Coordinate and deconflict electromagnetic spectrum usage.
    - Provide training during routine home station planning and training activities on appropriate electronic protection active and passive measures.
    - Take appropriate actions to minimize the vulnerability of friendly receivers to enemy jamming (such as reduced power, brevity of transmissions, and directional antennas).
- Electronic protection also includes spectrum management. The spectrum manager works for the G-6 or S-6 and plays a key role in the coordination and deconfliction of spectrum resources allocated to the force. Spectrum managers or their direct representatives participate in the planning for EW operations.
- The development and acquisition of communications and electronic systems includes electronic protection requirements to clarify performance parameters. Army forces design their equipment to limit inherent vulnerabilities. If electronic attack vulnerabilities are detected, then units must review these programs. (See DODI 4650.01 for information on the spectrum certification process and electromagnetic compatibility.)

Electronic Warfare Support

- *Electronic warfare support* is a division of electronic warfare involving actions tasked by, or under the direct control of, an operational commander to search for, intercept, identify, and locate or localize sources of intentional and unintentional radiated electromagnetic energy for the purpose of immediate threat recognition, targeting, planning, and conduct of future operations (JP 3-13.1).

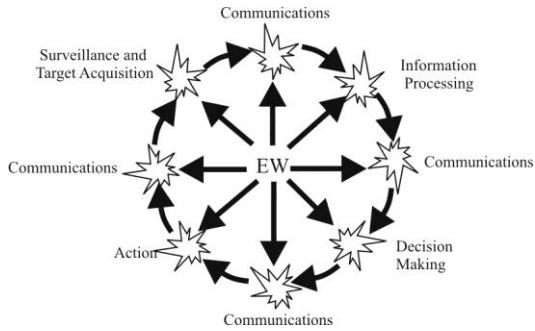


Figure 1.2 Command and control and EW. (From [15], © Artech House, 2001. Reprinted with permission.)

- Electronic warfare support systems are a source of information for immediate decisions involving electronic attack, electronic protection, avoidance, targeting, and other tactical employments of forces. Electronic warfare support systems collect data and produce information or intelligence to—
    - Corroborate other sources of information or intelligence.
    - Conduct or direct electronic attack operations.
    - Initiate self-protection measures.
    - Task weapon systems.
        - Support electronic protection efforts.
        - Create or update EW databases.
        - Support information tasks.
- Electronic warfare support and signals intelligence missions use the same resources. The two differ in the detected information's intended use, the degree of analytical effort expended, the detail of information provided, and the time lines required. Like tactical signals intelligence, electronic warfare support missions respond to the immediate requirements of a tactical commander. Signals intelligence above the tactical level is under the operational control of the National Security Agency and directly supports the overarching national security mission. Resources that collect tactical-level electronic warfare support data can simultaneously collect national-level signals intelligence. See FM 2-0 for more information on signals intelligence.

We will limit our concern here to electronic attack (EA) and electronic support (ES). More detailed information about EP can be found in [15].

In summary, EW is warfare conducted in the *radio frequency* (RF) domain. It is the collection of information about an adversary by the noncooperative intercept of RF emissions (ES). It is also the actions taken to preclude an adversary from effectively using their electronic systems (EA).

EW contributes to the counter C2 process. Figure 1.2 illustrates some of the areas to which EW can contribute. Surveillance and target acquisition assets must communicate their detections to an information processing center. The results of this analysis must be communicated to the appropriate DMs. Likewise, the decisions must be turned into actions via communications to the effectors. Thus, the C2 cycle continues and EW assets can intervene at any point where communications are required.

The loop shown in Figure 1.2 is very similar to the *observe, orient, decision, action* (OODA) loop we introduce in in Section 2.3. The communication functions are explicitly shown in Figure 1.2, however.

1.5.1 EW and Networking

There are two aspects when considering NCO and EW systems: (1) the opponent is using NCS concepts in their operations, and (2) friendly forces are using NCO concepts. Of course, both could be true as well. In (1), the analysis is therefore to consider EW against networked forces. In (2), friendly EW systems are networked with the rest of the friendly IT systems, and what are the considerations there?

While considering aspects of friendly NCO is an interesting subject, much has been written about it already [16–18]. We will concentrate herein mostly on the first aspect of NCO.

1.5.1.1 Friendly Use of NCO

Tactical and strategic EW systems have been networked essentially since the first one was deployed. If real-time geolocation is required, two or more ES systems must be used against the target at the same time; otherwise, obtaining directional information to the same target cannot be assured. This requires a datalink of some kind, also known as networking. Tasking and reporting into and out of EW systems via networks have been the mainstay for several years.

So the concept is not new to EW systems. What is new is the interconnection of the EW systems to other nodes directly and not through an analysis center. This can produce information for several nodes that heretofore were not privy to that information. Furthermore, EW systems can receive information not previously available.

Post Then Analyze

One of the oft-quoted aspects of NCO is "post then analyze" rather than "analyze then post." The idea is to share information as soon as it becomes available, irrespective of whether it is for sure true.

In classical EW operations any single intercept is usually only a small part of the overall picture and usually must be verified by other means. To be sure, there are some instances when this is not true, and post then analyze concepts might apply. For example, if one of the tasks of the EW system is to locate and identify a particular CP, for example, and that information is obvious from the intercept, that piece of information might qualify for post then analyze. However, even then, deception is possible, so subsequent verification after analysis is probably in order.

Most single intercepts make little sense on their own since they are usually only a snippet of information. It is through verification and repeated intercepts followed by analysis for which most ES information is used.

ES information can be, and is, used for real-time steerage of EA assets. When this is the case, usually the ES and EA assets are collocated. We provide analyses in Chapter 9 of cases when this is not true. When these functions are separated, however, real-time coordination between the two functions is difficult.

Collaboration

Collaboration between two or more nodes is one of the advantages of NCO. Operators in several locations can work together to solve a problem using the network. This is a function that comes new with NCO (at least for short-time problems). We discuss this subject is finer detail in Chapter 6.

Agility

Agility is defined as the ability to change directions/positions rapidly due to changing conditions. That is, rapidly adapting to the environment to improve friendly conditions. As mentioned previously, agility is one of the principal goals of NCO. NCO brings with it the capability to be agile—at least more agile than the opponent. We should point out, however, that the goal of NCO is to make the teeth (infantry, armor, and so forth) more agile, not necessarily the EW system. It is when the maneuver elements can rapidly adapt that the goals of NCO are met. An exception to this is when the EA assets are integrated into the weapons mix (and they should always be). EA capabilities in this case must be agile as well.

Self-Synchronization

Self-synchronization is the capability to be aware of and observe the existing conditions, particularly the geometrical relationships of the adversary vis-à-vis the friendly forces, and adjusting friendly behavior to support/enhance the friendly tactical position. Ideally this is accomplished without specific guidance to the effect from higher echelons.

ES produces information that can contribute to a more accurate and timely *common operating picture* (COP) at all levels, but particularly at the edge, in the red zone. Again it is important to point out that the idea is to facilitate the teeth to self-synchronize, not the EW system.

Massing Effects

At the edge is where the action can be fast and furious and timely targeted EA can be a significant force multiplier in several ways. The effects produced by EA

include inducing confusion and delaying C2 in opposition force operations. Call for support and/or fires can be precluded for long enough a time for friendly forces to maneuver into favorable positions.

The idea is to allow the teeth to mass effects, not the EA system, although EA is a fires function that should be included in the analysis. EA is part of the teeth. By targeting specific C2 nodes at critical times (during maneuvering, for example), critical information exchange is denied, potentially causing confusion and denial of effective command.

EW operations facilitate full spectrum domination. Denying the *forward observer* (FO) to artillery *fire direction center* (FDC) communication link precludes adjustments in fires. *Pulse position information* (PPI) data cannot be exchanged between a tracking radar and its associated weapon system.

ES can provide information on intentions and contributes to generation and maintenance of the COP. Targets can be identified and located with ES assets. ES can be used to confirm targets identified by other sensors.

As we will discuss in Chapter 5, fusion of information in the forward area reduces the amount of data that otherwise overwhelms the data analysis centers and DMs at all echelons.

Cooperative Engagement

When combining the effects of agile EA with other forms of fires in realtime, combat multiplier effects ensue. When direct attack weapons, such as tanks, are combined with indirect attack weapons, such as artillery and EW systems, the best of each can be taken advantage of. Calling for reserves, for example, can be denied to an adversary allowing the kinetic weapons to maximize their effects.

1.5.1.2 NCO Communication Architectures

Typically, high-capacity trunk networks have been used to provide operational and tactical communications at EAB as illustrated in Figure 1.3. These links were point-to-point for the most part and were almost always encrypted. For years they have used *low probability of intercept* (LPI) technology for EP.

More modern trunked networks use Internet technology, such as that illustrated in Figure 1.4. The *radio access points* (RAPs) are the interfaces of the headquarters indicated into the tactical Internet cloud. *Combat net radio* (CNR) interfaces to the cloud are also provided to form the tactical Internet services at the edge.

Ad Hoc Networks

Mobile ad hoc networks (MANETs) are a form of network where all users on the networks potentially serve as routers for all other users. The networks are self-

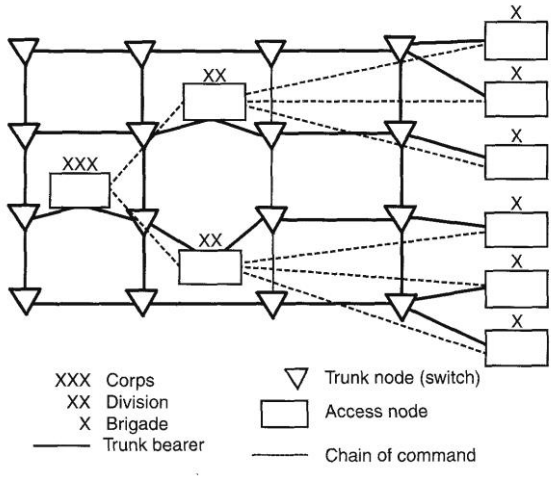


Figure 1.3 Traditional trunked network services form a mesh. (From: [15], © Artech House, 2001. Reprinted with permission.)

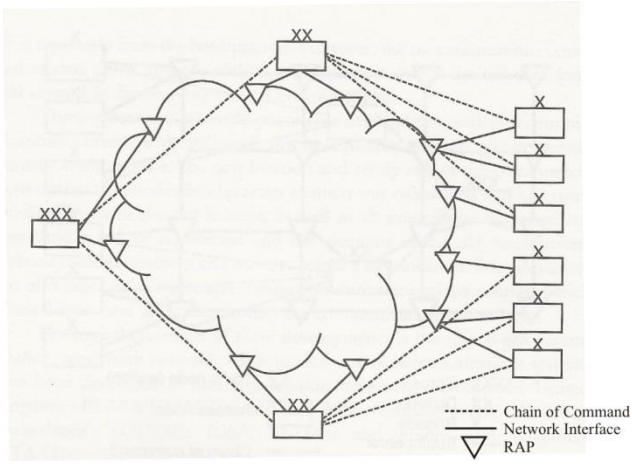


Figure 1.4 Modern trunked networks form a trunked cloud. (From: [15], © Artech House, 2001. Reprinted with permission.)

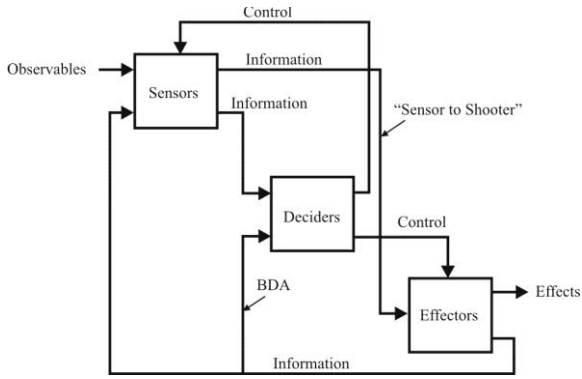


Figure 1.5 Logical model of warfare in the information age. (Source: [4].)

managed and self-healing, with users dropping out and entering frequently. There is considerable administrative overhead in such networks, as each node must maintain information about all the networks to which it is to communicate. We discuss ad hoc network technology in Chapter 6.

1.6 EW Systems

The Alberts logical model of information age warfare, shown in Figure 1.5, has as one of its three blocks, the effectors of the actions taken by the DM [19]. These effectors are responsible for causing the desired effects on an adversarial force. Kinetic and explosive weapons are certainly effectors but there are more effectors than these. EW/EA systems can be considered effectors and communication EW/EA can be considered indirect fire weapons.

We delve into the elements of EW systems in Chapter 8.

1.6.1 ES Systems

The topic of primary concern here is how communications EW/ES systems are analyzed, designed, and used. An EW/ES system is one of the sensors in Figure 1.5 and can be used for several purposes. In general, an ES system searches the RF spectrum for energy looking for targets of interest. In many cases the frequencies of these targets are known ahead of time and searching for which ones are active is referred to as *directed search*. There may be 10 or more of such frequencies assigned to each operator in an ES system. When not all of the frequencies are known but must be found by scanning the RF spectrum, this is referred to as *general search*.

In general search, some amount of intercept is necessary to ascertain whether the signal is a *signal of interest* (SOI). For example, the tasking may be to locate artillery nodes, so signals must be monitored to determine whether they are associated with an artillery network.

One of the critical functions of an ES system is geolocation of the targets. Usually this requires two or more systems to be linked together, typically with a datalink. Each system computes a *line of position* (LOP) of the target and where these LOPs intersect is an estimate of the target location. This technique is known as *triangulation*, and the more LOPs available to use, the more accurate the geoposition estimate is. Alternately, if the ES system is airborne, determining LOPs when the aircraft has moved some distance yields triangulation results as well. There are other forms of determining target locations than triangulation, but it is a popular method. For the reader interested in more depth into geoposition calculations, [20] is recommended.

There are a plethora of operational considerations when employing ES systems. For example, the coverage range of ground systems is typically substantially less than that of an airborne sensor. We discuss some of these in Chapter 8.

ES systems produce information, for the generation of combat information, the intelligence process, as well as for steerage of jammers. Their contribution to the COP generation process is to locate emitters and attempt to identify who the unit is that is attempting to communicate. Measuring external signal parameters[2] is undertaken to assist in this process. Intentions can also sometimes be gleaned from the intercept of C2 communications. They also generate BDA for the effectiveness of jammer activity. They belong to the "sensors" category in Figure 1.5.

1.6.2 EA Systems

Jammers (EA system) are consumers of information and are in the "effectors" category. They can be considered indirect fire weapons, just as artillery is an indirect fires weapon. An EA system works by injecting energy into an adversary's receiver of sufficient strength to overcome the intended signal at the receiver. There are numerous way to accomplish this and we will discuss the principal ones in Chapter 8.

The targets for communication EA are generally C2 nodes although new entries into the target list are IEDs and other roadside bombs. These weapons are frequently controlled remotely by RF devices such as cellular phones and garage door openers.

Proper coordination between the ES and EA processes is essential for effective use of EA assets. Furthermore, EA activities must be fully integrated into the battle planning process if its effectiveness is to be maximized.

[2] Externals parameters consist of signal characteristics that can be measured automatically such as baud rate for digital signals and modulation type (e.g., AM, FM, PM).

1.7 Concluding Remarks

This book is structured as follows. This introductory chapter is followed by a fairly detailed introduction to information and *information warfare* (IW) in Chapter 2. That is followed by Chapter 3, which introduces the fundamental concepts in information theory, the approach that we use in Chapter 8 to evaluate the performance of EW systems. Included in Chapter 3 is an introduction to the most common types of channel models. Chapter 4 introduces a useful model for IW. It can be used to understand some of the basic characteristics of the fundamentals of IW. Chapter 5 introduces EW systems and how they fit into the IW paradigm. Chapter 6 discusses the basics of networking, including an introduction to the Internet and mobile networking concepts. The key aspects of MANET communication systems are discussed, the key communication technology that makes NCW work. One of the fundamental principles of NCO is the sharing of a common picture of the battlespace among all active participants (scaled, of course, to their individual needs). This common picture displays the current situation, and is the basis for *situation assessment* (SA). Chapter 7 presents the key aspects of SA. Chapter 8 contains an overview of the salient characteristics and elements that comprise EW systems. Chapter 9, one of the more important chapters in the book, contains theoretical analyses of EW system performance, based on the tenets of information theory introduced in Chapter 3. Chapter 10 presents the results of engineering and operational simulations of EA systems in typical scenarios.

References

[1] Deakin, R. S., *Battlespace Technologies: Network-Enabled Information Dominance*, Norwood, MA, Artech House, 2010.
[2] CJCS, "Joint Vision 2010." (Joint Staff) and CJCS, "Joint Vision 2020." (Joint Staff), 2010.
[3] U.S. Army. "Future Combat System (Brigade Combat Team (FCS(BCT)) 14+1+1 Systems Overview." (March 14, 2007) p. 2.
[4] Alberts, D. S., J. J. Garstka, and F. P. Stein. *Network Centric Warfare*, Washington D.C.: Command and Control Research Program, 2003, p. 187.
[5] LandWarNet 2015, The United States Army's Concept of Operations, February 11, 2008.
[6] www.globalsecurity.org/military/systems/ship/systems/forcenet.htm.
[7] http://www.globalsecurity.org/military/systems/aircraft/systems/c2-constellation.htm.
[8] Shannon, C. E., and W. Weaver, *The Mathematical Theory of Communication*, Urbana, IL: University of Illinois Press, 1963.
[9] Cebrowski, A. K., and J. J. Garstka. "Network-Centric Warfare: Its Origin and Future." *Naval Institute Proceedings.* 1998.
[10] Cebrowski, A. K., and J. J. Garstka. "Network-Centric Warfare: Its Origin and Future." *Naval Institute Proceedings*,1998, p. 28.

[11] Alberts, D. S., and R. E. Hayes. *Power to the Edge, Information Age Transformation*, Washington D.C.: Command and Control Research Program, 2004, p. 125.
[12] Alberts, D. S., and R. E. Hayes. *Power to the Edge, Information Age Transformation*, Washington D.C.: Command and Control Research Program, 2004, p. 127.
[13] Alberts, D. S., J. J. Garstka, and Frederick P. Stein. *Network Centric Warfare*, Washington D.C.: Command and Control Research Program, 2003, pp. 157–183.
[14] U.A. Army Field Manual 3-36 (February 2009), The Pentagon, Washington, D.C.
[15] Frater, M. R., and M. Ryan, *Electronic Warfare for the Digitized Battlefield*, Norwood, MA: Artech House, 2001, Ch. 3.
[16] Alberts, D. S., *Understanding Command and Control*, Washington, D.C.: CCRP Publication Series, 2006.
[17] Alberts, D. S., J. J. Garstka, R. E. Hayes, and D. A. Signori, *Understanding Information Age Warfare*, Washington, D.C.: CCRP Publication Series, 2001.
[18] Alberts, D. S., J. J. Gartstka, and F. P. Stein, *Network Centric Warfare*, Washington, D.C.: CCRP Publication Series, 1999.
[19] Alberts, D. S., J. J. Garstka, and F.P. Stein, *Network Centric Warfare*, Washington, D.C.: CCRP Publication Series, 1999, pp. 87–114.
[20] Poisel, R. A., *Electronic Warfare Target Location Methods*, 2nd Ed., Norwood, MA: Artech House, 2012.

Chapter 2

Information and Information Operations

2.1 Introduction

Information is developed in the military setting so that a DM can make decisions. This process generally involves developing and understanding the situation through SA. We will delve much more into these topics later. The key point to understand now is that these processes are based on the availability of information. Therefore, we investigate some of the basic characteristics of information in this chapter.

As we discussed in Chapter 1, Shannon separated the content (meaning) of communications from the means of communications (the channel). This chapter addresses some of the attributes of the meaning of communications in a military context. We delve into the means of communications (the channel) in the next chapter.

The remaining chapters of this book investigate many aspects of information and how it is processed in a military setting. Information is key to most military actions, so we begin our journey by investigating the nature of information itself. Later we focus specifically on one (broad) aspect of processing information—that due to EW operations. In particular we examine what impacts ES and EA have on the channel itself.

This chapter is structured as follows. We first examine several attributes of information and define their meaning. This allows us to start with a common baseline of definitions. We then introduce the elements of IO and IW. This is followed with the introduction to the *observe, orient, decide, action* (OODA) loop and cognitive hierarchy. The three domains of conflict, as defined by the *Command and Control Research Program* (CCRP), are then introduced, including discussions on the military information environment and the meaning of information advantage/superiority.

2.2 Information

2.2.1 The Importance of Information to Warfare

The importance of information in warfare goes without further elaboration. What is relatively new is how information is processed and used. Great amounts of battlespace information are routinely available due to recent leaps in information processing technology (primarily computers and communications).

The importance of information to success in war has been understood from ancient times. For example, Sun Tzu (circa 300 BC) wrote, [1–4]:

- Warfare is the art of deceit. Therefore, when able, seem to be unable; when ready, seem unready; when nearby, seem far away; and when far away, seem near. If the enemy seeks some advantage, entice him with it. If he is in disorder, attack him and take him. If he is formidable, prepare against him. If he is strong, evade him. If he is incensed, provoke him. If he is humble, encourage his arrogance. If he is rested, wear him down. If he is internally harmonious, sow divisiveness in his ranks. Attack where he is not prepared; go by way of places where it would never occur to him you would go.
- The ultimate skill in taking up a strategic position is to have no form. If your position is formless, the most carefully concealed spies will not be able to get a look at it, and the wisest counselors will not be able to lay plans against it.
- Unless you know the intentions of the rulers of the neighboring states, you cannot enter into preparatory alliances with them; unless you know the lay of the land its mountains and forests, its passes and natural hazards, its wetland and swamps—you cannot deploy the army on it; unless you can employ local scouts, you cannot turn the terrain to your advantage.
- [A major military operation is a severe drain on the nation.] Two sides will quarrel with each other for several years in order to fight a decisive battle on a single day. If, begrudging the outlay of ranks, emoluments, and a hundred pieces of gold, a commander does not know the enemy's situation, his is the height of inhumanity. Such a person is no man's commander, no ruler's counselor, and not master of victory.

  Thus the reason the farsighted ruler and his superior commander conquer the enemy at every move, and achieve successes far beyond the reach of the common crowd, is foreknowledge. Such foreknowledge cannot be had from ghosts and spirits, educed by comparison with past events, or verified by astrological calculations. It must come from people—people who know the enemy's situation.

2.2.2 Information Sources

Information in warfare has always been present as long as wars have been fought. Sources of information in a military setting include:

- Intelligence: the gathering of information for the purpose of gleaning the status of an adversary.

- Surveillance: the oversight of a region for the purpose of detecting the use of that region by adversarial forces. This oversight is not necessarily by directl viewing of that region by a person, but may include many different types of sensors for detecting the presence of adversarial entities.
- Reconnaissance: the exploring an area to gain information about enemy forces or environmental features for later analysis and examination.
- Weather: gleaned from the observation of weather sensors (radars, ionosonds, weather balloons, and so forth).
- Geographic: information gained from observing maps (paper, electronic, or otherwise).
- Other.

2.2.3 Information Attributes

This section lists many general attributes of information. The importance of a given attribute depends on the manner in which the information network is implemented. For example, relevance and clarity may have less importance when unit-level commanders pull information from the net, compared with when the network pushes information onto commanders.

Many of the properties listed below have two aspects, one that is intrinsic to the piece of information and always applies and one that depends on the context: the use to which the information is being put, or the relationship of the piece of information in question to other pieces.

Fewell and Hazen list 13 attributes of information that are relevant to decision-making [5]:

- Relevance;
- Clarity;
- Timeliness;
- Age;
- Currency;
- Accuracy;
- Consistency;
- Completeness;
- Comprehensibility;
- Secrecy/security;
- Authenticity;
- Value;
- Degree of interoperability.

We add an additional attribute of brevity to this list.

These attributes are examined further in the following paragraphs. Since our primary interests deal with EW, these attributes are discussed here in the context of moving information—a network.

2.2.3.1 Characteristics of Information

Relevance

Relevance is the extent to which an item of information applies to the situation at hand. Each individual piece of information has an intrinsic relevance on its own in a given scenario, but this may be altered, in either direction, by subsequent information. That is, several pieces of information may together have a relevance different (more or less, in either direction) from that of any one of them alone or any subset of the group.

Relevance is an inherently binary quantity—it seems that a piece of information is either relevant or irrelevant—but uncertainty leads to interpreting relevance as a continuous variable. For example, consider a piece of information that is relevant in situation A and irrelevant in situation B, but the commander does not yet know which situation exists. If the two situations are equally likely, then it makes sense to say that the piece of information has a relevance of 0.5; if situation A has probability 0.1, then the relevance of the information is 0.1. That is, relevance can be a probability-weighted mean.

Clarity

Clarity reflects the degree of clutter, that is, the extent to which a relevant item of information is obscured by a plethora of irrelevant items. It is a contextual property.

Timeliness

Timeliness reflects the difference between the time at which an item of information is required and the time at which it is available. Time of availability is an intrinsic property, while time at which the piece of information is required is a contextual property.

Age

Age, an intrinsic property, is the time since the item of information was created or last updated and when it is used.

Currency

Currency is the contextual property that relates to the time when a piece of information becomes so outdated that it can no longer reliably be used in any analysis.

Accuracy

When an element of information is created, it has an intrinsic accuracy determined by the characteristics of its source. Contextual accuracy is the level of uncertainty in the recipient's mind of the degree to which the item of information corresponds to truth, based on what else is known or believed at the time. By taking into account a recipient's beliefs, we include the effects of circumstances that may cause the recipient to suspect that the accuracy of an item of information is more or less than its intrinsic accuracy.

Consistency

Consistency is the degree to which a new item of information agrees with previous items, or with the local COP. This is clearly a contextual property. Lack of consistency is one of the reasons a DM may downgrade the contextual accuracy of an item of information. This raises the complicated question of the processes by which misconceptions in a local COP can be corrected by additional information. Information and logic are not necessarily monotonic—increased data may not lead to increased information. A new element of information may very well contradict the existing epistemology of the DM.

Completeness

Completeness means the extent to which all the required elements of an item of information are available. Completeness is an intrinsic property acquired by information when it is created; it degrades as the information passes along unreliable communications channels or is held in storage locations that are less than 100% reliable. It could also change with time, as new items of information are received that changes the estimate of the situation.

Comprehensibility

Comprehensibility, or understandability, is the ease with which the recipient can fuse the item of information into the local COP. It is an intrinsic property, despite its definition by way of a context. Low comprehensibility may mean that the information does not make sense to the recipient.

Secrecy/Security

Intrinsic secrecy is a function of the source of information and the security of the communication channel over which the information was transferred; it is degraded by transmission through a channel that is not 100% secure. The contextual aspect to secrecy concerns the extent to which the recipient suspects that an adversary may have intercepted the information.

While it has always been true that keeping military information secure is important, with the proliferation of automation services and digital communications on the modern battlefield, as well as in the commercial sector, making this happen has become extremely difficult. Attempts to obtain military sensitive information occur every day.

The four fundamental tenets to information security are confidentiality, integrity, availability, and authenticity. Without any one, or in fact all of them, information can become unreliable, untrustworthy, and uncertain.

Confidentiality

Confidentiality means, at the core of the concept, that the data is hidden from those who are not supposed to see it. Confidentiality can be achieved in a number of ways. These methods are complementary. First, strong authentication for any access to data is required. Second, strict access controls are required. In communications only the sender and intended recipient should be able to access the data. In file systems and data repositories, only the creator and intended users can access the data. Third, the data should be encrypted, both when communicated and when stored, so that it cannot be intercepted, and cannot be accessed during transmission, transportation, and storage. Encryption is frequently what is thought of first when considering confidentiality. While encrypting data is surely a way of keeping it confidential, it is not the only way.

Integrity

Integrity as a concept means that there is resistance to alteration or substitution of data, and/or that such changes are detected and provable. Integrity, in IT vernacular, means that data remains unchanged while stored or transmitted. Unauthorized changes to stored data violate integrity. The information should only be changed by an authorized agent. Once in place, changes to the data should only be possible if the change is authorized.

On the modern battlefield, as in modern business, enormous amounts of information are created, transmitted, and stored daily. When the information is digital, ensuring integrity usually involves the use of checksums, one-way hashes, or other algorithmic validation of the data. Whether the data might be changed by accident or intention, as in EW activities, preventing that change is the foremost

concern and detecting if it has changed is secondary. Integrity can be maintained at many levels, from the hardware all the way to the application.

Verifying and Retaining Integrity. Computational techniques for verifying data integrity include: comparisons, checksums, *message authentication and integrity codes* (MAC/MIC), and message digests such as the *Message Digest* 5 (MD5) hash. The MD5 hash is a mathematical algorithm that produces a unique 128-bit number (a hash) created from the data input. If even one bit of data changes, the hash value will change. An example of this in use is that most open source programs and packages are distributed along with an MD5 hash. Before installing, the recipient can generate the MD5 hash and compare it with the (known good) hash provided by the source. If the generated and provided hashes are not the same, the program or package has been changed.

Availability

For data to be of use to us, it has to be accessible when and where we need it. In the ubiquitous Internet and wireless access era, information must be available 24/7, or whenever it is needed. All the effort spent securing data from unauthorized access or integrity failures may go to waste.

High availability solutions, including load balancing, fail-over, and quick backup and restoration are all involved. These topics are network and systems architecture concerns, operations concerns, and not truly a primary security component.

Authenticity

Authenticity is the extent to which the recipient of information can verify that it has come from the purported source unaltered, as opposed to having been subjected to adversary IO. This is an extreme aspect of secrecy; its degradation requires not only that adversaries can access the network, but also that they can alter information.

Intrinsic authenticity may be conveyed by trusted signatures, where they exist (this is fairly common on the Internet). As before, contextual authenticity refers to doubts that may be raised in the mind of a recipient. Degraded contextual authenticity does not require that the adversary actually was successful in altering an item of information; it is frequently enough for the DM to suspect that it may have happened.

Authenticity is assurance that a message, transaction, or other exchange of information is from the source it claims to be from. Authenticity involves proof of identity. At first glance it might seem that Authenticity is included in the concept of Integrity. Integrity is more specifically about the content of the data itself. Authenticity means that when I get a message from Bob, it is verifiably Bob who

sent it. The message is of no value if Bob did not send it. So, Authenticity involves assurance that the data was created or sent by the source from which it appears.

Value

The value of information is usually defined as the extent to which possession of an item of information enables the recipient to perform more effectively. With this definition, value is a higher-level characteristic than the others in this section—more of a measure of system performance than a measure of performance since it depends on most, perhaps all, of the other properties of information. Value is a contextual property; but the context there is more physical than cognitive.

The definition has been interpreted as implying that a particular piece of information is of little value to a force if the force would probably win regardless of the availability of the information, or would probably lose even if it receives the information.

Degree of Interoperability

Degree of interoperability is also a higher characteristic in the hierarchy, where the distinction between information-flow interoperability and information-usage interoperability is discussed. The second is the high-level property; here we refer to the first. As regards to information flow, degree of interoperability means the efficiency with which a DM can obtain information from, or provide information to, the network. It is a contextual property in that it depends not only on the characteristics of the piece of information, but also on the nature of the recipient.

Brevity

Brevity refers to the degree to which an item of information contains only the level of detail required. Superfluous amounts of information frequently confuse the understanding and implications of the information. Yet too little content also has a detrimental effect.

2.2.4 EW and Its Effects on Information

We discuss EW at length throughout the rest of the book. It is informative, or at least motivational, at this juncture to examine some of the effects EW operations can have on these attributes of information.

We can see that simply denying the transfer of information between two points affects all of the attributes mentioned since no information is exchanged. However, by judiciously applying EW, totally denying the exchange of information is frequently not necessary.

The timeliness of information can be affected by EA by delaying its delivery. The speed at which information is made available is slowed by degrading the channel, as we will see in Chapter 9. Less network speed is available on degraded channels, and in some cases, can be reduced to zero.

Similarly the accuracy of the information can be affected by deleting critical information from the message and inserting false information.

When parts of an information exchange are denied, the completeness of pieces of information can be affected, even though some information gets through. A similar effect on the comprehensibility occurs when some of the information is missing.

Clearly ES can affect the security of information. The very function of ES activities is to glean information from messages to: (1) ascertain the direction an adversary is following, (2) generate the COP, (3) determine the priority nodes for EA activities, (4) separating important nodes from those less important, (5) updating the *electronic order of battle* (EOB),[1] (6) verifying target detections by other sensors, (7) *battle damage assessment* (BDA),[2] target development and target identification, and (8) queuing other sensors, to mention some uses.

2.3 OODA Loop and Cognitive Hierarchy

The OODA loop and *cognitive hierarchy* are shown in Figure 2.1. The OODA loop is a model of the human thought process and begins with the process of observing the environment [6]. Based on this observation, a human categorizes what is being observed with their epistemology and decides on what action(s) is (are) appropriate to respond to the situation. That action is taken, and in so doing, the environment changes or responds in some way. That change is observed and the loop continues.

There is both experimental and theoretical evidence that there are two general types of decisions that are made: *analytical* and *recognition-primed* [7]. The first of these is based on analysis of the data at hand, generating options based on this data, and selecting the optimum decision from the options thus analyzed. The second of these decision types is based on recognition of the situation at hand based on similarity to previous situations. The former requires substantially more cognitive effort than the latter, the processing of which can be immediate. It is estimated that over 90% of tactical decisions are recognition-primed [5]. This further supports the notion that military forces require extensive and continuous training. Such training exposes decision-makers at all levels to as many tactical situations as possible, allowing for later recognition-primed decision-making.

---

[1] The EOB is a graphical description of the layout of electronic systems in the battlespace, and it can be red, blue, or both.

[2] BDA is an evaluation of the damage inflicted by some battlefield action(s).

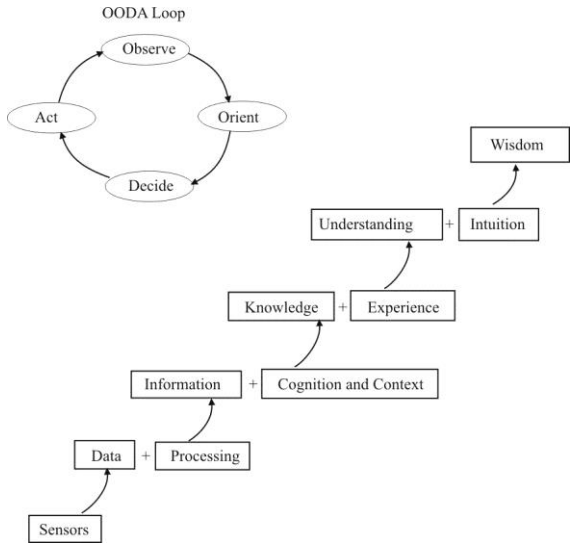


Figure 2.1 The OODA loop and cognitive hierarchy.

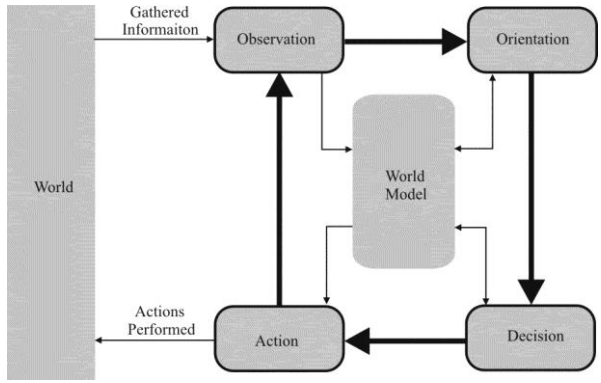


Figure 2.2 Boyd's OODA loop model.

Orientation is the most cerebral of the nodes in that it requires the use of intellect and thought. On the other hand, decide is probably the most visceral of the nodes. Once the DM thinks he/she understands the situation, training and experience take over for the most part, and the decision is made on the course of action.

2.3.1 The OODA Loop Model

Brumley, Kopp, and Korb describe a useful model that includes the OODA loop as the basis of decision-making [8]. The OODA loop model was initially developed by Boyd [6] to understand the decision-making process of fighter pilots in aerial combat. However, it can be generalized to model decision-making in any field. Its name comes from the four steps of the loop—Observation, Orientation, Decision, and Action. A representation of the OODA loop can be seen in Figure 2.2.

The loop starts with the DM observing his or her environment. The DM uses information receivers to collect information about the current state of their environment. Information receivers may include eyes, ears, noses, video cameras, microphones, or pressure sensors, and may measure properties such as light, temperature, sound, pressure, or vibration. The observed information and the information receivers are both potential targets for IW attacks.

Once the individual has collected the new state information, it is analyzed during the Orientation step. During this process the new information is interpreted in the context of its existing knowledge, before updating the model of the world. The model now reflects the understanding of the current state of the world. Boyd [6] states that the individual combines the new information, previous experience, cultural traditions, genetic heritage, and analysis and synthesis methods to update the model. Since existing knowledge affects the interpretation of new information,

individuals who possess differing models of the world can develop different interpretations of the same event. Boyd stresses the importance of the Orientation step, describing it as the "schwerpunkt," or focal point, of the OODA loop model [6].

Once the individual has updated its model of the world, he or she can decide which actions to perform. With an updated model of the world, the individual can see how the actions will affect the future state of the world. These changes can be assessed in terms of their benefit and penalty. If a rational decision-making method is used, then actions will be chosen that maximize the expected benefits and minimize the expected penalties. The OODA loop model does not restrict the individual's decision-making methods in any way.

Finally, during the Action step, the chosen actions are executed. Actions typically affect the state of the world, by manipulating objects in the world or communicating information to others. Changes to the state of the world are observable by others. Individuals may also partially or completely fail to perform their intended actions correctly.

The OODA loop models a feedback loop between the DM and the environment. Actions alter the state of the world, which is then perceived during future iterations of the loop. Advantages over opponents can be gained by "operating inside their OODA loop"—deciding and acting faster than they can [6].

2.3.1.1 Orientation

Brumley, Clayton, and Kopp make the case that the orient node is the most analytical [8]. The Orientation step is an integration step where new information and existing knowledge are combined to produce an updated model of the world [6]. There are four main tasks that an individual performs during the Orientation step (see Figure 2.3). The first task is recognizing previously observed objects, events, and relationships, which allows an individual to retrieve existing information about these elements. The second is to analyze the new information with previously known processing methods to produce an updated model of the world. The model of the world stores the individual's perception of the past, present, and future state of the environment. The third task is to determine whether goals are being met and to develop new aims, based on the predicted future state of the world. These aims are specific outcomes that the individual wishes to occur, which direct current and future behavior. The fourth task is to determine what options possibly may be performed and how these options will affect the environment.

From this determination of the tasks during the Orientation step, the Orientation step can be partitioned into four substeps: identification, interpretation, aims derivation, and options generation (Figure 2.3). This model also shows the internal information channels between the stored model of the world and the

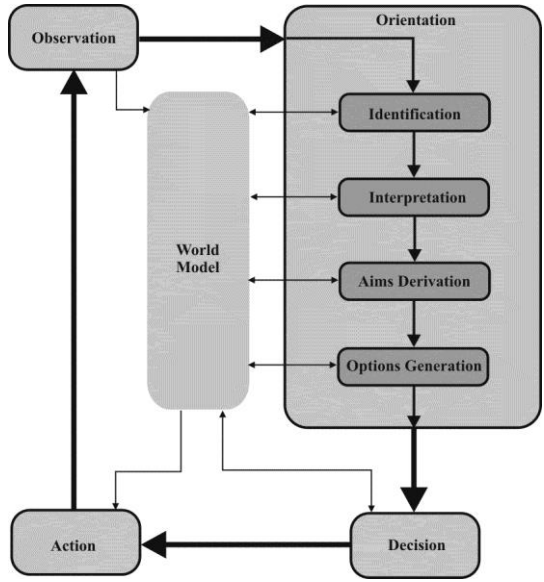


Figure 2.3 OODA loop with expanded Orientation step.

substeps that access this information. All of the substeps are capable of both retrieving information from and placing information into the model of the world.

Identification

The Orientation step starts with an Identification substep. Here the newly gathered information is compared to stored information, allowing known objects and events to be recognized. If an element is not recognized, then a placeholder entry can be created for it, which will have associated information linked to it as the individual learns about the element. A recognized element is something that has previously been observed and can be recognized. Any existing knowledge for this element can be used in the analysis. In this step the question that is being asked is what is it for each of the various elements that have been perceived in its environment.

Interpretation

In interpretation the existing world model is compared with the objects, both those identified and those not, to determine how the world has changed since the last loop iteration and to predict how it will change in the future. Information about identified objects is retrieved and unidentified elements may have temporary attributes assigned to them, which later need to be refined. The observed elements

and their related known information are used to predict the future (and past) state of the world. As newly gathered and analyzed information updates the model of the world, learning takes place. Interpretation is where any existing errors in the model of the world can affect the analysis of the new information. What does this imply is the question being asked; while the newly gathered information is analyzed and contrasted with existing knowledge.

Aims Derivation

In the Aims Derivation substep the updated model of the world is tested to determine whether the goals have been met. New aims are developed to guide future behavior. Achieved and impossible aims are removed, unachieved aims are updated, and new aims are developed. In this step the questions being asked are what are my aims and am I achieving them?

Options Generation

The final substep is Options Generation. Here the aims and updated model of the world are used to determine what options may be performed and the expected outcomes of each of them. If there are errors or omissions in the model of the world, then the outcomes expected will not match what actually occurs. In this step the questions being asked are what can I do and what will that cause in order to determine what options are believed to be possible to perform and how each option will influence the world.

Completion

After the completion of the Orientation step, the individual has determined how their environment has changed since they last observed it, predicted how it will change in the future, determined what their aims are, and determined the possible options they may perform and the outcomes that they will lead to. During the Decision step, the options and outcomes are examined and assessed based on the perceived value of the expected outcomes of each option. The best options are selected and then performed in the Action step.

2.3.1.2 Corruption Attacks

Corruption attacks and self-deception both cause errors during the Orientation step [9]. Corruption attacks are IW attacks where the attacker transmits a corrupted signal to the defender that mimics a valid signal expected by the defender. The attack is successful if the defender believes that the corrupted signal is actually a valid signal. Corruption attacks are one of the canonical IW strategies that we will discuss in Chapter 4.

In a Corruption attack, the attacker's corrupted information enters the defender's system through Observation and is then analyzed during the Identification substep, where the error occurs. The defender believes that the corrupted information is genuine. The defender misidentifies the corrupted information as the element it is mimicking, which is then used during the Interpretation substep as though it were valid.

2.3.1.3 Self-Deception

Self-deception is really a misnomer. It is more accurately described as an intentional misinterpretation by the target that aims to support a favored, but inaccurate, belief [10]. Self-deception is a self-inflicted corruption attack that specifically targets the information processing methods. Instances of self-deception typically occur when the desired environmental state cannot be achieved through actions and so instead that target manipulates the model of the world to produce the illusion of the desired state.

Self-deception may, however, reduce the cognitive dissonance [11] and therefore produce some positive use. Cognitive dissonance is when an individual holds beliefs that are inharmonious [12], resulting in psychological discomfort. Individuals can use self-deception to reduce the dissonance between the beliefs and thereby reduce their discomfort.

In self-deception, the self-deceiving individual correctly gathers information during Observation and then correctly Identifies known objects, events and relationships. During the Interpretation substep, when the information is analyzed, it is found to be dissonant with the individual's existing knowledge and discomfort ensues. This discomfort is reduced by misinterpreting the information in such a way that it is no longer dissonant with its existing knowledge.

2.3.2 Cognitive Hierarchy Model

The cognitive hierarchy [13, private communication with P. J. Berenson, 1998] provides an alternative way to view information processing. As illustrated in Figure 2.1, it begins with sensing the environment, which is loosely equivalent to the observe node in the OODA loop. That sensing produces data, which is processed. This processing of the raw data produces information and, when combined with cognition and then put into context, produces knowledge. Adding a person's experience to the new knowledge produces an understanding of the situation. Adding intuition to that understanding produces wisdom. Not all models of the cognitive hierarchy include this last step generation of wisdom (see Figure 2.4 [14]).

This graphic is not intended to imply where these functions are performed. For example, processing the raw data may be accomplished at the sensor site. If

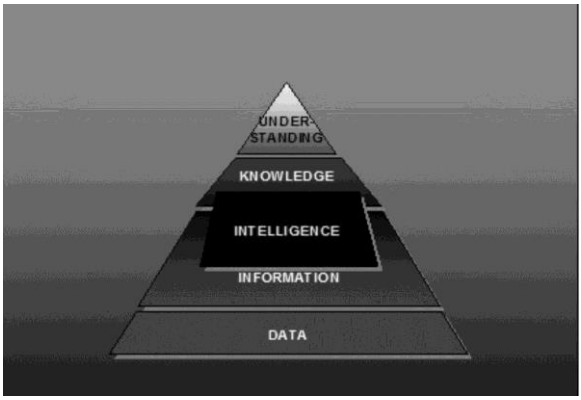


Figure 2.4 Cognitive hierarchy with the intelligence function overlay. (Source: [14].)

there is more than one sensor at that site, or multiple observations with a single sensor are available at that site, sensor data fusion may occur at that site as well.

2.4 Information Operations

Information operations is the appellation applied to warfighting concepts that rely on knowing information about an adversary as well as trying to keep information about friendly forces from that adversary. Knowing all that is necessary about an adversary can, in theory, keep friendly forces out of harm's way. Knowing the location of an adversary's tank column out of range of those forces' weapons reach, for example, can facilitate indirect fires against the column thereby precluding a meeting engagement.

Two related titles have surfaced in recent years that refer to the handling of information and what it is used for: information operations and information warfare. The U.S. DoD Joint Staff defines the two fields as [15]:

- Information Operations: "Actions taken to affect adversary information and information systems while defending one's own information and information systems."
- Information Warfare: "Information operations conducted during time of crisis or conflict to achieve or promote specific objectives over a specific adversary or adversaries."

We can see that the primary difference is when and where the operation is conducted—IW in terms of crisis and IO at any appropriate time. EW, being

composed of the three legs of a stool, ES, EA, and EP, can be considered part of IO. Certainly in peacetime, or the time leading up to conflict, ES and EP can play significant roles in terms of updating the EOB and contributing to the generation of the COP and protecting friendly forces by minimizing an adversary's ability to establish blue EOB. It is probably fair to say that EA is a function delegated to IW situations, for the most part.

2.4.1 Information Warfare/Information Operations

IW/IO are different disciplines from information altogether. Libiki delineates seven forms of IW/IO as follows [16]:

> There are [instead] several forms of information warfare, each laying claim to the larger concept. Seven forms of information warfare—conflicts that involve the protection, manipulation, degradation, and denial of information—can be distinguished: (i) command-and-control warfare (which strikes against the enemy's head and neck), (ii) intelligence-based warfare (which consists of the design, protection, and denial of systems that seek sufficient knowledge to dominate the battlespace, (iii) electronic warfare (radio-electronic or cryptographic techniques), (iv) psychological warfare (in which information is used to change the minds of friends, neutral, and foes),) (v) "hacker" warfare (in which computer systems are attacked), (vi) economic information warfare (blocking information or channeling it to pursue economic dominance), and (vii) cyberwarfare (a grab bag of futuristic scenarios). All these forms are weakly related.

2.4.1.1 Influence Attitudes

Otherwise known as "perception management," influencing a group's attitudes about a situation is an attempt to convince others about a viewpoint by favorable spin. Perception management involves actions taken to convey and/or deny selected information and indicators to foreign audiences to influence their emotions, motives, and objective reasoning. This can be accomplished by using:

- Public diplomacy: Be open and public about the information, whether false or true.
- Private diplomacy: Be private about the information, sharing it with just a few in an attempt to convince them.
- *Psychological operations* (PSYOP): Deliberately attempt to employ psychological principles in arguments such as appealing to a sense of loyalty or pity.
- Media relations [*public affairs/public information* (PA/PI)]: Let the media carry the message to the masses, thereby giving it a sense of foundation or truth.
- Education: Attempt to employ teaching the masses as if they are in the dark about the topic.

- Counter influence/propaganda: Convince the targets by using statements in favor of your position, exaggerated or not.
- Fabricated truth: Make up "facts" that support your arguments.

2.4.1.2 Deny/Protect

Denying information to groups and protecting sensitive information are elements of IW. They are passive measures to preclude the intentional or unintentional release of information. Essential activities to accomplish this protection include:

- *Operational security* (OPSEC): Do not publically or privately release information that is sensitive or do not discuss it in public.
- *Information assurance* (IA): Manage information so that it is not tampered with or otherwise changed or revealed to those to whom it should not be.
- *Computer network defense* (CND): A whole myriad of activities that protect commuter networks from invaders and hackers.
- *Counterintelligence* (CI): Actions are taken to prevent or at least minimize an adversary's attempts to gain intelligence against friendly forces.

2.4.1.3 Deception

To deceive an adversary is to intentionally take actions to cause incorrect conclusions to be drawn. It is also a passive measure. Ways to accomplish this include:

- Spoofing: In IW, a spoofing attack is actions taken where one node masquerades as another by falsifying data and thereby gaining an illegitimate advantage. It causes the element being spoofed to think that it is dealing with someone it actually is not.
- Imitation: Similar to spoofing, imitation is when a node takes action to pretend it is who it is not.
- Distortion: Deliberately change or exaggerate the facts.
- Deception: Actions are taken to deliberately mislead a target and to thereby cause it to take (or not take) specific actions that aid information protection.

2.4.1.4 Exploit/Attack

While exploitation of information is a passive attack measure on information, deliberate attack by, for example, EW means, is an active measure. A few ways to exploit or attack information are:

- Electronic warfare: EW, the subject of this book, consists of the three elements previously noted: electronic protect, electronic support, and electronic attack. Electronic protect is covered elsewhere. Electronic support is exploitation of electronic emissions. Electronic attack is the injection of energy into a target receiver to prevent an intended message from being received.
- Ballistic: Ballistic attacks are actions that employ kinetic energy to damage or destroy targeted information systems.
- EMP: Electromagnetic pulses (EMP) can be used to destroy the electronic components in targeted information systems as well. In this case it is specifically the electronics that are targeted—usually the sensitive front ends of receiving and computer networking systems.

2.4.1.5 Computer Network Operations

Computer network attack, computer network defense, and related computer network exploitation enabling operations [17] are the components of *computer network operations* (CNO). Additionally, CNO is an element of IW/IO.

- *Computer network attack* (CNA). Actions taken through the use of computer networks to disrupt, deny, degrade, or destroy information resident in computers and computer networks, or the computers and networks themselves.
- *Computer network defense* (CND). Actions taken through the use of computer networks to protect, monitor, analyze, detect and respond to unauthorized activity within Department of Defense information systems and computer networks.
- *Computer network exploitation* (CNE). Enabling operations and intelligence collection capabilities conducted through the use of computer networks gather data from target or adversary automated information systems or networks.

2.4.2 Three Domains of Conflict

Alberts et al. [18] identify three domains within which IO and decision-making transpire. These domains are the *physical domain*, the *information domain*, and the *cognitive domain* (see Figure 2.5). The physical domain is where physical entities exist and actions occur. This is where warfighting takes place. The information domain is where information is moved and stored. The cognitive domain is where thinking occurs and the decisions are made.

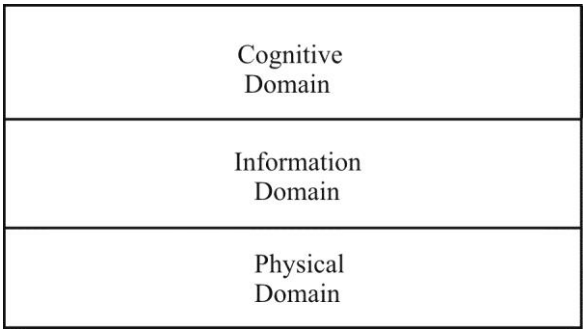


Figure 2.5 The three domains of conflict.

The domains of conflict are usually depicted as a two dimensional figure consisting of three equal blocks stacked one on top of another (see Figure 2.5). The physical domain is placed on the bottom, the information domain in the middle, and the cognitive domain on top, implying an equality and structural hierarchy between the domains. This is just a visual representation and should not imply anything other than the information domain's role as the linkage between the real world and the human mind. The information domain can be depicted as a line where the physical and cognitive domains meet, or perhaps as a space created by the overlap between the two domains.

Key characteristics of the three domains are depicted in Table 2.1. Even though the physical, information, and cognitive domains are often portrayed as separate entities as in Figure 2.5, in reality they are closely connected. The interrelationship becomes clearer when a decision-making or action-reaction cycle (also known as the OODA loop) is superimposed on the domain structure as in Figure 2.6 [17].

Battlespace monitoring with sensors in the physical domain produces the observations, which are passed to the information domain in the form of data. SA produces orientation and situation awareness. Decisions are made in the cognitive domain based on the understanding produced by the situation awareness. These decisions, guided by the commander's intent, facilitate battlespace management in the form of actions, which are synchronized at the information/physical domain boundary. These actions change the world, and the changes are observed. Thus, the loop continues through the three domains on a continuous basis.

Unlike land, sea, air, and space, the information environment has minimal physical presence. Yet it is possible to visualize information's effects on military operations by portraying the structure of the information environment in a manner similar to how commanders and staff visualize the physical environments of their operational area.

Table 2.1 The Information Environment

Cognitive Domain	Where human decision making takes place
	Dimension of intangibles such as morals, unit cohesion, public opinion, situational awareness
	Key characteristics: perceptions, emotions, awareness, and understanding
Information Domain	Where information is collected, processed, stored, disseminated, displayed, and protected
	Dual nature—information itself and the medium
	Links physical and cognitive dimensions
	Key characteristics: information content and flow, information quality
	Where automated decision-making takes place
Physical Domain	Where the information environment overlaps with the physical world
	Information systems and networks
	Key characteristics: computers and communication systems, and supporting infrastructure

Source: [19].

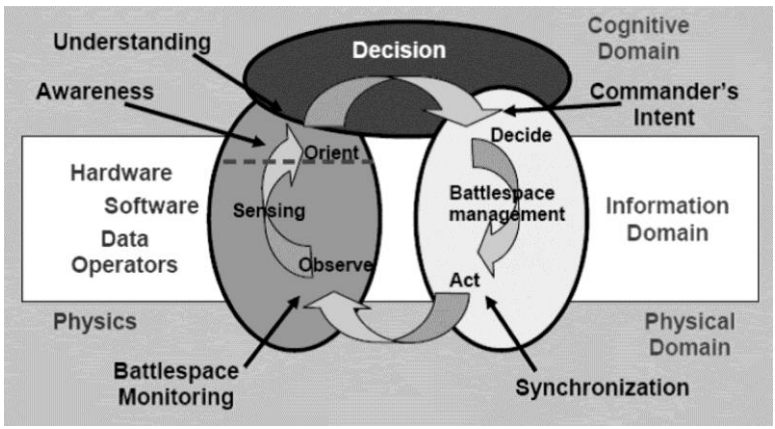


Figure 2.6 Relationships between the OODA loop and the three domainsof IO. (Source: [16].)

Activity in the physical domain generates data that is collected by information systems. These information systems create and direct the flow of information through the information domain. In turn, the information is used by humans in the cognitive domain to form perceptions and to ultimately make decisions. These decisions are subsequently communicated through the information domain via information systems to the physical domain and then converted into human activity. As a result, activity in one domain can produce subsequent effects in the other domains. Furthermore, because of the physical domain's connection to the rest of the physical world, information content and flow can manifest themselves in very real ways [20]. Thus, despite the information domain's intangible nature, its effects are very tangible.

The key to using information as a military capability lies in the information domain. This is because the information domain is the means by which physical domain activity and decision-making interrelate. As such, information content and flow are essential to both the formation of decisions and the execution of decisions as physical activity or behavior.

In reality, the relative importance of the domains to military operations is not as simple as a series of coequal geometric shapes. For example, the domains' relevance can vary by echelon of operation (i.e., tactical, operational, and strategic). At the tactical level of operations, the nature of the information environment is very physical. Data collecting and the resulting information content are dominated by visual observation and face-to-face human contact. Surveillance and reconnaissance are the predominant sources of data. Information flow is greatly impacted by terrain and physical objects. In this environment, IO uses short-range information means and the profile and posture of maneuver forces to change the immediate and short-term behavior of discrete target audiences.

At the operational and strategic levels, the information environment becomes more conceptual, an exchange of broad competing ideas and ideologies. At these echelons, IO uses mass communication means to change mid- and long-term beliefs and attitudes of broad target audiences.

Furthermore, the importance of each domain to military forces may change according to mission and area of operation. During conventional combat operations, the destruction of enemy information systems and networks may dominate non-lethal measures to influence adversary trust, will, and decision-making. At the opposite end of the operational spectrum, during peace operations, key leader and populace group perceptions and attitudes may be more important than physical world reality.

2.4.2.1 Primitives in the Three Domains of IO/IW

Alberts et al. [18] described several primitives that exist within these three domains that form the fundamental tenents of IO. These primitives are: (1) sensing, (2) awareness, (3) decisions, (4) observations, (5) understanding, (6)

actions, (7) information, (8) sharing, (9) synchronization, (10) knowledge, and (11) collaboration. We add a twelfth one: communications.

Sensing

Sensing is the measuring of some attribute in the environment. It can be accomplished by an individual with the five senses most of us possess or it can be accomplished by a sensor of some variety. In the former case, what is sensed is changed immediately into knowledge (which may or may not represent truth), whereas in the latter case the sensor gathers data for subsequent conversion to information and possible knowledge. In the case of interest herein, the sensors involved are EW sensors and may or may not have humans involved. Such sensors detect and receive RF energy of some variety. That energy is processed in several possible ways to extract information. In the case of a clear communication signal, an operator may listen to the conversation and prepare a gist of the conversation, for example.

Awareness

Awareness refers to perceptions of the current situation and is a complex combination of current observations and epistemology. [3] People are aware of a state of nature if they are cognizant of it—it has attracted their attention and they have some degree of knowledge about its existence.

Decisions

Decisions are the result of analysis of the current state of knowledge and generate subsequent actions. They are a determination arrived at after due consideration of the state of affairs.

Observations

Observations are the result of sensing described above. Observations need not be common between two or more individuals. A person's experiences may cause that person to observe a state of affairs completely differently from someone else observing the same phenomenon.

Understanding

Understanding is achieving an adequate level of combined knowledge about a state of nature that decisions can be made about the situation. It is grasping the

[3] A person's epistemology consists of what that person perceives to be true (knowledge and beliefs), whether they are in fact true or not.

meaning of that state so that decisions can be made. This understanding requires the ability to forecast into the near future the results of decisions made.

Actions

Actions are the result of making decisions. They always occur in the physical domain in a warfighting context. Actions are accomplished by actors that change the state of the world in some way.

Information

Information results from adding *context* to observations. It places what is currently being observed into one or more possible scenarios. The context is the surrounding state of affairs that causes the observations to make sense.

Sharing

Sharing refers to the exchange of information and/or knowledge between two or more entities. It always requires some form of communication between the entities, and in the battlespace context of the military at the edge, is frequently accomplished by radios. Sharing of information implies that data is exchanged, although the receiver of the information may not have the same degree of knowledge about the data that is being exchanged.

Synchronization

Synchronization is the process of coordinating the actions of two or more physical entities to achieve a common purpose. *Swarming*, the result of self-synchronization, is an example of synchronization. Swarming results from two or more bodies viewing the state of affairs similarly and similar actions of the bodies result.

Knowledge

Knowledge is the result of adding context to observations, and combining the results with experience and training. It represents the current belief about what is being observed.

Collaboration

Collaboration is the sharing of information and knowledge among two or more entities for the purpose of accomplishing a common purpose. Again, collaboration implies some means of communication between parties.

Communications

Communications is at the heart of most of these definitions. Without the means to communicate, *dominate battlespace knowledge* (DBK) [21] is difficult, if not impossible, to establish. If future combat is to depend on knowing more about the enemy than the enemy knows about friendly forces, then the ability to communicate is of paramount importance.

2.4.2.2 Data and Information

On the IW battlefield, there are only four tasks to be performed: *Data* is:

- Collected;
- Moved;
- Stored;
- Used to reduce uncertainty (entropy) (perform SA).

In the process of using data to perform SA, information is generated. The efficiency with which we can do this depends on the amount of data available (the information completeness) and the clarity in the data.

As pointed out by Borden [22], *information* is not collected, stored, moved, or used to reduce uncertainty. Information is *generated* in the course of reducing uncertainty so that decisions can be made [23]. In fact, as we investigate in Chapter 3, information *is* the reduction in uncertainty—which we normally measure in bits.

2.4.3 Applying the Domains of Conflict to IO

Developed primarily as a way to explain the process of decision-making, the three domains of conflict provide a general framework for explaining how information affects the performance of military operations. In brief, the CCRP's description of the domains are as follows [18]:

- The physical domain is the real, tangible world: the environments of land, sea, air, and space. "It is the domain where strike, protect, and maneuver take place across the different environments. It is the domain where physical platforms and the communication networks that connect them reside."
- The information domain is "where information lives. It is the domain where information is created, manipulated, and shared. It is the domain that facilitates the communication of information."
- The cognitive domain is in the mind of human beings. It is where "perceptions, awareness, understanding, beliefs, and values reside and

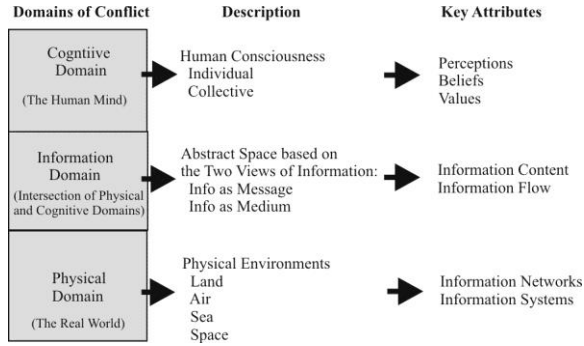


Figure 2.7 The information environment. (Source: [20].)

where, as a result of sense-making, decisions are made." It is an inherently human environment.

The domains of conflict depict the relationship among physical action, information, and decisions. For the purpose of understanding how to use information as a military capability, the domains also represent aptly the structure of the information environment.

2.4.3.1 A Model for the Information Environment

The information environment, in contrast to the other environments in which military forces operate—land, sea, air, and space—is largely nonphysical and abstract. It is a man-made paradigm based on the belief that the existence and proliferation of information and information systems has created a new operating dimension or environment. However, even though a portion of the information environment is composed of physical information systems (mainly computers and communications), the primary component of the information environment—information—is intangible.

Any model of the information environment must accommodate the tangible and intangible parts. The model must also include information's dual nature, namely, its utility as a vessel that contains meaningful content and its existence as a medium by which data and information are created, manipulated, and exchanged.[4] Combining the three domains of conflict with the two views of information generates the view of the information environment depicted in Figure 2.7 [20].

[4] For a discussion of the three views of information, see [24].

The domains can be further examined to describe the information environment as follows:

- The physical domain is the tangible portion of the information environment that is part of the physical environments of land, sea, air, and space. Technology or human-based networks and information systems exist in this domain. Individuals and organizations employ information systems in this domain. For the purposes of IO, the physical domain is where information systems are attacked and defended.
- At the interface of the physical and cognitive domains resides the information domain. It is an abstract space [25]. Individuals and organizations communicate in this domain and it is where the functions of physical information systems occur (i.e., information collection, processing, and dissemination). Perhaps most importantly, the information domain is where information resides. Governed by information theory (which we discuss in Chapter 3), the domain has two principal components: information-as-message and information-as-medium. This results in a duality of information content and flow.
- The cognitive domain is also abstract. It exists in the minds of human beings and collective consciousness of groups and organizations. This domain is intangible, consisting of those elements of human thought that influence decision-making and behavior. In this domain, IO seeks to affect the interpretation and use of information by decision-makers, other specific audiences, and sometimes, whole population groups.

2.4.3.2 A Possible Fourth Domain—Cultural or Social

Culture is a dynamic that influences the attributes of all three domains. In the physical domain, the social structures and human networks (i.e., organizations and institutions) that impact the use of information systems are created. Language and cultural symbols impact the content and flow of information in the information domain. In the cognitive domain, mental programming (i.e., values, beliefs, epistemology in general) affects how information is used for decision-making.

It is widely believed that there is a strong cultural component or aspect when applying the domains of conflict to an adversary. This is a possible fourth domain consisting of cultural or social factors that impact the creation, processing, dissemination, and use of information. However, culture remains an amorphous and elusive concept that is difficult to integrate.

2.4.3.3 The Information Domain

The information domain is the most intangible part of the information environment. Existing at the intersection of the physical and cognitive domains, it

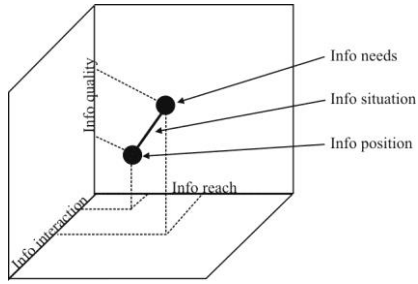


Figure 2.8 Information needs, position, and situation.

is an abstract, nonphysical space. The domain can be characterized as having three primary attributes—information quality, reach, and interaction.

Information quality, reach, and interaction are the elements that connect the information domain to the physical and cognitive domains. However, more importantly, the three attributes form the basis of information's utility to military organizations. From this perspective, the attributes can be loosely described as follows:

- Information quality. The value of information to an organization in terms of accuracy, relevancy, and timeliness. Organizations require information that is useful to their mission and current situation.
- Information reach. The degree of interoperability to which an organization exchanges information both internally and with the rest of the information environment. An organization must share and distribute information to collaborate or synchronize activities.
- Information interaction. The quality of information exchange (e.g., face-to-face discussion, radio, print, telephone, computer network, and so forth) available to an organization for the generation and distribution of information. The employment of information technology and process affects an organization's ability to use information and interface with the information domain.

Information Needs, Position, and Situation

A constant flow of relevant, accurate information is needed in all organizations in order to operate successfully. An organization's information needs are defined as [18]: "the measurable set of information required to plan and/or execute a mission or task." We refer to the information an organization possesses at any point in time is its *information position*. Needs and position can both be expressed in terms of information quality, reach, and interaction, and depicted as a point in three-dimensional space in Figure 2.8.

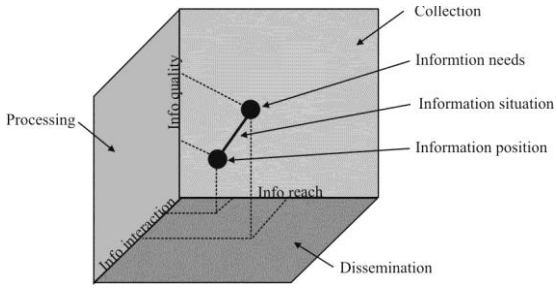


Figure 2.9 Information needs, position, and situation planes. Collection plane = information quality × information reach; processing plane = Information interaction × information quality; and dissemination plane = information interaction × information reach. Note that collection + processing reduces entropy and thereby generates information.

An organization is unlikely to possess all the information it requires to operate optimally so there will always be a disparity between its information needs and position. This gap is referred to as its *information situation*. The information situation constantly fluctuates because needs and position change with each mission or task, over time. The challenge for an organization is to reduce the disparity between information needs and position as much as possible—that is, to make the information situation as small as possible.

Although the cube shown in Figure 2.8 may adequately describe information's utility to organizations, it does not explain what an organization needs to do to maintain or improve its information situation. To change focus from simply possessing information to how organizations operate in the information domain, it is necessary to consider the planes formed by the three axes (see Figure 2.9). To this end, the three planes can represent information system functions: information collection equals quality and reach, information processing equals quality and interaction, and information dissemination equals reach and interaction. Together, these functions can represent how an organization maintains or improves its information situation.

Information Advantage

Information and the information environment are not static. A distinct advantage to an organization relative to its opponents can be produced by the possession and use of information. Therefore, organizations not only produce and disseminate information to meet their own information needs, but also, when in a competitive situation, attempt to gain and maintain an information advantage over an opponent.

The CCRP defines *information advantage* as the ability to use information better than an opponent. Information advantage means being in a superior

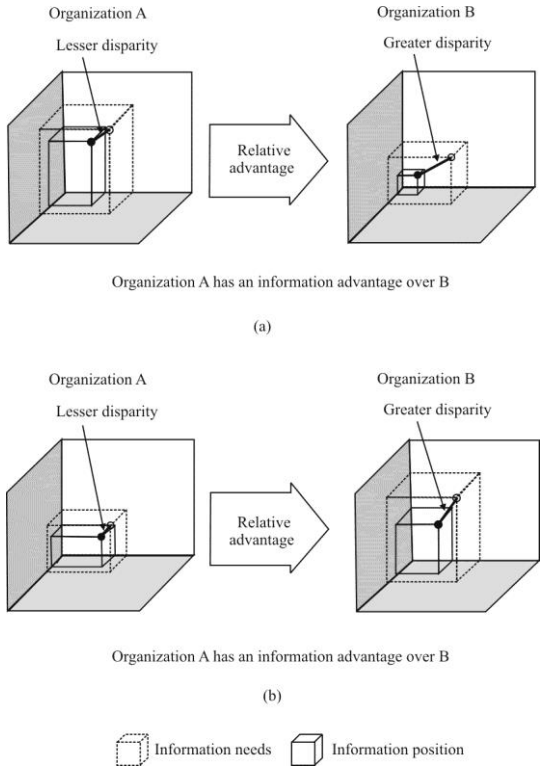


Figure 2.10 (a, b) Information advantage. In (a), Organization A's information needs and position are larger than B's but there is less disparity between them. Therefore Organization A has an information advantage over Organization B. In (b), Organization A's information needs are smaller than Organization B's but its position more closely matches its needs. Therefore again it has an information advantage over Organization B.

information situation relative to another, perhaps opposing, perhaps not, organization. Information advantage is determined by comparing the disparity between each side's information situations. "[The] information situation can be described in terms of the volumetric difference between needs and position" [18]. For some simple illustrations, see Figure 2.10. In Figure 2.10(a), Organization A has information needs greater than those of Organization B; however, its information position is closer to its needs than Organization B so it has an information advantage. Likewise in Figure 2.10(b), Organization A's information position is considerably less than that of Organization B's; however, its information needs are substantially less as well. Again, Organization A has an information advantage.

Even though two organizations may occupy the same operating environment, they are unlikely to have the same information needs, position, and situation, or even the same capabilities to use the information domain. Therefore, information advantage is measured in terms of one's own information situation relative to that of the opponent. Furthermore, because characteristics of the operating environment impact different organizations in different ways, relativity extends to how the information environment affects opposing organizations, thus the situations depicted in Figure 2.10.

An information advantage can be created by the ability to generate and use information better than one's adversary, the reduction of the adversary's information position, or the leveraging of the information environment for one's own purposes. Whichever methods are used to produce an information advantage, both the content and flow of information must be addressed if an exploitable information advantage is to be realized.

Information Superiority

A useful definition of information superiority is [26]: "the operational advantage gained by the ability to generate and disseminate an uninterrupted flow of information while denying an adversary's ability to do the same." In a hostile setting, information superiority is also known as DBK [21]. As suggested previously, information is generated in the process of collecting and processing data. As the entropy (randomness) in the data is reduced, information is produced. Information superiority results from an information advantage.

The operational advantage resulting from information superiority can generally manifest itself in two ways: in the physical domain as a force or position advantage, or in the cognitive domain as decision-making advantage. However, an advantage in the information environment does not automatically equate to information superiority. An operational advantage will only result from information advantage if it is achieved for a specific purpose as part of the overall operational plan.

Both information advantage and superiority are localized and transitory conditions. This is because the respective information situations of opposing forces, as well as information content and flow in, and through, a specific geographic area are dynamic. Therefore, to have value to a military force, information advantage and superiority are sought at certain places and times in the operational area.

2.4.4 The Efficiency of Decision-Making

IW is all about measures to improve (or degrade) the efficiency of decision-making. The maximum theoretical efficiency depends on the amount and quality of data available and on the amount of ambiguity in the data. The achievable

efficiency depends also on the strategy used to generate information from data. For example, if the introduction of a dummy radar transmitter introduces ambiguity into the radar parametric database, we would wish to measure the effect on our ability to identify and classify radars in seconds or in bits of information generated per second.

For another example, if the adversary introduces an LPI communications system, our ability to determine the amount of traffic on the link is degraded, thereby reducing the amount of data we have available to generate information and make decisions. The reduction in efficiency would be a good measure of the effectiveness of the LPI IW measure.

The task of determining the theoretical efficiency of information generation (uncertainty reduction) is not easy. The difficulty of designing good decision-making strategies comes from two facts:

- Each source of data has a different cost, typically measured in time.
- Each source of data makes a different information contribution (has a different amount of ambiguity) and this depends on the current state of the problem.

Taking these two facts into account means that we must consider the contribution of data to the timely solution of the problem. Computing the bits per second of information to be derived from a data source requires heavy computation involving many conditional probabilities.

2.4.5 Summary

IO can be described as activities to impact the content, flow, and use of information in order to gain an operational advantage over an adversary. The purpose of any information operation is information superiority—an operational advantage resulting from the use of information. While IO operates in all three domains of the information environment, to be most effective, it must focus on the information domain where an information advantage is achieved. Thus, the broad objective of IO is information advantage.

The creation of an information advantage is linked to specific activities in the physical world. IO is conducted by affecting and protecting the means of information content and flow in the physical domain (i.e., information systems and networks). These actions are directed at affecting the adversary's functions in the information domain (i.e., information generation and dissemination). This manipulation of the information domain and attacking of adversary information capabilities creates an information advantage, that when synchronized with other military operations provides information superiority at a specific place and time in either the cognitive (i.e., a decision-making advantage) or physical (i.e., a force advantage) domains.

We must recognize that the three-domain model is not exclusive to the information environment. All military operations, not just IO, occur within the framework of these domains. Every military action has the potential to convert information into a military capability, and any asset or capability that can affect content and flow of information is a possible contributor to an information operation. For this reason, IO should not be viewed as a stand-alone operation or finite, discrete set of assets and capabilities. IO must be integrated with the other battlespace activities to be effective. Therefore, at a minimum, IO should represent all methods and means that can impact the information environment.

2.5 Concluding Remarks

We investigated some of the characteristics of information early in this chapter. This discussion characterized Levels B and C of the Shannon model of communication discussed in Chapter 1: Level B—the semantic problem, and Level C—the effectiveness problem. In the remainder of the book we will concentrate on Level A—the technical problem, and specifically, how effective EW can be applied to impact on this aspect.

In this chapter we examined the nature of IO. We discussed the three domains of conflict and showed how the OODA loop interacts in each of the domains. Information advantage and dominance were explained, showing how they can be achieved and applied to IW.

References

[1] Sun Tzu, *The Art of War*, translated by R. A. Ames, New York: Ballantine Books, 1993, p. 104.

[2] Sun Tzu, *The Art of War*, translated by R. A. Ames, New York: Ballantine Books, 1993, p. 126.

[3] Sun Tzu, *The Art of War*, translated by R. A. Ames, New York: Ballantine Books, 1993, p. 161.

[4] Sun Tzu, *The Art of War*, translated by R. A. Ames, New York: Ballantine Books, 1993, p. 169.

[5] Fewell, M. P., and M. G. Hazen, "Network-Centric Warfare—Its Nature and Modeling," Report DSTO-RR-0262, Maritime Operations Division, System Sciences Laboratory, Defense Science and Technology Organization, Australian Department of Defense, September 2002, http://www.dsto.defence.gov.au/corporate/reports/dsto-rr-0262.

[6] Boyd, J. R., "A Discourse on Winning and Losing," a collection of unpublished briefings and essays, Maxwell AFB, AL: Air University Library, 1976–1992. http://www.belisarius.com/ modern_business_strategy/boyd/essence/eowl_frameset.htm, January 1996.

[7] Klein, G. A., *Sources of Power: How People Make Decisions*, Cambridge, MA: MIT Press, 1988, pp. 1-30.

[8] Brumley, L., C. Kopp, and K. Korb, "The Orientation step of the OODA loop and Information Warfare," Clayton School of Information Technology, Monash University, Australia, 1991.
[9] Brumley, L., C. Kopp, and K. Korb, "Misperception, Self-Deception and Information Warfare," *Proceedings of the 6th Australian Information Warfare & Security Conference 2005*, Geelong, Victoria, pp. 125–130.
[10] Szabados, B., "Self-Deception," *Canadian Journal of Philosophy*, Vol. 4, No. 1, pp. 51–68.
[11] Ramachandran, V. S., "The Evolutionary Biology of Self-Deception, Laughter, Dreaming and Depression: Some Clues from Anosognosia," *Medical Hypotheses*, Vol. 47, 1996, pp. 347–362.
[12] Festinger, L., *A Theory of Cognitive Dissonance*, Stanford, CA: Stanford University Press, 1957.
[13] Berenson, P. J., Unpublished notes, February 1998.
[14] White Paper "Objective Force Fusion," U.S. Army Intelligence Center, Directorate of Combat Developments, Ft. Huachuca, AZ, March 2003.
[15] Joint Publication 3-13, Infomrmation Operations Doctrine, The Pentagon, Washington D.C., February 2006.
[16] Libiki, M. C., "What Is Information Warfare?," Center for Advanced Concepts and Technology, Institute for National Strategic Studies, National Defense University, 1996.
[17] IO Primer, U.S. Army Carlyle Barracks, Carlyle, PA, 2012.
[18] Alberts, D. S., et al., *Understanding Information Age Warfare,* DoD Command and Control Research Program, Washington, D.C., 2004.
[19] Joint Pub. 3-14, *Space Operations*, U.S. DoD Joint Chiefs, January 6, 2009.
[20] Romanych, M. J., "A Theory-Based View of Information Operations," *IO Sphere*, Spring 2005, pp. 12–16.
[21] Johnson, S. E., and M. C. Libiki (eds.), *Dominant Battlespace Knowledge*, Washington D.C.: CCRP Publications, 1995.
[22] Borden, A., "What Is Information Warfare?" http://www.airpower.au.af.mil/airchronicles /cc/borden.html.
[23] Campen, A. D., USAF (Ret.), "Rush to Information-Based Warfare Gambles with National Security," *Signal Magazine*, July 1995, pp. 67–69.
[24] Arquilla, J., and D. Ronfeldt, *In Athena's Camp: Preparing for Conflict in the Information Age*, Santa Monica, CA: RAND, 1997, pp. 144–149.
[25] Sparling, B. N., "Information Theory as a Foundation for Military Operations," Fort Leavenworth, Kansas: Scholl of Advanced Military Studies, U.S. Army Command and Staff College, 2002, pp. 12–19.
[26] FM 1-02, *Operational Terms and Graphics*, Headquarters, Department of the Army, Washington, D.C., September 2004.

Chapter 3

Information Theory

3.1 Introduction

This chapter presents a brief tutorial on information theory, as formulated by Shannon [1]. It is well beyond the scope of this book to engage in a comprehensive discussion of that field; however, it is worthwhile to have a short reference of the relevant concepts.

The chapter is structured as follows. We begin with a discussion of the relevant concepts from probability theory including the meaning of entropy (randomness). We then introduce information and how it relates to entropy reduction. Next we introduce information channels including Shannon's channel coding theorem and describe the notion of channel capacity. Finally we present some useful channel models.

3.2 Random Variables and Probabilities

A *variable* is an object, X, that can take on any value from a set of values Ω_x (called its *domain*). These values may be discrete and finite, such as the letters of the alphabet or {0, 1}, or they may be continuous and infinite, such as any real number. A random variable is a variable whose value is unpredictable. A particular value that a random variable has taken on is called a *trial*. A collection of trials is called a *sample*. A common example of a random variable is one representing the flip of a coin. The random variable may take on one of two values $\{H, T\}$. Each time we flip the coin, we have a trial. A series of coin flips is then a sample. A particular instantiation of X is denoted x.

Associated with every random variable is a (possibly unknown) *probability distribution*, $\Pr\{X\}$.[1] The probability of a particular value is the proportion of the

[1] The formal notation for the probability of event E occurring is given by

number of times you expect to see that value over a very large sample. This distribution maps every possible value of Ω_x to a value in [0, 1]. As Pr{X} is a probability distribution,

$$\sum_{x \in \Omega_x} \Pr\{x\} = 1$$

In our coin example, we might assume a fair coin, such that $\Pr\{X = H\} = \Pr\{X = T) = 1/2$. For a two-headed coin, we would have $\Pr\{X = H\} = 1$ and $\Pr\{X = T\} = 0$.

For continuous random variables, there is a subtle problem. Because X can take on an infinite number of values, the probability of any given value will almost always be zero. Instead of probability distributions, we use *probability densities* and integrate over ranges of possible values to determine probabilities; however, the distinction is not important for our purposes.[2]

In addition to discussing a single random variable, we have a vocabulary for discussing several at once. This is useful because random variables may be dependent upon one another. For example, we may define a new variable, $Y = F(X)$, where $F()$ is a deterministic function. In this way, knowing X means that we always know the value of Y. On the other hand, if X and Y represent two separate coin flips then we might expect that knowing the value of one will not tell us anything about the other. If this is true, they are said to be *independent*. Of course, there are states between complete dependence and complete independence.

We can formalize these notions using *joint distributions*. A joint distribution, Pr{X, Y}, tells us everything about the co-occurrence of events from X and Y. In fact, we can derive Pr{X} [and Pr{Y}] from the joint by computing the *marginal distribution*.

Definition 3.1: The *marginal distribution* is given by

Pr{E}

However, when it is convenient and it does not otherwise cause confusion, in this book this is sometimes shortened to P_E. Both refer to the exact same concept and property. Such confusion can occur, for example, when discussing signal power. P is frequently used in engineering contexts to represent this power. When such confusion can occur, we revert back to the formal notation and hopefully the context will indicate which is being used for what.

[2] This is very important in general, however. In particular, many of the theorems that hold for discrete random variables do not hold for continuous variables. Where this is a problem, we will mention it; otherwise, when thinking of continuous random variables use integrals instead of summations.

$$\Pr\{X\} \triangleq \sum_{y \in \Omega_y} \Pr\{X, Y = y\} \tag{3.1}$$

Property 3.1: Two variables are independent if and only if

$$\Pr\{X, Y\} = \Pr\{X\}\Pr\{Y\} \tag{3.2}$$

Closely related to the joint distribution is the *conditional distribution*.

Definition 3.2: The *conditional distribution* is given by

$$\Pr\{Y \mid X\} \triangleq \frac{\Pr\{X, Y\}}{\Pr\{X\}} \tag{3.3}$$

from which we get the probability of Y if we already know X.

Note that (3.3) gives us another definition for the joint, namely, that $\Pr\{X, Y\} = \Pr\{Y \mid X\}\Pr\{X\}$. Because joints are not order dependent, this also means that $\Pr\{X, Y\} = \Pr\{X \mid Y\}\Pr\{Y\}$. This observation leads us to *Bayes' rule*:

Property 3.2: Bayes' Rule.

$$\Pr\{X \mid Y)\} = \Pr\{Y \mid X\} \frac{\Pr\{X\}}{\Pr\{Y\}} \tag{3.4}$$

We can construct joint and conditional distributions over three random variables, Pr**{***X*, *Y*, *Z*}, as well. We can also compute the marginal:

$$\Pr\{X\} = \sum_{z \in \Omega_z, y \in \Omega_y} \Pr\{X, Y = y, Z = z\} \tag{3.5}$$

We can even define the joints in terms of conditionals:

$$\Pr\{X, Y, Z\} = \Pr\{X \mid Y, Z\}\Pr\{Y \mid Z\}\Pr\{Z\} \tag{3.6}$$

A definition of independence

$$\Pr\{X, Y, Z\} = \Pr\{X\}\Pr\{Y\}\Pr\{Z\} \tag{3.7}$$

follows naturally since $\Pr\{X|Y,Z\} = \Pr\{X\}$, and $\Pr\{Y|Z\} = \Pr\{Y\}$. In general, we can define joints and conditionals for any number of random variables.

3.2.1 Moments

There are several statistics we might want to use to describe the behavior of random variables. When the random variable ranges over numbers, one of the most common statistics is the "average." We can define the mean or expected value of a random variable as:

> **Definition 3.3:** The *average*, *mean*, *expected value*, or *first moment* of a random variable is defined as

$$\mathcal{E}_X\{X\} = \sum_{x \in \Omega_x} x \Pr\{X = x\} \tag{3.8}$$

In a common abuse of notation, we will usually dismiss the subscript and refer to the expectation of X as simply $\mathcal{E}\{X\}$, unless doing so creates confusion.

If $\{\ldots x_i \ldots\}$ refers to a series of trials of X, then we can compute the sample mean:

> **Definition 3.4:** The *sample mean* is given by

$$\mathcal{E}\{X\} = \frac{1}{N} \sum_{x \in \Omega_x} x \tag{3.9}$$

> where N is the number of trials in the sample.

It is worth noting that the true mean is a deterministic function of the distribution of X while the sample mean is not. Because the samples are themselves random, we might calculate a different sample mean each time we pick a sample. Therefore, the sample mean is also a random variable However, the law of large numbers[3] allows us to prove that as we take more and more trials of X, we approach an estimation of the true distribution. Thus, in the limit, the sample mean approaches the true expectation.

[3] The law of large numbers is a property that describes the result of performing the same experiment a large number of times. According to the law, the average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performed.

There are other statistics that are useful when describing random variables. One is the variance.

> **Definition 3.5:** The *variance* of random variable X measures the variation of values about the mean:

$$\mathrm{Var}\{X\} \triangleq \mathcal{E}\{(X - \mathcal{E}\{X\})^2\} = \mathcal{E}\{X^2\} - \mathcal{E}^2\{X\} \tag{3.10}$$

The variance is often denoted by $\sigma^2(X)$. It is closely related to the standard deviation.

> **Definition 3.6:** The *standard deviation* of a random variable is the square root of the variance:

$$\sigma(X) \triangleq \sqrt{\mathrm{Var}\{X\}} \tag{3.11}$$

The mean is the first *moment* of the random variable *X*. In general, there are *k* moments, each denoted by $\mathcal{E}\{X^k\}$. When we subtract the mean from *X* before taking the expectation, $\mathcal{E}\{(X - \mathcal{E}\{X\})^k\}$, we have a *central moment*. The variance is therefore the second central moment of *X*. Often, in order to control scale, we compute a *normalized central moment:*

$$\frac{1}{\sigma^k(X)} \mathcal{E}\{(X - \mathcal{E}\{X\})^k\} \tag{3.12}$$

Each increasing moment can be used to further classify the behavior of a random variable We often use kurtosis—the fourth normalized central moment—as a convenient measure of the peakedness of a distribution, for example.

3.2.2 Entropy

Although it is a very old concept, information entropy is generally credited to Shannon because it is the fundamental measure in information theory. Entropy can be defined as an expectation:

> **Definition 3.7:** The *entropy* of a random variable *X* is given by:

$$H(X) \triangleq -\mathcal{E}\{\log_2 \Pr\{X\}\} = -\sum_{x \in \Omega_x} \Pr\{X = x\} \log_2 \Pr\{X = x\} \tag{3.13}$$

where[4] $0 \log(0) = 0$.

The base of the logarithm is generally 2, but others can be used. When this is the case, the units of entropy are *bits*.

Entropy captures the amount of randomness or uncertainty in a variable. This, in turn, is a measure of the average length of a message that would have to be sent to describe a sample. Recall our fair coin from Section 3.3.1. Its entropy is $H = -(0.5\log_2 0.5 + 0.5\log_2 0.5) = 1$; that is, there is one bit of information in the random variable This means that on average we need to send one bit per trial to describe a sample. This should fit your intuitions: if we flip a coin 100 times, we'll need 100 numbers to describe those flips, if order matters. By contrast, our two-headed coin has entropy $H = -(1\log_2 1 + 0\log_2 0) = 0$. Even if we flip this coin 100 times, it doesn't matter because the outcome is always heads. We don't need to send any message to describe a sample.

There are other possibilities besides being completely random and completely determined. Imagine a weighted coin, such that heads occurs 75% of the time. The entropy would be: $H = -(0.75\log_2 0.75 + 0.25\log_2 0.25) = 0.8113$. After 100 trials, we would only need a message of about 82 bits on average to describe the sample. Shannon showed that there exists a coder that can construct messages of length $H(X) + 1$, nearly matching this ideal rate.

We can compute joint and conditional entropies, just as with probabilities. Joint entropy is the randomness contained in two variables, while conditional entropy is a measure of the randomness of one variable given knowledge of another.

Definition 3.8: *Joint entropy* is defined as:

$$\begin{aligned} H(X,Y) &\triangleq -\mathcal{E}_X\{\mathcal{E}_Y\{\log_2 \Pr\{X,Y\}\}\} \\ &= -\sum_{x\in\Omega_x, y\in\Omega_y} \Pr\{x = X, y = Y\}\log_2 \Pr\{X = x, Y = y\} \end{aligned} \tag{3.14}$$

Definition 3.9: *Conditional entropy* is defined by:

$$\begin{aligned} H(Y|X) &\triangleq -\mathcal{E}_X\{\mathcal{E}_Y\{\log_2 \Pr\{Y|X\}\}\} \\ &= -\sum_{x\in\Omega_x, y\in\Omega_y} \Pr\{Y = y|X = x\}\log_2 \Pr\{Y = y|X = x\} \end{aligned} \tag{3.15}$$

[4] We define $0\log_2 0 \triangleq \lim_{x\to\infty} (1/x)\log(1/x)$ which equals 0 since $1/x$ approaches 0 faster than $\log(1/x)$ approaches $-\infty$ as x approaches ∞.

We now list some properties associated with the entropy function.

Property 3.3: Two random variables, *X* and *Y*, are considered *independent* if and only if

$$H(Y|X) = H(Y) \text{ or } H(X,Y) = H(X) + H(Y)$$

Property 3.4: Knowing more information can never increase our uncertainty. That is:

$$H(Y|X) \leq H(Y)$$

Similarly,

Property 3.5:

$$H(X,Y) \leq H(X) + H(Y)$$

We also have that:

Property 3.6:

$$H(X,Y) = H(Y|X) + H(X) = H(X|Y) + H(Y)$$

All of these relationships hold in the general case of more than two variables.

There are several facts about discrete entropy, *H*(), that do not hold for continuous or *differential entropy*, *h*(). The most important is that while, $H(X) \geq 0$, *h*() can actually be negative. Worse, even a distribution with an entropy of $-\infty$ can still have uncertainty. Even though differential entropy cannot provide us with an absolute measure of randomness, it is still the case that if $h(X) \geq h(Y)$ then *X* has more randomness than *Y*.

3.3 Information

Information can only be produced in the presence of knowledge. In the absence of knowledge, all messages from the environment are noise. Moreover, knowledge, in the form of a database, for example, is not enough. There must be a procedure for using the knowledge to reduce uncertainty when messages from the environment are received.

It could be that the information-producer has the option to select types of messages from the environment. Each type of message could make a different information contribution, depending on the database, the current uncertainty, and how noisy the message communication channel is. Moreover, each type of message could have a different cost, perhaps time. It is clearly necessary then to develop a strategy for selecting messages to reduce the entropy most efficiently.

3.3.1 Entropy and Information

> **Definition 3.10:** *Information* is the degree to which uncertainty (entropy) is reduced.

We define "information" in terms of uncertainty because, when performing any type of SA, we begin with relative uncertainty and attempt to replace it with certainty. Therefore, uncertainty is the starting point for all information theoretic definitions.

3.3.2 Measuring Information

If uncertainty is measured, an action is taken and uncertainty measured again—the difference in measurements corresponds to information generated. The unit of measurement is in bits. The ratio of information to time is in bits per second (bps).

Since the key to measuring information is to measure uncertainty repeatedly, it is important to understand the mathematical characterization of uncertainty. Uncertainty is always associated with a probability distribution: $\{P_i\}$ i = 1, 2, 3... where $\Pr\{\text{Event } i\} = P_i$, each $P_i \geq 0$, and the P_i's sum to 1. For discrete scenarios, each P_i corresponds to the a priori probability of event i happening. This uncertainty (also called *entropy*) is given by [1]

$$H = -\sum_i P_i \log_2 P_i \tag{3.16}$$

This concept is illustrated by the following example.

> **Example 3.1: Paul Revere**
>
> In the mind of Paul Revere, land and sea attacks were equiprobable. That is, Pr{Land} = Pr{Sea} = 1/2. A lookout in a nearby tower was to observe the approach of the British forces and encode the information about the method of approach as:

Show one lantern if by land, two if by sea.

> Since $\log_2(1/2) = -1$, (3.16) shows that Paul Revere had one bit of uncertainty. History tells us that Paul Revere saw two lanterns (data). He applied his knowledge of the code given above to deduce that the British were approaching by sea (information). His uncertainty had been reduced to zero. This intuitive situation is confirmed again by (3.16) with the Pr{Land} = 0 and the Pr{Sea} = 1. We conclude that Paul Revere received one bit of information because his uncertainty had been reduced by one bit. This is consistent with the Shannon definition of uncertainty.

Data becomes information as follows. An active memory compares new data with a static database. The active memory and the database taken together function as an associative memory, adjusting the values of a nearness function or metric to reduce the uncertainty about the meaning of the data.

In this case, Paul Revere had the benefit of a noiseless communications channel and unambiguous decoding of the message. In IW, we rarely have an ideal situation like this. There is usually a great deal of ambiguity, which the DM must accommodate. In situations with prodigious ambiguity (the normal case in an active battlespace), it is a significant challenge to develop a strategy for decision-making, which produces on-time, high confidence decisions.

If "land" and "sea" were equally probable, Paul Revere's initial entropy was exactly one bit. The signal from the Old North Church (one lantern if by land, two if by sea) contained exactly one bit of information, reducing the entropy to zero. For another example, if there are four equally probable possibilities (each with probability 0.25), our initial entropy is:

$$H = -[0.25\times(-2) + 0.25\times(-2) + 0.25\times(-2) + 0.25\times(-2)] = 2\,\text{bits}$$

[because $\log_2(0.25) = -2$].

As you might expect, entropy is low when one of the possibilities has a very high probability and all the others have low probabilities. The converse is also true. Entropy is high when all possibilities are almost equally probable. Our objective in SA is to reduce entropy, that is, to sharpen the probability distribution so that we can select one of the possibilities with a specified level of confidence.

3.3.3 Mutual Information

A message from an environment is the result of a measurement of one or more parameters of something in the environment. The "information" contained in a message is the reduction in entropy produced by the message. So, if we receive a new message from the environment, we may revise our assessment of the current

probabilities and recompute the entropy. The difference between the new and old entropies is the information (in bits) contributed by the message.

If there are several parameters which could be measured, we want to choose the one that is likely to provide the most information (entropy reduction). The expected amount of information to be derived from a new parameter measurement is called the mutual information between the new information source and our current knowledge. Mutual information is measured in bits.

> **Definition 3.11:** *Mutual information* between a new message source and our current knowledge is the expected value of the information (entropy reduction) to be obtained by evaluating a message from the new source.

For example, suppose that, as the result of making a parametric observation of a situation, we have new probabilities for the situation in the above example: 0.125, 0.125, 0.25, 0.5. The new entropy is

$$H = -[0.125\times(-3) + 0.125\times(-3) + 0.25\times(-2) + 0.5\times(-1)] = 1.75\,\text{bits}$$

So this parameter measurement has reduced our entropy by 0.25 bit. The weighted average entropy reduction over all possible values of the new parameter is the mutual information between the new parameter and what we already know. In other words, the mutual information of a candidate information source is the amount to which we can expect the uncertainty (entropy) to be reduced by the source.

Although conditional entropy can tell us when two variables are completely independent, it is not an adequate measure of dependence. A small value for $H(Y|X)$ may imply that X tells us a great deal about Y or that $H(Y)$ is small to begin with. Thus, we measure dependence using mutual information.

> **Definition 3.12:** The *mutual information* between random variables X and Y is given by

$$\mathbb{I}(X;Y) \triangleq H(Y) - H(Y|X) \tag{3.17}$$

Mutual information is a measure of the reduction of randomness of a variable given knowledge of another variable. Using properties of logarithms, we can derive several equivalent definitions.

> **Property 3.7:** Equivalent definitions of mutual information:

$$\mathbb{I}(X;Y) = H(Y) - H(Y|X) \tag{3.18}$$

$$= H(X) - H(X|Y) \tag{3.19}$$

$$= H(X) + H(Y) - H(X,Y) \tag{3.20}$$

$$= \mathbb{I}(Y;X) \tag{3.21}$$

3.3.3.1 Information Payoff

If each candidate data source has a cost associated with it, we would, of course, compute the ratio of mutual information to cost. If the cost is time, we would come up with a rate of information production form the new source in bits per second.

> **Definition 3.13:** The *information payoff* from a candidate message source is the ratio of mutual information to the cost of using the candidate source.

Suppose we have a number of intelligence resources available to assess a situation. If the knowledge base includes statistical descriptions of each possible situation, we can compute the expected value of the information payoff from each resource. In this case, the cost might not be only time. It might be in lack of covertness, fuel, risk, or some other commodity. We would select the intelligence resources that have the highest information payoff, based on what we already know. This is done iteratively until the classification of the situation is successful or until all the intelligence resources have been used without success.

3.4 Information Channels

We introduce information channels in this section, and examine some of the properties of the more commonly used ones. A channel is a mathematical model of communications between two or more nodes. In our case these nodes represent battlespace entities, usually two or more OPFACS (operational facilities).

3.4.1 Channels

A *channel* is a correlation between two random variables. In a channel, source symbols from some finite alphabet are mapped into some sequence of channel symbols, which then produces the output sequence of the channel. Each of the possible input sequences induces a probability distribution on the output sequence. Shannon's model of the channel is illustrated in Figure 3.1.

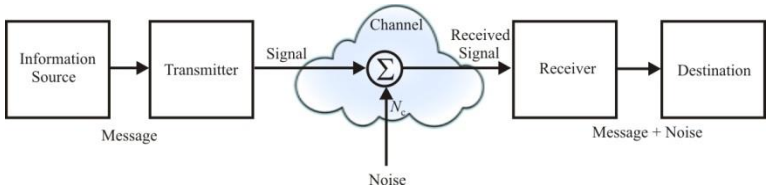


Figure 3.1 Shannon model of a communication channel.

Information theory is a probabilistic theory. The communications goal is to reconstruct the input from the output with a negligible probability of error. The maximum rate at which this can be done is called the *capacity* of the channel. The EW goal is to reduce this channel capacity to zero—that is, to cause as many errors as possible thereby precluding communications on that channel at all.

3.4.2 Discrete Channels

A *discrete channel* is a system consisting of an input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ and a probability transition matrix $\mathbf{P}(y|x)$ that expresses the probability of observing the output symbol y given that we send the symbol x. The channel is said to be *memoryless* if the probability distribution of the output depends only on the input at that time and is conditionally independent of previous channel inputs or outputs. A common model for a channel is therefore a *discrete memoryless channel* (DMC). The usual notation that is used to represent a discrete channel is the three-tuple $[\mathcal{X}, \mathbf{P}(y|x), \mathcal{Y}]$.

3.4.3 Coding

In theory, the channel capacity can be attained. First, however, we must introduce codes (another name for redundancy).

Codes

> **Definition 3.14:** An $(n, \mathcal{M})$ *code* for a DMC $[\mathcal{X}, \mathbf{P}(y|x), \mathcal{Y}]$ consists of the following:
>
> - An index set $\{1, 2, \ldots, \mathcal{M}\}$ representing the messages to be sent.

- An encoding function f: $\{1, 2, \ldots, \mathcal{M}\} \rightarrow X^n$, yielding *codewords* $x^n(1), x^n(2), \ldots, x^n(\mathcal{M})$. The set of codewords is called the *codebook*.
- A decoding function $\phi : Y^n \rightarrow \{1, 2, \ldots, \mathcal{M}\}$, which is a deterministic rule that assigns a guess to each possible received vector.

Example 3.2: The repetition code: $0 \rightarrow 000$, $1 \rightarrow 111$. ($\mathcal{M} = 2^1$, $n = 3$) code.

3.4.4 Channel Capacity

The highest rate of information that can be transmitted through a channel is called the *channel capacity*, and is denoted by C. The channel capacity is defined as:

Definition 3.15: The *channel capacity* is the maximum rate of reliable (error-free) information transmission through the channel.

Channel capacity is concerned with the information handling capacity of a given channel. It is affected by:

- The attenuation of a channel, which varies with frequency as well as channel length;
- The noise induced into the channel;
- Nonlinear affects such as clipping on the signal.

Some of the effects may change with time, for example, the frequency response of a copper cable changes with temperature and age, or a car driving down a road may change the reflection characteristics of a signal. We need a way to model a channel in order to estimate how much information can be passed through it.

Channel Capacity of a Discrete Memoryless Channel [1]

A classical theorem due to Shannon identifies the channel capacity of a DMC.

Definition 3.16: A channel $\Pr\{y^n \mid x^n\}$ is *memoryless* if

$$\Pr\{y^n \mid x^n\} = \prod_{i=1}^{n} \Pr\{y_i \mid x_i\} \tag{3.22}$$

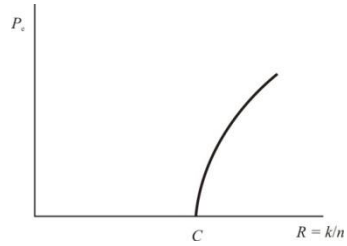


Figure 3.2 Channel capacity converse. If the rate R exceeds the channel capacity C, then the error rate increases from zero. Conversely, if the rate R is less than the channel capacity C, then there exists a data encoding method that ensures error-free transmission.

or, equivalently,

$$\Pr\{y_i \mid y^{i-1}, x^i\} = \Pr\{y_i \mid x_i\} \tag{3.23}$$

This definition means that the chance of being in state y_i, given that the previous state y^{i-1} is dependent on the input x_i and not on the previous state.

Property 3.8: [1**]** The channel capacity of a DMC is given by

$$C = \max_{p(x)} \mathbb{I}(X;Y) \tag{3.24}$$

where the maximum is taken over all possible input distributions $p(x)$ and $\mathbb{I}(X; Y)$ is the mutual information between the random variables X and Y. X is the input to the channel and Y is the output. Both are random variables.

Example 3.3: Noiseless Binary Channel. C = 1 bit.

Example 3.4: Noisy Channel with Nonoverlapping Outputs. C = 1 bit.

Example 3.5: Binary Symmetric Channel (BSC). $C = 1 - H(P)$ bits.

Example 3.6: If W = 3 kHz and S/N is maintained at 30 dB for a typical telephone channel, the channel capacity C is about 30 kbps.

3.4.5 Shannon's Channel Coding Theorem

Perhaps Shannon's most important contribution to the field of information theory is given by the channel coding theorem.

Shannon's noisy channel coding theorem states that if the information rate, R (bps) is equal to or less than the C, $(R \le C)$, then there is a coding technique that enables transmission over the noisy AWGN channel with no errors. The inverse of this is that if $R > C$, then the probability of error is close to 1 for every symbol (see Figure 3.2).

> **Definition 3.17**: A rate R is said to be *achievable* in a channel if there exists a sequence of $(n, 2^{nR})$ codes with $P_e^{(n)} \to 0$.

> **Definition 3.18:** The *capacity region* is the closure[5] of the set of achievable rates.

> **Property 3.9:** Noise-Free Channel Coding Theorem [Shannon]: For a DMC, all rates below capacity C are achievable. Specifically, for every rate $R < C$, there exists a sequence of (n, 2^{nR}) codes with negligible maximum probability of error. Conversely, any sequence (n, 2^{nR}) codes with negligible probability of error must have $R \le C$.

Although the existence of good codes has been proven, there is no method (as yet) for constructing them for arbitrary channels mainly because the existence proof is based on the idea of random coding.

The theorem implies that error-free transmission is possible if we do not send information at a rate greater than the channel capacity. Thus, the information capacity theorem defines the fundamental limit on the rate of error-free transmission for a power limited, bandlimited Gaussian channel. Letting W denote the RF bandwidth, S the signal power at the receiver, and $N/2$ the (two-sided) noise power PSD at the receiver (see Figure 3.1), we have:

> **Property 3.10:** Shannon's Noisy Channel Capacity Theorem: The channel capacity of a power limited, bandlimited AWGN channel is given by

[5] Given any subset, A, of a vector space X, the smallest closed set containing A is called the *closure* of A and is denoted by $\bar{A}$ or cl(A).

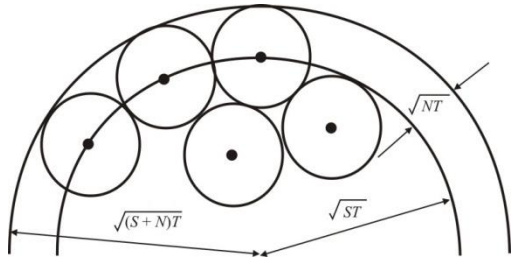


Figure 3.3 Signal space for calculating channel capacity. In this case $n = 3$.

$$C = W \log_2 \left(1 + \frac{S}{N}\right) \qquad \text{bps} \tag{3.25}$$

where

$$N = \int_{-W}^{W} \frac{N_0}{2} dw = N_0 W \tag{3.26}$$

Proof [2]

Suppose that we transmit one of a set of M equiprobable signals of bandwidth W in time T. Each signal thus represents $\log_2 M$ *bits*. According to the sampling theorem, each signal can be flawlessly represented by $n = 2WT$ samples in T seconds. Assume that the average signal power is S and the noise power is N. In the geometrical representation, all the transmitted signals must be restricted to an n-dimensional hypersphere of radius $\sqrt{ST}$ around the origin corresponding to their maximum energy. Similarly, all the received signals are restricted to an overall signal space of radius $\sqrt{(S+N)T}$. This is shown in two dimensions in Figure 3.3.

A noise power greater than NT will cause incorrect detection as this will cause the circles in Figure 3.3 to overlap, making distinguishing between the correct symbol and the adjacent symbol impossible if it falls into the region of overlap. In the presence of noise, the channel capacity can be determined by the number of signals that can be accommodated in the signal space.

The volume of an *n*-dimensional hypersphere is proportional to rn, where r is the radius of the hypersphere. Hence the number of signals with radius $(NT)^{n/2}$ that can be accommodated in an *n*-dimensional signal space of radius $[(S+N)T]^{n/2}$ is

$$M = \frac{[(S+N)T]^{n/2}}{(NT)^{n/2}} = \left[\frac{(S+N)T}{NT}\right]^{n/2} = \left(1+\frac{S}{N}\right)^{n/2}$$

Therefore, the information per signal is

$$\log_2 M \le \frac{n}{2}\log_2[1+S/N)]$$

and the channel capacity is

$$C = \frac{1}{T}\log_2 M$$
$$\le \frac{n}{2T}\log_2(1+S/N) = W\log_2(1+S/N)$$

This theorem states that signals can be sent through the channel as long as the information rate does not exceed *C*. It also implies that there is a code[6] that can be applied to the information stream that can achieve this capacity (although the theorem does not yield any clues as to how to find that code).

If the noise in the channel is characterized by the (two-sided) PSD given as $N_0/2$ watts/Hz, then the total noise power at the receiver is given by

$$N = N_0 W \tag{3.27}$$

(Note that *S* and *N* are numeric, not decibels.)

We can see from (3.25) that *C* increases as the available bandwidth increases and as SNR increases (improves). Equation (3.25) applies to information in any format and to both analog and data communications, but its application is most common in the latter. The channel capacity theorem is one of the most important results of information theory. In a single formula it highlights the interplay between three key system parameters:

[6] A code in this case is a generalization of what is typically called a code. Indeed, FEC coding is a coding technique, but in this case varying the amplitude may also be considered a code, for example. A code here means a "representation" of the signal.

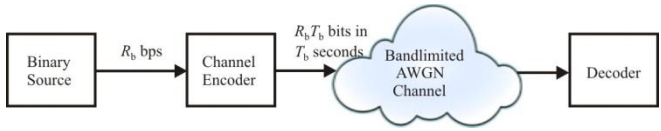


Figure 3.4 Shannon's error-free communication transmission system model.

- Channel bandwidth.
- Average transmitted or received signal power.
- Noise power at the channel output.

Figure 3.4 shows the general form of encoding scheme suggested by Shannon. A binary sequence of length R_b bits in a second are encoded into a binary sequence of length R_bT_b bits in T_b seconds before transmission. However, the design of the encoder and decoder is left unspecified by Shannon.

We see that the encoding time is T_b seconds. Therefore there is a encoding delay of T_b seconds in transmission and a decoding delay of T_b seconds at the receiver yielding a total delay of $2T_b$ seconds to encode these symbols. We can reduce the delay by decreasing the value of T_b, but we require more channel bandwidth for transmission.

3.4.6 Capacity Versus Bandwidth

It appears from (3.25) that as the bandwidth increases the capacity should increase proportionately. But this does not happen, because increasing the bandwidth, W, also increases the noise power $N = N_0W$, yielding:

$$\begin{aligned} C &= W\log_2\left(1+\frac{S}{N}\right) \\ &= W\log_2\left(1+\frac{S}{N_0W}\right) \\ &= \frac{S}{N_0}\frac{N_0W}{S}\log_2\left(1+\frac{S}{N_0W}\right) \\ &= \frac{S}{N_0}\log_2\left(1+\frac{S}{N_0W}\right)^{\frac{N_0W}{S}} \end{aligned}$$

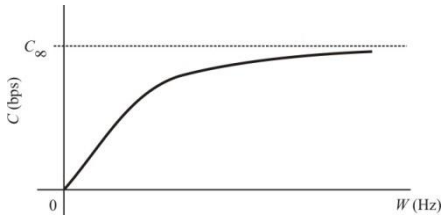


Figure 3.5 Channel capacity versus bandwidth.

$$= \frac{S}{N_0} \log_2 \left(1 + \frac{S}{N_0 W}\right)^{\frac{1}{S/N_0 W}}$$

Consider the case where an infinite bandwidth is available. Increasing W towards ∞ means that $S / N_0 W \to 0$. We know that

$$\lim_{x \to 0} (1 + x)^{1/x} = e$$

This means that as the bandwidth goes to infinity, S/N_0W goes to zero and $(1 + S / N_0 W)^{N_0 W / S}$ goes to e. The channel capacity therefore goes to

$$\lim_{W \to \infty} C = C_\infty = \lim_{W \to \infty} \frac{S}{N_0} \log_2 (1 + S / N_0 W)^{N_0 W / S}$$

$$= \frac{S}{N_0} \log_2 e$$

$$= 1.44 \frac{S}{N_0}$$

So as the bandwidth goes to infinity the capacity goes to $1.44S/N_0$, that is, it goes to a finite value and is not infinite.

The channel capacity variation with bandwidth is shown notionally in Figure 3.5.

> **Example 3.7:** A communication channel with a bandwidth of 4 kHz has a signal to noise ratio of 7. The bandwidth is reduced by 25%. To maintain the same channel capacity, how much must the signal power be increased?

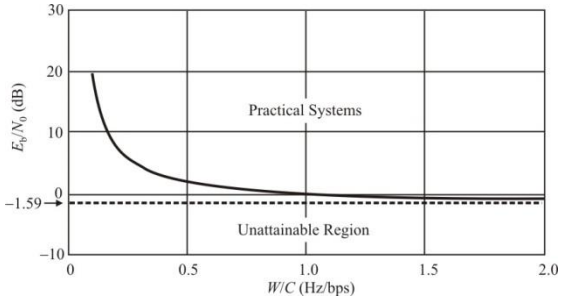


Figure 3.6 Shannon limit.

$$\frac{C}{C'} = 1 = \frac{4{,}000 \log_2\left(1 + \frac{S}{N}\right)}{3{,}000 \log_2\left(1 + \frac{S'}{N}\right)}$$

$$1 = \frac{4}{3} \frac{\log_2(8)}{\log_2\left(1 + \frac{S'}{N}\right)}$$

$$\frac{S'}{N} = 2^{\frac{4}{3}\times 3} - 1 = 2^4 - 1 = 15$$

So going from $S/N = 7$ to $S'/N = 15$ requires increasing S by 15–7 = 8 (9 dB).

3.4.7 Shannon Limit

For an ideal system that transmits data at a rate $R_b = C$, we have $S = E_b R_b = E_b C$. From (3.25) we have

$$\frac{C}{W} = \log_2\left(1 + \frac{E_b}{N_0}\frac{C}{W}\right)$$

Therefore,

$$\frac{E_b}{N_0} = \frac{2^{C/W} - 1}{C/W} \tag{3.28}$$

For infinite bandwidth, the E_b/N_0 approaches a limiting value as

$$\left.\frac{E_b}{N_0}\right|_{W\to\infty} = \lim_{W\to\infty}\frac{E_b}{N_0} = \ln 2 = 0.693 \leftrightarrow -1.59\,\text{dB} \tag{3.29}$$

Expression (3.29) is known as the *Shannon limit*.

A plot of (3.28) is shown in Figure 3.6. We can see that E_b/N_0 approaches –1.59 dB as the bandwidth ratio W/C gets larger.

3.4.8 Capacity of *M*-Point QAM Signals

Assume that in a channel, bandlimited to W Hz, each M-ary (M-point) *quadrature amplitude modulation* (QAM) signal symbol has a duration of T seconds [3, 4]. We can represent each M-point QAM signal by $\log_2 M$ bits. Thus, we have $\log_2 M$ bits/symbol and $1/T$ symbols/s, so the transmission rate R_b is

$$R_b = \frac{\log_2 M}{T} \qquad \text{bps} \tag{3.30}$$

Suppose that the bandwidth of the M-ary QAM signal is set equal to the channel bandwidth, W. The null-to-null bandwidth of the M-ary QAM signals is $W = (f_c + 1/T) - (f_c - 1/T) = 2/T$, where f_c is the carrier frequency. Therefore, we can express the transmission rate of (3.30) as

$$R_b = \frac{W}{2}\log_2 M \quad \text{bps} \tag{3.31}$$

For a fixed spacing between adjacent signals, increasing the value of M also increases the average transmitted signal power S. Accordingly, we increase the SNR.

Let $M = K'S/N$, where K' varies with error rate and is a constant small enough to achieve negligible error rate. Then

$$R_b = \frac{W}{2}\log_2\left(K'\frac{S}{N}\right) \qquad \text{bps} \tag{3.32}$$

The capacity of an M-ary QAM system approaches the Shannon channel capacity C if the average transmitted signal power in the QAM system is increased by a factor of $1/K'$.

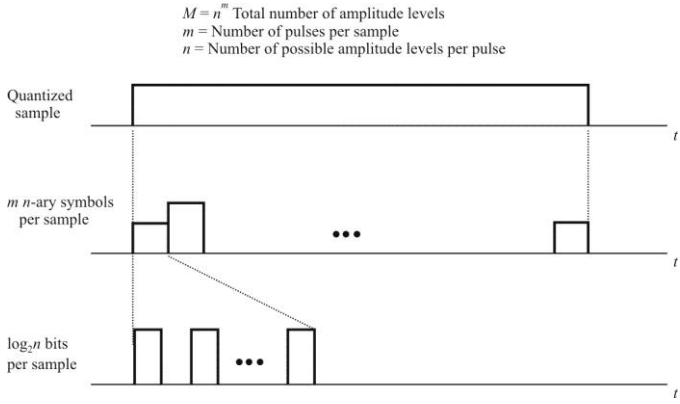


Figure 3.7 Representations of quantized sample.

3.4.9 Capacity of an *n*-ary PCM System

Assume that an input analog signal of bandwidth W Hz is sampled at the minimum Nyquist sampling rate of $2W$ samples/s and the samples are uniformly quantized to $M = n^m$ levels. We can represent each M-level signal sample by m n-ary symbols as illustrated in Figure 3.7.

Thus, we have $2W$ samples/s, $M = n^m$ levels/sample, m symbols/sample, $\log_2 n$ bits/n-ary symbol, and $m \log_2 n$ bits/sample. The symbol rate is $2W$ m symbol/s and the information transmission rate is

$$R_b = 2Wm \log_2 n \qquad \text{bps} \tag{3.33}$$

For error-free transmission, the channel capacity $C \geq R_b$. We can see that for fixed values of n and m, the capacity R_b is proportional to W.

Let S be the average transmitted signal power and a be the spacing between the n-levels. We assume that the n discrete levels are equally likely and have the values $\pm a/2, \pm 3a/2, \ldots, \pm(n-1)a/2$. The average transmitted signal power is

$$\begin{aligned} S &= \frac{1}{n}\left\{\left(\frac{a}{2}\right)^2 + \left(\frac{3a}{2}\right)^2 + \ldots + \left[\frac{(n-1)a}{2}\right]^2\right\} \times 2 \\ &= a^2 \frac{n^2 - 1}{12} \end{aligned} \tag{3.34}$$

Rearranging (3.34) to express n in terms of S and substituting into (3.33), we get

$$R_b = W \log_2 \left(1 + \frac{12S}{a^2}\right) \tag{3.35}$$

To maintain a negligible error rate, there must be a finite and adequate separation a between adjacent n-ary levels. Call this separation $a = K\sigma$, where K varies with the error rate and is a constant large enough to allow recognition of individual levels with negligible error rate, and $\sigma^2 = N$ is the noise power. Then

$$R_b = W \log_2 \left(1 + \frac{12S}{K^2 N}\right) \quad \text{bps} \tag{3.36}$$

From (3.36) we see that bandwidth can be traded for SNR for a system with given channel capacity $C = R_b$.

Equation (3.36) is identical to the Shannon channel capacity expression if the average transmitted signal power in the *pulse code modulation* (PCM) system is increased by a factor of $K^2/12$.

For a given average transmitted power S and channel bandwidth W, we can transmit information at the rate C bps with no error, by employing sufficiently complex coding systems. It is not possible to transmit at a rate higher than C bps by any coding system without a definite probability of error. Hence the channel capacity theorem defines the fundamental limit on the rate of error-free transmission for a power-limited, band-limited channel.

Example 3.8: A *public switched telephone network* (PSTN) has a bandwidth of 3.4 kHz.

(a) The capacity of the channel for an S/N = 30 dB is given by:
$(S / N_{\text{dB}} = 30\,\text{dB} \Rightarrow S / N = 1{,}000)$

$$C = 3{,}400 \log_2 (1 + 1{,}000) = 33{,}898\,\text{bps}$$

[Note that $\log_2(1{,}001) = 9.97$.]

(b) The minimum SNR required for information transmission through the channel at the rate of 4,800 bps is

$$\text{SNR} = 2^{C/W} - 1 = 2^{4{,}800/3{,}400} - 1 = 2.66 \leftrightarrow 4.2\,\text{dB}$$

(c) The minimum SNR required for information transmission through the channel at the rate of 9,600 bps is

$$\text{SNR} = 2^{C/W} - 1 = 2^{9{,}600/3{,}400} - 1 = 6.1 \leftrightarrow 7.9\,\text{dB}$$

3.4.10 Capacity of Frequency-Hopped Code-Division Multiple-Access Channels

Goh and Maric considered the capacities of *frequency-hopped* (FH) *code-division multiple-access* (CDMA) channels [5]. In this section we present a summary of some of their findings.

The type of modulation scheme usually used in FH systems is *M*-ary *frequency-shift keying* (MFSK), where frequently $M = 2$. In this scheme, Q MFSK channels are used to transmit the messages. The *M*-ary modulated messages are transmitted using orthogonal FSK signals, and then hopped to one of the Q MFSK channels. The capacity of the multiple-access communication system is calculated by reducing the problem to a single-user channel. The single-user channel is then modeled as a multiple access interference channel subjected to AWGN and Rayleigh fading.

3.4.10.1 MFSK FH-CDMA Channel

In this section we investigate the capacity region of the MFSK FH-CDMA communication system. We assume the hopping patterns used by all the users are independent sequences, and the *i*th receiver is only interested in the message transmitted by the *i*th transmitter, that is, no cooperation between users either at the encoder or at the decoder is allowed.

In an MFSK FH system, two types of hopping are possible: *fast FH* (FFH) where the hop (or chip) rate is an integer multiple of symbol rate, and *slow FH* (SFH) where the symbol rate is an integer multiple of hop rate. It is assumed that both types of systems have the same chip rate R_c, where R_c is defined as $\max(R_h, R_s)$. We see that FFH is no different from a repetition coding and, since a repetition code is not a good coding method and we are interested in finding the capacity of the system, we only consider SFH.

The single-user model is as follows. There are K users, each transmitting over a bandwidth which is divided into Q MFSK channels. The data is then encoded and modulated using one of the *M*-ary signals. The symbols are then frequency-hopped to one of the Q MFSK channels such that there are N_S encoded symbols per hop. We assume chip asynchronization. We further assume random synchronous hopping patterns, and the codewords are fully interleaved. It was shown in [6] that the probability of a hit (partial hit or full hit) is given by

$$p_{\mathrm{h}} = \frac{1}{Q}\left[1 + \frac{1}{N_{\mathrm{S}}}\left(1 - \frac{1}{Q}\right)\right] \tag{3.37}$$

Since $N_{\mathrm{S}}, Q >> 1, \quad p_{\mathrm{h}} \approx 1/Q.$ Then the probability of one or more hits from the other $K - 1$ signals is

$$P \approx 1 - \left(1 - \frac{1}{Q}\right)^{K-1} \tag{3.38}$$

For the AWGN and flat fading channels, the symbol error probability P_{e} is given by

$$P_{\mathrm{e}} = P_0(1 - P) + P_1 P \tag{3.39}$$

where P_0 is the conditional probability of error for one of the symbols in a codeword given that there are no hits, which can be expressed as [7]

$$P_0 = \begin{cases} \displaystyle\sum_{m=1}^{M-1} \frac{(-1)^{m+1}\binom{M-1}{m}}{m+1} \exp\left(-\frac{m\gamma}{m+1}\right), & \text{AWGN} \\ \displaystyle\sum_{m=1}^{M-1} \frac{(-1)^{m+1}\binom{M-1}{m}}{m+1+m\gamma_0}, & \text{Rayleigh fading} \end{cases} \tag{3.40}$$

and P_1 is the conditional probability of error of that symbol given that there is at least one hit, γ denotes the SNR, and γ_0 denotes the mean SNR in the case of Rayleigh fading.

After dehopping, the demodulator in each receiver consists of M-branch bandpass filters followed by envelope detectors. Hard decisions are then made at the output of the envelope detectors. If the multiple-access interference is ignored, each transmitter–receiver pair can be modeled as an individual M-ary single-user channel as illustrated in Figure 3.8(a). The frequency–time diagram of this MFSK SFH-CDMA scheme is shown in Figure 3.8(b).

An upper bound for the symbol error probability can be obtained by assuming that the conditional error probabilities of a symbol equals 1/2 given that there is one hit and $(M-1)/M$ if there at least two hits from other users. The symbol error probability can then be expressed as

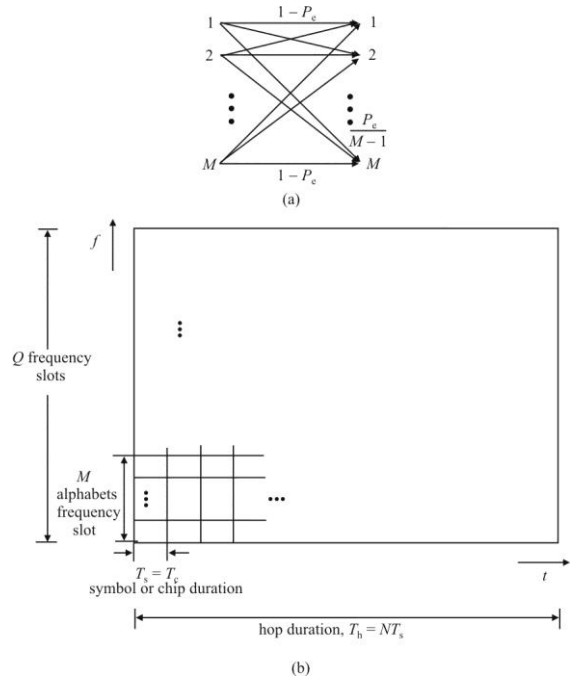


Figure 3.8 (a) Channel model of the MFSK frequency-hopping system. (b) Frequency–time diagram.

$$P_e \le P_0(1-P) + (K-1)\frac{1}{Q}\left(1-\frac{1}{Q}\right)^{K-2}\left(\frac{1}{2}+\frac{M-2}{2(M-1)}P_0\right) + \frac{M-1}{M}\left[P-(K-1)\frac{1}{Q}\left(1-\frac{1}{Q}\right)^{K-2}\right] \tag{3.41}$$

The first, second, and third terms in (3.41) correspond to error probabilities due to no hits, one hit, and at least two hits, respectively.

For BFSK FH-CDMA (a common case), an approximation for (3.39) is given in [6] which considers the situation when there is only either no hit or one hit. This approximation is accurate only when Q/K is large, however. To prevent the symbol error probability from being too pessimistic, a more accurate approximation of (3.39) for $M = 2$ (which assumes the conditional probability of error of the bit is 1/2 only if there are two hits or more than two hits) is used

$$P_e \approx P_0(1-P) + (K-1)\frac{1}{Q}\left(1-\frac{1}{Q}\right)^{K-2}\left(\frac{1}{8}+\frac{3}{4}P_0\right) + \frac{1}{2}\left[P-(K-1)\frac{1}{Q}\left(1-\frac{1}{Q}\right)^{K-2}\right] \tag{3.42}$$

$$= P_0(1-P) + (K-1)\frac{1}{Q}\left(1-\frac{1}{Q}\right)^{K-2}\left(\frac{3}{4}P_0-\frac{3}{8}\right)+\frac{1}{2}P \tag{3.43}$$

3.4.10.2 Capacity Region

In the previous section, the multiple-access channel was modeled as K individual single-user channels, since there is no cooperation between the users at the encoder and decoder. We can therefore calculate the capacity of the multiple-access channel as the sum of the capacities of each individual single-user channel.

Capacity of MFSK FH-CDMA Channel

The capacity of the ith equivalent single-user MFSK FH-CDMA channel is given by

$$C_i = \log_2 M - h_M(P_e) \tag{3.44}$$

where

$$h_M(x) = -x\log_2\left[\frac{x}{M-1}\right]-(1-x)\log_2(1-x) \tag{3.45}$$

The sum of the K individual capacities yields the *sum capacity*, which is given by

$$C_{sum} = K[1-h(P_e)] \tag{3.46}$$

where

$$h(x) = -x\log_2 x-(1-x)\log_2(1-x)$$

is the binary entropy function. The *normalized sum capacity* is defined as the capacity per channel (where we have Q MFSK channels). Letting $K=\lambda Q$ with λ fixed, the *maximum normalized sum capacity* is defined as

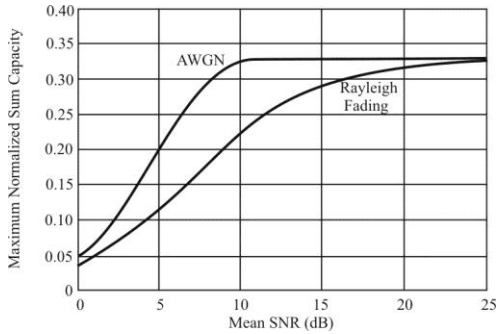


Figure 3.9 Maximum normalized sum capacity as a function of the mean SNR of FSK for a slow FH system.

$$\frac{C}{K} \triangleq \max_{q}\left[h(q^{K}) - qh(q^{K-1})\right] \qquad \text{bits/dimension} \qquad (3.47)$$

where $q = 1 - 1/Q$.

C / K reflects the maximum bit rate that can be transmitted per channel with arbitrarily small probability of error. It was shown in [8] that when K is large, the capacity (bits/dimension) of the channel approaches $(\ln 2)/K$. C can be expressed as

$$C = \max_{\lambda} \lim_{K\to\infty} \frac{\lambda C_{\text{sum}}}{K} \max_{\lambda} \lim_{K\to\infty} \lambda\left[1 - h_{\text{M}}(P_{\text{e}})\right] \qquad (3.48)$$

3.4.10.3 Numerical Results

For the BFSK SFH-CDMA scheme, the degradation in Rayleigh fading is shown in Figure 3.9. We can see that the capacity degradation at mean SNR of 20 dB and higher is insignificant. The capacity decreases by more than 20% only when the mean SNR is below 12 dB.

The normalized sum capacity as a function of the number of users with Q = 64 for the case of no additive noise is shown in Figure 3.10, while Figure 3.11 illustrates the normalized sum capacity in the case of AWGN and Rayleigh fading with mean SNR = 10 dB and Q = 63. The curve in Figure 3.10 peaks at a slightly higher value than the AWGN curve in Figure 3.11.

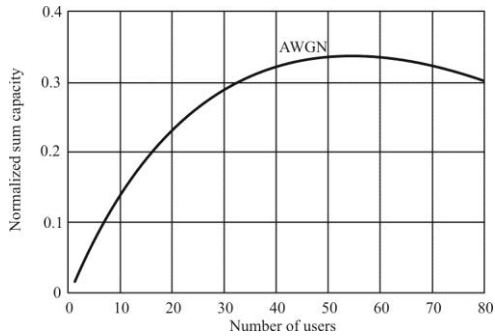


Figure 3.10 Normalized sum capacity as a function of the number of users, with $\text{SNR} = \infty$ and Q = 63.

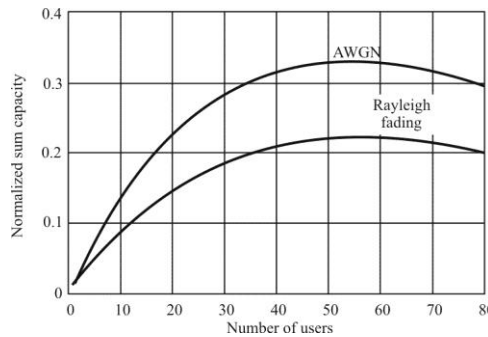


Figure 3.11 Normalized sum capacity as a function of the number of users for BFSK SFH in AWGN and Rayleigh fading channels, with mean SNR = 10 dB and Q = 63.

3.4.10.4 Summary

In this section, we investigated the multiple-access capabilities of SFH MFSK CDMA channels, with particular emphasis on BFSK. This emphasis is because that is the most prolific modulation scheme for this class of communications. We approached the problem by first determining the single user capacity and modeled the multi-user channel as the sum of the individual single user channels.

Channel degradation consisting of AWGN and Rayleigh fading were considered. As expected, ceteris paribus, Rayleigh fading caused more degradation than AWGN.

3.4.11 Data Processing Theorem

It has been mathematically proven that processing information after it has been generated cannot increase the amount of information present. This is known as the *data processing theorem* in information theory [9]. This is not to be taken to imply that more information cannot be extracted from an element of available information. It could very well be the case that such an element may contain inherent information that has not previously been extracted. The data processing theorem simply states that if all the information has been discovered, no amount of additional processing of the element will increase the information content.

3.5 Common Channel Models

In this section we describe some common channel models and discuss some of their properties, such as capacity.

We are primarily concerned about two types of channels: the AWGN channel and the discrete memoryless channel. A practical communication system is depicted in Figure 3.12. One of 2^k messages is to be sent from the transmitter to the receiver. The message is first encoded with the sender's codebook. This codebook contains 2^k codewords of length n where k is the number of information bits transmitted in n channel uses.

> **Definition 3.19:** Code Rate. The *rate R* of a code is the ratio k/n where k is the number of information bits transmitted in n channel uses.

> **Property 3.11:** Channel Capacity [1]. For $R \le C,$ encoding methods exist with decoding error probability $P_e \to 0.$

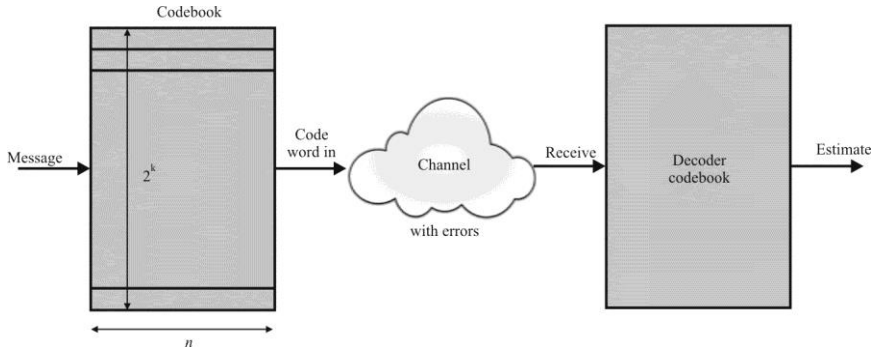


Figure 3.12 Communication system.

3.5.1 Encoding and Decoding

Let the code consist of 2^k binary codewords where Pr{0} = Pr{1} = ½. Channel errors occur with probabilities $\Pr\{0 \to 1\} = \Pr\{1 \to 0\} = P$. That is, the number of error sequences ~$2^{nh(P)}$. The decoder searches around the received sequence for a codeword with ~nP differences (see Figure 3.13).

A decoding error occurs for either of the cases shown in Figure 3.14.

- In Figure 3.14(a), for t-errors $|t/n - P| > \varepsilon$. By the law of large numbers (see Appendix A), this error approaches zero as $n \to \infty$.
- In Figure 3.14(b) there is more than one code word in the region.

$$\Pr\{>1\} \approx (2^k - 1)\frac{2^{nh(P)}}{2^n}$$
$$\to 2^{-n[1-h(P)-R]} = 2^{-n(C_{\text{BSC}}-R)} \to 0$$

for

$$R = \frac{k}{n} < 1 - h(P)$$

and

$$n \to \infty$$

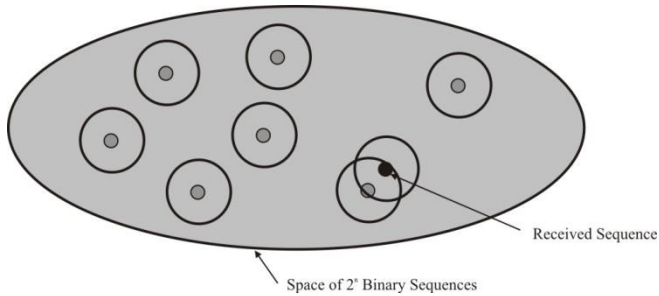


Figure 3.13 Space of 2^n code sequences.

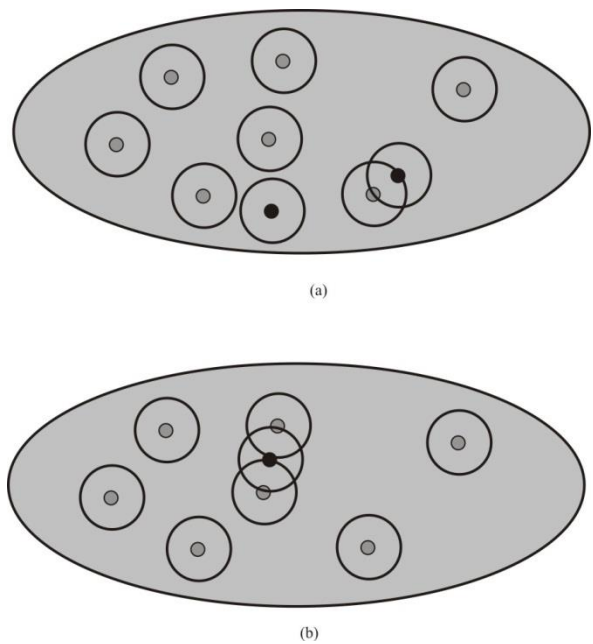


Figure 3.14 Decoding error probability: (a) *t*-errors and (b) more than one codeword in region.

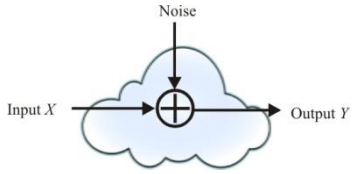


Figure 3.15 AWGN channel.

3.5.2 Capacity for Additive White Gaussian Noise Channels

Consider the channel shown in Figure 3.15. The capacity of this channel is given by

$$C = \sup_{p(x)} \left[H(Y) - H(\text{Noise}) \right]$$

The input X is Gaussian with variance (power) $\sigma_x^2 \le S/2W$. The noise is Gaussian with variance (power) σ_{noise}^2. The output Y is Gaussian with variance (power) $\sigma_y^2 = S/2W + \sigma_{\text{noise}}^2$.

Now

$$\begin{aligned} C &= \frac{1}{2}\log_2[2\pi(\sigma_x^2 + \sigma_{\text{noise}}^2)] - \frac{1}{2}\log_2(2\pi\sigma_{\text{noise}}^2) \text{ bits/transmission} \\ &= \frac{1}{2}\log_2\left(\frac{\sigma_{\text{noise}}^2 + \sigma_x^2}{\sigma_{\text{noise}}^2}\right) \quad \text{bits/transmission} \\ &= W\log_2\left(\frac{\sigma_{\text{noise}}^2 + S/2W}{\sigma_{\text{noise}}^2}\right) \quad \text{bits/sec} \end{aligned}$$

so

$$p(y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-y^2/2\sigma_y^2} \tag{3.49}$$

and

$$H(Y) = \frac{1}{2}\log_2(2\pi e\sigma_y^2) \quad \text{bits} \tag{3.50}$$

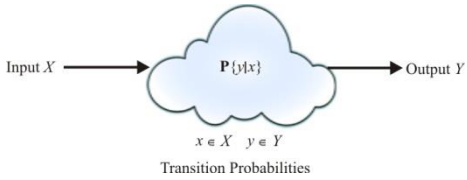


Figure 3.16 Discrete memoryless channel model.

3.5.3 Memoryless Channels

For a memoryless channel, output at time *i* depends only on input at time *i* and the input and output alphabets are finite. Such a channel is depicted in Figure 3.16. The channel is completely characerized by the transition probabilities $\Pr\{y|x\}$, which are the probabilities that *y* emerges given that *x* was input to the channel.

3.5.4 Binary Channels

The channel corresponding to the transition diagram shown in Figure 3.17 is known as a *binary channel* because there are two inputs {0, 1} and only two outputs {0, 1}. When errors do not occur, we have[7] $\Pr\{y=1|x=1\}=1-p$ and $\Pr\{y=0|x=0\}=1-q$. p and q are known as the error probabilities or error rates. Errors occur in the channel and we have $\Pr\{y=1|x=0\}=p$ while $\Pr\{y=1|x=0\}=q$.

3.5.5 Binary Symmetric Channel

The channel shown in Figure 3.18 is known as the *binary symmetric channel* (BSC). *E* is the binary error sequence such that $\Pr\{1\} = 1 - \Pr\{0\} = p$. *X* is the binary information sequence, while *Y* is the binary output sequence. The probability of error is given by $\Pr\{e\}=\Pr\{y=0|x=1\}=\Pr\{y=1|x=0\}=p$. The channel is *symmetric* because these two probabilities are equal, that is, the probability of a zero emerging given that a one was input is the same as the probability of a one emerging given that a zero was input.

Converting an AWGN channel to a BSC is illustrated in Figure 3.19. The symbol pdfs are given by

[7] We use a lowercase *p* and *q* in this and subsequent sections to denote probabilities as that is common in the information theory literature. $p(x)$ denotes the pdf of *x*.

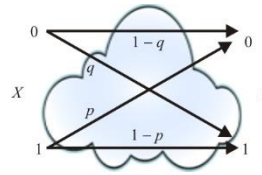


Figure 3.17 Binary channel.

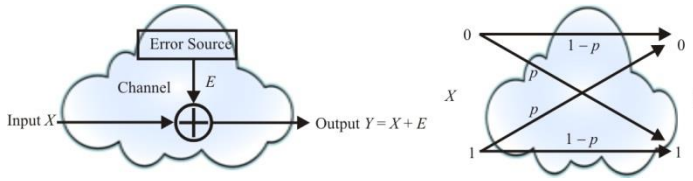


Figure 3.18 Binary symmetric channel.

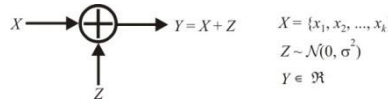


Figure 3.19 AWGN to BSC conversion.

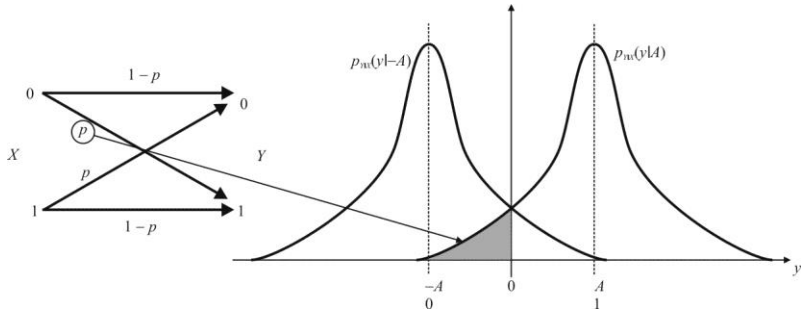


Figure 3.20 AWGN to BSC conversion. The crossover probability in the BSC flow diagram on the left is given by the shaded area in the AWGN tail on the right.

$$p_{Y|X}\left(y\middle|X=x_k\right)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(y-x_k)^2}{2\sigma^2}\right] \tag{3.51}$$

The crossover probability of the BSC, given by p, is illustrated as the tail probability in Figure 3.20 as the shaded area. That is,

$$p=\int_{-\infty}^{0} p_{Y|X}\left(y\middle|X=A\right)dy \tag{3.52}$$

The binary entropy $h(p)$ is given by

$$h(p)=-p\log_2 p-(1-p)\log_2(1-p) \tag{3.53}$$

Equation (3.53) is sketched in Figure 3.21. Note that $h(p) = h(1 - p)$.

From (3.24) we know that the channel capacity is given by $\max_{p(x)} I(X;Y)=H(X)-H(X\,|Y)$. The maximum of $H(X) = 1$ and since X is binary, $H(X\,|Y)=h(p)$ so the capacity for the BSC is

$$C_{\text{BSC}}=1-h(p) \tag{3.54}$$

This channel capacity is illustrated in Figure 3.22.

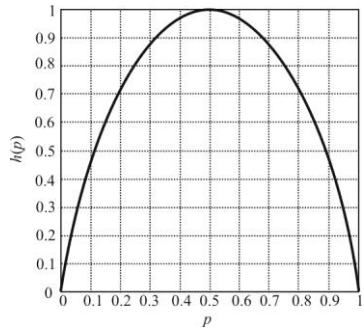


Figure 3.21 Binary entropy $h(p)$.

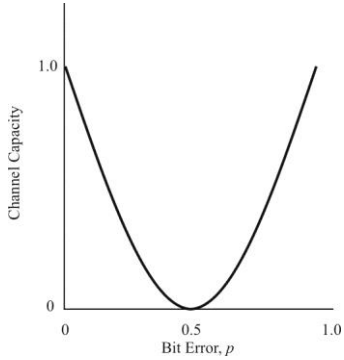


Figure 3.22 BSC channel capacity.

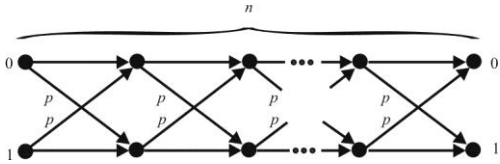


Figure 3.23 Cascade of *n* BSCs.

Property 3.12: A cascade of *n* identical BSCs, that is *n* BSCs in series (see Figure 3.23) each with crossover probability *p* is equivalent to a BSC with crossover probability

$$p_n = \frac{1}{2}[1-(1-2p)^n] \tag{3.55}$$

and hence

$$\lim_{n\to\infty} I(X_0; X_n) = 0$$

3.5.6 Erasure Channel

The erasure channel shown in Figure 3.24 allows for errors in the decoder so that if the decoder cannot decide which of 0 or 1 is present at the receiver, instead of deciding $\Pr\{0|1\}$ with probability *p* or $\Pr\{1|0\}$ with probability *p*, the decoder does not decide on either. It essentially says "I don't know." That symbol is then dropped and no decision is made. Therefore, $\Pr\{E|0\} = \Pr\{E|1\} = e$.

The erasure channel is a popular model for a CDMA channel. The reason for this is illustrated in Figure 3.25. CDMA signals share the same segment of the spectrum and are noise-like in character. Therefore, the detection decision region is frequently around zero, represented by the two vertical lines in Figure 3.25(b). An erasure channel would erase decisions in that region, as illustrated in Figure 3.25(c).

3.5.6.1 Channel Capacity

The channel capacity of the erasure channel is found as follows:

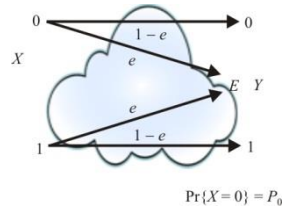


Figure 3.24 Erasure channel model.

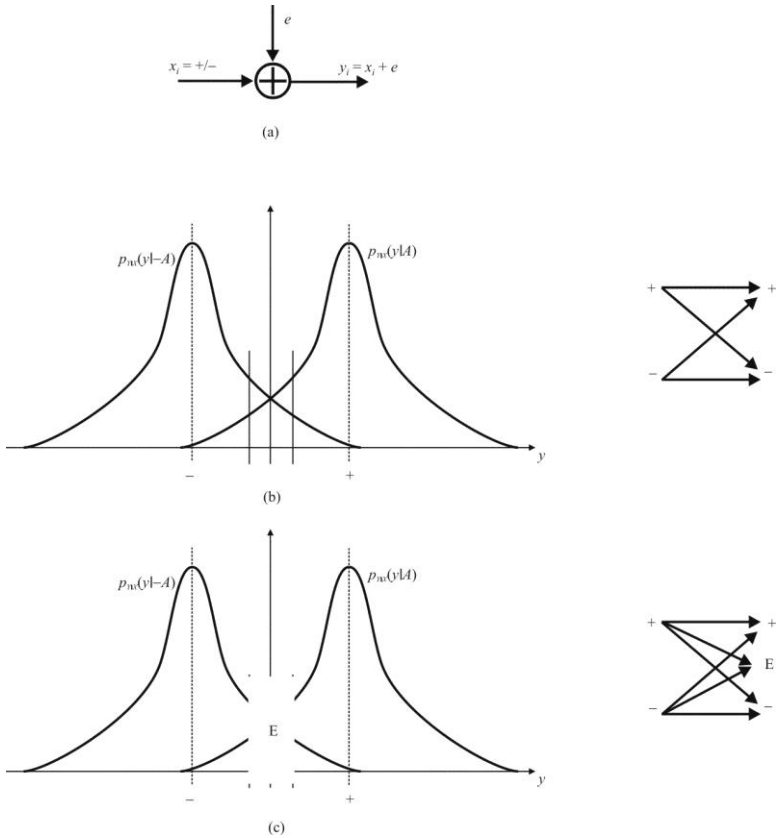


Figure 3.25 AWGN to BSC to erasure. (a) The AWGN channel model; (b) the BSC without erasure demonstrating the region around zero where decisions can be problematic; and (c) the erasure channel where the region around zero is erased.

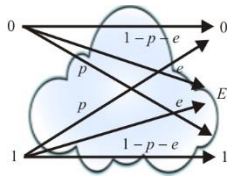


Figure 3.26 Erasure channel with errors.

$$I(X;Y) = H(X) - H(X|Y)$$
$$H(X) = h(P_0)$$
$$H(X|Y) = eh(P_0)$$

Therefore,

$$C_{\text{erasure}} = 1 - e \tag{3.56}$$

An adaptation of the erasure channel is the erasure channel with errors depicted in Figure 3.26. The functioning is the same as the BSC except that erasures are allowed with probability *e*.

3.5.6.2 Capacity and Coding for the Erasure Channel

The code is 2^k binary codewords where $P(0) = P(1) = 1/2$. The channel errors are characterized by $\Pr\{0 \to E\} = \Pr\{1 \to E\} = e$. The decoder searches around the received sequence for a codeword with $\approx ne$ differences (see Figure 3.13 again).

For *t* erasures such that

$$|t/n - e| > \varepsilon$$

which $\to 0$ as $n \to \infty$ by the law of large numbers [Figure 3.14(a)]. We see in Figure 3.14(b) that more than one candidate codeword agrees in $n(1-e)$ positions after the *ne* positions are erased (the codewords are random). We have

$$P(>1) \approx (2^k - 1)2^{-n(1-e)}$$
$$\to 2^{-n(1-e-R)} = 2^{-n(C_{\text{erasure}} - R)} \to 0$$

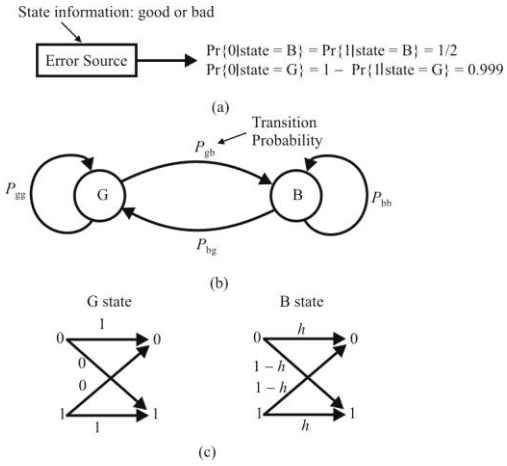


Figure 3.27 Burst error channel. (a) The error source outputs are dependent on previous outputs. (b) Flow diagram for burst error channel. (c) Flow diagram.

for

$$R = \frac{k}{n} < 1 - e, \qquad n \to \infty$$

3.5.7 Burst Error Model (Gilbert-Elliott Channel)

The Gilbert-Elliot channel [10] is a time-varying BSC the crossover probabilities of which are determined by the current state of a discrete time stationary binary Markov process (see Figure 3.27). The states are appropriately designated G for good and B for bad. Bit error models generate a sequence of noise bits (where 0s represent good bits and 1s represent bit errors) to produce output bits. These models are grouped into two classes: memoryless models and models with memories. In memoryless models the noise bits are produced by a sequence of independent trials where each trial has the same probability Pr{0}of producing a correct bit and probability Pr{1} = 1 – Pr{0} of producing a bit error.

To model channels with memory, a commonly used technique is a Markov chain. This technique helps to make the bit error probability depend on the states. The use of Markov chains in bit error models was introduced by Gilbert and Elliott [11, 12]. In state G, transmission is error-free and in state B the link has the probability *h* of transmitting a bit correctly. Figure 3.27 shows a transition diagram and bit error probabilities for the Markov chain. The model has three independent parameters (P_{bg}, P_{gb}, and h) to describe the error performance. Transition jumps

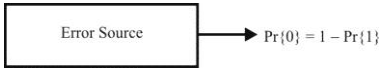


Figure 3.28 Random error channel. The error source outputs are independent.

from B to G with probability P_{bg} and transition jumps from G to B with probability P_{gb}. The states B and G tend to persist and the model simulates bursts of errors.

Due to the underlying Markov nature of the channel, it has memory that depends on the transition probabilities between the states. The capacity of the channel is denoted C_μ, where μ is a measure of memory. When the one-dimensional statistics of the channel are fixed, C_μ increases monotonically with μ and converges asymptotically to a value C^{SI} which is the capacity of the same channel when side information about its instantaneous state is available to the receiver.

In the random error channel, the outputs from the error source are independent (see Figure 3.28). In the burst error channel, the current output is dependent on previous outputs [see Figure 3.27(a)]. The flow diagram for the Gilbert-Elliot channel is shown in Figure 3.27(b).

For the Gilbert-Elliot channel we have

$$P_{gb} = 1 - P_{gg} \tag{3.57}$$

and

$$P_{bg} = 1 - P_{bb} \tag{3.58}$$

Clearly we have

$$\Pr_X\{X(i) = g \,|\, X(i-1) = g\} = P_{gg}$$
$$\Pr_X\{X(i) = b \,|\, X(i-1) = g\} = P_{gb}$$
$$\Pr_X\{X(i) = g \,|\, X(i-1) = b\} = P_{bg}$$
$$\Pr_X\{X(i) = b \,|\, X(i-1) = b\} = P_{bb}$$

We know that reliable communication over a finite state channel is theoretically possible at any rate below capacity. In use, however, two practical difficulties arise. First, much less is known about good codes for such channels than for memoryless ones; second, the length (and therefore the decoding complexity) of such codes depends on the length of the channel memory. This is

apparent from the fact that the error exponent for channels with memory depends on the block length N_B whereas for memoryless channels it is independent of N_B.

The parameters P_{bg}, P_{gb}, and h are not directly observable and therefore must be determined from statistic measurements of the error process. It is also important to note that runs of G alternate with runs of B. The run length has geometric distributions, with mean $1/P_{gb}$ for the G-runs and $1/P_{bg}$ for the B-runs.

3.5.7.1 Geometric Distribution

The time fraction in both G and B states based on persistence in each state can be calculated. For example, the fraction of time spent in the B state is

$$P(B) = \frac{P_{gb}}{P_{gb} + P_{bg}} \tag{3.59}$$

The sequence of states cannot be reconstructed from the sequence of bits in the error process, because both 0s and 1s (the good bits and bad bits) are produced in the B state and since bit errors happen only in state B with probability of $1 - h$ then the probability of error is

$$P(1) = P(1, B) = P(B)P(1/B) = (1-h)\frac{P_{gb}}{P_{gb} + P_{bg}} \tag{3.60}$$

The bits of the error process (runs of 0s and 1s) and the distribution of run lengths of 0s (error gaps) and 1s (error bursts) are observable to determine model parameters.

3.5.8 Broadcast Channels

The broadcast channel involves the simultaneous communication of information from one sender to multiple receivers as shown in Figure 3.29.[8] The goal is to find the capacity region, that is, the set of simultaneously achievable rates (R_1; R_2) [13]. Military tactical push-to-talk communications can be well modeled as broadcast channels. We will show that this is a good model for analyzing communication channels with an ES function present, intercepting the communication transpiring over that network. In this section we present some of their salient characteristics in

[8] In general, we can have $k > 2$ receivers. We are interested in ES performance, where there is likely only one ES receiver involved, so limiting the analysis to two receivers (one target, one ES) is adequate.

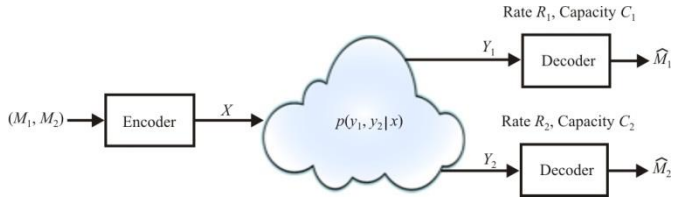


Figure 3.29 Broadcast channel model.

preparation for the performance analysis presented in Chapter 9 on intercept channels.

Intuitively, it is clear that it is possible to transmit to both receivers at a rate equal to the minimum of the two capacities, C_1 and C_2 (i.e., the transmission rate is limited by the worst channel). At the other extreme we could transmit on the best channel at a rate equal to its capacity and transmit no information on the other channel. Assuming $C_1 \le C_2$, the rates $R_1 = \alpha C_1$ and $\alpha C_1 + (1-\alpha)C_2$ can be achieved; this is called *time-sharing*. Cover showed that it is possible to do better than time-sharing [14]. The formal definitions for the broadcast channel are discussed in the next section.

3.5.8.1 Broadcast Channel Model

Definition 3.20: A *broadcast channel* (BC) consists of an input alphabet $\mathcal{X}$ and two output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ and a probability transition function $p(y_1, y_2|x)$. For a *memoryless* BC

$$p(y_1^n, y_2^n|x^n) = \prod_{i=1}^{n} P(y_{1i}, y_{2i}|x_i) \tag{3.61}$$

Definition 3.21: A $[n, (2^{nR_1}, 2^{nR_2})]$ code for a BC with independent information consists of an encoder,

$$X : (1, 2, \cdots, 2^{Rn_1} \times 1, 2, \cdots, 2^{nR_2}) \to \mathcal{X}^n \tag{3.62}$$

and two decoders,

$$\phi_i : \mathcal{Y}_i^n \to 1, 2, \cdots, 2^{nR_2}, \qquad i = 1, 2 \tag{3.63}$$

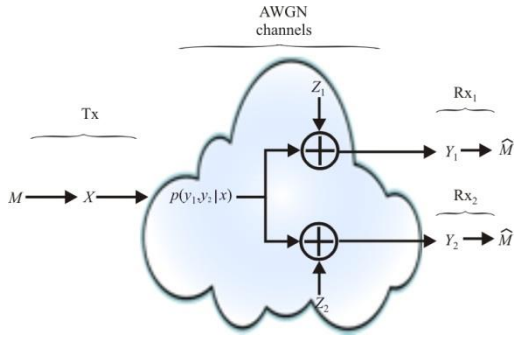


Figure 3.30 AWGN broadcast channel.

The average probability of error is defined as

$$P_e^{(n)} \triangleq \Pr\{\phi_1(Y_1^n) \neq M_1 \text{ or } \phi_2(Y_2^n) \neq M_2\} \tag{3.64}$$

where (M_1, M_2) are assumed to be uniformly distributed over $2^{nR_1} \times 2^{nR_2}$.

Definition 3.22: A rate pair (R_1, R_2) is said to be *achievable* for the BC channel if there exists a sequence of $[n,(2^{nR_1},2^{nR_2})]$ codes with $P_e^{(n)} \to 0$. The *capacity region* is the closure of the set of achievable rates.

Note that the event $E_1 = \{\phi_1(Y_1^n) \neq M_1\}$ and the event $E_2 = \{\phi_2(Y_2^n) \neq M_2\}$ imply the event $E = \{\phi_1(Y_1^n) \neq M_1 \text{ or } \phi_2(Y_2^n) \neq M_2\}$, hence $\Pr\{E_1\} \leq \Pr\{E\}$ and $\Pr\{E_2\} \leq \Pr\{E\}$. Also $\Pr\{E\} \leq \Pr\{E_1\} + \Pr\{E_2\}$ by the union bound. This implies $\Pr\{E\} \to 0 \Leftrightarrow \Pr\{E_1\} \to 0$ and $\Pr\{E_2\} \to 0$. Therefore, the capacity region depends only on the conditional distributions $p(y_1|x)$ and $p(y_2|x)$.

3.5.8.2 Gaussian Broadcast Channels

A model for the AWGN broadcast channel is shown in Figure 3.30. We assume that Tx with power P_{Tx} and two distant receivers, Rx_1 and Rx_2, one with Gaussian noise power N_1 and the other with Gaussian noise power N_2. We assume that $N_1 < N_2$. Thus receiver Rx_1 is less noisy than receiver Rx_2. We therefore have

$$Y_1 = X + Z_1 \qquad\qquad Y_2 = X + Z_2$$

where Z_1 and Z_2 are arbitrarily correlated Gaussian random variables with variances N_1 and N_2, respectively. Tx wishes to send independent messages at rates R_1 and R_2 to receivers Rx_1 and Rx_2, respectively.

We find that the capacity region of the Gaussian BC is

$$R_1 \le C\left(\frac{\alpha P_{Tx}}{N_1}\right) \tag{3.65}$$

$$R_2 \le C\frac{(1-\alpha)P_{Tx}}{\alpha P_{Tx} + N_2} \tag{3.66}$$

where α may be arbitrarily chosen $(0 \le \alpha \le 1)$ to trade off rate R_1 for rate R_2 as Tx wishes. P_{Tx} here denotes the average power of the signal at the transmitter.

To encode the messages, Tx generates two codebooks, one with power αP_{Tx} at rate R_1, and another codebook with power $(1-\alpha)P_{Tx}$ at rate R_2, where R_1 and R_2 lie in the capacity region specified in (3.65) and (3.66). Then to send an index $i \in \{1, 2, \ldots, 2^{nR_1}\}$ and $j \in \{1, 2, \ldots, 2^{nR_2}\}$ to Rx_1 and Rx_2, respectively, Tx takes the codeword $\vec{X}(i)$ from the first codebook and codeword $\vec{X}(j)$ from the second codebook and computes the sum. This sum is then sent over the channel.

The receivers must now decode their messages. First consider the bad receiver Rx_2. He merely looks through the second codebook to find the closest codeword to the received vector $\vec{Y}_2$. His effective SNR is $\gamma_2 = (1-\alpha)P_{Tx}/(\alpha P_{Tx} + N_2)$, since Rx_1's message acts as noise to Rx_2.

The good receiver Rx_1 first decodes Rx_2's codeword, which it can accomplish because of its lower noise N_1. It subtracts this codeword $\hat{\vec{X}}_2$ from $\vec{Y}_1$. It then looks for the codeword in the first codebook closest to $\vec{Y}_1 - \hat{\vec{X}}_2$. The resulting probability of error can be made as low as desired.

3.5.8.3 Physically Degraded BC

> **Definition 3.23:** A BC channel is said to be *physically degraded* if $p(y_1, y_2|x) = p(y_1|x)\,p(y_2|y_1)$.

The motivation for defining the degraded BC is depicted in Figure 3.31. The broadcast BC shown in Figure 3.31(a) can be converted into the cascaded (degraded) broadcast channel shown in Figure 3.31(b). This transformation

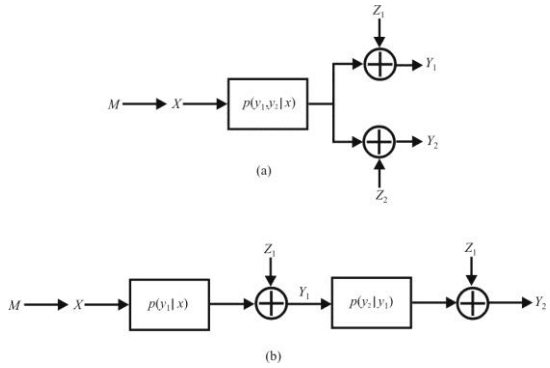


Figure 3.31 (a, b) BC to degraded channel transformation.

requires Y_1 to be less noisy that Y_3. Tx wishes to send independent messages at rates R_1 and R_2 to receivers Rx_1 and Rx_2, respectively.

Property 3.13: The capacity region for sending independent information over the degraded BC $X \to Y_1 \to Y_2$ is the convex hull of the closure of all (R_1, R_2) satisfying

$$\begin{aligned} R_2 &\le \mathbb{I}(U;Y_2) \\ R_1 &\le \mathbb{I}(X;Y_1|U) \end{aligned} \tag{3.67}$$

for some joint distribution $p(u)p(x|u)p(y,z|x)$, where the auxiliary random variable U has cardinality bounded by $|U| \le \min\{|\mathcal{X}|, |\mathcal{Y}_1|, |\mathcal{Y}_2|\}$.

Example 3.9: The Gaussian Channel

As an application of Property 3.6 consider the physically degraded Gaussian BC channel

$$\begin{aligned} Y_1 &= X + Z_1 \\ Y_2 &= X + Z_2 = Y_1 + Z_2' \end{aligned}$$

where $Z_1 \sim \mathcal{N}(0, N_1)$ and $Z_2' \sim \mathcal{N}(0, N_2 - N_1)$ with a power constraint P_{Tx}. The capacity region for the channel is given by

$$R_1 \le C\left(\frac{\alpha P_{\text{Tx}}}{N_1}\right)$$
$$R_2 \le C\frac{(1-\alpha)P_{\text{Tx}}}{\alpha P_{\text{Tx}} + N_2} \tag{3.68}$$

where $0 \le \alpha \le 1$ and $C(x) = 0.5\log_2(1+x)$.

Rx_2 with power $(1 - \alpha)P_{\text{Tx}}$ decodes in the presence of his ambient noise N_2 and also the "corruption" in X due to part of the power being used to communicate to Rx_1. However, Rx_1 can decode the message intended for Rx_2 and hence only has to combat a noise power of N_1.

3.5.8.4 Characteristic of Broadcast Channels

A broadcast channel is characterized by having one source, and $k \ge 1$ receivers. The goal is to send information with negligible probability of error to all receivers. More formally, we want to find the set of simultaneously achievable rates (R_1, R_2, . . . R_k), or the capacity region. The theory of the broadcast channel resides in a larger taxonomy with the appellation *network information theory* [15], which examines all forms of networks. Push-to-talk networks are the primary concern here as that is the typical tactical C2 communication method. Tactical C2 countermeasures are one of the most likely applications of EW on the battlefield.

Minimax Schemes

Suppose the transmission channels to the receivers have respective channel capacities C_1, C_2, . . . , C_k bits per second. One simple approach could be to send at rate $C_{\min} = \min(C_1, C_2, \ldots, C_k)$. However, even this is only possible when the channels are "compatible" (uncorrelated, or orthogonal). The transmission rate is limited by the worst channel. At the other extreme, we may try to send at rate $R = C_{\max}$, with resulting rates $R_i = 0$ for all but the best channel (say, the kth one), and $R_k = C_{\max}$ for the best channel. Neither of these schemes looks optimal.

Time-Sharing Approaches

Time-sharing the resources is another possible approach. This is where portions of time $\lambda_1,\lambda_2, \ldots, \lambda_k$, $\lambda_i \ge 0$, $\sum \lambda_i = 1$, are allocated to sending at rates C_1, C_2, . . . , C_k . Assuming compatibility of the channels and assuming $C_1 \le C_2 \le \ldots \le C_k$, we find that the rate of transmission through the ith channel is

$$R_i = \sum_{j \le i} \lambda_j C_j, \qquad i = 1, 2, \ldots, k \tag{3.69}$$

Example 3.10: Capacity Region (or Rate Trade-off) Diagrams
Rate trade-off examples for $1 \rightarrow 2$ broadcast channels, with $\mathcal{X} = \{1, 2, 3, 4\}$, $\mathcal{Y}_1 = \{1, 2\}$, $\mathcal{Y}_2 = \{1, 2\}$.

Orthogonal channels:

$$\mathbf{W}_1 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \qquad \mathbf{W}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Completely incompatible channels:

$$\mathbf{W}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \qquad \mathbf{W}_2 = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The switch-and-talk channel. Analogy with a speaker fluent in Spanish and English who must speak simultaneously to two listeners, one of whom understands only Spanish and the other only English.

Superimposing Information

Can we do better than time-sharing for the switch-and-talk channel? This is accomplished with "superimposing" the information. We make use of the fact that the English receiver does not understand Spanish and vice versa but both realize when the sender does not broadcast in their language, to send extra information, common to both parties. If channel 1 is used in proportion α of the time, then αC_1 bits/transmission are received by Rx_1 and $(1 - \alpha)C_2$ by Rx_3. However, $H(\alpha)$ additional common bits/transmission may be transmitted, by choosing the ordering in which the channels are used (constrained by α).

In other words, modulation of the switch-to-talk button, subject to the time-proportion constraint α, allows the perfect transmission of $2^{nH(\alpha)}$ additional messages to both Y_1 and Y_3. Thus, all rates (R_1, R_2) of the form

$$(R_1, R_2) = [\alpha C_1 + H(\alpha), (1-\alpha)C_2 + H(\alpha)]$$

can be achieved by choosing the subset of n transmissions devoted to the use of channel 1 in one of the $\binom{n}{\alpha n} \approx 2^{nH(\alpha)}$ possible ways.

3.5.8.5 Marton's Lower Bound on the Achievable Region

The Deterministic Broadcast Channel

Property 3.14: The capacity region of the deterministic memoryless BC with $y_1 = f_1(x)$, $y_2 = f_2(x)$, is given by the convex closure of the union of the rate pairs satisfying

$$\begin{aligned} R_1 &\le H(Y_1) \\ R_2 &\le H(Y_2) \\ R_1 + R_2 &\le H(Y_1, Y_2) \end{aligned} \tag{3.70}$$

Marton's Theorem

Property 3.15: The rates (R_1, R_2) are achievable for the BC channel $[\mathcal{X}, p(y_1, y_2 | x), \mathcal{Y}_1 \times \mathcal{Y}_2]$ if

$$\begin{aligned} R_1 &\le I(U;Y_1) \\ R_2 &\le I(V;Y_2) \\ R_1 + R_2 &\le I(U;Y_1) + I(V;Y_2) - I(U,V) \end{aligned} \tag{3.71}$$

for some $p(u, v, x)$ on $\mathcal{U} \times \mathcal{V} \times \mathcal{X}$.

3.5.9 Channel Models—General Diagram

The general diagram for channel models is illustrated in Figure 3.32. The input alphabet is $X = \{x_1, x_2, \ldots, x_n\}$ and the output alphabet is $Y = \{y_1, y_2, \ldots, y_m\}$. The

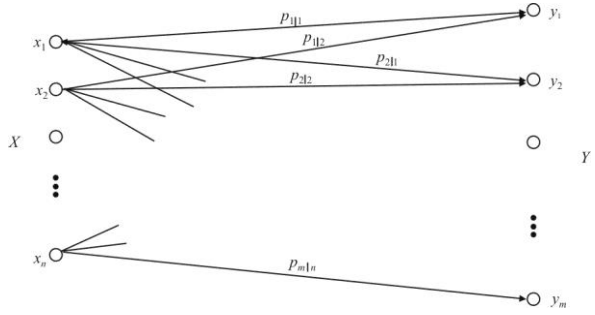


Figure 3.32 Channel models general diagram.

transition probabilities are given by $p_{j|i} = \Pr_{Y|X}\{y_j \,|x_i\}$. In general, computing the capacity needs more information. The statistical behavior of the channel is completely defined by the channel transition probabilities.

The mutual information $I(X;Y)$ is a convex function (denoted as $\cap$) in the input probabilities so finding a maximum is usually simple.

3.6 Concluding Remarks

We introduced the fundamental tenents of information theory in this chapter. We began with a discussion of probability and Shannon's theory of information, including entropy and mutual information measures. Random variables were discussed and some of their properties were investigated.

The amount of information contained in a message or produced by a process is measured by the change in entropy the process produced. The entropy is computed or measured prior to execution of the process. The process is then performed and the entropy subsequently computed or measured. The difference between these two entropy values is the amount of information produced.

In this the chapter we investigated some of the salient characteristics of common communication channel models. We first investigated what a channel is, and then examined its channel capacity. Coding was introduced as a way of increasing the channel capacity.

The common channel models that we discussed included memoryless channels and binary channels, including the all-important binary symmetric channel. The erasure channel was introduced, which, as pointed out, is a common model used for CDMA communications channels. The Gilbert-Elliott channel, a common model for channels experiencing burst errors, was discussed. This channel is a time-varying channel and not as much is known on how to use these channels as there is for the other channels discussed. Finally, the broadcast

channel was discussed that we will use substantially in later chapters to analyze the performance of EW systems targeted against communication systems. Using these models puts the theory of EW systems on a firm mathematical and technical basis.

References

[1] Shannon C. E., "A Mathematical Theory of Communication," *The Bell System Technical Journal*, Vol. 27, July, October 1948, pp. 379–423 and 623–656.

[2] Burr, A., *Modulation and Coding for Wireless Communications*, Pearson Education, 2001.

[3] Haykin, S., *Communication Systems*, 4th Ed., New York: Wiley, 2001.

[4] Schwartz, M., *Information Transmission, Modulation, and Noise*, 4th Ed., McGraw-Hill, 1990.

[5] Goh, J. G., and S. V. Maric, "The Capacities of Frequency-Hopped Code-Division Multiple-Access Channels," *IEEE Transactions on Information Theory*, Vol. 44, No. 3, May 1998, pp. 1204–1211.

[6] Geraniotis, E. A., and M. B. Pursley, "Error Probabilities for Slow Frequency-Hopped Spread-Spectrum Multiple-Access Communications over Fading Channels," *IEEE Transactions on Communications*, Vol. COM-30, No. 5, May 1983, pp. 996–1009.

[7] Proakis, J. G., *Digital Communications*, 2nd Ed. New York: McGraw- Hill, 1989.

[8] Cohen, A. R., J. A. Heller, and A. J. Viterbi, "A New Coding Technique for Asynchronous Multiple Access Communication," *IEEE Transactions on Communication Technology*, Vol. COM-19, No. 5, October. 1971, pp. 849–855.

[9] Gallager, R. G., *Information Theory and Reliable Communications*, New York: Wiley, 1968, p. 80.

[10] Mushkin, M., and I. Bare-David, "Capacity and Coding for the Gilbert-Elliot Channels," *IEEE Transactions on Information Theory*, Vol. 35, No. 6, November 1989, pp. 1277–1290.

[11] Gilbert, E. N., "Capacity of Burst-Noise Channels," *Bell System Technical Journal*, Vol. 39, September 1960, pp. 1253–1265.

[12] Elliott, E. O., "Estimates of Error Rates for Codes on Burst-Noise Channels," *Bell System Technical Journal*, Vol. 42, September 1963, pp. 1977–1997.

[13] Cover, T. M., "Comments on Broadcast Channels," *IEEE Transactions on Information Theory*, Vol. IT-44, October 1998, pp. 2524–2530.

[14] Cover, T. M., "Broadcast Channels," *IEEE Transactions on Information Theory*, Vol. IT-18, January 1972, pp. 2–13.

[15] Cover, T. M., and J. A. Thomas, *Elements of Information Theory*, New York: Wiley, 1991, Chapter 14.

Appendix 3A: Weak Law of Large Numbers

Assume we have a binary sequence where Pr{0} = 1 – Pr{1} = 1 – p. Let t denote the number of 1's in the sequence. Then as $n \to \infty$, and with ε > 0, the weak law of large numbers says that

$$\Pr\{|t/n - p| > \varepsilon\} \to 0$$

or

$$\frac{t}{n} - p = 0$$

That is, we expect with high probability that there are pn 1's.

As a result of this we have the following:

- $n(p-\varepsilon) < t < n(p+\varepsilon)$ with high probability
- $\sum_{n(p-\varepsilon)}^{m(p+\varepsilon)} \binom{n}{t} \approx 2n\varepsilon \binom{n}{pn} \approx 2n\varepsilon 2^{nh(p)}$
- $\lim_{n\to\infty} \frac{1}{n} \log_2 2n\varepsilon \binom{n}{pn} \to h(p)$

where the binary entropy $h(p)$ is given by

$$h(p) = -p\log_2 p - (1-p)\log_2(1-p)$$

This equation is sketched in Figure 3A.1. Note that $h(p) = h(1-p)$.

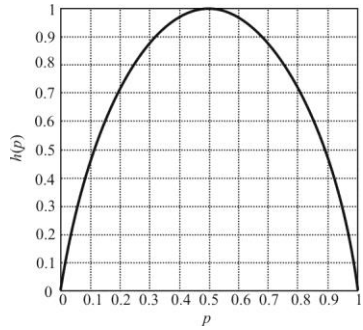


Figure 3A.1 Binary entropy $h(p)$.

Chapter 4

A Model of Information Warfare[1]

4.1 Introduction

For many years IW existed in the repertoire of the IO cognoscenti without a fundamental mathematical theory to support it. Recently, Borden and Kopp related four canonical IW strategies to Shannon's information theory, to provide a mathematically quantifiable theoretical basis for the discipline [1, 2]. The limitation of this theoretical model is that it can model the effects of IW actions upon an information carrying channel, but provides little insight into how those actions might affect the outcome of an engagement between adversaries. This chapter will review the existing model, based upon Shannon's theory, and further extend it through the application of *hypergames*, thus providing a more powerful technique for explaining and modeling the system level effects of IW actions.

This chapter is structured as follows. First, we provide a definition of IW. Then we present four canonical IW strategies and discuss some of their properties. We then introduce hypergames and discuss how they can be used as a model of IW, including how the four canonical strategies of IW can be mapped onto hypergames. This section contains a discussion of how game theory can be applied to overcome many of the limitations imposed by Shannon's model. We conclude the chapter with an appendix that discusses the basic properties of Turing machines.

4.2 Defining Information Warfare

One U.S. DoD definition of IW is as follows:

[1] Substantial portions of this chapter are directly cited from a series of research publications by Dr. Carlo Kopp at Monash University, and are used with permission.

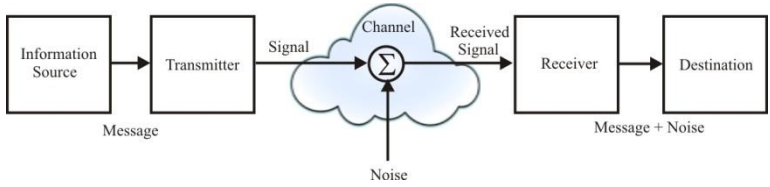


Figure 4.1 Shannon noisy channel model.

> Information Warfare is any action to deny, exploit, corrupt, or destroy the enemy's information and its functions; protecting ourselves against those actions, and exploiting our own military information functions.

In this definition, IW is defined as "actions" which yield intended outcomes of "denial," "exploitation," "corruption," and "destruction" of an opponent's "information." The definition unfortunately does not provide a quantifiable basis or measure of information. Borden [2] and Kopp [1] argued that Shannon [3] provided a model to address this limitation. Shannon's channel capacity model relates useful channel capacity to bandwidth and the ratio of available signal to AWGN.

Shannon's model is defined in terms of a communication channel, with a source, destination, transmitter (Tx), receiver (Rx), and a noise source that impairs the channel capacity, otherwise bounded by bandwidth and signal. The model that Shannon used is shown in Figure 4.1. Shannon derived the model from first principles, showing that the capacity, C (bps), of a channel to transmit information in a signal of power S across a channel of a given bandwidth (W, Hz) in the presence of noise of power N is upper-bounded by

$$C = W \log_2\left(1 + \frac{S}{N}\right) \tag{4.1}$$

where S and N are power in compatible units (not decibels).

Borden and Kopp showed that Shannon's model can be easily mapped onto four canonical offensive IW strategies:

1. Denial of information;
2. Disruption and destruction;
3. Deception and mimicry;
4. Subversion.

We discuss this mapping in detail in this chapter and provide an extension: exploitation.

4.2.1 Limitations of the Shannon Model

The Shannon model, described by (4.1), provides a powerful tool for capturing the interactions between adversaries and the information carrying channel. The Shannon model, however, does not capture how the manipulation of the channel might be reflected in the behavior of the adversaries. It does not model the effects that may flow from manipulating the channel. In fact, Shannon went to considerable effort to preclude the necessity of knowing what information is contained within the message and therefore what possible actions might ensue based on this information [1]. Weaver's interpretation of Shannon's original paper divides the subject of communications into three levels [4], as mentioned in Chapter 1. IW is concerned with all three of these levels; Shannon's theory, however, deals with only Level A. Hypergames, investigated later in this chapter, deals primarily with Level C.

4.3 Information Warfare Strategies

IW/IO is any organized use or manipulation of information or knowledge that is intended to produce an advantage in a contest with an opponent. Whether the use or manipulation is applied against the thought processes of an opponent or the software and hardware comprising an opponent's information systems is a matter of application [1];

> The fundamental paradigm of IW/IO appears to be a basic evolutionary adaptation resulting from competition in the survival game. Whether it is the game of chemical deception played by a micro-organism against an immune system, or the use of camouflage and deception by prey and predator alike in every tier of the natural world, or whether it is some part of the complex structures we use to describe the modern IW/IO paradigm, the fundamental paradigm is essentially one and the same.

IW/IO applications can be divided into four simple categories described as [5–9]:

1. Denial of information (DoI)/passive denial. Denial can be passive or active. Passive denial includes concealment and camouflage, or stealth. DoI makes the signal sufficiently noise-like that a receiver cannot discern its presence from that of the noise in the channel.
2. Disruption and destruction (D&D)/active denial is the insertion of information which produces a dysfunction inside the opponent's system, or alternately, the outright destruction of the system. Jamming (EA) fits in this category.

3. Deception and mimicry (D&M) is the insertion of intentionally misleading information. In a successful D&M attack the known signal is mimicked so well that a receiver cannot distinguish the phony signal from the real signal.
4. Subversion (SUB) is the insertion of information that triggers a self-destructive process in the target system. At the simplest level, SUB amounts to the diversion of the thread of execution within a Turing machine, which maps on to the functional behavior of the victim system, that is, flipping specific bits on the tape to alter the behavior of the victim Turing machine. (See Appendix 4A for a description of Turing machines.)

We can note that these strategies are neither mutually exclusive, nor confined to either side of the predator/prey or offensive/defensive game. In principle either player can use any or all. It is worth noting that ample examples of these strategies naturally occur in nature [10].

4.3.1 Four Canonical IW Strategies

Whereas the IW model we are discussing here applies to much more than the communication channel problem, it has not been proven that (4.1) applies to other than that problem when the added noise is Gaussian. Shannon's model has been used to describe many situations beyond that originally envisioned by Shannon. Shannon himself issued words of caution about applying (4.1) to other than that for which it was originally derived [3]. In these cases, (4.1) is assumed to apply and the analysis proceeds from there. This is the case for the Kopp/Borden model described here. For such a model, indeed, for any model, its value depends on the useful results it can produce, not if it were derived from first principles as Shannon's theory was.

When used as a model of IW, the "message" and "transmitter" components of Shannon's model, as well as the "receiver" and "destination" components, have broader meanings than those in the original theory. Consider for instance the laser designators used to guide bombs or missiles toward ground targets. A laser is used to illuminate the target and the reflected energy is detected in laser energy receivers in the nose of the missiles. The transmitter in this case is the target itself that reradiates the laser energy. The receiver is the laser designation receiver in the missile. The message is embedded in the laser signal from the laser designator.

IW in actuality amounts to manipulation of a channel carrying information in order to achieve a specific goal [1, 2]. Borden describes this effect as the "battle for bandwidth"—a contest over the available capacity in an information bearing channel.

The specific implementation of an attack measure depends on the means being used to perform the attack. The specific implementation of a protect measure

depends on the specific attack measure being used. For example, beamforming and adaptive nulling are protect measures against a jammer.

To establish a fundamental theoretical model, the starting point must be fundamental information theory, which is centered in Shannon's channel capacity theorem, expressed as (4.1).

If an attacker intends to manipulate the flow of information to an advantage, the game will revolve around controlling the capacity of the channel, given by *C*. To achieve this, the attacker must manipulate the remaining variables in the equation, the bandwidth *W*, the signal power *S*, and/or the noise power *N*. Three of the four canonical strategies involve direct manipulation of bandwidth, signal power, and/or noise power.

Shannon's communication channel model is given by Figure 4.1. It consists of an information source that generates a message that is to be transferred to a destination. That message is sent to the transmitter that puts out a signal into the channel. AWGN is added to that signal by one source or many noise sources in the channel. The sum of the signal and noise is received by the receiver at the destination. The message, mixed with the noise, is subsequently transferred from the receiver to the destination.

Viewed in the context of Shannon's theorem, IW is the battle for channel bandwidth and therefore channel capacity [2].

4.3.1.1 First Canonical Form—Passive Denial

The first canonical form is denial of information/degradation or destruction, that is, concealment and camouflage, or stealth.

To deny means to deny completely by a direct attack on the means of accomplishing the denial. The use of a high-energy laser to blind or destroy an electro-optic sensor is an example of denial by direct attack. Another example is a virus that destroys operating systems in a computer used to do SA. To degrade means to reduce the ability of the targeted nodes to transfer information. An example of this is to jam a portion of the bandwidth being used in a communication channel.

The degradation strategy involves manipulation of the *S*/*N* term in (4.1). The flow of information between the source and destination is impaired or even stopped by burying the signal in noise causing usable $S \to 0$ and subsequently driving $C \to 0$.

There are two forms of this strategy, the first being the "camouflage/stealth" or "passive" form, and the second being the "jamming" or "active" form. The first of these strategies makes up the first canonical form.

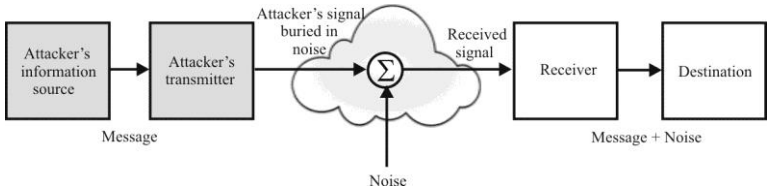


Figure 4.2 First canonical form: passive denial.

The first form involves forcing $S \to 0$ to force $C \to 0$. In effect the signal is made so faint that it cannot be distinguished from the noise floor of the receiver. Such a condition is depicted in Figure 4.2.

Examples of Degradation via Passive Denial

- Passive form—biological or military camouflage patterns. For example, a chameleon changing its colors to match that of its surroundings.
- Passive form—the stealth fighter that uses shape and absorbers to hide from radar, and a cooled jet exhaust to hide from infrared equipment, and camouflaged tents to avoid detection from the air.
- Passive form—encryption and concealment to prevent unwanted parties from reading or finding what they ought not to.
- Passive form—the use of encryption and concealment to prevent unwanted parties from gleaning the contents of transmissions.

4.3.1.2 Second Canonical Form—Active Denial

The second denial category involves the insertion of information that produces a dysfunction inside the opponent's system (see Figure 4.3). Jamming/EA fit in this category since noise jamming yields to the fact that the receiver knows the jammer is there; however, enough noise is introduced into the receiver that demodulation of the signal is not possible or at least degraded. Doing so increases *N* in (4.1). By

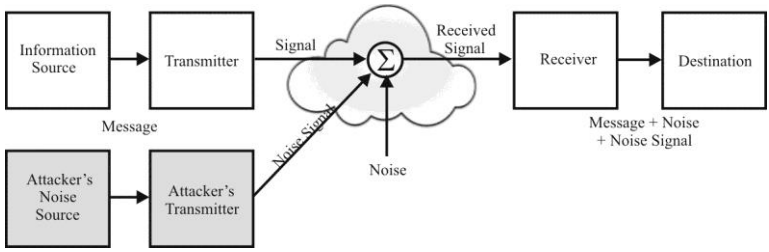


Figure 4.3 Second canonical form, active denial; DOI degradation strategy.

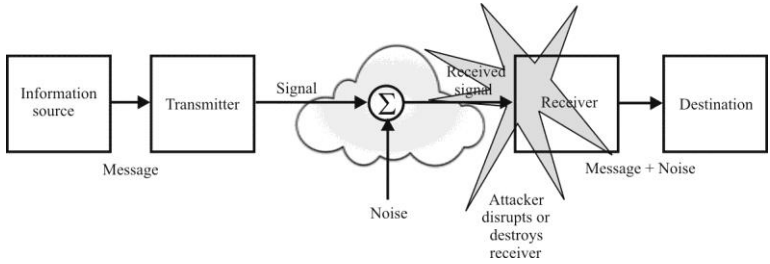


Figure 4.4 Second canonical form—destruction; D&D denial strategy.

jamming a portion of the spectrum and yielding it unusable by the communication system also decreases W in (4.1).

Data can be degraded either by delaying it until its usefulness is reduced or by destroying it in full or part. For example, the use of concealment is an attack measure (degradation) against the collection task. The use of jamming to reduce the capacity of a communications channel (thereby delaying transmission) is another example.

Active denial involves the injection of an interfering signal into the channel, to make $N >> S$ [causing $S/N \to 0$, and $\log_2(1)=0$], and thus force $C \to 0$. In effect, the interfering signal drowns out the real signal flowing across the channel. In actuality, the requirement for $N >> S$ is not really required for jamming digital communication. $S/N \sim -6$ dB ($N/S \sim 1/4$) or so is sufficient to produce a *bit error rate* (BER) of 10^{-1}, a condition that precludes successful communication in most cases [11]. This condition is depicted in Figure 4.3. The equivalent conditions for analog FM denial occur at $S/N \sim +6$ dB or lower and $S/N \sim -5$ dB or lower for analog AM [12].

Destruction also fits into the active denial category. Clearly, destroying the receiver eliminates the possibility of using any of the frequency spectrum for exchanging information, causing $C \to 0$ when $W \to 0$. See Figure 4.4.

The distinction between using EA in form 1 and EA in form 2 is similar to *low probability of detection* (LPD), LPI, and *low probability of exploitation* (LPE). For LPD the goal is for the receiver to not know the signal is there while LPI is yielding to the possibility of detection of the signal, but intercept is made difficult. Lastly, LPE yields to the possibility of detection and intercept of the signal, but exploitation is difficult (perhaps by encryption). For each of these stages, implementation is easier from the last to the first (LPD is the most difficult to achieve, while LPE is the easiest).

There is an important distinction between the passive and active forms of degradation strategy as contained in forms 1 and 2. In the passive form of this attack, since the signal is submerged in noise and cannot be detected, this form is "covert" in the sense that no information is conveyed to the victim. In the active form of this attack, the signal which jams or interferes with the messages carried

by the channel will be detected by the victim. Therefore this form is "overt" in the sense that information is conveyed to the victim, telling the victim that an attack on the channel is taking place. Both forms are widely used in biological survival contests and in social conflicts.

Examples—Degradation Via Active Denial

- Active form—Barrage jamming wireless radio broadcasts or communication links.
- Active form—Smoke screens hides troops from enemy gunfire.
- Active form—Octopus squirts ink at a predator.
- Active form—The *improvised explosive device* (IED) jammer precludes the detonation of the bomb when triggered.
- Active form—Defensive jamming equipment on a fighter emits signals similar to radar returns from a hostile radar, but including an erroneous position measurement. The use of cardboard cutouts of tanks and artillery that appears to be a large force in an area.
- Destruction—Organisms spray noxious fluids on predators, thereby blinding and numbing the predators' visual and olfactory senses, temporarily or permanently.
- Destruction—Very high-power RF weapons can permanently or temporarily impair the function of victim receivers by overloading input circuits.
- Destruction—Destroy the receiver system by direct attack, for instance, by fire, bombing, or other such means.
- Destruction—In the IT domain, that is, any temporary or permanent denial of service attack, such as induced packet storms, cutting data or power cables, or using electromagnetic weapons.

4.3.1.3 Third Canonical Form—D&M Corruption Strategy (Mimicry)

Deception and mimicry/corruption is the insertion of intentionally misleading information. This also effectively reduces usable S in (4.1). To corrupt is to insert false data. For example, the use of dummies on the battlefield is an attack measure against the collection (observation in the OODA loop) function. Intrusion into a communications channel and spoofing is another example. Psyops is an example of corrupting information being stored in the human mind of the target.

The corruption strategy involves the substitution of a valid message in the channel with a deceptive message, created to mimic the appearance of a real message. The attack is successful if the defender believes that the corrupted signal is actually a valid signal. During a corruption attack, the attacker's corrupted information enters the defender's system through Observation and is then examined during Identification. It is at this point that the error occurs. The

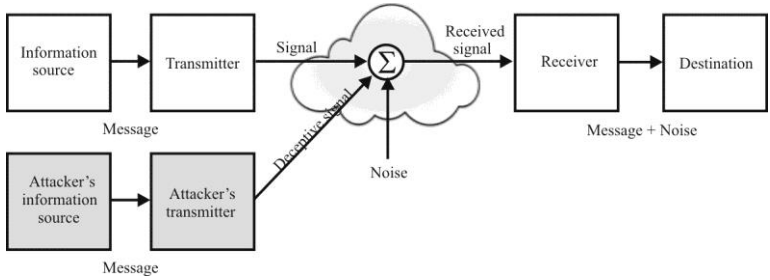


Figure 4.5 Third canonical form; D&M corruption strategy (mimicry).

corrupted information mimics a signal that the defender believes is genuine. The defender then misidentifies the corrupted information as the element it is mimicking. It is then used during the individual's Interpretation substep as though it were a valid signal.

In terms of (4.1), S_{actual} is replaced with S_{mimic}, while the W and N terms remain unimpaired. The victim receiver cannot then distinguish the deception from a real message, and accepts corrupted information as the intended information. Success requires that the deceptive message emulates the real message well enough to deceive the victim. Corruption is inherently "covert" since it fails in the event of detection by the victim receiver. Corruption is used almost as frequently as degradation in both biological and social conflicts. This strategy is depicted in Figure 4.5.

The degradation and corruption strategies both focus on the S and N terms in (4.1). The denial strategy manipulates the W term, by effecting an attack on the transmission link or receiver to deny the reception of any messages, by removing the means of providing bandwidth W. This means that $W \to 0$ or $W = 0$ if the attack is effective. The denial strategy is inherently "overt" in that the victim will know of the attack very quickly, as the channel or receiver is being attacked. A denial attack may be temporary or persistent in effect, depending on how the channel or receiver is attacked. Numerous biological and social examples exist as illustrated in the following:

Examples—Corruption

- Biological examples of organisms mimic the appearance of harmful, predatory, or toxic species to deceive predators.
- A Harris hawk family hunts together. One of the hawks intentionally exposes his presence to prey so that the others can attack from different directions.
- Biological predators mimic the appearance of prey organisms to attract lesser predators.

- Deception jamming techniques are used against radars, producing errors in angle and/or range measurements, or producing false (nonexistent) targets.
- Deceptive propaganda radio broadcasts or deceptive radio transmissions emulating real messages are used.
- Deceptive advertising is used in the commercial and political domains.
- Identity theft, phishing, phracking, hacker use of stolen user codes such as user names and passwords, and spammer e-mail address substitution are used.

Self-Deception

The phrase self-deception suggests an instance of deception targeted at the individual [13]. However, self-deception is more accurately described as an intentional misinterpretation with the intent of supporting a favored, but unrealistic belief [14]. Self-deception can therefore be viewed as a self-inflicted corruption attack that specifically targets an individual's information processing. Instances of self-deception typically occur when an individual is unable to change his or her environment to a desired state through actions and instead manipulates his or her world model to produce the illusion of the desired state.

One property of self-deception is that it may reduce cognitive dissonance [15]. Cognitive dissonance occurs when an individual holds beliefs that are incongruous [1, 16]; psychological discomfort is caused by holding dissonant beliefs. Individuals can use self-deception to reduce the dissonance between the beliefs and thereby reduce discomfort.

Construing self-deception as intentional misinterpretation suggests that it causes errors during the Interpretation substep of the Orientation step in the OODA loop (see Figure 2.3). In this case, the self-deceiving individual correctly gathers information during Observation and then correctly Identifies known objects, events and relationships. During the Interpretation substep, when the information is analyzed, it is found to be dissonant with the individual's existing knowledge and thus causes discomfort. The individual can reduce this uneasiness by misinterpreting the information in such a way that it is no longer at odds with existing knowledge. Once the individual has reduced the incompatibilities, the Aims can be derived and Options and Outcomes can be generated, albeit they are likely to be in error.

While misidentification could also reduce dissonance, it needs to be intentional to cause self-deception. This requires that the individual has some reason for the misidentification, which can only be determined by interpreting the new information. Therefore, new information cannot be found to be dissonant until it has been analyzed, so misidentification is therefore not responsible for self-deception.

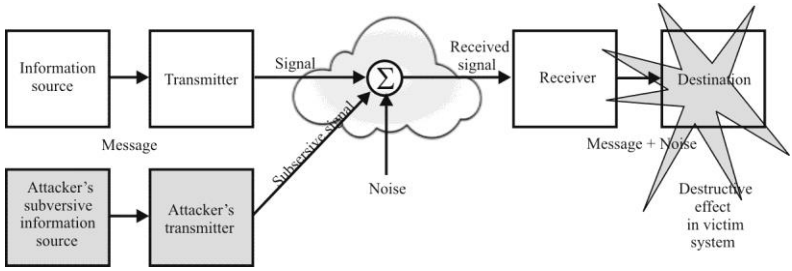


Figure 4.6 Fourth canonical form; sub denial strategy.

4.3.1.4 Fourth Canonical Form—Denial Via Subversion

The fourth canonical form is Subversion/Denial. It is the insertion of information which triggers a self-destructive process in the targeted system.

Denial via subversion differs from the first three forms in that it does not involve an attack on the message, its contents, or the channel/receiver. Subversive attacks involve the insertion of information at the receiver that triggers a self-destructive process in the victim system. At the most basic level, this is the diversion of the thread of execution within a Turing machine,[2] which maps onto the functional behavior of the victim system. It amounts to surreptitiously flipping one or more specific bits on the tape, to alter the behavior of the machine. The attack may impair or destroy the victim system. Numerous biological, social, and technological examples exist. See Figure 4.6.

Examples—Denial Via Subversion

- Parasites emit chemicals that alter the internal function of the victim organisms to favor the parasite, such as the production of favorable nutrients or weakening of immune defenses.
- The use of deceptive radio or optical signals triggers the premature initiation of weapon fuses, such as proximity fuses on guided missiles or artillery shells.
- Logic bombs, viruses, worms, and other destructive programs use system resources to damage the system itself.
- The use of deceptive signals which trigger the premature initiation of weapon fuses, such as proximity fuses on guided missiles and artillery shells.

[2] A Turing machine is an abstract model for a computation device. The universal Turing machine can compute all the computable functions. See Appendix 4A for a further description of Turing machines.

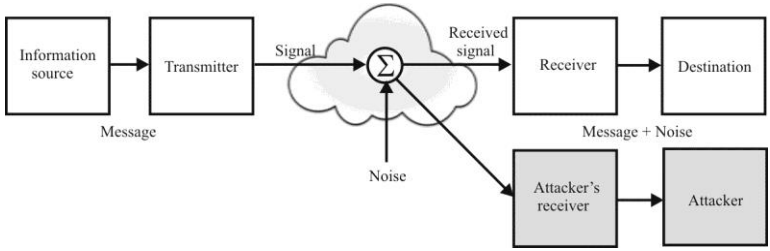


Figure 4.7 Exploitation model**.**

- Most examples of subversion rely on the attacker's use of corruption to penetrate the victim's defenses and create conditions to affect the subversive attack.

4.3.1.5 Exploitation

To exploit is to collect against the adversary's movement of data. This increases the data available for friendly SA and makes the generation of friendly information more efficient.

A model for exploitation is depicted in Figure 4.7. The U.S. DoD definitions of the four strategies of information attack include exploitation, which is eavesdropping on victim messages. Kopp argues that eavesdropping is a wholly passive activity which does not involve a direct attack on the victim channel, receiver or system, thus impairing or altering the function of the victim. Therefore, it cannot be a canonical strategy defining a mode or type of attack on a system.

The U.S. DoD model lumps destruction of the opponent's receiver function and destruction of the opponent's system through subversion into one category, while the U.S. DoD model includes passive exploitation as an active offensive measure. Exploitation is defined as gathering an adversary's flow of information to facilitate friendly information generation. Exploitation amounts to attaching a receiver in parallel with the target receiver. Since it does not in itself produce an immediate causal effect in the function of the target channel, it cannot be classified as an offensive IW strategy in the sense of the four defined strategies. Rather, it is an information-gathering technique, albeit one which may facilitate the application of an offensive IW strategy.

At the very least, exploitation is an enabler for the other four canonical forms. It is through exploitation that an attacker determines which of the other four forms will work best. But exploitation is more than that. We argue in Chapter 9 that, with broadcast channel models, messages can be intercepted to the point that the channel (privacy) capacity from Tx to Rx is reduced to zero, and no information can flow across the channel that is free from intercept. This can effectively reduce the utility of the channel to zero for the target network. This is precisely what the

first four canonical forms aspire to accomplish. Therefore, we will include exploitation in the model.

4.3.1.6 Unique Canonical Strategies

There are only three variables in (4.1), each accounting for one of the first three strategies. Therefore, manipulation of additional parameters in this model is not possible. In the fourth canonical form where the model incorporates the functioning of a Turing machine, information can be used to alter the functioning of the program (by flipping bits), but not the nature of the machine. Hence, there are no obvious candidates for further canonical strategies within the mathematical model. We include exploitation as a canonical form because it can be used to reduce the (privacy) capacity of the channel just as the other strategies reduce capacity.

4.3.1.7 Properties of the Four Canonical Strategies

There are three properties of the canonical strategies that are obvious from their definitions.

Orthogonality: Because each strategy attacks the victim system in different ways, a canonical strategy cannot be formed by combining any number of the other canonical strategies.

Indivisibility: Canonical strategies cannot be further divided or decomposed because each of them represents the simplest way to affect their respective modes of attack.

Concurrency: A victim system can be subjected to any number of concurrent attacks. When the attacks are alike, the effects on the victim system are additive. However, when the attacks are dissimilar, the effects on the victim system are orthogonal.

4.3.1.8 Compound and Chained Strategies

The four canonical strategies can be used in compound strategies where more than one of the basic strategies are employed [17]. An example of such a structure is shown in Figure 4.8. Each of the labeled circles represents one of the basic strategies and they can be employed in series or parallel. Figure 4.8(a) illustrates application of a single strategy, while Figure 4.8(b) shows employment of multiple strategies.

A straightforward example of a compound strategy is a tactical EW system that uses a remoted antenna. Perhaps the equipment is at the base of a hill in a

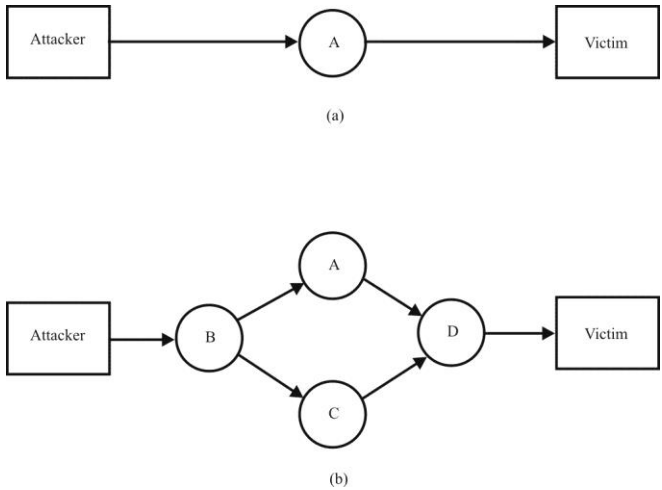


Figure 4.8 Compound strategies: (a) simple application of a single strategy and (b) simultaneous application of several strategies.

camouflaged shelter and the antenna is at the top of that hill. The camouflaged shelter is passive denial (type 1), while the remoted antenna is an example of deception (type 3), since an adversary attempting to geolocate the EW system can only locate the radiating element—the antenna.

Likewise, the strategies may be chained as illustrated in Figure 4.9. In this case there is an intermediate victim that is attacked, which subsequently attacks a second victim using one or more of the basic strategies.

4.3.2 Summary

The canonical strategies define all modes of attack involving information in terms of basic manipulation of fundamental models—the Shannon channel model and the Turing machine. All attacks on information processing or transmission systems comprise one or more of the canonical strategies. The canonical strategies are ubiquitous in the biological and social domains, with IW being one of the latter. The canonical strategies discussed in this section provide a mathematically useful and robust model for conflicts involving the use of information.

As with the preceding example of the laser designator, in the interests of generality the depicted models separate the attacker's message generation and transmitter functions from the channel proper. This reflects the reality that the topology of the physical system may not map directly on to the channel model. Strategy 1 could, for instance, involve a support jamming aircraft flying at a

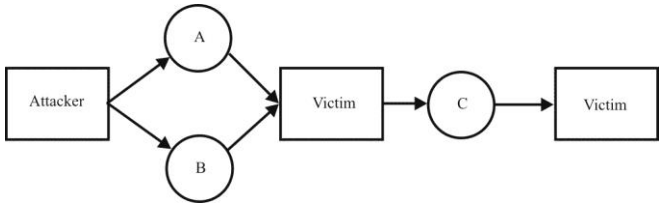


Figure 4.9 Chained strategy.

distance of several miles from the target aircraft it is protecting by jamming the victim radar.

A range of contemporary examples describing the four strategies are discussed in [1, 2]. There is no shortage of case studies in the domain of EW and particularly EA. Of interest is the fact that all basic EW and EA techniques predate Shannon's research by several years. Contemporary examples are discussed in [18–21]. Shannon's information theory provides a powerful model for describing the interaction between adversaries applying IW techniques and the information carrying channel itself. What the model cannot describe is how the manipulation of the channel may be reflected in the behavior of the respective adversaries. We cover next an extension to the model that focuses on that aspect.

4.4 Hypergames and IW

To address the most significant shortcoming of the Shannon model for IW, that is, discerning the impact on the decisions made by the target based on the attack used, a different approach is required. One possibility, and the one described here, is based on game theory, in particular, hypergames.

Hypergames are games in which the respective adversaries (players) may not be fully aware of the nature of the engagement in which they are participating, or indeed, that they are even participating in an engagement at all. Hypergames provide an alternate approach to modeling the fundamental paradigm of IW.

Considerations in a hypergame include [5]:

- The intent or goals of the other players are likely unknown or misunderstood by a player.
- The choices available to other players are likely to be unknown to a player.
- Who all the other players in the game may be is likely to be unknown to a player.
- There are differences in player knowledge and expertise.

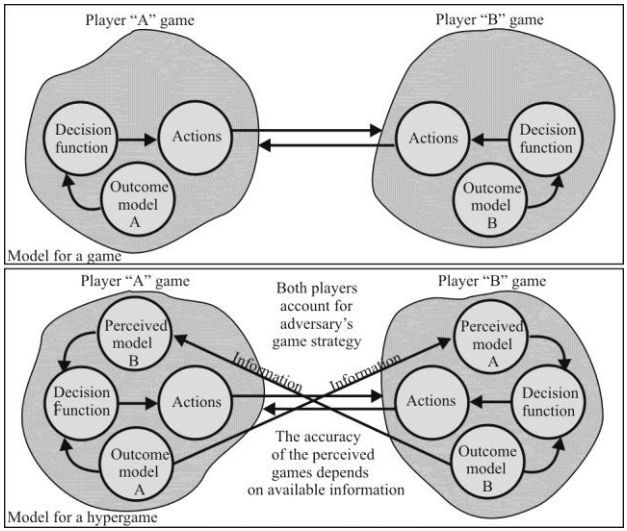


Figure 4.10 In a hypergame, the players perceive their opponent's game. How accurate that perception might be depends on the information available to the respective player. Inaccurate information leads to a misperception of the game state and may lead to actions that do not gain the player an advantage. (Source: [6], © C. Kopp, 2005. Reprinted with permission.)

- There are differences in player starting SA.
- There are differences in player ongoing assessment capability (evidence processing).
- There are differences in player understanding of plan projection (what beats what?).
- There are differences in player information (both at the commitment phase and during the operations).
- Player time constraints.
- There are differences in player creativity (what tricks can be added) such as feints, hidden reserves, denial, and deception operations.

Players in hypergames have perceptions of the engagement that may not reflect the true nature of the situation, resulting in decisions that may not be in the best or even good interests or intent of the players. See Figure 4.10. Players typically have *perfect information* about the state of the game in classical game theory, resulting in no misperceptions of previous moves. The perfect information assumption does not hold for a hypergame.

In terms of the OODA loop, a player's perception of a game is described by the Observation-Orientation phase of an OODA loop. A player's choices in a game are described by the Decision-Action phase of an OODA loop. Boyd's

OODA loop describes the basic dynamic in a game/hypergame. IW is a means to an end in a hypergame—it permits alteration of an opponent's perception of the game in a manner yielding an advantage to the player using it.

4.4.1 Hypergames

A general description of a hypergame is given in [22], in which *N* players each perceive a particular game:

$$G = \{G_1, G_2, \ldots, G_N\} \tag{4.2}$$

Each game perceived by the participating players can be described with a set of outcomes, as perceived by that player:

$$G_i = \{O_1, O_2, \ldots, O_M\} \tag{4.3}$$

Each outcome, in turn, comprises a set of possible actions (moves) by respective players, as perceived by player *i*:

$$O_i = \{\{A_1, A_2, \ldots, A_q\}_1, \{A_1, A_2, \ldots, A_p\}_2, \ldots, \{A_1, A_2, \ldots, A_r\}_N\} \tag{4.4}$$

Assuming that the players are rational, each player will seek to execute actions that yield a set of outcomes most favorable to that player. This model can be related to the well-established Boyd OODA model insofar as a player's perception of the game is the outcome of the Observation-Orientation phases of the loop, and the Decision-Action phases of the loop reflect the choices made by the player, based upon the player's perception of the game and what constitutes the best choice to make [23].

In the context of a hypergame, IW is a means to an end. Applying IW in a hypergame is to try to alter an opposing player's perception of the game such that an advantage ensues to the player who applies the means of IW.

A special case in the hypergame model is *strategic surprise*, where a player may be wholly unaware of another player's presence in the game, or may be unaware of action other players have the option of taking.

A hypergame can be specified by the following elements:

Players: They are the parties (individual agents, groups, coalitions) that may affect the multiagent situation that we want to study using the hypergame.

Strategies: Each player may see a number of combinations of actions available to himself or herself and to each of the other players. Notice that all players may not

recognize the same actions as being available for each given player since they do not perceive the same actions as relevant.

Preferences: For each player, the various strategies define a set of perceived outcomes. Usually the player prefers some outcomes to others and has some beliefs about other players' preferences.

> **Definition 4.1:** An *N-person hypergame* is a system consisting of the following:

1. A set P_N of N players,
2. For each $p,q \in P_N$, a nonempty finite set S_p^q that reflects the set of strategies for player p as perceived by player q.
3. For each $p,q \in P_N$, an ordering relationship O_p^q is defined over the product space $S_1^q,\ldots,S_M^q$ and represents the preference ordering of p's strategies as perceived by q.

Thus, S_p^q and O_p^q express q's perception of p's options and aims. The set $S_1^q,\ldots,S_M^q$ makes up q's strategy matrix and together with P_N and the ordering $O_1^q,\ldots,O_M^q$ reflect player q's game G^q within the hypergame G. Thus, a hypergame G can be considered as a set of N game, $G^1,\ldots,G^N$, one for each player. We assume that each player i makes her strategy choice with full knowledge of her own game G^i. Obviously, a player may realize that others may perceive the situation differently: if so, the player may have more or less an idea as to what games they are trying to play. Or the player may see only his or her own game, which he or she assumes to represent his or her perception shared by all.

To give an initial illustration, here is an example of a two-player hypergame, in game-normal form,[3] for which q perceives an option available to player p, an option (i.e., option c) that is not available for himself or herself:

Agent p's Game G^p

$S_p^p \setminus S_q^p$	α	β
a	1,4	2,3
b	4,1	3,2

Agent q's Game G^q

$S_p^q \setminus S_q^q$	α	β
a	1,4	2,3
b	4,1	3,2
c	3,2	5,0

[3] Game-normal form refers to a description of a game in matrix form. An extensive-form game is a game where players move sequentially rather than simultaneously.

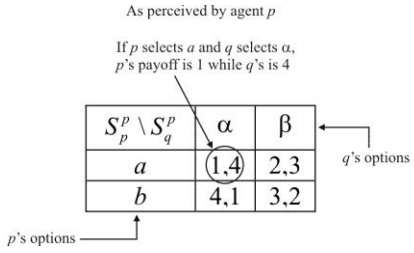


Figure 4.11 Hypergame normal mode.

The way to read these normal-mode matrices is explained in Figure 4.11.

Having defined our hypergame, the final step is, of course, to analyze it using general principles, and hence to draw some conclusions. One could hope to define a uniquely rational course of action for each agent-player. If used in a normative way, the hypergame approach would thus provide a very definite prescription for the decision-maker to follow; if used descriptively—under an assumption that agents will act rationally—it would give a prediction of the outcome to be expected.

In order to analyze a hypergame, we must introduce some set of decision rules for the players. Such rules are based on the notion of a "dominant" strategy as specified by classical authors of game theory (see, for instance, [24]).

> **Definition 4.2:** A strategy is called a *dominant strategy* for a player if choosing it leads to an outcome at least as highly preferred by that player as those obtained using any other of her strategies, whatever the strategy choice of the other player(s).

Note that according to this definition, it is theoretically possible for a player to have several such strategies. Starting from the dominance, we can introduce the following:

> **Definition 4.3:** We say one allocation of payoffs *Pareto-dominates* another, or is *Pareto-superior* to another, if all players are at least as well off in the first as in the second, and at least one is better off.

> **Definition 4.4:** We say an allocation is *Pareto-efficient* or *Pareto-optimal* if it is not dominated by any other allocation.

Now we can formulate some decision rules for a hypergame [25].

> **Rule 1:** If a player has a dominant strategy, then this player chooses that strategy.
>
> **Rule 2:** If a player perceives that another player has a dominant strategy, he or she chooses the most preferred outcome of those available when the other player uses his or her dominant strategy.
>
> **Rule 3:** In a nonconflict game, if a player perceives an outcome that is most preferred by every player, then he or she chooses that strategy enabling this outcome to be reached.

Notice that Rule 1 says if an agent has a dominant strategy, then he or she should use it; whereas Rule 2 says that if an agent believes that another has a dominant strategy, he or she assumes that he will use it and act accordingly. In the case where neither player has a dominant strategy and players nevertheless have a preferred outcome, the decision-maker can facilitate the preferred outcome in the non-conflict game case. This is what Rule 3 suggests.

> **Definition 4.5:** An outcome is *stable* for an individual player if it is not reasonable for him her to change the outcome by switching her strategy.

One criterion for the stability is the *Nash equilibrium*, which can be expressed by [24]

> An outcome of a game is a Nash equilibrium if no player has incentive to deviate from her strategy given that the other players do not deviate.

This equilibrium does not refer to other's preferences and consequently we can assume it is also valid in a hypergame.

Suppose that the two players p and q are two agents representing two companies, each desiring "not to be aggressive about the other (in the sense of market)" but suspicious of the other. We can have a hypergame model of this situation by assuming that each player has a choice between a cooperative (C) strategy and an aggressive (A) one. Player p, we suppose, places the four possible outcomes in the following order of decreasing preference:

(C, C) Coexistence;
(A, C) Attack without q retaliating;
(A, A) Mutual aggression;
(C, A) Attack by q without reply.

However, these preferences are not correctly perceived by q. In fact, q believes[4] p to have the following preference order:

(A, C) Attack without q retaliating;
(C, C) Coexistence;
(A, A) Mutual aggression;
(C, A) Attack by q without reply.

However, q has the same preferences as p and these preferences are also not correctly perceived by p which perceived them as q perceived those of p. This situation can be represented by the following two-person game:

Agent p's Game G^p

$S_p^p \setminus S_q^p$	C	A
C	4,3	1,4
A	3,1	2,2

Agent q's Game G^q

$S_p^q \setminus S_q^q$	C	A
C	3,4	1,3
A	4,1	2,2

Consider the situation from p's point of view looking at the game G^p. In this game, p does not have a dominant strategy and consequently, p can use rule 2 since q has a dominant strategy, which is A. In these conditions, p assumes that q will adopt this aggressive strategy and consequently is faced with outcomes (C, A) and (A, A). According to rule 2, it chooses to be aggressive also, that is, it chooses (A, A) which seems to be for p a Nash equilibrium. q reasons similarly on G^q.

In a classical game, we cannot see the players' differing perceptions and consequently we cannot understand exactly why players deviate from cooperation. In fact, if each player had not mistaken each other's preferences, both would converge on the cooperation option.

Now suppose that p and q want to verify their misperceptions by communicating or by consulting a mediator. It is clear here that mediation and communication are both important in the presence of suspicious perceptions. If players can communicate, or they have motive to lie, or they do not trust each other, a mediator may be able to help by suggesting a Pareto-efficient allocation. The players have no reason not to take this suggestion, and might use the mediator [24].

Now suppose that p and q communicate their actions. In the case where p trusts q and this latter does not reciprocate the feeling, the matrices reflecting p's perception and q's perception are the following:

[4] In hypergames, we are taking into consideration "high-order" beliefs; that is, players' perceptions of each other's perceptions of the situation.

Agent p's Game G^p

$S_p^p \setminus S_q^p$	C	A
C	4,4	1,3
A	3,1	2,2

Agent q's Game G^q

$S_p^q \setminus S_q^q$	C	A
C	3,4	1,3
A	4,1	2,2

Looking at the situation from p's point of view, we see that neither p nor q has a dominant strategy and as the game considered by p is a nonconflict game, this player applies rule 3 and chooses (C, C), which is a Nash equilibrium that dominates (A, A). Looking now at the situation from q's point of view now, we see that this player has not been convinced by p and consequently q maintains his or her misperception on p. q's reasoning is: p has a dominant strategy A and q must act on the assumption that p will adopt this strategy (according to rule 2). In this situation, q is faced with two choices (C, A) and (A, A). As q is rational, q will opt for (A, A). From an external point of view, p and q have opted for (C, A), that is, that p will cooperate and q will attack. This is a very bad choice for p.

Thus, communication between agents is very risky in the case where agents are motivated to lie, or they do not trust each other. In this specific case, it is better to consider a mediator who might suggest a Pareto-efficient allocation.

To achieve that, each agent communicates his or her "exact" preferences to the mediator since this latter is in charge to find the Pareto-efficient allocation for agents. As external observer, this mediator m sees the "exact" perceptions of p and q represented by the following matrix:

m's perception of p and q

$S_p^q \setminus S_q^p$	C	A
C	4,4	1,3
A	3,1	2,2

Now, p and q both trust m and their perceptions are the following:

Agent p's Game G^p

$S_p^p \setminus S_q^p$	C	A
C	4,4	1,3
A	3,1	2,2

Agent q's Game G^q

$S_p^q \setminus S_q^q$	C	A
A	4,4	1,3
C	3,1	2,2

Now each agent supposes she is in a cooperative game and applies rule 3 that leads her to the dominant strategy (C, C), a Pareto-efficient allocation which dominates (A, A).

We have assumed here that the "exact" perception was the perception of *p*. If conversely, the mediator has received from *p* and *q* as the "exact" perception the perception of *q*, that is, (*A*, *C*), (*C*, *C*), (*A*, *A*), and (*C*, *A*), then we obtain as final perceptions of *p* and *q*:

Agent *p*'s Game G^p

$S_p^p \setminus S_q^p$	*C*	*A*
C	3,3	1,4
A	4,1	2,2

Agent *q*'s Game G^q

$S_p^q \setminus S_q^q$	*C*	*A*
C	3,3	1,4
A	4,1	2,2

The game now is the famous Prisoner's Dilemma (PD) for which the dominant strategy equilibrium is (*A*, *A*), which is worse than the strategy (*C*, *C*). To force *p* and *q* to adopt the strategy (*C*, *C*), we add a new rule.

> **Rule 4:** If two players *x* and *y* agree to choose an outcome under the supervision of a mediator *m*, then as soon as one of them deviates from this outcome, *m* informs the other.

If our players *p* and *q* follow this rule, they adopt the dominant strategy forced equilibrium (*C*, *C*) [that Pareto-dominates (*A*, *A*)] since they know if one of them deviates from this "forced equilibrium," the other knows it (informed by *m*) and both switch to (*A*, *A*). Rule 4 reduces in fact the DP matrix to only two outcomes, (*C*, *C*) and (*A*, *A*), and the first one dominates the second one. In this case, choices of players *p* and *q* are facilitated.

Thus, the PD usually used to model many different situations does give a rationale for some behaviors. But without a hypergame representation, the essential element of the story—misunderstanding—is left out.

4.4.2 Gaining Advantage from Differences in Perception

Suppose a two-player hypergame for which *p* perceives two options *c* and γ that are not available for *q*. In *p*'s point of view, option *c* is an option for *p* and γ is an option for *q*:

Agent *p*'s Game G^p

$S_p^p \setminus S_q^p$	α	β	γ
a	1,3	2,3	2,3
b	4,1	3,2	3,2
c	3,2	6,0	2,3

Agent *q*'s Game *Gq*

$S_p^q \setminus S_q^q$	α	β
a	1,3	2,3
b	4,1	3,2

From *q*'s point of view, we see that *q* believes that *p* will play strategy *b* and he or she will play β in order to obtain the stable outcome (*b*, β). Player *p* is far

from this point of view since he or she perceives two additional strategies that *q* does not see. From his or her point of view, *q* has a dominant strategy which is γ and as he or she assumes that *q* is rational, he or she believes that *q* will opt for that strategy. Knowing that, *p* will opt for *b* so that he or she gains the best payoff. We are faced with two points of view; according to *p*, the stable outcome is (*b*, γ), whereas according to *q*, the stable outcome is (*b*, β).

Suppose now that *p* is curious and wants to know if *q* has the same perceptions or not. In this case, *p* could ask a third party who knows *p* and *q* for instance, and this third party informed *p* (the latter can do that, if *p* for instance shares the advantage gained) that *q* has a limited view and *p* does not view options *c* and γ. Knowing that, *p* might let *q* opt for β with the intention to choose *c* in order to obtain a more preferable outcome (*c*, β) rather than (*b*, β).

Notice that this case is similar to the case where *q* sees two options that *p* does not perceive and which can be represented by the following matrices:

Agent *p*'s Game G^p

$S_p^p \setminus S_q^p$	α	β
a	1,3	2,3
b	4,1	3,2

Agent *q*'s Game G^q

$S_p^q \setminus S_q^q$	α	β	γ
a	1,3	2,3	3,2
b	4,1	3,2	0,6
c	3,2	2,3	3,2

Notice that the reasoning is similar for the following cases:

1. *p* (or *q*) perceives one option *c* (or γ) for himself or herself but which is not available for *q* (or *p*).
2. *p* (or *q*) perceives one option γ (or *c*) for the other agent but which is not available for herself.

Suppose now that the points of view *p* and *q* are the following:

Agent *p*'s Game G^p

$S_p^p \setminus S_q^p$	α	β	γ
a	1,3	2,3	2,3
b	3,1	3,2	4,3

Agent *q*'s Game G^q

$S_p^q \setminus S_q^q$	α	β
a	1,3	2,3
b	4,1	3,2

In this case, *q*'s reasoning is the same as previously and *q* believes that stable outcome is (b, β). *p* believes that *q* has a dominant strategy that is γ, and consequently, *p* will opt for the outcome *b*. However, as *p* is uncertain about what *q* perceives as outcomes, *p* communicates with *q* in order to tell *q* the different options that *p* perceives: α, β, and γ. Once *q* is convinced, both agents perceive the same options and the same preferences and in this case, *p* and *q* opt for (*b*, γ).

4.4.3 Mapping the Canonical IW Strategies onto Hypergames

The mappings from the four canonical IW strategies onto the context of a simple two player hypergame are as follows:

1. *Denial of information* (*DoI*)*/degradation or destruction, passive denial* is when either the presence of a player, or the intent of a player, is to be concealed from another.

2. *Disruption and destruction* (*D&D*)*/denial active denial* attempts to prevent another player from perceiving the state of the game. Unlike DoI, D&D can show the intent and possibly identity of the player using it, and thus may convey this information to the victim player.

3. *Deception and mimicry* (*D&M*)*/corruption* attempts to change another player's perception of the game at hand.

4. *Subversion/denial* is a strategy where an action by a player attempts to alter the perception of the game state of a victim player in such a way that elicits a self-destructive action by the victim player.

We can see that the hypergame model is a very good fit to the fundamental paradigm of IW, insofar as the four canonical strategies map directly into models that are well represented by hypergames. Figure 4.12 depicts these mappings. Higher-level hypergames, in which a player's perception of an opponent's perceptions is incorporated into the model, are an important refinement.

It is important to note that game players observe the actions of their opponents. This is the same notion as exploitation mentioned above. It is by observation that success or failure is judged.

4.5 Concluding Remarks

In this chapter we examined the modeling of four canonical strategies in IW using Shannon's channel capacity model expressed by (4.1) and the Turing machine, both of which are well founded in theory and well described mathematically. Further, we analyzed the effects of exploitation with these models, and investigated using the model of a hypergame for IW. With these models, we can glean an understanding of the basic nature of IW, as well as provide a means for directly quantifying the effects of an IW action, and determining likely outcomes of an IW action, respectively. The combined use of these techniques offers a

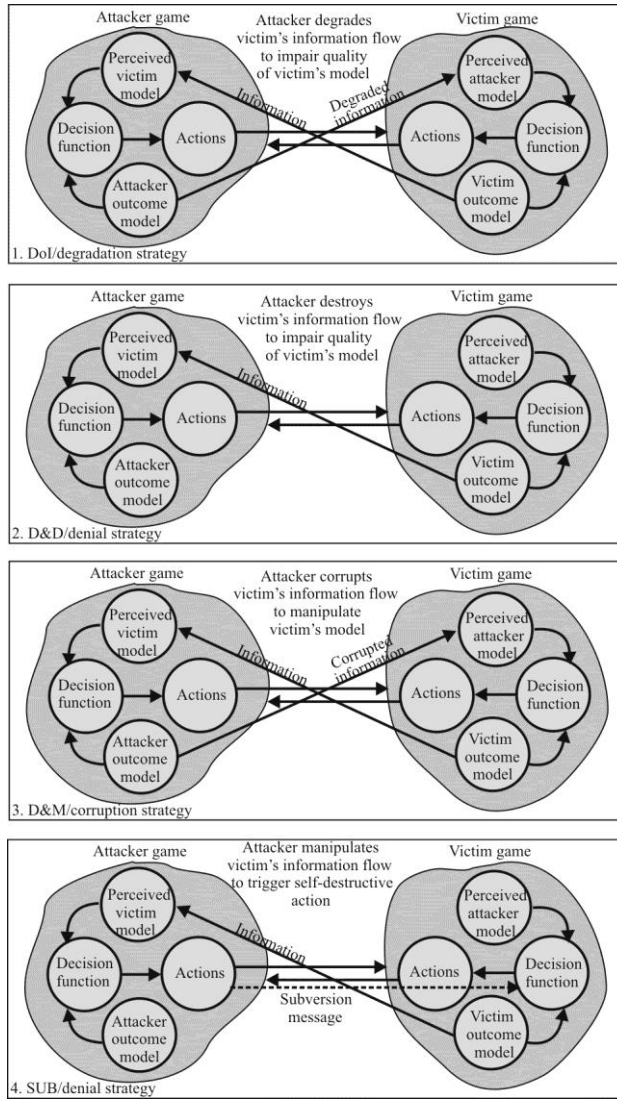


Figure 4.12 Hypergame models for the four canonical IW strategies. (Source: [5], © C. Kopp, 2003. Reprinted with permission.)

robust and flexible means of modeling the effects of an IW action along the whole chain comprised of the information carrying channel and the decision processes of

an opponent, thus permitting the modeling of substantial portions of an opponent's OODA loop.

References

[1] Kopp, C., "Information Warfare: A Fundamental Paradigm of Infowar," *Systems: Enterprise Computing Monthly*, Sydney: Auscom Publishing, February 2000, pp. 46–55, http://www.infowar.com/.

[2] Borden, A., "What Is Information Warfare?" *Air & Space Power Journal*—Chronicles Online Journal, November 1999, http://www.airpower.au.af.mil/airchronicles /cc/borden .html.

[3] Shannon, C. E., "A Mathematical Theory of Communication," *The Bell System Technical Journal*, Vol. 27, July, October 1948, pp. 379–423, 623–656.

[4] Weaver, W., "Recent Contributions to the Mathematical Theory of Communication," *The Mathematical Theory of Communication*, Urbana, IL: University of Illinois Press, 1963, pp. 3–28.

[5] Kopp, C., "Shannon, Hypergames, and Information Warfare," *Journal of Information Warfare*, Vol. 2, No. 2, 2003, pp. 108–118.

[6] Kopp, C., "Classical Deception Techniques and Perception Management vs. the Four Strategies of Information Warfare*," Proceedings of the 6th Australian Information Warfare & Security Conference*, (IWAR 2005), Deacon University School of Information Systems, Geelong, Victoria, November 2005, pp. 81–89.

[7] Kopp, C., "Considerations on Deception Techniques Used in Political and Product Marketing," *Proceedings of the 7th Australian Information Warfare & Security Conference*, (IWAR 2006), Edith Cowan University of Computer and Information Science, Perth, Western Australia, November 2006, pp. 62–71.

[8] Kopp, C., "The Four Strategies of Information Warfare and Their Applications," *IO Journal*, Vol. 1, Issue 4, Association of Old Crows, Alexandria, VA, February 2010, pp. 28–33.

[9] Kopp, C., *NCW101: An Introduction to Network Centric Warfare*, Melbourne: Air Power Australia, 2009.

[10] Kopp, C., and B. Mills, "Information Warfare and Evolution," *Proceedings of the 3rd Australian Information Warfare & Security Conference*, (IWAR 2002), Edith Cowan University, Perth, Western Australia, 2002, pp. 352–360.

[11] Poisel, R. A., *Modern Communication Jamming Principles and Techniques* 2nd Ed., Norwood, MA: Artech House, 2011, Ch. 8.

[12] Poisel, R. A., *Introduction to Communication Electronic Warfare Systems*, 2nd Ed., Norwood, MA: Artech House, 2008, Ch. 4.

[13] Demos, R., "Lying to Oneself," *The Journal of Philosophy*, Vol. 57, No. 18, 1960, pp. 588–594.

[14] Szabados, B., "Self-Deception," *Canadian Journal of Philosophy*, Vol. 4, No. 1, 1974, pp. 51–68.

[15] Ramachandran, V. S., "The Evolutionary Biology of Self-Deception, Laughter, Dreaming and Depression: Some Clues from Anosognosia," *Medical Hypotheses*, Vol. 47, 1996, pp. 347–364.

[16] Festinger, L., *A Theory of Cognitive Dissonance*, Stanford, CA: Stanford University Press, 1957.

[17] Kopp, C., "The Analysis of Compound Information Warfare Strategies," *Proceedings 6th Australian Information Warfare & Security Conference*, (IWAR 2005), Deakin University, School of Information Systems, Geelong, Victoria, November 2005, pp. 90–97.

[18] Fitts R. E., (ed.), *The Strategy of Electromagnetic Conflict*, Los Altos, CA: Peninsula Publishing, 1980.
[19] Schlesinger R. J., *Principles of Electronic Warfare*, Los Altos, CA: Peninsula Publishing, 1979.
[20] Knott, E. F., J. F. Schaeffer and M. T. Tuley, *Radar Cross Section*. Dedham, MA: Artech House, 1985.
[21] Ball, R. E., *The Fundamentals of Aircraft Combat Survivability Analysis and Design*, New York: American Institute of Aeronautics and Astronautics, Inc., 1985.
[22] Fraser N. M., and K. W. Hipel, *Conflict Analysis, Models and Resolution*, New York: Elsevier Science Publishing Co., 1984.
[23] Boyd, J. R., "A Discourse on Winning and Losing," a collection of unpublished briefings and essays, Maxwell AFB, AL: Air University Library, 1976–1992, http://www.belisarius.com/modern_business_strategy/boyd/essence/eowl_frameset.htm.
[24] Rasmussen, E., *Games and Information: An Introduction to Game Theory*, 2nd ed., Oxford: Basil Blackwell, 1989.
[25] Bennett, P. G., "Bidders and Dispenser: Manipulative Hypergames in a Multinational Context," *European Journal of Operations Research*, Vol. 4, 1980, pp. 293–306.
[26] Turing, A., 1948, "Intelligent Machinery," 1948. Reprinted in *Cybernetics: Key Papers*, Evans, C. R., and Robertson, A. D. J., (eds.), Baltimore, MD: University Park Press, 1968, p. 31.

Appendix 4A

4A.1 Turing Machines

A Turing machine is a theoretical device that is used to manipulate symbols contained on a strip of tape. It consists of an input/output tape (which consists of multiple cells, each of which contain the symbol "0" or "1"), the Turing machine (which consists of a read-write head, which scans the tape cells according to its current "state"), and a list of instructions or "transitions" (which can also be understood as the machine's program) that tell the machine which way to move (left or right) and according to which state to operate. The read-write head can also write a symbol into the cell under the head. An example of a simple Turing machine is shown in Figure 4A.1.

A succinct definition of the thought experiment was given by Turing in his 1948 essay, "Intelligent Machinery." Referring back to his 1936 publication, Turing writes that the Turing machine, here called a Logical Computing Machine, consisted of [26]:

> ...an infinite memory capacity obtained in the form of an infinite tape marked out into squares on each of which a symbol could be printed. At any moment there is one symbol in the machine; it is called the scanned symbol. The machine can alter the scanned symbol and its behavior is in part determined by that symbol, but the symbols on the tape elsewhere do not affect the behavior of the machine. However, the tape can be moved back and forth through the machine, this being one of the elementary operations of the machine. Any symbol on the tape may therefore eventually have an innings.[5]

Although they were intended to be technically feasible, Turing machines were not meant to be a practical computing technology, but a gedankenexperiment (German for thought experiment, popularized by Einstein early in the twentieth century) about the limits of mechanical computation. A gedankenexperiment is a construct for an experiment that would test an hypothesis or theory. Studying their abstract properties yields many insights into computer science and complexity theory.

4A.1.1 The Turing Machine as a Model for Computing

The Turing machine was described by Alan Turing in 1937. Despite its simplicity, a Turing machine can be adapted to simulate the logic of any computer algorithm, and is particularly useful in explaining the functions of a CPU inside of a computer. Turing machines are not intended as a practical computing technology,

[5] *Innings* is a term borrowed from cricket. It refers to being in a position of power, or the object of the current focus.

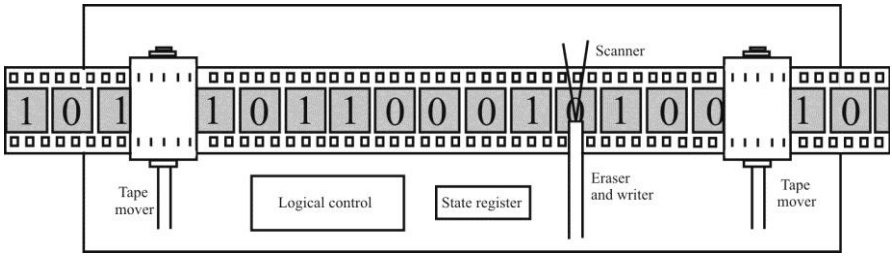


Figure 4A.1 Turing machine.

but rather as a thought experiment representing a computing machine. They help computer scientists understand the limits of mechanical computation.

In the 1930s (before the advent of the digital computer), several mathematicians began to think about what it means to be able to compute a function. Alonzo Church and Alan Turing independently arrived at equivalent conclusions. As we might phrase their common definition now:

A function is computable[6] *if and only if it can be computed by a Turing machine.*

A Turing machine is a very simple machine, but, logically speaking, has all the power of any digital computer. It may be described as follows: A Turing machine processes an infinite tape. This tape is divided into squares, any square of which may contain a symbol from a finite alphabet, with the restriction that there can be only finitely many nonblank squares on the tape. At any time, the Turing machine has a read/write head positioned at some square on the tape. Furthermore, at any time, the Turing machine is in any one of a finite number of internal states. The Turing machine is further specified by a set of instructions of the following form:

(*current_state*, *current_symbol*, *new_state*, *new_symbol*, *left/right/no movement*)

This instruction means that if the Turing machine is now in *current_state*, and the symbol under the read/write head is *current_symbol*, change its internal state to *new_state*, replace the symbol on the tape at its current position by *new_symbol*, and move the read/write head one square in the given direction (*left* or *right*). If a Turing machine is in a condition for which it has no instruction, it halts.

[6] Simplistically, a *computable function* is any function for which there is an algorithm that computes it. The particular algorithm for computing a computable function may not be known, however. A function can be shown to be computable, and therefore computable by a Turing machine, by specifying an algorithm to compute it.

It is natural to think of the set of instructions as a program for the Turing machine.

There are several conventions commonly used in Turing machines (and several slightly different, though of course logically equivalent, definitions of them). We adopt the convention that numbers are represented in unary notation, that is, that the non-negative integer n is represented by a string of $n+1$ successive 1s. Furthermore, if we want to compute a function $f(n_1,n_2, \ldots ,n_k)$, we assume that initially the tape consists of n_1, n_2, ... , n_k, properly encoded, with each separated from the previous one by a single blank, and with the tape head initially poised over the leftmost bit of the first argument, and the state of the Turing machine some initial specified value. We say that the Turing machine has computed $m = f(n_1,n_2,\ldots,n_k)$ if, when the machine halts, the tape consists of n_1, n_2, ... , n_k, m, properly encoded, and separated by single blanks, and the read/write head back at the leftmost bit of the first argument.

For example, suppose we wish to create a Turing machine to compute the function

$$m = multiply(n_1, n_2) = n_1 \times n_2$$

Suppose the input tape reads

_<1>1 1 1 _ 1 1 1 1 1 _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

which encodes 3 and 4, respectively, in unary notation. (Here the position of the read/write head is marked with < >.) Then the Turing machine should halt with its tape reading

_<1>1 1 1 _ 1 1 1 1 1 _ 1 1 1 1 1 1 1 1 1 1 1 1 1 _ _ _ _ _

which encodes 3, 4, and 12 in unary notation.

The very simplicity of a Turing machine makes it a challenge to program one to perform a specific computation.

Very precise (in a mathematical sense) descriptions are available for Turing machines. Proof of their operations has long since been available. Because of this, and the fact that the SUB canonical form can be modeled as a Turing machine, the canonical models are described very precisely as well.

Chapter 5

Electronic Warfare Systems and Network-Centric Warfare[1]

5.1 Introduction

Networking of EW systems is a particularly important issue when discussing the virtues of thin and thick EW configurations (defined subsequently). The detailed networking discussion is deferred to Chapter 6.

This chapter presents some of the important characteristics of NCW and how EW systems integrate with the other systems on the network.

This chapter is structured as follows. We begin with a description and discussion of some of the characteristics of NCW. That is followed by a high level synopsis of EW systems. More in-depth information on EW systems is presented in Chapter 8. We include operational considerations here. Next we investigate how EW systems integrated with the notions of *effects-based operations* (EBO). Then considerations of collaboration are investigated. The end of the chapter is devoted to discussions of data and information fusion and how it is used for the development of combat information and intelligence.

5.2 Network-Centric Warfare

Network-centric warfare as a topic has generated much debate. Each area within the military domain is trying to work out how it should alter and change to meet the challenges posed by operating in a fully networked environment. Attempts are being made to understand not only the change to equipment, but also to develop the new ways in which the warfighter will need to operate.

[1] The contributions of Derek Elsaesser, DREO, Canada, for this chapter are gratefully acknowledged.

This chapter concentrates just on *communications electronic warfare* (CEW) sensors and, to a lesser extent, jammers. It should not be considered to be definitive or proscriptive; rather it is a discourse on work in progress. This is just one view of what is a complex problem. Alternate views may well be as equally viable as those expressed here.

5.2.1 Concept of Network-Centric Warfare

The central idea of networking military forces and systems together to enhance their warfighting capability remains unchallenged. Reduction in the fog of war,[2] if it can be achieved, will be beneficial. To know where the enemy is, to be able to track him and understand his intent, when improved, will benefit the warfighter. The CEW sensor provides a view of the battlefield from the intercepted communications. A mobile enemy will inevitably use RF communications, which can be intercepted and geolocated. The individual sensors can cover a wide area and combining three or four sensors together into a baseline allow the positions of many of the transmissions to be determined. This would be from just one system, but what if a number of ES systems could be deployed and interlinked? The vision then takes us further to linking disparate sensors together, achieving a multiplier effect.

The area over which a commander has an interest is too big for just one ground-based system to survey. Airborne systems are seen as a means to fill in the gaps and to extend the range. They also help in providing coverage in difficult terrain. Sensors mounted on a ship also have an additional source of different information, particularly when deployed in a littoral environment. Information is derived from the collected ES data and put into reports. Traditionally, it is these reports that are exchanged. Often this has been a slow and, perhaps, haphazard and unreliable process. NCW seeks to change this by networking the sensors so that not only is there much better exchange at the information level, but there is also data exchange between systems providing mutual support and increasing the value of the information provided. The effort does not stop there. It continues its evolutionary path with other sensor's information output being fused into a COP.

From this it is possible to see that NCW is not just showing that two sensors can be linked and exchange data in a meaningful way. This is a first step, yes, but NCW requires systems to be able to integrate together and appear as one. This will be at all levels, from the basic input data that has been acquired about a signal, through to the output information in a form that will make it easy for commanders to assimilate. On a dynamic battlefield, movement will be constant, so that the systems need to be able to combine and split apart transparently. This demands much more from the systems.

[2] The fog of war is the confusion that always exists on a battlefield.

5.2.2 Definition of NCW

This section presents some commonly used definitions of network-centric warfare (NCW) and discussion of its potential benefits to military operations involving CEW.

Alberts et al. define NCW as [1]:

> Network Centric Warfare is the best term developed to date to describe the way we will organize and fight in the information age. … We define NCW as an information superiority-enabled concept of operations that generates increased combat power by networking sensors, decision makers, and shooters to achieve shared awareness, increased speed of command, higher tempo of operations, greater lethality, increased survivability, and a degree of self-synchronization. In essence, NCW translates information superiority into combat power by effectively linking knowledgeable entities in the battlespace.

5.2.3 Dissenting Views

It should be noted that not everyone subscribes to the virtues of NCW. In a series of three papers, Reid and Giffin [2] espoused an alternative due to the belief that the theoretical underpinnings surrounding NCW are flawed. The fundamental point is that NCW is based on a business model, developed in the 1990s, called the New Economy Theory, and an epistemology that has been proven wrong and therefore cannot possibly be the correct path for developing new ways for armies to equip themselves and fight[3] [3].

The New Economic Theory business model is based on two fundamental laws and one physical system development: Moore's law, Metcalf's law, and the Internet. Discussions about the Internet need not be included here as the reader is assumed to be familiar enough with that for our purposes. Some technical details about the Internet are presented in Chapter 6.

Moore's law dictates that semiconductor technology has advanced and will continue to advance at a pace such that every 18 months, capability approximately doubles and costs are cut in half. Moore's law has been applicable since the 1970s and is projected to continue for the foreseeable future, but at least until 2015. At that time, semiconductor technology will have progressed to the point that the science behind semiconductors will no longer apply, and the quantum mechanical aspects of atoms will have to be taken into account. Moore's law is true, at least as of the time of this writing.

Metcalf's law purports that the value of a computer network is proportional to the square of the number of nodes on the network, specifically,

[3] Alberts subsequently addressed the views espoused in [2]. See [3] for more details.

$$\text{Network Value} = \frac{1}{2}n(n-1) \tag{5.1}$$

where n is the number of network members.

An individual's epistemology is composed of the knowledge (those things known to be true) and beliefs (those things that are not necessarily true, but the individual believes them to be) that individuals possess as well as the process through which that knowledge and belief are obtained. The epistemology behind NCW, as delineated by Alberts et al. [1], is based on a philosophical construct called *naive intuitivism*. In this view, people learn knowledge in four steps:

1. Objective observation of the facts regarding a phenomenon under investigation;
2. Inductive generalization to produce a universal theory of the phenomenon;
3. Continued empirical justification of the theory, whereby it eventually attains the status of a law;
4. Deductive prediction of future events.

In more basic terms, these steps equate to:

1. Collect data;
2. Generalize based on this data;
3. Continue generalization of the old data and beliefs until the beliefs are accepted as fundamental truths (laws);
4. Use the laws to predict what will happen in life's situations.

Reid and Giffin develop the argument that NCW, as the new way for armies to fight, is based on rational military thinking, which, in turn, is based on naive inductivism and its tenets. They also put forth, based on a theory developed by Popper, that this epistemology has been proven incorrect and that it is not the way humans learn [4–7].

They put forth the proposition that the way people learn is by a system known by the appellation critical *rationalism*. In this epistemology, the characteristics of learning are:

1. Generalization—The question of the validity of inferring universal statements from singular statements.
2. Elimination of psychologism—Removing reliance on intuition as the explanation of knowledge development, by delineating between the creative conception of a new idea and the examination of its internal structure and consequences.

3. Deductive testing of theories—Retrieving the consequences of a theory by logical deduction, and comparing those consequences with experimental or observational evidence.
4. The problem of demarcation—The way in which empirical statements are distinguished from those that are not empirical in nature.
5. Falisfiability—The logical property of a theory by which it may in principle be refuted by singular statements. Falsifiability is the criterion of demarcation between empirical and non-empirical theories.
6. Falsification—The methodological decision to admit singular statements in apparent breach of a theory, thereby falsifying the theory. The community in which our investigation is conducted defines guidelines for when we may consider a theory to be falsified.
7. Objectivity—The objectivity of theories lies in their testability in a reproducible way. In principle, the theory can be tested by anyone.

At the risk of oversimplification, this worldview can be stated that the way humans generate knowledge and beliefs is to generate theories with corresponding hypotheses that can potentially be proven wrong (falsifiability) and then to try to prove the hypotheses incorrect (falsification). The theories that survive this process are eventually accepted as truths, or as close to truths as we can get.

The distinction between these two epistemologies is manifest in the virtues of NCW and the devotion of one's total resources to develop and employ a large network as a way to fight, or to adopt another view that not all nodes in a battlespace must have access to all the data. The data needed by a particular node is that data, and only that data, necessary to disprove a theory.

The dissenters believe that Metcalf's law breaks down at some value of *n*, and adding more net members actually can decrease the value of adding those. The cited reason is that the information search time increases as *n* increases, and, while no proof is provided, users of the Internet can attest to the frustration of searching many nodes and not finding the information desired.

5.3 Thick and Thin Sensors

The NCW paradigm creates a new possibility in thinking about how to execute EW. Traditional EW sensors have considerable capability inherent on the sensor platform. All of the necessary functionality to do the EW mission is contained on the same platform, save for the few that required more than one system to execute, such as position fixing. This type of sensor could be termed *thick*, in analogy with the thick and thin networking terminology.

The new idea is to create *thin* sensors. Such sensors would have limited capability but could be deployed in far greater number. They could be deployed on many platforms such as tanks and *armored personnel carriers* (APCs). The RF

sensing that they provide would be limited to energy detection, perhaps creating LOPs, and perhaps extracting other external information about the intercepted signal. This information could be fused with other sensed information at the platform level to create better information products for the platform. The information could also be sent, via the sensor network, to fusion centers where it is combined with similar data collected from other platforms, to include intelligence collectors and higher echelon fusion centers. This information would or could then be sent back to the platform where the initial data came from for better situation awareness.

Deployment of RF sensors in this way extends the area of coverage of the EW system as a whole because the additional platforms cover a much larger physical area than the EW systems alone can, albeit the information collected is less. Such sensors would be required to operate within the performance envelope of the host platform and would provide ranges commensurate with other sensors likewise incorporated. Thus, the RF sensor would collect signals to a nominal range of approximately 5 km from the host platform. Such a range would not require an elevated antenna, which would normally be incorporated on an EW sensor platform.

The loading on the network due to the distributed sensors would be dependent on the RF environment encountered by the conglomerate host platforms. Due to the limited range of the sensors, it would be expected to be minimal. Some sort of mechanism for sorting friendly transmissions would be required so that detections of such would not necessarily be reported unless that functionality is desired.

Although the exact architectural configuration needs to be determined by further analysis and experimentation, it is envisioned that the thick sensors would form a peer-to-peer architecture. Due to the limited functionality at the thin sensor, that architecture would probably be in a client/server arrangement.

The minimal information required from thin sensors would be:

- Signal presence indication;
- *Time of arrival* (TOA);
- Amplitude;
- Frequency;
- Confidence;
- Platform self-location.

The TOA should be as precise as possible, and if derived from *global positioning system* (GPS) timing, it should be accurate enough to provide PFs via *time difference of arrival* (TDOA) algorithms when combined with sensed information from other sensors. This list of information about sensed RF transmissions is referred to as external data. Additional possible information sensed could consist of

- Modulation type;
- Signal duration; and
- Data for correlation with sensed RF signals from other sensors.

In all cases, however, the security classification of such information must be consistent with the environment of the host platform.

5.4 EW Contributions

In this section we examine what contributions EW capabilities bring to networked forces.

5.4.1 EW Contribution to Situation Assessment

Situation assessment (SA) will be examined in detail in Chapter 7. Suffice it to say at this point that EW, and particularly ES, contributes to SA by:

- Contributing to updates to the EOB;
- Locating emitters;
- Potentially determining the intent of an adversary.

5.4.2 EW Contribution to Targeting

5.4.2.1 ES for EA Targeting

Communications EW addresses timeliness, age, and currency, all of which deal with temporal characteristics of tactical information. Jamming a target passing C2 or targeting information precludes that information from reaching the intended receiver, at least temporarily. Depending upon the criticality of the information, this could have a major impact on the outcome of an encounter. Thus, of the attributes in Fewell and Hazen's list discussed in Chapter 2, jamming is applicable to 23% of them.

Clearly, ES is used to establish targets for these EA functions. At the beginning of an encounter, it would reasonably be expected that the spectrum locations of critical targets as well as other information on the targets are largely provided to the EW systems from communications *intelligence preparation of the battlefield* (IPB) and the EOB maintained by the *Defense Intelligence Agency* (DIA) (the EOB generation and maintenance process is described shortly). Once the adversary is engaged and EA is applied to these targets, it is not uncommon for them to move in frequency as well as possibly physical locations. It is the function

of the ES capability to keep track of where these targets move. This is provided both by spectrum search as well as target geolocation.

5.4.2.2 Kinetic Weapons (Artillery and Rocket) Targeting

Just as ES can be used for EA targeting, keeping track of movement in both frequency and location, the ES sensor can be used for targeting kinetic weapons. As with EA, it would be expected that many, if not all, of the targets for kinetic weapons would be established ahead of an encounter. However, these targets can and do move after the commencement of hostilities. The ES sensor can participate in the friendly sensor suite trying to track the targets as long as they communicate.

5.4.3 Electronic Support

The purpose of ES is to obtain information from the intercept of communication signals. A *concept of operation* (CONOP) for tactical ES is shown in Figure 5.1[4] [8]. Targets are sensed by the ES sensors in the *reconnaissance, surveillance, and target acquisition* (RSTA) systems as well as the RF sensor field consisting of sensors that can detect RF energy and measure parameters associated with that energy. Such parameters might consist of LOB, TOA, signal type, modulation type, and so forth. Information on the signals detected by the RF sensors are sent to the RSTA systems via the network for further analysis. This information might consist of some of the raw parameters just listed, geolocations, gisting,[5] SA, EOB determination, and so forth.

The EW (ES) process is illustrated in the architecture shown in Figure 5.2 [9]. ASC refers to *all source correlation*, combining multiple information sources together. Blue force tracking is knowing where friendly forces are located. The EW sensors, whether unattended or attended, intercept signals based on the steerage/tasking from the EW target analysis. The EW target analysis generates emitter information based on steerage from the EW intelligence analysis process. Updates to the EOB, also known as the EW situation assessment, result from the EW intelligence analysis. This analysis is based on the *priority intelligence requirements* (PIR) that have been assigned to the EW discipline to answer. The EOB comprises the ES single source input to the all source process, the latter of which combines all the intelligence disciplines together. The all source analysis (correlation) is tasked by and is responsive to the PIR based on the information needs of the commander.

[4] We use the U.S. DoD C4ISR nomenclature for architectural views in this chapter. SV stands for system view and OV stands for operational view. For more information on these views, see [8, Chapter 11].

[5] A gist is a summary of an intercepted transmission.

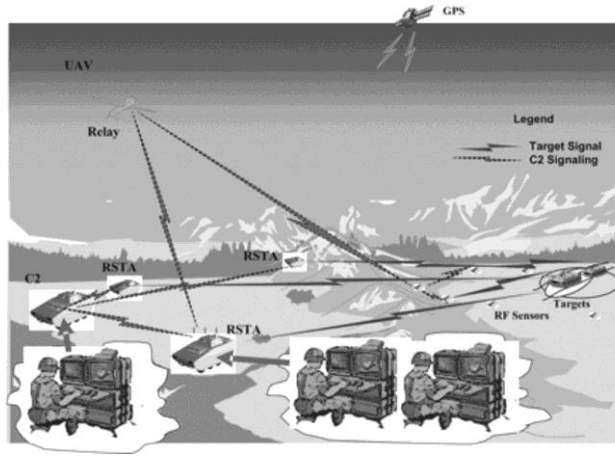


Figure 5.1 Forward electronic support CONOP (OV–1).

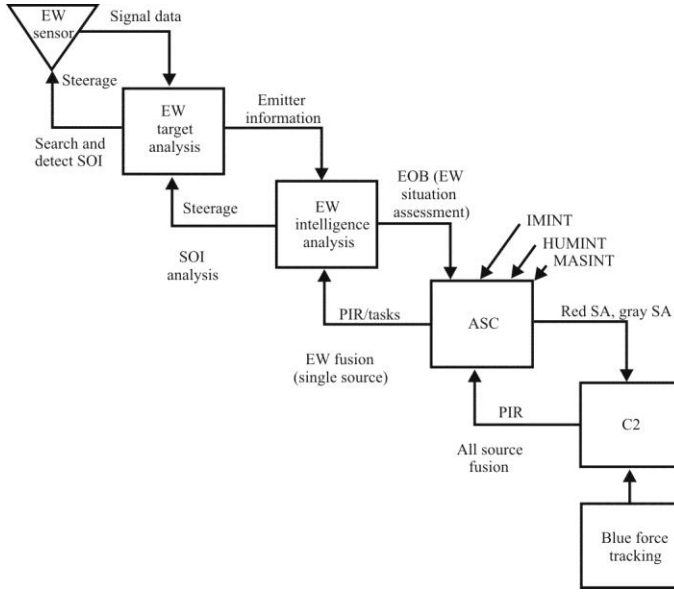


Figure 5.2 The ES process (OV–5). (Source: [9], © British Crown, 2003. Reprinted with permission.)

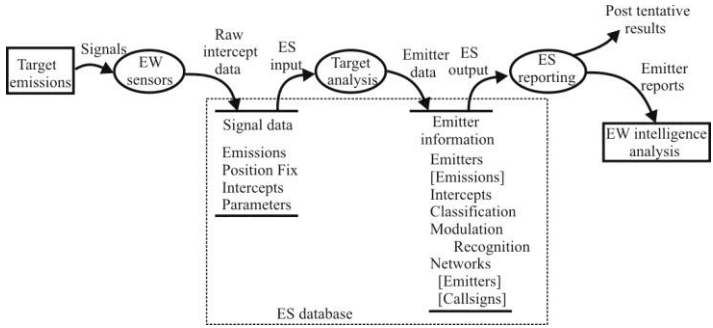


Figure 5.3 EW target analysis process flow. (Source: [10], © British Crown, 2001. Reprinted with permission.)

5.4.4 EW Target Analysis

Signals are intercepted by the suite of sensors, whether they are unattended or attended. The raw intercept data that results, as illustrated in Figure 5.3 [10], consists of information about the signal. Such data may include but is not limited to:

- Data about the emissions
  - Identification of the emitter
  - Frequency and RF bandwidth
  - Times that the signal came on and when it turned off
- Data about the emitter location
  - LOBs or LOPs
  - PF
- Data about the intercepts

This information forms the *signal database*, which is part of the ES database. The other part of the ES database consists of the *emitter information database* which results from analysis of the target. This database may include but is not limited to:

- Emitter information
  - Emitter location
  - Emitter intercepts
- Emitter classification
- Modulation classification
- Network information
  - Constituent emitters
  - Call signs

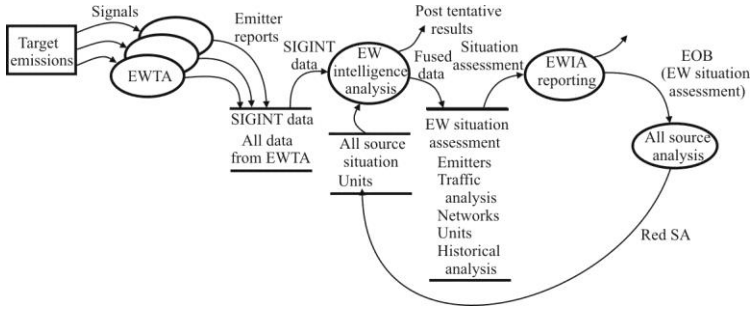


Figure 5.4 EW intelligence analysis process flow (SV–4). (Source: [10], © British Crown, 2001. Reprinted with permission.)

The ES analysis output comes from this database, and is reported. It is at this stage that ES information can reliably be posted to the GIG. Such postings can then be used for EW intelligence analysis.

5.4.5 EW Intelligence Analysis

The ES process is pictured in Figure 5.2. The commander generates information requirements. The G3/S3 converts these information requirements into PIRs and sends these to the all source fusion center. If a PIR can be satisfied with information already in the all source database, then the information is provided immediately. If it is not, then tasking is prepared for the EW systems (and/or other intelligence systems and/or higher echelon assets). This generates steerage information for the EW sensors and analysis systems. As targets are detected by the EW sensors, the signal data is provided to the target analysis cells that generate emitter information. This data is provided to single source analysis systems which generate the EOB information that is the EW situation assessment. This information, satisfying PIRs, is provided to the all source analysis element, which generates the red, gray, and brown situation assessment provided to the commander and G3/S5.

The architecture for EW (ES) intelligence analysis is shown in Figure 5.4 [10]. The signals are collected by ES sensors and processed by the *EW target analysis* (EWTA). This process generates emitter reports, which are forwarded over the network to the EW intelligence analysis node where they are posted for use by all with authority to access them. These new reports are combined with the all source database to generate the EW picture and fused data. The EW situation assessment is reported to the all source intelligence center, where it is combined with other single source intelligence estimates.

The signal analysis function follows the architecture shown in Figure 5.3. The EW sensors intercept the signals and the raw intercepts form part of the ES database. If there is an operator involved, then those intercepts can be listened to and information extracted. That information is put into report form and posted as well as forwarded to the EW intelligence analysis function shown in Figure 5.4. If there is not operator involvement, as might be the case for remoted RF sensors, then whatever automated processing is available is applied to the intercepted signal, and that information is forwarded to the EW intelligence analysis process.

It should be noted that usually it is unwise to post most raw information that has been intercepted by an ES system but not examined by an analyst. The exception to this might be signal technical external information such as frequency and modulation type, which is likely to be of little use outside of the ES processing chain anyway. The geolocations of emitting targets also comprise information that can be posted without analysis. Posting other than such external information can lead to erroneous conclusions about what the information means without putting that information into context, which is what the ES operators and EW analysts do.

5.4.6 Communications EW Contribution

This section explains the potential major contributions of CEW sensors and systems to the concepts involved with NCW. CEW sensors gather information from the frequency spectrum and the specific information depends on the scenario at hand.

There are a myriad of communication devices deployed to regions of potential interest to coalition forces. In any given situation, however, coalition forces will only face a limited subset of these. A system designed to perform against all of the potential threats would be prohibitively large and expensive. Therefore, CEW systems must be mission tailorable.

5.4.6.1 Information Available from Communications EW Systems

There is a great deal of data and information that is potentially available from a communications EW system. Information has the attributes of timeliness, accuracy, and relevance [11]. The following is an overview of some of the potential products from CEW that could contribute to a network-centric force.

Enemy Electronic Order of Battle Updating

The *electronic order of battle* (EOB) is a description of the electronic devices in use by an adversary and where on the battlefield they are located. With modern display and processing technology, this is perhaps best displayed on a map, with icons indicating unit types. Other information available from the EOB can consist of unit strength, echeloning, unit identification, types of communication and

noncommunication systems in use by the unit, and unit affiliation. Combining this information, an analyst can frequently determine hostile intents and movements. One of the key capabilities/technologies that can assist in the EOB analysis/generation process is the ability to automatically recognize the type of modulation used by the electronic system. A digital signal that is transmitting the computer format V.23, for example, determined by establishing the modulation as *frequency shift keying* (FSK) with specific protocols, almost ensures that one computer is communicating with another over narrowband radios, which, in the low *very high frequency* (VHF, 30–88 MHz), typically is limited to about 600 or 5.2 kbps. Such information implies non-real time data that can be slowly exchanged.

EOB Production

Generating an EOB requires identifying emitters in an *area of responsibility* (AOR), determining their geographic location or range of mobility, characterizing their signals, and, where possible, determining their role in the broader organizational order of battle. EOB covers both *communication intelligence* (COMINT, radios) and *electronic intelligence* (ELINT, radars) targets—that is, both communication and noncommunication (typically radar) emitters. In the United States, the *Defense Intelligence Agency* (DIA) maintains an EOB by location for most of the world. The *Joint Spectrum Center* (JSC) of the *Defense Information Systems Agency* (DISA) supplements this location database with five more technical databases:

- FRRS: Frequency Resource Record System;
- BEI: Background Environment Information;
- SCS: Spectrum Certification System;
- EC/S: Equipment Characteristics/Space;
- TACDB, *tactical database*: platform lists, sorted by nomenclature, which contain links to the C-E equipment complement of each platform, with links to the parametric data for each piece of equipment, military unit lists, and their subordinate units with equipment used by each unit.

The EOB is produced from the databases that hold data on emitters, platorms, threat systems, and signatures, among other things (see Figure 5.5). When it is decided to deploy assets into a particular theater of operations, an AOR will be allocated and an EOB generated for that AOR.

Data is extracted from the data bases based on various source products linked to the AOR and the EOB for that AOR. The data thus extracted is used in the

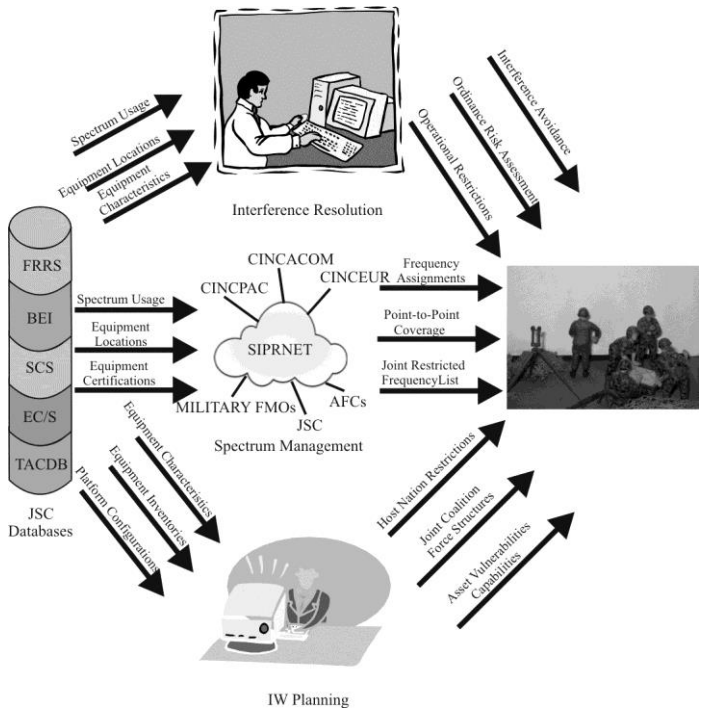


Figure 5.5 EOB generation data flow.
Commander in Chief Atlantic Command (CINCACOM);
Commander in Chief Pacific Command (CINCPAC);
Commander in Chief European Command (CINCEUR);
Area Frequency Coordinator (AFC);
Frequency Management Office (FMO).

programming mission libraries for the particular EW systems to be deployed into the AOR. However, the mission libraries are only as accurate as the available data allows. Therefore, there is the facility to feedback data on emitters encountered in theater to the mission library production stage for rapid reprogramming of systems, and the EOB so that it is updated to reflect the current emitter situation in theater. In this way the EOB is continually updated by the original source products and feedback for the AOR.

Threat Warnings

The intercept of communications can provide *indication and warning* (I&W) of threat actions. The intercept of non-communication signals can also carry I&W information. A good example of this is a weapons radar that switches from search mode to tracking mode.

Combat Information

Combat information is information gathered by tactical CEW sensors as well as other sensors that is immediately useful for some process. This is opposed to intelligence information that takes some amount of processing to produce useful information. Of course, combat information feeds the intelligence process just as all other data collected or provided from existing databases.

Geolocations/PFs of electronic equipment can be key indicators of impending enemy actions, especially when combined with other indicators such as target identification and/or specific emitter identification. PFs combined with *specific emitter identification* (SEI) can also indicate unit movement. More accurate PFs, of course, lead to better such indications.

When precoordinated and preplanned, direct sensor to shooter links can be established so that when the CEW (or other) sensor detects specific target types, (typically) indirect fires can be tasked. This is also true when a specific target is detected, such as a TOC associated with a specific unit.

Communication Signals Externals and Internals

The internal information from a signal refers to the content or meaning of the information being exchanged over a communication link. Technology is available to automatically determine this information only to a very limited extent, so such can generally only be generated by a human analyst. External information refers to data gleaned by automatically measuring parameters based on features that can be determined without determining the information meaning or content. These external parameters are not solely associated with measurements made on raw intercept of signals, but may be based on demodulated signals. PSK baud rate, for example, requires first the ability to determine that the signal is PSK (modulation

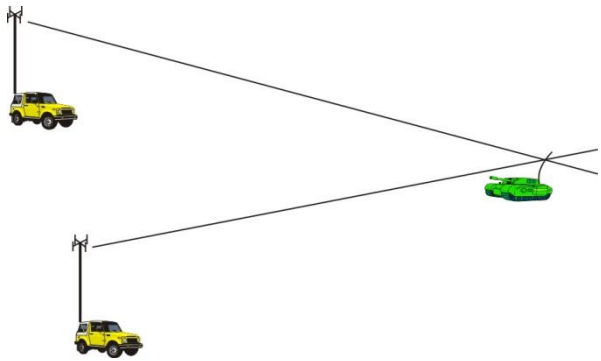


Figure 5.6 Triangulation with multiple ES systems.

recognition) and then to determine the order of the modulation (binary, quaternary, and so forth). Once these are determined, the baud rate can be determined. All of this can be done automatically without the intervention of a human analyst, and, depending on the particular target and scenario, could provide important target identification information and therefore combat information.

Combat information, requiring no human intervention to use its value, determined by CEW sensors is largely based on external parameters associated with signals. There could be exceptions to this—the intercept of an unencrypted message that identifies a particular unit of intense interest, for example, would qualify as combat information. This example would require an analyst to determine the content of the message, but it does not require extensive analysis and correlation with other data to measure its importance.

Intelligence

Communication EW can make a significant contribution to the intelligence process. The following are a few types of information that can be provided for intelligence.

Geolocations/PF/LOBs

Communication EW systems provide geolocations on intercepted signals. This is accomplished either by measuring the angle of arrival of a signal at two or more systems and calculating where the resultant LOBs intersect (triangulation; see Figure 5.6) or by measuring the TDOA and/or the differential frequency at two or more systems. Typical geolocation accuracies provided by these techniques are illustrated in Table 5.1.

Table 5.1 Emitter Geolocation Accuracy Provided by CES Sensors

Range to Target	Measurement Accuracy	Accuracy
10 km	LOB: 5° RMS	1 km
	TDOA: 50 ns DD: 0.01 Hz	50m
20 km	LOB: 5° RMS	2 km
	TDOA: 50 ns DD: 0.01 Hz	100m

The TDOA between two systems results in iso-TDOA contours [12] as illustrated in Figure 5.7. These isochrones provide approximate curves upon which the emitter lies. The FDOA measurements also provide isochrones upon which the emitter lies, but of a different shape as illustrated in Figure 5.8 [12]. In order to generate FDOA contours, one or more of the systems in question must be moving, else there is no differential Doppler generated. The moving system can be any one or more of the sensors or the target.

Combining these techniques can yield the geolocation of the SOI. The intersection of the TDOA and FDOA contours provides an estimate of the geolocation of the target. With only two ES systems, ambiguous results ensue. The emitter can lie at any of four locations. These ambiguities can be removed by providing other information such as LOBs from the sensors to indicate the approximate target location or providing additional TDOA or FDOA measurements by adding another platform. The set of isochrones intersect at a single place (assuming no noise and no errors in the measurements).

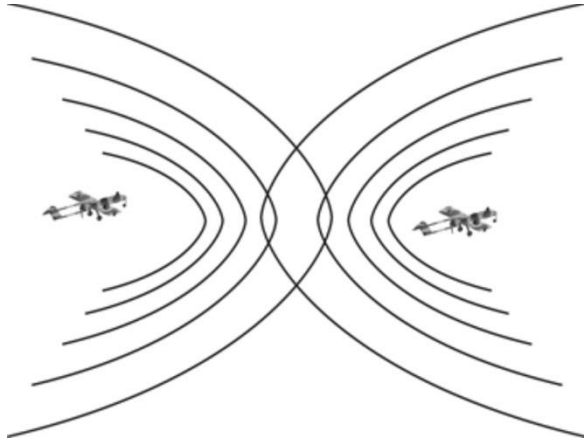


Figure 5.7 TDOA contours.

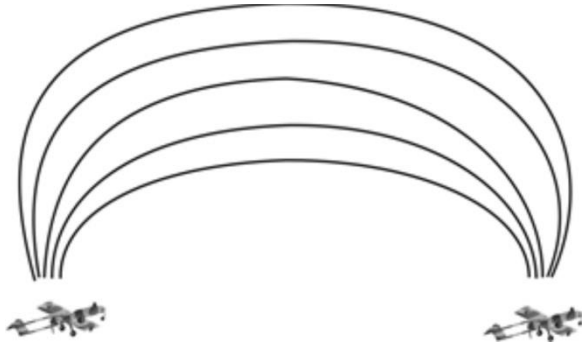


Figure 5.8 FDOA contours.

Electronic Map

The geolocations provided by communication ES systems, combined with other pertinent information such as target type and unit affiliation, can be conveniently plotted on a map of the target area. This electronic map provides a quickly absorbed view of the disposition of target (and friendly) forces. Such a map provides the EOB at a glance.

Modulation Recognition

A parameter useful for several purposes in ES systems is the modulation of a signal. This information can be used to help identify a target by type and affiliation. It can also be used to automatically assign the correct demodulator *within the ES system.*

Azzouz and Nandi [13] describe a particularly thorough and well-performing statistical modulation recognition method for communication signals. Examples of the performance of that algorithm are illustrated in Figure 5.9 [14].

Frequencies in Use

By scanning the RF spectrum and making energy measurements, the frequencies in use by opposition and friendly forces can be determined.

Radio Types

The types of radios being used by opposition forces can sometimes be determined by ES systems. The frequency range is easily determined. That provides an initial

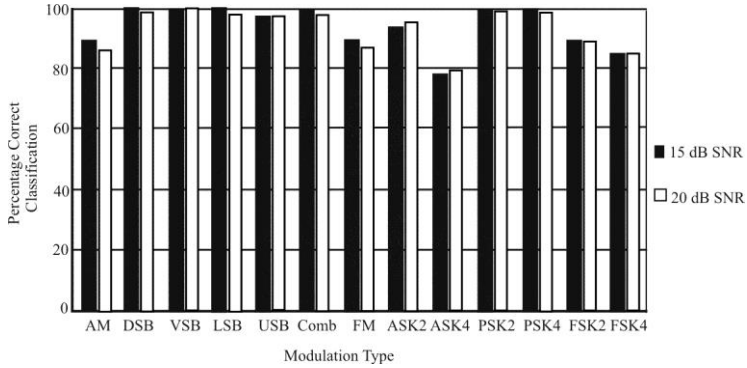


Figure 5.9 Azzouz Nandi modulation recognition algorithm performance.

assessment of the type of radio. The modulation measurements discussed in Section 5.4.4 provide an additional parameter.

Digital/Analog Communications

Most significant modern communication systems are digital, although some analog communications systems are still available. Probably the most notable of these are the commercial radio broadcasts in both the AM and FM frequency ranges. These remain important targets for the military as witnessed by the TV broadcasts in the second Gulf War, sending information to the citizenry on the survivability of Saddam Hussain.

There is also a class of radios that are not digital and are quite prolific. These are small, inexpensive, short-range walkie-talkies. They can be used for short-range squad communications in less developed countries. Important information associated with digital communications include the data rate, protocols (e.g., V 32), and data format. These parameters and others can be automatically measured and provided to higher echelon intelligence systems from ES systems.

5.4.7 Electronic Attack

The SV-4 for EA is shown in Figure 5.10 [9]. The EA target list is formulated by the S3, EWO, and supporting team members based on the current EOB as delineated in FM 3-36. The coordinating group for EW activities is shown in Figure 5.11 [15].

Target frequencies are examined based on the prioritized target list. The active frequency with the highest priority on this list is selected for jamming. That target jamming continues until the preestablished criteria for cessation of jamming is

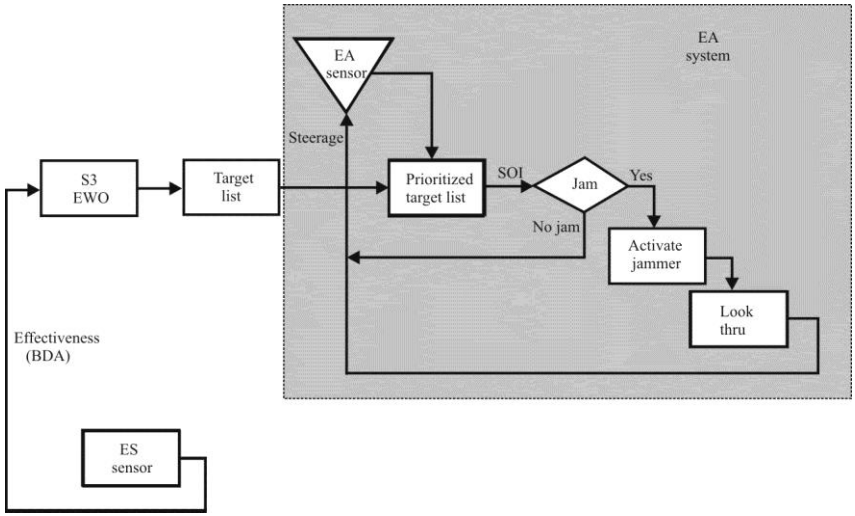


Figure 5.10 Electronic attack process flow (SV-4). (Source: [9]. © British Crown, 2003. Reprinted with permission.)

met, or the target has moved to a new frequency. This is normally determined by a look-through process in the EA sensor shown.

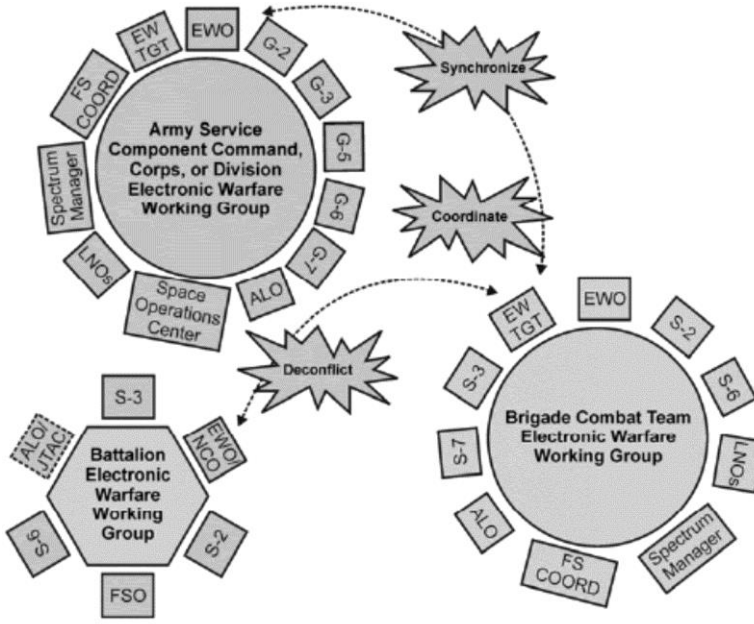
Depending on whether thin or thick jammers are employed, the operational deployment is different. The OV-1 diagram for thick jammers is illustrated in Figure 5.12. In the scenario shown, the thick jammers are deployed with two EW operators in a RSTA platform, usually in a stand-off configuration.

In a thin-jammer configuration, both UAS and ground-based jamming platforms are illustrated in the OV-1 in Figure 5.13. In this case a single EW operator (the EWO or one of the staff), residing in a C2 platform controls the jammers remotely, probably not in realtime. The associated EW sensors at the jammers determine active SOIs based on assignments from the EWO.

Both of these OV-1s are notional, of course, and there are several different ways these EA assets can be employed.

5.4.8 Virtual CEW Organizations

As the information age overtakes military operations, it becomes more important that the use of assets be optimized. A unit commander need not own all of the assets necessary to conduct an information operation. However, the assets that he or she does possess may be part of a larger, virtual organization, constructed so that each individual component contributes in an optimal way [16].



ALO	Air liaison officer	G-7	Assistant chief of staff, information management
EW	Electronic warfare	JTAG	Joint terminal attack controller
EWO	Electronic warfare officer	LNO	Liaison officer
FSCOORD	Fire support coordinator	NCO	Noncommissioned officer
FSO	Fire support officer	S-2	Intelligence staff officer
G-2	Assistant chief of staff, intelligence	S-3	Operations staff officer
G-3	Assistant chief of staff, operations	S-6	Signal staff officer
G-5	Assistant chief of staff, plans	S-7	Information engagement staff officer
G-6	Assistant chief of staff, signal	TGT	Targeting

Figure 5.11 EW coordinating group. (Source: [15].)

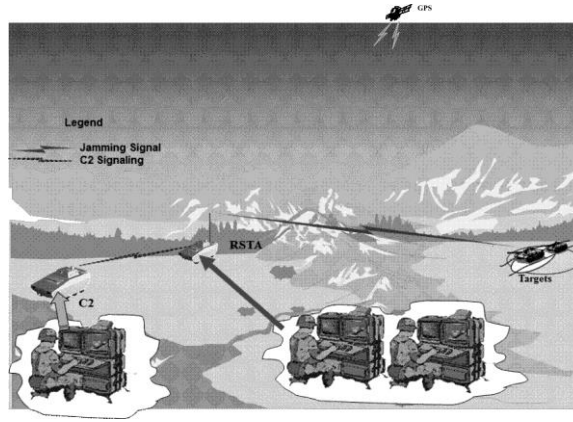


Figure 5.12 Forward EA CONOP (OV-1) for thick jammers.

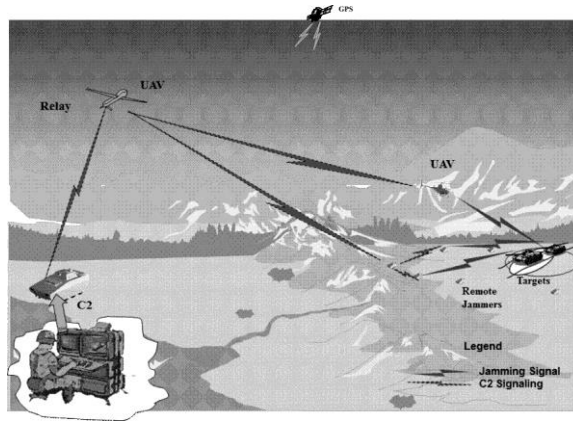


Figure 5.13 Forward EA CONOP (OV-1) for thin jammers, ground and airborne.

5.4.9 Information Required by Communications EW Systems

Communications EW systems do not operate in a vacuum. For these (often expensive) systems to be effective in producing threat warnings and intelligence for the commander, they must be provided with various forms of information to direct their search and provide context for analysis of the data they collect. We cover those requirements in this section.

5.4.9.1 Priority Intelligence Requirements

The *priority intelligence requirements* (PIRs) are those elements of information required to answer a commander's information needs. Commanders generally ask fairly high-level questions about a situation, and it is up to the S3/G3 staff to generate answers to these questions. They do this by converting the high-level questions into PIRs that, assuming that the answers are not already available in the database, can be used to query databases in higher-level echelons, or to specifically task sensors, both organic or external.

5.4.9.2 Friendly Force Disposition

As part of the COP, the disposition of friendly forces is important information for a CEW system. Target signals received from specific areas occupied by friendly forces could be an important indication of intent.

5.4.9.3 CEOI

The CEOI defines the frequencies in use by friendly forces and at what times they are used. The particular value to CEW systems is avoidance or at least minimization of fratricide for EA actions and avoidance of detecting friendly signals for ES. It should be noted that some frequencies in use by friendly forces could be jammed by EA activities if a target use of the same frequency is deemed to be of higher importance than that friendly use.

5.4.9.4 Common Operating Picture

The *common operating picture* (COP) is essential for CEW analysis to put the data into context. It is also especially needed for EA to avoid fratricide. The COP is developed at the Level 2 fusion step described in Section 5.7.

5.4.9.5 Mission Tasking

The EW systems must have a firm understanding of the mission upon which they deploy. Mission tasking includes, for example, whether the specific EW targets are

all known or if there is some target searching required to find some of the targets. If the preliminary EOB is sketchy, prior to engagement the mission may be only a general search.

5.4.9.6 Start/Stop Times

EW activities are fully integrated with other battlespace planning and as such they have applicability at designated times if their effect is to be maximized. This is especially true for EA, as EA is indirect fire and is useful only at specific times coordinated with other maneuver and fire activities.

5.4.9.7 Technical Information

Certain technical information about target types is necessary for the EW system if the ES and EA components are to be programmed correctly. This programming establishes the details about the collection activity, for example.

Frequencies

The frequencies to examine in directed search must be known. If the tasking includes some general search (see Section 5.4.9.4), then the frequency bands to be searched must be known or estimated. For example, if tactical PTT networks comprise the targets of interest, the low VHF military band (30–88 MHz) would be the likely frequency region to investigate.

Modulations

The modulations used by the targets of interest is another parameter that is required for programming the collection or jamming elements. Quite often the modulations are established by default depending on the frequencies involved. For example, if the targets are employing frequency-hopping LPI methods in the low VHF range, chances are the modulation is BFSK (U.S. SINCGARS radios, for example, employ BFSK as the modulation method).

5.4.9.8 Coverage Regions

The geographical area to be included in the EW processes is useful information for EW systems. This is particularly true when directional antennas are employed. Directional antennas have higher gain in particular directions than other directions and therefore must be pointed. When the coverage area can be provided to the EW system, its performance can be optimized. Of course, if the coverage region cannot be specified because the locations of all the targets of interest are not known, directional antennas cannot be used.

Coverage region information is also useful for eliminating signals not of interest (for example, friendly signals). Reporting on such signals can be precluded and performance optimized.

5.4.9.9 Movement/Deployment

As engagements transpire, chances are that EW systems must move along with the supported force. In fact, operation *on the move* (OTM) is often required where EW activities are conducted when the system is moving (we examine the importance of this capability in Chapter 10 by simulation). Such movements may be dictated based on preestablished events and provided prior to activities or they may be provided in real-time as the events happen.

5.4.9.10 Weather Data/Mobility/Terrain Data

All tactical vehicles have limitations on their mobility. Airborne systems have even more weather constraints than ground-based systems. The weather affects this mobility.

In addition the weather can affect the ability to conduct EW activities. Signal propagation is affected by the weather, especially at higher frequencies where rainfall rates can significantly attenuate signals.

Terrain also affects signal propagation. Signals do not propagate well through mountains. However, signals propagate better over water than over land. Where to deploy EW systems can be significantly affected by terrain limitations because of these considerations. So if particular regions are to be geographically covered, placement of the EW systems must take signal propagation into account. Usually higher ground is better than in valleys. Reverse slopes can be effectively used, however the use of ground reflection of signals immediately in front of the antenna must be considered if target geolocations are required.

5.5 Effects-Based Operations and the Role of EW

The CCRP defines EBO this way [17]:

> Effects based operations are coordinated sets of operations directed at shaping the behavior of friends, neutrals, and foes in peace, crisis, and war.

However, RAND suggests the following definition of EBO [18]:

> Effects-based operations are operations conceived and planned in a systems framework that considers the full range of direct, indirect, and cascading effects, which may–with different degrees of probability–be achieved by the application of military, diplomatic, psychological, and economic instruments.

Whichever definition is used, four principal tenets of EBO are:

1. *Options*. A wide range of options previously unavailable to commanders are provided by network-centric operations and EBO. The actions within these options can be tailored to specific situations at hand to include those observing the actions. Essentially, the probability of kill metric, P_k, associated with platform-centric operations is replaced with the probability of an option being effective, P_o, associated with network-centric operations.
2. *Agility*. With the shared awareness and speed of command facilitated by network-centric operations the networked forces can be more responsive with increased agility to dynamically adapt to an adversary's actions.
3. *Coordination*. Complex actions can be coordinated due to shared situational awareness, mutual understanding of command intent, capacity for synchronization, and self-synchronization leading to unity of effects across all levels.
4. *Knowledge mobilization*. Rapid movement of knowledge allows timely and relevant support to DMs at all levels, from commanders to combat crews. Responsive networking brings a breadth of knowledge to bear.

It has been said that EBO is the end to which NCW is the means to that end.

5.5.1 EW and EBO

The effect provided by EW systems depends on the role into which they are cast. EW is comprised of three principal components: EA, ES, and EP.

EA is the application of RF energy into the receiver of an adversary for the purpose of preventing the adversary's electronic equipment from working. Frequently EA systems contain their own sensors for the purpose of determining whether a particular radio frequency is active. Such sensors are relatively unsophisticated compared to ES sensors (when they are different systems). ES sensors are intended to extract more information from transmissions than EA sensors.

EP is composed of those actions taken to protect our own transmissions from the EA and EP activities of an adversary.

The fundamental principle behind effects-based operations is to mass effects, not systems or personnel. Bringing the appropriate amount of the correct effects to a situation to accomplish the mission is the goal. This measure might be steel on target, it might be pamphlets in an attempt to persuade a crowd, or it may be application of EA to deny adversarial C2, as examples.

EW is one of the effects in a commander's toolkit. EA can be used to deny effective C2 of troops. Over-the-air RF communications has been a mainstay

method to accomplish C2 since World War II. Such communications may be vulnerable to interception, where intelligence and battle data can be gleaned, as well as denial by active countermeasures.

Communication EW is normally targeted against the physical layer of communication networks. The purpose of EA is to deny the reception of network traffic by electronically capturing the receiving equipment. There is a capture effect in analog as well as digital communications. For the latter, it is usually around 0 dB *jam-to-signal ratio* (JSR) or so [19].

5.5.2 Ability to Conduct Effects-Based Operations

Some describe *effects-based operations* (EBO) as outcome-oriented activities directed at enemy behavior, so that the objectives are psychological rather than simply physical, although physical means such as application of kinetic weapons is part of EBO. These activities are focused on the adversary's decision-making process and ability to take action in some coherent manner. Put simply, EBO seek to defeat an adversary's strategy and resolve instead of merely attriting his armed forces. Planning must focus on effects rather than means. For example, targets should be selected for psychological and strategic impact in addition to the level of attrition likely to be achieved. A good example of the application of this notion is from Operation Desert Storm, where Iraqi soldiers would surrender to reconnaissance UAV flying overhead.

The concept is that EBO would be facilitated by the improved understanding of the battlespace, faster command cycles, and precision targeting that are expected to result from NCO. EBO imply using these to identify and target enemy C2 networks, with the aim of cutting the connections between their sensors, shooters, and command hierarchy. In addition, elements of IO are likely to be involved, to target the actual commanders and not just physical C2 systems. Here again, the increased knowledge of the battlespace and of the enemy expected to be available through networked information sources and reachback to cultural experts should facilitate these operations. However, access to current HUMINT is a critical element for EBO, since the primary targets are people's perceptions.

5.5.3 Cueing Other Sensors

The CEW sensor, in its modern form, has the capability to acquire signals virtually as they switch on. Depending on the detail of the system design, there can be considerable bandwidths over which the probability of intercept is virtually 100%, if the received field strength is above threshold. For a land system, a baseline would be able to provide geolocations on emitters very quickly. Further exploitation of the acquired signals would allow their value to be determined. While a single sea or air system alone would not be able to produce geolocations, if operated over a network with other systems, it would be possible to obtain such

information. Thus, CEW sensors arranged to give maximum area of coverage and to produce geolocations provide a very powerful means of locating areas of interest. It is, however, a single sensor view. In general, commanders prefer to have confirming information and, ideally, images of the area in question.

The problem with other sensors is that, although frequently with excellent performance, finding targets is difficult if they do not know where to look. It is like looking through a straw at a panoramic scene; in the end, a picture can be built up, but it takes a long time and consumes significant levels of resources. Using the CEW sensor to provide the cues to other sensor systems has been shown in experiments in the United States to be highly effective.

The need is to link the output from the CEW sensor in some way to other sensor systems. Just providing all the geolocations to the other sensors could prove to be more confusing than enlightening. The dots on the map representing emitters need to be interpreted and the CEW systems use skilled analysts for this. They use their knowledge to identify from all the intercepted signals which are from targets of importance and which are not. The other sensors need the CEW view of the world, but filtered through this screen so that their efforts may be directed to seek the targets of highest value.

As an example, consider the use of a UAS with an imaging sensor. When planning the flight path, information from the CEW systems will be used to indicate where the UAS should be flown to identify high priority targets. Once the planning is over, the CEW analysts will need to know the flight path so that they are able to update the UAS systems controllers with any new locations and targets that are found coming up along that flight path. The CEW analysts need to be looking 10 minutes ahead of the UAS, not 10 minutes behind. This emphasizes the need for a two-way flow of information if this type of operation is to be performed efficiently. The CEW system will need to be able to take the UAS systems flight plan and other details, make sense of them and provide the analyst with a meaningful display. Equally, the UAS system operators will need their computer systems to be able to understand and display all that comes from the CEW system.

The above example uses a sensor outside the EW domain. Probably the greatest number of interactions will be with the radar EW sensor. The systems may frequently be deployed in the same area on the battlefield. Integration of these systems should be an obvious first step, each cueing the other. As well as cueing, the information derived by the analysts in each system can be used to support or cast doubt on interpretations of what the emitter maps actually mean and in defining the intent of the enemy.

5.6 Collaboration

Moffat presented the approach delineated by Perry [20] for quantifying the benefit of collaborating across an information network [21]. We summarize the salient

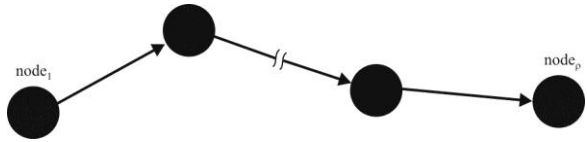


Figure 5.14 Critical collaboration subnet.

characteristics of those results in this section. Full details of the general approach and other areas of application are contained in [20].

We model collaboration as a network of C2 nodes that are involved in coordinating a time-critical operation. Each of these nodes has a number of information processing tasks to perform that contribute to the collaboration. We denote the mean time for node$_i$ to complete all of its tasks by $1/\lambda_i$. We assume an exponential distribution[6] for the completion time of these tasks. Denoting the probability of completing all tasks at node *i* by time *t* by $f_i(t)$, then

$$f_i(t) = \lambda_i e^{-\lambda_i t} \tag{5.2}$$

In general, there will be a number of sequential and parallel nodes in the network involved in the operation being considered. Let this total number be denoted by τ. While in general there can be several, in the simplest case there is a single *critical subnet* consisting of ρ nodes $\{\text{node}_i\}_{i=1}^{\rho}$ that is a subset of $\{\text{node}_i\}_{i=1}^{\tau}$, as shown in Figure 5.14.

The total latency of the sequential path is defined as the sum of the delays (latencies) at each of the nodes, plus the time, denoted by t_m, required to move an attack system (such as an aircraft) to the attack zone. In this sequential case the total expected latency $\mathcal{E}\{T\}$ is simply the sum of the expected latencies at each node on the critical subnet, plus the time t_m

$$\mathcal{E}\{T\} = \sum_{i=1}^{\rho} \frac{1}{\lambda_i} + \mathcal{E}\{t_m\} \tag{5.3}$$

If there are sequential and parallel nodes on the critical subnet, the latencies of the critical subnet are combined as shown in Figure 5.15. In this example

[6] An exponential distribution is used to model the time between events or how long it takes to complete a task for a Poisson process, a reasonable model of such processing times. That is, it describes the time between events in a Poisson process, that is, a process in which events occur continuously and independently at a constant average rate.

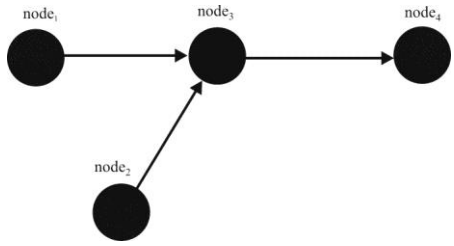


Figure 5.15 Parallel nodes in the critical subnet.

$$\mathcal{E}\{T\} = \max\left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}\right) + \frac{1}{\lambda_3} + \frac{1}{\lambda_4} + \mathcal{E}\{t_m\} \tag{5.4}$$

For the set of nodes that constitute the critical subnet as in Figure 5.15, for each such node$_i$ on the critical subnet the *indegree* d_i is the number of edges for which node$_i$ is a terminal node. For example, $d_3 = 2$ while $d_4 = 1$ in Figure 5.15.

For each node$_j$ in the C2 network, in order to quantify the amount of knowledge available at node$_j$ concerning its ability to process the information and provide quality collaboration, we assume it is a function of the uncertainty in the distribution of *information processing time* $f_j(t)$ at node$_j$.

Let $H_j(t)$ be the Shannon entropy of the function $f_j(t)$. Then $H_j(t)$ is a measure of this (residual) uncertainty defined in terms of a lack of knowledge. This Shannon entropy is

$$H_j(t) = -\int_0^\infty \ln\left(\lambda_j e^{-\lambda_j t}\right)\lambda_j e^{-\lambda_j t} dt$$

but since

$$\frac{d(xe^x - e^x)}{dx} = xe^x$$

we have

$$H_j(t) = \ln\frac{e}{\lambda_j} \tag{5.5}$$

Denote the minimum rate of task completions at node j by $\lambda_{j,\min}$. Then $1/\lambda_{j,\min}$ corresponds to the maximum expected time to complete all tasks at node$_j$.

The normalized knowledge $K_j(t)$ available at node$_j$ in terms of the Shannon entropy is defined as

$$K_j(t) \triangleq \ln\left(\frac{e}{\lambda_{j,\min}}\right) - \ln\left(\frac{e}{\lambda_j}\right) = \begin{cases} 0, & \lambda_j < \lambda_{j,\min} \\ \ln\left(\frac{\lambda_j}{\lambda_{j,\min}}\right), & \lambda_{j,\min} \le \lambda_j \le e\lambda_{j,\min} \\ 1, & \lambda_j > e\lambda_{j,\min} \end{cases} \quad (5.6)$$

Suppose now that node$_i$ is on the critical subnet, and node$_j$ is another network node feeding node$_i$. Let c_{ij} represent the quality of collaboration obtained by including node$_j$. If this is high, $K_j(t)$ should be close to 1. We assume that the effective latency at node$_i$ is reduced by the factor $[1-K_j(t)]^{\omega_j}$ because of this high quality of collaboration. ω_j is assumed to be 1 if j is one of the nodes directly involved in the time-critical operation (but not in the critical subnet). It is assumed to be 0.5 if node$_j$ is one of the other network nodes, to reflect a lower level of collaboration quality with these nodes. These values are, of course, somewhat arbitrary but are indicative of reasonable values.

The total (equivalent) reduction in latency at node$_i$ due to collaboration with the network nodes connected to node$_i$ is then given by

$$\begin{aligned} c_i &= \prod_{j=1}^{d_i} c_{ij} \\ &= \prod_{j=1}^{d_i} [1-K_j(t)]^{\omega_j} \end{aligned} \quad (5.7)$$

Thus, the total effective latency along the critical path, accounting for the positive effects of collaboration, is given by

$$\begin{aligned} \mathcal{E}\{T_c\} &= \sum_{i=1}^{\rho} \frac{c_i}{\lambda_i} + \mathcal{E}\{t_m\} \\ &= \left\{ \sum_{i=1}^{\rho} \frac{1}{\lambda_i} \prod_{j=1}^{d_i} [1-K_j(t)]^{\omega_j} \right\} + \mathcal{E}\{t_m\} \end{aligned} \quad (5.8)$$

The balance sought is that between such enhanced collaboration and the effects of information overload due to increasing network complexity which we consider next.

5.6.1 Information Saturation

The impact of information *saturation* is the negative effect of collaboration. Additional network connectivity and complexity can lead to such information overload effects. This can lead to an increase in effective latency in the critical subnet. Let N_c denote the total number of network connections accessed by nodes in the critical subnet. For each node$_i$ on the critical subnet, this is the indegree d_i, and we have

$$N_c = \sum_{i=1}^{\rho} d_i \tag{5.9}$$

The value of N_c is then a measure of the complexity of the network.

Denote by g the measure of the normalized benefit to collaboration. Normalization is with respect to the maximum benefit possible. Assume that the rate of change of g is proportional to g as well as the amount of benefit yet to be realized so that

$$\frac{dg}{dN_c} = b\left(\frac{g}{1-g}\right) \tag{5.10}$$

Denote by N_{c0} the number of collaborations when g is 50%. Then we get the benefits of collaboration displayed as

$$\int_{N_{c0}}^{N_c} b\left(\frac{g}{1-g}\right) dN_c \tag{5.11}$$

Solving (5.11) for g, we get

$$g(N_c) = \frac{1}{1+\exp[-b(N_c - N_{c0})]}$$

or

$$g(N_c) = \frac{1}{1+c\exp(-bN_c)} \tag{5.12}$$

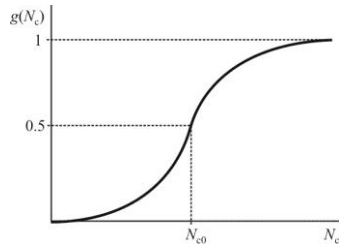


Figure 5.16 Logistic S-curve.

Equation (5.12) is the expression for the nonlinear *S*-shaped *logistic curve* shown in Figure 5.16. This chart implies that as the complexity increases (N_c increases), at some point there begins a reduction in the increased value of including new nodes. In the limit, there is no increase in value by adding nodes.

The penalty for information overload is then defined as

$$P(N_c) = \frac{1}{1 - g(N_c)} \tag{5.13}$$

Example 5.1: Suppose $N_{c0} = 5$. Then the collaboration benefits for a few values of the variables are illustrated in Figure 5.17. In this case, when $b = 0.5$, the benefit of additional collaborators is still increasing when $N = 10$, but when $b = 0.9$, cutting the number at around 7 or 8 is reasonable. The corresponding penalties are shown in Figure 5.18 and support the same

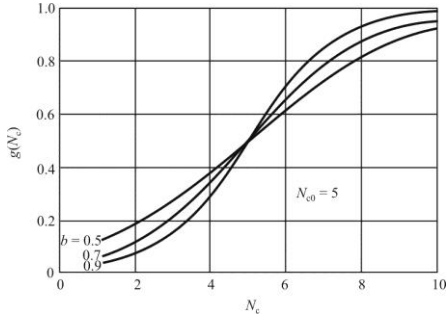


Figure 5.17 Collaboration performance as given by the normalized benefit of collaboration, *g*. After a point, the benefits of adding collaborators starts to decrease.

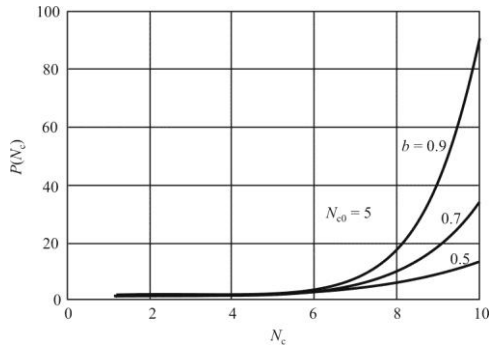


Figure 5.18 Collaboration performance as given by the penalty for information overload. The penalty for adding collaborators can bet quite large.

conclusion. The penalties for including more than seven or eight collaborators when $b = 0.9$ can get quite large.

The average total effective latency, taking account of both the positive and negative effects of C2 network collaboration, is then

$$\mathcal{E}\{T_{c,N_c}\} = \frac{\mathcal{E}\{T_c\} - \mathcal{E}\{t_m\}}{1 - g(N_c)} + \mathcal{E}\{t_m\} \tag{5.14}$$

We see that $P(N_c)$ is a multiplier to $\mathcal{E}\{T_c\} - \mathcal{E}\{t_m\}$ so there is considerable incentive to keep it as small as possible.

5.6.2 Network-Centric Benefit

This network-enabled approach allows us to compute the distribution of the response time of the system as a function of the network assumptions. As the collaboration throughout the network is increased in going from platform-centric to network-centric, the positive effects of enhanced collaboration have to be balanced against the downside effects of information overload and increased network complexity.

5.7 Data and Information Fusion

All military activities produce "observables" related either to time, space, and mass, or to a specific platform or system. It is the observable that sensors detect and that the fusion process develops into meaning. This is the friendly, adversary,

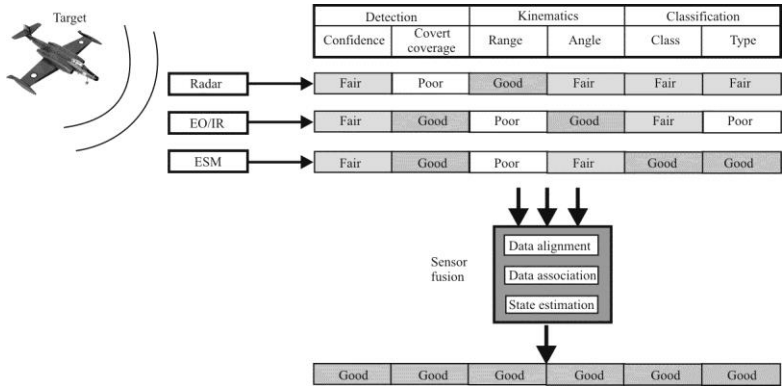


Figure 5.19 Data association uses overlapping sensor capabilities so that wtate estimation can exploit their complementary capabilities.

and neutral information about which IW is concerned. Currently there is so much information available to a decision maker that sifting out the important items is very problematic. In the future, we must process and analyze a variety of collected enemy observables from different, but complementary systems, and more rapidly produce actionable information and intelligence for DMs.

> **Definition:** *Fusion* is a series of processes performed to transform observational data into more detailed and refined information, knowledge, and understanding. These processes, by their very nature, involve both automation and human cognition.

The advantage of fusing data and information together is illustrated in Figure 5.19 [22]. Each sensor in this scenario has its advantages and limitations. The radar sensor's confidence is fair at detection of the target and good at covert coverage. The radar is good at measuring the range and range rate and fair at measuring the angle. It is fair at classification of the class of the target (e.g., fighter or bomber) and fair at determining the type of target. The other sensors in the mix are measured similarly. However, when all three are combined together, where one sensor is poor or fair, another may be good. So the conglomerate sensor is good at all categories once the information is fused.

Fusion takes place at multiple levels. Some sensors with on-board processing capability will fuse information independently (such as an ASE sensor linking a radar signature to its associated air defense system), while others require the network. Fusion systems can assist, but not replace, human intervention. This is because:

- Opponents will develop countermeasures designed to deceive or confuse friendly sensors.
- Effects of weather and terrain can degrade the capability of collection systems.
- The fog of war creates a dirty sensor environment that may limit some capabilities.
- Some degree of error is inherent in every form of collection means.
- Not every sensor will be capable of collecting every observable. A static, well-camouflaged enemy may avoid detection by an EO UAS and in radio silence, signals intelligence (SIGINT) collectors as well. A thermal collector may identify the enemy, but may introduce ambiguity as to specific identity and purpose.

Fusion of data, information, intelligence, and knowledge will remain an integrated organizational, technological, and human endeavor.

Also included as part of fusion are the databases, human interfaces and information portrayal, and the control and feedback of the fusion process. To be relevant, the products of fusion must be accurate, timely, usable, complete, precise, and reliable as discussed in Chapter 2. These products can be presented in forms like the COP, running estimates, and any other form that assists the DM in visualizing and understanding the battlefield or at an individual weapon system such as fusing various MASINT sensors with a UAS imagery feed to produce a better picture for the combat crew.

5.7.1 The Need for Fusion

The average human span of attention and control is limited to three to five things. Current and anticipated advances in information technology, battlefield sensors (quantity and quality), and communications have combined to literally overwhelm the ability of human beings to achieve a sufficient level of understanding about the battlefield environment. As illustrated in Figure 5.20, literally thousands of reports are available in the battlespace at all levels that may contain potentially valuable information. This flood of information combined with ever compressed decision cycle times places a paralyzing information burden on the decision-makers.

Fusion at the lowest possible level can help to manage this flood of information as depicted in Figure 5.21. Fusion can help to alleviate some of this information overload by providing a means to merge the various sources of data into a more coherent picture.

Knowledge is the key enabler of future forces. These forces conduct dominant maneuvers based on information superiority and decision dominance, as we discussed in Chapter 3. The sensors will generate a significant increased

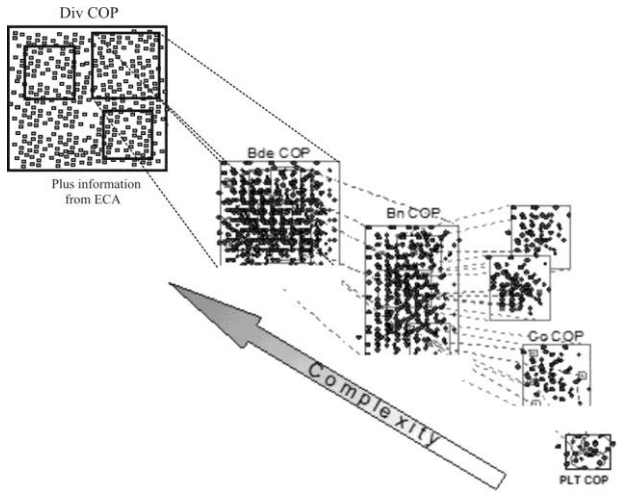


Figure 5.20 Reports going up-echelon without fusion.

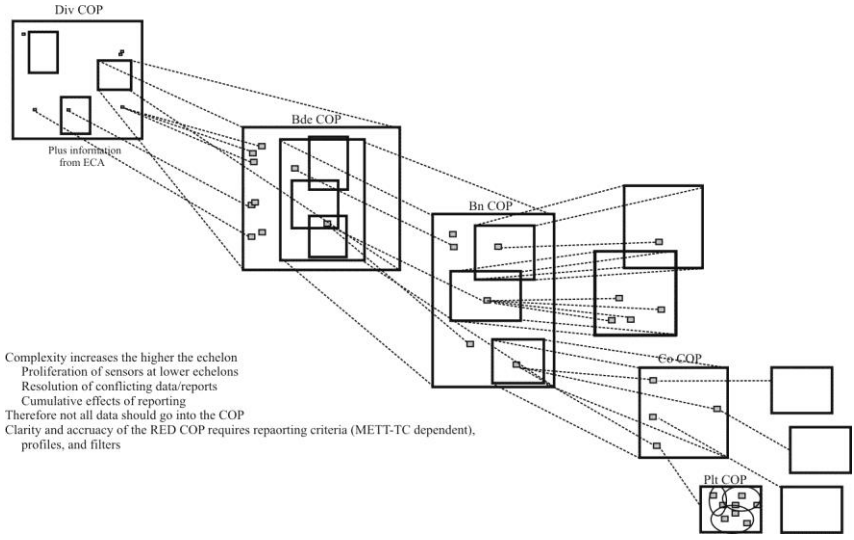


Figure 5.21 Reports with fusion at all levels.

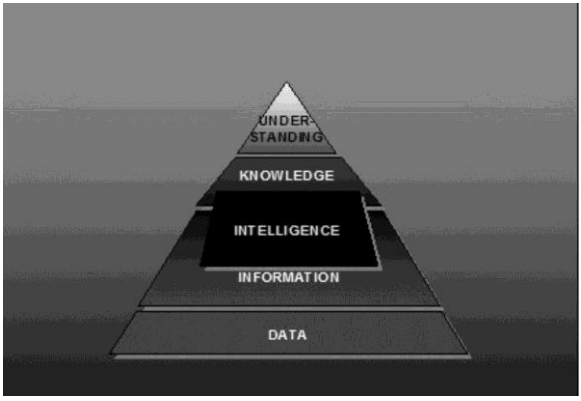


Figure 5.22 Cognitive hierarchy. (Source: [23].)

stream of discrete data that, unaltered, is of limited value. The ability to gain and maintain information superiority depends upon our ability to generate knowledge.

Information superiority is the operational advantage derived from the ability to collect, process, and disseminate an uninterrupted flow of relevant and accurate information while exploiting or denying the adversary's ability to do the same.

5.7.2 Cognitive Hierarchy—Revisited

The basic building block of fusion is the cognitive hierarchy discussed in Chapter 2. Cognition is the act of learning from and integrating various pieces of information. The cognitive hierarchy depicts the different levels of meaning assigned as fusion takes place. Data, information, intelligence, and knowledge are transformed, through fusion, by adding progressively greater meaning at each level. This transformation occurs from the lowest level (data) to the highest (understanding).

Data is the lowest level in the cognitive hierarchy. Data alone is of limited use until it is processed to give it meaning.

Information is composed of data to which meaning and relevance has been applied. Data is processed and/or analyzed to become information. This processing and analysis may include filtering, formatting, organizing, collating, correlating, plotting, translating, categorizing, and arranging. Some automated processing as well as humans can create information from data.

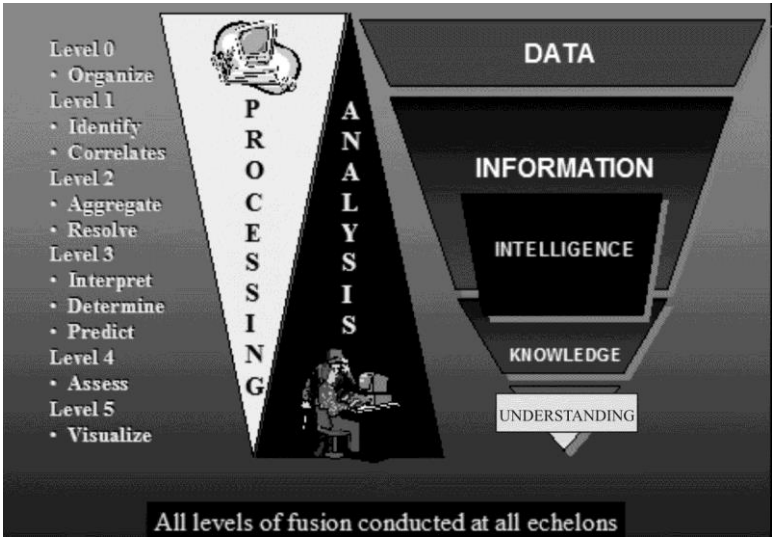


Figure 5.23 U. S. Army Intelligence Center of Excellence–defined fusion levels. (Source: [23].)

Intelligence concerns adversaries or entities of interest to the commander, weather effects, and terrain considerations. While not considered a level in the cognitive hierarchy shown in Figure 5.22 [23–26], intelligence results from the application of cognitive skills to provide relevance and meaning to data and information. It is important to distinguish between information and intelligence. Information is anything that can be known, regardless of how it may be discovered. Intelligence refers to information that meets the stated or understood requirements of commanders. All intelligence is information, but not all information is intelligence.

Knowledge results from the evaluation and analysis of information and/or intelligence.

Understanding is the ultimate goal of analysis and fusion and purely a human endeavor.

As data moves up the cognitive hierarchy, it is transformed through fusion. As illustrated in Figure 5.23, as the fused data/information moves up the cognitive hierarchy, less automated data fusion is applied while the amount of human analysis increases. Since machines can process and categorize many types of data more quickly and efficiently than people, much of this process can be and is automated.

Although some machines and processes can be taught to learn, cognition (learning) is primarily a human mental activity. Knowledge is generated by the DM and staff via this cognition. Various pieces and categories of information that have been integrated and interpreted are used to begin to build a picture of the situation.

The DM applies judgment to transform knowledge into understanding and estimates what is happening. Judgment is based on training, experience, expertise, and intuition. When the DM achieves an understanding of the situation, patterns emerge from events and he or she anticipates consequences of the operational environment, and the interaction between his or her force and that of the adversary. While understanding should be the basis for the DM's decisions, recognition of uncertainty and time typically precludes perfect understanding before deciding and acting (the fog of war). The DM normally conveys his understanding through dissemination of his or her intent, planning guidance, or information requirements.

5.7.3 Fusion Levels

The U. S. Army Intelligence Center of Excellence has established that there are six levels of fusion (0–5); we define their characteristics here [23]. However, levels 1 through 3 add progressively greater meaning and involve more analysis. Level 4 is continuous and occurs at all levels of fusion. The fusion levels are:

- Level 0: Organize;
- Level 1: Identify/Correlate (Beginning of situational awareness);
- Level 2: Aggregate/Resolve (Situational awareness increases and beginning of situational understanding);
- Level 3: Interpret/Determine/Predict (Situational understanding achieved);
- Level 4: Assess (Review performance; Adjust accordingly);
- Level 5: Visualize (Feedback; Redirect activities).

Level 0 fusion is the initial processing accomplished at or near the sensor that organizes the collected data into a usable form for the system that will receive it. For example, an imagery sensor digitizes an analog picture with a sensor processor. The data is then formatted and transmitted to another system for processing. With ES sensors, this data may be automatically extracted, such as modulation type, or, when there is an operator, some human cognition may be employed. For a single sensor platform, level 0 fusion is not really fusion in the sense of the higher levels since data is not really combined at this level, just organized for transmission elsewhere. In the case of a sensor platform with multiple sensors, level 0 fusion by combining sensor reports can, and likely does, take place at the sensor platform. It is probably safe to say that when possible,

level 0 fusion *should* take place on a multisensor platform because that typically is where alignment[7] is best accomplished.

Level 1 fusion takes new inputs and correlates it into an existing entity database, and updates that database. This fusion level reduces redundancy, provides the last known disposition or status of an element, and makes the information available in a database. At this level, the information is assigned an identity and, through correlation, is subjected to a series of comparison referencing characteristics such as unit identification, item name, location, date/time of information, and key word descriptions. Information is combined and represented as a single element if the comparisons indicate an identical (or close enough) match. The correlated element of information retains information from the originating report(s) for future reference and cross-reference. However, it would appear as a single piece of information in the database or a single icon on the COP. This fusion level can be largely automated, but may require some analysis for the information that does not exactly match the correlated parameters. The results of level 1 fusion indicate the threat that is being observed and can result in actionable information.

Level 2 fusion aggregates individual entities or elements, analyzes those aggregations, and resolves conflicts. Dissimilar information is analyzed and combined into a product. Level 2 fusion indicates how two or more target entities are working together and what they are doing. This is where we begin to develop situational awareness.

Level 3 fusion interprets enemy events and actions, determines enemy objectives and how enemy elements operate, and predicts enemy future actions and their effects on friendly forces. Level 3 fusion is a threat refinement process that projects current situations (friendly, threat, and neutrals) into the future, drawing inferences on the threat and vulnerabilities for both sides with the goal of predicting intent and strategy.

Level 4 fusion consists of assessing the entire process and related activities to improve the timeliness, relevance, and accuracy of information and/or intelligence. It is where the performance of sensors and collectors, as well as analysts, information management systems, and staffs involved in the fusion process is reviewed. Based on the assessment, it is decided what adjustments are required concerning the procedures and systems involved in the fusion process. Level 4 fusion indicates what must be done to improve the products from fusion levels 1–5.

Level 5 fusion is processes that allow the user to control fusion levels 1–4. The user can visualize the fusion products and generate feedback/control. It is at level 5 that knowledge gaps are discovered and turned into feedback for the other

[7] Alignment is the association of multiple sensor reports with one another, either in time or space. For example, aligning an EO sensor report with a RF DOA in time could support the notion that the two reports are on the same target.

fusion levels. It is similar to the cognitive process where humans apply their experience to assess the current or new situations to refine solutions.

5.7.4 Human Interaction

In the fusion processes, at all levels, the human provides attributes that automation alone cannot: curiosity, judgment, experience, and intuition. The priority of automated fusion is to rapidly present actionable information to DMs and shooters at every level. In addition, it is important to feed information to analysts and provide them automated support so they can quickly present intelligence to DMs and shooters.

5.7.5 Summary

The principal purpose of fusion at all levels is to produce actionable information and intelligence to the commander. As we discussed, some characteristics of fusion lend themselves to automation while others require human cognitive abilities. The goal is to automate as much as possible. However, for the foreseeable future it will not be possible to automate all aspects of fusion. Technology will enhance the capability to automate.

Information technology systems presenting information and enabling situational awareness development must support the human cognitive process by conveying information in a form that can be understood by the user and assist the user by projecting future events based on known information. Understanding is conveyed from, between, and within echelons as level 2 situational awareness with enablers that assist the user in rapid formulation of level 3 situational awareness.

5.8 Concluding Remarks

The basic characteristics of NCW were presented in this chapter, including the dissenting views of Reid and Giffin. The EW contributions to NCW were the focus, including the contributions to SA and targeting. Also discussed were methods of EW target analysis and EW intelligence analysis. We considered what information is available from EW systems and what information is required by EW systems to perform their functions. We discussed the role of EW functionality in EBO.

A discourse on the benefits and drawbacks of collaboration was included. The penalty incurred with including too many collaborators was demonstrated.

We concluded the chapter with a discussion of data and information fusion, pointing out why it is necessary. We also discussed the six levels of fusion and their basic characteristics, including their relationship with the levels in the cognitive hierarchy.

References

[1] Alberts, D. S., J. J. Garstka, and F. P. Stein, *Network Centric Warfare,* Washington D.C.: CCRP Publications, 1999, p. 2.

[2] Reid, D. J., and R. E. Giffin, "A Woven Web of Guesses, Canto One: Network Centric Warfare and the Myth of the New Economy," "A Woven Web of Guesses, Canto Two: Network Centric Warfare and the Myth of Inductivism," and "A Woven Web of Guesses, Canto Three: Network Centric Warfare and the Virtuous Revolution," *Proceedings 8th International Command and Control Research & Technology Symposium*, 2003.

[3] Alberts, D. S., *The Agility Advantage*, Washington, D.C.: CCRP Publications, 2011.

[4] Popper, K. R., *The Open Universe: An Argument for Indeterminism*, Routledge, London and New York, 1988.

[5] Popper, K. R., *Conjectures and Refutations: The Growth of Scientific Knowledge*, London and New York: Routledge, 1995.

[6] Popper, K. R., *All Life is Problem Solving*, London and New York: Routledge, 2001.

[7] Popper, Z. R., *The Logic of Scientific Discovery*, London and New York: Routledge, 2005.

[8] Maier, M. W., and E. Rechtin, *The Art of Systems Architecting*, 2nd ed., New York: CRC Press, 2002, Ch. 11.

[9] Elsaesser, D. S., and F. Rivest, "Network Centric Concepts for EW in the Canadian Land Forces (U)," Working Paper for TTCP EWS AG4 and TP4 Workshop, Ft. Monmouth, NJ, April 2–4, 2003.

[10] Elsaesser, D. S., and R. G. Brown, "An Intelligence and Electronic Warfare Simulation (IEWSIM) Facility (U)," Unclassified, DREO Technical Report 2001-116, Defense Research Establishment, Ottawa, CA, November 2001.

[11] Alberts, D. S., J. J. Garstka, and F. P. Stein, *Network Centric Warfare,* Washington, D.C.: CCRP Publications, 1999, pp. 54–58.

[12] Poisel, R. A., *Introduction to Communication Electronic Warfare Systems*, Norwood, MA: Artech House, 2002.

[13] Azzouz, E. E., and A. K. Nandi, *Automatic Modulation Recognition of Communication Signals*, Boston, MA: Kluwer Academic Publishers, 1995.

[14] Nandi, A. K., and E. E. Azzouz, "Algorithms for Automatic Modulation Recognition of Communication Signals," *IEEE Transactions on Communications*, Vol. 46, No. 4, April 1998, pp. 431–435.

[15] FM 3-36, *Electronic Warfare in Operations*, Washington, D.C.: Headquarters Department of the Army, February 26, 2009.

[16] Alberts, D. S., J. J. Garstka, and F. P. Stein, *Network Centric Warfare,* Washington, D.C.: CCRP Publications, 1999, pp. 38–40.

[17] Smith, E. A., *Effect Based Operations, Applying Network Centric Warfare in Peace, Crisis, and War*, Washington, D.C.: CCRP Publications, 2002, p. 108.

[18] Davis, P. K., *Effects-Based Operations, A Grand Challenge for the Analytical Community*, Santa Monica, CA: RAND, 2001.

[19] Poisel, R. A., *Modern Communications Jamming Principles and Techniques*, Norwood, MA: Artech House, 2005.

[20] Perry W., et. al, *Measures of Effectiveness for the Information-Age Navy: The Effects of Network-Centric Operations on Combat Outcomes*, Santa Monica, CA: RAND, 2002.

[21] Moffat, J., *Complexity Theory and Network Centric Warfare*, Washington, D.C.: CCRP Publications, 2003, pp. 130–136.

[22] Steinberg, A. N., et. al, "Revisions to the JDL Data Fusion Model," *Proceedings Joint NATO/IRIS Conference*, Quebec, October 1998.

[23] "White Paper: Objective Force Fusion," US Army Intelligence Center, Directorate of Combat Developments, Ft. Huachuca, AZ, March 2003.

[24] Kessler, O., K. Askin, N. Beck, J. Lynch, F. White, D. Buede, D. Hall, and J. Llinas, *Functional Description of the Data Fusion Process*, Warminster, PA report prepared for the Office of Naval Technology Data Fusion Development Strategy, Naval Air Development Center, November, 1991.

[25] Steinberg, A. N., C. L. Bowman, and F. E. White, "Revisions to the JDL Model," *Joint HATO/IRIS Conference Proceedings*, Quebec, October 1998.

[26] Hall, D. L. and S. A. H. McMullen, *Mathematical Techniques in Multisensor Data Fusion*, Norwood, MA: Artech House, 2004.

Chapter 6

Networking

6.1 Introduction

Data communications at the tactical edge are characterized by intermittent communications between many elements that are frequently not within radio contact with one another. Almost constant movement is to be expected with two elements that can communicate now, being unable to the next minute. NCW requires that the situation awareness be available to the maximum number of battlefield facilities (FACs) as possible in more or less real time. This requires adequate communication among these elements to include the EW resources in the area, which are a source of a significant portion of the SA information.

The technical answer that has evolved to address these communication requirements at the tactical edge are referred to as *mobile ad hoc networks* (MANETs). Mobile ad hoc networks, simply stated, are unplanned, self-organizing networks composed of mobile nodes that utilize mesh networking principles for interconnectivity. In this chapter we will examine the advantages and disadvantages of MANETs, and then decompose a MANET into its major functional components.

MANETs offer several significant advantages to a military force. A MANET's ability to self-form and self-manage eliminates the need for intensive central management of network links, thus reducing support personnel and equipment requirements in forward areas. By their very nature, MANET technologies allow a force of mobile nodes to more easily share data and attain greater SA than a non-networked force. This increased SA is the cornerstone-enabling capability for the NCW tenets of cooperative engagement and self-synchronization [1]. These benefits, however, do not come without some disadvantages.

MANETs suffer from the same limitations as fixed wireless mesh networks, but also are vulnerable to additional challenges resulting from their inherent mobility. One of the strengths of traditional wireless networks is the ease of user

node mobility. The critical distinction between a typical wireless network and a MANET is that the wireless network's primary routing infrastructure tends to be static around a fixed entry point into the Internet. In a MANET, the entire network infrastructure is moving along with the user nodes. As the nodes move, point-to-point links may be dropped due to terrain interference or simply because they move beyond range of other nodes. Network stability is continually stressed as nodes drop in and out of the mesh. MANETs may also have limited access to fixed GIG entry points, which ultimately diminishes, but does not eliminate, the overall capability of a MANET while "disconnected" from the broader GIG.

Encryption of the MANET links is facilitated by standard means, allowing protection of sensitive tactical information in the forward areas.

Before we delve into the details of MANETs, we discuss computer networking in general, as a MANET is an implementation of such networking, albeit with some unique characteristics. Specifically, we present the OSI seven-layer model for networks, and then we discuss some details of the Internet and how they relate to this model. After that, we introduce MANETs and their relationships to both the OSI seven-layer model and the Internet.

This chapter is also devoted to MANET security issues. We investigate the major security concerns surrounding the use of MANETs in a military setting where adequately securing information is a must, while not precluding communications at the tactical edge altogether. (It was once said that the ultimate communication security system can be built with absolute security assured. Unfortunately, no information could be exchanged over such a system.)

6.2 Computer Networks

A computer network is a system for communication between computers. These networks may be fixed (cabled, permanent) or temporary (as via modems). In its most basic form, a computer network is three or more computers connected via a communications system for the purpose of sharing data and/or resources, such as a printer. Although the communications systems used to build a computer network can vary, the most common types are based by far on the open systems interconnection basic reference model or, OSI model. The OSI model was developed by the *International Standards Organization* (ISO) beginning in 1977 and is characterized by the seven-layer network abstract model as shown in Figure 6.1.

Each layer provides specific functionality for the overall networking protocol. A given layer's functionality is implemented by one or more entities (either software, hardware, or both, dependent upon the specific layer) which provide services to the neighboring higher layer and communicate directly only with entities in the next lower level. The entities in each layer in the model may only communicate with the layer immediately above or immediately below inside the

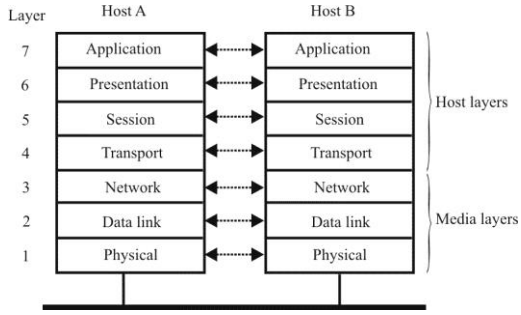


Figure 6.1 The seven-layer OSI model of message transport.

same host (computer or node) or with the same layer of a different host (e.g., the network layer in host A in may communicate only with the transport layer or data link layer in host A or the network layer in host B).

The OSI model separates the communications process into seven layers, which divides the tasks involved with moving information between networked computers into seven smaller, more manageable task groups. A task or group of tasks is then assigned to each of the seven OSI layers. Each layer is reasonably self-contained so that the tasks assigned to each layer can be implemented independently. This enables the solutions offered by one layer to be updated without adversely affecting the other layers.

The OSI seven-layer model has clear characteristics. Layers 7 through 4 deal with end-to-end communications between data source and destinations. Layers 3 to 1 deal with communications between network devices. See Figure 6.1.

However, the seven layers of the OSI model can be divided into two groups: upper layers (layers 7, 6, and 5) and lower layers (layers 4, 3, 2, and 1). The upper layers of the OSI model deal with application issues and generally are implemented only in software. The highest layer, the application layer, is closest to the end user. The lower layers of the OSI model handle data transport issues. The physical layer and the data link layer are implemented in hardware and software. The lowest layer, the physical layer, is closest to the physical network medium (the wires, for example) and is responsible for placing data on the medium.

The major functions for each layer are as follows:

- *Layer 7*: *Application Layer.* Defines interface to user processes for communication and data transfer in the network. Provides standardized services such as virtual terminal, file and job transfer, and operations.
- *Layer 6*: *Presentation Layer.* Masks the differences of data formats between dissimilar systems. Specifies architecture-independent data transfer format. Encodes and decodes data, encrypts and decrypts data, and compresses and decompresses data.

- *Layer 5*: *Session Layer*. Manages user sessions and dialogues. Controls establishment and termination of logic links between users. Reports upper layer errors.
- *Layer 4*: *Transport Layer.* Manages end-to-end message delivery in network. Provides reliable and sequential packet delivery through error recovery and flow control mechanisms. Provides connectionless-oriented packet delivery.
- *Layer 3*: *Network Layer*. Determines how data are transferred between network devices. Routes packets according to unique network device addresses. Provides flow and congestion control to prevent network resource depletion
- *Layer 2*: *Data Link Layer.* Defines procedures for operating the communication links. Frames packets. Detects and corrects packet transmit errors.
- *Layer 1*: *Physical Layer.* Defines physical means of sending data over network devices. Interfaces between network medium and devices. Defines optical, electrical, and mechanical characteristics.

The OSI seven-layer model underlies every popular networking protocol. It was developed as an extension of the original five-layer *Transmission Control Protocol* and *Internet Protocol* (TCP/IP) suite that served as the heart of the ARPANET, the first large-scale, long-distance computer network developed by DARPA that morphed into today's Internet. The Application, Presentation, and Session layers of the OSI Model are encapsulated in the Application layer of the TCP/IP. The remaining layers are consistent between the OSI model and TCP/IP.

The bottom layer, or layer 1, of both the OSI model and the TCP/IP, is the physical layer. The physical layer defines all the electrical and physical specifications for connectivity in a network. One of the primary functions that this layer provides is the conversion of digital data into the appropriate electrical signal for transmission over a communications channel. This signal may be particular voltage on a wired copper cable, a certain wavelength of light for a fiber optic cable or open-air laser, or a specific analog signal for a radio link. The key is that from an overall network perspective, the actual physical connection from one host to another is, ideally, invisible/immaterial to the network as a whole.

6.2.1 The Internet

The Internet has been a great success at interconnecting communication devices across the globe. It has done this by using a homogeneous set of communication protocols, called the TCP/IP suite. All devices on the hundreds of thousands of subnets that make up the Internet use these protocols for routing data and insuring the reliability of message exchanges.

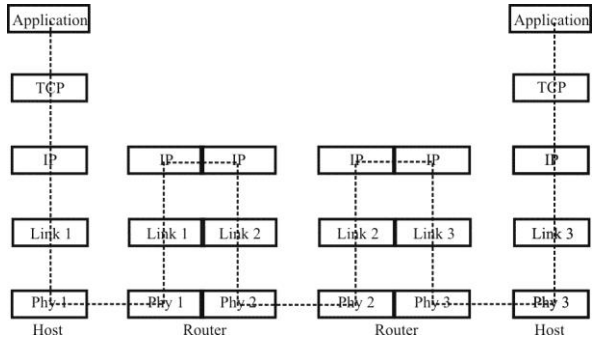


Figure 6.2 Message flow over the Internet.

Connectivity on the Internet has relied primarily on wired links, including the wired telephone networks. However, new wireless technologies, such as short-range mobile connectivities such as WiFi, have appeared and are ubiquitous. Packet switching is used to route messages from one point to another. These links are continuously connected in end-to-end, low-delay paths between sources and destinations. They have low error rates and relatively symmetric bidirectional data rates.

6.2.1.1 Internet Protocol Layers

Messages are moved through the Internet by protocol layers, a set of functions performed by network nodes on data communicated between nodes. Hosts (computers or other communication devices that are the sources or destinations of messages) usually implement at least five protocol layers, which perform the following functions (see Figure 6.2):

- *Application Layer.* Generates or consumes user data (messages).
- *Transport Layer.* Source-to-destination (end-to-end) segmentation of messages into message pieces and reassembly into complete messages, with error control and flow control. On the Internet, the TCP performs these functions.
- *Network Layer.* Source-to-destination routing of addressed message pieces through intermediate nodes, with fragmentation and reassembly if required. On the Internet, the IP is used.
- *Link Layer.* Link-to-link transmission and reception of addressed message pieces, with error control. Common link-layer protocols include Ethernet for *local-area networks* (LANs) and *point-to-point* (PPP) for dial-up modems or very high-speed links.

- *Physical Layer.* Link-to-link transmission and reception of bit streams. Common physical media include category 5 (cat5) cable, *unshielded twisted pair* (UTP) telephone cable, coaxial cable, fiber-optic cable, and RF.

6.2.1.2 Packet Switching

Communication on the Internet is based on *packet switching*. *Packets* are pieces of a complete block of user data (e.g., pieces of an e-mail message or a Web page) that travel independently from source to destination through a network of links connected by routers. The source, destination, and routers are collectively called *nodes*.

Each packet that makes up a message can take a different path through the network. If one link is disconnected, packets take another link. Packets contain both application program user data (the payload part) and a header (the control part). The header contains a destination address and other information that determines how the packet is switched from one router to another. Routing of these packets is determined by the network (IP) layer. The packets in a given message may arrive out of order, but the designation's transport (TCP) layer reassembles them in correct order.

Successful operation of the Internet depends on some fundamental assumptions:

- *Continuous, bidirectional, end-to-end path.* A continuously available bidirectional connection between source and destination to support end-to-end interaction.
- *Short round trips.* Small and relatively consistent network delay in sending data packets and receiving the corresponding ACK packets.
- *Symmetric data rates.* Relatively consistent data rates in both directions between source and destination.
- *Low error rates.* Relatively little loss or corruption of data on each link.

Internet routing is depicted in Figure 6.3.

6.2.1.3 Routers

Routers—in their function of forwarding data, as shown in Figure 6.2—implement only the lower three protocol layers. However, routers also implement the higher layer for routing-table maintenance and other management purposes.

As illustrated in Figure 6.2, the IP runs on all nodes and the TCP runs only on source and destination end points. Each hop on a path can use a different link-layer and physical layer technology. Several other IPs and applications are also used to provide routing-path discovery, path selection, and error recovery services.

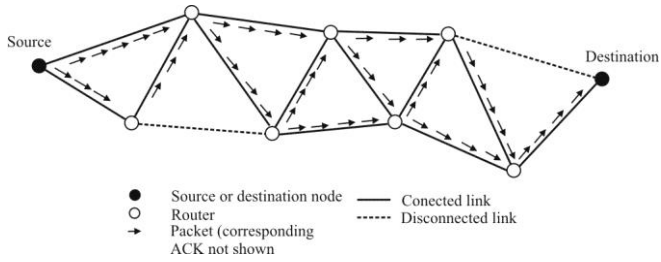


Figure 6.3 Internet routing.

6.2.1.4 Encapsulation

The term *packet* is applied to the objects actually sent over the physical links of a network. They are often called IP packets because the IP—the only protocol used by all nodes on the path—is primarily responsible for directing them, node-by-node, from source to destination along their entire path.

Packets consist of a hierarchy of data-object encapsulations that are performed by the protocol layers. Higher-level data and its header are enclosed (encapsulated) in a lower-layer data object, which is given its own header. The headers are used by their respective protocol layers at nodes along the link to control the processing of the encapsulated data. Successive headers are added at the source as user data moves down the layer structure from source application to physical layer. Headers are removed at the destination end as data moves up the layer structure to the destination application.

TCP breaks user data into pieces called *segments*. IP encapsulates the TCP segments into *datagrams*, and it may break the segments into pieces called *fragments*. The link-layer protocol encapsulates IP datagrams into *frames*. The physical layer then transmits and receives a sequence of frames as a continuous bit stream.

6.2.1.5 Conversational Protocols

The TCP is said to be *conversational* (interactive), because a complete one-way message involves many source-to-destination signaling round trips:

- *Setup.* A three-way "hello" handshake.
- *Segment Transfer and Acknowledgement.* Each TCP segment (or a few segments) sent by the source is acknowledged by the destination.
- *Take Down.* A four-way "goodbye" handshake.

See Figure 6.4.

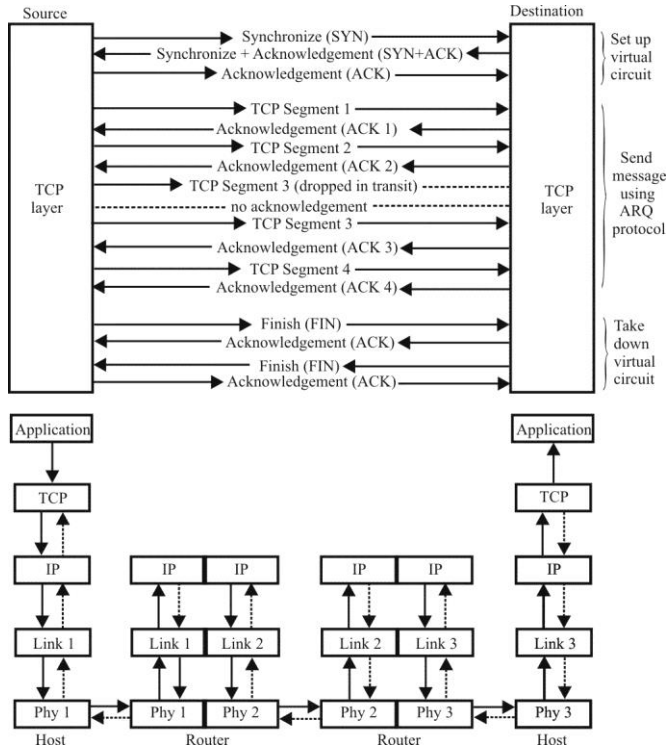


Figure 6.4 The Internet uses a conversational protocol. The complete setup, messaging, and take-down processes are shown here.

The use of positive or negative acknowledgements to control retransmission of lost or corrupt segments is called an *automatic repeat request* (ARQ) protocol.

6.2.2 Mobile Computer Networks

As mentioned, there are currently two variations of mobile wireless networks. The first is known as infrastructured networks, that is, those networks with fixed and wired gateways. The bridges for these networks are known as base stations. A mobile unit within these networks connects to, and communicates with, the nearest base station that is within its communication radius. As the mobile travels out of range of one base station and into the range of another, a "handoff" occurs from the old base station to the new, and the mobile is able to continue communication seamlessly throughout the network. Typical applications of this type of network include wireless local area networks (WLANs, e.g., WiFi).

The second type of mobile wireless network is the infrastructureless mobile network, commonly known as an ad hoc network. Infrastructureless networks have no fixed routers; all nodes are capable of movement and can be connected dynamically in an arbitrary manner. Nodes of these networks function as routers which discover and maintain routes to other nodes in the network. Example applications of ad hoc networks are mobile military data networks, emergency search-and-rescue operations, meetings or conventions in which persons wish to quickly share information, and data acquisition operations in inhospitable terrains. The remainder of this chapter is devoted to the analysis of the first of these applications as that is the application into which forward-edge, tactical EW systems fall.

6.2.3 Evolving Wireless Networks Outside the Internet

Communication outside of the Internet—where power-limited mobile wireless, satellite, and interplanetary communications are developing—is accomplished on independent networks, each supporting specialized communication requirements. These networks do not use Internet protocols and they are mutually incompatible—each is good at passing messages within its network, but not able to exchange messages between networks.

Each network is adapted to a particular communication region, in which communication characteristics are relatively homogeneous. The boundaries between regions are defined by such things as link delay, link connectivity, data-rate asymmetry, error rates, addressing and reliability mechanisms, *quality of service* (QoS) provisions, and trust boundaries. Unlike the Internet, these wireless networks support long and variable delays, arbitrarily long periods of link disconnection, high error rates, and large bidirectional data-rate asymmetries.

Examples of wireless networks outside of the Internet include:

- Terrain civilian networks connecting mobile wireless devices, including personal communicators, intelligent highways, and remote Earth outputs;
- Wireless military battlefield networks connecting troops, aircraft, satellites, and sensors (on land or on water);
- Outer-space networks, such as the Interplanetary (IPN) Internet project, described at http://www.ipnsig.org.

We are particularly interested in the second of these.

Spanning two network regions requires the intervention of an agent that can translate between incompatible networks characteristics and act as a buffer for mismatched network delays.

6.3 Mobile Ad Hoc Networks

6.3.1 Ad Hoc Networks Versus Mobile Ad Hoc Networks

Ad hoc networks form spontaneously without a need of an infrastructure or centralized controller. This type of peer-to-peer system infers that each node, or user, in the network can act as a data endpoint or intermediate repeater. Thus, all users work together to improve the reliability of network communications. These types of networks are also popularly known as "mesh networks" because the topology of network communications resembles a mesh.

The fault tolerance for the network is significantly improved by the redundant communication paths provided by ad hoc mesh networks. Additionally, the ability for data packets to move from one user to another effectively extends the network coverage area and provides a solution to overcome *nonline-of-sight* (NLOS) issues.

Mobile applications present challenges for mesh networks as changes to the network topology are swift and widespread. Such scenarios require the use of MANET technology to ensure communication routes are maintained quickly and accurately. MANETs are self-forming, self-maintained, and self-healing, allowing for extreme network flexibility.

While MANETs can be completely self-contained with no outside interface, they can also be tied to an IP-based global or local network (e.g., Internet or private networks) where they are referred to as hybrid MANETs.

A MANET is a self-configuring network of mobile routers (and associated hosts) connected by wireless links—the collection of which forms a random topology. The data sources, destinations, and routers are free to move at will and organize themselves at random; thus, the network's wireless topology may change rapidly and unpredictably. Minimal static configuration and quick deployment make ad hoc networks highly suitable for tactical-edge data communications in military conflicts.

6.3.2 History of MANETs

The earliest MANETs were called packet radio networks, and were developed by DARPA in the early 1970s. BBN Technologies and SRI designed, built, and experimented with these earliest systems. It is interesting to note that these early packet radio systems predated the Internet, and indeed were part of the motivation of the original Internet protocol suite. Later DARPA experiments included the Survivable Radio Network (SURAN) project, which took place in the 1980s. Another third wave of academic activity started in the mid-1990s with the advent of inexpensive 802.11 radio cards for personal computers, which are currently included in virtually all laptop computers. Current MANETs are designed primary for military utility, of which JTRS is an example.

The popular IEEE 802.11 ("Wi-Fi") wireless protocol incorporates an ad hoc networking capability when no wireless access points are present, although it is considered a very low-grade ad hoc protocol. The IEEE 802.11 ad hoc functionality only handles traffic within a local "cloud" of wireless devices. Each node transmits and receives data, but does not route anything between the network's systems. However, higher-level protocols can be used to aggregate various 802.11 ad hoc networks into MANETs.

6.3.3 MANET Layers

In order to evaluate the specific technologies that enable MANETs, it is useful to functionally decompose a MANET into the first four layers of the OSI model—specifically, the hardware and software that implement the physical layer (layer 1) and the hardware and software that implement the datalink (layer 2), network (layer 3), and transport (layer 4) layers. The physical layer is the actual physical manifestation of the communications bit stream. For MANETs, the bit stream can consist of RF signals or photons (in free-space lasers). In order to simplify the discussion, we will collectively refer to the physical layer implementations as radios, and the datalink, network, and transport layer implementations as the network.

6.3.4 Routing Protocols for MANETs

In order to facilitate communication within the network, a routing protocol is used to discover and maintain routes between nodes. The primary goal of such an ad hoc network routing protocol is correct and efficient route establishment between a pair of nodes so that messages may be delivered in a timely manner. Route construction should be done with a minimum of overhead (time) and bandwidth (frequency spectrum) consumption. An ad hoc routing protocol is a convention or standard that controls how nodes come to agree which way to route packets between computing devices in a MANET. In ad hoc networks, nodes do not have a

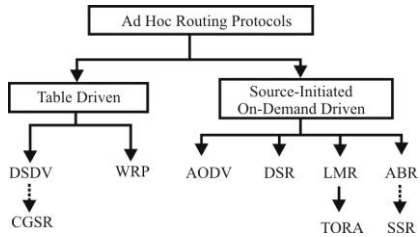


Figure 6.5 Ad hoc routing protocols:
destination-sequenced distance-vector (DSDV); lightweight mobile routing (LMR);
cluster-head gateway switch routing (CGSR); temporally ordered routing algorithm (TORA);
wireless routing protocol (WRP); associatively based routing (ABR);
ad hoc on-demand distance vector (AODV); signal stability routing (SSR);
dynamic source routing (DSR).

priori knowledge of topology of network around them; they have to discover it. The basic idea is that a new node announces its presence and listens to broadcast announcements from its neighbors. The node learns about new near nodes and ways to reach them, and announces that it can also reach those nodes. As time goes on, each node knows about all other nodes and one or more ways how to reach them.

Routing algorithms have to:

- Keep routing tables reasonably small;
- Choose the best route for given destination (this can be the fastest, most reliable, highest throughput, cheapest, or most secure route);
- Keep tables up-to-date when nodes die, move, or join;
- Require small amount of messages/time to converge.

Figure 6.5 lists several of the protocol that have been developed.

Since the advent of DARPA packet radio networks in the early 1970s, numerous protocols have been developed for mobile ad hoc networks. Such protocols must deal with the typical limitations of these networks, which include high power consumption, low bandwidth, and high error rates. As shown in Figure 6.6, these routing protocols may generally be categorized as table-driven, also known as *proactive protocols* and source-initiated on-demand driven, also known as *reactive protocols*. Solid lines in this figure represent direct descendants while dotted lines depict logical descendants.

6.3.4.1 Table-Driven Routing Protocols

The table-driven routing protocols attempt to maintain consistent, up-to-date routing information from each node to every other node in the network. These protocols require each node to maintain one or more tables to store routing

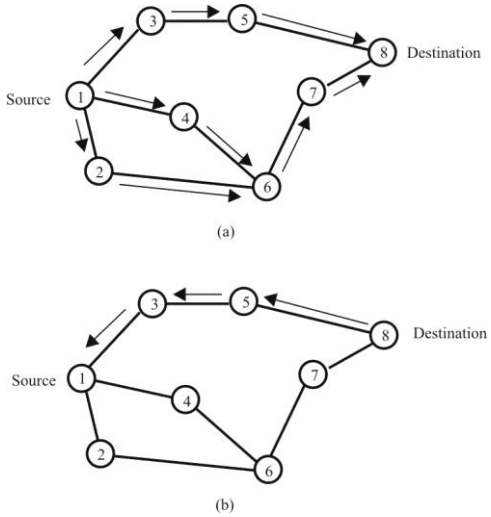


Figure 6.6 AODV route discovery: (a) propagation of RREQ and (b) path of RREP to source.

information, and they respond to changes in network topology by propagating updates throughout the network in order to maintain a consistent network view. The areas where they differ are the number of necessary routing-related tables and the methods by which changes in network structure are broadcast.

Destination-Sequenced Distance-Vector Routing

An example of a table-driven protocol is the *destination-sequenced distance-vector* (DSDV) routing protocol. All mobile nodes in the network maintain a routing table in which all of the possible destinations within the network and the number of hops to each destination are kept. Routing table updates are periodically transmitted throughout the network in order to maintain table consistency and currency. To help ease the potentially large amount of network traffic that such updates can generate, route updates employ two possible types of packets. The first is known as a "full dump" that carries all available routing information and can require multiple *network protocol data units* (NPDUs). During periods of occasional movement, these packets are transmitted infrequently. Smaller "incremental" packets are used to relay only that information that has changed since the last full dump.

New route broadcasts contain the address of the destination, the number of hops to reach the destination, the sequence number of the information received regarding the destination, as well as a new sequence number unique to the broadcast. The route labeled with the most recent sequence number is always used.

In the event that two updates have the same sequence number, the route with the smaller metric is used in order to optimize the path.

6.3.4.2 Source-Initiated On-Demand Routing

A different approach from table-driven routing is *source-initiated on-demand routing* which creates routes only when desired by the source node. When a node requires a route to a destination, it initiates a route discovery process within the network. This process is completed once a route is found or all possible route permutations have been examined. Once a route has been established, it is maintained by a route maintenance procedure until either the destination becomes inaccessible along every path from the source or until the route is no longer required.

Ad Hoc On-Demand Distance Vector Routing

The *ad hoc on-demand distance vector* (AODV) routing protocol builds on the DSDV algorithm described above (see Figure 6.5). AODV minimizes the number of required broadcasts by creating routes on an on-demand basis, as opposed to maintaining a complete list of routes as in the DSDV algorithm. Nodes that are not on a selected path do not maintain routing information or participate in routing table exchanges.

When a source node is to send a message to some destination node and does not already have a valid route to that destination in its route table, it initiates a *path discovery process* to locate the other node. It broadcasts a *route request* (RREQ) packet to its neighbors, which then forward the request to their neighbors, and so on, until either the destination or an intermediate node with a "fresh enough" route to the destination is located.

During the process of forwarding the RREQ, intermediate nodes record in their route tables the address of the neighbor from which the first copy of the broadcast packet is received, thereby establishing a reverse path. If additional copies of the same RREQ are later received, these packets are discarded. Once the RREQ reaches the destination or an intermediate node with a fresh enough route, the destination/intermediate node responds by unicasting a *route reply* (RREP) packet back to the neighbor from which it first received the RREQ. As the RREP is routed back along the reverse path, nodes along this path set up forward route entries in their route tables which point to the node from which the RREP came.

Routes are maintained as follows. When a source node moves, it is able to reinitiate the route discovery protocol to find a new route to the destination. If a node along the route moves, its upstream neighbor notices the move and propagates a link failure notification message to each of its active upstream neighbors to inform them of the erasure of that part of the route. These nodes, in

Table 6.1 Overall Comparison of On-Demand Versus Table-Driven Routing Protocols

Parameter	On-Demand	Table Driven
Availability of routing information	Available when needed	Always available regardless of need
Routing philosophy	Flat	Mostly flat
Periodic route updates	Not required	Required
Coping with mobility	Use localized route discovery	Inform other nodes to achieve a consistent routing table
Signaling traffic generated	Grows with increasing mobility of active routes	Greater than that of on-demand routing
Quality of service support	Few can support QoS, although most support shortest path	Mainly shortest path as the QoS metric

Source: [2].

turn propagate the link failure notification to their upstream neighbors, and so on until the source node is reached.

6.3.4.3 Table-Driven Versus On-Demand Routing

The table-driven ad hoc routing approach has no regard as to when and how frequently routes are desired. It relies on an underlying routing table update mechanism that involves the constant propagation of routing information. This is not the case, however, for on-demand routing protocols. When a node using an on-demand protocol desires a route to a new destination, it must wait until such a route can be discovered. However, because routing information is constantly propagated and maintained in table-driven routing protocols, a route to every other node in the ad hoc network is always available, regardless of whether or not it is needed. This feature incurs substantial signaling traffic and power consumption at all nodes in the network. Since both bandwidth and battery power are scarce resources in mobile computers, this is a serious limitation. Table 6.1 [2] lists some of the basic differences between the two classes of algorithms.

6.3.4.4 Energy Constraints

From the above discussion it should be clear that there is considerable overhead in the operation of MANETs. Significant time is spent just maintaining routing information, over and above the time spent sending useful SA information. This time equates to energy consumption at all the nodes. This is particularly a problem with small, battery powered devices. The participants in MANETs are essentially on and/or transmitting continuously. Transmitters on data links tend to consume substantial energy, not to mention the considerable power consumption of the nodes computing the network configuration information [the power consumption of some semiconductor technologies (CMOS, for example) increases as the square

of the operating frequency], and significant continuous computations are required at each node.

6.4 MANET Security

The characteristics of MANETs presented in the previous section clearly make a case for building multilevel security approaches that achieve both broad protection and desirable network performance. First, the security approach should spread across many individual components and rely on their collective protection power to secure the entire network. This is because the security scheme adopted by each device has to work within its own resource limitations in terms of computation capability, memory, communication capacity, and energy supply. Second, the security approach adopted should span different layers of the protocol stack, with each layer contributing to a line of defense. No single layer is capable of thwarting all potential attacks. Third, threats from both outsiders who launch attacks on the wireless channel and network topology, and insiders who sneak into the system through compromised devices and gain access to certain system knowledge need to be considered. Fourth, all three components of prevention, detection, and reaction that work in concert to guard the system from collapse should be included. Last, the approach should be practical and affordable in a highly dynamic and resource constrained networking scenario.

6.4.1 Security Issues

Early research efforts on MANETs assumed a friendly and cooperative environment and focused on problems such as wireless channel access and multihop routing. However, security became a primary concern in order to provide protected communication between nodes in a potentially hostile environment. The unique characteristics of MANETs present a new set of nontrivial challenges to security design. These challenges include open network architecture, shared wireless medium, stringent resource constraints, and highly dynamic network topology.

The goal of the security for MANETs is to provide the security services of authentication, confidentiality, integrity, anonymity, and availability, to mobile users. In order to achieve this goal, complete protection spanning the entire protocol stack is required. Table 6.2 [3] describes the major security issues in each layer.

Multihop connectivity is provided in MANETs through two steps: (1) ensuring one-hop connectivity through link-layer protocols [e.g., wireless *medium access control* (MAC)], and (2) extending connectivity to multiple hops through network layer routing and data forwarding protocols (e.g., ad hoc routing).

Table 6.2 Security Approaches for MANETs Should Provide Complete Protection Spanning the Entire Protocol Stack

Layer	Security Issues
Application	Detecting and preventing viruses, worms, malicious codes, and application abuses
Transport	Authenticating and securing end-to-end communications through data encryption
Network	Protecting the ad hoc routing and forwarding protocols
Link	Protecting the wireless MAC protocol and providing link-layer support
Physical	Preventing signal jamming denial-of-service attacks

Source: [3].

Unlike wired networks that have dedicated routers, each mobile node in an ad hoc network functions as a router and forward packets for other peer nodes. The wireless channel is accessible to both legitimate network users and malicious attackers. The existing ad hoc routing protocols, such as AODV [2] and wireless MAC protocols, such as 802.11, typically assume a trusted and cooperative environment. As a result, a malicious attacker can readily become a router and disrupt network operations by intentionally disobeying the protocol specifications.

The two approaches to protecting MANETs are proactive and reactive. The proactive approach attempts to prevent an attacker from launching attacks in the first place, typically through various cryptographic techniques. In contrast, the reactive approach seeks to detect security threats a posteriori and react accordingly. A complete security approach for MANETs should integrate both approaches and encompass all three components: prevention, detection, and reaction.

The prevention component deters the attacker by increasing the difficulty of penetrating the system. However, the history of security has shown that a completely intrusion-free system is infeasible. This is especially true in MANETs, consisting of mobile devices that are subjected to compromise or physical capture. Therefore, the detection and reaction components that discover the occasional successful intrusions and take reactions to avoid persistent adverse effects are required.

The prevention component is mainly achieved by secure ad hoc routing protocols that prevent the attacker from installing incorrect routing states at other nodes. This is usually accomplished with the message authentication primitives described below. The detection component discovers ongoing attacks through examination of abnormal behavior exhibited by malicious nodes. Such misbehavior is detected either in an end-to-end manner, or by the neighboring nodes through overhearing the channel and reaching collaborative consensus. Once an attacker node is detected, the reaction component makes adjustments in routing and forwarding operations, ranging from avoiding the node in route selection to collectively excluding the node from the network.

Security is never free. When more security features are introduced into the network, in addition to the enhanced security strength is the ever-increasing

Table 6.3 The Components in the Multifence Security Approach

Network-layer security approaches				
Secure ad hoc routing Proactive protection through message authentication primitives			Secure packet forwarding Reactive protection through detection and reaction	
Source routing	Link state routing	Distance vector routing	Misbehavior detection	Misbehavior reaction
Link-layer security approaches				
Secure wireless MAC Reactive protection through detection and reaction		Next-generation WEP Modification to existing protocol to fix the cryptologic loopholes		

Source: [3].

computation, communication, and management overhead. Consequently, network performance, in terms of scalability, service availability, robustness, and so on becomes an important concern in a resource-constrained ad hoc network. Both dimensions of security strength and network performance are equally important, and achieving a good trade-off between two extremes is one fundamental challenge in security design for MANETs.

6.4.2 A Multilevel Security Approach

Because multihop connectivity is provided in MANETs through distributed protocols in both the network and link layers, the ultimate multilevel security approach naturally spans both layers, as illustrated in Table 6.3 [3].

6.4.2.1 Network-Layer Security

The network-layer security designs for MANETs protect the network functionality to deliver packets between mobile nodes through multihop ad hoc forwarding. Therefore, they seek to ensure that the routing message exchanged between nodes is consistent with the protocol specification, and the packet forwarding behavior of each node is consistent with its routing states. Accordingly, the existing proposals can be classified into two categories: *secure ad hoc routing protocols* and *secure packet forwarding protocols*. Before we describe these security approaches in detail, we first introduce several cryptographic primitives for message authentication.

Message Authentication Primitives

There are three cryptographic primitives widely used to authenticate the content of messages exchanged among nodes (not unique to MANETs).

HMAC (*Message Authentication Codes*)[1]

If two nodes share a secret symmetric key K, they can efficiently generate and verify a message authenticator $hK(\cdot)$ using a cryptographic one-way hash function h. The computation is very efficient, even affordable for low-end devices. However, an HMAC can be verified only by the intended receiver, making it unappealing for broadcast messaging. In addition, establishing the secret key between any two nodes is a nontrivial problem. If the pairwise shared key is used, a total of $N(N-1)/2$ keys will be maintained in a network with N nodes.

Digital Signature

Digital signature is based on asymmetric key cryptography (e.g., RSA), which involves much more computation overhead in signing/decrypting and verifying/encrypting operations. Each node needs to keep a *certificate revocation list* (CRL) of revoked certificates. However, a digital signature can be verified by any node given that it knows the public key of the signing node. This makes digital signature scalable to large numbers of receivers. Only a total number of N public/private key pairs need be maintained in a network of N nodes.

One-Way HMAC Key Chain

Many cryptographic one-way functions exist such that given the output $f(x)$, it is computationally infeasible to find the input x. By applying $f(\cdot)$ repeatedly on an initial input x, we can obtain a chain of outputs $f_i(x)$. These outputs can be used in the reverse order of generation to authenticate messages: a message with an HMAC using $f_i(x)$ as the key is proven to be authentic when the sender reveals $f_{i-1}(x)$.

Secure Ad Hoc Routing

The secure ad hoc routing protocols take the proactive approach and enhance the existing ad hoc routing protocols, such as AODV, with security extensions. In these protocols, each mobile node proactively signs its routing messages using the cryptographic authentication primitives described above. This way, collaborative nodes can efficiently authenticate the legitimate traffic and differentiate the unauthenticated packets from outsider attackers. However, an authenticated node may have been compromised and controlled by the attacker. Therefore, we have to

[1] In networking literature, MAC normally refers to the medium access control protocol at the link layer. To avoid ambiguity, we use MAC to refer to link-layer medium access control and HMAC to refer to keyed hashing for message authentication.

further ensure proper compliance with the routing protocols even for an authenticated node.

Secure Packet Forwarding

The protection of routing message exchange is only part of the network-layer security for MANET networking. It is possible for a malicious node to correctly participate in the route discovery phase but fail to correctly forward data packets. The security approach should ensure that each node forwards packets according to its routing table. This is typically achieved by the reactive approach because attacks on packet forwarding cannot be prevented: an attacker may simply drop all packets passing through it, even though the packets are carefully signed. At the heart of the reactive solutions are a detection technique and a reaction scheme that can effectively be used.

Detection

Because the wireless channel is open, each node can perform localized detection by overhearing ongoing transmissions and evaluating the behavior of its neighbors. However, its accuracy is limited by a number of factors such as channel error, interference, and mobility. A malicious node may also abuse the security and intentionally accuse legitimate nodes as being malicious. In order to address such issues, the detection results at individual nodes can be integrated and refined in a distributed manner to achieve consensus among a group of nodes as one possibility.

An alternative detection approach relies on explicit acknowledgment from the destination and/or intermediate nodes to the source so that the source can figure out where the packet was dropped.

Reaction

Once a malicious node is detected, certain actions are triggered to protect the network from future attacks launched by this node. The reaction component typically is related to the prevention component in the overall security system. For example, the malicious node may have its certificate revoked, or be chosen with smaller probability in future forwarding paths. Based on their scope, the reaction schemes can be categorized as global reaction and end-host reaction. In the former scheme, all nodes in the network react to a malicious node as a whole. In other words, the malicious node is excluded from the network. However, in the end-host reaction scheme, each node may make its own decision on how to react to a malicious node (e.g., putting this node in its own blacklist or adjusting the confidentiality weight of this node).

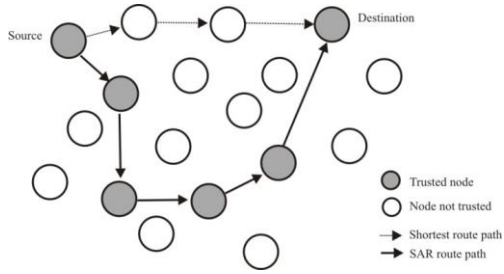


Figure 6.7 Variation of shortest path route selection between SAR and other routing algorithms.

6.4.3 Trusted Node Routing

If a mechanism is in place to determine which nodes can be trusted, then a scheme for using only those nodes can be devised (see Figure 6.7). Instead of using the shortest path from source to destination that most ad hoc routing protocols use, the trusted node path may be, and probably is, longer.

6.5 EW Attacks on MANETs

Traditional attacks on MANETs are fundamentally no different from attacks on other forms of communication. Short of physical destruction, which would probably include physical destruction of the host platform as well, these attacks consist of some form of soft kill, also known as jamming, attacks. Included in this group are [4]:

- Noise jamming [*broadband noise jamming* (BBN)];
- Tone jamming (single tone);
- Multitone jamming;
- Barrage jamming;
- *Partial band noise* (PBN) jamming.

Traditional attacks attempt, in several different ways, to raise the noise floor in the communication channel to reduce the channel capacity of the communication network. We will discuss in detail the most common form of attack: that of raising the BBN thermal noise background at the network receivers. The other forms produce similar performance.

Nontraditional attacks are those that are unique to the particular characteristics of MANETs and, in particular, their various protocols.

6.5.1 Traditional Attacks/Channel Capacity for MANETs

As mentioned, we consider only BBN jamming here; other interference and noise sources are not considered [4]. In addition real terrain is not considered. We assume that the Earth is round and smooth. In order to consider real terrain, either the actual communication paths must be known or a statistical averaging of likely terrain properties must be used. In the scenarios usually considered for tactical communication links, it is this latter approach that is typically used, since the actual terrain is not usually known in advance since the mobile nodes can move.

Many MANET links are DSSS (in particular, CDMA) and therefore share bandwidth. DSSS signals are noise-like in character and therefore are noise sources to one another. We do not consider these noise sources here. Such effects are considered elsewhere [5].

Cellular phone CDMA systems require a base station for synchronization. A MANET does not have one of these. Technological solution for synchronization of mobile CDMA nodes without a base station requires timing sources that are very precise (such as GPS). Such solutions for inexpensive, low-end devices may be cost prohibitive. This is a particular problem when some of the nodes cannot "see" each other. We assume that the mobile CDMA nodes are synchronized.

The biggest real-world problem with MANET networks is the overhead involved with connectivity; that is, the amount of energy it takes to maintain connectivity. Most of the energy is involved with this connectivity, a particular problem with battery-powered nodes where available energy is severely limited.

In order to evaluate the performance of MANETs in realistic environments, we present representative examples in this section. We examine air-to-air and ground-to-ground mobile networks.

6.5.1.1 Assumptions

For this analysis, we assume that the antennas of the mobile nodes are isotropic (radiate and receive equally well in all directions). The data rate (R_b) = 2 Mbps, and the operating frequency f = 1 GHz. The link bandwidth is W = 100 MHz and the receiver NF = 10 dB. These parameters yield the processing gain = W/R_b = $100\times10^6/2\times10^6$, or 17 dB where W is the bandwidth of the DSSS signal.

6.5.1.2 Receiver Noise Floor

While noise can emanate from many sources, when looking purely at the receiver, the noise is dependent upon a number of elements. The first is the minimum equivalent input noise for the receiver. This can be calculated from:

$$P = k_B TW \tag{6.1}$$

where:

P is the power in watts
k_B is Boltzmann's constant (1.38×10^{-23} J/K)
W is the bandwidth in hertz
T is the temperature in Kelvins

It is then possible to calculate the thermal noise floor, where $W = 1$ Hz:

$$N_{0,\text{dBmperHz}} = -174 \text{ dBm/Hz} \tag{6.2}$$

Representing the receiver noise figure (in decibels) as NF_{dB}, the receiver noise floor is given by

$$\text{NoiseFloor} = N_{0,\text{dBmperHz}} + 10\log_{10} W + \text{NF}_{\text{dB}} \quad \text{dBm} \tag{6.3}$$

6.5.1.3 Receive Antenna Effective Area

The effective area of an isotropic antenna is given by [6]

$$A_{\text{eff}} = \lambda^2 / 4\pi \qquad \text{m}^2 \tag{6.4}$$

where λ is the wavelength given by

$$\lambda = c / f \qquad \text{m} \tag{6.5}$$

and where c is the speed of propagation, assumed to be the speed of light.

6.5.1.4 Received Power

The power density radiated from an isotropic antenna at a distance d from the antenna is given by

$$P_{\text{den}} = \frac{S_{\text{t}}}{2\pi d^{\alpha}} \qquad \text{watts/m}^2 \tag{6.6}$$

where α is the path loss coefficient. The resultant power received out of the receive antenna is given by

$$P_{\text{R}} = P_{\text{den}} A_{\text{eff}} \qquad \text{watts} \tag{6.7}$$

Combining (6.3) (numerically, not decibels) with (6.7) yields the received SNR as

$$\text{RxSNR} = \frac{P_R}{N} \tag{6.8}$$

where

$$N = 10^{\text{NoiseFloor}/10} \times 10^{-3} \quad \text{watts} \tag{6.9}$$

6.5.1.5 Air-to-Air Channel

Air-to-air channel propagation characteristics are essentially those of free space as long as there are no significant objects between the transmitter and receiver. Free-space propagation is characterized by having a path loss coefficient, $\alpha = 2$.

No Jammer

The channel capacity for the air-to-air network when there is no jammer present is given by

$$C = W \log_2 (1 + \text{RxSNR}) \quad \text{bps} \tag{6.10}$$

The results are illustrated in Figure 6.8. For this case S_t = 1W and 2W. As we see for S_t = 1W, the distance achieved for a rate of 2 Mbps is about 100 km while that for S_t = 2W, is about 150 km. These high results are a result of the ideal channel assumption ($\alpha = 2$) and 100 MHz bandwidth.

A MATHCAD program to compute the above is given in Figure 6.9.

With Jammer

With a broadband jammer present ($W_{\text{jammer}} = W$), the jammer noise adds to the thermal noise at the receiver. The transmitted signal power remains the same; thus, there is a decrease in the received SNR.

If J_0 denotes the jammer noise PSD, then the jammer power at the receiver is given by

$$J = J_0 W_{\text{jammer}} \tag{6.11}$$

Because we are assuming DSSS modulation on the communication signal, the receiver *processing gain* (P_g) decreases the effective jamming power at the

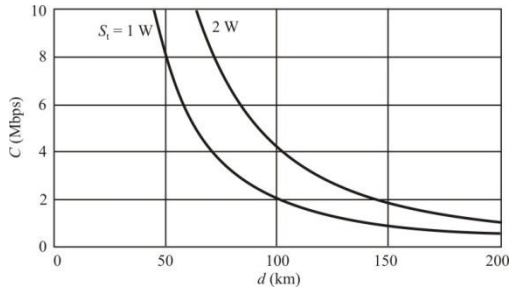


Figure 6.8 Capacity for air-to-air channel with no jammer.

$f := 10^9$ $d := 1, 2 .. 200$

$\lambda := 3 \cdot \frac{10^8}{f}$ $DataRate := 2 \cdot 10^6$ $D(d) := d \cdot 10^3$

$S_t := 1$ $A_{eff} := \frac{\lambda^2}{4\pi}$

$\alpha := 2$

$W := 100\,10^6$ $NF_{dB} := 10$

$N0_{dBmperHz} := -174$

$P_{den}(d) := \frac{S_t}{4\pi\, D(d)^{\alpha}}$ $\frac{watts}{m^2}$ $Noise_{dBm} := N0_{dBmperHz} + 10\log(W) + NF_{dB}$

$P_R(d) := P_{den}(d) \cdot A_{eff}$ watts $N := 10^{\frac{Noise_{dBm}}{10}}$ milliwatts

$RxSNR(d) := \frac{P_R(d)}{N \cdot 10^{-3}}$

$C(d) := W \cdot \log(1 + RxSNR(d), 2)$

Figure 6.9 Air-to-air no-jammer program.

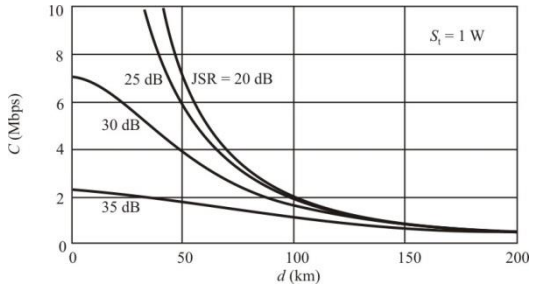


Figure 6.10 Capacity of air-to-air channel with jamming.

receiver by the amount of P_g being used. Then, since we assume that the DSSS signal and the jamming noise are uncorrelated, then

$$P_g = \frac{W}{R_b} \tag{6.12}$$

The effective noise at the receiver becomes

$$N = N_{\text{NoJam}} + J / P_g \qquad \text{watts} \tag{6.13}$$

yielding

$$\text{RxSNR} = \frac{P_R}{N} = \frac{P_R}{N_{\text{NoJam}} + J / P_g} \tag{6.14}$$

We can rewrite (6.14) as

$$\text{RxSNR} = \frac{1}{1/\text{RxSNR}_{\text{NoJam}} + \text{JSR} / P_g} \tag{6.15}$$

The rest of the analysis for the channel capacity from above remains the same, yielding

$$C = W \log_2(1 + \text{RxSNR}) \quad \text{bps} \tag{6.16}$$

Calculated results of (6.16) when $\alpha = 2$ are shown in Figure 6.10 when S_t = 1 W. We can see that at JSR = 20 dB, there is essentially no effect. The processing gain (P_g = 17 dB) must be significantly overcome before the jamming has much affect. Even at 30 dB, the distance for the required channel capacity has only

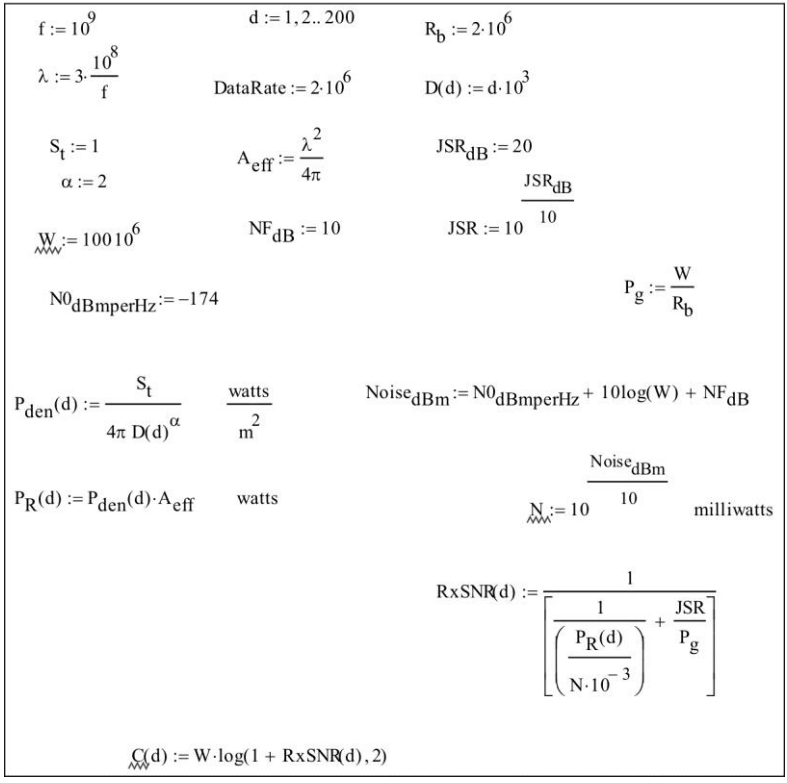


Figure 6.11 Air-to-air with jamming program.

decreased by about 15 km or so. At JSR = 35 dB (a factor of 2 higher than P_g), the distance for the required channel capacity has decreased to about 35 km.

A MATHCAD program for calculating the effects of broadband jamming on air-to-air communications is given in Figure 6.11.

6.5.1.6 Ground-to-Ground Channel

For ground-to-ground communications the attenuation is much higher—free-space conditions do not apply. Attenuation is much higher with distance, and the path-loss coefficient, α, is larger than that for free-space communications. Furthermore, fading is prevalent in ground communications. This fading is a result of multipath reflections that cause the transmitted signal to take several paths to get to the receiver. Each of these paths has different amplitude and phase change effects on the signal that traverse over them, resulting in varying degrees of supporting and

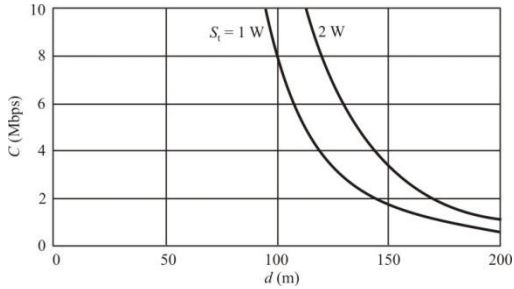


Figure 6.12 Capacity of ground-to-ground with no jammer.

opposing interactions of the signals at the receiver. These amplitude and phase characteristics can even vary with time, making it a nonstationary problem. We will assume that the statistics remain constant over the time interval of interest, however.

In addition, the propagation constant comes into play for ground communications.

Using the same considerations that led to (6.8), we get for the channel capacity

$$\text{RxSNR} = \frac{K \times P_{\text{R}}}{N} \tag{6.17}$$

where K is a constant incorporating the effects of the propagation constant and also taking the fade margin into account. Thus,

$$K = K_{\text{prop}} \times \text{Margin}_{\text{fade}} \tag{6.18}$$

where

K_{prop} = propagation constant
$\text{Margin}_{\text{fade}}$ = fade margin being considered

No Jammer

Performance results are depicted in Figure 6.12 for ground-to-ground communications and for the following parameters

$K_{\text{prop}} = 0.01$
$\text{Margin}_{\text{fade}} = 6$ dB
$\alpha = 4$

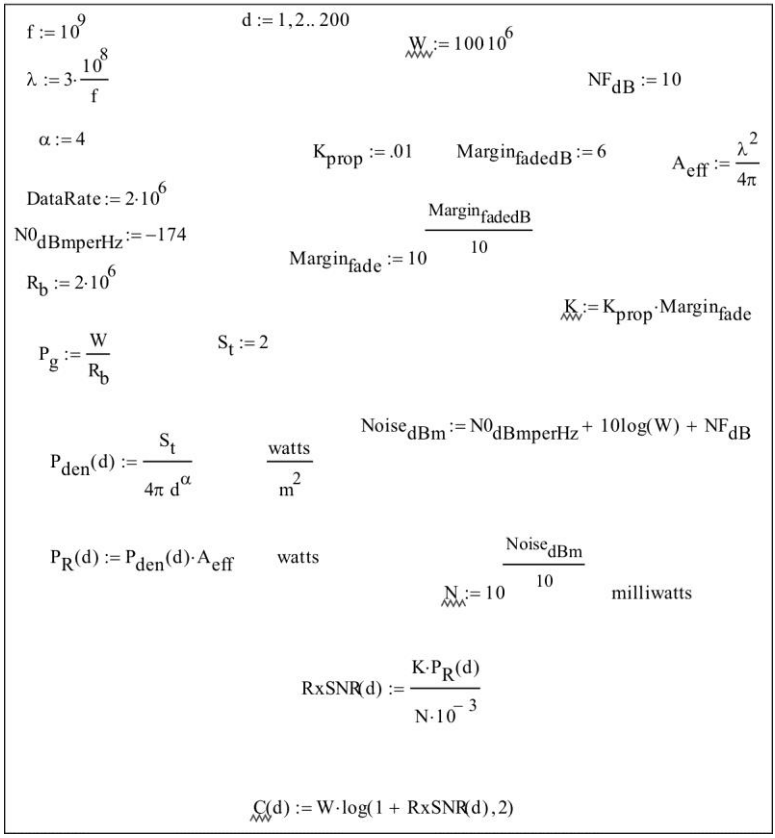


Figure 6.13 Ground-to-ground with no jammer program.

All other factors are the same as previously specified.

From Figure 6.11 we can see that when S_t = 1W, the 2 Mbps required data rate (channel capacity) is met at about d = 145m, and when S_t = 2W, it was met at about d = 170m. These values are substantially less than for air-to-air, and are reflective of the higher path-loss coefficient ($\alpha = 4$) for ground-based communication networks (which also applies to intercept distances). It is also indicative of how close the nodes in a ground-based MANET system must be in order to effectively communicate.

A MATHCAD program for calculating the channel capacity versus distance is shown in Figure 6.13.

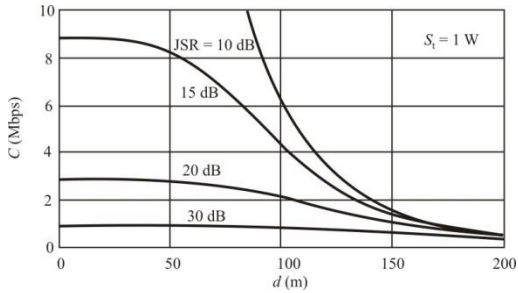


Figure 6.14 Capacity of ground-to-ground channel with jammer.

With Jammer

Following the same philosophy as above to the ground situation, we get

$$\text{RxSNR} = \frac{K \times P_{\text{R}}}{N} \tag{6.19}$$

where now N has the jammer noise signal imbedded so that

$$\text{RxSNR} = \frac{K}{1/\text{RxSNR}_{\text{NoJam}} + \text{JSR}/P_g} \tag{6.20}$$

and

$$C = W \log_2(1 + \text{RxSNR}) \tag{6.21}$$

Results are plotted in Figure 6.14 for when $\alpha = 4$, $K_{\text{prop}} = 0.01$, and $\text{Margin}_{\text{fading}} = 6$ dB. We can see that at JSR = 10 dB, there is no jammer effect on the distance to achieve the required channel capacity. At JSR = 15 dB, the distance decreases slightly to about 130m. At JSR = 20 dB, the distance drops to about 100m and at JSR = 30 dB, the capacity requirement is not met at any distance for S_t = 1W. Again, these results indicate how close the nodes in a ground-based MANET system must be in order to maintain a reasonable channel capacity, especially against a dedicated jamming threat. It is also indicative of how effective an EA program against an adversary depending on MANET communications can be.

A MATHCAD program for calculating the channel capacity is given in Figure 6.15.

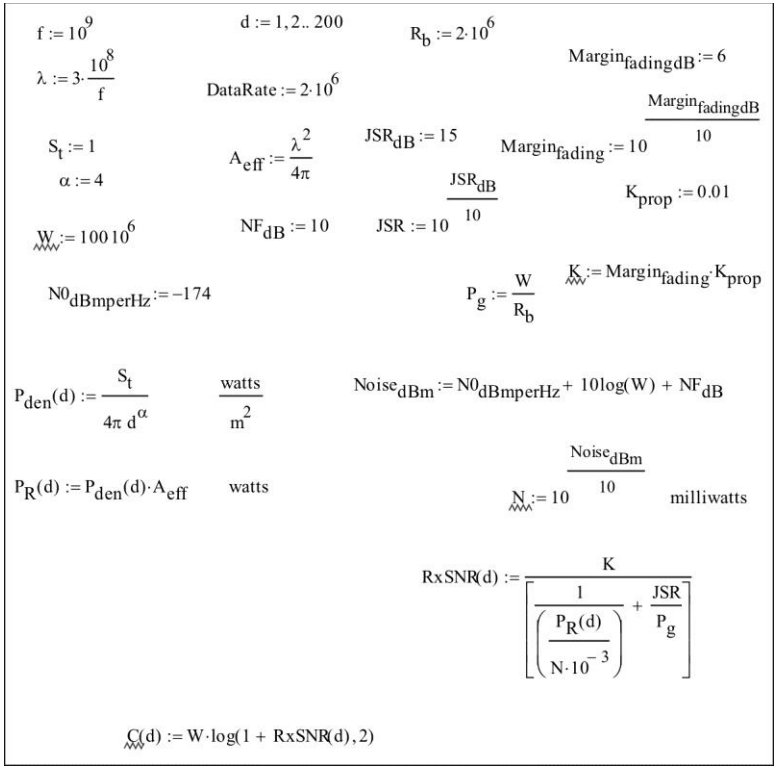


Figure 6.15 Ground-to-ground with jammer program.

As previously mentioned, providing adequate security for MANET systems is an absolute must, and it is very difficult to do. We discuss the common methods of attacking MANETs and methods to thwart such attacks in this section.

6.5.2 Nontraditional MANET Attacks

In this section, we define a taxonomy of types of attackers and discuss specific attacks against ad hoc network routing. This approach allows us to categorize the security of an ad hoc network routing protocol based on the strongest attacker it withstands.

6.5.2.1 Attacker Model

There are two main attacker classes, *passive* and *active*, that are commonly considered. The passive attacker does not send messages; it only eavesdrops on the

network. Passive attackers are mainly threats against the privacy of communication, rather than against the functioning of the network or its routing protocols; as such we do not discuss them further here.

An active attacker generally eavesdrops and injects packets into the network. An attacker can be characterized based on the number of nodes that it owns in the network and based on the number of those that are good nodes it has compromised. We assume that the attacker owns all the cryptographic key information of compromised nodes and distributes it among all its nodes. We denote such an attacker Active-*n*-*m*, where *n* is the number of nodes it has compromised and *m* is the number of nodes it owns. An attacker hierarchy (with increasing strength) is a way to measure routing protocol security: Active-0-1 (the attacker owns one node), Active-0-*x* (the attacker owns *x* nodes), Active-1-*x* (the attacker owns one compromised node and distributes the cryptographic keys to its $x - 1$ other nodes), and Active-*y*-*x*. An attacker that owns all nodes on a vertex cut through the network that partitions the good nodes into multiple sets is called an Active-VC attacker. An attacker that has compromised nodes is also called an Active-VC attacker if it can force good nodes in different partitions to communicate only through an attacker node. This attacker is particularly powerful, as it controls all traffic between nodes of the disjoint partitions.

A protocol may require a trusted *key distribution center* (KDC) in the network, and some ad hoc networks may use one for key setup, as mentioned in Section 6.3. We do not consider the case in which an attacker compromises the KDC, since the KDC is a central trust entity, and a compromised KDC compromises the entire network.

6.5.2.2 Attacks on Ad Hoc Network Routing Protocols

A MANET provides network connectivity between mobile nodes over potentially multihop wireless channels mainly through link-layer protocols that ensure one-hop connectivity, and network-layer protocols that extend the connectivity to multiple hops. These distributed protocols typically assume that all nodes are cooperative in the coordination process. This assumption is unfortunately not true in a hostile environment. Because cooperation is assumed but not enforced in MANETs, malicious attackers can easily disrupt network operations by violating protocol specifications.

Attacks on ad hoc network routing protocols generally fall into one of two categories: routing disruption attacks and resource consumption attacks. In a routing disruption attack, the attacker attempts to cause legitimate data packets to be routed in dysfunctional ways. In a resource consumption attack, the attacker injects packets into the network in an attempt to consume valuable network resources such as time and bandwidth, or to consume node resources such as memory (storage) or computation power. From the application layer perspective (the user), both attacks are instances of a *denial-of-service* (DoS) attack.

An example of a routing disruption attack is for an attacker to send forged routing packets to create a routing loop, causing packets to traverse nodes in a cycle without reaching their destinations, consuming energy, time, and available bandwidth. An example of a resource consumption attack is for an attacker to inject extra data packets into the network, which will consume bandwidth resources when forwarded, especially over detours or routing loops. Similarly, an attacker can inject extra control packets into the network, which may consume even more bandwidth or computational resources as other nodes process and forward such packets. An example of a DoS attack is where the attacker sends a single packet that results in a packet flood throughout the network.

The family of routing attacks refers to any action of advertising routing updates that does not follow the specifications of the routing protocol. The specific attack behaviors are related to the routing protocol used by the MANET. For example, when distance-vector routing protocols such as AODV [7] are used, the attacker may advertise a route with a smaller distance metric than its actual distance to the destination, or advertise routing updates with a large sequence number and invalidate all the routing updates from other nodes. By attacking the routing protocols, the attackers can attract traffic toward certain destinations in the nodes under their control, and cause the packets to be forwarded along a route that is not optimal or even may be nonexistent.

In addition to routing attacks, an adversary may launch attacks against packet forwarding operations as well. Such attacks do not disrupt the routing protocol and poison the routing states at each node. Instead, they cause the data packets to be delivered in a way that is intentionally inconsistent with the routing states. For example, an attacker along an established route may drop the packets, modify the content of the packets, or duplicate the packets it has already forwarded.

6.5.3 MANET Security Challenges

The fundamental vulnerability of MANETs comes from their open peer-to-peer architecture. Unlike wired networks that have dedicated routers, each mobile node in an ad hoc network may function as a router and forward packets for other nodes. The wireless channel is accessible to both legitimate network users and malicious attackers.

Portable devices, as well as the system security information they store, are vulnerable to compromises or physical capture, especially low-end devices with weak protection. Attackers may sneak into the network through these subverted nodes, which pose the weakest link.

The stringent resource constraints in MANETs constitute another nontrivial challenge to security design. The wireless channel is bandwidth-constrained and shared among multiple networking entities. The computation capability of a mobile node is also constrained. For example, some low-end devices, such as PDAs, can hardly perform computation-intensive tasks like asymmetric

cryptographic computation. Because mobile devices are often powered by batteries, they typically have very limited energy resources.

The wireless medium and node mobility pose far more dynamics in MANETs compared to the wireline networks. The network topology is highly dynamic as nodes frequently join or leave the network and roam in the network. The wireless channel is also subject to interferences and errors, exhibiting volatile characteristics in terms of bandwidth and delay.

6.6 MANETs and EW Systems

Networking EW systems brings many advantages. It facilitates geolocation of assigned targets by triangulation or other means (over intranetworks). It allows for collaboration among ES operators when and if required. Communicating OTM allows the EW systems to keep pace with the supported forces.

Internetworking EW systems is not a new concept. EW systems have been networked both with each other as well as with command and control systems ever since datalinks have been on the battlefield.

In this section we outline some of the characteristics of networks as they are applied to the operation of EW systems.

6.6.1 Command and Control

Just as every other OPFAC in the battlespace, EW systems must be managed and integrated into the battle plan. They can be a force multiplier but only if used correctly. C2 is the method of this management.

Typically the mission is planned and assigned prior to its execution. However, plans can change as a mission is executed and these changes must be communicated to the tactical systems, especially if the changes affect the operation of the systems.

When the targets change, if they do during execution, be they EA or ES targets, the tasking for the EW systems will change.

Such C2 can, of course, be in the form of analog communications and often is. Reliability of communications however, is higher when the C2 can be in digital form. This prevents misunderstandings that can easily occur in voice communications, especially over narrowband channels. When time allows, C2 should be exercised digitally.

6.6.2 Reporting

The results of ES activities are normally put into standard report form and passed to an analysis center, or, more recently, immediately posted to the network if appropriate conditions are met (post then analyze).

6.6.3 Target Tasking/Dynamic Retasking

In many cases, dynamic retasking of EW systems is required. This dynamic target tasking will be different depending on whether the EW system is a thin or thick configuration.

Thin EW systems have very little organic ES capability, relying on separate systems to perform this function. EA tasking for thin EW systems comes from a central location over the network. The ES function for thin systems simply detects energy at a frequency, and perhaps performs some simple additional externals measurements (such as modulation recognition). Tasking for these systems consists of passing a frequency to cover to the EW system.

However, thick EW systems possess considerable organic ES functionality, and their tasking can be at a higher level. The targets in this case can be tasked based on their battlefield function, for example. Such retasking might be to find artillery fire direction centers, identify their affiliation, and their geographical locations.

Dynamic target tasking can be the result of target detections by other sensor systems that need to be verified. Sometimes EW systems can be tasked to do this verification.

6.6.4 On-the-Move Communications

On-the-move (OTM) communications has become a modern requirement as a result of the highly dynamic force movements. The EW systems must keep up with the forces they support. In the past, because of the requirement to erect antennas from ground-based systems, EW systems would have to stop to operate. This, of course, is overcome with aerial EW systems, but these systems are not all-weather so they are not the total answer.

Modern ground-based EW systems must operate OTM. This causes some issues. Communicating digitally OTM is an issue as discussed elsewhere in this chapter. Joining and dropping from digital links dynamically requires the functionality of MANETs or equivalent. Another issue is limited geographical coverage. Erecting antennas is precluded so the EW antenna heights are basically the top of the associated vehicle. This results in limited range. With such limited range, the EW systems must be repositioned frequently, again generating the requirement for communications OTM.

MANETs, by design, have very limited range from one node to the next, while overall reach can be quite large due to the relay functioning of all the nodes. Distances between nodes on the same network are on the order of a few hundred meters maximum. As such, links are established and dis-established dynamically and frequently due to the movement of the nodes. Clearly this dictates OTM communications capability.

6.6.5 Sensor Networks

EW systems are sensor systems and can be used as such interacting with other tactical sensors to detect and verify targets. EW systems detect assigned targets, but typically detection reports from a single sensor are not adequate. Other sensors are frequently used for verification, and targets detected by EW systems are no exception. The sensor network can be tasked directly to execute this verification if appropriate.

It is important to note that MANETs are probably not the answer for intranetworking ES systems. The links used for this purpose would saturate the MANET links quickly as the digital traffic over these paths is typically voluminous.

6.6.6 Location Reporting

Blue force tracking requires the ability of tactical systems, including EW systems, to report their locations to appropriate C2 nodes. This allows for total blue force SA. MANETs provide the path for such reporting.

Blue force SA allows for repositioning sensor assets as required. This repositioning might be required to provide for additional coverage areas or to avoid being overrun.

6.7 Concluding Remarks

After briefly discussing some of the characteristics of networking in general, and in particular, the Internet, the concept of MANETs for use among battlespace entities to implement essential parts of NCW was introduced in this chapter. MANETs are particularly important for edge communications where networks form and reform rapidly and often.

MANET fundamental characteristics were discussed as were the particularly troublesome facets of securing the routing protocols.

Included were issues associated with securing MANETs. To be sure, this is not an easy process. Distribution of the various keys is problematic in itself. Of particular issue is the attacks on the routing protocols because they are transferred unencrypted.

In May 2011, the U. S. DoD approved the use of PKE in a *public key infrastructure* (PKI), for distribution of classified information at the SECRET level and below, for all networks connected to the GIG [8]. Hence, the security strategy using the PKI described here applies.

References

[1] Alberts, D. S., J. J. Garstka, and F. P. Stein, *Network Centric Warfare: Developing and Leveraging Information Superiority*, 2nd ed., DoD C4ISR Cooperative Research Program, Washington, D.C.: CCRP Publications, 1996.

[2] Royer, E. M., and C-K. Toh, "A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks," *IEEE Personal Communications*, April 1999, pp. 46–55.

[3] Yang, H., et al., "Security in Mobile Ad Hoc Networks: Challenges and Solutions," *IEEE Wireless Communications*, February 2004, pp. 38–47.

[4] Poisel, R. A., *Modern Communications Jamming Principles and Techniques*, 2nd ed., Norwood, MA: Artech House, 2011.

[5] Viterbi, A. J., *CDMA: Principles of Spread Spectrum Communication*, New York: Addison-Wesley, 1995, Ch. 6.

[6] Poisel, R. A., *Antenna Systems and Electronic Warfare Applications*, Norwood, MA: Artech House, 2012.

[7] Perkins, C., and E. Royer, "Ad hoc On-Demand Distance Vector Routing," *2nd IEEE Workshop on Mobile Computing Systems and Applications*, 1996.

[8] DoD Instruction 8520.02 May 24, 2011, Public Key Infrastructure (PKI) and Public Key (PK) Enabling, http://www.dtic.mil/whs/directives/corres /pdf/852002p.pdf.

Chapter 7

Situation Assessment

7.1 Introduction

Situation assessment is the *process* of evaluating the current state of affairs within the DM's AOR. In order to perform SA, information is needed. Performing SA leads to *situation awareness*, which is the state of being. SA includes the three stages depicted in Figure 7.1, which is a model of the SA as applied to dynamic decision making proposed by Endsley in 1995 [1]. The first step to being aware (Stage 1) is to perceive the environment (Observation). The second step is to understand what is being observed (Stage 2) (Orientation). The last step (Stage 3) is to project a short time into the future what is likely to happen. Based on this projection and the awareness of the current situation, a decision is made as to what action is appropriate and then that action is taken (Decide, Act). The action likely changes the environment in some way, providing feedback to what is being observed. Then the process is repeated. We note how closely this process follows the OODA loop.

7.2 Situation Awareness and Fusion Levels

Research and experimental studies have shown that situation awareness is the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future (see Figure 7.2 [2]). Figure 7.2 is an attempt to illustrate this relationship between the situational awareness model compared to the levels of fusion and cognitive hierarchy discussed in Chapter 2.

Stage 1 Situational Awareness (*Perception*) is the individual's perception of information. This stage is composed of disaggregate elements of information.

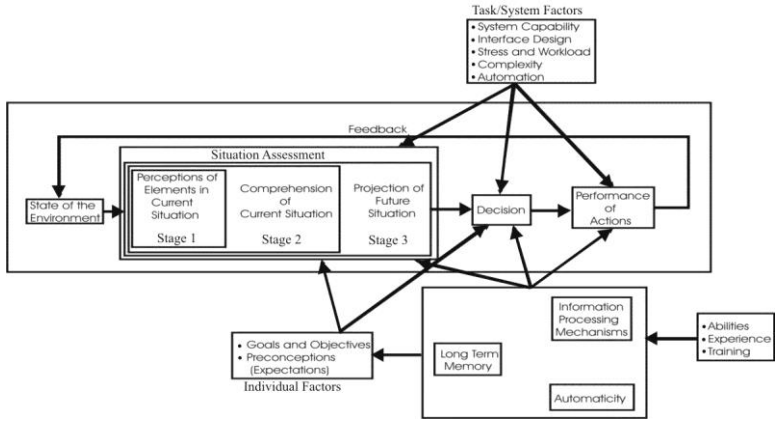


Figure 7.1 Situation assessment model. (Source: [1].)

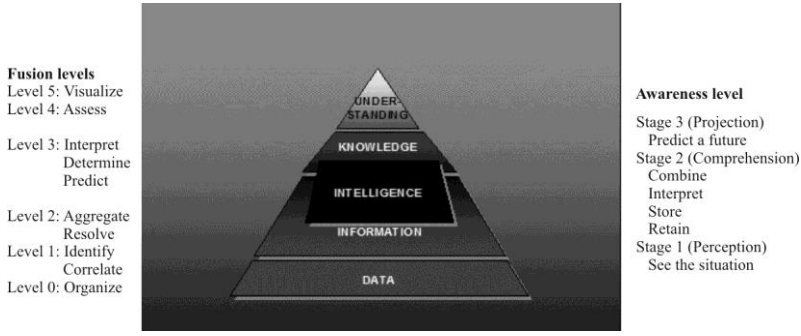


Figure 7.2 Fusion levels compared to awareness levels. (Source: [2].)

The DM sees the situation. Level 0 fusion from humans acting as sensors and level 1 fusion are roughly equivalent to Stage 1 situational awareness in human cognition terms. Stage 1 Situation Awareness is also roughly equivalent to the Observation step in the OODA loop.

Stage 2 Situational Awareness (*Comprehension*) is what results when the individual combines, interprets, and stores the information. It includes integrating and correlating multiple pieces of the information perceived and then determining the relevance of the individual pieces to the person's goals and to achieving the desired end-state. When the DM compares and combines what he or she receives in Stage 1 Situational Awareness against what he or she knows, he or she places and relates what he or she has seen in such a way that he or she comes to understand. Using the cognitive process of "comprehending," he or she now understands what he or she has perceived. Level 2 fusion is roughly equivalent to Stage 2 Situation Awareness in human cognition terms. Stage 2 Situation Awareness is roughly equivalent to the Orientation step in the OODA loop.

Stage 3 *Situational Awareness* (*Projection*) is reached by using Stage 1 and Stage 2 Situational Awareness to project possible future events arising from that situation and to anticipate their outcomes. Individuals project a future as an outcome of their environment in confluence with anticipated events that will impact on their desired end state. Projection is an extension of understanding. We must "know" before we can project the effects of intended actions. Based on what we see and project, we decide on a course of action. Individuals decide using the cognitive process of selecting a best-projected outcome based on what they perceive, how they understand that perception, and what they expect the actions will do toward achieving the desired end state. Level 3 fusion and, perhaps, level 4 fusion are roughly the equivalent of Stage 3 situational awareness in human cognition terms. Stage 3 Situation Awareness is roughly equivalent to the Decide step in the OODA loop.

Factors affecting this SA process are comprised of task or system factors that are outside of the DM's control and individual factors. As indicated in Figure 7.1, the elements of the former are the system capacity, the interface through which the DM interacts with the system, the stress and workload of the DM, the complexity of the system or task, and the amount of inherent automation. The elements of the latter include the goals and objectives of the DM in the current situation and the DM's preconceptions or expectations. Affecting the human factors implications are the long-term memory of the DM, the DM's information processing ability, and the degree of automatic response that the DM possesses (automaticity). These, in turn, are influenced by the DM's abilities, experience, and training.

Although not specifically included in Figure 7.1, significant communication is implied in order to perform SA. While we have indicated that the process is

performed by the DM, in fact only in the simplest of cases is the flow shown in Figure 7.1 performed by a single person at a single location. With the complexity of combat, many people are involved in the process and they must communicate with one another. This provides significant EW opportunities.

7.3 Situation Assessment Strategies

Borden documented an approach to designing SA strategies [3]. We discuss this approach in this section.

"Situation" in the context of warfare consists of the composition, readiness, location, status, and identification (if not by unit name, at least by type) of adversary systems and forces. "Assessment" takes place within a "frame of discernment." A frame of discernment is a set of distinguishable possibilities, one of which is the actual situation. The objective of SA is to make the distribution of probabilities on the frame of discernment asymmetrical so that one possible situation is determined to have a higher probability than the others. A threshold is usually applied so that the possibility selected with the highest probability exceeds it. If the threshold is not reached for any possibility then we defer the selection until new information becomes available. In that way, we begin the process at some level of uncertainty (entropy) and we reduce the uncertainty until we have sufficient confidence to make the assessment. The reduction of uncertainty is information, as we will discuss at length in Chapter 9. The "assessment" is accomplished by the systematic generation and use of information.

7.3.1 Knowledge Acquisition and Database Development

Figure 7.3 illustrates how passive knowledge (a database) is developed, then becomes a knowledge base or associative memory. Independent verification of the meaning of messages (parametric measurements) is required in order to properly bin the measurements. Clustering within the parameter measurement space is required. Hypotheses are generated on which values of different parameters appear to occur together and may be related. With independent verification, rough clusters can be refined and related to distinguishable objects in the frame of discernment.

To develop a relatively complete, usable database two, not necessarily consecutive, steps are needed. The first is a dedicated collection effort focusing on collecting the parameters of interest. The second is the design of a sound database format, based on the intended use of the information. If the intended use is to facilitate smart, structured queries and sorts, the best format is highly relational.

Note that, in Figure 7.3, there is overlap between the clusters of parameter measurements. In many cases, and for one or more reasons, it is impossible to distinguish between objects in the frame of discernment by using the database. Often the reason for this is that not enough parameters are available to make a

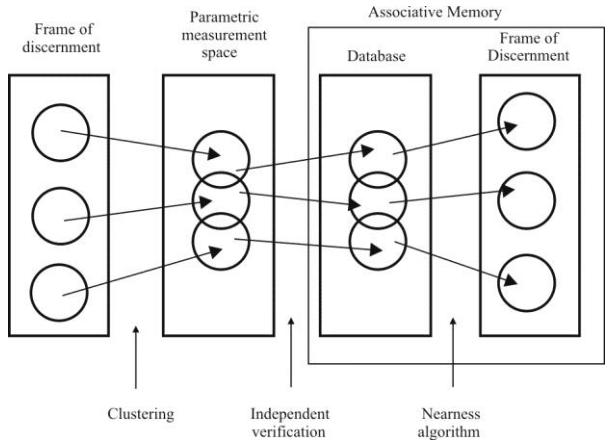


Figure 7.3 Knowledge development.

classification decision with zero uncertainty. This is almost certainly the case when only one parameter measurement is available to evaluate. The overlap can be regarded as noise in the channel through which the parameter was sent.

If we have one or more parameter measurements for an object, how do we use the database to reduce entropy? The answer is that we use some sort of "nearness" function to identify the best candidate classification. There are a number of ways to specify nearness, but the one that is consistent with Shannon's mathematics was developed by Bayes. That is, we compute the conditional probabilities for each candidate classification, given the parameter that we have measured. Suppose the set of objects in the database is given by $\{A_1, A_2, \ldots, A_N\}$ and the parameter we measured is given by x. We generate the conditional probabilities $\left\{\Pr\{A_1|x\}, \Pr\{A_2|x\}, \ldots, \Pr\{A_N|x\}\right.$ and select the A_i with the largest $\Pr\{A_i|x\}$. The candidate with the highest conditional probability is the nearest, so is the best guess based on what we currently know. When we get near enough to one object (when the probability gets high enough, say, it rises above some pre-determined threshold), we select that object.

7.3.2 Development of an Active Memory

Application of the SA strategy is illustrated in Figure 7.4. In Figure 7.4, the parameter messages coming from the environment are shown as dotted lines to indicate that one or more, but not necessarily all, of the parameters may be selected. We compute the nearness function only with the tables in the database that are relevant to the selected parameters.

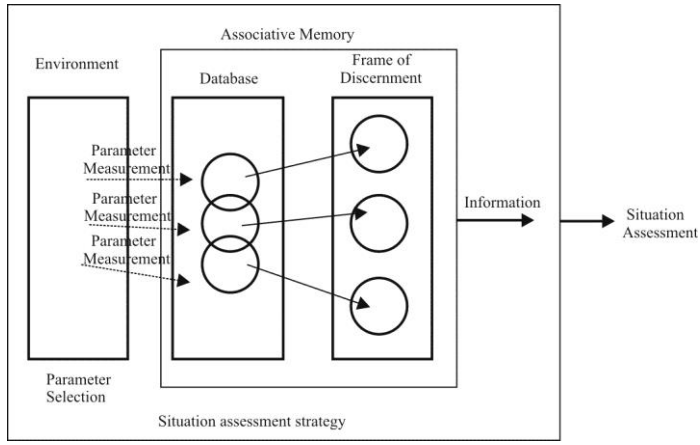


Figure 7.4 Applying the SA strategy.

The database with the nearness function is an active, associative memory. It takes a partial description of an object and finds the nearest classification—the one having the highest probability of being the correct one. As we discussed in Chapter 3, as the probability distribution sharpens (as one object gets very near to our parameters), the entropy becomes less, so information is being produced.

7.3.3 Summary

- Information is created or produced when data in the form of parametric messages from the environment are processed by an active memory.
- The active (associative) memory consists of a database and an algorithm that computes a "nearness" function. Bayes formula is such a typical function.
- The nearness computation reduces entropy and produces information.
- Developing a good SA strategy means selecting parameters whose values will produce the most information (reduce uncertainty) at the least cost.
- This process can be automated.

7.4 Bayesian Logic and Bayesian Belief Networks

Bayes was an English clergyman who lived in the eighteenth century. He invented what has become one of the most popular systems of logic in use today.

7.4.1 Introduction to Bayesian Logic

7.4.1.1 Bayes' Theorem

Bayes' rule was introduced in Chapter 3 as Property 3.2. Bayesian logic systems are a collection of logical properties based upon this rule. Bayes rule is repeated here for convenience as

$$\Pr\{A|B\} = \frac{\Pr\{A,B\}}{\Pr\{B\}} \tag{7.1}$$

where this expression means "the probability of event A happening given that B has happened is…." $\Pr\{A,B\}$ is the probability of both event A and event B occurring while $\Pr\{B\}$ is the probability of event B alone occurring. Rearranging this expression, we get

$$\Pr\{A,B\} = \Pr\{A|B\}\Pr\{B\} \tag{7.2}$$

or, more generally, we get the chain rule

$$\begin{aligned}\Pr\{A_1,A_2,...,A_n\} = \Pr\{A_1|A_2,A_3,\cdots,A_n\}\times\Pr\{A_2|A_3,A_4,\cdots,A_n\}\times \\ \cdots\times\Pr\{A_{n-1}|A_n\}\Pr\{A_n\}\end{aligned} \tag{7.3}$$

which is obtained by repeated application of the above basic formula. Note that these A_is can occur in any order. This joint probability function is a statement about the world which includes these A_is as random variables. This equation then forms a prescription on how to compute this state of the world by considering the conditional probabilities on the right.

Suppose that H_1, H_2, ..., H_n make up a (complete and mutually exclusive) set of hypotheses that can explain the occurrence of an event E. Then

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{\sum_i P(E|H_i)P(H_i)} \tag{7.4}$$

where:

$P(H_i|E)$ = the a posteriori probability of hypothesis H_i, given E has been observed,
$P(H_i)$ = the a priori probability of H_i being true at all,

$P(E|H_i)$ = the probability of observing E given that H_i is true, also called the *likelihood function*. [4]

while

$$\sum_i P(H_i) = 1 \tag{7.5}$$

What these equations say is that the probability of hypothesis H_i being true, given that event E has been observed, is given by the product of the probability of event E happening if hypothesis H_i is true and the probability of hypothesis H_i happening at all, divided by the sum of the products of all of these probabilities.

One of the significant benefits of Bayesian inference, as represented by these equations is that a prior probabilities are used which represent the probability of events occurring at all, out of a general population. Not all logic systems provide this, such as classical inference probabilities.

There are three types of probabilities that are usually discussed:

> *Empirical probability*: This is the probability based on large numbers of occurrences of the events under consideration.
>
> *Classical probability*: Probability based on the law of large numbers.
>
> *Subjective probability*: The probability most often used by people in their everyday life. It is a person's own judgment as to the likelihood of outcomes of events, and therefore it varies with personal opinion.

7.4.2 Modeling Knowledge and Conflict Using Bayes' Reasoning

Moffat developed a process and procedure for modeling knowledge and conflict [5]. He used Bayes' reasoning in the approach so there is a firm mathematical foundation in the development. We describe the salient characteristics of his methods in this section.

We investigate applying Bayes' reasoning to the situation assessment problem by way of an example. This example has all the properties we wish to examine yet is fairly simple.

The situation under consideration is illustrated in Figure 7.5. A friendly commander is faced with the possibility of attack by adversarial forces as shown. The question to be answered (estimated) is how many hostile units will attack. The friendly commander has an airborne ES sensor at his or her disposal. Time is divided into frames, with each frame consisting of one pass through the decision

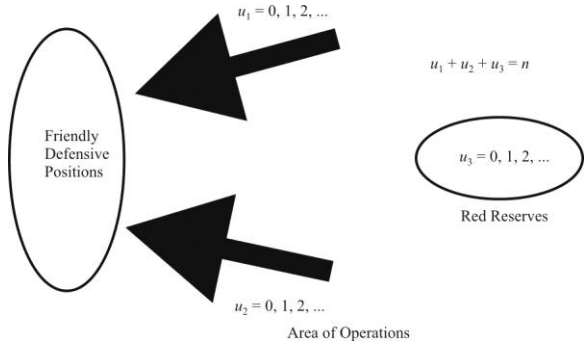


Figure 7.5 Blue commander's situation assessment problem.

(OODA) loop. Initially reports on unit detections are provided by the sensor once per frame.

7.4.2.1 Decision Uncertainty

An ES sensor's primary function is to contribute to tactical situation assessment by observing the battlefield, detecting and identifying enemy units, and reporting on its findings. In the United States, an airborne ES system might be Compass Call or Rivet Joint. For the examples used here, however, these systems would not likely be available. The examples deal with small numbers of battalions, such as might be associated with a Brigade-sized unit. As such, an airborne UAV ES system in direct support of the Brigade is more likely.

An estimate of the degree of confidence that the commanders have that they possesses an accurate picture of the battlefield in his AOR—in other words, the amount of decision uncertainty—is of interest. We would expect that the greater their knowledge about the location, size, and composition of the enemy force, the greater their confidence in making decisions concerning the allocation of his weapons and the movement of their forces. We also recognize that information of this type is not all that they would require. Information concerning enemy movement such as that available from JSTARS and known enemy fighting doctrine would also assist in completing the picture. Force movement pictures such as from JSTARS would be provided from higher echelons as part of the COP.

Let U represent the competing hypotheses that any number of enemy units are arrayed against the friendly commander at time cycle t so that $U = \{0,1,2,...,n\}$. For our purposes a unit is taken to be a battalion. Omitting the cycle index, t, for now allows us to focus instead on analysis within a time step. Figure 7.5 depicts a notional defensive campaign situation.

We assume that the friendly commanders know the number of enemy units that might be brought to bear against them during the campaign. That is, we

assume that they knows *n*. This is a reasonable assumption in that it is highly likely that the IPB would yield this information. What is unknown is the tactical deployment of the units at each time step. Tactical SA then is the process of estimating the enemy's tactical deployment at time *t* and the effectiveness of this estimate is the degree of uncertainty associated with his current state of knowledge.

7.4.2.2 Bayesian Decision Making

Single Sweep

We begin by analyzing the intelligence gathering process at each time step. We first assume that a Bayesian update methodology for tactical SA is appropriate within a frame, but not between frames.[1] Consequently, the process described here is repeated prior to each decision to commit forces.

1. *Input distribution:* The friendly commanders may or may not have some idea of the likely disposition of enemy units. If they do, the corresponding probability distribution can be used. Here however, we assume that the friendly commanders are ignorant of the enemy commander's intentions. This provides us with a worst case situation, corresponding to the assumed lack of memory between time steps (memoryless between time steps). We let $\Pr\{U = u\}$ represent the probability that the enemy commander will commit u of his n units in a specified AOR in the AO (avenue of approach in Figure 7.5). Assuming that the enemy commander is equally likely to deploy any number of units, we have $\Pr\{U = u\} = 1/(n + 1)$.
2. *The sensor model:* Let $V = \{0,1,2,...,n\}$ represent the number of units detected by the sensor[2] [5]. Therefore, $\Pr\{V = v\}$ is the probability that the sensors will detect v of the enemy units arrayed against the friendly forces. This number is conditioned on the number of hostile units deployed in the AOR. Consequently, the probability of interest is the conditional probability, $\Pr\{V = v \mid U = u\}$. We further assume that the sensor is capable of detecting a unit with probability q, and that there are no false detections from the sensor or elsewhere[3] [6]. Consequently,

[1] We later exploit Bayesian updating by assuming multiple sensor sweeps within a single decision cycle.

[2] By "detect," we mean that sufficient information is provided to allow the unit to be targeted by a weapon. Since we have an ES sensor, this means that an enemy unit has been detected broadcasting at a frequency that was either known ahead of time based on SIGINT IPB or found during a general search. Detection in this case means identified and geolocated to sufficient accuracy for weapon application..

[3] It is possible to relax this assumption and allow for the possibility that the sensor detections/identifications are false, that the command and control system used to transmit the sensor

$$\Pr\{V = v \mid U = u\} = \begin{cases} \binom{u}{v} q^v (1-q)^{u-v}, & v \le u \\ 0, & \text{otherwise} \end{cases} \tag{7.6}$$

which is the binomial distribution.

3. *Sensor operations:* The sensor observations are used to clarify the enemy disposition by refining the friendly commander's initial and subsequent probability distributions on U. That is, we wish to calculate $\Pr\{U = u \mid V = v_d\}$, where v_d is the number of detections reported in the cycle. Initially, we assume that the sensor sweeps the AOR once in a cycle. As a detection occurs, it is immediately reported so that there are $v_d + 1$ reports from the sensor per cycle. The additional report accounts for the fact that a report of 0 detections is sent initially. We assume a uniform distribution of reports; that is, a report of no detections occurs at time $t/(v_d + 1)$, a report of one occurs at $2t/(v_d + 1)$, and so forth. The estimate is refined at every subinterval using Bayes' formula as

$$\Pr\{U = u \mid V = v\} = \frac{\Pr\{U = u \mid V = v-1\} \Pr\{V = v \mid U = u\}}{\sum_{i=0}^{n} \Pr\{U = i \mid V = v-1\} \Pr\{V = v \mid U = i\}} \tag{7.7}$$

In (7.7), $\Pr\{U = u \mid V = v-1\}$ is the prior probability, $\Pr\{V = v \mid U = u\}$ is the knowledge contributed by the latest report (the probability that one more unit is detected), and $\Pr\{U = u \mid V = v\}$ is the posterior probability on U given the last report. Note that $\Pr\{U = u\} \mid V = -1\} = \Pr\{U = u\} = 1/(n+1)$; that is, the prior distribution before sensors are deployed is flat, as described above. This process is repeated for $v = 0, 1, \ldots, v_d$. Making the appropriate substitutions in (7.7), we get

$$\Pr\{U = u \mid V = v\} = \frac{\Pr\{U = u \mid V = v-1\} \binom{u}{v} q^v (1-q)^{u-v}}{\sum_{i=0}^{n} \Pr\{U = i \mid V = v-1\} \binom{i}{v} q^v (1-q)^{i-v}}$$

information may report a false detection/identification as real, and that the intelligence processing center may interpret a false detection/identification as real. See [6].

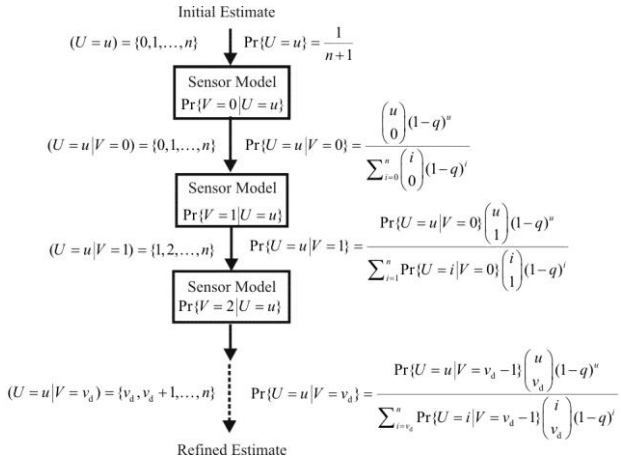


Figure 7.6 Developing refined estimate. (Source: [7].)

$$= \frac{\Pr\{U=u|V=v-1\}\binom{u}{v}(1-q)^u}{\sum_{i=0}^{n}\Pr\{U=i|V=v-1\}\binom{i}{v}(1-q)^i} \tag{7.8}$$

where $v = 0, 1, ..., v_d$ is the number of units detected by the sensor and $u \geq v$ at each iteration. Figure 7.6 [7] depicts the flow diagram of this process. Note the difference between no sensor sweep in progress and a report of no detections. The former is depicted by a flat probability distribution on U, whereas the latter is a refinement to the flat distribution.

> **Example 7.1** [5]: Table 7.1 summarizes the results of a simple situation in which three units are known to be available to the enemy commander. The sensor system has a probability of detection/identification of $q = 0.7$. The entries in the rows are the probabilities resulting from 0, 1, 2, and 3 detections, where the first row is the uniform a priori probability assessment on U. Figure 7.7 depicts the results graphically.

Multiple Sweeps

We now examine the effects of multiple sensor sweeps within the same cycle. We assume that the sensors are capable of k sweeps of the AOR within the

Table 7.1 Refined Probability Assessments: Example 7.1

v	$\Pr\{U=0 \mid V=v\}$	$\Pr\{U=1 \mid V=v\}$	$\Pr\{U=2 \mid V=v\}$	$\Pr\{U=3 \mid V=v\}$
-	0.250	0.250	0.250	0.250
0	0.8013	0.1603	0.0321	0.0064
1	0	0.9218	0.0736	0.0046
2	0	0	0.9634	0.0366
3	0	0	0	1

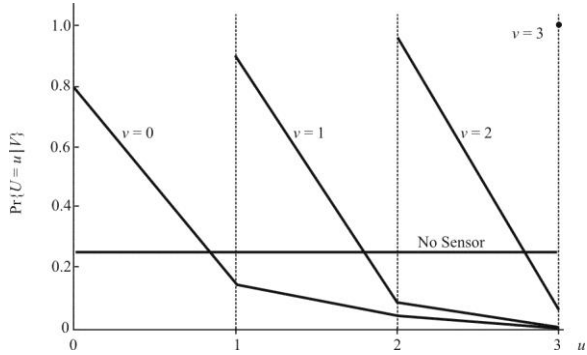


Figure 7.7 Refined probability assessments for Example 7.1.

commander's decision cycle. That is, the sensor can perform k sweeps of the AOR before the enemy commanders can move their units in any significant way. In sweep i, v_{di} enemy units are detected where, $i = 1,2,...,k$. We further assume that the probability estimates are made sequentially, and that the sweep time is sufficiently small to allow for a single "end of sweep" report. Using Bayes' formula, we get

$$\begin{aligned}&\Pr\{U = u \,|\, V = v_{di}\} \\ &= \frac{\Pr\{U = u \,|\, V = v_{d(i-1)}\}\Pr\{V = v_{di} \,|\, U = u\}}{\sum_{j=0}^{n} \Pr\{U = j \,|\, V = v_{d(i-1)}\}\Pr\{V = v_{di} \,|\, U = j\}}\end{aligned} \tag{7.9}$$

$\Pr\{U = u \,|\, V = v_{d0}\) = \Pr(U = u\} = 1/(n+1)$. So

$$\Pr\{U = u \,|\, V = v_{di}\} = \frac{\Pr\{U = u \,|\, V = v_{d(i-1)}\}\binom{u}{v_{di}}(1-q)^{u}}{\sum_{j=v_{di}}^{n} \Pr\{U = j \,|\, V = v_{d(i-1)}\}\binom{j}{v_{di}}(1-q)^{j}} \tag{7.10}$$

Bayesian updating has a tendency to converge rather rapidly—especially in cases such as this where false detections/identifications are not allowed because it is impossible to overstate the number of units actually present. The effect is that subsequent detections that report fewer units than the previous can be ignored.

> **Example 7.2** (continued): Suppose that three sweeps were conducted resulting in three sequential detections using a sensor with probability of detection: $q = 0.7$. Table 7.2[4] summarizes the results of (7.10) with $k = 3$. The number of units detected each time is listed in the second column of the table. The number of units in the AOR is actually three and subsequent observations that two units were detected/identified are ignored.
>
> Now consider a second case with a somewhat different history as depicted in Table 7.3. In this case, four sweeps were conducted resulting in the sequential detections depicted in the

[4] $\Pr\{U = u \,|\, V\}$ in these tables means $\Pr\{U = u \,|\, V = v_{di}\}$.

Table 7.2 Multiple Sweeps Case 1

i	v_{di}	$\Pr\{U=0\|V\}$	$\Pr\{U=1\|V\}$	$\Pr\{U=2\|V\}$	$\Pr\{U=3\|V\}$
0	-	0.250	0.250	0.250	0.250
1	1	-	0.658	0.263	0.079
2	2	-	-	0.637	0.207
3	3	-	-	-	1

Table 7.3 Multiple Sweeps Case 2

i	v_{di}	$\Pr\{U=0\|V\}$	$\Pr\{U=1\|V\}$	$\Pr\{U=2\|V\}$	$\Pr\{U=3\|V\}$
0	-	0.250	0.250	0.250	0.250
1	1	0	0.658	0.263	0.079
2	1	0	0.767	0.122	0.111
3	1	0	0.925	0.059	0.016
4	2	0	0	0.855	0.145

table. The detection of one unit persisted for three reports. Note the rapid convergence of $\Pr\{U=1|V\}$. However, the single detection of two units in sweep 4 shifts the mode of the distribution to U = 2. Because we exclude false detections, all reports less than the current number detected are ignored.

7.4.2.3 Knowledge Representation

We now determine the degree of uncertainty existing in the mind of the DM at the time he must take a force employment decision. There are two components of his current knowledge: (1) a number of enemy units were detected by his sensor suite in his AOR; and (2) the refined pdf over the possible number of enemy units that might be in his AOR based on his most recent sensor report. The first component depends upon whether false detections are possible while the second depends upon the number of enemy units detected and the reliability of the sensor system. We need a knowledge metric that incorporates these two components.

Information Entropy

The information entropy measures the amount of uncertainty in a probability distribution. It is a function of the average information present in the set of all possible uncertain events. The amount of information available from the known occurrence of the event, $U = u$, that is, that u enemy units are arrayed against the friendly force, is inversely proportional to the likelihood that the event will occur. From Chapter 2, information is defined as:

$$\mathcal{I}(U=u) \triangleq \ln \frac{1}{\Pr\{U=u\}} = -\ln \Pr\{U=u\} \tag{7.11}$$

Each of the events in the refined set $\{U|V=v_{\mathrm{d}}\}$, occurs with probability $\Pr\{U=u|V=v_{\mathrm{d}}\}$. Therefore, the information available from the occurrence of each event is

$$\mathcal{I}(U=u|V=v_{\mathrm{d}}) = -\ln \Pr\{U=u|V=v_{\mathrm{d}}\} \tag{7.12}$$

and the expected information from the occurrence of each event is this information quantity multiplied by the probability of the event occurring:

$$\Pr\{U=u|V=v_{\mathrm{d}}\}\,\mathcal{I}(U=u|V=v_{\mathrm{d}}) = -\Pr\{U=u|V=v_{\mathrm{d}}\}\ln \Pr\{U=u|V=v_{\mathrm{d}}\} \tag{7.13}$$

Consequently, the average amount of information in the probability distribution $\Pr\{U|V=v_d)\}$ can be expressed as

$$H[\Pr\{U|V=v_d\}] = H(U|V=v_d)$$
$$= -\sum_{i=0}^{n} \Pr\{U=i|V=v_d\} \ln[\Pr\{U=i|V=v_d\}] \quad (7.14)$$

The entropy $H(U|V=v_d)$ is the *residual uncertainty* regarding *U* given that *V* = v_d. The average uncertainty then is the sum of the residual uncertainties weighted by the probability distribution on the sensor detection/ identifications

$$H(U|V) = -\sum_{j=0}^{n} \Pr\{V=j\} \sum_{i=0}^{n} \Pr\{U=i|V=j\} \ln[\Pr(U=i|V=j)] \quad (7.15)$$

Properties of Information Entropy

Information entropy has several favorable properties for measuring the DM's uncertainty prior to making a decision and for measuring the uncertainty in the entire campaign:

- *Maximum entropy*: The entropy is maximized when the uncertainty in the distribution is greatest. Maximum uncertainty occurs when the friendly commander has no sensor assets to deploy and therefore no knowledge about the number of units that might be arrayed against him. Thus, we have that $\Pr\{U=u\} = 1/(n+1)$. The entropy in this case is

$$H(U) = -\sum_{i=0}^{n} \frac{1}{n+1} \ln \frac{1}{n+1} = \ln(n+1) \quad (7.16)$$

  Thus, the maximum uncertainty in *P(U)* is ln(*n* + 1). With the lack of any sensor reports, we would expect the entropy to go up since we have no idea of the number of units that the adversary has. In general, a probability distribution with a wide variance exhibits high entropy.

- *Minimum entropy*: The entropy function is minimized at 0. This occurs when $P(U = u_i) = 1.0$ and $P(U = u_j) = 0$ for all $j \neq i$. This represents total certainty or minimum uncertainty.

- *Campaign entropy*: The total campaign entropy, denoted $H(U_1,U_2, \ldots U_m)$, where m is the total number of campaign cycles satisfies

$$H(U_1,U_2,\ldots,U_m) \le \sum_{i=1}^{m} H(U_i) \tag{7.17}$$

The equality condition holds when the process is memoryless, when the situation being considered is rapidly changing across the time span of the campaign.

7.4.2.4 Combat Cycle Knowledge

We now develop a measure incorporating the *residual uncertainty* in the refined pdf, and the *detection information* gained by the sensor report. Let $K(U,V = v_d)$ represent the knowledge gained from detecting v_d enemy units when there are U enemy units in the AOR. Thus

$$K(U,V = v_d) = K(U \mid V = v_d)K(V = v_d) \tag{7.18}$$

where $K(U \mid V = v_d)$ is the knowledge associated with the residual uncertainty in the refined probability distribution given a sensor report of v_d units, and $K(V = v_d)$ is the knowledge gained by detecting/identifying v_d enemy units. We can think of $K(U,V = v_d)$ as a probability[5] representing the likelihood that the DM has a complete picture of the battlefield at the time he makes a decision.

1. *Residual Knowledge*: Since the maximum uncertainty in $\Pr\{U \mid V = v_d\}$ is $\ln(n+1)$, we can define maximum certainty as $\ln(n+1) - H(U \mid V = v_d)$[6]. Normalizing this quantity to the maximum uncertainty we get the following definition of residual knowledge

$$K(U \mid V = v_d\} = \frac{\ln(n+1) - H(U \mid V = v_d)}{\ln(n+1)} \tag{7.19}$$

We see that residual knowledge is maximized (= 1) when residual entropy is 0 and it is minimized (= 0) when residual entropy is

[5] $K(U,V = v_d)$ satisfies the probability axioms and therefore can be thought of as a subjective probability.

[6] In general, the change in information resulting from detecting $V = v_d$ units is

$$\Delta I(U \mid V = v_d) = H(U) - H(U \mid V = v_d).$$

$\ln(n + 1)$. Residual knowledge reflects the amount of uncertainty in the refined probability distribution.

2. *Detection knowledge*: Given that v_d enemy units were detected, we are now concerned with the likelihood that there are actually v_d or more enemy units in the AOR. Thus we are interested in the information content for the event: $U \geq v_d \,|\, V = v_d$. That is, the information that will be provided from the detection reports this cycle, or the prior information content of the event, $V = v_d$. This is given by

$$\begin{aligned} \mathcal{I}(U \geq v_d \,|\, V = v_d) &= -\ln\left[\Pr\{U \geq v_d \,|\, V = v_d)\right] \\ &= -\ln\left[\sum_{i=v_d}^{n} \Pr\{U = i \,|\, V = v_d\}\right] \end{aligned} \tag{7.20}$$

If $v_d = 0$, no information is produced because $\Pr\{U \geq 0\} = 1$. However, if $v_d = n$, the information content is maximized at

$$-\ln[\Pr\{u = n \,|\, V = v_d\}]$$

because

$$\Pr\{U \geq u \,|\, V = v_d\}$$

monotonically decreases with increasing u and therefore is smallest for $u = n$. This suggests the following definition for $K(V = v_d)$

$$K(V = v_d) = \frac{\ln\left[\sum_{i=v_d}^{n} \Pr\{U = i \,|\, V = v_d - 1\}\right]}{\ln\left[\Pr\{U = n \,|\, V = v_d - 1\right]} \tag{7.21}$$

(We use $v_d - 1$ to ensure that the denominator never goes to zero.)

The total knowledge gained is then given by the product of residual and detection knowledge, (7.20) and (7.21)

$$K(U, V = v_d) = \frac{\left[\ln(n+1) - H(U \,|\, V = v_d)\right]}{\ln(n+1)} \frac{\ln\left[\sum_{i=v_d}^{n} \Pr\{U = i \,|\, V = v_d - 1\}\right]}{\ln\left[\Pr\{U = n \,|\, V = v_d - 1\right]} \tag{7.22}$$

Table 7.4 Total Knowledge Example 7.3

V	$\Pr\{U=0\|V\}$	$\Pr\{U=1\|V\}$	$\Pr\{U=2\|V\}$	$\Pr\{U=3\|V\}$	$H(U\|V)$	$K(U,V)$
-	0.250	0.250	0.250	0.250	1.3863	0
0	0.8013	0.1603	0.0321	0.0064	0.6130	0
1	0	0.9218	0.0736	0.0046	0.2918	0.1638
2	0	0	0.9634	0.0366	0.1570	0.4434
3	0	0	0	1	0	1

Example 7.3 (continued): The first five columns of Table 7.4 repeat the information in Table 7.1 for convenience. The last two columns contain the entropy and knowledge figures based on the refined distributions at each iteration and the intermediate values of *V*. Figure 7.8 depicts the results graphically.

7.4.3 Bayesian Belief Networks

Bayesian networks are a form of graph that portray the dependencies that hypotheses have on one another in Bayesian logic systems [8]. Note that sometimes a random variable is independent of other random variables in the problem at hand. This is denoted by

$$\Pr\{A_1 \,|\, A_2, A_3, ..., A_n, B\} = \Pr\{A_1 \,|\, A_2, A_3, ..., A_n\} \tag{7.23}$$

where A_1 is independent of the random variables in set *B*.

Each hypothesis in the world model is represented by a node in a directed, acyclic graph. If hypothesis *A* depends on hypothesis *B* then an arc is constructed that connects node *B* to *A* as in Figure 7.9. The weight assigned to this arc is $\Pr\{A|\,B\}$.

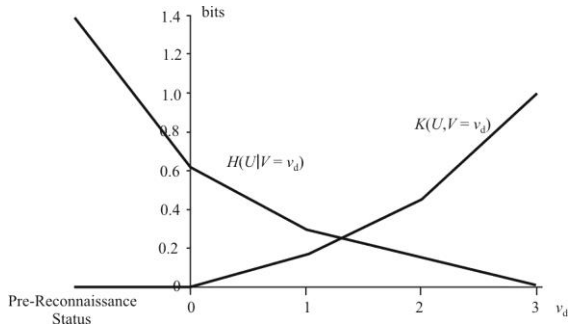


Figure 7.8 Knowledge and entropy for Example 7.3.

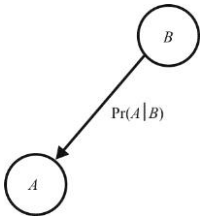


Figure 7.9 When the occurrence of event *A* depends on the occurrence of event *B*, an arc in an acyclic Bayesian belief graph is used to represent this dependency.

One of the often-quoted shortcomings of Bayesian inference is its requirement to know the a priori knowledge of the probabilities, but in many cases these can be approximated with reasonable success. The *principle of indifference* can be applied when the a priori probabilities $P(H_i)$ are not known. This principle assigns equal probabilities to these quantities, because it is implicitly assumed that all events are equally likely, lacking any evidence to the contrary. In fact, the Bayesian system updates the probabilities as the logic proceeds with time—as new evidence enters, the probabilities are recomputed. One of the effects of this is that as time goes by the effects of the a priori probabilities becomes less and less, and eventually can become negligible. Therefore, why not set them all equal?

Example 7.4: Suppose the following four hypotheses have been ascertained relative to an enemy unit being opposed by a blue force as depicted in Figure 7.10:

H_1: the enemy is likely to advance to the left
H_2: the enemy is likely to advance to the right
H_3: the enemy will stay in its current location
H_4: the enemy will retreat

Furthermore, it has been determined that these are the only four possibilities, so the set of hypotheses is complete and mutually exclusive. Intelligence has estimated that the probabilities of each of these hypotheses are given by:

$P(H_1) = 0.35$
$P(H_2) = 0.35$
$P(H_3) = 0.2$
$P(H_4) = 0.1$

That is, the enemy is more likely to attack than to stay put or retreat, but whether it is to the left or right is unknown, and it is

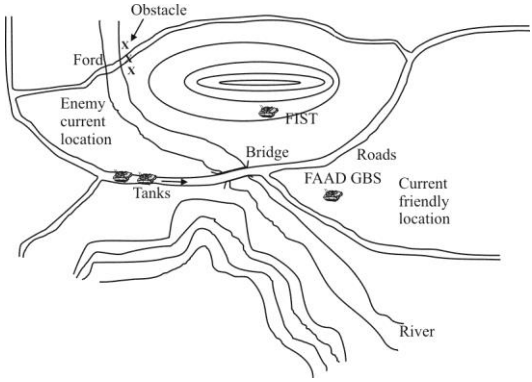


Figure 7.10 Example scenario.

equally likely to be either way. Suppose that a message has been received from a FIST that there is a column of ten tanks moving down a road as indicated in Figure 7.10.

$P(\text{Observe Tanks On Road} \mid H_1) = 0.35$
$P(\text{Observe Tanks On Road} \mid H_2) = 0.5$
$P(\text{Observe Tanks On Road} \mid H_3) = 0.1$
$P(\text{Observe Tanks On Road} \mid H_4) = 0.05$

Observing the tanks on the road as shown supports the conjecture that the enemy will advance to the right is predicated on the notion that if the enemy were to go left, it is more likely that he would go around the left side of the hill, rather than the right side.

Thus the probabilities computed based on Bayesian statistics would be

$$P\left(H_1 \mid \text{Observed Tanks On Road}\right) = \frac{0.35 \times 0.35}{0.35 \times 0.35 \; + \; 0.35 \times 0.5 \; + \; 0.2 \times 0.1 \; + \; 0.1 \times 0.05} = \frac{0.1225}{0.3225} = 0.380$$

$$P\left(H_2 \mid \text{Observe Tanks On Road}\right) = \frac{0.35 \times 0.5}{0.3225} = 0.543$$

$$P\left(H_3 \left| \text{Observe Tanks On Road} \right.\right) = \frac{0.2 \times 0.1}{0.3225} = 0.062$$

$$P\left(H_4 \left| \text{Observe Tanks On Road} \right.\right) = \frac{0.1 \times 0.05}{0.3225} = 0.0155$$

Therefore, the notion of an attack to the right is supported more than any other possibility in this case.

This example points out some of the basic issues with Bayesian logic. First, the set of hypotheses must be complete and mutually exclusive. We assumed these for the example but, if this example were a real situation, there are shades of gray as to whether these hypotheses represent a complete set. Furthermore, part of the enemy force may go right and some may go left, so the hypotheses are not necessarily mutually exclusive.

Another issue is that there is no room for uncertainty. A probability must be assigned to every hypothesis, and there can be no combining of the hypotheses together in an "OR" fashion. There is no room for "I don't know for sure."

These things aside, however, the logic involved can produce useful results—we just accept the risk of not being very mathematically correct in the assumptions.

At each node a truth table is maintained that displays the value of the hypothesis depending on the values of the incoming arcs. For example, if node A has three incoming arcs a_1, a_2, a_3, then the truth table would be as in Table 7.5 where the A_i are either T or F.

These networks are primarily used to calculate two entities: (1) the belief in a hypotheses based on evidence, accumulated over time, for example, and (2) the best explanation for a given belief in a hypothesis.

In general, calculating with these networks is difficult because of the complexities involved. They fit the class of problems called "NP-hard," which means that they cannot be solved in polynomial time, but the solutions expand

Table 7.5

a_1	a_2	a_3	A
F	F	F	A_0
F	F	T	A_1
F	T	F	A_2
F	T	T	A_3
T	F	F	A_4
T	T	F	A_5
T	F	T	A_6
T	T	T	A_7

exponentially as more nodes are added. Certain classes of networks, however, can be solved in reasonable time. One of these classes is "causal polytrees" or "singly connected networks," which means that there are no loops. (Analysis with loops is possible but would only complicate this discussion [8]).

Given some evidence $\vec{e}$, which is an instantiation of one or more nodes in the network, then the goal in belief management is to determine the probabilities at nodes of interest that is the most consistent with this evidence; that is it has the highest probability of being true. It is frequently assumed, without loss of generality, that the evidence is observed only at edge, or leaf, nodes. $\vec{e}$ is frequently expressed as a vector of zeros and ones, representing false and truth for the instantiated nodes. For example, if there are six edge nodes, then one instantiation would be $\vec{e} = (e_1, e_2, e_3, e_4, e_5, e_6)^{\mathrm{T}} = (0,0,1,0,1,1)^{\mathrm{T}}$. This particular instantiation says that e_1 = false, e_2 = false, e_3 = true, e_4 = false, e_5 = true, and e_6 = true.

When the evidence is assigned to a root node (one with only arcs leading away from it), then the evidence is said to be *predictive evidence*—such nodes are frequently referred to as evidence nodes. It is possible to have evidence observed or assigned to output nodes—those nodes with only incoming arcs. In that case, the evidence is called *diagnostic evidence*, and the associated nodes are called anticipatory nodes.

The propagation of the effects of the evidence $\vec{e}$ proceeds into the network using the above chain rule and independence property.

These results, then, allow propagating information throughout the network so that the belief at each hypothesis node can be updated.

Example 7.5: Suppose the last example is extended and modified as shown in Figure 7.11. A *forward area air defense*

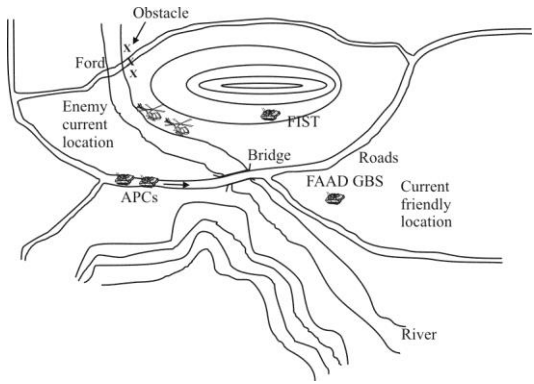


Figure 7.11 Continuation of the previous example.

(FAAD) *ground based sensor* (GBS) is added to the friendly sensor mix, and helicopter scouts are flying for the opposing forces. Obstacles have been added just to the friendly side of the road where the river ford is located. Lastly, the tanks have been changed to APCs, which can be a little less foreboding.

It is assumed that all sensor reports are relayed to the decision-makers without delay and they are understood, with probability 1.

It is important to keep in mind that the view here is that of a friendly analyst. For example, the node in the graph below that refers to "advance to the left" does not reflect whether the enemy actually advances to the left but whether the friendly analyst "believes" that to be the case. The modeling does allow, however, for incorporation of "ground truth" such as this, by assigning a value of 1 to the belief. That is, if in fact the enemy forces moved to the left in this example, the probability associated with that node is set at one and the resultant propagation of that fact through the network occurs, updating the beliefs at each node.

The weather plays a big role in most combat situations in which the military is involved. As noted elsewhere, it determines visibility, transversability, weapon effectiveness, and other things. In this example, a heavy rain might make a river crossing at the ford untenable. A heavy rain or heavy snow, for example, might degrade the performance of all of the RISTA sensors. For this example it will be assumed that all it does is impact on whether a fording is possible allowing an attack on the left or not (H_{10}). Alternately the weather may be such that the bridge is icy and therefore would not safely allow wheeled vehicles to cross yielding a probability of whether the weather supports a bridge crossing (H_{11}). So that this example can proceed, these probabilities are assigned as follows:

H_{10}: Weather allows an attack left:

H_{10}	Pr{H_{10}}
F	0.1
T	0.9

H_{11}: Weather allows an attack right:

H_{11}	Pr{H_{11}}
F	0.1
T	0.9

Hypotheses H_9 refers to whether or not the enemy goals dictate the necessity of advancing at this time. Clearly this is a guess and normally would not be known for sure. Even if information seemed to indicate that this is true, such information could be based on deception activities and cannot be totally trusted. Friendly analysis must reach a conclusion on their best guess that this is the case, however, and this guess would be based on past experience, both of the adversary and friendly commander's experience, the tactical situation, the state of the overall combat situation, and so forth. In addition, since it is not possible normally to know the truth absolutely, human judgment enters and some threshold of belief is necessary in order to declare this belief true or false. We will assume that this threshold is $P = 0.8$ here. In this case it is assumed that the conclusion was that the probability that an attack is included in the enemy's plans is:

H_9: Enemy goals include an advance at this time:

H_9	Pr{H_9}
F	0.2
T	0.8

The mapping of real situations to binary truth tables is performed routinely by humans in their thought process. If a situation seems to be true to the satisfaction of an individual (the subjective probability discussed previously), then a value of true is determined, which normally means more or less true for the case at hand. In the case for this example, the truth conditions might be that belief in the hypothesis associated with sensor assets, both friendly and enemy, H_6, H_7, and H_8, are if the indications that a target detection is made, then if the probability is greater than, say, 0.3, a value of true is declared and if less than that value, it is declared false.

The probability that the scout helicopters will detect the obstacle at the ford would depend on many factors, such as the terrain, whether the weather was such that an obstacle could be seen from where the scouts were flying, and so forth. Here we

will simply apply a probability but noting that in a real circumstance it would be a conditional probability.

H_5: Scout helicopters observed the obstacle at the ford:

H_5	Pr{H_5}
F	0.2
T	0.8

Assume that the probabilities associated with the FIST observing tanks on the road, unconditioned on anything unstated, is given by:

H_7: *Fire Support Team* (FIST) observes APCs on road:

H_7	Pr{H_7}
F	0.1
T	0.9

In the theory of Bayesian nets this is known as evidence and would correspond to a sensor report arriving. The probability is not associated with whether a report arrived or not—that is known with probability one. The probability is a measure of the confidence in the report. Either past history would indicate that the sensor and/or operator was very reliable, or less so, would be some of the parameters associated with this probability.

Similarly assume that the probability of observing the APCs at the ford point is given by:

H_8: FIST observes APCs at ford:

H_8	Pr{H_8}
F	0.3
T	0.7

This probability might be lower because the fording point is further away than the road and there may be more obstructions, such as trees, big rocks, and so forth., so a clear view of the ford is not possible.

H_6: FAAD detects scout helicopters:

H_6	Pr{H_6}
F	0.1
T	0.9

H_{12}: The opposition is attacking

The conditional probability matrix associated with hypothesis 12 is given in Table 7.6.

For the output hypothesis, H_1 through H_4, the following truth tables are assumed to apply:

H_1: Advance to the left (Table 7.7). The bridge route is assumed to be the preferred way, if everything else is equal since vehicles can move faster over the bridge than in water. Therefore, the truth table is biased in this direction. A true value for weather then indicates that the weather does not preclude an advance to the left. In normal circumstances it would be difficult to know a priori if the enemy scouts detect the obstacle or not, so its probability threshold is set at 0.7. An advance is an important event to recognize so a relatively low threshold is put on this variable. Note that "advance" at this point does not indicate left or right.

H_2: Advance to the right (Table 7.8). Similar dialogue applies in this case as above from the slant of this being the preferred route of advance.

H_3: Stay put (Table 7.9). Whether it is believed that the enemy forces are staying in place or not depends on the analysts' views on the meaning of tactical activities—siting of APCs and scouts—and whether the friendly commander believes that an attack is imminent.

H_4: Retreat (Table 7.10). The only condition that would indicate that the enemy might retreat is if there is absolutely no activity detected and the likelihood that an attack is necessary is nil. Then H_4 is given in Table 7.10.

The Bayesian belief network associated with this example is shown in Figure 7.12.

Table 7.6 H_{12} Probabilities

H_7:APCs on Road	H_8:APCs at Ford	H_9:Goals Req Advance	H_6:FAAD Detects Scouts	Pr{H_{12}= T}: Advance
F	F	F	F	0.1
F	F	F	T	0.1
F	F	T	F	0.1
F	F	T	T	0.2
F	T	F	F	0.2
F	T	F	T	0.2
F	T	T	F	0.5
F	T	T	T	0.7
T	F	F	F	0.2
T	F	F	T	0.2
T	F	T	F	0.5
T	F	T	T	0.7
T	T	F	F	0.2
T	T	F	T	0.4
T	T	T	F	0.5
T	T	T	T	0.8

Table 7.7 H_1 Probabilities

H_5:Scouts Detect Obstacle	H_{10}:Attack Weather Left	H_9:Advance	Pr{H_1= T}: Advance to Left
F	F	F	0.1
F	F	T	0.3
F	T	F	0.1
F	T	T	0.6
T	F	F	0.1
T	F	T	0.3
T	T	F	0.2
T	T	T	0.8

Table 7.8 H_2 Probabilities

H_5:Scouts Detect Obstacle	H_{11}:Attack Weather Right	H_9:Advance	Pr{H_2 = T}: Advance to Right
F	F	F	0.1
F	F	T	0.7
F	T	F	0.1
F	T	T	0.7
T	F	F	0.1
T	F	T	0.7
T	T	F	0.1
T	T	T	0.9

Table 7.9 H_3 Probabilities

H_7:APCs on Road	H_8:APCs at Ford	H_9:Enemy Req Adv	H_6:FAAD Det Scouts	Pr{H_3= T} Stay Put
F	F	F	F	0.9
F	F	F	T	0.8
F	F	T	F	0.6
F	F	T	T	0.4
F	T	F	F	0.7
F	T	F	T	0.5
F	T	T	F	0.2
F	T	T	T	0.1
T	F	F	F	0.7
T	F	F	T	0.6
T	F	T	F	0.2
T	F	T	T	0.1
T	T	F	F	0.5
T	T	F	T	0.4
T	T	T	F	0.4
T	T	T	T	0.1

Table 7.10 H_4 Probabilities

H_7:APCs on Road	H_8:APCs at Ford	H_9:Enemy Req Adv	H_6:FAAD Det Scouts	Pr{H_4= T} Retreat
F	F	F	F	0.9
F	F	F	T	0.2
F	F	T	F	0.2
F	F	T	T	0.2
F	T	F	F	0.2
F	T	F	T	0.2
F	T	T	F	0.2
F	T	T	T	0.2
T	F	F	F	0.2
T	F	F	T	0.2
T	F	T	F	0.2
T	F	T	T	0.2
T	T	F	F	0.2
T	T	F	T	0.2
T	T	T	F	0.2
T	T	T	T	0.1

The leaf nodes in this network are H_1, H_2, H_3, H_4, H_5, H_6, H_7, H_8, H_9, H_{10}, and H_{11}. Assuming that the weather is good then in the quiescent state all nodes are at the quiescent value. This corresponds to $\vec{e} = (0,0,1,0,0,0,0,0,0,1,1)^{\mathrm{T}}$.

Using the independence property above, by inspection:

$$\Pr\left\{H_{12}\middle|H_1,H_2,\cdots,H_{11}\right\} = \Pr\left\{H_{12}\middle|H_6,H_7,H_8,H_9\right\}$$

$$\Pr\left\{H_1\middle|H_2,H_3,\cdots H_{12}\right\}$$
$$= \Pr\left\{H_1\middle|H_5,H_{10},H_{12}\right\}\Pr\left\{H_{12}\middle|H_6,H_7,H_8,H_9\right\}$$

$$\Pr\left\{H_2\middle|H_1,H_3,\cdots H_{12}\right\}$$
$$= \Pr\left\{H_2\middle|H_5,H_{11},H_{12}\right\}\Pr\left\{H_{12}\middle|H_6,H_7,H_8,H_9\right\}$$

$$\Pr\left\{H_3\middle|H_1,H_2,\cdots H_{12}\right\}$$
$$= \Pr\left\{H_3\middle|H_6,H_7,H_8,H_9\right\}\Pr\left\{H_{12}\middle|H_6,H_7,H_8,H_9\right\}$$

and

$$\Pr\{H_4\middle|H_1,H_2,\cdots H_{12}\} = \Pr\{H_4\middle|H_6,H_7,H_8,H_9\}$$
$$\times\Pr\{H_{12}\middle|H_6,H_7,H_8,H_9\}$$

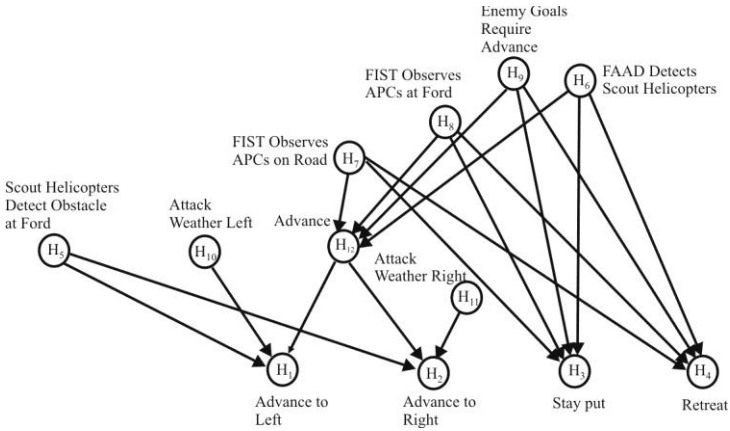


Figure 7.12 Bayes' network associated with the example.

Suppose that a new evidence vector arrives at time = 1 that updates two hypotheses: (1) the friendly commander has determined that it is in the best interest of the enemy commander to attack at this time (e_9 changes from 0 to 1) and (2) the FAAD GBS has detected scout aircraft in the vicinity of the river (e_6 changes from 0 to 1). At this point, there is no evidence available to the analyst that would indicate that the scout helicopter has detected the obstacle. Therefore $\vec{e} = (0,0,1,0,0,1,0,0,1,1,1)^{\mathrm{T}}$.

At time = 2, suppose a message from the FIST is received that indicates APCs have been detected moving forward at the ford in the river (e_8 changes from 0 to 1). There is still no evidence that the obstacle at the ford has been detected, however. The evidence vector thus is $\vec{e} = (0,0,1,0,0,1,0,1,1,1,1)^{\mathrm{T}}$. The resultant changes are shown in Table 7.11. The shorthand P_i is used for $\Pr\{H_i = \text{True}\}$.

Thus, we see that the likelihood of an attack increased with

Table 7.11 Probabilities at $t = 2$

Time	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}
0	0.1/F	0.1/F	0.9/T	0.9/T	0.2/F	0.1/F	0.1/F	0.1/F	0.1/F	0.9/F	0.9/F	0.1/F
1	0.1/F	0.1/F	0.4/F	0.2/F	0.2/F	0.9/F	0.1/F	0.1/F	0.8/F	0.9/F	0.9/F	0.2/F
2	0.6/T	0.7/T	0.1/F	0.2/F	0.2/F	0.9/T	0.1/F	0.7/T	0.8/T	0.9/T	0.9/T	0.7/T

these reports and the logic of the network correctly indicated that fact.

7.5 Concluding Remarks

In this chapter we discussed some of the aspects of SA. The SA model shown in Figure 7.1 follows the OODA loop quite well. It illustrates how SA can be conducted, preparing an action list based on environmental observations. We showed how external factors, as well as individual factors affect the SA process. The relationship of the fusion levels to the assessment levels were illustrated.

We discussed how Bayesian reasoning can be used to model knowledge and conflict and to generate situation awareness in a quantifiable way.

This chapter also discussed how Bayesian logic can be used to generate situation awareness.

References

[1] Endsley, M. R., "A Taxonomy of Situation Awareness Errors," in R. Fuller, N. Johnston, and N. McDonald (eds.), Human Factors in Aviation Operations, pp. 287–292, Aldershot, England: Avebury Aviation, Ashgate Publishing Ltd.

[2] "White Paper: Objective Force Fusion," U. S. Army Intelligence Center, Directorate of Combat Developments, Ft. Huachuca, AZ, March 2003.

[3] Borden, A., "The Design and Evaluation of Situation Assessment Strategies," *Information & Security*, Vol. 1, No. 1, 1998, pp. 63–77.

[4] Waltz, E., and J. Llinas, *Multisensor Data Fusion*, Norwood, MA: Artech House, 1990, p. 242.

[5] Moffat, J., *Complexity Theory and Network Centric Warfare*, Washington D. C.: CCRP Publications, 2003.

[6] Perry, W., and J. Moffat, "Measuring the Effects of Knowledge in Military Campaigns," *Journal of Operational Research Society*, Vol. 48, 1997, pp. 965–972.

[7] Moffat, J., *Complexity Theory and Network Centric Warfare*, Washington D.C.: CCRP Publications, 2003, p. 114.

[8] Pearl, J., *Probabilistic Reasoning in Intelligence Systems: Networks of Plausible Inference*, San Francisco, CA: Morgan Kaufmann, 1987.

Chapter 8

EW Systems

8.1 Introduction

EW system configurations are discussed at length in several sources [1–6]. We need not duplicate that information here. Rather, we include in this chapter a brief synopsis of the essential characteristics of EW systems that is suitable for understanding the performance results presented in the next two chapters.

We recall that an EW system is composed of essentially two subsystems: (1) electronic support subsystem and (2) electronic attack subsystem.

This chapter is structured as follows. We begin with a discussion on notional EW system architectures. Abbreviated descriptions of the major components of ES and EA susbsystems are provided. The chapter is concluded with a discussion of some operational deployment considerations for EW systems.

8.2 EW System Architectures

As mentioned, an EW system is comprised of two fundamental parts: an ES subsystem and an EA subsystem. In this section we briefly discuss some fundamental architectures for these two subsystems.

A generic EW system architecture is shown in Figure 8.1. The two subsystems are delineated as shown. Frequently some of the system equipment is shared among the two principal subsystems—usually the system control, *human-computer interface* (HCI), and communications subsystems.

The system represented in Figure 8.1 assumes that the two principal subsystems are colocated. This is not a necessity. The configuration shown in Figure 8.2 shows the two geographically separated. In fact, the two subsystems need not coordinate activities with each other (a situation that can lead to chaos in

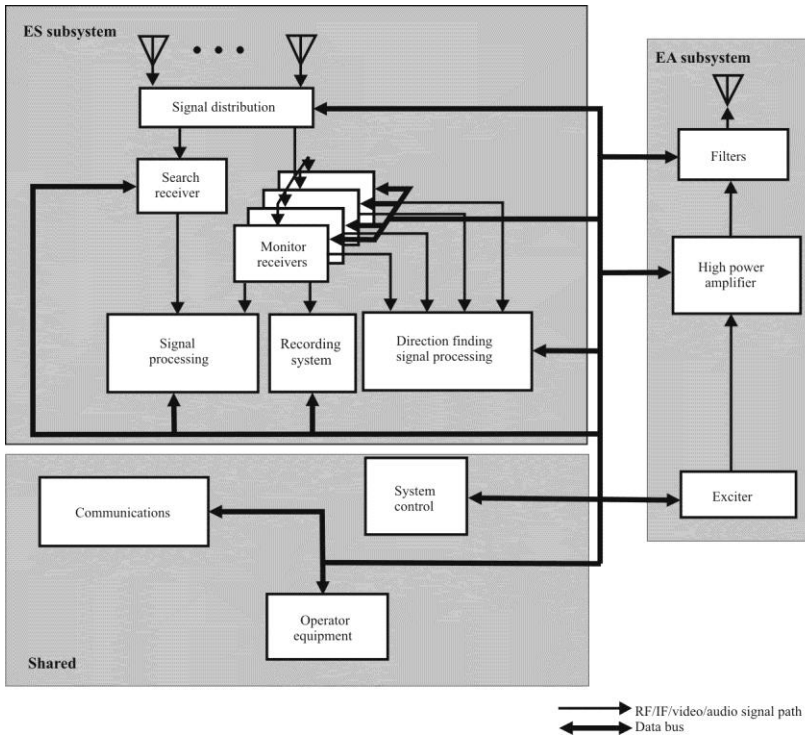


Figure 8.1 Block diagram of a typical EW system.

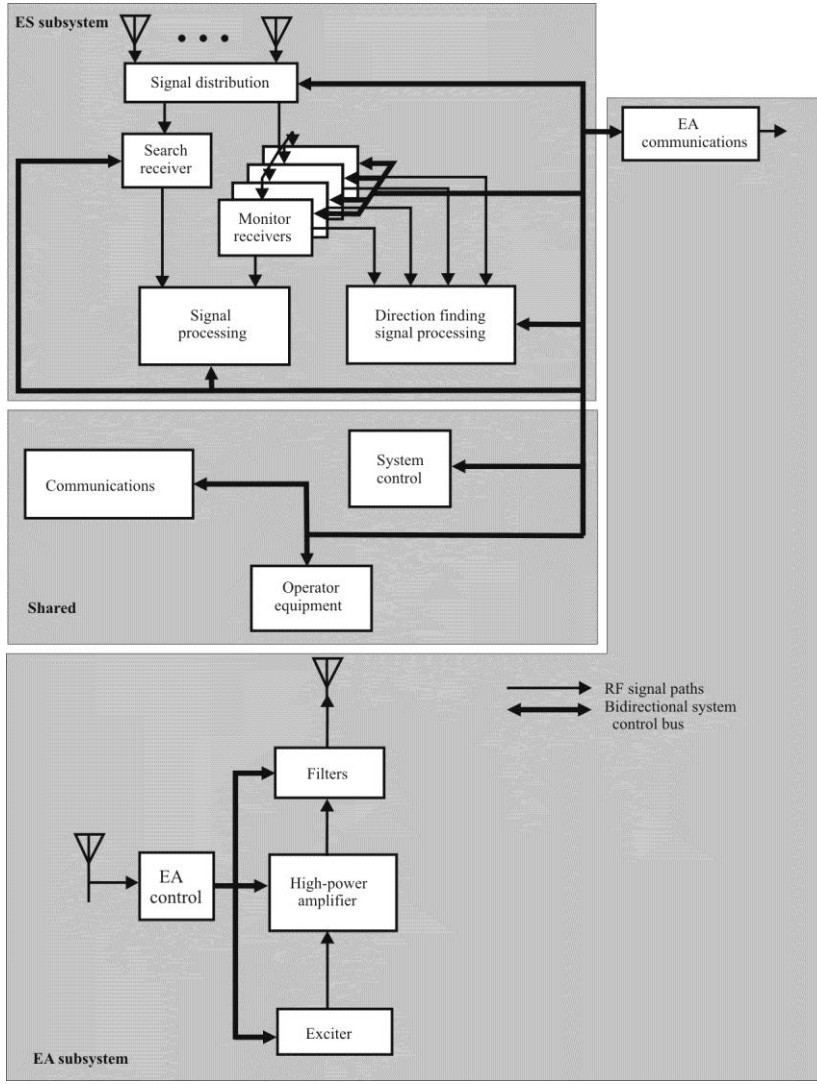


Figure 8.2 Block diagram of an EW system when the EA and ES components are geographically separated.

real operational deployments). We will delve into the performance of the two architectures in Chapter 9.

8.2.1 ES System Architectures

As ES system is used to search and intercept the frequency spectrum searching for and monitoring SOIs. It does this by executing the following functions:

- Detect radiation energy (usually through some form of automated search mechanism);
- Intercept signals;
- Monitor signals;
- Geolocate targets;
- Record signals;
- Analyze transmissions;
- Characterize transmissions into one or several bins (e.g., single channel/multichannel, narrowband, or LPI);
- Prepare gists of transmissions;
- Provide steerage for EA systems;
- Report on information gleaned.

By performing these tasks, ES systems contribute critical information to SA and therefore NCO by providing input into the electronic map as well as the distribution of forces on the battlespace. The signals that are detected and located are used to update the EOB as well.

An important function of a communication ES receiving system is geolocation functionality (mentioned above). That is, for those SOI that are received, the geographical location of the transmitter emitting those signals is an essential piece of information. That function is not covered in this book. A companion but separate book is available which covers that topic in detail [7].

We discuss each of the blocks in the ES subsystem shown in Figure 8.1 in this section.

8.2.1.1 Antennas

Antennas are used to extract the EM energy from the propagation medium. As a minimum the antenna must cover the complete frequency range of operation of the EW system which, in fact, may entail more than one physical antenna. Antennas convert the EM energy into electrical signals for use of subsequent stages.

Antennas are also used in the reverse fashion for EA applications, converting electrical signals into EM energy that can be propagated through the atmosphere to the target.

8.2.1.2 Signal Distribution

The signal distribution function routes the signals from the antenna to different places in an ES system. Simply splitting the signals is seldom adequate and concerns about, for example, impedance matching, must be taken into account. The antenna signals are normally quite weak, and routing such weak signals must be accomplished properly. Signal distortion can arise if impedances are mismatched as well.

8.2.1.3 Search Receiver

There is usually a function to be performed in an RF ES system that searches the frequency spectrum looking for a SOI. This is accomplished by a search receiver. This receiver typically scans through the spectrum of interest, looking for energy. When it finds energy, measurements are made to characterize it. Those measurements can be performed as part of the search receiver, or by using an intercept receiver.

In the simplest of ES systems, the search function may be accomplished manually by an operator tuning a narrowband intercept receiver or perhaps automatically scanning the intercept receiver through the spectrum of interest.

8.2.1.4 Monitor Receivers

A monitor receiver is used for relatively long-term analysis of signals detected by other means. In the simplest of ES systems, there may be only one of these. There is usually one per operator; however, there can be many more than this when signals are recorded for later analysis. They are tuned either manually by an operator or automatically based on energy detected by the search receiver. When the search receiver is a digital implementation, the intercept receivers may, in fact, be nothing more than channelized filters using the search receiver as the RF portion. The outputs of these receivers are used to measure parameters of signals or, in the case of analog communications, for example, for the operators to listen to.

Modern intercept receivers are typically digitally controlled where the digital control word can change any of the parameters of the receiver, such as frequency and IF bandwidth attempting to match the parameters of the SOI.

Superheterodyne receivers are the most prolific form of intercept receivers in use. Figure 8.3 shows the architecture a single-conversion superheterodyne receiver architecture. More conversion stages are possible and have certain advantages. The following elements are common to all superhet circuits: a receiving antenna, a tuned stage that may contain an RF amplifier, a variable frequency local oscillator, a bandpass IF filter and amplifier, and a demodulator

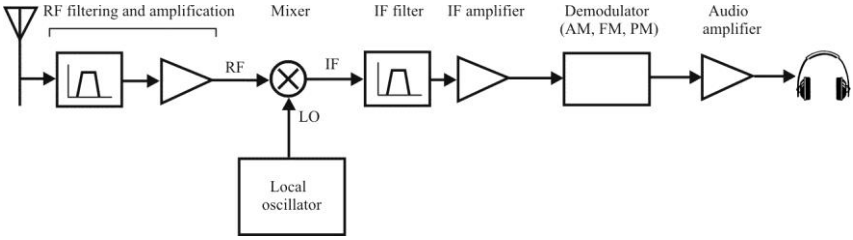


Figure 8.3 Block diagram of a typical superheterodyne receiver. IF: Intermediate frequency; LO: local oscillator.

plus additional circuitry to extract the original audio signal (or other transmitted information).

8.2.1.5 Signal Processing

There are several types of signal processing in EW systems. Typical functions include detection of the presence of energy at a particular frequency and within a specified bandwidth, determination of the modulation on a signal, measuring the baud rate of a digital communication signal, noise reduction, and so forth.

One particularly important form of signals processing is determining the DOAs of the targets of interest. DOAs from physically separated EW systems are used to estimate the geographical location of target transmitters.

8.2.1.6 Communications

The communication subsystem is the means for command and control of the system as well as the means for tasking and reporting. If the system is remotely controlled, then this subsystem is the means to exercise that control.

8.2.1.7 Recording Subsystem

The recording subsystem is included in the architecture because frequently it is necessary to save intercepted audio for later, more in-depth analysis.

Analog Recording

Analog tape recorders are mentioned here for historical reasons only. They are no longer used in EW systems. In the past, however, before the development of large digital disks and later semiconductor *random access memory* (RAM) with capacities in the tens of gigabyte range, analog tape recorders were used to store intercepted audio.

Digital Recording

A great deal more flexibility and functionality is possible by storing signals in a digital format and that is the way it is currently accomplished. Audio can be accessed randomly rather than by rewinding or forwarding a tape recorder.

A 16-GByte semiconductor disk (thumb drive) with audio recorded at 16 kbps can store over 2,000 hours of audio; 128 kbps is common speed for MP3 encoding of music; however, speech audio can be stored at substantially lower speed than that.

8.3 Receiving Systems

In this section we present some of the notions applying to receiving systems in general.

8.3.1 Basic Architecture

A diagram of a simple (maybe the simplest) receiving system is shown in Figure 8.4. The receiving subsystem consists of an antenna, a receiver (Rx), a *spectral display unit* (SDU), and a *receiver control unit* (RCU). The operator subsystem consists of a system computer and the necessary equipment for an operator to interface with it. The ability to communicate with the outside world is also included. Signal paths (RF, IF/video, and audio in this case) are shown with thin lines while thick lines indicate control lines.

The Rx may actually consist of several receivers, each of which covers a portion of the total frequency spectrum (receivers for the HF range are typically much different from those for the VHF range, for example, and two receivers would be implemented in this case). The SDU displays the portion of the frequency spectrum of current interest to the operator. The signal sent from the receiver to the SDU is normally the IF signal in the receiver or some other video (predetected) signal useful for the purpose. The presence of a signal at a frequency/channel is indicated by some mechanism on the screen of the SDU. The functionality of the SDU is frequently implemented with the computer and displayed on the operator display, rather than implemented as a separate unit. The RCU performs the function of controlling the receiver and may, in reality, be the receiver front panel or a function implemented by the system computer, in which case the operator control of the receiver is via the keyboard and display. Control signals are sent from the RCU to the receiver to change the receiver settings, such as frequency. In return, status signals are sent from the receiver back to the RCU to indicate the health of the receiver.

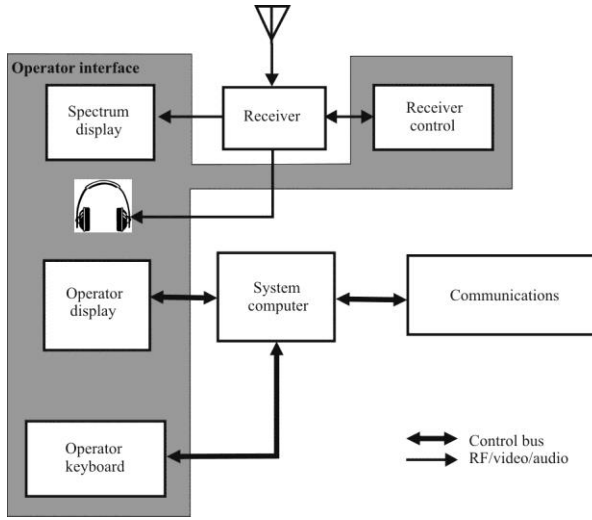


Figure 8.4 Block diagram of a simple receiving system.

The operator subsystem, while not literally part of the receiving subsystem, is shown for completeness to illustrate where the operator fits into the receiving process. The keyboard and display permit the operator to interface with the system computer to facilitate receiving tasking from the C2 source as well as to communicate the results of the receiving process to a destination.

The architecture, while simple in this case, is extensible and not much changes. A receiving system with many operators each with one or more receivers would look like Figure 8.4 replicated the necessary number of times. The system computer would likely be more powerful. The operator interfaces would likely be some sort of control bus such as MIL-STD-1555 or 10baseT or 100baseT Ethernet.

In the largest of such receiving system, the functions of searching the spectrum and subsequent copying of targets of interest would typically be separated and executed by operators specializing in one subset of the overall functions.

8.3.1.1 Signal Searching

The system shown in Figure 8.4 could be used to search the spectrum for signals of interest, or it could be used to tune to specific frequencies to see if there is current activity at a frequency. The former is referred to as *general search* and typically the search is executed from some start frequency to some stop frequency, typically linearly.

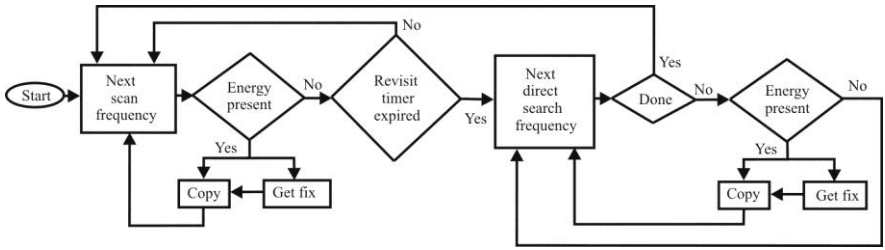


Figure 8.5 Process flow for the simple receiving system.

The latter search method is referred to as *directed search*, and specific frequencies are programmed into the RCU. This is the mode when the target frequencies of interest are a priori known.

The two search modes can be mixed. One implementation of an algorithm that combines the two is illustrated in the process flow diagram in Figure 8.5. In this implementation a revisit timer is used to cycle between the two search modes. It represents the amount of time the system is spent in the general search mode before entering the directed search mode. Starting in the general search mode, the next scan frequency is selected to examine for energy presence. If there is energy present at that channel, the signal is copied, which may or may not be accompanied by recording of the signal. Other functions may also be executed at that time, such as obtaining a geoposition (fix) of the target, as illustrated in Figure 8.5. If the target is one of interest, typically the channel is entered onto the directed search list. When copying is completed, the scan process proceeds. If there is no energy detected at the tuned channel, the revisit timer is examined to see if it has expired. If not, the next channel is chosen. If it has expired, the directed search mode is entered. In this mode, each directed search frequency is examined in sequence. If energy is detected at one of these frequencies, then, as in the general search mode, the target is copied, and possibly other functions are executed. Once all the directed frequencies are examined, the general search mode is started again.

A somewhat more extensive receiving system is illustrated in Figure 8.1. In this case the search receiver/receiving subsystem is separate from the receiver used for the copy function. The copy receivers are typically queued from the search process and tuned automatically to where signal energy is detected. The signals thus received are recorded into the receiving subsystem, which could be analog tape recorders, or, more recently, digitally recorded onto computer disk storage. The operators in this case would process the recorded signals rather than the signals directly from the receivers.

Recording the signals digitally provides for considerably more flexibility at playback, such as the ability to loop forward and backward and enter a recorded segment anywhere instantly, rather than rewinding an analog tape. Control over the playback process could be provided as a function within the system computer or with a separate control unit, much like the RCU discussed above.

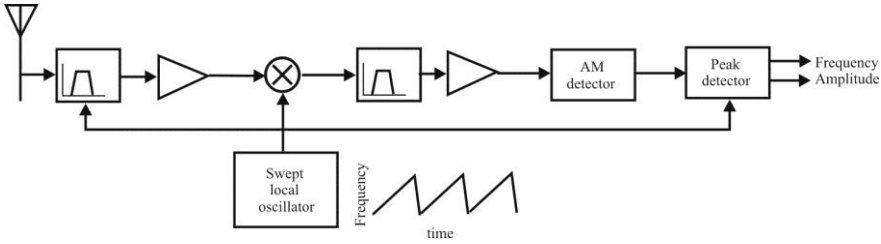


Figure 8.6 Simplified block diagram of a scanning superheterodyne receiver. This flow diagram is similar to that shown in Figure 6.3, except the LO is swept in time versus fixed.

The signal distribution unit takes the input from the antenna(s) and routes, with minimum loss, the antenna signals to all the receivers. It consists primarily of signal splitters, a topic that is revisited in detail in [8].

8.3.1.2 Search Receiver Architectures

There are several configurations of receivers that can serve as search receivers. If search speed is not an issue, the simple narrowband receiver described in Section 8.3.1 can be manually tuned through the spectrum of interest to search for targets. In modern EW systems, this is not normally the case, however. When faster searching is required, one of the methods described in this section is usually implemented.

Scanning Narrowband Receiver

The simplest is probably the scanning superheterodyne receiver, which is simply a superheterodyne receiver whose frequency is linearly changed with time. With modern, digitally tuned narrowband receivers, this scanning is not really scanning at all, but stepping from one frequency to the next (not necessarily linearly) by sending tuning commands to the receiver. The receiver dwells on a frequency for a time, measures the energy present at that frequency, and then moves on to the next frequency. If energy is detected on that dwell, the system reacts accordingly, such as tuning a narrowband copy receiver to that frequency so an operator can process the signal.

This type of receiver is used for searching the spectrum, looking for energy in frequency channels. Frequencies of SOIs as well as an estimate of their amplitudes can be determined. A block diagram of such a receiver is shown in Figure 8.6. Torrieri [9] provides a detailed analysis of this receiver.

The swept local oscillator causes signals within a frequency band to be mixed within the mixer. The preselector filters must also be tuned along with the local oscillator so that synchronization is maintained. A narrowband signal at the input

will be mixed whenever the local oscillator tunes to the IF offset from the signal (recall that the mixer output is a constant frequency, the IF). At that point in time the AM detector will detect the peak amplitude of the signal and the peak detector will measure the amplitude. By measuring where the peak occurs in time, that time can be compared with where the scanning local oscillator is at that time so a frequency estimate can be computed.

If μ represents the scanning rate of the receiver in hertz per second and W represents the 3 dB bandwidth of the bandpass filter after the mixer, the normalized peak value α (normalized relative to the amplitude of the input signal) is given by

$$\alpha = \left(1 + 0.195\frac{\mu^2}{W^4}\right)^{-1/4} \tag{8.1}$$

and the frequency resolution is given by

$$\Delta = B\left(1 + 0.195\frac{\mu^2}{W^4}\right)^{1/2} \tag{8.2}$$

Equations (8.1) and (8.2) are plotted in Figures 8.7 and 8.8, respectively, for W = 25 kHz. Thus, the normalized amplitude peak value decreases as the scan rate increases while the resolution increases in value (selectivity decreases). In a dense RF environment, a resolution bandwidth of less than 50 kHz or so is desirable in order to minimize adjacent channel interference. Thus, in the military VHF range (30–90 MHz) where the signals typically have a bandwidth of 25 kHz, the scanning rate must be kept at about 3 GHz per second or less in order to maintain the required resolution.

Most modern tuned superheterodyne receivers are digitally controlled for maximum flexibility. They can be controlled directly with an RCU or via computer as required. These receivers actually dwell on a frequency channel as opposed to scanning as in an analog equivalent.

The frequency resolution achievable with a digitally tuned narrowband receiver is inversely proportional to dwell time. For example, if the receiver dwells at a frequency for 10 ms and the instantaneous bandwidth of the receiver is 200 kHz, then the maximum channelization possible is to divide the 200 kHz bandwidth into 200, 1 kHz cells.

Compressive Receiver

The compressive receiver is another common form of search receiver. Several megahertz can be scanned by such a receiver per microsecond.

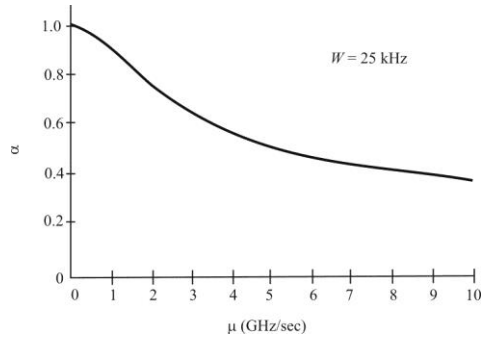


Figure 8.7 Example plot for the normalized peak value for a scanning superheterodyne receiver when W = 25 kHz. (Source: [10], © 2002, Artech House. Reprinted with permission.)

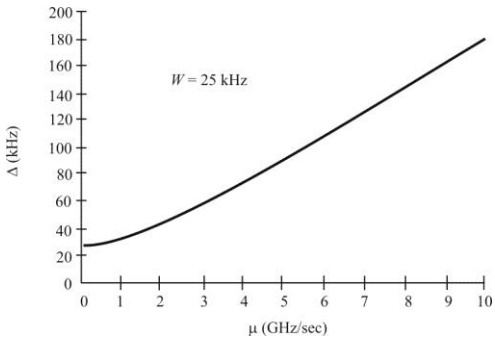


Figure 8.8 Scanning superheterodyne receiver frequency resolution when W = 25 kHz. (Source: [10], © 2002, Artech House. Reprinted with permission.)

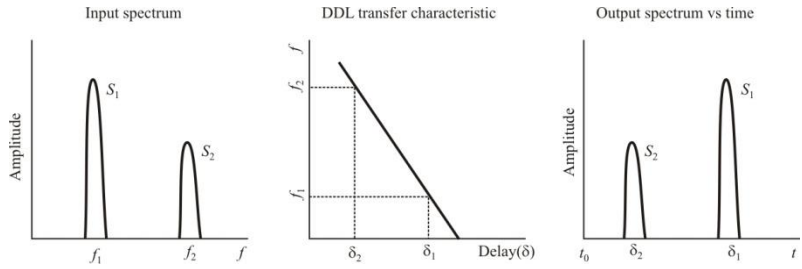


Figure 8.9 DDL characteristics.

The principal component in a compressive receiver is the *dispersive delay line* (DDL), one version of which is available as a SAW device. A DDL is one implementation of a *chirp filter* and has an impulse response with a linearly varying frequency characteristic.

The operation of a DDL is illustrated in Figure 8.9. The output is (ideally) a replication of the spectrum of the input but the various signals at different input frequencies emerge from the DDL at different times due to the linear transfer characteristic of the DDL shown.

There are two fundamental architectures for compressive receivers. The first is the *multiply-convolve-multiply* (MCM) architecture shown in Figure 8.10(a) and the second is the *convolve-multiply-convolve* (CMC) architecture shown in Figure 8.10(b). They both compute the *short-term Fourier transform* (STFT) of the input signal; which one is chosen is largely an implementation/manufacturing issue. The STFT exits the receivers serially. If the true FT is not required, but simply the amplitude versus frequency (spectrum) characteristic of the input signal, then the second convolve in Figure 8.10(b) is not required. If this is the case, however, the phase response is not available so no subsequent processing that uses phase information (e.g., DF) is possible.

If we let T denote the duration of a single scanning pulse and T_c the duration of the response of the DDL, then when

$$T_L = \min(T, T_c) \tag{8.3}$$

we have the frequency resolution of the compressive receiver, Δ, given approximately by [9]

$$\Delta \approx \frac{1}{T_L}, \qquad |\mu| T_L^2 >> 1 \tag{8.4}$$

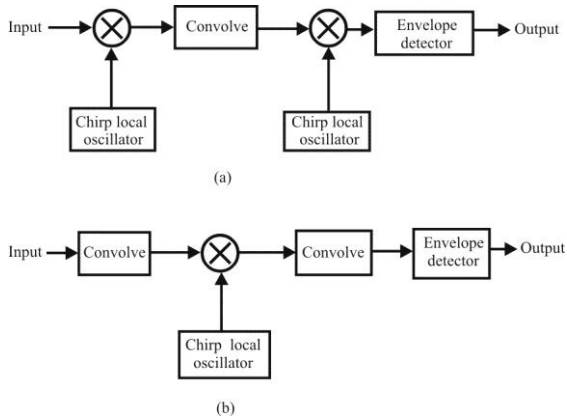


Figure 8.10 Compressive receiver architectures. (a) MCM configuration. (b) CMC configuration.

where μ is the scan rate. For example if $\mu = 10\,\text{GHz/sec}\,(10\,\text{kHz}/\mu\text{sec})$ and $T_L = 100\,\mu\text{sec}$, then $|\mu| T_L^2 = 100$, and $\Delta \approx 10\,\text{kHz}$. We see that, unlike the superhet, the frequency resolution of the compressive receiver is independent of the scan rate. Furthermore, the resolution of the compressive receiver is considerably improved over the scanning superhet.

One disadvantage of the compressive receiver is shared with all signal processing techniques that sample time waveforms. Sidelobes are generated due to the sampling. Without careful design, the sidelobes generated by strong, sometimes friendly transmissions can mask the SOI that may be in the adjacent channel. Without weighting these first sidelobe levels are about –13 dB (relative to the in-channel signal level). This level can be substantially reduced by weighting the response of the DDL, with the tradeoff being increasing the width of the in-channel response [10]. With simple weighting, –35 dB can be achieved and much lower levels are possible if the in-channel bandwidth expansion can be tolerated.

Digital Receiver

A receiver that digitizes the RF or IF signal as early in the receive process as possible which also has a wide bandwidth also serves as a search receiver. Digitizing the RF signal followed by an *N*-point DFT essentially creates *N* narrowband filers/receivers and energy detection can be performed in each of these channels essentially in parallel.

Optimal energy detection is typically accomplished with a radiometer at the output of each of these channels. Actually, at this point the signals are digital, so a single radiometer can be, and normally is, shared among several of the channels.

In addition, any one of the channels can function as an intercept receiver essentially deleting the requirement for a separate set for the copy function.

The frequency resolution of this approach is only limited by the duration of the sample time. If the sample time is T_s, then the frequency resolution is approximately

$$\Delta \approx 1/T_s.$$

For example, if the sample time is 10 ms, then the frequency resolution is about 100 Hz.

8.4 EA System Architectures

One fundamental classification for jammer architectures is whether they are thin or thick [Chapters 5 and 11]. Thin jammers are basically remotely controlled and have, in and of themselves, very little in the form of ES capability. (Most jammers have some sort of built-in ES in order to ascertain the presence of the SOI.) The jamming frequency(ies) are supplied remotely for these jammers, and their ES function is simply to measure the frequency spectrum for the presence of the SOI. The frequency assignments are sometimes supplied prior to a mission, but typically are tasked over the network real-time.

A thick jammer on the other hand has considerable embedded ES functionality, and usually an operator to operate the systems. They can operate relatively autonomously. Assigning targets is typically performed before the beginning of a mission.

8.4.1 Jamming Techniques

There are several EA techniques that jammers can employ. We delineate some of those here [11].

8.4.1.1 Spot Jamming

Spot, or tone, jamming utilizes a waveform at a single frequency, which is the frequency of the narrowband target. Normally this tone is modulated (FM is typical) with Gaussian noise. All of the power from the jammer is concentrated at this single frequency for the duration of the jamming mission. With proper filtering, there is minimum fratricide for friendly communications. It is an inefficient method, however, because communication signals do not need to be jammed 100% of the time to degraded [12].

Spot jamming is a useful technique against some forms of LPI targets. DSSS targets can sometimes be disrupted if the processing gain provided by the DSSS technique is overcome.

8.4.1.2 Comb Jamming

This method is similar to spot jamming, except that several frequencies are jammed at the same time. Obviously there is less jammer power at each of the frequencies. It is more efficient than spot jamming but still relatively inefficient.

DSSS targets are susceptible to comb jamming, although the effects of decreased power per tone when using multiple tones needs to be weighed against the greater power available with a single tones.

8.4.1.3 Barrage Jamming

A wideband waveform, usually noise, is broadcast from the jammer in barrage jamming. In any given channel the power is significantly reduced. In addition, there is typically considerable fratricide since the frequency spectrum is an asset shared by both adversarial and friendly communications. It is probably the most inefficient use of jammer power.

Barrage jamming may not require jamming, for example, the entire low VHF spectrum (30–90 MHz). *Partial band noise* (PBN) jamming is employed when only a portion of a given band contains the SOI.

8.4.1.4 Swept Jamming

When a relatively narrowband jamming waveform, modulated with noise, is swept across a frequency band, much of the effect of barrage jamming can be achieved. At any given instant, the waveform is covering a single channel so the entire power of the jammer is concentrated on that channel. The jammer power is thus shared among the channels. The result, however, produced considerable fratricide. It is obviously more efficient than barrage jamming, however.

PBN jamming can also be implemented with swept techniques rather than sweeping through the whole band.

8.4.1.5 Responsive Jamming

In this type of strategy, the jammer is tuned to a target frequency and the energy in the spectrum is measured. If there is energy present, the jammer commences jamming. At a point shortly after that, the jammer signal is turned off, and the spectrum is again measured to ensure that the target signal is still transmitting. If it is, then jamming begins again. This process is repeated until the transmitter stops transmitting.

With this approach the full power of the jammer is applied to each target. It is more efficient than some of the other techniques and with proper filtering fratricide can be avoided. Sophisticated ES at the jammer is required to perform the energy detection and decision making, however.

Follower jamming, where the spectrum is measured and a jamming signal is applied only where new energy appears (and the signal satisfies other criteria as well, such as DOA), is a technique used to apply EA countermeasures to some types of spread spectrum communications, such as frequency hopping.

8.4.2 Asset Sharing

From the above brief discussions of some of the basic jamming techniques, it is clear that sharing the jamming assets is possible and increases the efficiency of the jammer operation. The two most common methods for this sharing is to share the jammer in time and utilization of its power.

8.4.2.1 Time Sharing

The jammer time is shared, for example when the jammer waveform is swept through the frequency spectrum. The jammer spends a brief moment at each channel with full power. Spot and comb jamming can also share the jamming waveform in time by spending time at multiple target frequencies one after the other. For example, in digital communications a 10% BER can be generated in the communication signal by disrupting only 10% of the waveform. Theoretically, 10 channels could be jammed with a single jammer. For analog FM communications, the communications can be significantly degraded if about 30% of the transmission can be jammed [12]. Thus, a single jammer can jam about three analog FM targets essentially simultaneously by time sharing the jammer.

8.4.2.2 Power Sharing

Jammer power can be shared as in comb and barrage jamming. Multiple targets are attacked simultaneously this way. The power at each frequency is reduced however. In fact, the power at each frequency decreases faster than $1/N$ when N target frequencies are attacked. This can be seen as follows [12]. Consider sharing the jammer power between two signals $s_1(t)$ and $s_2(t)$. The transmitted signal is thus

$$s(t) = s_1(t) + s_2(t)$$

The radiated power is

$$
\begin{aligned}
P &= \frac{1}{T}\int_0^T s^2(t)dt \\
&= \frac{1}{T}\int_0^T [s_1(t)+s_2(t)]^2 dt \\
&= \frac{1}{T}\int_0^T [s_1^2(t)+2s_1(t)s_2(t)+s_2^2(t)]dt \\
&= \frac{1}{T}\int_0^T s_1^2(t)dt + \frac{1}{T}\int_0^T 2s_1(t)s_2(t)dt + \frac{1}{T}\int_0^T s_2^2(t)dt
\end{aligned}
$$

The first and last terms are the radiated powers at the two target frequencies while the middle term is a signal radiated at neither of the target frequencies. This power is lost as far as the jamming function is concerned.

8.4.3 Jamming Systems

Jamming systems perform their function by executing the following principal tasks:

- Selects frequencies to be jammed from a prioritized target list.
- Form jamming signals by generating low power exciter signals and applies the appropriate EA waveform to the signal (modulation, usually noise or tones).
- Amplifies the jamming signals to appropriate levels.
- Jam SOIs by emitting high-powered signals at appropriate frequencies.
- Look through to ensure the target SOI is still on the tasked frequency. (This function may be performed by an associated ES system.)

The intent is to input the jamming signals into the target receivers at sufficient levels to overwhelm the intended signal at the receiver.

The EA subsystem shown in Figure 8.1 consists of four principal components: (1) the exciter, (2) the power amplifier, (3) the filters, and (4) the transmit antenna. We will briefly discuss these subsystems in this section.

8.4.3.1 Exciter

An exciter is an RF signal generator that generates the jamming signal at the desired frequency. That signal is a tone, and it can be used directly (after amplification) as the jamming signal. More often, however, the RF signal is modulated. The modulation can take many forms, with the jamming performance depending on the modulation as well as the modulation of the target signal. The

modulation used most frequently for communication signals is random noise. This signal raises the noise level at the target receiver, thus decreasing the SNR. The impacts of the SNR at the receiver on a communication system's performance have been heavily studied. Therefore accurate prediction of degradation of performance due to jamming a signal can be estimated. Other forms of modulation that can be used include tone jamming (no modulation at all or perhaps a tone offset from the carrier), sometimes effective against FSK signals. For a complete analysis of the impacts of the common jammer modulations on target signals, see [13].

8.4.3.2 Power Amplifier

The *high power amplifier* (HPA) amplifies the signals(s) from the exciter. The signal from the exciter is normally quite moderate—typically 0 dBm, or 1 mW. The HPA significantly increases this signal to a level adequate for jamming purposes. For the communication EA applications we are interested in, the signal level sent to the antenna can range from 1W to over 10 kW.

HPAs for communication EW applications must be broadband, for the same reasons the other components in the EA subsystem must be broadband. A priori knowledge of the spectrum location of target systems is almost always unknown so the EW systems must be capable of adapting to the target environment.

The propagation paths over which the intended signal travels from the transmitter to the receiver is generally not the same as the path taken by the jammer signal from the jammer to the receiver. As such, the jammer must assume the worst-case path will be taken by its signals. Jammers are therefore typically considerably more powerful than the transmitters in the target communication network.

8.4.3.3 Filters

EA operations in the presence of friendly communication networks can significantly impinge on these networks if the power from the HPA is not properly managed. Such interference is usually called RF fratricide. To avoid this fratricide, filters on the output of the amplifier are frequently needed, since perfect amplifiers with no out-of-band spurious responses are not a reality. Connected at the output of the HPA, these filters must be able to handle the power levels associated with the output of the PA.

For general purpose EA applications, these filters must be tunable. It is typically not a priori known where in the frequency spectrum a target will appear (although it is generally true that the *range* of target frequencies is probably known). Therefore, the jammer and the associated filters must be able to change the transmit frequency. For frequency-hopping targets, the filters must tune rapidly.

Some jammers are designed to operate against multiple targets simultaneously (or apparently so). When targeted against multiple targets at different frequencies, multiple filters are needed. There is an operational mode in some jammers, however, where jamming multiple targets is accomplished by time sharing the EA assets. A target is jammed in this instant, and moved to the next frequency in the next instant, coming back to the first target, next, and so on. Poisel [13] contains a description of how such systems perform. In this case a single filter is sufficient; however, it must be rapidly tunable.

The output filters may be required to have tunable bandwidths as well. Some modes of EW require (relatively) broadband noise waveforms. The output filters must be capable of passing these waveforms with minimal distortion.

8.4.3.4 Transmit Antenna

The antenna is used to convert the electrical signals from the power amplifier/filter chain into EM waves for propagation through the transmission medium. They typically have some gain due to directionality. An antenna that is isotropic (transmits the same in all directions) in the horizontal plane is a monopole or dipole. It has about 1.5 dBi+ gain in the horizontal plane. In ground applications, it is not normally used as transmit antennas, however, because antennas with more gain are available. Such antennas are directional and must be pointed towards the targets. A log-periodic is an example of such an antenna. It exhibits 6 dBi+ typically. An extensive description of antennas used for EW applications is given in [8].

8.5 EW System Operational Considerations

8.5.1 Means Versus Effects

EW means are applied against targets to create a full range of lethal and nonlethal effects. (See Figure 8.11 [15].) Choosing a specific EW capability depends on the desired effect on the target and other considerations, such as time sensitivity or limiting collateral damage. EW capabilities provide commanders with additional options for achieving their objectives. During major combat operations there may be circumstances in which commanders want to limit the physical damage on a given target. Under such circumstances, a clear understanding of the lethal and nonlethal effects EW capabilities can be achieved. For example, a target might be enemy radar mounted on a fixed tower. Two EW options to defeat the radar could be to jam the radar or destroy it with antiradiation missiles. If the commander desired to limit damage to the tower, an electronic attack jamming platform would be preferred. In circumstances in which commanders cannot sufficiently limit undesired effects such as collateral damage, they may be constrained from

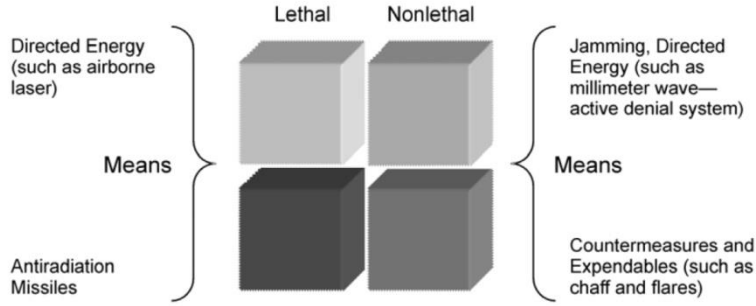


Figure 8.11 Means versus effects (Source: [15].)

applying physical force. The EW staff articulates succinctly how EW capabilities can support actions to achieve desired effects and provide lethal and nonlethal options for commanders.

8.5.2 Radio Propagation Issues

Radio propagation over land, especially in mountainous or urban environments, is highly variable and extremely difficult to predict even with the most advanced analytical propagation models because all of the parameters affecting propagation cannot be fully known. Radio propagation models may be used to estimate the coverage distances of ES sensors against land-based communications emitters, the *jamming-power-to-signal-power ratio* (JSR) at target receivers, and air-ground-air (A/G/A) data link distances for UASs. These propagation models can also be used to estimate the number of cochannel signals that a UAS may encounter due to radio frequency reuse by terrestrial communications systems. Radio propagation path loss can be modeled using various models, empirical, based on measured value, and analytical, based on wave and optic theory. There are a large number of models available, some that take into account terrain-specific features (such as digital terrain elevation data) and some that do not.

8.5.3 Wartime Reserve Modes

We have mentioned the EOB several times. Indeed, it is the starting point for EW operations. However, wartime reserve modes can be employed by an adversary that tend to minimize the value of the EOB. Wartime reserve modes are characteristics and operating procedures of sensors, communications, navigation aids, threat recognition, weapons, and countermeasures systems that will contribute to military effectiveness if unknown to or misunderstood by opposing commanders before they are used, but could be exploited or neutralized if known

in advance. By definition, wartime reserve modes are deliberately held in reserve for wartime or emergency use and seldom, if ever, applied or intercepted prior to such use.

8.5.3.1 EW Reprogramming

Adapting to wartime reserve modes is one reason for EW reprogramming. Electronic warfare reprogramming refers to modifying friendly EW or target sensing systems in response to validated changes in enemy equipment and tactics or the electromagnetic environment. Reprogramming EW and target sensing system equipment includes changes to self-defense systems, offensive weapons systems, and intelligence collection systems. During joint operations, swift identification and reprogramming efforts are critical in a rapidly evolving hostile situation. The key consideration for EW reprogramming is joint coordination.

8.5.4 Employment Considerations

EW has specific ground-based, airborne, and functional (electronic attack, electronic warfare support, or electronic protection) employment considerations. Each capability employed has certain advantages and disadvantages.

8.5.4.1 Ground-Based EW Considerations

Ground-based EW capabilities support the commander's scheme of maneuver. Ground-based EW equipment can be employed by a dismounted soldier or on highly mobile platforms. Due to the short-range nature of tactical signal direction finding, electronic attack assets are normally located in the forward areas of the battlefield, with or near forward units.

Ground-based EW capabilities have certain advantages. They provide direct support to maneuver units (for example, through counter-radio-controlled improvised-explosive-device EW and communications or sensor jamming). Ground-based EW capabilities support continuous operations and respond quickly to EW requirements of the ground commander. However, to maximize the effectiveness of ground-based EW capabilities, maneuver units must protect EW assets from enemy ground and aviation threats. EW equipment should be as survivable and mobile as the force it supports. Maneuver units must logistically support the EW assets, and supported commanders must clearly identify EW requirements.

Ground-based EW capabilities have certain limitations. They are vulnerable to enemy attack and can be masked by terrain. They are vulnerable to enemy electromagnetic deceptive measures and electronic protection actions. In addition, they have distance or propagation limitations against enemy electronic systems.

8.5.4.2 Airborne EW Considerations

While ground-based and airborne EW planning and execution are similar, they significantly differ in their EW employment time. Airborne EW operations are conducted at much higher speeds and generally have a shorter duration than ground-based operations. Therefore, the timing of airborne EW support requires detailed planning.

Airborne EW requires the following:

- A clear understanding of the supported commander's EW objectives;
- Detailed planning and integration;
- Ground support facilities;
- Liaisons between the aircrews of the aircraft providing the EW support and the aircrews or ground forces being supported;
- Protection from enemy aircraft and air defense systems.

Airborne EW capabilities have certain advantages. They can provide direct support to other tactical aviation missions such as *suppression of enemy air defenses* (SEAD), *destruction of enemy air defenses* (DEAD), and employment of high-speed antiradiation missiles. They can provide extended range over ground-based assets. Airborne EW capabilities can provide greater mobility and flexibility than ground-based assets. In addition, they can support ground-based units in beyond line-of-sight operations.

The limitations associated with airborne EW capabilities are time-on-station considerations, vulnerability to enemy electronic protection actions, electromagnetic deception techniques, and limited assets (support from nonorganic EW platforms need to be requested).

8.5.5 ES Operational Considerations

The distinction between whether a given asset is performing a signals intelligence or EW support mission is determined by who tasks and controls the assets, what they are tasked to provide, and the purpose for which they are tasked. Operational commanders task assets to conduct EW support for the purpose of immediate threat recognition, targeting, planning the conduct of future operations, and other tactical actions (such as threat avoidance and homing). The EWO coordinates with the G-2 or S-2 to ensure all EW support needed for planned EW operations is identified and submitted to the G-3 or S-3 for approval by the commander. This ensures that the required collection assets are properly tasked to provide the EW support. In cases where planned electronic attack actions may conflict with the G-2 or S-2 intelligence collection efforts, the G-3, S-3, or commander decides which has priority. The EWO and the G-2 or S-2 develop a structured process

within each echelon for conducting this intelligence gain-loss calculus during mission rehearsal exercises and predeployment workups.

System operators are usually but not always deployed with ES assets. A receiving system could be deployed at a location and used for recording intercepts that are replayed later at the same or a different location.

8.5.5.1 Siting Considerations

In general, the higher the ES antenna, the better. Frequently the ES targets are a considerable distance away from the EW system, and increasing the antenna height increases the amount of target signal that can be collected.

In addition, site locations must be traversable. This is not only for EW system siting but for support vehicle access as well.

DF Baselines

The geometry of the baseline of both ground-based and airborne ES systems affects the coverage area and, in particular, the accuracy that is achievable over this area. Shown in Figure 8.12 [14] is a chart showing the DF accuracy coverage of a baseline of three EW systems. These systems are spaced evenly along the baseline that extends along the *y*-axis. They are separated by *y*/*L* (in consistent units). The parameter on the chard is $CEP / L\sigma_L$ where *CEP* is the circular error probable,[1] *L* is the length of the baseline, and σ_L is the standard deviation of the DF measurements. For example if *L* = 20 km and $\sigma_L = 5^o$ (0.09 radians), a target located at *y*/*L* = 0.5 (*y* = 10 km) and *x*/*L* = 2 (x = 40 km) would be located with an accuracy of $CEP / L\sigma_L = 20$, or a CEP = 3,600m. The dashed parts of the contours represent areas where the CEP calculation is suspect because of numerical issues. For example, off either end of the baseline the CEP is unbounded. In addition, the contours for the fourth quadrant are mirror images of the contours in the first quadrant.

A similar chart for a W-shaped baseline is illustrated in Figure 8.13 [14]. In this case we can see that the CEP becomes unbounded off the end of both of the baselines made up of the EW system above the *x*/*L* axis with the system on the *x*/*L* axis, and the baseline made up of the system below the *x*/*L* axis with the system on the *x*/*L* axis.

Therefore the geometry of the DF baseline is very important in determining the achievable accuracy from a set of DF systems.

Taking a somewhat different perspective on the DF coverage issue, consider the long and narrow AOR illustrated in Figure 8.14. To optimize the accuracy of the fixes computed for the region of interest, the EW systems should form a

[1] The circular error probable is the circle contour within which the target lies with a specified probability (usually 50% or 90%). An elliptical error probable (EEP) can be defined similarly.

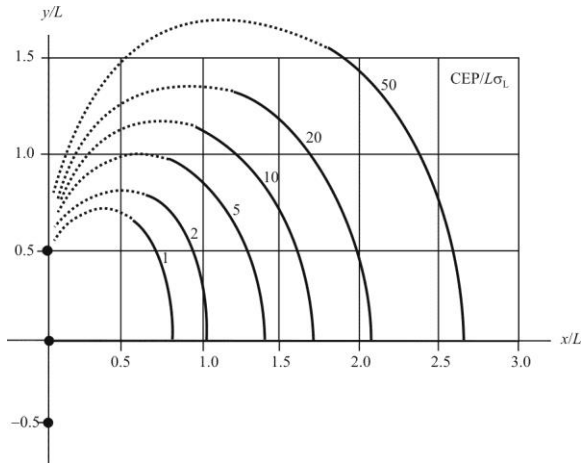


Figure 8.12 Linear baseline isocontours. (Source: [14]. © Artech House, 2008. Reprinted with permission.)

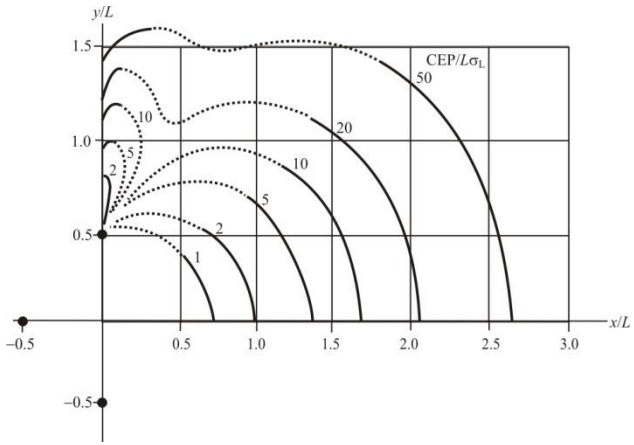


Figure 8.13 Nonlinear baseline isocontours. (Source: [14]. © Artech House, 2008. Reprinted with permission.)

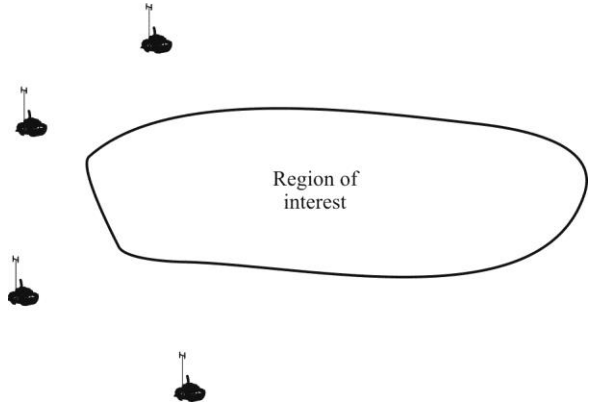


Figure 8.14 Narrow region of interest.

concave contour around the region as shown (this assumes that the EW systems are in a stand-off posture). With this shape, the probability of forming unintended baselines and their attendant high CEPs can be minimized.

Likewise, the region shown in Figure 8.15 is shallow and long. In this case, a convex baseline geometry will tend to minimize the unintended baselines.

Airborne ES systems have the advantage of being able to move to compute bearings (see Figure 8.16). The track in this case can cover most of the region of interest. They have better coverage but potentially more interference as well.

8.5.5.2 Noise Considerations

ES systems are usually designed to be very sensitive. This is so they can intercept signals that are a considerable distance away. However, being sensitive, they also intercept more noise than otherwise. This background noise can be atmospheric, man-made (arc-welders and neon lights are two examples of man-made noise), thermal, and galactic. It is best to locate an operational EW system as far away from potential noise sources as possible. Parking a system next to an office building, for example, is not a good idea.

8.5.5.3 Metallic Structures

Metallic structures reflect electromagnetic signals. These reflections cause two affects. Ghosts can be formed at baseband, which may cause the signals to be difficult to listen to. The other is that the DF subsystem cannot tell the difference between the original signal and the reflected signal. Most DF systems will

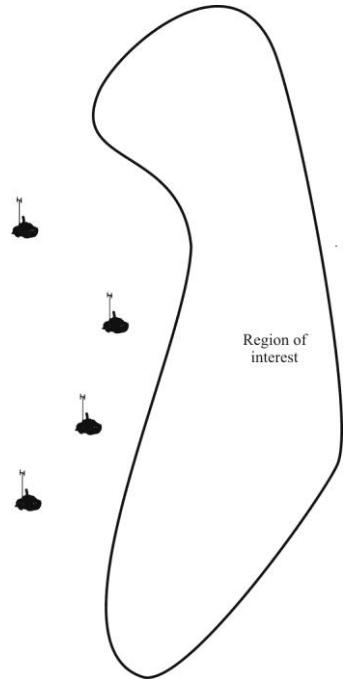


Figure 8.15 Long and narrow region of interest.

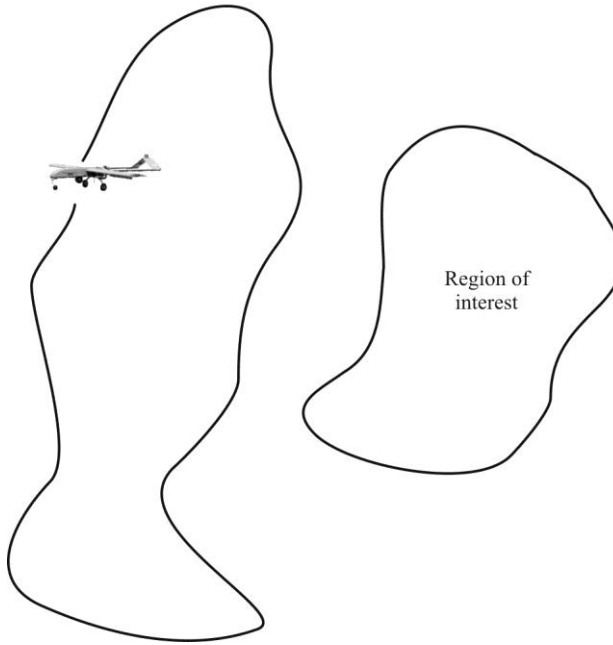


Figure 8.16 Aircraft coverage.

compute some sort of average direction of arrival based on both the original signal and the reflection.

8.5.6 EA Operational Considerations

Electronic attack includes both offensive and defensive activities. These activities differ in their purpose. Defensive electronic attack protects friendly personnel and equipment or platforms. Offensive electronic attack denies, disrupts, or destroys enemy capability. In either case, certain considerations are involved in planning for employing electronic attack:

- Friendly communications;
- Intelligence collection;
- Other effects;
- Nonhostile local electromagnetic spectrum use;
- Hostile intelligence collection;
- Persistency of effect.

There are several ways to categorize the operational deployment of jammers. We will consider several of them here.

A *stand-in jammer* is a system that is deployed into operation in the midst of the target networks. As such, the targeted receiver(s) can be anywhere in the region around the jammer. A *stand-off jammer* is deployed somewhat away from the target networks and the targets are generally in a single (maybe large) region so the jamming direction can be focused.

An *escort jammer* is the term applied to a stand-in airborne jammer that accompanies an aircraft sortie into the adversarial air space. Their purpose is to deny the adversary use of the RF spectrum, and, in particular, for their air defense units (*suppression of enemy air defense*, SEAD).

Some types of airborne jammers are mounted on unattended aerial systems (UAS). As such, they can either be fixed into the aircraft or they can be discharged for operation on the ground. Jammers for the latter of these configurations are referred to as expendable. Deploying expendable jammers is not limited to UAS, however. They can also be hand-emplaced or artillery-delivered.

8.5.6.1 Signal Fading

Signal fading occurs frequently in both ES and EA situations, and is primarily a ground-based deployment problem. It is caused by the target signals being affected by reflections off of reflective surfaces. Such surfaces include moving vehicles, road signs that are waving in the wind, and so forth.

This fading is manifest in the signal fading from strong to weak signal strength and back again. It also affects the DF results for ES.

8.5.6.2 Fratricide

At all echelons, the staff needs to coordinate closely to avoid friendly communications interference that can occur when using EW systems on the battlefield. Coordination ensures that electronic attack systems, frequencies are properly deconflicted with friendly communications and intelligence systems or that ground maneuver and friendly information tasks are modified accordingly.

The number of information systems, EW systems, and sensors operating simultaneously on the battlefield makes deconfliction with communications systems a challenge. The EWO, the G-2 or S-2, the G-6 or S-6, and the spectrum manager plan and rehearse deconfliction procedures to quickly adjust their use of EW or communications systems.

Electronic attack operations depend on EW support and signals intelligence to provide targeting information and battle damage assessment. However, EWOs must keep in mind that not all intelligence collection is focused on supporting EW. If not properly coordinated with the G-2 or S-2 staff, electronic attack operations may impact intelligence collection by jamming or inadvertently interfering with a particular frequency being used to collect data on the threat, or by jamming a given enemy frequency or system that deprives friendly forces of that means of collecting data. Either can significantly deter intelligence collection efforts and their ability to answer critical information requirements

Other forms of effects rely on electromagnetic spectrum. For example, psychological operations may plan to use a given set of frequencies to broadcast messages, or a military deception plan may include the broadcast of friendly force communications. In both examples, the use of electronic attack could unintentionally interfere or disrupt such broadcasts if not properly coordinated. To ensure electronic attack does not negatively impact planned operations, the EWO coordinates between fires, network operations, and other functional or integrating cells as required.

Like any other form of electromagnetic radiation, electronic attack can adversely affect local media and communications systems and infrastructure. EW planners consider unintended consequences of EW operations and deconflict these operations with the various functional or integrating cells. For example, friendly jamming could potentially deny the functioning of essential services such as ambulance services or firefighters to a local population. EWOs must synchronize electronic attack with the other functional or integrating cells responsible for the information tasks. In this way, they ensure that electronic attack efforts do not cause fratricide or unacceptable collateral damage to their intended effects.

The potential for hostile intelligence collection also affects electronic attack. A well-equipped enemy can detect friendly EW capabilities and thus gain intelligence on friendly force intentions. For example, the frequencies that Army forces jam could indicate where they believe the enemy's capabilities lie. The EWO and the G-2 or S-2 develop an understanding of the enemy's collection

capability. Along with the red team (if available), they determine what the enemy might gain from friendly force use of electronic attack. (A *red team* is an organizational element comprised of trained and educated members that provide an independent capability to fully explore alternatives in plans and operations in the context of the operational environment and from the perspective of adversaries and others.)

The effects of jamming only persist as long as the jammer itself is emitting and is in range to affect the target. Normally this time frame is a matter of seconds or minutes, which makes the timing of such missions critical. This is particularly true when jamming is used in direct support of aviation platforms. For example, in a mission that supports suppression of enemy air defense, the time on target and duration of the jamming must account for the speed of attack of the aviation platform. They must also account for the potential reaction time of enemy air defensive countermeasures. The development of directed-energy weapons may change this dynamic in the future. However, at present (aside from antiradiation missiles), the effects of jamming are less persistent than effects achieved by other means.

8.6 Concluding Remarks

In this chapter we provided a brief overview of the hardware elements in EW system, both ES and EA. These configurations are, of course, notional and many other possible architectures exist.

We concluded the chapter with a discussion of some of the more relevant operational consideration for EW systems. In particular, siting of DF systems is the most critical operational parameter.

References

[1] Poisel, R. A., *Introduction to Communication Electronic Warfare Systems*, 2nd ed., Norwood, MA: Artech House, 2008.

[2] Poisel, R. A., *Modern Communications Jamming Principles and Techniques*, 2nd ed., Norwood, MA: Artech House, 2011.

[3] Adamy, D., *EW101: A First Course in Electronic Warfare*, Norwood, MA: Artech House, 2001.

[4] Adamy, D., *EW102: A Second Course in Electronic Warfare*, Norwood, MA: Artech House, 2004.

[5] Adamy, D., *EW103: Tactical Battlefield Communications Electronic Warfare*, Norwood, MA: Artech House, 2009.

[6] Frater, M. R., and M. Ryan, *Electronic Warfare for the Digitized Battlefield*, Norwood, MA: Artech House, 2001.

[7] Poisel, R. A., *Electronic Warfare Target Location Methods*, 2nd ed., Norwood, MA: Artech House, 2012.

[8] Poisel, R. A., *Antenna Systems and Electronic Warfare Applications*, Norwood, MA: Artech House, 2012, Ch. 21.
[9] Torrieri, D. J., *Principles of Secure Communication Systems*, 2nd ed., Norwood, MA: Artech House, 1992, p. 331.
[10] Poisel, R. A., *Introduction to Communication Electronic Warfare Systems*, 1st ed., Norwood, MA: Artech House, 2002, Ch. 9.
[11] Poisel, R. A., *Modern Communications Jamming Principles and Techniques*, 2nd ed., Norwood, MA: Artech House, 2011, Ch. 17.
[12] Poisel, R. A., *Introduction to Communication Electronic Warfare Systems*, 1st ed., Norwood, MA: Artech House, 2002, Ch. 13
[13] Poisel, R. A., *Modern Communications Jamming Principles and Techniques*, 2nd ed., Norwood, MA: Artech House, 2011, Ch. 8.
[14] Poisel, R. A., *Introduction to Communication Electronic Warfare Systems*, 2nd ed., Norwood, MA: Artech House, 2008, Ch. 8.
[15] FM 3-36, *Electronic Warfare in Operations*, Washington, D.C.: Headquarters Department of the Army, February 26, 2009.

Chapter 9

Electronic Warfare System Performance

9.1 Introduction

We examine some fundamental performance measures of electronic warfare systems in this chapter.[1]

Five separate approaches are pursued. The first examines ES performance based on what is referred to in the information theory literature as a broadcast channel, also referred to as a wiretap channel, where a transmitter is trying to communicate with one or more receivers and an EW system is trying to intercept the transmission. The *measure of effectiveness* (MOE) in this case is the *privacy capacity* of the channel.

The second review considers a jammer's performance when the channel can be accurately modeled as one with AWGN. The MOE in this case is C/W.

The next two investigate the performance of the wiretap channel, and, in particular, what effects ES and EA have on the capacity of the channel. As opposed to the approach mentioned above where C/W is forced to drop below one, these approaches measure the effect on the privacy capacity of the channel, recognizing that some capacity may remain even when intercept and jamming are applied. Both consider the problem of *multiple-input multiple-output* (MIMO) channel performance, where all three nodes in the scenarios have multiple antennas. The first approach divides the EW antennas into two groups, one set for ES and the other for EA. The EW system therefore performs both functions simultaneously. The effects of the number of antennas at the transmitter, receiver, and EW system are examined. The last investigation assumes that the EW system can either intercept or jam, but not both at the same time. In addition, the transmitter can devote some of its assets (antennas and power) to transmitting

[1] For notational convenience, we adopt the following symbology: transmitter = Tx, legitimate receiver = Rx, electronic warfare system = EW, electronic support (intercept) = ES, and electronic attack (jamming) = EA. Note that EW = ES + EA when they are collocated and both functions are represented.

artificial noise (a type of jamming signal) to the EW system. The investigation approach for the last scenario uses results from game theory, where the model is a zero sum game between the transmitter and the EW system and the payoff function is the privacy capacity of the channel.

The last approach considered for evaluating the performance of EW systems applies the concept of an arbitrarily varying wiretap channel to the analysis. Such a channel assumes that there is a jammer impinging on the channel along with an intercept, so both EA and ES capabilities are included. These EA and ES nodes are assumed to be independent in the sense that they do not coordinate their actions when attempting to attack the channel. This situation is a close model to what frequently occurs in practice.

This chapter reviews confidentiality and integrity as two important targets of a communication EW system. To examine how well an EW system can attack these two properties, we adopt some results from the field of communication theory, and, in particular, from the theory of information. Those results are generally couched in terms of how well a communication channel can protect these properties. We turn those thought processes around to see how well an EW system can attack them. That is, how well can the EW cause poor performance of the communication channel?

9.1.1 Confidentiality from Eavesdropping

Wyner [1] proposed the wiretap channel as an information-theoretic model for reliable information transmission over noisy media with confidentiality from eavesdropping. For this model, he coined the term secrecy capacity as a counterpart of channel capacity where the secrecy constraint is also considered. We use the term privacy capacity here, since in EW the term secret means something different. Wyner in his new privacy paradigm asked for a weak asymptotic independence. His main idea was then to exploit the noise of the communication channels along with proper physical layer coding schemes to guarantee privacy against a computationally unlimited eavesdropper. Csiszar and Korner [2] and Leung-Yan-Cheong and Hellman [3] further extended this model to general broadcast channel with confidential messages and Gaussian wiretap channels, respectively. The recent focus of research studies on this physical-layer approach to information-theoretic security stems from the ubiquitous application of wireless communications systems that are highly susceptible to eavesdropping due to their broadcast nature. Examples include the wiretap versions of relay channels [4], fading channels [5, 6], and MIMO channels [7, 8]. An extensive survey of various techniques and results on wiretap channels can be found in [9].

We will investigate wiretap channels in this chapter and examine how effectively an ES system can perform in them. The MOE we will use is the resulting privacy capacity of the channel, with the goal of driving it to zero so that any communications crossing the channel can be intercepted.

9.1.2 Jammer Effects on Communication Reliability

Communication reliability in the presence of jamming is a principal goal of many communication problems, especially for the military. It is frequently modeled as a decision making (detection/estimation) problem in a game-theoretic setting where Tx and Rx try to maximize the reliability performance (measured in terms of probability of correct decision) and the jammer tries to minimize it (measured in terms of probability of error) [10, 11]. Then optimal solutions for this min-max problem are investigated, usually in the form of a saddle point or *Nash equilibrium* (NE). Some have added an information-theoretic flavor to the problem by taking mutual information as the performance metric [12, 13].

An information-theoretic model for reliable communication in the presence of jamming was introduced by Blackwell, Breiman, and Thomasian [14]. In their AVC model, EA has a class of attacks, of which one is selected in a time-varying manner. Tx then attempts to reliably communicate a message to Rx irrespective of the unknown attacks that EA selects in the course of transmission of that message. We will investigate the characteristics of AVCs in this chapter.

> **Notation:** We will use $\mathcal{CN}(\mathbf{0}, \mathbf{Z})$ to denote a circularly symmetric complex Gaussian distribution[2] with zero mean and covariance matrix **Z**. **H** represents channel matrices with entries from $\mathcal{CN}(., .)$. Each entry is a complex number and represents the "gain" of the channel, which can be complex and therefore have effects on the phase. **W** represents channel matrices for DMCs. Each entry is a transition probability from an input symbol to an output symbol. Finally, $H_b(\_)$ denotes the binary entropy function.

9.2 The Wiretap Channel

We examine the wiretap channel model in this section with emphasis on how it applies to the EW function. A wiretap channel is a particular type of broadcast

---

[2] Let $\vec{z} = (z_1, z_2, \ldots, z_n)^T$ be a complex jointly Gaussian random vector. That is, $\Re(z_k)$ and $\Im(z_k)$ for $1 \le k \le n$ comprise a set of $2n$ jointly Gaussian (real) random variables. For many situations where it is useful to view $2n$ jointly Gaussian random variables as a vector of n complex jointly Gaussian random variables, these vectors have an additional property called *circular symmetry*. By definition, z is circularly symmetric if $e^{j\phi}z$ has the same probability distribution as z for all real ϕ. For $n = 1$, that is, for the case where z is a complex Gaussian random variable z, circular symmetry holds if and only if $\Re(z_k)$ and $\Im(z_k)$ are statistically i.i.d. with zero mean, that is, if and only if $\Re(z_k)$ and $\Im(z_k)$ are jointly Gaussian with equi-probability-density contours around 0.

channel (see Section 3.5.8) where one or more of the receivers are ES receivers, and the interest of Tx and Rx is to minimize the information sent ES receivers, while maximizing the communication performance between Tx and Rx. The interest of the EW systems is, of course, to maximize this information exchange and/or minimize the information exchange between the transmitters and target receivers (in the case of EA). The primary difference is that in the wiretap model it is assumed that the communicators are cooperating. In some cases, feedback on the channel state (referred to as *channel state information*, CSI) is even possible from the receivers to the transmitter in order to optimize the transmitted symbols from symbol to symbol. CSI can include the state of interference (either unintentional or intentional, that is, jamming).

Note that the theory about broadcast channels is based on a transmitter sending information to two (or more) receivers. It is generally couched in terms of "desiring" to transmit optimally to the set of receivers. Since an ES receiver is a noncooperative receiver, it is unlikely that a receiver would want to optimize the transmission to this receiver. Alas, the theory does not depend on all the receivers being "friendly." Therefore the results we discuss below based on the broadcast channel theory apply.

9.2.1 Wyner's Wiretap Channel

Wyner was the first to examine the characteristics of wiretap channels [1]. The problem, as stated there, is the following (put into the vernacular we are using):

> Referring to Figure 9.1, the source is discrete and memoryless with entropy H_{M}. The "main channel" and the "wiretap channel" are DMCs with transition probabilities $\mathbf{W}_{\mathrm{Rx}}(\cdot|\cdot)$ and $\mathbf{W}_{\mathrm{ES}}(\cdot|\cdot)$, respectively. The source and the transition probabilities $\mathbf{W}_{\mathrm{Rx}}$ and $\mathbf{W}_{\mathrm{ES}}$ are given and fixed. The encoder is a channel with the k-vector M^k as input and the n-vector X^n as output. The vector X^n is, in turn, the input to the main channel. The main channel output and the wiretap channel input is Y_{Rx}^n. The wiretap channel output is Y_{ES}^n. The decoder associates a k-vector $\hat{M}^k$ with Y_{Rx}^n, and the error probability P_{e} is given by
>
> $$P_{\mathrm{e}} = \frac{1}{k}\sum_{i=1}^{k} \Pr\{M^k \neq \hat{M}^k\} \tag{9.1}$$
>
> The equivocation Δ is given by

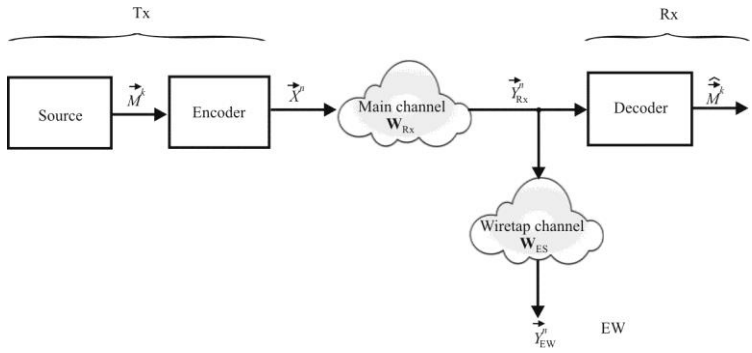


Figure 9.1 Wyner's wiretap channel.

$$\Delta \triangleq \frac{1}{k} H(M^k \mid Y_{\mathrm{ES}}^n) \tag{9.2}$$

and the transmission rate is kH_M/n source bits per channel input symbol. Roughly speaking, a pair (d, R) is achievable if it is possible to find an encoder-decoder with arbitrarily small P_e, and kH_M/n about R, and Δ about d (with perhaps n and k very large). Our main problem is the characterization of the family of achievable (d, R) pairs. It turns out that, in nearly every case, there exists a "privacy capacity," $C_p > 0$, such that (C_p, H_M) is achievable [while for $R > C_p$, (H_M, R) is not achievable]. Thus, it is possible to reliably transmit information at the positive rate C_p in essentially perfect secrecy.

The wiretap channel was modeled as a BSC and therefore adds "noise" to the data sequence. Therefore, the data sequence at EW is degraded from that available to Rx.

Several extensions to Wyner's wiretap channel model have been developed and we will discuss some of them in the remainder of this book. These models form the basis for modeling the ES functions associated with the EW systems we are investigating.

9.2.2 Discrete Memoryless Wiretap Channel

As illustrated in Figure 9.2, a *discrete memoryless wiretap channel* is characterized by a finite input alphabet $\mathcal{X}$, two finite output alphabets $\mathcal{Y}_{Rx}$ and $\mathcal{Y}_{ES}$,

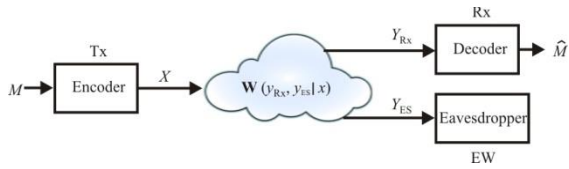


Figure 9.2 Model of the wiretap channel.

and a transition probability matrix $\mathbf{W}(y_{\mathrm{Rx}}, y_{\mathrm{ES}} | x)$ from $X \to Y_{\mathrm{Rx}} \times Y_{\mathrm{ES}}$. In addition, the *n*th extension of the channel law for input $\vec{x} = (x_1, \ldots, x_n) \in X^n$ and outputs $\vec{y}_{\mathrm{Rx}} = (y_{\mathrm{Rx},1}, \ldots, y_{\mathrm{Rx},n}) \in Y_{\mathrm{Rx}}^n$ and $\vec{y}_{\mathrm{ES}} = (y_{\mathrm{ES},1}, \ldots, y_{\mathrm{ES},n}) \in Y_{\mathrm{ES}}^n$ is characterized as

$$\mathbf{W}^n(y_{\mathrm{Rx}}, y_{\mathrm{ES}} | x) \triangleq \prod_{i=1}^{n} \mathbf{W}(y_{\mathrm{Rx},i}, y_{\mathrm{ES},i} | x_i) \tag{9.3}$$

This channel is depicted in Figure 9.3. The node observing the output $\vec{y}_{\mathrm{Rx}} \in \mathcal{Y}_{\mathrm{Rx}}^n$ is Rx and the marginal channel $\mathbf{W}^n(\vec{y}_{\mathrm{Rx}} | \vec{x}) = \sum_{\vec{y}_{\mathrm{ES}} \in Y_{\mathrm{ES}}} \mathbf{W}^n(\vec{y}_{\mathrm{Rx}}, \vec{y}_{\mathrm{ES}} | \vec{x})$ is referred to as the *main channel*, while the node observing the output $\vec{y}_{\mathrm{ES}} \in Y_{\mathrm{ES}}^n$ is EW (ES) and the marginal channel $\mathbf{W}^n(\vec{y}_{\mathrm{ES}} | \vec{x}) = \sum_{\vec{y}_{\mathrm{Rx}} \in Y_{\mathrm{Rx}}} \mathbf{W}^n(\vec{y}_{\mathrm{Rx}}, \vec{y}_{\mathrm{ES}} | \vec{x})$ is referred to as the *intercept channel*.

Secure communication over wiretap channels is realized by using wiretap coding schemes according to the following definition.

> **Definition 9.1:** An $(n; \mathcal{M})$ *wiretap code* consists of a message set $\mathcal{M} = \{1, 2, \ldots, \mathcal{M}\}$, a stochastic encoder $f: \mathcal{M} \to \mathcal{X}^n$, and a deterministic decoder $\phi: \mathcal{Y}_{\mathrm{Rx}}^n \to \mathcal{M}$.

By *stochastic encoding*, we mean that the codeword $\vec{x}$ associated with the message

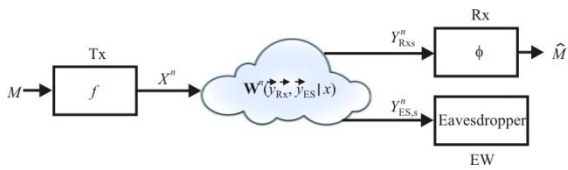


Figure 9.3 Extended model of wiretap channel.

m is selected randomly according to some conditional probability $p(\vec{x}\,|\,m)$ and may change for different uses of the channel.

Now, we assess the performance of the wiretap code (f, ϕ) over the wiretap channel $\mathbf{W}^n(\vec{y}_{\text{Rx}}, \vec{y}_{\text{ES}}\,|\,\vec{x})$. Assume that the message m is selected uniformly at random from the message set $\mathcal{M}$, encoded for transmission as X^n, and received at Rx and ES as Y_{Rx}^n and Y_{ES}^n, respectively.

Definition 9.2: The *reliability performance* of the wiretap code is given by the *error probability*

$$\overline{e}(\mathbf{W}^n, f, \phi) \triangleq \frac{1}{\mathcal{M}} \sum_{m \in \mathcal{M}} \sum_{y_{\text{ES}} \in Y_{\text{ES}}^n} \mathbf{W}^n\{[\phi^{-1}(m)]^c, y_{\text{ES}}\,|\,f(m)\} \tag{9.4}$$

Definition 9.3: The *privacy performance* of the wiretap code is given by the *leakage rate*

$$L(\mathbf{W}^n, f, \phi) \triangleq \frac{1}{n} I(M; Y_{\text{ES}}^n) \tag{9.5}$$

Definition 9.3 says that the confidentiality of message m is measured in terms of the amount of information gleaned by ES through the wiretapping observation Y_{ES}^n, or since $H(m)$ is fixed, the amount of ES's uncertainty $H(m\,|\,Y_{\text{ES}}^n)$ about the message m after observing Y_{ES}^n.

Ideally, the communicator wants the wiretap code to convey some information rate with vanishing error probability and vanishing leakage rate. These conditions will mean that Rx recovers an almost error-free version of the message, while ES observes a channel output almost independent of the message. In the following, we make this statement precise. For ES functionality, the primary goal is to maximize the leakage rate while the primary EA goal is to maximize the error probability. We shall see that these two goals are contradictory.

9.2.3 Privacy Capacity

The degree to which a channel can provide privacy in communications is indicated by its privacy capacity. More formally we have:

Definition 9.4: A privacy rate R_p is called achievable for the wiretap channel $\mathbf{W}^n(\vec{y}_{Rx}, \vec{y}_{ES} | \vec{x})$ if for every $\varepsilon > 0$ there exists an $(n, 2^{nR_p})$ wiretap code[3] such that

$$\bar{e}(\mathbf{W}^n, f, \phi) \le \varepsilon \tag{9.6}$$

$$L(\mathbf{W}^n, f, \phi) \le \varepsilon \tag{9.7}$$

The *privacy capacity* C_p of the wiretap channel $\mathbf{W}^n(\vec{y}_{Rx}, \vec{y}_{ES} | \vec{x})$ is defined as the supremum of all achievable privacy rates.

The privacy capacity is a counterpart of the more common point-to-point channel capacity; the latter is concerned with the reliability of communication, while the former also accounts for the privacy. In addition, we notice that the privacy capacity depends on the joint transition probability $\mathbf{W}^n(\vec{y}_{Rx}, \vec{y}_{ES} | \vec{x})$ only through its marginal transition probabilities since the reliability constraint (9.6) only concerns the main channel and the privacy constraint (9.7) only involves the intercept channel.

We are now ready to examine the privacy capacity of wiretap channels. The first privacy capacity is for the special degraded case, in which $X \to Y_{Rx} \to Y_{ES}$ forms a Markov chain,[4] as Wyner initially considered [1].

Property 9.1: The privacy capacity of the discrete memoryless degraded wiretap channel is

$$C_p = \max_{P_X(x)} \left[I(X; Y_{Rx}) - I(X; Y_{ES}) \right] \tag{9.8}$$

The privacy capacity for a general (possibly non-degraded) wiretap channel was later established by Csiszar and Korner [2] as follows.

---

[3] $(n, 2^{nR_s})$ is shorthand notation for $(n, \lceil 2^{nR_s} \rceil)$.

[4] For $n = 3,4,5,\ldots$, we say that the sequence of random variables $\{X_i\}_{i=1}^n$ is a Markov chain if $(X_1, X_2, \ldots, X_{j-1})$, and $(X_{j+1}, \ldots, X_n)$ are conditionally independent, given X_j $(1 < j < n)$. We make repeated use of the fact that, if X_1, X_2, X_3 is a Markov chain, then

$$H(X_3 | X_1, X_2) = H(X_3, X_2)$$

When X, Y_s, Z_s form a Markov chain, then

$$p(y_s, z_s | x) = p(y_s | x) p(z_s | y_s)$$

with $y_s \in Y_s$, $z_s \in Z_s$, and $x \in X$.

Property 9.2: The privacy capacity of the discrete memoryless wiretap channel is

$$C_{\mathrm{p}} = \max_{P(u,x)} [\mathbb{I}(U;Y_{\mathrm{Rx}}) - \mathbb{I}(U;Y_{\mathrm{ES}})] \tag{9.9}$$

where U is an auxiliary random variable satisfying the Markov chain $U \to X \to Y_{\mathrm{Rx}} Y_{\mathrm{ES}}$.

These properties suggest that the amount of privacy obtained over a wiretap channel is related to the "excess information" that Rx collects about the message relative to ES, or equivalently, the "excess noise" that the intercept channel exhibits relative to the main channel. We see from (9.9), if ES has the same observation as Rx, that is, $Y_{\mathrm{Rx}} = Y_{\mathrm{ES}}$ so that $\mathbb{I}(U;Y_{\mathrm{Rx}}) = \mathbb{I}(U;Y_{\mathrm{ES}})$, then the privacy capacity is zero.

The following example illustrates these ideas.

Example 9.1: For a wiretap channel with $\mathcal{X} = \mathcal{Y}_{\mathrm{Rx}} = \{0,1\}$, let the main and intercept channels be BSCs with crossover probabilities p and q, respectively. If $p < q$, the intercept channel is stochastically degraded[5] with respect to the main channel; thus, the privacy capacity is positive and equal to the difference of the channel capacities

$$C_{\mathrm{p}} = [1 - H_{\mathrm{b}}(p)] - [1 - H_{\mathrm{b}}(q)] = H_{\mathrm{b}}(q) - H_{\mathrm{b}}(p) \tag{9.10}$$

If $p \geq q$, the main channel is stochastically degraded with respect to the intercept channel; in that case, Rx has no advantage over ES and the privacy capacity is zero.

9.3 Arbitrarily Varying Channels

In this section we consider AVCs and investigate some of their properties. The AVC is the channel model we use later to examine the effects of jamming on the discrete, memoryless communication channel.

[5] A broadcast (and specifically a wiretap) channel is said to be *stochastically degraded* if its conditional marginal distributions are the same as that of a physically degraded one, that is, if there exists a distribution $p'(z|y)$ such that $p(z|x) = \sum_{y \in \mathcal{Y}} p'(z|y)\, p(y|x)$.

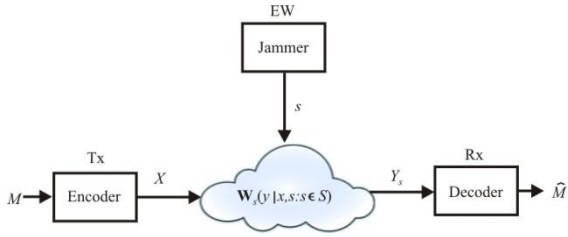


Figure 9.4 Model of arbitrarily varying channel.

9.3.1 Arbitrarily Varying Channels

A discrete memoryless AVC, depicted in Figure 9.4, is characterized by a finite input alphabet $\mathcal{X}$, a finite output alphabet $\mathcal{Y}$, an arbitrary state space S,[6] and a family of transition probability matrices from X to Y indexed by S:

$$\mathcal{W} = \left\{ \mathbf{W}_s(y|x) \triangleq \mathbf{W}(y|x;s) : s \in S \right\} \tag{9.11}$$

In addition, the nth extension of the channel law for input $\vec{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$, output $\vec{y} = (y_1, \ldots, y_n) \in \mathcal{Y}^n$, and state sequence $\vec{s} = (s_1, \ldots, s_n) \in \mathcal{S}^n$, is characterized as

$$\mathbf{W}_s^n(\vec{y}|\vec{x}) \triangleq \prod_{i=1}^{n} \mathbf{W}_{s_i}(y_i|x_i) = \prod_{i=1}^{n} \mathbf{W}(y_i|x_i;s_i) \tag{9.12}$$

Such a scenario is depicted in Figure 9.5. Notice that the output is related to the input without memory, whereas the channel state is arbitrarily selected, without any presumed a priori distribution, and possibly with memory. In addition, the jammer selects the attack sequence $\vec{s}$ without knowledge of the transmitted message m (as opposed to when the jammer has an associated ES capability so that coherent jamming is possible). Lastly, we assume that the Tx and Rx are aware of the state space $\mathcal{S}$, but not the actual state sequence $\vec{s}$.

There are two concepts that characterize the behavior of an AVC. The first is the notion of "averaged states" and the associated convex closure of these states. The second is when EA can mimic Tx to the point that Rx cannot tell them apart. We discuss these concepts here.

[6] In theory, the state space need not originate from a jammer. For our purposes, we assume that it does.

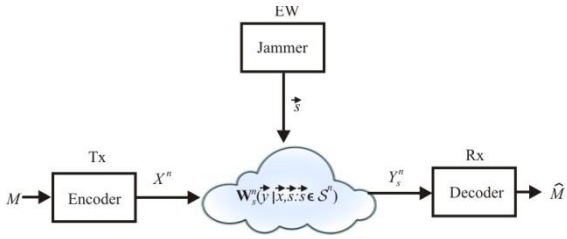


Figure 9.5 Extended AVC model.

9.3.1.1 Averaged States

Since the channel is memoryless, we expect the behavior of an AVC to be related to the number of times each individual attack is imposed, and not to their ordering. Thus we can define "averaged" states and the convex closure of an AVC.

Definition 9.5: For any number $r \in \mathbb{N}$, any set of states $\{s_k \in S\}_{k=1}^{r}$, and any assigned probability law $\{p(s_k)\}_{k=1}^{r}$ with $p(s_k) \geq 0$ and $\sum_{k=1}^{r} p(s_k) = 1,$, the associated *averaged state* $\bar{s}$ is defined as[7]

$$\bar{s} = \sum_{k=1}^{r} s_k \, p(s_k) \tag{9.13}$$

The *averaged state space* $\bar{S}$ is defined as the closure[8] of the set of all such averaged states

$$\bar{\mathcal{S}} = \mathrm{cl}\left(\left\{ \bar{s} = \sum_{k=1}^{r} s_k \, p(s_k) : r \in \mathbb{N}, s_k \in \mathcal{S}, p(s_k) \geq 0, \sum_{k=1}^{r} p(s_k) = 1 \right\} \right) \tag{9.14}$$

Example 9.2: Suppose $\mathcal{S} = \{0, 1, 2\}$, and $p(0) = p(1) = p(2) = 1/3$. Then $s_1 = 0$, $s_2 = 1$, and $s_3 = 2$ while:

when $r = 1$

$$\bar{s} = \sum_{k=1}^{1} s_k \, p(s_k) = 0 \times 1/3 = 0; 1 \times 1/3 = 1/3; 2 \times 1/3 = 2/3$$

[7] The set of natural numbers, $\mathbb{N} = \{1, 2, \ldots\}$.

[8] Given any subset, *A*, of a vector space *X*, the smallest closed set containing *A* is called the *closure* of *A* and is denoted by $\bar{A}$ or cl(*A*).

when $r = 2$

$$\bar{s} = \sum_{k=1}^{2} s_k\, p(s_k) = 0\times 1/3 + 1\times 1/3 = 1/3;$$

$$0\times 1/3 + 2\times 1/3 = 2/3; 1\times 1/3 + 2\times 1/3 = 1$$

when $r = 3$

$$\bar{s} = \sum_{k=1}^{3} s_k\, p(s_k) = 0\times 1/3 + 1\times 1/3 + 2\times 1/3 = 1$$

so $\bar{\mathcal{S}} = \{0, 1/3, 2/3, 1\}$.

Any given state sequence $\vec{s} \in \mathcal{S}^n$ imposes a probability distribution on the state space $\mathcal{S}$. Therefore, we can use the above definition to associate some averaged state $\bar{s}$ to $\vec{s}$. Accordingly, to the channel $\mathbf{W}_{\vec{s}}^{n}(.,.|.)$, we can associate the nth extension of the averaged channel $\mathbf{W}_{\bar{s}}(.,.|.)$. Thus, we define:

Definition 9.6: The *convex closure* of the AVC $\mathcal{W}$ is defined as the closure of the set of all averaged channels

$$\bar{\mathcal{W}} = \mathrm{cl}\left(\left\{\mathbf{W}_{\bar{s}}(\vec{y}|\vec{x}) = \sum_{k} p(s_k)\mathbf{W}_{s_k}(\vec{y}|\vec{x}) : \bar{s} = \sum_{k} s_k\, p(s_k) \in \bar{\mathcal{S}}\right\}\right) \qquad (9.15)$$

We will refer to the AVC $\mathcal{W}$ also as the AVC $\mathcal{S}$ and the convex closure $\bar{\mathcal{W}}$ also as $\bar{\mathcal{S}}$, provided there is no ambiguity.

9.3.1.2 Symmetrizability

Another important concept for AVCs is that of *symmetrizability* [15]. Consider the case with $\mathcal{S} = \mathcal{X}$, and assume the condition $\mathbf{W}_s(y|x) = \mathbf{W}_x(y|s)$ holds for all $x \in \mathcal{X}, y \in \mathcal{Y}, s \in \mathcal{S}$. In such a case, Rx cannot distinguish between whether the letter x is sent by Tx and the attack s is applied by EA and whether the letter s is sent by Tx and the attack x is applied by EA. Therefore, EA is capable of forging the operations of Tx for all input letters in all states, with the result that Rx will always make a mistake in detecting the message. This situation can be generalized to the case in which $\mathcal{S} \neq \mathcal{X}$ by allowing EA to excite the AVC via some auxiliary channel $U : \mathcal{X} \to \mathcal{S}$.

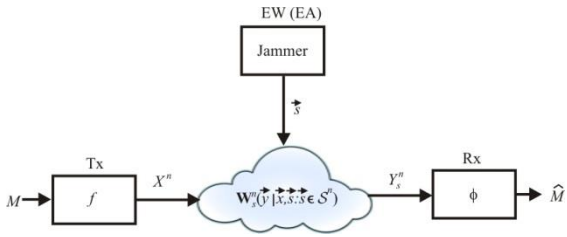


Figure 9.6 Arbitrarily varying channel with a deterministic code.

Definition 9.7: An AVC is called *symmetrizable* if there exists some stochastic matrix $\mathbf{U}(s|x)$ such that

$$\sum_{s\in\mathcal{S}} \mathbf{W}_s(y|x)\,\mathbf{U}(s|x') = \sum_{s\in\mathcal{S}} \mathbf{W}_s(y|x')\,\mathbf{U}(s|x), \quad \forall (x,x',y)\in X\times X\times Y \tag{9.16}$$

and is called *nonsymmetrizable* if no such $\mathbf{U}(s|x)$ exists.

We can see that a symmetric AVX is symmetrizable, that is, the number of its inputs and states are both equal to r and for all $1\le i, j\le r$ with $i\ne j$, the ith row of the jth matrix is identical to the jth row of the ith matrix; especially, for an AVC with two inputs and two states, if the AVC is represented by

$$\mathcal{W}=\left\{\begin{bmatrix} p & 1-p \\ q & 1-q \end{bmatrix}, \begin{bmatrix} q & 1-q \\ r & 1-r \end{bmatrix}\right\} \tag{9.17}$$

or

$$\mathcal{W}=\left\{\begin{bmatrix} p & 1-p \\ q & 1-q \end{bmatrix}, \begin{bmatrix} r & 1-r \\ p & 1-p \end{bmatrix}\right\} \tag{9.18}$$

for any $0\le p,q,r\le 1$.

9.3.2 Coding Scheme

We consider the common deterministic code for the AVC, illustrated in Figure 9.6, that assigns a unique codeword to each message, and the usual average error probability can be used to assess its performance. A formal definition of this coding scheme is the following.

Definition 9.8: An $(n, \mathcal{M})$ *deterministic code* consists of a message set $\mathcal{M} = \{1, \ldots, \mathcal{M}\}$ a deterministic encoder $f : \mathcal{M} \to \mathcal{X}^n$ and a deterministic decoder $\phi : \mathcal{Y}^n \to \mathcal{M}$.

To assess the behavior of a deterministic code (f, ϕ) over the AVC $\mathcal{S}$, we assume that the message m is selected uniformly at random from the message set $\mathcal{M}$, encoded for transmission as X^n, and received at Rx Y_s^n. Then the reliability performance of the deterministic code is given by the error probability averaged over the messages

$$\bar{e}(\mathbf{W}_s^n, f, \phi) = \frac{1}{\mathcal{M}} \sum_{m \in \mathcal{M}} \mathbf{W}_s^n \{[\phi^{-1}(m)]^c \mid f(m)\} \tag{9.19}$$

where ^{c} represents complement; that is, the output is not what it should be given that $f(m)$ was transmitted. We notice that the error probabilities defined in (9.19) are a function of the state sequence $\vec{s}$.

Reliable communication over an AVC requires that reliability of the message be guaranteed regardless of the particular attack $\vec{s}$ of EA. Therefore, a good code for an AVC must have a vanishing error probability for all state sequences, as captured by the following definition.

Definition 9.9: A code rate R is called *achievable* for the AVC $\mathcal{S}$ if for every $\varepsilon > 0$ there exists an $(n, 2^{nR})$ code such that

$$\bar{e}(\mathbf{W}_s^n, f, \phi) \le \varepsilon \qquad \forall \vec{s} \in \mathcal{S}^n \tag{9.20}$$

The *code capacity* C of the AVC $\mathcal{S}$ is the supremum of all achievable code rates.

9.3.3 AVC Capacities

The fundamental capacity limit for AVCs is that of Ericson, and Csiszar and Narayan on the deterministic-code capacity of an AVC given by [15, 16].

Property 9.3: The code capacity of the discrete memoryless AVC $\mathcal{S}$ is given by:

$$C_{\mathrm{p}} = \begin{cases} 0, & \text{iff symmetrizable} \\ \max\limits_{P_X(x)} \min\limits_{\bar{s}\in\bar{\mathcal{S}}} \mathbb{I}(X;Y_{\bar{s}}) = \min\limits_{\bar{s}\in\bar{\mathcal{S}}} \max\limits_{P_X(x)} \mathbb{I}(X;Y_{\bar{s}}), & \text{iff nonsymmetrizable} \end{cases} \tag{9.21}$$

Property 9.3 shows that the capacity of an AVC is characterized by the worst (for the communicator, best for the jammer) averaged channel. We can interpret this by assuming that the application of any state sequence $\vec{s} \in \mathcal{S}^n$ by EA effectively gives rise to an equivalent discrete memoryless channel characterized by the corresponding averaged state $\bar{s}$, namely the averaged channel $\mathbf{W}_{\bar{s}}^n(\vec{y}|\vec{x})$. Because the reliability of communication must be guaranteed for all state sequences, the worst averaged channel must be used to determine the channel capacity.

Example 9.3: The XOR AVC is defined by $\mathcal{X} = \mathcal{Y} = \mathcal{S} = \{0,1\}$, and the family of transition probability matrices

$$\mathcal{W} = \{\mathbf{W}_1, \mathbf{W}_2\} = \left\{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right\} \tag{9.22}$$

This AVC is called the XOR AVC because it can be described by the deterministic relationship $Y = X \oplus S$. The capacity of this channel in both states $s = 0$ and $s = 1$ is equal to 1. However, any channel in the convex closure $\bar{\mathcal{S}} = [0,1]$ of this AVC is a BSC with crossover probability $p = \bar{s}$ and capacity $1 - H_{\mathrm{b}}(\bar{s})$; in particular, the averaged channel corresponding to the averaged state $\bar{s} = 1/2$ is a BSC with cross-over probability $p = 1/2$ and zero capacity. Therefore, the code capacity is zero.

Example 9.4: Consider the AVC defined by $\mathcal{X} = \mathcal{S} = \{0,1\}$, $\mathcal{Y} = \{0,1,2\}$, and the family of transition probability matrices

$$\mathcal{W} = \left\{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right\} \tag{9.23}$$

This AVC is known as the "additive" AVC since it can be described by the deterministic relationship $Y = X + S$. Any channel in the convex closure of this AVC is given by the transition probability matrix

$$\mathbf{W} = \begin{bmatrix} 1-p & p & 0 \\ 0 & 1-p & p \end{bmatrix} \tag{9.24}$$

where $p \in \bar{\mathcal{S}} = [0,1]$. Assigning a probability distribution (q, 1 – q) to the binary input, we can use (9.21) to conclude that

$$C = \max_{q \in 0,1]} \min_{p \in [0,1]} H[(1-q)(1-p), q * p, qp] = 1/2 \tag{9.25}$$

where $a * b \triangleq a(1-b) + b(1-a)$ and $H(\vec{p}) \triangleq -\sum_{i=1}^{m} p_i \log_2 p_i$ for a probability distribution $\vec{p} = (p_1, \ldots, p_m)$. The deterministic-code capacity of this AVC is zero since it is clearly symmetrizable.

9.4 Electronic Support Performance

The ES problem can be modeled as a wiretap channel problem in that there is a single transmitter broadcasting to (possibly several) other users on the target network. The ES receiver is also listening, but this is typically unknown to Tx. The ES receiver is a noncooperating participant to the network exchanges. The Tx would not normally attempt to maximize the channel throughput to the ES, along with the other receivers.

The scenario under consideration is shown in Figure 9.7. Tx is attempting to communicate with Rx over the channel. ES is attempting to recover the messages being sent over the communication channel. Both of these links are corrupted by noise. This noise can be of several varieties including thermal noise (caused by heating of the atmosphere by the sun), manmade noise, atmospheric noise, and others [17]. We will consider only thermal noise here as that is the most amenable to theoretical analysis. The other forms of noise are important but such problems are scenario/situation dependent and are usually addressed by simulation and experimental measurements.

We show a ground-to-ground link in Figure 9.7 [18], but air-to-air and air-to-ground scenarios are just as important, and in many cases more so. The only real difference is in the signal propagation characteristics, which we will discuss.

The communication link shown in Figure 9.7 is referred to in the information theory field as a broadcast channel, because, even though there is only one target receiver shown in Figure 9.7, in reality in most tactical C2 scenarios a single push-to-talk transmitter is sending information to several receivers on the network. It thus is “broadcasting” to several receivers (including the intercept receiver which the transmitter may or may not know is present). The transmitter/receivers do, or at least can, cooperate with one another to improve the network communication

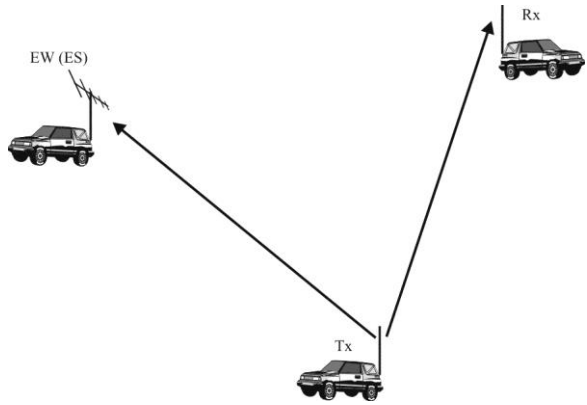


Figure 9.7 ES scenario. (Source: [18], © Artech House, 2008. Reprinted with permission.)

exchange performance. However, the transmitter/intercept link is a noncooperating link and no such cooperation can be expected.

9.4.1 ES Performance—Privacy Capacity

Barros and Rodrigues provided an analysis of whether messages sent over a channel were intereceptable or not [19]. Their approach was based on the wiretap channel first proposed by Wyner [1] and further developed by Csiszar and Korner [2]. The basic principle of *information-theoretic security*—widely accepted as the strictest notion of security—calls for the combination of cryptographic schemes with channel coding techniques that exploit the randomness of the communication channels at the physical level (see Chapter 6 for a discussion of networking) to guarantee that the sent messages cannot be decoded by a third party maliciously eavesdropping on the wireless medium.

We examine the interceptability issue here based on the concept of the wiretap channel but do not address the cryptographic issues involved. We are concerned with quantifying whether messages can be intercepted by an ES system based on information-theoretic concepts; that is, whether the capacity of the channel has an impact on the interceptability of messages or not. We will show that the answer to this question is affirmative, and it is possible to provide levels of capacity over the wiretap channel that essentially preclude private communication over the target communication channel.

As discussed in Section 9.2, in the wiretap channel proposed by Wyner, Tx and Rx communicate over a main channel and ES has access to degraded versions of the channel outputs that reach Rx. In [3] it was shown that if both the main channel and the intercept channel are AWGN channels, and the latter has less capacity than the former, the privacy capacity (that is the maximum transmission

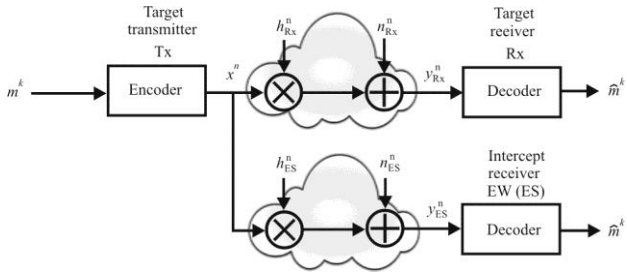


Figure 9.8 Model of the scenario depicted in Figure 9.1.

rate at which ES is unable to decode any information) is equal to the difference between the two channel capacities.

9.4.1.1 ES Model

Consider the scenario illustrated in Figure 9.7. Tx wants to send messages m to the target receiver (Rx). An information-theoretic model of this configuration is depicted in Figure 9.8.

The message block m^k is encoded into the codeword $\vec{x}^n = (x_1, \ldots, x_n)$ to be transmitted over a discrete-time Rayleigh fading channel (the main channel) with output

$$y_{\mathrm{Rx},i} = h_{\mathrm{Rx},i} x_i + n_{\mathrm{Rx},i}$$

where $h_{\mathrm{Rx},i}$ is the (possibly time-varying) complex fading coefficient and $n_{\mathrm{Rx},i}$ denotes the zero-mean circularly symmetric complex Gaussian noise. The coefficient $h_{\mathrm{Rx},i}$, also referred to as CSI, is independent from the channel output and assumed to be drawn i.i.d. according to the pdf $p(h_{\mathrm{Rx}})$, which is zero-mean complex Gaussian for Rayleigh fading. We assume quasi-static fading, that is, the fading coefficients are constant for all uses of the channel (or, equivalently, for all time), that is, $h_{\mathrm{Rx},i} = h_{\mathrm{Rx}}$, $\forall i$.

ES is capable of intercepting the signals sent by Tx by observing the channel output

$$y_{\mathrm{ES},i} = h_{\mathrm{ES},i} x_i + n_{\mathrm{ES},i}$$

of an independent Rayleigh fading channel, with quasi-static fading coefficient $h_{\mathrm{ES},i} = h_{\mathrm{ES}}$, $\forall i$, and zero-mean circularly symmetric complex Gaussian noise $n_{\mathrm{ES},i}$.

The channel is power limited in the sense that

$$\frac{1}{n}\sum_{i=1}^{n}\mathcal{E}\{|X_i|^2\} \le P_{\text{Tx}} \tag{9.26}$$

where P_{Tx} corresponds to the average transmit signal power. Furthermore, we denote the power of the noise in the main channel and the intercept channel as N_{Rx} and N_{ES}, respectively. The instantaneous SNR at Rx is thus given by

$$\gamma_{\text{Rx},i} = \frac{|h_{\text{Rx},i}|^2 P_{\text{Tx}}}{N_{\text{Rx}}} = \frac{|h_{Rx}|^2 P_{\text{Tx}}}{N_{\text{Rx}}} = \gamma_{\text{Rx}}$$

and the average SNR is

$$\overline{\gamma}_{\text{Rx},i} = \frac{\mathcal{E}\{|h_{\text{Rx},i}|^2\}P}{N_{\text{Rx}}} = \frac{\mathcal{E}\{|h_{\text{Rx}}|^2\}P_{\text{Tx}}}{N_{\text{Rx}}}\overline{\gamma}_{\text{Rx}}$$

Likewise, the instantaneous SNR at ES is given by

$$\gamma_{\text{ES},i} = \frac{|h_{\text{ES},i}|^2 P_{\text{Tx}}}{N_{\text{ES}}} = \frac{|h_{\text{ES}}|^2 P_{\text{Tx}}}{N_{\text{ES}}} = \gamma_{\text{ES}}$$

and the average SNR is

$$\overline{\gamma}_{\text{ES},i} = \frac{\mathcal{E}\{|h_{\text{ES},i}|^2\}P_{\text{Tx}}}{N_{\text{ES}}} = \frac{\mathcal{E}\{|h_{\text{ES}}|^2\}P_{\text{Tx}}}{N_{\text{ES}}}\overline{\gamma}_{\text{ES}}$$

Since the channel fading coefficients h are zero-mean complex Gaussian random variables [20] and the instantaneous SNR, $\gamma \propto |h|^2$, the pdfs of γ_{Rx} and γ_{ES} are exponentially distributed, given by

$$p(\gamma_{\text{Rx}}) = \frac{1}{\overline{\gamma}_{\text{Rx}}}\exp\left(-\frac{\gamma_{\text{Rx}}}{\overline{\gamma}_{\text{Rx}}}\right) \tag{9.27}$$

and

$$p(\gamma_{\text{ES}}) = \frac{1}{\overline{\gamma}_{\text{ES}}}\exp\left(-\frac{\gamma_{\text{ES}}}{\overline{\gamma}_{\text{ES}}}\right) \tag{9.28}$$

The transmission rate between Tx and Rx is $R = H(m^k)/n$ and the error probability is defined as $P_\varepsilon^k = \Pr\{m^k \neq \hat{m}^k\}$, where $\hat{m}^k$ denotes Rx's estimate of the sent messages. We are interested in minimizing ES's uncertainty about *m*, that is the equivocation rate

$$\Delta = \frac{H(m^k \left| Y_{\text{ES}}^n \right.)}{H(m^k)} \tag{9.29}$$

We say that (R', d') is *achievable* if for all $\varepsilon > 0$ there exists an encoder-decoder pair such that $R \geq R' - \varepsilon, \Delta \geq d' - \varepsilon,$ and $P_\varepsilon^k \leq \varepsilon.$ Our goal is to characterize the privacy capacity C_{p} defined as the maximum transmission rate *R* at $\Delta = 1.$

We will assume henceforth that Tx and Rx have perfect CSI about the main channel but no CSI about the intercept channel. ES, in turn, has CSI on the intercept channel.[9]

9.4.1.2 Privacy Capacity of Quasi-Static Rayleigh Fading Channels

This section characterizes the privacy capacity of a quasi-static Rayleigh fading channel in terms of intercept probability. First, we consider a single realization of the fading coefficients and compute its privacy capacity. Then we discuss the existence of (strictly positive) privacy capacity in the general case, and characterize the intercept probability and the intercept privacy capacity.

Single Realization

We start by deriving the privacy capacity for one realization of a pair of quasi-static fading channels with complex noise and complex fading coefficients. We take the relative distances between Tx and Rx and between Tx and ES into consideration since EW could be closer to Tx than Rx or further away. The path losses on the paths have an effect on the amount of received power at both receivers and therefore the SNR values are affected.

Suppose that both the main and the intercept channel are complex AWGN channels, that is, transmit and receive symbols are complex and both additive noise processes are zero mean circularly symmetric complex Gaussian. The power of the complex input *X* is constrained according to (9.26). Since each use of the complex AWGN channel can be viewed as two uses of a real-valued AWGN channel [21], the privacy capacity of the complex intercept channel is

[9] Note that CSI on the main channel is of no use to ES, because the equivocation rate depends only on the output of the wiretap channel.

$$C_{\mathrm{p}} = C_{\mathrm{Rx}} - C_{\mathrm{ES}} \tag{9.30}$$

Assume that the nodes are static over the time interval of concern. As indicated above, we assume that the fading is constant (these are actually related concepts). h_{Rx} is the path loss coefficient that includes the effects of fading as well as distance between Tx and Rx. Likewise, h_{ES} is the path loss coefficient that includes the effects of fading as well as distance between Tx and EW. Since the nodes are static, h_{Rx} and h_{ES} are quasi-static Gaussian random variables that we can consider constants over our interval.

We know that

$$C_{\mathrm{Rx}} = \log_2(1+\gamma_{\mathrm{Rx}}) \tag{9.31}$$

and

$$\gamma_{\mathrm{Rx}} = \left|h_{\mathrm{Rx}}\right|^2 \frac{P_{\mathrm{Tx}}}{N_{\mathrm{Rx}}} \tag{9.32}$$

so that

$$C_{\mathrm{Rx}} = \log_2\left(1+\left|h_{\mathrm{Rx}}\right|^2 \frac{P_{\mathrm{Tx}}}{N_{\mathrm{Rx}}}\right) \tag{9.33}$$

Similarly, we have

$$C_{\mathrm{ES}} = \log_2\left(1+\left|h_{\mathrm{ES}}\right|^2 \frac{P_{\mathrm{Tx}}}{N_{\mathrm{ES}}}\right) \tag{9.34}$$

Based on (9.30)–(9.34) and that the channel capacity must be nonnegative, we may write the privacy capacity for one realization of the quasi-static fading scenario as

$$C_{\mathrm{p}} = \begin{cases} \log_2(1+\gamma_{\mathrm{Rx}}) - \log_2(1+\gamma_{\mathrm{ES}}), & \gamma_{\mathrm{Rx}} > \gamma_{\mathrm{ES}} \\ 0, & \gamma_{\mathrm{Rx}} \le \gamma_{\mathrm{ES}} \end{cases} \tag{9.35}$$

ES Effects on Privacy Capacity

We will now consider the existence of a nonzero private capacity between Tx and Rx, which is the goal that ES is attempting to negate. As explained above, for

specific fading realizations, the main channel (from Tx to Rx) and the intercept channel (from Tx to ES) can be viewed as complex AWGN channels with SNR γ_{Rx} and γ_{ES}, respectively. Moreover, from (9.35) we know that the privacy capacity is positive when $\gamma_{Rx} > \gamma_{ES}$ and is zero when $\gamma_{Rx} \leq \gamma_{ES}$. Since the main channel and the intercept channel are independent and knowing that the random variables γ_{Rx} and γ_{ES} are exponentially distributed with pdfs given by (9.27) and (9.28), respectively, we may write the probability of a nonzero privacy capacity as

$$\begin{aligned}\Pr\{C_p > 0\} &= \Pr\{\gamma_{Rx} > \gamma_{ES}\} \\ &= \int_0^{\infty}\int_0^{\gamma_{Rx}} p(\gamma_{Rx}, \gamma_{ES}) d\gamma_{ES} d\gamma_{Rx} \\ &= \int_0^{\infty}\int_0^{\gamma_{Rx}} p(\gamma_{Rx})\, p(\gamma_{ES}) d\gamma_{ES} d\gamma_{Rx} \\ &= \frac{\overline{\gamma}_{Rx}}{\overline{\gamma}_{Rx} + \overline{\gamma}_{ES}}\end{aligned} \tag{9.36}$$

Considering the point node locations, note that $\overline{\gamma}_{Rx} \propto 1/d_{Tx,Rx}^{\alpha}$ and $\overline{\gamma}_{ES} \propto 1/d_{Tx,EW}^{\alpha}$ where $d_{Tx,Rx}$ is the distance between Tx and Rx, $d_{Tx,EW}$ is the distance between Tx and EW, and α is the path loss exponent[10] [22], the probability in (9.36) is given by

$$\Pr\{C_p > 0\} = \frac{1}{1 + (d_{Tx,Rx}/d_{Tx,EW})^{\alpha}} \tag{9.37}$$

Note that when $\gamma_{Rx} \gg \gamma_{ES}$ (or $d_{Tx,EW} \ll d_{Tx,Rx}$) then $\Pr\{C_p > 0\} \approx 1$ (or $\Pr\{C_p = 0\} \approx 0$). Conversely, when $\gamma_{ES} \gg \gamma_{Rx}$ (or $d_{Tx,EW} \ll d_{Tx,Rx}$), then $\Pr\{C_p > 0\} \approx 0$ (or $\Pr\{C_p = 0\} \approx 1$).

To guarantee a zero privacy capacity with probability greater than p_0, then from (9.36) and (9.37), we require that

$$\frac{\overline{\gamma}_{Rx}}{\overline{\gamma}_{ES}} < \frac{p_0}{1 - p_0}$$

or

[10] The path loss exponent is a reflection of how rapidly the signal power density falls with range. For air-to-air links, $\alpha \sim 2$, for ground-to-ground links, $\alpha \sim 4$, and for air-to-ground links, $\alpha \sim 3$ (see [22] for a detailed discussion of the path loss exponent).

$$\frac{d_{\text{Tx,Rx}}}{d_{\text{Tx,EW}}} > \sqrt{\frac{1-p_0}{p_0}}$$

Note that a nonzero privacy capacity may exist even when

$$\overline{\gamma}_{\text{Rx}} < \overline{\gamma}_{\text{ES}} \tag{9.38}$$

albeit with probability less than 0.5, due to the fading characteristics of the channels. That is, the instantaneous SNRs may be such that $\gamma_{\text{Rx}} > \gamma_{\text{ES}}$ for a period of time even though (9.38) is satisfied. The advantage of this, however, in practical terms is minimal since the time intervals over which this is true are unknown to the transmitter.

Intercept Probability

We are now ready to characterize the intercept probability $\text{Pr}_{\text{int}}(R_{\text{p}}) = \text{Pr}\{C_{\text{p}} < R_{\text{p}}\}$, that is, the probability that the instantaneous privacy capacity is less than a target privacy rate $R_{\text{p}} > 0$.[11] The significance of this definition of intercept probability is that when setting the privacy rate R_{p}, Tx assumes that the capacity of the intercept channel is given by $C'_{\text{ES}} = C_{\text{Rx}} - R_{\text{p}}$. If $R_{\text{p}} > C_{\text{p}}$, then $C_{\text{ES}} > C'_{\text{ES}}$ and intercept is possible and information theoretic privacy is compromised On the other hand, when $R_{\text{p}} < C_{\text{p}}$, EW's channel will be worse than Tx's estimate, that is, $C_{\text{Rx}} < C'_{\text{ES}}$, and so the wiretap codes used by Tx will ensure privacy.

We know from the law of total probability that

$$\begin{aligned}\text{Pr}_{\text{int}}(R_{\text{p}}) = \text{Pr}\{C_{\text{p}} < R_{\text{p}} \,|\, \gamma_{\text{Rx}} > \gamma_{\text{ES}}\}\,\text{Pr}\{\gamma_{\text{Rx}} > \gamma_{\text{ES}}\} \\ + \text{Pr}\{C_{\text{p}} < R_{\text{p}} \,|\, \gamma_{\text{Rx}} \le \gamma_{ES}\}\,\text{Pr}\{\gamma_{\text{Rx}} \le \gamma_{\text{ES}}\}\end{aligned} \tag{9.39}$$

Now, from (9.36) we know that

$$\text{Pr}\{\gamma_{\text{Rx}} > \gamma_{\text{ES}}\} = \frac{\overline{\gamma}_{\text{Rx}}}{\overline{\gamma}_{\text{Rx}} + \overline{\gamma}_{\text{ES}}} \tag{9.40}$$

Consequently, we have

[11] Wiretapability or the ability to wiretap is what is referred to in reality since whether a message is wiretapped or not depends on many other factors beyond simply channel capacity (e.g., whether there are resources available at the wiretap receiver to process the message).

$$\Pr\{\gamma_{\mathrm{Rx}} \le \gamma_{\mathrm{ES}}\} = 1 - \Pr\{\gamma_{\mathrm{Rx}} > \gamma_{\mathrm{ES}}\} = \frac{\overline{\gamma}_{\mathrm{ES}}}{\overline{\gamma}_{\mathrm{Rx}} + \overline{\gamma}_{\mathrm{ES}}} \tag{9.41}$$

We also have that

$$\begin{aligned}
&\Pr\{C_{\mathrm{p}} < R_{\mathrm{p}} \,|\, \gamma_{\mathrm{Rx}} > \gamma_{\mathrm{ES}}\} \\
&\quad = \Pr\{\log_2(1+\gamma_{\mathrm{Rx}}) - \log_2(1+\gamma_{\mathrm{ES}}) < R_{\mathrm{p}} \,|\, \gamma_{\mathrm{Rx}} > \gamma_{\mathrm{ES}}\} \\
&\quad = \Pr\{\gamma_{\mathrm{Rx}} < 2^{R_{\mathrm{p}}}(1+\gamma_{\mathrm{ES}}) - 1 \,|\, \gamma_{\mathrm{Rx}} > \gamma_{\mathrm{ES}}\} \\
&\quad = \int_0^{\infty} \int_{\gamma_2}^{2^{R_{\mathrm{p}}}(1+\gamma_{\mathrm{ES}})-1} p(\gamma_{\mathrm{Rx}}, \gamma_{\mathrm{ES}} \,|\, \gamma_{\mathrm{Rx}} > \gamma_{\mathrm{ES}}) d\gamma_{\mathrm{ES}} d\gamma_{\mathrm{Rx}} \\
&\quad = \int_0^{\infty} \int_{\gamma_{\mathrm{ES}}}^{2^{R_{\mathrm{p}}}(1+\gamma_{\mathrm{ES}})-1} \frac{p(\gamma_{\mathrm{Rx}})\, p(\gamma_{\mathrm{ES}})}{\Pr\{\gamma_{\mathrm{Rx}} > \gamma_{\mathrm{ES}}\}} d\gamma_{\mathrm{ES}} d\gamma_{\mathrm{Rx}} \\
&\quad = 1 - \frac{\overline{\gamma}_{\mathrm{Rx}} - \overline{\gamma}_{\mathrm{ES}}}{\overline{\gamma}_{\mathrm{Rx}} + 2^{R_{\mathrm{p}}}\overline{\gamma}_{\mathrm{ES}}} \exp\left(-\frac{2^{R_{\mathrm{p}}} - 1}{\overline{\gamma}_{\mathrm{Rx}}}\right)
\end{aligned} \tag{9.42}$$

and, since $R_{\mathrm{p}} > 0$,

$$\Pr\{C_{\mathrm{p}} < R_{\mathrm{p}} \,|\, \gamma_{\mathrm{Rx}} \le \gamma_{\mathrm{ES}}\} = 1 \tag{9.43}$$

Combining (9.39)–(9.43), we get

$$\Pr_{\mathrm{int}}(R_{\mathrm{p}}) = 1 - \frac{\overline{\gamma}_{\mathrm{Rx}}}{\overline{\gamma}_{\mathrm{Rx}} + 2^{R_{\mathrm{s}}}\overline{\gamma}_{\mathrm{ES}}} \exp\left(-\frac{2^{R_{\mathrm{p}}} - 1}{\overline{\gamma}_{\mathrm{Rx}}}\right) \tag{9.44}$$

Asymptotic Behavior

From (9.44) it follows that when $R_{\mathrm{p}} \to 0$,

$$\Pr_{\mathrm{int}} \to \frac{\overline{\gamma}_{\mathrm{Rx}}}{\overline{\gamma}_{\mathrm{Rx}} + \overline{\gamma}_{\mathrm{ES}}}$$

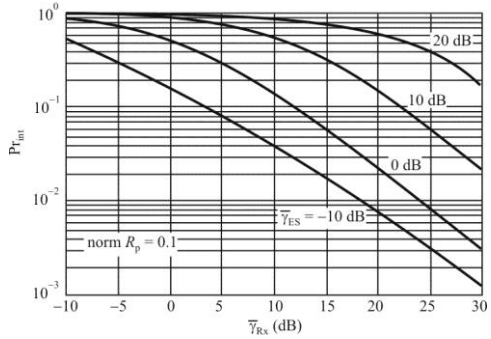


Figure 9.9 Intercept probability versus $\overline{\gamma}_{\text{Rx}}$.

and when $R_{\text{p}} \to \infty$, we have that $\Pr_{\text{int}} \to 1$, such that it becomes impossible for Tx and Rx to transmit private information (at very high rates anyway[12]).

We now examine the asymptotic behavior of the intercept probability for extreme values of the average SNRs of the main channel and the interceptor's channel. When $\overline{\gamma}_{\text{Rx}} >> \overline{\gamma}_{\text{ES}}$, (9.44) yields

$$\Pr_{\text{int}}(R_{\text{p}}) \approx 1 - \exp\left(-\frac{2^{R_{\text{p}}} - 1}{\overline{\gamma}_{\text{Rx}}}\right)$$

and when $\overline{\gamma}_{\text{Rx}}$ is high, $\Pr_{\text{int}} \approx (2^{R_{\text{p}}} - 1)/\overline{\gamma}_{\text{Rx}}$, the intercept probability decays as $1/\overline{\gamma}_{\text{Rx}}$. Conversely, when $\overline{\gamma}_{\text{Rx}} << \overline{\gamma}_{\text{ES}}$,

$$\Pr_{\text{int}}(R_{\text{p}}) \approx 1$$

and interceptability is assured. Figure 9.9 depicts the intercept probability versus $\overline{\gamma}_{\text{Rx}}$, for selected values of $\overline{\gamma}_{\text{ES}}$, and for a normalized target privacy rate of 0.1. We can see that the higher $\overline{\gamma}_{\text{Rx}}$, the lower the intercept probability, and the higher $\overline{\gamma}_{\text{ES}}$, the higher the probability of intercept.

With respect to the asymptotic behavior of the intercept privacy capacity, we see that $C_{\text{int}} \to 0$ yields $\Pr_{\text{int}} \to \overline{\gamma}_{\text{ES}}/(\overline{\gamma}_{\text{Rx}} + \overline{\gamma}_{\text{ES}})$, and when $C_{\text{int}} \to \infty$, we have $\Pr_{\text{int}} \to 1$.

[12] The fading characteristics are statistical in nature and therefore there are times when $\gamma_1 > \gamma_2$ and nonzero channel rates are possible. These times depend on the SNRs at the two receivers, and when $\gamma_1 < \gamma_2$, no private communication is possible.

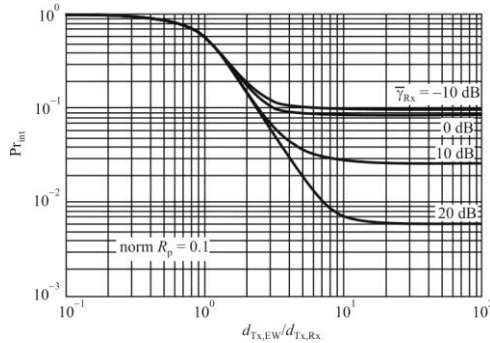


Figure 9.10 Intercept probability versus distance with the normalized R_p = 0.1. Normalization is effected with respect to the capacity of an AWGN channel with SNR equal to $\bar{\gamma}_{Rx}$.

Distance Ratio

The impact of the distance ratio on the performance is illustrated in Figure 9.10, which depicts the intercept probability versus $d_{Tx,EW}/d_{Tx,Rx}$, for selected values of $\bar{\gamma}_{Rx}$, and for a normalized target privacy rate of 0.1. The path loss exponent is set to be equal to a typical value of 3. When $d_{Tx,EW}/d_{Tx,Rx} \to \infty$ $(\bar{\gamma}_{Rx} / \bar{\gamma}_{ES} \to \infty)$, then $\Pr_{int} \to 1 - \exp[-(2^{R_p} - 1)/\bar{\gamma}_{Rx}]$. If $d_{Tx,EW}/d_{Tx,Rx} \to 0$ $(\bar{\gamma}_{Rx} / \bar{\gamma}_{ES} \to 0,$ then $\Pr_{int} \to 1$.

Summary

We provided a characterization of the intercept privacy capacity of wireless channels with quasi-static fading. Specifically, we assumed that Tx—having access to the CSI of the main channel only—chooses a target privacy rate R_p (without knowing the intercept channel) and we investigated the intercept probability defined as $\Pr\{R_p > C_p\}$. Our results reveal that for reasonable levels of SNR at Rx and ES, $\Pr_{int}$ can approach 1, thus assuring intercept at ES. As expected, the interceptability is also a function of the range ratio $d_{Tx,EW}/d_{Tx,Rx}$.

9.5 Jamming Performance in AWGN Channels

In this section we evaluate the effects of including a jammer in the EW scenario. A relatively simple measure of performance is used, which is facilitated by assuming that the channel is an AWGN channel. That is, the noise in the channel is zero-mean Gaussian. When that is the case, the denominator in the expressions for the SNR in (3.25) is simply increased to account for the additional noise.

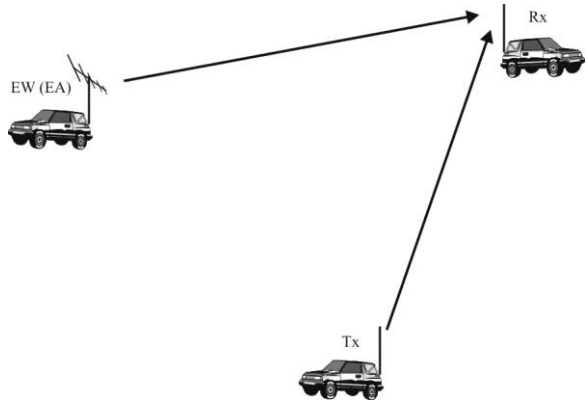


Figure 9.11 EA scenario. (Source: [23]. © Artech House, 2008. Reprinted with permission.)

9.5.1 Jammer Scenario

When a jammer is added to the communication channel, the scenario is as depicted in Figure 9.11. As before, the target transmitter is attempting to communicate with the target receiver and the jammer is attempting to thwart that communication by reducing the channel capacity available to the transmitter/receiver pair. Artificial thermal noise is the most common type of interfering signal available to the jammer.

Adding a jammer to the scenario depicted in Figure 9.12 that adds a thermal noise-like signal to the channel is one effective way to conduct EA in RF systems. In fact, the characteristics of the resulting configuration follows Shannon's basic theorem properties as indicated by (3.25) so the jammer effectiveness on the channel capacity reduction can be theoretically evaluated.

9.5.2 Broadband Noise Jamming

The power, J, in the noise-like signal produced at the receiver by a jammer as illustrated in Figure 9.12, adds to N_c, and the total noise in the denominator of the last term in (3.25) is given by

$$N = N_c + J \tag{9.45}$$

Equation (3.25) can thus be manipulated to yield

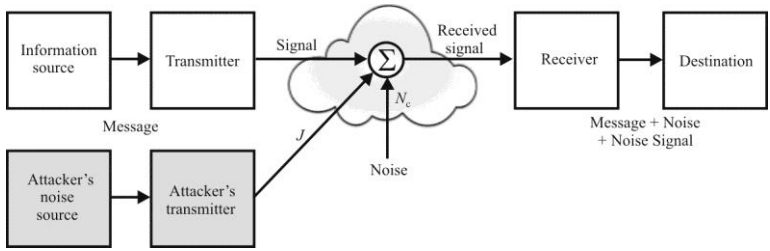


Figure 9.12 DOI active degradation strategy.

$$C/W = \log_2\left(1+\frac{1}{\frac{1}{\text{SNR}}+\text{JSR}}\right) \tag{9.46}$$

where $\text{SNR} = S/N$ and $\text{JSR} = J/S$, both numeric (not dB), and calculated at the receiver. Graphs of *C*/*W* are illustrated in Figure 9.13 for a few values of SNR. We can see that at reasonable tactical values of SNR (greater than 10 dB), the channel capacity begins to be significantly affected at JSR values greater than about 0 dB.

It should be noted, however, that no matter what the JSR level, there is always a level of SNR and coding scheme that will transmit information from the transmitter to the receiver. The data rates may be quite low, but some information will get through.

Other forms of jamming waveforms are also possible, such as single or multiple tones. The jamming performance of these waveforms cannot theoretically be evaluated using Shannon's basic theorem because they are not noise waveforms occupying all of *W*. Their performance can be evaluated nevertheless using different techniques [23].

9.5.3 Partial-Band Noise Jamming

Partial-band noise (PBN) jamming places a noise waveform over a portion of the bandwidth occupied by the communication system as illustrated in Figure 9.14. We denote the fraction of the jammer spectral density to the noise spectral density by η; that is $W_J = \eta W_0$ or

$$\eta = \frac{W_J}{W_0} \tag{9.47}$$

With the addition of the jammer noise to the thermal noise floor, (3.25) becomes

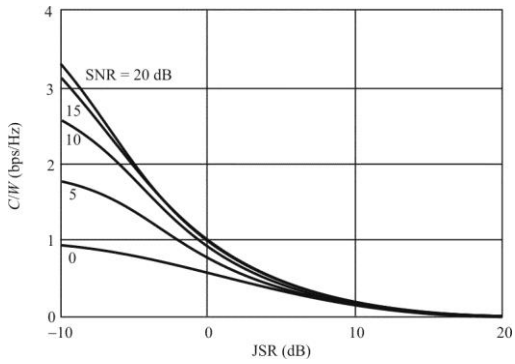


Figure 9.13 Capacity over bandwidth.

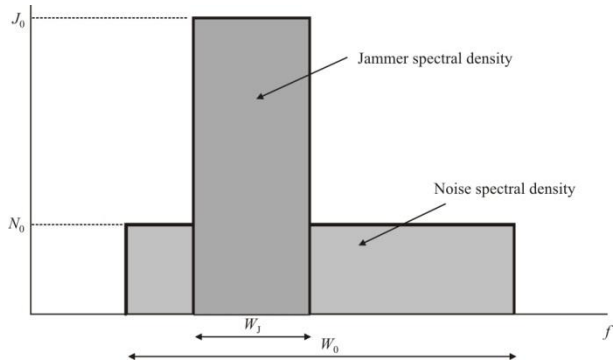


Figure 9.14 Spectral densities of the noise and jammer.

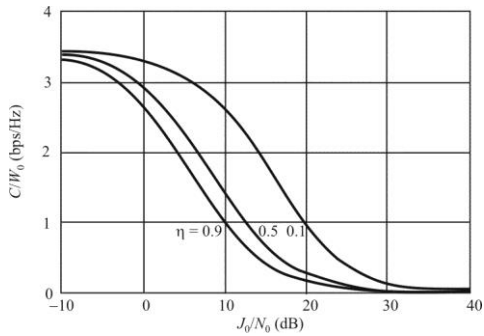


Figure 9.15 Capacity over bandwidth when SNR = 10 dB.

$$C = W_0 \log_2 \left(1 + \frac{S}{W_0 N_0 + W_J J_0} \right) \tag{9.48}$$

so that

$$C = W_0 \log_2 \left\{ 1 + \left[\frac{1}{\text{SNR}} \left(1 + \underset{\eta}{\frac{W_J}{W_0} \frac{J_0}{N_0}} \right) \right]^{-1} \right\} \tag{9.49}$$

The SNR in this case is what the SNR is without the jammer.

C/W_0 in (9.49) is plotted in Figures 9.15–9.17 for typical values of SNR and JSR. As expected, as the SNR increase, the jamming effectiveness decreases for fixed noise levels. We can note that a ten-fold increase in η produces a tenfold increase in jamming performance.

9.6 Spatially Duplexed EW System Performance with Multiple Antennas

Mukherjee and Swindlehurst approached the combined problem of a transmitter communicating with a receiver in the presence of an intercept receiver and a

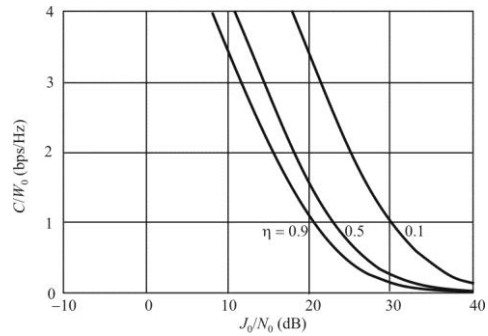


Figure 9.16 Capacity over bandwidth when SNR = 20 dB.

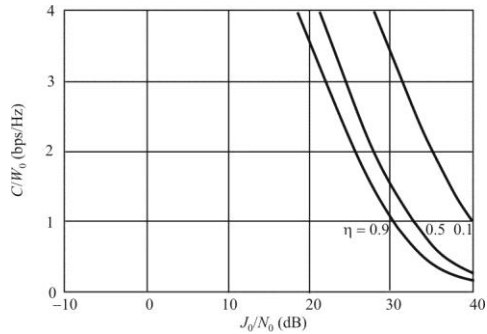


Figure 9.17 Capacity over bandwidth when SNR = 30 dB.

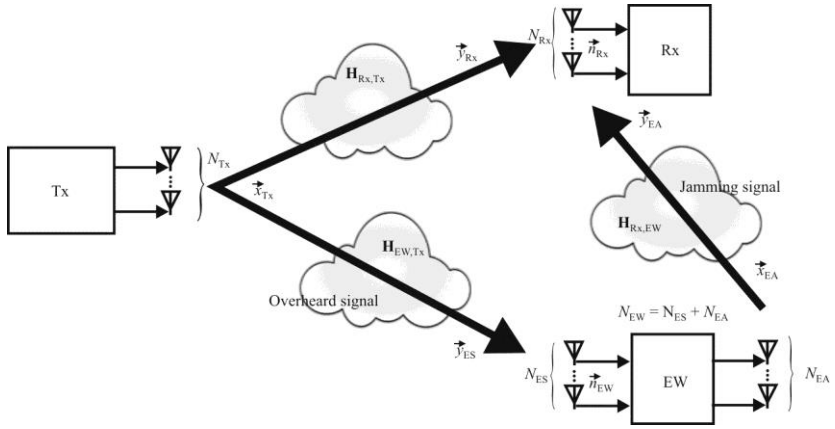


Figure 9.18 MIMO scenario for spatially duplexed antennas.

jammer over MIMO channels[13] [24]. We outline their approach and conclusions in this section.

9.6.1 Active Intercept Channel

We begin by considering a three-node network with a transmitter (Tx) with N_{Tx} antennas, an N_{Rx}-antenna receiver (Rx), and a malicious user (EW) with N_{EW} antennas. EW uses N_{ES} of the N_{EW} elements in the antenna array to intercept, and the remaining $N_{EA} = N_{EW} - N_{ES}$ antennas for jamming Rx. The scenario is depicted in Figure 9.18.

In the strict sense of the term, full-duplex operation at EW is interpreted as all the antennas being used simultaneously for intercept and jamming: $N_{EA} = N_{ES} = N_{EW}$. However, due to practical hardware limitations and to improve circuit isolation, it is more realistic to assume the array N_{EW} is partitioned into intercept and jamming subarrays of sizes N_{ES}, N_{EA}, respectively, $(N_{EA} + N_{ES} = N_{EW})$ and with a nonnegligible cross-coupling matrix $\mathbf{H}_{si} \in \mathbb{C}^{N_{ES} \times N_{EA}}$, that is, a "quasi" full-duplex mode.

The signals received by Rx and EW at time t are

$$\vec{y}_{Rx}(t) = \mathbf{H}_{Rx,Tx}\vec{x}_{Tx}(t) + \mathbf{H}_{Rx,EA}\vec{x}_{EA}(t) + \vec{n}_{Rx}(t) \tag{9.50}$$

$$\vec{y}_{ES}(t) = \mathbf{H}_{EW,Tx}\vec{x}_{Tx}(t) + \sqrt{\rho}\mathbf{H}_{si}\vec{x}_{EA}(t) + \vec{n}_{ES}(t) \tag{9.51}$$

[13] Recall that one or more of the nodes in a MIMO channel have more than one antenna. They therefore have the property of space diversity, in addition to possibly frequency and time diversity.

respectively, where $\mathbf{H}_{\text{Rx,Tx}} \in \mathbb{C}^{N_{\text{Rx}} \times N_{\text{Tx}}}$, $\mathbf{H}_{\text{EW,Tx}} \in \mathbb{C}^{N_{\text{Rx}} \times N_{\text{ES}}}$ are the channels from Tx, $\mathbf{H}_{\text{Rx,EW}} \in \mathbb{C}^{N_{\text{Rx}} \times N_{\text{EA}}}$ is the channel from EW to Rx, $\mathbf{H}_{\text{si}}$ is the loop interference channel at EW due to full-duplex operation, and ρ, $0 \le \rho \le 1$, represents the self-interference parameter of the intercept after self-interference cancellation. Dropping the time index for brevity, the transmitted information signal is $\vec{x}_{\text{Tx}} \in \mathbb{C}^{N_{\text{Tx}} \times 1}$, $\vec{x}_{\text{EA}} \in \mathbb{C}^{N_{\text{EA}} \times 1}$ is the jamming signal, and $\vec{n}_{\text{Rx}}$, $\vec{n}_{\text{ES}}$ are the independent background additive white complex Gaussian noise vectors: $\mathcal{E}\{n_k n_k^{\text{H}}\} = \sigma_k^2 \mathbf{I}$, where $k \in \{\text{Rx}, \text{ES}\}$.

We assume a fading scenario where the channels stay constant over a certain number of channel uses, and then transition independently to a new realization. Tx has perfect knowledge of $\mathbf{H}_{\text{Rx,Tx}}$ and a potentially imperfect estimate of $\mathbf{H}_{\text{EW,Tx}}$; EW possesses knowledge of $\mathbf{H}_{\text{Rx,Tx}}$, $\mathbf{H}_{\text{Rx,EW}}$, $\mathbf{H}_{\text{EW,Tx}}$ and a possibly imperfect estimate of $\mathbf{H}_{\text{si}}$. These estimates can be obtained by EW by intercepting training signals emanating from Tx and Rx to learn $\mathbf{H}_{\text{Rx,EW}}$, $\mathbf{H}_{\text{EW,Tx}}$ and intercepting the CSI feedback from Rx to Tx to acquire $\mathbf{H}_{\text{Rx,Tx}}$, while the estimation of $\mathbf{H}_{\text{si}}$ is discussed in Section 9.6.4.

Tx's transmit power is assumed to be fixed at P_{Tx}:

$$\mathcal{E}\{\vec{x}_{\text{Tx}} \vec{x}_{\text{Tx}}^{\text{H}}\} = \mathbf{Q}_{\text{Tx}} \qquad \text{Tr}(\mathbf{Q}_{\text{Tx}}) = P_{\text{Tx}} \tag{9.52}$$

When in the jamming mode, EW has a maximum power constraint, P_{EW}, as well, given by

$$\mathcal{E}\{\vec{x}_{\text{EA}} \vec{x}_{\text{EA}}^{\text{H}}\} = \mathbf{Q}_{\text{EA}} \qquad \text{Tr}(\mathbf{Q}_{\text{EA}}) \le P_{\text{EA}} \tag{9.53}$$

When the input $\vec{x}_{\text{Tx}}$ is drawn from an arbitrary distribution, the MIMO privacy rate without prefix coding [7, 25] is given by

$$R_{\text{p}} = [\mathcal{I}(\vec{x}_{\text{Tx}}; \vec{y}_{\text{Rx}}) - \mathcal{I}(\vec{x}_{\text{Tx}}; \vec{y}_{\text{ES}})]^{+} \tag{9.54}$$

where $[a]^{+} = \max(0, a)$.

For this scenario, EW seeks to solve

$$\min_{N_{\text{EA}}} \min_{\mathbf{Q}_{\text{EA}} \in \mathbf{\Omega}} R_{\text{p}} \tag{9.55}$$

where $\mathbf{\Omega} = \{\mathbf{Q}_{\text{EA}} \,|\, \mathbf{Q}_{\text{EA}} \in \mathbb{C}^{N_{\text{EA}} \times N_{\text{EA}}}, \mathbf{Q}_{\text{EA}} \ge \mathbf{0}, \text{Tr}(\mathbf{Q}_{\text{EA}}) \le P_{\text{EA}}\}$ is the nonempty, convex set of feasible jamming covariances. Note that due to our definition of full-duplex

EW operation, the optimal attack strategy consists of an outer optimization over the subset of antennas allocated for jamming versus eavesdropping (that is, an antenna selection problem), and an inner optimization over $\mathbf{Q}_{\mathrm{EA}}$ for a given choice of N_{EA}. In the remainder of this section we focus on the inner optimization of $\mathbf{Q}_{\mathrm{EA}}$ for a fixed set N_{EA}, bearing in mind that an exhaustive search over N_{EA} has complexity that grows exponentially with N_{EW}. Later we examine the antenna selection issue assuming a fixed structure for $\mathbf{Q}_{\mathrm{EA}}$.

Property 9.4 [24]: For a fixed jamming subarray N_{EA}, the optimal jamming covariance for the MIMO wiretap channel with arbitrary input distribution satisfies

$$\lambda \mathbf{Q}_{\mathrm{EA}} = -[\nabla \mathbf{Q}_{\mathrm{EA}}(R_{\mathrm{p}})]\mathbf{Q}_{\mathrm{EA}} \tag{9.56}$$

where

$$\begin{aligned}\nabla \mathbf{Q}_{\mathrm{EA}}(R_{\mathrm{p}}) = &-\mathbf{H}_{\mathrm{Rx,EW}}^{\mathrm{H}}\mathbf{K}_{\mathrm{Rx}}^{-1}\mathbf{H}_{\mathrm{Rx,Tx}}\mathbf{E}_{\mathrm{Rx}}\mathbf{H}_{\mathrm{Rx,Tx}}^{\mathrm{H}}\mathbf{K}_{\mathrm{Rx}}^{-1}\mathbf{H}_{\mathrm{Rx,EW}} \\ &+\rho\mathbf{H}_{\mathrm{si}}^{\mathrm{H}}\mathbf{K}_{\mathrm{ES}}^{-1}\mathbf{H}_{\mathrm{ES,Tx}}\mathbf{E}_{\mathrm{ES}}\mathbf{H}_{\mathrm{ES,Tx}}^{\mathrm{H}}\mathbf{K}_{\mathrm{ES}}^{-1}\mathbf{H}_{\mathrm{si}}\end{aligned} \tag{9.57}$$

$\mathbf{E}_i = \mathcal{E}\left\{(\vec{x}_{\mathrm{Tx}} - \mathcal{E}\{\vec{x}_{\mathrm{Tx}} \,|\, \vec{y}_i\})(\vec{x}_{\mathrm{Tx}} - \mathcal{E}\{\vec{x}_{\mathrm{Tx}} \,|\, \vec{y}_i\})^{\mathrm{H}}\right\}, i \in \{\mathrm{Rx}, \mathrm{ES}\}$, are the associated minimum mean-square error matrices, λ is a non-negative scale factor, and $\mathbf{K}_{\mathrm{Rx}}$, $\mathbf{K}_{\mathrm{ES}}$ are the receive interference-plus-noise covariance matrices at Rx and ES, respectively

$$\begin{aligned}\mathbf{K}_{\mathrm{Rx}} &= \mathbf{H}_{\mathrm{Rx,EW}}\mathbf{Q}_{\mathrm{EA}}\mathbf{H}_{\mathrm{Rx,EW}}^{\mathrm{H}} + \sigma_{\mathrm{Rx}}^2\mathbf{I} \\ \mathbf{K}_{\mathrm{ES}} &= \rho\mathbf{H}_{\mathrm{si}}\mathbf{Q}_{\mathrm{EA}}\mathbf{H}_{\mathrm{si}}^{\mathrm{H}} + \sigma_{\mathrm{ES}}^2\mathbf{I}\end{aligned} \tag{9.58}$$

Property 9.4 characterizes a jamming strategy by EA for an arbitrary input distribution chosen by Tx. For the MIMO wiretap channel with ES present, a Gaussian input signal for EA is known to be optimal. Likewise, Gaussian signaling is the capacity achieving input distribution against worst-case Gaussian jamming in the MIMO channel without privacy considerations [26]. Therefore we consider Gaussian input signals of the form $\vec{x}_{\mathrm{Tx}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q}_{\mathrm{Tx}})$. The *generalized singular value decomposition* (GSVD) precoder is the asymptotically optimal transmission strategy for Tx (assuming knowledge of $\mathbf{H}_{\mathrm{EW,Tx}}$ and noise pre-

whitening of $\vec{y}_{\text{Rx}}$ at Rx) in the Gaussian MIMOME[14] wiretap channel with passive ES. This is a potential candidate for $\mathbf{Q}_{\text{Tx}}$.

We are interested in the optimization at ES, which seeks to solve the following

$$\min_{\mathbf{Q}_{\text{EA}} \in \boldsymbol{\Omega}} R_{\text{p}} \tag{9.59}$$

For the proposed system with Gaussian signaling, define the MIMO privacy rate as

$$R_{\text{p}}(\mathbf{Q}_{\text{EA}}) \triangleq \log_2 \left| \mathbf{I} + \mathbf{H}_{\text{Rx,Tx}} \mathbf{Q}_{\text{Tx}} \mathbf{H}_{\text{Rx,Tx}}^{\text{H}} \mathbf{K}_{\text{Rx}}^{-1} \right| - \log_2 \left| \mathbf{I} + \mathbf{H}_{\text{EW,Tx}} \mathbf{Q}_{\text{Tx}} \mathbf{H}_{\text{EW,Tx}}^{\text{H}} \mathbf{K}_{\text{ES}}^{-1} \right| \tag{9.60}$$

[$\mathbf{K}_{\text{Rx}}$ and $\mathbf{K}_{\text{ES}}$ are functions of $\mathbf{Q}_{\text{EA}}$ as seen in (9.58).]

If ES does not adopt any self-interference cancellation techniques as a worst-case assumption, then $\rho = 1$. This would be the case, for example, if the jamming and intercept arrays are actually collocated but without coordination between ES and EA.

Let $\mathbf{H}_{\text{Rx,Tx}} = \mathbf{U}\mathbf{diag}(d_{ii})\mathbf{V}^{\text{H}}$ be the SVD of the Tx-Rx channel. We can then evaluate the following special cases.

Property 9.5(a) [24]: When $\rho = 0$, the optimal jamming covariance given uniform power allocation at Tx is

$$\mathbf{H}_{\text{Rx,EW}} \mathbf{Q}_{\text{EA}} \mathbf{H}_{\text{Rx,EW}}^{\text{H}} = \mathbf{U}\mathbf{diag}(\sigma_i^*)\mathbf{U}^{\text{H}}$$

where

$$\sigma_i^* = -\frac{d_{ii}^2 P_{\text{Tx}}}{2\sigma_{\text{Rx}}^2} + \frac{d_{ii}^2 P_{\text{Tx}}}{2\sigma_{\text{Rx}}^2} \sqrt{1 + \frac{4\sigma_{\text{Rx}}^2}{\alpha d_{ii}^2 P_{\text{Tx}}}} - 1, \qquad \alpha \le \frac{d_{ii}^2 P_{\text{Tx}}}{\sigma_{\text{Rx}}^2 + d_{ii}^2 P_{\text{Tx}}}$$

and α is chosen to satisfy the Tx power constraint.

Property 9.5(b) [24]: For the MISOSE channel ($N_{\text{Tx}} \ge 1, N_{\text{Rx}} = N_{\text{EW}} = 1$) where Tx employs a beamformer $\mathbf{W} \triangleq \vec{w}\vec{w}^{\text{H}}$, EW dedicates its antenna to jamming if and only if

[14] We use the notation MIMOME = multiple input, multiple output, multiple intercept; MISOSE = multiple input, single output, single intercept; and SIMOSE = single input, multiple output, single intercept; where the single, multiple terms refer to the number of antennas.

$$\frac{1+\dfrac{P_{\mathrm{Tx}}\mathbf{H}_{\mathrm{Rx,Tx}}\mathbf{W}^{\mathrm{H}}\mathbf{H}_{\mathrm{Rx,Tx}}^{\mathrm{H}}}{\sigma_{\mathrm{Rx}}^2}}{1+\dfrac{P_{\mathrm{Tx}}\mathbf{H}_{\mathrm{Rx,Tx}}\mathbf{W}^{\mathrm{H}}\mathbf{H}_{\mathrm{Rx,Tx}}^{\mathrm{H}}}{\sigma_{\mathrm{Rx}}^2+\left|\mathbf{H}_{\mathrm{Rx,EW}}\right|^2 P_{\mathrm{EA}}}} < 1+\frac{P_{\mathrm{Tx}}\mathbf{H}_{\mathrm{EW,Tx}}\mathbf{W}^{\mathrm{H}}\mathbf{H}_{\mathrm{EW,Tx}}^{\mathrm{H}}}{\sigma_{\mathrm{ES}}^2}$$

otherwise EW intercepts.

Note: Property 9.5(b) describes a very common tactical EW scenario where the number of antennas on the mobile nodes is restricted to one, but the larger, more static transmit site may be able to facilitate a number of them.

Property 9.5(c): For the SIMOSE channel (N_{Tx} =N_{EW} = 1, $N_{\mathrm{Rx}} \geq 1$)) EW dedicates its antenna to jamming if and only if

$$\frac{1+\dfrac{\left|\mathbf{H}_{\mathrm{Rx,Tx}}\right|^2 P_{\mathrm{Tx}}}{\sigma_{\mathrm{Rx}}^2}}{1+\dfrac{\left|\mathbf{H}_{\mathrm{Rx,Tx}}\right|^2 P_{\mathrm{Tx}}}{\sigma_{\mathrm{Rx}}^2+\left|\mathbf{H}_{\mathrm{Rx,EW}}\right|^2 P_{\mathrm{EA}}}} < 1+\frac{\left|h_{\mathrm{EW,Tx}}\right|^2 P_{\mathrm{Tx}}}{\sigma_{\mathrm{ES}}^2}$$

otherwise EW intercepts.

Again, Property 9.5(c) describes a common tactical scenario. In this case Tx, representing a mobile transmitter, has one antenna as does the mobile EW. Rx has multiple antennas, representing a relatively static node where multiple antennas may be allowed. Since Tx has a single antenna, implementation of a beamformer is not possible, although a receive beamformer at Rx is within reason.

Property 9.6 [24]: For a fixed jamming subarray N_{EA}, the optimal jamming covariance for the MIMO wiretap channel with Gaussian input signaling satisfies

$$\mu\mathbf{Q}_{\mathrm{EA}} = -\left[\frac{d}{d\mathbf{Q}_{\mathrm{EA}}}R_{\mathrm{p}}(\mathbf{Q}_{\mathrm{EA}})\right]\mathbf{Q}_{\mathrm{EA}} \tag{9.61}$$

where the derivative $dR_{\mathrm{p}}(\mathbf{Q}_{\mathrm{EA}})/d\mathbf{Q}_{\mathrm{EA}}$ is

Algorithm 1 Gradient Projection+Fixed-Point Iteration (GP-FP) search for optimum $\mathbf{Q}_{\mathrm{EA}}$.

Require: $P_{\mathrm{EA}} > 0$

Initialize $\mathbf{Q}_{\mathrm{EA}}^{(0)}, k = 0$

do

Compute gradient $\nabla R_{\mathrm{p}}[\mathbf{Q}_{\mathrm{EA}}^{(k)}]$

$\tilde{\mathbf{Q}}_{\mathrm{EA}}^{(k)} = \mathbf{Q}_{\mathrm{EA}}^{(k)} + \alpha^{(k)} \nabla R_{\mathrm{p}}[\mathbf{Q}_{\mathrm{EA}}]$

$\mathbf{Q}'^{(k)}_{\mathrm{EA}}$ = projection of $\tilde{\mathbf{Q}}_{\mathrm{EA}}^{(k)}$ onto Ω

$\mathbf{Q}_{\mathrm{EA}}^{(k+1)} = \mathbf{Q}'^{(k)}_{\mathrm{EA}} + \beta^{(k)}[\mathbf{Q}'^{(k)}_{\mathrm{EA}} - \mathbf{Q}_{\mathrm{EA}}^{(k)}]$

if $\left\|\mathbf{Q}_{\mathrm{EA}}^{(k+1)}\right\|_{\mathrm{F}} - \left\|\mathbf{Q}_{\mathrm{EA}}^{(k)}\right\|_{\mathrm{F}} > \varepsilon$ then

$k = k + 1$

end if

Initialize $y = 0, \mathbf{Q}_{\mathrm{EA,FP}}^{(y)} = \mathbf{Q}_{\mathrm{EA}}^{(k+1)}$

Update $\mathbf{Q}_{\mathrm{EA,FP}}^{(y+1)}$ from (6.31) until convergence.

Source: [24].

$$\frac{d}{d\mathbf{Q}_{\mathrm{EA}}} R_{\mathrm{p}}(\mathbf{Q}_{\mathrm{e}}) = \frac{1}{\ln 2}\left\{\mathbf{H}_{\mathrm{Rx,EW}}^{\mathrm{H}}\left(\mathbf{Z}_{\mathrm{Rx}} + \mathbf{H}_{\mathrm{Rx,EW}}\mathbf{Q}_{\mathrm{EA}}\mathbf{H}_{\mathrm{Rx,EW}}^{\mathrm{H}}\right)^{-1}\mathbf{H}_{\mathrm{Rx,EW}}\right\}^{\mathrm{T}}$$

$$-\frac{1}{\ln 2}\left\{\mathbf{H}_{\mathrm{Rx,EW}}^{\mathrm{H}}\left(\sigma_{\mathrm{Rx}}^2\mathbf{I} + \mathbf{H}_{\mathrm{Rx,EW}}\mathbf{Q}_{\mathrm{EA}}\mathbf{H}_{\mathrm{Rx,EW}}^{\mathrm{H}}\right)^{-1}\mathbf{H}_{\mathrm{Rx,EW}}\right\}^{\mathrm{T}}$$

$$-\frac{\rho}{\ln 2}\left\{\mathbf{H}_{\mathrm{si}}^{\mathrm{H}}\left(\mathbf{Z}_{\mathrm{EW}} + \rho\mathbf{H}_{\mathrm{si}}\mathbf{Q}_{\mathrm{EA}}\mathbf{H}_{\mathrm{si}}^{\mathrm{H}}\right)^{-1}\mathbf{H}_{\mathrm{si}}\right\}^{\mathrm{T}}$$

$$+\frac{\rho}{\ln 2}\left\{\mathbf{H}_{\mathrm{si}}^{\mathrm{H}}\left(\sigma_{\mathrm{EW}}^2\mathbf{I} + \rho\mathbf{H}_{\mathrm{si}}\mathbf{Q}_{\mathrm{EA}}\mathbf{H}_{\mathrm{si}}^{\mathrm{H}}\right)^{-1}\mathbf{H}_{\mathrm{si}}\right\}^{\mathrm{T}} \qquad (9.62)$$

In [24] a relatively simple two-stage numerical search algorithm for $\mathbf{Q}_{\mathrm{EA}}$ is included (here listed as Algorithm 1). In the first step, a *gradient projection* (GP) method is applied to generate a candidate covariance [27]. The output of the GP algorithm is then used as an initialization to a set of iterations that exploit Property 9.6.

The gradient in this case is

$$\nabla R_{\mathrm{p}}[\mathbf{Q}_{\mathrm{EA}}^{(k)}] = 2\left\{\frac{d}{d\mathbf{Q}_{\mathrm{EA}}^{(k)}} R_{\mathrm{p}}\left[\mathbf{Q}_{\mathrm{EA}}^{(k)}\right]\right\}^{*}$$

since $R_{\mathrm{p}}[\mathbf{Q}_{\mathrm{EA}}^{(k)}]$ is a real-valued function.

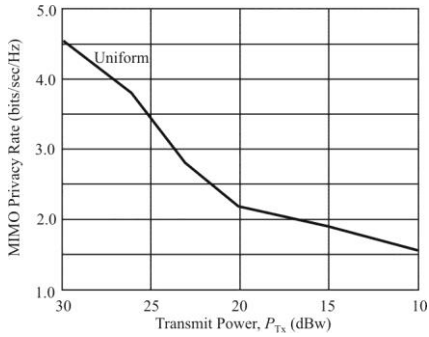


Figure 9.19 Privacy rate with precoding. Uniform power allocation, P_{EA} = 10 dBw, $N_{Tx} = N_{Rx} = 4$, $N_{EW} = 3$.

The step sizes $\alpha^{(k)}$, $\beta^{(k)}$, can be chosen based on Armijo's rule[15] along the feasible direction [27, 28]. The projection step onto set Ω involves scaling the eigenvalues of the Hermitian matrix $\tilde{\mathbf{Q}}_{EA}^{(k)}$ so as to satisfy the trace constraint of P_{EA}.

> **Example 9.5:** In Figure 9.19, we see a numerical example for i.i.d. Rayleigh fading channels comparing the privacy rate with GSVD precoding at Tx (best-case scenario where Tx knows $\mathbf{H}_{EW,Tx}$) and uniform power allocation. The optimal jamming subarray set N_{EA} is chosen via exhaustive search, and we observe that the optimal $\mathbf{Q}_{EA}$ allows EW to best suppress the channel privacy rate.

9.6.2 Jamming Waveforms

The jamming signal $\vec{x}_{EA}(t)$ can either be an arbitrary function of the information signal $f[\vec{x}_{Tx}(t-\tau)]$ where τ is the processing delay of EW's decoder, that is, a

---

[15] Armijo's rule is a method of selecting the search direction and size in line searches. Let $f:\mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function. The goal is to find $\min_{\vec{x}\in\mathbb{R}^n} f(\vec{x})$. With $\vec{g}(x) \triangleq \nabla f(x)$, Armijo's rule is implemented as follows: Set scalars s_k, β, L >0, μ, and σ as [28]:

- $s_k = -\vec{g}^{\mathrm{T}}(\vec{x}_k)\vec{p}_k / (L\|\vec{p}_k\|^2)$
- $\beta \in (0,1)$
- $\sigma \in (0,1/2)$

Then let α_k be the largest α in $\{s_k, \beta s_k, \beta^2 s_k, \ldots\}$ such that

$$f(x_k + \alpha\vec{p}_k) - f(\vec{x}_k) \le \sigma\alpha\vec{g}^{\mathrm{T}}(\vec{x}_k)\vec{p}_k$$

correlated jamming signal, or independent, spatially white Gaussian noise, or some linear combination of the two. As shown in the previous section, finding a closed-form solution for the optimal interference covariance can be elusive. Let us assume the jamming signal $\vec{x}_{\mathrm{EA}}(t)$ is spatially white complex Gaussian noise uncorrelated with the source signal $\vec{x}_{\mathrm{Tx}}(t)$. Thus, we can frame EW's optimization problem as an antenna subset selection issue.

9.6.3 Antenna Selection

There is a tradeoff between allocating resources for eavesdropping versus jamming. Allotting a large number for N_{ES} allows EW to overhear more of the information signal, but leaves fewer antennas for jamming Rx. Similarly, jamming Rx with full power and a large N_{EA} can significantly degrade the received signal at both Rx and EW (ES). A low-complexity approach when $P_{\mathrm{EA}} \ll P_{\mathrm{Tx}}$ or $N_{\mathrm{EW}} \gg N_{\mathrm{Tx}}$ is for EW to initialize its eavesdropping subarray with the "strongest" receive antenna (largest column norm of $\mathbf{H}_{\mathrm{EW,Tx}}$), and progressively enlarge N_{ES} in a greedy manner, thereby exploiting its natural advantage for eavesdropping. If $P_{\mathrm{EA}} \gg P_{\mathrm{Tx}}$ or $N_{\mathrm{EW}} \gg N_{\mathrm{Rs}}$, then priority is given to jamming Rx and the greedy antenna subset selection is initialized with EW's strongest transmit antenna to Rx.

9.6.4 Self-Interference Cancellation

EW is assumed to possess the following estimate of the self-interference channel

$$\hat{\mathbf{H}}_{\mathrm{si}} = \mathbf{H}_{\mathrm{si}} + \Delta\mathbf{H}_{\mathrm{si}} \tag{9.63}$$

where $\Delta\mathbf{H}_{\mathrm{si}}$ is a (deterministic) error term whose size is assumed to be bounded by EW's chosen jamming power:[16]

$$\left\|\Delta\mathbf{H}_{\mathrm{si}}\right\|_{\mathrm{F}}^{2} \le P_{\mathrm{EA}} \tag{9.64}$$

The additional difficulty in estimating $\mathbf{H}_{\mathrm{si}}$ as compared to the other channels in the network is due to the limited dynamic range of the RF front end: the difference in self-interference and target signal powers can be as large as 100 dB or greater [29].

Since in principle EW knows its jamming signal perfectly, it can be canceled out from the received signal to provide

$$\vec{y}_{\mathrm{ES}} = \mathbf{H}_{\mathrm{EW,Tx}}\vec{x}_{\mathrm{Tx}} - \Delta\mathbf{H}_{\mathrm{si}}\vec{x}_{\mathrm{EA}} + \vec{n}_{\mathrm{ES}} \tag{9.65}$$

---

[16] $\left\|\cdot\right\|_{\mathrm{F}}$ is the Frobenius norm.

However, practical considerations such as amplifier nonlinearities preclude total cancellation of the jamming signal from the received signal. The cross-coupling factor ρ quantifies the impact of imperfect self-interference channel estimation and cancellation process, and is assumed to be directly proportional to the variance of the residual interference-plus-noise at EW [30].

EW can also use its spatial dimensions to mitigate the self-interference. Let the jamming signal be preprocessed by the $N_{\text{EA}} \times N$ precoding matrix $\mathbf{P}$ as $\vec{x}_{\text{EA}} = \mathbf{P}\vec{z}$, where $\vec{z}$ is an N-dimensional spatially white Gaussian noise signal. If EW applies a linear receive filter $\mathbf{D} \in \mathbb{C}^{N_{\text{EA}} \times N'}$ to detect $\vec{x}_{\text{Tx}}$ from (9.51), then we can require that $\mathbf{D}^{\text{H}}\mathbf{H}_{\text{si}}\mathbf{P} = \mathbf{0}$.

9.6.5 Summary

We have examined optimal jamming strategies for a (pseudo-) full duplex active intercept that seeks to minimize the channel privacy rate of the three-node MIMOME wiretap channel. We characterized the worst-case jamming covariance for arbitrary and Gaussian input signaling and a numerical algorithm was presented to compute that covariance. We then examined suboptimal active eavesdropping schemes that comprise essentially an antenna selection problem where EW allocates some of its spatial dimensions (antennas) for jamming Rx.

9.7 EW Performance with Collocated EA and ES and Multiple Antennas

In some realistic cases, the ES and EA are collocated and their operation is synchronized or coordinated. We present some results here when that is the case.

Mukherjee and Swindlehurst [31] provided an analysis of a channel with a sophisticated adversary, consisting of an interceptor (active intercept) with the dual capability of either eavesdropping passively or jamming any ongoing transmission. All three of the nodes in the channel have multiple antennas so we have a MIMO communication link. Tx faces the dilemma of establishing a reliable communication link to Rx that is robust to potential jamming, while also ensuring maximum confidentiality from interception. Since it is not clear a priori what strategies should be adopted by Tx or ES per channel use, a game-theoretic formulation of the problem is used due to the mutually opposite interests of the agents. The game payoff function in our application is the achievable MIMO privacy rate between Tx and Rx. Tx and Rx try to maximize this rate while EW (ES and EA) try to minimize it. The scenario is depicted in Figure 9.20. We present the relevant results of that analysis in this section.

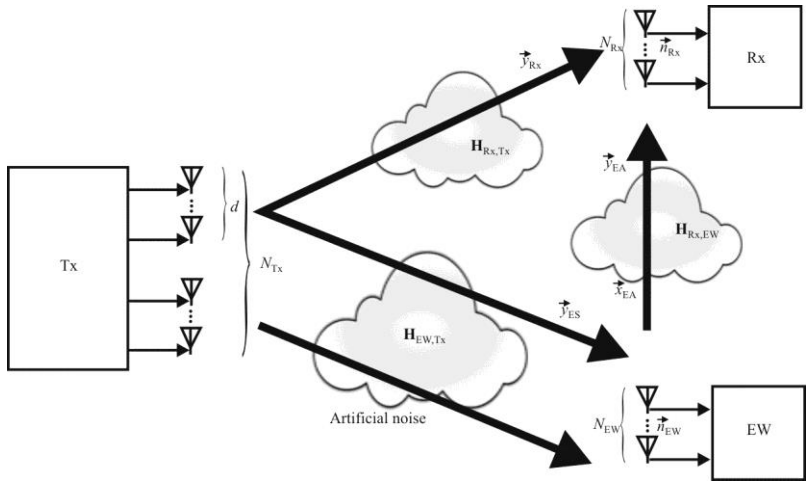


Figure 9.20 MIMO scenario with collocated ES and EA.

9.7.1 Channel Scenario

We examine the MIMO intercept problem in which three multiple-antenna nodes are present: Tx has N_{Tx} antennas, Rx has N_{Rx} antennas, and EW has N_{EW} antennas (see Figure 9.20). We assume that Tx does not have knowledge of the instantaneous CSI of the intercept channel, only an estimate of its statistical distribution (zero-mean with a scaled-identity covariance). Therefore, Tx has the option of utilizing all its resources (power and antennas) for transmitting data to Rx, regardless of channel conditions or potential intercepts, or splitting its power and simultaneously transmitting the information vector and an "artificial interference" signal that jams any receivers other than Rx. The artificial interference scheme does not require knowledge of EW's instantaneous CSI, which makes it suitable for deployment against passive intercepts [7, 32–35]. EW also has two options for disrupting the private information rate between Tx and Rx: it can either eavesdrop on Tx or jam Rx, under a half-duplex constraint.

The signal received by Rx from Tx while simultaneously being jammed by EW can be represented as

$$\vec{y}_{\text{Rx}} = \mathbf{H}_{\text{Rx,Tx}}\vec{x}_{\text{Tx}} + \sqrt{g_2}\mathbf{H}_{\text{Rx,EW}}\vec{x}_{\text{EA}} + \vec{n}_{\text{Rx}} \tag{9.66}$$

while the received signal at EW when eavesdropping is

$$\vec{y}_{\text{ES}} = \sqrt{g_1}\mathbf{H}_{\text{EW,Tx}}\vec{x}_{\text{Tx}} + \vec{n}_{\text{ES}} \tag{9.67}$$

where $\vec{x}_{\text{Tx}}$ is the signal vector transmitted by Tx, $\vec{x}_{\text{EA}}$ is the Gaussian jamming signal from EW, $\vec{n}_{\text{Rx}}, \vec{n}_{\text{ES}}$ are the naturally occurring additive noise at Rx and EW, respectively, and $\mathbf{H}_{\text{Rx,Tx}}$, $\mathbf{H}_{\text{Rx,EW}}$, $\mathbf{H}_{\text{EW,Tx}}$ are the corresponding $N_{\text{Rx}}\times N_{\text{Tx}}$, $N_{\text{Rx}}\times N_{\text{EW}}$, $N_{\text{EW}}\times N_{\text{Tx}}$ channel matrices whose elements are independent and drawn from the complex Gaussian distribution $\mathcal{CN}(0, 1)$. That is, $\mathbf{H}_{\text{Rx,Tx}}$, $\mathbf{H}_{\text{Rx,EW}}$, and $\mathbf{H}_{\text{EW,Tx}}$ are fixed, complex gain matrices that model the channel gains between the transmitter and the receivers. The receive and transmit channels of the intercept have gain factors $g_1^{1/2}$ and $g_2^{1/2}$, respectively. These scale factors may be interpreted as an indicator of the relative distances between EW and the other nodes.

The background noise at all receivers is assumed to be spatially white and zero-mean complex Gaussian: $\mathcal{E}\{\vec{n}_k \vec{n}_k^H\} = \sigma_k^2 \mathbf{I}, k \in \{\text{Rx}, \text{ES}\}$. Tx's total transmit power is assumed to be fixed at P_{Tx}. When the transmit sequence is given by $\vec{x}_{\text{Tx}}$, we have

$$\mathcal{E}\{\vec{x}_{\text{Tx}} \vec{x}_{\text{Tx}}^{\text{H}}\} \triangleq \mathbf{Q}_{\text{Tx}}, \qquad \text{Tr}\{\mathbf{Q}_{\text{Tx}}\} = P_{\text{Tx}}$$

and similarly, EW has a fixed power of P_{EA} when in jamming mode. In the most general scenario where Tx jams EW (ES) by transmitting artificial interference, we have

$$\vec{x}_{\text{Tx}} = \mathbf{T}\vec{z} + \mathbf{T}'\vec{z}' \tag{9.68}$$

where $\mathbf{T}$, $\mathbf{T}'$ are $N_{\text{Tx}}\times d$, $N_{\text{Tx}}\times(N_{\text{Tx}} - d)$ precoding matrices for the $d\times 1$ information vector $\vec{z}$ and uncorrelated $(N_{\text{Tx}}-d)\times 1$ jamming signal $\vec{z}'$, respectively. To ensure that the artificial noise does not interfere with the information signal, a common approach is to make these signals orthogonal when received by Rx. If Tx knows $\mathbf{H}_{\text{Rx,Tx}}$, this goal can be achieved by choosing $\mathbf{T}$ and $\mathbf{T}'$ as disjoint sets of the right singular vectors of $\mathbf{H}_{\text{Rx,Tx}}$. The matrix $\mathbf{Q}_{\text{Tx}}$ may be expressed as

$$\mathbf{Q}_{\text{Tx}} = \mathbf{T}\mathbf{Q}_{\text{z}}\mathbf{T}^{\text{H}} + \mathbf{T}'\mathbf{Q}_{\text{z}}'\mathbf{T}'^{\text{H}} \tag{9.69}$$

where $\mathbf{Q}_{\text{z}}$, $\mathbf{Q}_{\text{z}}'$ are the covariance matrices associated with $\vec{z}$ and $\vec{z}'$, respectively. Let ρ denote the fraction of the total power available at Tx that is devoted to the information signal, then $\text{Tr}\{\mathbf{T}\mathbf{Q}_{\text{z}}\mathbf{T}^{\text{H}}\} = \rho P_{\text{Tx}}$ and $\text{Tr}\{\mathbf{T}'\mathbf{Q}_{\text{z}}'\mathbf{T}'^{\text{H}}\} = (1-\rho)P_{\text{Tx}}$. The covariance matrices of the received interference-plus-noise at Rx and EW are

$$\mathbf{Q}_{\text{Rx}} = g_2 \mathbf{H}_{\text{Rx,EW}} \mathbf{Q}_{\text{Rx,EA}} \mathbf{H}_{\text{Rx,EW}}^{\text{H}} + \sigma_{\text{Rx}}^2 \mathbf{I} \tag{9.70}$$

$$\mathbf{Q}_{\mathrm{EA}} = g_1 \mathbf{H}_{\mathrm{EW,Tx}} \mathbf{T}' \mathbf{Q}_{\mathrm{z}}' \mathbf{T}'^{\mathrm{H}} \mathbf{H}_{\mathrm{EW,Tx}}^{\mathrm{H}} + \sigma_{\mathrm{ES}}^2 \mathbf{I} \tag{9.71}$$

where $\mathbf{Q}_{\mathrm{Rx,EA}}$ is the covariance of the jamming signal transmitted by EW. Note that we have assumed that Tx's jamming signal (if any) is orthogonal to the information signal received by Rx, and hence, from the point of view of mutual information, can be ignored in the expression for $\mathbf{Q}_{\mathrm{Rx}}$.

We indicated that Tx knows $\mathbf{H}_{\mathrm{Rx,Tx}}$ in order to appropriately precode the jamming and information signals via $\mathbf{T}$ and $\mathbf{T}'$. At EW we will assume that EW knows the channel to Tx, $\mathbf{H}_{\mathrm{EW,Tx}}$, and the covariance $\mathbf{Q}_{\mathrm{EA}}$ of the interference and noise, and similarly we will assume that Rx knows $\mathbf{H}_{\mathrm{Rx,Tx}}$ and $\mathbf{Q}_{\mathrm{Rx}}$. All other CSI at the various nodes is assumed to be unavailable; the only available information is the assumption that the channels are composed of independent $\mathcal{CN}(0, 1)$ random variables. Therefore when EW jams Rx the half-duplex constraint prevents it from detecting the transmitted signal $\vec{z}$ and applying correlated jamming [26]. Consequently, it will uniformly distribute its available power over all N_{EW} transmit dimensions, so that $\mathbf{Q}_{\mathrm{Rx,EW}} = (P_{\mathrm{EA}} / N_{\mathrm{EW}})\mathbf{I}$. Similarly, when Tx transmits a jamming signal, it will also be uniformly distributed across the $N_{\mathrm{Tx}} - d$ available dimensions: $\mathbf{Q}_{\mathrm{z}}' = (1-\rho)P_{\mathrm{Tx}} / (N_{\mathrm{Tx}} - d)\mathbf{I}$. While Tx could use its knowledge of $\mathbf{H}_{\mathrm{Rx,Tx}}$ to perform power loading, for simplicity we will assume that the power of the information signal is also uniformly distributed, so that $\mathbf{Q}_{\mathrm{z}} = (\rho P_{\mathrm{Tx}} / d)\mathbf{I}$. In most situations, the degradation due to this latter assumption is minimal.

Given the above, (9.69)–(9.71) simplify to

$$\mathbf{Q}_{\mathrm{Tx}} = \frac{\rho P_{\mathrm{Tx}}}{d} \mathbf{T}\mathbf{T}^{\mathrm{H}} + \eta_{\mathrm{Tx}} \mathbf{T}'\mathbf{T}'^{\mathrm{H}} \tag{9.72}$$

$$\mathbf{Q}_{\mathrm{Rx}} = \frac{g_2 P_{\mathrm{EW}}}{N_{\mathrm{EW}}} \mathbf{H}_{\mathrm{Rx,EW}} \mathbf{H}_{\mathrm{Rx,EW}}^{\mathrm{H}} + \sigma_{\mathrm{Rx}}^2 \mathbf{I} \tag{9.73}$$

$$\mathbf{Q}_{\mathrm{EA}} = g_1 \eta_{\mathrm{Tx}} \mathbf{H}_{\mathrm{EW,Tx}} \mathbf{T}' \mathbf{T}'^{\mathrm{H}} \mathbf{H}_{\mathrm{EW,Tx}}^{\mathrm{H}} + \sigma_{\mathrm{ES}}^2 \mathbf{I} \tag{9.74}$$

where η_{Tx} is defined as

$$\eta_{\mathrm{Tx}} = \frac{(1-\rho)P_{\mathrm{Tx}}}{N_{\mathrm{Tx}} - d} \tag{9.75}$$

The MIMO privacy capacity between Tx and Rx is obtained by solving

$$C_{\mathrm{p}} = \max_{\mathbf{Q}_{Tx} \succeq \mathbf{0}} I(X_{\mathrm{Tx}}; Y_{\mathrm{Rx}}) - I(X_{\mathrm{Tx}}; Y_{\mathrm{ES}}) \tag{9.76}$$

where X_{Tx}, Y_{Rx}, Y_{ES} are the random variable counterparts of the realizations $\vec{x}_{\mathrm{Tx}}, \vec{y}_{\mathrm{Tx}}, \vec{y}_{\mathrm{ES}}$. Such an optimization cannot be performed, however, since Tx is unaware of the instantaneous values of all channels and interference covariance matrices. Consequently, we work with the achievable lower bound on the MIMO privacy capacity based on Gaussian inputs and uniform power allocation at all transmitters:

$$C_{\mathrm{p}} \geq \mathcal{E}_{\mathbf{H}} \left\{ \begin{array}{l} \log_2 \left| \mathbf{I} + \dfrac{\rho P_{Tx}}{d} \mathbf{H}_{\mathrm{Rx,Tx}} \mathbf{T}\mathbf{T}^{\mathrm{H}} \mathbf{H}_{\mathrm{Rx,Tx}}^{\mathrm{H}} \mathbf{Q}_{\mathrm{Rx}}^{-1} \right| \\ \qquad -\log_2 \left| \mathbf{I} + \dfrac{\rho P_{\mathrm{Tx}}}{d} \mathbf{H}_{\mathrm{EW,Tx}} \mathbf{T}\mathbf{T}^{\mathrm{H}} \mathbf{H}_{\mathrm{EW,Tx}}^{\mathrm{H}} \mathbf{Q}_{\mathrm{EA}}^{-1} \right| \end{array} \right\} \tag{9.77}$$

where $\mathbf{H} \triangleq \{\mathbf{H}_{\mathrm{Rx,Tx}}, \mathbf{H}_{\mathrm{Rx,EW}}, \mathbf{H}_{\mathrm{EW,Tx}}\}$. Note that the expectation is taken over all channel matrices to provide Tx and EW with a common objective function, since neither possesses complete knowledge of **H**. While EW's decision is binary; whether to allocate resources to eavesdroping or allocate all power to jamming, Tx's options include determining how the spatial dimensions are to be split between the data and artificial noise (if any), and what is the optimal fraction ρ that determines the transmit power allocated to them. It turns out that the achievable privacy rate is not very sensitive to these parameters, and good (communication) performance can be obtained for a wide range of reasonable values. The general approach of this section is applicable to essentially any value for ρ and *d*, although the specific results we present assume that optimal values have been chosen. Consequently, as with EW, this reduces the number of Tx's decisions to two: either transmitting the information signal with full power to Rx, or devoting an optimal amount of resources, that is, power and spatial dimensions, to a jamming signal aimed towards EW while transmitting some energy toward Rx.

9.7.2 Privacy Rate Approximations

We denote EW's set of possible actions as {*E*, *J*} to indicate either "Eavesdropping" or "Jamming," while Tx's will be expressed as {*F*, *A*} to indicate "Full-power" devoted to the information signal, or a nonzero fraction of the power allocated to "Artificial noise." The privacy rates that result from the resulting four possible scenarios will be denoted by R_{ik}, where $i \in \{F, A\}$ and $k \in \{E, J\}$. While (9.77) could be directly evaluated to determine the set of possible privacy rates for a given scenario, in this section we will investigate the problem of finding simpler approximate expressions that will facilitate comparisons between different scenarios.

9.7.2.1 Asymptotic MIMO Rates

Let $\mathbf{X}$ represent an $N_{\text{Tx}}\times N_{\text{Rx}}$ MIMO channel matrix composed of $\mathcal{CN}(0,\ 1)$ elements over which a Tx transmits to Rx with signal-to-noise ratio (SNR) at Rx, γ. When interference is absent (including jamming) and thermal noise is the only impairment at the receiver, the MIMO information rate is given by

$$R = \mathcal{E}\left\{\log_2\left|\mathbf{I} + \frac{\gamma}{N_{\text{Rx}}}\mathbf{X}\mathbf{X}^{\text{H}}\right|\right\}$$

assuming a uniform power allocation. Let λ represent an arbitrary eigenvalue of the matrix $(1/N_{\text{Rx}})\mathbf{X}\mathbf{X}^{\text{H}}$. Since the determinant of a matrix is equal to the product of its eigenvalues, from the identical distribution of the eigenvalues we can write

$$R = \min(N_{\text{Tx}}, N_{\text{Rx}})\mathcal{E}_{\lambda}\{\log_2(1+\gamma\lambda)\} \tag{9.78}$$

The closed-form expression for this expectation is well known in the literature [31]. However, a more tractable expression for the MIMO capacity is available based on the limit of a large number of antennas, as described next.

For a matrix $(1/N_{\text{Rx}})\mathbf{X}\mathbf{X}^{\text{H}}$ where $\mathbf{X}$ is $N_{\text{Tx}}\times N_{\text{Rx}}$, the asymptotic marginal pdf of an arbitrary (unordered) eigenvalue is known to be

$$p(\lambda_{\text{a}}) = \begin{cases} \dfrac{1}{\pi}\sqrt{\dfrac{\beta}{\lambda_{\text{a}}} - \dfrac{1}{4}\left(1+\dfrac{\beta-1}{\lambda_{\text{a}}}\right)^2}, & (\sqrt{\beta}-1)^2 \le \lambda_{\text{a}} \le (\sqrt{\beta}+1)^2 \\ 0, & \text{otherwise} \end{cases} \tag{9.79}$$

where $\beta = N_{\text{Tx}}/N_{\text{Rx}}$. Based on (9.79), a useful closed-form expression for the capacity is [36, 37]

$$\min(N_{\text{Tx}}, N_{\text{Rx}})\mathcal{E}_{\lambda_a}\{\log_2(1+\gamma\lambda_a)\} = N_{\text{Tx}}F(\beta,\gamma) \tag{9.80}$$

where

$$F(\beta,\gamma)=\log_2\left[1+\gamma\left(\sqrt{\beta}+1\right)^2\right]+(\beta+1)\log_2\left(\frac{1+\sqrt{1-a}}{2}\right)$$
$$-\log_2(e)\sqrt{\beta}\frac{1-\sqrt{1-a}}{1+\sqrt{1-a}}+(\beta-1)\log_2\left(\frac{1+\kappa}{\kappa+\sqrt{1-a}}\right) \tag{9.81}$$

with

$$a=\frac{4\gamma\sqrt{\beta}}{1+\gamma(\sqrt{\beta}+1)^2} \qquad \kappa=\frac{\sqrt{\beta}-1}{\sqrt{\beta}+1} \tag{9.82}$$

Though originally derived under an asymptotic assumption, (9.80) has been shown to be very accurate even for small- and medium-sized antenna arrays.

Now, in addition to the target signal over channel **X** with SNR γ, let **Y** be the $N_{\text{Rx}}\times N_{\text{EW}}$ channel from EA to Rx with *jammer-to-noise ratio* (JNR) η and $\mathcal{CN}(0, 1)$ elements. Assuming uniform power allocation at both Tx and EA, the MIMO information rate with interference and Gaussian background noise is given by

$$R_{\text{I}}=\mathcal{E}_{\mathbf{X},\mathbf{Y}}\left\{\log_2\left|\mathbf{I}+\frac{\gamma}{N_{\text{Rx}}}\mathbf{X}\mathbf{X}^{\text{H}}\left(\mathbf{I}+\frac{\eta}{N_{\text{EW}}}\mathbf{Y}\mathbf{Y}^{\text{H}}\right)^{-1}\right|\right\} \tag{9.83}$$

Property 9.7 [38]: In a MIMO channel where Tx, Rx, and EW have N_{Tx}, N_{Rx}, N_{EW} antennas, respectively, the asymptotic MIMO information rate with receive SNR γ and JNR η can be bounded as

$$R_{\text{I}}\le(N_{\text{Tx}}+N_{\text{EW}})F\left[\frac{N_{\text{Rx}}}{N_{\text{Tx}}+N_{\text{EW}}},(\gamma+\eta)\right]-N_{\text{EW}}F\left(\frac{N_{\text{Rx}}}{N_{\text{EW}}},\eta\right) \tag{9.84}$$

where $F(\beta, \gamma)$ is defined in (9.81).

9.7.2.2 MIMO Privacy Rate Analysis

Define the effective channels conveying information $\vec{z}$ from Tx to Rx and EW as $\tilde{\mathbf{H}}_{\text{Rx,Tx}}\triangleq\mathbf{H}_{\text{Rx,Tx}}\mathbf{T}$ and $\tilde{\mathbf{H}}_{\text{EW,Tx}}\triangleq\mathbf{H}_{\text{EW,Tx}}\mathbf{T}$, respectively. Since **T** is a submatrix of an isotropically random unitary matrix, $\tilde{\mathbf{H}}_{\text{Rx,Tx}}$ and $\tilde{\mathbf{H}}_{\text{EW,Tx}}$ are also zero-mean complex

Gaussian matrices with i.i.d elements. The elements of $\tilde{\mathbf{H}}_{\mathrm{Rx,Tx}}$, however, will, in general, have a variance greater than unity because the data is concentrated in a subset of the spatial subchannels corresponding to the stronger singular values. In order to apply the random matrix results stated previously, it is necessary to normalize the effective channel $\tilde{\mathbf{H}}_{\mathrm{Rx,Tx}}$ to obtain elements with unit variance. The exact normalization constant is difficult to obtain analytically, and since we are dealing with an upper bound on the achievable rate, our results will be based on scaling $\tilde{\mathbf{H}}_{\mathrm{Rx,Tx}}$ by an approximate factor $\sqrt{d / N_{\mathrm{Tx}}}$. The inverse of this factor, which represents an upper bound on the increase in the variance of the elements of $\tilde{\mathbf{H}}_{\mathrm{Rx,Tx}}$, is absorbed into the transmit power constraint.

Assuming Gaussian inputs $\vec{z}$ and $\vec{z}\,'$, the MIMO privacy rate between Tx and Rx when EW is in eavesdropping mode is

$$R_{\mathrm{iE}} = \mathcal{E}_{\mathbf{H}} \left\{ \begin{array}{l} \log_2 \left| \mathbf{I} + \dfrac{\rho P_{\mathrm{Tx}}}{d\sigma_{\mathrm{Rx}}^2} \mathbf{H}_{\mathrm{Rx,Tx}} \mathbf{T}\mathbf{T}^{\mathrm{H}} \mathbf{H}_{\mathrm{Rx,Tx}}^{\mathrm{T}} \right| \\ \qquad - \log_2 \left| \mathbf{I} + \dfrac{g_1 \rho P_{\mathrm{Tx}}}{d} \mathbf{H}_{\mathrm{EW,Tx}} \mathbf{T}\mathbf{T}^{\mathrm{H}} \mathbf{H}_{\mathrm{EW,Tx}}^{\mathrm{H}} \mathbf{Q}_{\mathrm{EA}}^{-1} \right| \end{array} \right\} \tag{9.85}$$

whereas the privacy rate when EW is jamming reduces to

$$R_{\mathrm{iJ}} = \mathcal{E}_{\mathbf{H}} \left\{ \log_2 \left| \mathbf{I} + \frac{\rho P_{\mathrm{Tx}}}{d} H_{\mathrm{Rx,Tx}} \mathbf{T}\mathbf{T}^{\mathrm{H}} \mathbf{H}_{\mathrm{Rx,Tx}}^{\mathrm{H}} \mathbf{Q}_{\mathrm{Rx}}^{-1} \right| \right\} \tag{9.86}$$

where $i \in \{F, A\}$ denotes the transmission strategies available to Tx. When Tx transmits with full power, then $d = r$, where $r = \min(N_{\mathrm{Tx}}, N_{\mathrm{Rx}})$, and the precoder $\mathbf{T}$ consists of the right singular vectors of $\mathbf{H}_{\mathrm{Rx,Tx}}$ corresponding to the r largest singular values. In view of (9.80) and (9.84), the asymptotic MIMO privacy rate outcomes therefore are

$$\begin{aligned} R_{\mathrm{AE}} \le\ & dF\left(\frac{N_{\mathrm{Rx}}}{d}, \frac{\rho P_{\mathrm{Tx}} N_{\mathrm{Tx}}}{d\sigma_{Rx}^2} \right) \\ & - \left[N_{\mathrm{Tx}} F\left(\frac{N_{\mathrm{EW}}}{N_{\mathrm{Tx}}}, \frac{g_1 P_{\mathrm{Tx}}}{\sigma_{\mathrm{ES}}^2} \right) - (N_{\mathrm{Tx}} - d) F\left(\frac{N_{\mathrm{EW}}}{N_{\mathrm{Tx}} - d}, \frac{g_1 (1-\rho) P_{\mathrm{Tx}}}{\sigma_{\mathrm{ES}}^2} \right) \right] \end{aligned} \tag{9.87}$$

$$R_{\mathrm{AJ}} \le (N_{\mathrm{EW}} + d) F\left(\frac{N_{\mathrm{Rx}}}{N_{\mathrm{EW}} + d}, \frac{\rho N_{\mathrm{Tx}} P_{\mathrm{Tx}}}{d\sigma_{\mathrm{Rx}}^2} + \frac{g_2 P_{\mathrm{EA}}}{\sigma_{\mathrm{Rx}}^2} \right) - N_{\mathrm{EW}} F\left(\frac{N_{\mathrm{Rx}}}{N_{\mathrm{EW}}}, \frac{g_2 P_{\mathrm{EA}}}{\sigma_{\mathrm{Rx}}^2} \right) \tag{9.88}$$

$$R_{\mathrm{FE}} \le N_{\mathrm{Tx}} F\left(\frac{N_{\mathrm{Rx}}}{N_{\mathrm{Tx}}}, \frac{P_{\mathrm{Tx}}}{\sigma_{\mathrm{Rx}}^2}\right) - N_{\mathrm{Tx}} F\left(\frac{N_{\mathrm{EW}}}{N_{\mathrm{Tx}}}, \frac{g_1 P_{\mathrm{Tx}}}{\sigma_{\mathrm{ES}}^2}\right) \tag{9.89}$$

$$R_{\mathrm{FJ}} \le (N_{\mathrm{EW}} + N_{\mathrm{Tx}}) F\left(\frac{N_{\mathrm{Rx}}}{N_{\mathrm{EW}} + N_{\mathrm{Tx}}}, \frac{P_{\mathrm{Tx}} + g_2 P_{\mathrm{EW}}}{\sigma_{\mathrm{RX}}^2}\right) - N_{\mathrm{EW}} F\left(\frac{N_{\mathrm{Rx}}}{N_{\mathrm{EW}}}, \frac{g_2 P_{\mathrm{EA}}}{\sigma_{\mathrm{Rx}}^2}\right) \tag{9.90}$$

The asymptotic rates in (9.87)–(9.90) demonstrate reasonable accuracy even for relatively small antenna arrays.

In the following game theoretic analysis, we will assume that the following reasonable conditions are always satisfied for any MIMO wiretap scenario:

(C1) $R_{\mathrm{FE}} \le R_{\mathrm{AE}}$, which is based on the existing literature on the MIMO wiretap channel, which demonstrates that judicious use of artificial noise always improves the privacy rate;
(C2) $R_{\mathrm{AJ}} \le R_{\mathrm{FJ}}$, since allocating power for artificial noise when EW jams Rx can only decrease the privacy rate for Tx.

It is useful to examine the behavior of the rates for several limiting cases. Consider the scenario where N_{Rx} and N_{EW} both grow large with respect to N_{Tx}, that is, $N_{\mathrm{Rx}}/N_{\mathrm{Tx}} \to \infty$, $N_{\mathrm{EW}}/N_{\mathrm{Tx}} \to \infty$, and $N_{\mathrm{Rx}}/N_{\mathrm{EW}} \to 1$, while transmit powers and channel gains remain finite. We can show that $F(\beta,\gamma) \approx \log_2(\gamma\beta)\,\mathrm{as}\,\beta \to \infty$ and $F(1,\gamma) \approx \log_2(1+4\gamma) - 2 - \log_2(e)\,\mathrm{if}\,\beta \to 1.$ Consequently, for this large-antenna scenario we obtain

$$\begin{aligned} R_{\mathrm{AE}} \le\ & d \log_2\left(\frac{\rho N_{\mathrm{Rx}} N_{\mathrm{Tx}} P_{\mathrm{Tx}}}{d^2 \sigma_{\mathrm{Rx}}^2}\right) - N_{\mathrm{Tx}} \log_2\left(\frac{N_{\mathrm{EW}} g_1 P_{\mathrm{Tx}}}{N_{\mathrm{Tx}} \sigma_{\mathrm{ES}}^2}\right) \\ & + (N_{\mathrm{Tx}} - d) \log_2\left[\frac{N_{\mathrm{EW}} g_1 (1-\rho) P_{\mathrm{Tx}}}{(N_{\mathrm{Tx}} - d)\sigma_{\mathrm{ES}}^2}\right] \end{aligned} \tag{9.91}$$

$$R_{\mathrm{FE}} \le N_{\mathrm{Tx}} \log_2\left(\frac{N_{\mathrm{Rx}} P_{\mathrm{Tx}}}{N_{\mathrm{Tx}} \sigma_{\mathrm{Rx}}^2}\right) - N_{\mathrm{Tx}} \log_2\left(\frac{N_{\mathrm{EW}} g_1 P_{\mathrm{Tx}}}{N_{\mathrm{Tx}} \sigma_{\mathrm{ES}}^2}\right) \tag{9.92}$$

$$\begin{aligned} R_{\mathrm{AJ}} \le\ & (N_{\mathrm{EW}} + d) \log_2\left[\frac{N_{\mathrm{Rx}}}{\sigma_{\mathrm{Rx}}^2 (N_{\mathrm{EW}} + d)}\left(\frac{\rho N_{\mathrm{Tx}} P_{\mathrm{Tx}}}{d} + g_2 P_{\mathrm{EA}}\right)\right] \\ & - N_{\mathrm{EW}} \log_2\left(\frac{g_2 N_{\mathrm{Rx}} P_{\mathrm{EA}}}{\sigma_{\mathrm{Rx}}^2 N_{\mathrm{EW}}}\right) \end{aligned} \tag{9.93}$$

		EW	
		Eavesdrop (E)	Jam Rx (J)
Tx	Full Power (F)	R_{FE}	R_{FJ}
	Artificial Noise (A)	R_{AE}	R_{AJ}

Figure 9.21 Payoff matrix **R** of the strategic form of the MIMO wiretap game.

$$R_{FJ} \le (N_{EW} + N_{Tx}) \log_2 \left[\frac{N_{Rx}}{\sigma_{Rx}^2 (N_{EW} + N_{Tx})} (P_{Tx} + g_2 P_{EA}) \right] - N_{EW} \log_2 \left(\frac{g_2 N_{Rx} P_{EA}}{\sigma_{Rx}^2 N_{EW}} \right) \tag{9.94}$$

Comparing the positive terms in (9.91) and (9.92), we observe that the sum of the positive terms in (9.91) exceeds the positive term in (9.92), which implies (C1). Similarly, the pre-log factor of the positive term in (9.94) is at least as large as the corresponding pre-log factor in (9.93), leading to (C2). Similar conclusions can be reached for C1 and C2 by analyzing the other parameters.

9.7.3 Strategic Wiretap Game

In this section we construct a zero-sum model of the wiretap game using the rate results derived in the previous section. The payoff to Tx is defined as the achievable MIMO privacy rate between Tx and Rx (the larger the better for Tx and Rx). The strategic interactions between Tx and EW form a zero-sum game, where Tx tries to maximize its payoff and EW attempts to minimize it. As before, we define the following strategy sets $\mathcal{X}$, $\mathcal{Y}$ for the players: Tx chooses between transmitting with full power for data (F) or devoting some power to jam EW's receiver (A), denoted as $\mathcal{X} = \{F, A\}$. However, EW must select eavesdropping (E) or jamming Rx (J) at every channel use, represented by $\mathcal{Y} = \{E, J\}$.

9.7.3.1 Pure-Strategy Equilibria

When both Tx and EW move simultaneously without knowledge of the action taken by the other, the *strategic form* of the game can be represented by the 2 × 2 payoff matrix **R** in Figure 9.21. In this game, a pure strategy choice would entail Tx always playing a fixed $x \in \mathcal{X}$ and EW playing the same $y \in \mathcal{Y}$ at every transmission interval.

Property 9.8 [38]: For an arbitrary set of antenna array sizes, transmit powers, and channel gain parameters, a unique pure strategy NE exists in the MIMO wiretap game according to:

$$R(x^*, y^*) = \begin{cases} R_{\text{AE}}, & R_{\text{AE}} \le R_{\text{AJ}} \\ R_{\text{FJ}}, & R_{\text{FJ}} \le R_{\text{FE}} \end{cases} \tag{9.95}$$

Of the 24 possible orderings of the four rates, only six satisfy conditions (C1) and (C2). Furthermore, only two of these six mutually exclusive rates result in a pure NE.

If $R_{\text{AE}} \le R_{\text{AJ}}$, then (C1) and (C2) imply the following rate ordering

$$R_{\text{FJ}} \ge R_{\text{AJ}} \ge \underset{\text{NE}}{R_{\text{AE}}} \ge R_{\text{FE}} \tag{9.96}$$

In this case, R_{AE} represents an NE since neither Tx or EW can improve their respective payoffs by switching strategies.

Similarly, when $R_{\text{FJ}} \le R_{\text{FE}}$, then (C1) and (C2) result in the rate ordering

$$R_{\text{AE}} \ge R_{\text{FE}} \ge \underset{\text{NE}}{R_{\text{FJ}}} \ge R_{\text{AJ}} \tag{9.97}$$

and R_{FJ} will be the mutual best response for both players. Only one such ordering can be true for a given wiretap game scenario.

9.7.3.2 Mixed-Strategy Equilibria

Property 9.8 indicates that there is no single pure strategy choice that is always optimal for either player if the inequalities in (9.96) and (9.97) are not satisfied. This occurs in four of the six valid rate orderings of the entries of **R** that satisfy (C1) and (C2). Since the minimax theorem guarantees that any finite zero-sum game has a saddle-point in randomized strategies [39], in such scenarios Tx and EW must randomize over $\mathcal{X} \times \mathcal{Y}$; that is, they must adopt mixed strategies.

Let $\vec{p} = (p, 1-p)$ and $\vec{q} = (q, 1-q), 0 \le p, q \le 1$, represent the probabilities with which Tx and EW randomize over their strategy sets $\mathcal{X} = \{F, A\}$ and $\mathcal{Y} = \{E, J\}$, respectively. Tx obtains its optimal strategy by solving

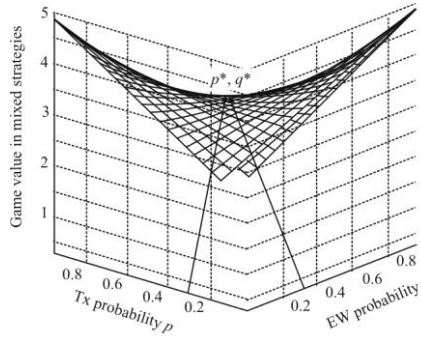


Figure 9.22 Game value in mixed strategies as the mixing probabilities at Tx and EW are varied. $N_{Tx} = 5$, $N_{Rx} = 3$, $N_{EW} = 4$, $d = 2$, $P_{EA} = 20$ dBw, and $g_1 = 1.1$, $g_2 = 0.9$.

$$\max_{p} \min_{q} \vec{p}^{\mathrm{T}} \mathbf{R} \vec{q} \tag{9.98}$$

while EW optimizes the corresponding minimax problem

$$\min_{p} \max_{q} \vec{p}^{\mathrm{T}} \mathbf{R} \vec{q} \tag{9.99}$$

For the payoff matrix **R** in Figure 9.21, the optimal mixed strategies and expected value v of the game are [38]

$$\begin{aligned} \vec{p}^{*} &= (p^{*}, 1-p^{*}) = (R_{AJ} - R_{AE}, R_{FE} - R_{FJ}) / D \\ \vec{q}^{*} &= (q^{*}, 1-q^{*}) = (R_{AJ} - R_{FJ}, R_{FE} - R_{AE}) / D \\ v(p^{*}, q^{*}) &= (R_{FE} R_{AJ} - R_{FJ} R_{AE}) / D \end{aligned} \tag{9.100}$$

where $D = R_{FE} + R_{AJ} - R_{FJ} - R_{AE}$. A graphical illustration of the saddle-point in mixed strategies as p and q are varied for a specific intercept channel is shown in Figure 9.22. For the specified parameters $N_a = 5$, $N_{Rx} = 3$, $N_{EW} = 4$, $d = 2$, $P_{Tx} = P_{EA} = 20$ dBw, $g_1 = 1.1$, $g_2 = 0.9$, the rate ordering turns out to be $R_{AE} = 5.04 > R_{FJ} = 5.02 > R_{AJ} = 2.85 > R_{FE} = 0$, which results in a mixed NE with optimal mixing probabilities ($p* = 0.3$, $q* = 0.3$) and value $v = 3.45$. Tx's bias towards playing $x = A$ more frequently is expected since that guarantees a privacy rate of at least 2.85, whereas playing $x = F$ risks a worst-case payoff of zero. EW is privy to Tx's reasoning and is therefore biased towards playing $y = J$ more frequently.

9.7.3.3 Wiretap Channel Configurations

Reference [38] delineates five wiretap channel cases to determine what the equilibrium outcomes of the corresponding game would be. In particular, either

N_{EW} or P_{EA} are specified relative to Tx and Rx's parameters, while assuming the other variables are of comparable magnitude. We examine these five cases in the following.

Case 1: All users have the same number of antennas, $N_{Tx} = N_{Rx} = N_{EW}$ and comparable power $P_{Tx} \approx P_{EA}$:

• If EW is near Tx ($g_1 \gg 1$, $g_2 \to 0$), then applying $F(\beta,0) \approx 0, F(\beta,\infty) \approx \log 2(\gamma)$ in (9.87)–(9.90), the resultant rate ordering is given by (9.96) with a pure NE in R_{AE}. This outcome is when EW is very close to Tx, and therefore being much more capable of affecting the privacy rate by intercepting as compared to jamming Rx from a distance.

• If EW is proximate to Rx ($g_2 \gg 1$, $g_1 \to 0$), the resultant rate ordering is given by (9.97), and we have a pure NE in R_{FJ}. EW can exploit its proximity to Rx to drown the signal received at Rx's antenna array from Tx via jamming.

Case 2: EW has more antennas than Tx and Rx ($N_{EW} > N_{Tx} = N_{Rx}$) and comparable power $P_{Tx} \approx P_{EA}$:

• If $N_{\mathrm{EW}} \geq 3(N_{\mathrm{Tx}} + N_{\mathrm{Rx}})$ in the large-antenna regime, then [7] showed that the MIMO privacy capacity with complete intercept CSI is zero, which implies that the privacy rate with artificial noise is also zero. Therefore, we have the rate ordering of (9.96), and a pure NE in R_{AE}. Thus, Tx should devote some resources to jamming EW while EW should eavesdrop.

• If $N_{\mathrm{Tx}} < N_{\mathrm{EW}} < 3(N_{\mathrm{Tx}} + N_{\mathrm{Rx}})$, as N_{EW} approaches N_{Tx}, that is, as R_{AE} and R_{FE} increase from zero, the game outcome changes to a mixed NE for the value of N_{EW} at which the rate ordering changes to $R_{\mathrm{FJ}} > R_{\mathrm{AE}} > R_{\mathrm{AJ}} > R_{\mathrm{FE}}$.

Case 3: EW has fewer antennas than Tx and Rx ($N_{EW} < N_{Tx}$, $N_{EW} < N_{Rx}$, $N_{Tx} \geq N_{Rx}$) and comparable power $P_{Tx} \approx P_{EW}$:

• If $N_{\mathrm{EW}} \leq (N_{\mathrm{Tx}} - d)$, then EW cannot suppress the artificial interference and therefore it chooses $y = J$ and the NE is R_{FJ}, where Tx devotes all resources to transmitting to Rx.

• If $(N_{\mathrm{Rx}} - d) \geq N_{\mathrm{EW}}$, then Rx can suppress EW if EW jams and also recover the private message from Tx, therefore EW should intercept and the NE is R_{AE}.

Case 4: Highly advantaged EW:

• $N_{\mathrm{EW}} \gg N_{\mathrm{Tx}} = N_{\mathrm{Rx}}$, $P_{\mathrm{EA}} = P_{\mathrm{Tx}}$: If EW has an overwhelming advantage in the size of its antenna array, then for standard values of (g_1, g_2) all four rate outcomes are zero; as a result both players are indifferent to the choice of strategies so any can be chosen.

• $P_{\mathrm{EA}} \gg P_{\mathrm{Tx}}$, $N_{\mathrm{EW}} \approx N_{\mathrm{Tx}} = N_{\mathrm{Rx}}$: As $P_{\mathrm{EA}} \to \infty$, we have the rate ordering of (9.98), and a pure NE in R_{FJ} so Tx should devote full resources to transmitting to Rx and EW should jam Rx.

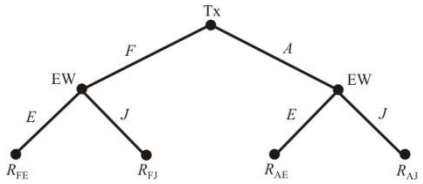


Figure 9.23 Extensive form game tree with perfect information $\Gamma^{e,1}$ where Tx moves first and EW moves second.

Case 5: Highly disadvantaged EW:

• The worst-case scenario for EW is $N_{EW} = 1$, $N_{Tx} = N_{Rx} \gg 1$. Since EW is unable to separate and decode the multiple streams sent by Tx with a single receive antenna, its best response is to play $y = J$ and we have a pure NE in R_{FJ} .

• When $P_{EA} \ll P_{Tx}$, EW must almost surely eavesdrop, and we have a pure NE in R_{AE}, so that Tx devotes some of its resources to jamming EW and some to communicate with Rx.

For general scenarios not covered above, a mixed-strategy NE as defined in (9.100) is the most probable outcome.

The following examples illustrate the use of the analysis delineated in the above cases.

> **Example 9.6(a):** $N_{EW} = 5$, $N_{Tx} = 4$, $N_{Rx} = 3$, $d = 2$, $P_{Tx} = P_{EA} =$ 20 dBw, $g_1 = 0.8$, $g_2 = 1.1$. From Section 9.7.3 this configuration yields the rate ordering $R_{FJ} = 3.56 > R_{AJ} = 1.69 > R_{AE} = 1.14 > R_{FE} = 0$, with a pure NE in R_{AE}. In this case, Tx should devote some resources to jamming EW and EW eavesdrops.

> **Example 9.6(b):** $N_{EW} = 2$, $N_{Tx} = N_{Rx} = 10$, $d = 8$, $P_{Tx} = 20$ dBw, $P_{EW} = 30$ dBw, $g_1 = 1.25$, $g_2 = 0.75$. This configuration yields the rate ordering $R_{FJ} = 39.88 > R_{AE} = 34.98 > R_{AJ} = 34.73 > R_{FE} = 32.69$, with a mixed NE of value $v = 34.85$ obtained by Tx mixing strategies as ($p* = 0.05$, $1 - p* = 0.95$), and EW mixing over ($q* = 0.48$, $1 - q* = 0.52$). Thus, Tx overwhelmingly would devote resources to jamming EW, while EW has a slight preference to jam over eavesdropping, although the tendency is small.

9.7.4 Extensive Form Intercept Game

While the strategic game with simultaneous moves provides some insight, in practice it is more reasonable to expect one of the players to move first, followed

by the opponent's response. Accordingly, we examine the sequential or *extensive form* of the MIMO intercept game in this section (see footnote 3 in Chapter 4). We begin with the worst-case scenario where Tx moves first by either playing F or A, which is observed by EW who responds accordingly. This would be the most common sequence in real situations, as EW would respond to the communication attempts of the target network.

It is convenient to represent the sequential nature of an extensive form game with a rooted tree or directed graph, as shown in Figure 9.23. The payoffs for Tx are shown at each terminal node, while the corresponding payoffs for EW are omitted for clarity due to the zero-sum assumption. We examine extensive-form games with and without perfect information, and the variety of equilibrium solution concepts available for them.

9.7.4.1 Perfect Information

When EW can distinguish which move was adopted by Tx, and furthermore determine the exact jamming power $(1-\rho)P_{\text{Tx}}$ if EW is being jammed by Tx, then the extensive game is classified as one of *perfect information*. In the following, we will make use of the notions of an information state and a subgame.

A player's *information state* represents the node(s) on the decision tree at which it must make a move conditioned on its knowledge of the previous move of the opponent. For the case of perfect information in Figure 9.23 where Tx moves first, Tx has a single information state, while EW has two information states (each with a single node) based on Tx's choice, since it has perfect knowledge of Tx's move.

A *subgame* is a subset (subgraph) of a game that starts from an information state with a single node, contains all and only that node's successors in the tree, and contains all or none of the nodes in each information state [40].

Next, we analyze *subgame-perfect equilibria* (SPE) of the extensive game, which are a more refined form of NE that eliminate irrational choices within subgames [39]. It is well known that in extensive games with perfect information, a sequential equilibrium in pure strategies is guaranteed to exist [40]. The equilibrium strategies can be obtained by a process of backward induction on the extensive game tree.

> **Property 9.9(a)** [38]: In the extensive form intercept game $\Gamma^{e,1}$ with perfect information where Tx moves first, the unique SPE in pure strategies is determined by the following:

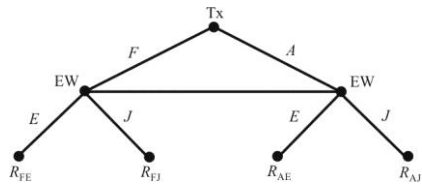


Figure 9.24 Extensive game tree with imperfect information.

$$\text{SPE}(\Gamma^{e,1}) = \begin{cases} R_{AE}, & R_{AE} \le R_{AJ} \\ R_{FJ}, & R_{FJ} \le R_{FE} \\ \max(R_{FE}, R_{AJ}) & R_{AJ} > R_{AE} \text{ or } R_{FJ} > R_{FE} \end{cases} \tag{9.101}$$

For completeness, we provide below the subgame-perfect equilibrium of the dual game where EW moves first but this is the less likely scenario in reality.

> **Property 9.9(b)** [38]**:** The extensive form game $\Gamma^{e,2}$ with perfect information where EW moves first and Tx moves second has the following SPE:
>
> $$\text{SPE}(\Gamma^{e,2}) = \min(R_{FJ}, R_{AE}) \tag{9.102}$$

9.7.4.2 Imperfect Information

Now consider the extensive game Γ_f^e of imperfect information between Tx and EW, where Tx moves first, but EW is uncertain of the exact strategy adopted by Tx. The game tree representation of $\Gamma_f^{e,1}$ is drawn by connecting the decision nodes of EW in Figure 9.23, as shown in Figure 9.24, to indicate its inability to correctly determine Tx's move in the initial phase of the game. Thus, in this case, there is only a single information state for EW. While no player has an incentive to randomize in the game with perfect information in Section 9.7.4.1, mixed strategies enter the discussion when the game is changed to one of imperfect information. The SPE solution is generally unsatisfactory for such games, since the only valid subgame in this case is the entire game $\Gamma_f^{e,1}$ itself. Therefore, *sequential equilibrium* is a stronger solution concept better suited for extensive games of imperfect information.

Consider the special case where it is common knowledge at all nodes that EW is completely unable to determine what move was made by Tx in the first stage of the game. Let EW assign the a priori probabilities $(\xi, 1-\xi)$ to Tx's moves over

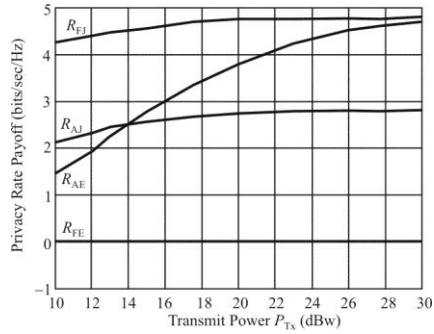


Figure 9.25 Strategic MIMO wiretap game for $P_{EA} = 4P_{Tx}$, $N_{Tx} = N_{EW} = 8$, $N_{Rx} = 6$, $g_1 = 1.2$, $g_2 = 0.75$, as a function P_{Tx}.

$\{F, A\}$ for some ρ and d, while EW randomizes over $\{E, J\}$ with probabilities $(\kappa, 1-\kappa)$. EW's expected payoff in this case can be expressed as

$$-\xi\left[\kappa R_{FE} + (1-\kappa)R_{FJ}\right] + (\xi - 1)\left[\kappa R_{AE} + (1-\kappa)R_{AJ}\right]$$

Using backward induction the equilibrium of $\Gamma_f^{e,1}$ can be computed, which turns out to be identical to the mixed-strategy NE in (9.100) [38].

The more realistic situation is when at each transmission time EW forms it's beliefs about Tx's move from its received signal $\vec{y}_{ES}$. This can be determined with the following hypothesis test:

$$\begin{aligned} \mathcal{H}_0 &: \vec{y}_{ES} = \sqrt{g_1}\mathbf{H}_{EW,Tx}\vec{z} + \vec{n}_{ES} \\ \mathcal{H}_1 &: \vec{y}_{ES} = \sqrt{g_1}\mathbf{H}_{EW,Tx}\mathbf{T}z + \sqrt{g_1}\mathbf{H}_{EW,Tx}\mathbf{T}'z' + \vec{n}_{ES} \end{aligned} \tag{9.103}$$

A simple energy detector cannot be employed [41] since both hypotheses have the same received signal power. Instead, EW would need to employ a detector that, for example, exploits the structure of the covariance matrices of the competing hypotheses [42] or additional side information, to generate its beliefs $(\xi, 1-\xi)$. Tx has no means of estimating the beliefs possessed by EW; therefore, Tx should stick to playing its maximin strategy, although the optimality of such a decision, and whether EW should assume Tx is playing pure or mixed strategies, has not been completely resolved [43]. For the dual game $\Gamma_f^{e,2}$ where EW moves first, Rx must carry out a hypothesis test to discern EW's move (in this case, an energy detector would suffice), and then report back to Tx to help it form a belief vector.

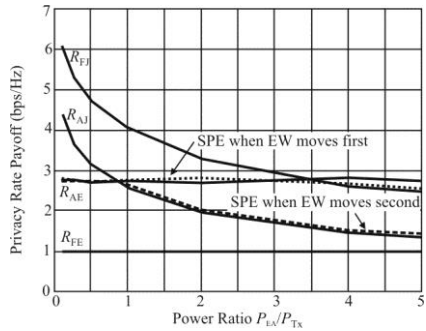


Figure 9.26 Extensive form game with perfect information, $N_{Tx} = N_{Rx} = N_{EW} = 3$, $P_{Tx} = 20$ dBw, $g_1 = 0.8$, $g_2 = 1.1$. The upper dotted line is the SPE when EW moves first, while the bottom dashed line represents the subgame-perfect SPE of the game when EW moves second.

9.7.5 Simulation Results

In this section examples that show the equilibrium privacy rate payoffs for various channel and user configurations are presented [40]. All displayed results are calculated based on an average of 5,000 independent trials per point. The power fraction $(1 - \rho)$ allocated for artificial interference and the optimal number of data streams d are computed using an exhaustive search by Tx. The background noise power was assumed to be the same for both Rx and EW: $\sigma_b^2 = \sigma_e^2 = 1$.

9.7.5.1 Strategic Game Results

Consider the strategic game in Figure 9.25 with $N_{Tx} = N_{EW} = 8$, $N_{Rx} = 6$, so that $d = 4$, and EW's total power is larger than Tx's: $P_{EA} = 4P_{Tx}$. The predicted cross-over point for rates R_{AE} and R_{AJ} is computed from (9.87) and (9.88) to be approximately 14.2 dBw. Prior to the crossover, a pure strategy NE in R_{AE} is the game outcome since the rate ordering is given by (9.96), whereas after the crossover it is optimal for both players to play mixed strategies according to (9.100). In this case, randomizing strategies clearly leads to better payoffs for the players as EW's jamming power increases, compared to adopting a pure strategy.

If Tx transmits with full power (R_{FJ} and R_{FE}), the highest potential payoff is possible (R_{FJ}), but it is also possible that the payoff is zero (R_{FE}). Below the R_{AJ}-R_{AE} crossover point, the payoffs are similar while above the crossover point R_{AE} produces significantly better (for Tx) results. Therefore it appears that the best move for Tx is to optimally send artificial noise to EW. However, if EW jams Rx (R_{FJ}, R_{AJ}), the payoff (for EW) is reduced. If EW eavesdrops on Tx (R_{FE}, R_{AE}), the possibility exists that the payoff is zero so it appears that to eavesdrop is the best

strategy. If those two strategies are followed, then payoff R_{AE} ensues, and Tx is the winner.

9.7.5.2 Extensive Game Results

The subgame-perfect outcomes of the two extensive form games $\Gamma^{e,1}$ and $\Gamma^{e,2}$ over a range of transmit power ratios P_{EA}/P_{Tx} are shown in Figure 9.26. The upper dotted line is the SPE when EW moves first, while the bottom dashed line represents the subgame-perfect SPE of the game when EW moves second, as defined in Property 9.9(a) and Property 9.9(b). Observe that prior to the crossover point of R_{AE} and R_{AJ} (about at $P_{EA}/P_{Tx} = 0.85$), both equilibria are equal as determined by Property 9.9(a), since a pure-strategy NE results. We see that it is always beneficial for EW to move second especially as EW's jamming power increases.

If Tx decides to transmit to Rx with full power (R_{FE}, R_{FJ}), it stands to achieve the best payoff (R_{FJ}) or the worst (R_{FE}). If Tx decides to allocate the optimum amount of resources to jamming EW (R_{AJ}, R_{AE}), then for low EW powers the payoff is significantly better whether EW decides to eavesdrop (R_{AE}) or jam (R_{AJ}). For higher EW powers, the return falls toward R_{FE}, so it appears that Tx's best move is to jam EW with optimum artificial noise (R_{Ai}), and if EW decides to eavesdrop (R_{AE}) the Tx payoff remains relatively good.

9.7.6 Summary

We discussed the interactions between a multiantenna communication link between Tx and Rx, and a dual-mode intercept/jammer as a zero-sum game with the MIMO privacy rate as the payoff function. We examined the conditions under which NE existed in pure and mixed strategies for the strategic version of the game. We also investigated subgame-perfect and sequential equilibria in the extensive forms of the game with and without perfect information. Our numerical results demonstrated that a change in a single parameter set while others remain constant can shift the equilibrium from a pure to a mixed NE outcome or vice versa.

9.8 Independent ES and EA System Performance

Some realistic scenarios place the EA and ES nodes at different places. Such would be the case, for example, when it is desired to minimize operator exposure to possible incoming fires that a jammer might attract due to its high power signal emanating from the antennas. Antiradiation missiles can be used to home in on the

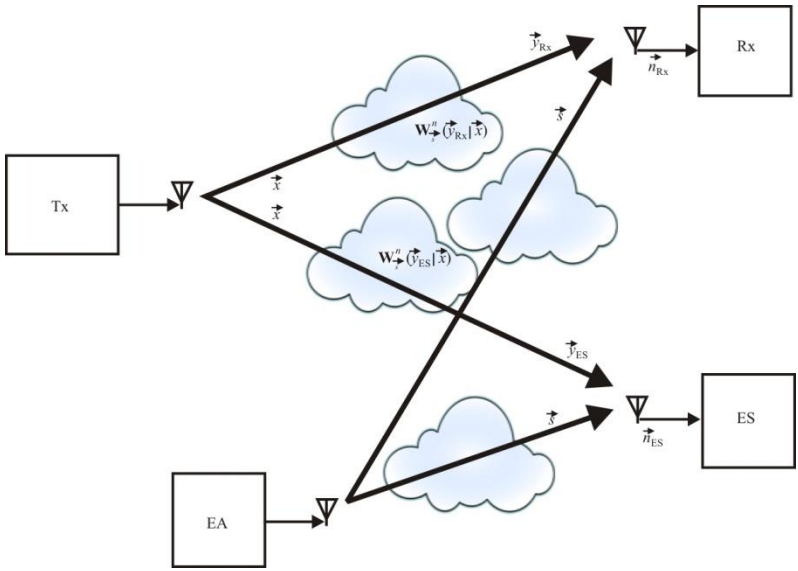


Figure 9.27 Channel model with independent ES and EA.

signal. Placing the ES at a different location minimizes such a threat to the ES operators.

MolavianJazi [44] considered the case of the presence of an eavesdropper as well as a jammer and the two are separate entities. Therefore, we have a wiretap channel between Tx and ES, as well as an AVC between Tx and Rx. The scenario is depicted in Figure 9.27. We will present the principal results of this analysis in this section.

9.8.1 Arbitrarily Varying Wiretap Channels

A discrete memoryless *arbitrarily varying wiretap channel* (AVWTC) combines the concept of the wiretap channel with that of an AVC. The wiretap channel models the intercept process as explained in Section 9.2, while the AVC models the jamming process introduced in Section 9.3. An AVWTC is illustrated in Figure 9.28, and is characterized by a finite input alphabet $\mathcal{X}$, two finite output alphabets $\mathcal{Y}_{\text{Rx}}$ and $\mathcal{Y}_{\text{ES}}$, an arbitrary state space $\mathcal{S}$, and a family of transition probabilities from X to $Y_{\text{Rx}} \times Y_{\text{ES}}$ indexed by $\mathcal{S}$ represented as (3.1)

$$\mathcal{W} = \left\{ \mathbf{W}_s(y_{\text{Rx}}, y_{\text{ES}} \,|\, x) \triangleq \mathbf{W}(y_{\text{Rx}}, y_{\text{ES}} \,|\, x; s) : s \in \mathcal{S} \right\} \tag{9.104}$$

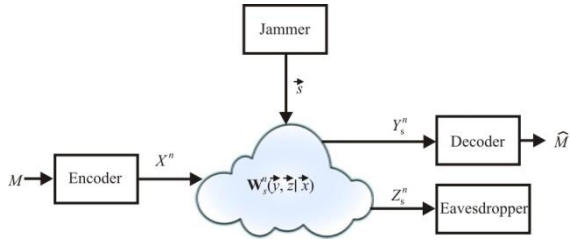


Figure 9.28 Arbitrarily varying wiretap channel model with an eavesdropper.

The *n*th extension of the channel law for input $\vec{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$ and outputs $\vec{y}_{\text{Rx}} = (y_{\text{Rx},1}, \ldots, y_{\text{Rx},n}) \in \mathcal{Y}_{\text{Rx}}^n$ and $\vec{y}_{\text{ES}} = (y_{\text{ES},1}, \ldots, y_{\text{ES},n}) \in \mathcal{Y}_{\text{ES}}^n$ under the state sequence $\vec{s} = (s_1, \ldots, s_n) \in \mathcal{S}^n$ is (3.2)

$$\mathbf{W}_{\vec{s}}^n(\vec{y}_{\text{Rx}}, \vec{y}_{\text{ES}} | \vec{x}) \triangleq \prod_{i=1}^n \mathbf{W}_{s_i}(y_{\text{Rx},i}, y_{\text{ES},i} | x_i) = \prod_{i=1}^n \mathbf{W}_{s_i}(y_{\text{Rx},i}, y_{\text{ES},i} | x_i ; s_i) \quad (9.105)$$

Note that the output is related to the input without memory, whereas there are no caveats on the channel state, that is, the channel state has no presumed a priori distribution and possibly has memory.

Note that in our notation for the nodes, $\text{EW} \neq \text{ES} + \text{EA}$ in this case. ES and EA represent two separate nodes, ES corresponds to the intercept while EA refers to the jammer.

The node observing the output $\vec{y}_{\text{Rx}} \in \mathcal{Y}^n$ is Rx and the marginal channel $\mathbf{W}_{\vec{s}}^n(\vec{y}_{\text{Rx}} | \vec{x}) = \sum_{y_{\text{ES}} \in \mathcal{Y}_{\text{ES}}^n} \mathbf{W}_{\vec{s}}^n(\vec{y}_{\text{Rx}}, \vec{y}_{\text{ES}} | \vec{x})$ is referred to as the *main channel*, and the node observing the output $\vec{y}_{\text{ES}} \in \mathcal{Y}^n$ is ES and the marginal channel $\mathbf{W}_{\vec{s}}^n(\vec{y}_{\text{ES}} | \vec{x}) = \sum_{y_{\text{Rx}} \in \mathcal{Y}_{\text{Rx}}^n} \mathbf{W}_{\vec{s}}^n(\vec{y}_{\text{Rx}}, \vec{y}_{\text{ES}} | \vec{x})$ is referred to as the *intercept channel*. The main and intercept channels are AVCs. The state sequence $\vec{s}$ is called an *attack*, which is selected by the jammer. It is assumed that the jammer and intercept are separate nodes, operating without coordination with one another.[17] The jammer selects its attack $\vec{s}$ in the absence of any knowledge about the message *m* or ES's

[17] While this is not necessarily always true in reality, it is a significant real-world problem. ES systems have the desire to intercept the channel and EA systems want to preclude any communications on the channel. These two goals are directly at odds with one another. Many military forces in developed countries have EA and ES functionality in separate organizations, and are therefore managed separately.

observation $\vec{y}_{\mathrm{ES}}$. The transmitter and the two receivers are cognizant of the state space $\mathcal{S}$, but not the actual state sequence $\vec{s}$.

In some cases, EA may be able to select the states of the main and intercept channels independently. This is formalized in the following definition [45].

> **Definition 9.10:** An AVWTC has *independent states* if the state space can be decomposed as $\mathcal{S} = \mathcal{S}_{\mathrm{y_{Rx}}} \times \mathcal{S}_{\mathrm{z_{ES}}}$ and the channel state as the pair $s = (s_{\mathrm{y_{Rx}}}, s_{\mathrm{y_{ES}}})$ such that the state $s_{\mathrm{y_{Rx}}}$ of the main AVC and the state $s_{\mathrm{y_{ES}}}$ of the intercept AVC are independently selected from the corresponding state spaces $\mathcal{S}_{\mathrm{y_{Rx}}}$ and $\mathcal{S}_{\mathrm{y_{ES}}}$, respectively.

As explained in Section 9.3, just as for AVCs, the same two notions that characterize jamming attacks and their effect on the individual channels apply to AVWTCs. The effects of ES must be included, however. Averaging over states and channel laws, leading to the set of averaged states and the convex closure of an AVWTC is the same concern as for AVCs, discussed in Section 9.3.

The convex closure for an AVWTC is defined similarly to that for an AVC:

> **Definition 9.11:** The *convex closure*[18] $\overline{\mathcal{W}}$ of the AVWTC $\mathcal{W}$ is the closure of the set of all averaged wiretap channels,

$$\overline{\mathcal{W}} = \mathrm{cl}\left(\left\{ \begin{array}{l} \mathbf{W}_{\bar{s}}(\vec{y}_{\mathrm{Rx}}, \vec{y}_{\mathrm{ES}} | \vec{x}) = \sum_k p(s_k) \mathbf{W}_{s_k}(\vec{y}_{\mathrm{Rx}}, \vec{y}_{\mathrm{ES}} | \vec{x}) : \\ \qquad\qquad \bar{s} = \sum_k s_k p(s_k) \in \bar{\mathcal{S}} \end{array} \right\} \right) \tag{9.106}$$

For convenience, we will refer to the AVWTC $\mathcal{W}$ also as the AVTWC $\mathcal{S}$ and its convex closure $\overline{\mathcal{W}}$ also as $\bar{\mathcal{S}}$, provided there is no ambiguity, since an AVWTC and its convex closure are essentially characterized by the state space $\mathcal{S}$ and averaged state space $\bar{\mathcal{S}}$.

Similar to the notion of the symmetrizable (Definition 9.7) [16] mentioned in Section 9.3, the AVWTC has the same concern. It is the potential for EA to forge the actions of the transmitter over the main channel. The same definition applies here. We recall from Section 9.3 that the code capacity of an AVC is zero if and only if it is symmetrizable.

---

[18] See footnote 7.

9.8.2 Degraded Channels

The relative quality of the main and intercept channels under different states is an important concept for AVWTCs. For this purpose, two notions of "degradedness" are defined:

Definition 9.12: The AVWTC $\mathcal{S}$ is *degraded* if the Markov chain $X \to Y_{\text{Rx},s} \to Y_{\text{ES},s}$ holds for every state $s \in \mathcal{S}$.

The convex closure of a degraded broadcast AVC can fail to be a degraded broadcast AVC as the following example illustrates.

Example 9.7: ([45]) Let $\mathcal{X} = \mathcal{Y}_{\text{Rx}} = \mathcal{Y}_{\text{ES}} = \mathcal{S} = \{0,1\}$, and define the broadcast AVC with the $\mathcal{X}$-$\mathcal{Y}_{\text{Rx}}$ channel being the XOR AVC

$$\mathcal{V} = \left\{ \mathbf{V}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{V}_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\} \tag{9.107}$$

and the $\mathcal{X}$-$\mathcal{Y}_{\text{ES}}$ channel being $\mathcal{W} = \{\mathbf{V}_1^2, \mathbf{V}_2^2\}$, that is, a cascade of two $\mathcal{X}$-$\mathcal{Y}$ channels. Thus,

$$\mathcal{W} = \{\mathbf{V}_1, \mathbf{V}_1\} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{9.108}$$

Although the $\mathcal{X}$-$\mathcal{Y}_{\text{ES}}$ channel is degraded with respect to the $\mathcal{X}$-$\mathcal{Y}_{\text{Rx}}$ channel for each state $s \in \{0,1\}$, this property does not hold for the averaged states satisfying $\overline{s} \in \{0,1\}$. We also observe that the capacity of the $\mathcal{X}$-$\mathcal{Y}_{\text{Rx}}$ AVC is zero, whereas the capacity of the $\mathcal{X}$-$\mathcal{Y}_{\text{ES}}$ AVC is equal to 1.

Motivated by this issue, the following stronger definition of degradedness is introduced:

Definition 9.13: The AVWTC $\mathcal{S}$ is *strongly degraded* if the Markov chain $X \to Y_{\text{Rx},\overline{s}} \to Y_{\text{ES},\overline{s}}$ is satisfied for every averaged state $\overline{s} \in \overline{\mathcal{S}}$.

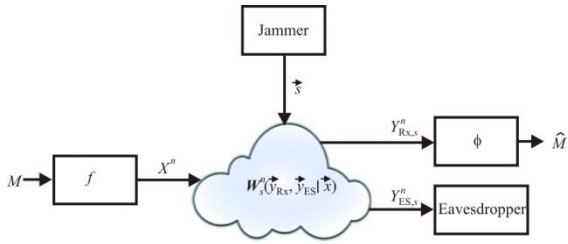


Figure 9.29 AVWTC model with wiretap code.

9.8.3 Coding Scheme and Performance Measures

As for AVCs, we consider traditional (deterministic) wiretap coding schemes. Wiretap codes, as illustrated in Figure 9.29, use a particular encoder-decoder for reliable and secure communication over the AVWTC. The precise definition of the coding scheme is given in the following:

> **Definition 9.14:** An $(n, \mathcal{M})$ *wiretap code* for an AVWTC consists of a message set $\mathcal{M} = \{1, \ldots, \mathcal{M}\}$, a stochastic encoder $f : \mathcal{M} \to \mathcal{X}^n$, and a deterministic decoder $\phi : \mathcal{Y}_{Rx}^n \to \mathcal{M}$.

We now describe the performance MOEs for this coding scheme. For this purpose, assume that the message m is selected uniformly at random from the message set $\mathcal{M}$, encoded for transmission as X^n, and received at Rx and ES terminals as $Y_{Rx,\vec{s}}^n$ and $Y_{ES,\vec{s}}^n$, respectively, where $\vec{s}$ is the actual state sequence.

The average error probability of the wiretap code (f, ϕ) under the state sequence $\vec{s}$ is

$$\bar{e}(\mathbf{W}_{\vec{s}}^n, f, \phi) \triangleq \frac{1}{\mathcal{M}} \sum_{m \in \mathcal{M}} \sum_{y_{ES} \in Y_{ES}^n} \mathbf{W}_{\vec{s}}^n \left\{ \left[\phi^{-1}(m) \right]^c, y_{ES} \,|\, f(m) \right\} \tag{9.109}$$

and the leakage rate of the wiretap code (f, ϕ) under the state sequence $\vec{s}$ is

$$L(\mathbf{W}_{\vec{s}}^n, f, \phi) \triangleq \frac{1}{n} \mathbb{I}[M; Y_{ES,\vec{s}}^n \,|(f, \phi)] \tag{9.110}$$

where the conditioning on (f, ϕ) accounts for the fact that the wiretap code (f, ϕ) is known to ES.[19]

9.8.4 Privacy Capacity

We can now delineate the privacy capacity performance limits of an AVWTC.

> **Definition 9.15:** A privacy rate R_{p} is *achievable* for the AVWTC $\mathcal{S}$ if for every $\varepsilon > 0$ there exists an $(n, 2^{nR_{\mathrm{p}}})$ wiretap code such that for every state sequence $\vec{s} \in \mathcal{S}^n$
>
> $$\bar{e}(\mathbf{W}_s^n, f, \phi) \le \varepsilon \tag{9.111}$$
> $$L(\mathbf{W}_s^n, f, \phi) \le \varepsilon \tag{9.112}$$
>
> The privacy capacity C_{p} of the AVWTC $\mathcal{S}$ is the supremum of all achievable privacy rates.

Taking into account the fact that neither Tx nor Rx are cognizant of the actual state sequence $\vec{s} \in \mathcal{S}^n$, Definition 9.15 ensures reliable and secure communication under any state sequence,

9.8.5 Performance of AVWTCs

This section summarizes the performance of AVWTCs in the presence of a jammer and an unrelated interceptor.

9.8.5.1 Lower Bound on the Privacy Capacity

Here we present a lower bound on the privacy capacity of AVWTCs, established for the following class of AVTWCs.

> **Definition 9.16:** An AVWTC $\mathcal{S}$ has a *worst channel* for ES if for any state $s \in \mathcal{S}$, the intercept channel is degraded with respect to the channel under some fixed state $s^* \in \bar{\mathcal{S}}$, that is,
>
> $$\exists s^* \in \bar{S} \quad \forall s \in \mathcal{S} \text{ s.t. } X \to Y_{\mathrm{ES},s^*} \to Y_{\mathrm{ES},s} \tag{9.113}$$

[19] It is worth noting that keeping the decoding function of Rx (ϕ) as an argument in our notations (3.10) and (3.14) for the leakage rate is only for the sake of notational symmetry and can actually be discarded since privacy only depends on the structure of the encoding function.

Four points are worth noting about this definition. First, this definition only concerns degradedness within the family of the intercept channels, whereas Definitions 9.12 and 9.13 involve degradation between the main and intercept channels.

Second, we can surmise some realistic situations satisfying this definition. A simple example is the class of AVWTCs in which the intercept channel is a fixed channel independent of the jammer state. This class models situations in which ES is robust to EA.

Third, (9.113) implies the most capable property of the state s^* in the sense that for all distributions $P_X(x)$

$$\mathbb{I}(X;Y_{\mathrm{ES},s^*}) = \sup_{s\in\mathcal{S}} \mathbb{I}(X;Y_{\mathrm{ES},s}) = \sup_{s\in\mathcal{S}} \mathbb{I}(X;Y_{\mathrm{ES},\bar{s}}) \tag{9.114}$$

This follows from the data processing inequality, the convexity of mutual information in the channel law, and the convexity of the set $\bar{\mathcal{S}}$. However, note that condition (9.114) does not necessarily imply the condition (9.113) of Definition 9.16.

Fourth, implied by the third point, the worst state $s^* \in \bar{\mathcal{S}}$ of ES is not actually an averaged state since if $|\mathcal{S}| < \infty$, it falls inside the state space $\mathcal{S}$ itself, and if $|\mathcal{S}| = \infty$, it falls in the boundary of the state space $\mathcal{S}$.

The following property states a lower bound result for the class of AVWTCs of Definition 9.16.

> **Proposition 9.1(a):** For the discrete memoryless AVTWC $\mathcal{S}$ that has a worst channel for ES, all randomized-code privacy rates satisfying
>
> $$R_{\mathrm{p}} < \max_{P_X(x)} \left[\min_{\bar{s}\in\bar{\mathcal{S}}} \mathbb{I}(X;Y_{\mathrm{Rx},\bar{s}}) - \max_{\bar{s}\in\bar{\mathcal{S}}} \mathbb{I}(X;Y_{\mathrm{ES},\bar{s}}) \right] \tag{9.115}$$
>
> are achievable.

Proposition 9.1(a) suggests that privacy rates are constrained by the worst averaged state of the main channel (from the point of view of Tx) and the best averaged state of the intercept channel (again from the point of view of Tx). For an AVWTC, the varying state of the channel is captured by the presence of an averaged state in the result. This worst-case scenario highlights the deleterious effect of arbitrary jamming on the achievable privacy. In fact, the randomized

wiretap coding must be carried out for the weakest case for Rx. Of course, the privacy rate introduced in Proposition 9.1(a) may be zero.

The following corollary provides an extension of Proposition 9.1(a).

> **Proposition 9.1(b):** For the discrete memoryless AVTWC $\mathcal{S}$ that has a worst channel for ES, all privacy rates satisfying
>
> $$R_{\mathrm{p}} < \max_{P(u,x)} \left[\min_{\bar{s} \in \bar{\mathcal{S}}} I(X; Y_{\mathrm{Rx},\bar{s}}) - \max_{\bar{s} \in \bar{\mathcal{S}}} I(U; Y_{\mathrm{ES},\bar{s}}) \right] \tag{9.116}$$
>
> are achievable, provided $U \to X \to Y_{\mathrm{Rx},\bar{s}} Y_{\mathrm{ES},\bar{s}}$ forms a Markov chain for every $\bar{s} \in \bar{\mathcal{S}}$.

9.8.5.2 Upper Bounds on the Randomized-Code Privacy Capacity

We present two upper bounds on the privacy capacity of an AVWTC. The first upper bound in the following proposition is the convex degraded-same marginal upper bound.

> **Proposition 9.2:** The privacy capacity of the discrete memoryless AVWTC $\mathcal{S}$ satisfies
>
> $$C_{\mathrm{p}} \le \max_{P_X(x)} \min_{\bar{s} \in \bar{\mathcal{S}}} I(X; Y_{\mathrm{Rx},\bar{s}} \,|\, Y_{\mathrm{ES},\bar{s}}) \tag{9.117}$$

The upper bound in the following proposition is the convex-compound upper bound.

> **Proposition 9.3:** The privacy capacity of the discrete memoryless AVWTC $\mathcal{S}$ is bounded by
>
> $$C_{\mathrm{p}} \le \min_{\bar{s} \in \bar{\mathcal{S}}} \max_{U \to X \to Y_{\mathrm{Rx},\bar{s}} Y_{\mathrm{ES},\bar{s}}} \left[I(U; Y_{\mathrm{Rx},\bar{s}}) - I(U; Y_{\mathrm{ES},\bar{s}}) \right] \tag{9.118}$$

Proposition 9.3 states that the privacy capacity of an AVWTC does not exceed the minimum of the privacy capacities of the family of discrete memoryless wiretap channels $\{\mathbf{W}_{\bar{s}}(y_{\mathrm{Rx}}, y_{\mathrm{ES}} \,|\, x) : \bar{s} \in \bar{\mathcal{S}}\}$.

These two upper bounds can be used to determine conditions under which the privacy capacity is zero, that is, situations in which secure communication over the AVWTC is impossible because ES can always intercept the messages. The convex degraded-same-marginal upper bound (9.117) implies that, if the AVWTC is

reversely degraded; that is, the Markov chain $X \to Y_{\mathrm{ES},\bar{s}} \to Y_{\mathrm{Rx},\bar{s}}$ holds for even a single averaged state $\bar{s} \in \bar{\mathcal{S}}$, then the privacy capacity is zero. Extending this result, the convex compound upper bound (9.118) asserts that the privacy capacity is zero if the intercept channel is less noisy [46] than the main channel for even a single averaged state $\bar{s} \in \bar{\mathcal{S}}$; that is, if there exists some $\bar{s} \in \bar{\mathcal{S}}$, for which $\mathbb{I}(U;Z_{\bar{s}}) \geq \mathbb{I}(U;Y_{\mathrm{Rx},\bar{s}})$ for every random variable U satisfying the Markov chain $U \to X \to Y_{\mathrm{Rx},\bar{s}} Y_{\mathrm{ES},\bar{s}}$. The latter condition is more general than the former condition since the less noisy constraint is a strictly weaker condition than the degree of degradation.

9.8.5.3 Strongly Degraded AVWTC with Independent States

There is a special class of AVWTC's for which the lower and upper bounds match, so that we obtain the actual privacy capacity. This class combines the strongly degraded AVWTC model given in Definition 9.11 with the AVWTC with independent states model given in Definition 9.8.

Definition 9.17: An AVWTC $\mathcal{S}$ is *strongly degraded with independent states* if (1) the state space can be decomposed as $\mathcal{S} = \mathcal{S}_{\mathrm{y_{Rx}}} \times \mathcal{S}_{\mathrm{y_{ES}}}$ where the state of the main channel $s_{\mathrm{y_{Rx}}} \in \mathcal{S}_{\mathrm{y_{Rx}}}$ and that of the intercept AVC $s_{\mathrm{y_{ES}}} \in \mathcal{S}_{\mathrm{y_{ES}}}$ are selected independently and (2) the Markov chain

$$X \to Y_{\mathrm{Rx},\bar{s}_{\mathrm{y}}} \to Y_{\mathrm{ES},\bar{s}_{\mathrm{z}}} \tag{9.119}$$

is satisfied for all averaged states

$$\bar{s}_{\mathrm{y_{Rx}}} \in \bar{\mathcal{S}}_{\mathrm{y_{Rx}}} \text{ and } \bar{s}_{\mathrm{y_{ES}}} \in \bar{\mathcal{S}}_{\mathrm{y_{ES}}}$$

The following property identifies the privacy capacity of this special class of AVWTCs.

Proposition 9.4: The privacy capacity of a strongly degraded discrete memoryless AVWTC with independent states is

$$C_{\mathrm{p}} = \max_{P_X(x)} \left[\min_{\bar{s}_{\mathrm{y}} \in \bar{\mathcal{S}}_{\mathrm{y}}} \mathbb{I}(X;Y_{\mathrm{Rx},\bar{s}_{\mathrm{y}}}) - \max_{\bar{s}_{\mathrm{z}} \in \bar{\mathcal{S}}_{\mathrm{z}}} \mathbb{I}(X;Y_{\mathrm{ES},\bar{s}_{\mathrm{z}}}) \right] \tag{9.120}$$

provided the AVWTC has a worst channel for ES.

9.8.5.4 Wiretap Codes

The following provides a characterization of the privacy capacity of a discrete memoryless AVWTC based on the propositions above.

> **Property 9.10(a):** The privacy capacity of the discrete memoryless AVWTC $\mathcal{S}$ is limited by the privacy capacities delineated in Propositions 9.1–9.4 if the main channel is nonsymmetrizable, and is zero otherwise.

Note that Property 9.10(a) establishes the privacy capacity of an AVTWC based solely on the symmetrizability of the AVC channel. This means that the wiretap channel plays no role in establishing this MOE for a deterministic channel.

The significance of Property 9.10(a) is that it guarantees the existence of wiretap codes for secure communication over the AVWTC without any common randomness between the legitimate parties. Therefore, a single encoder-decoder known to all parties, including the ES and EA, is shown to be capable of approaching the privacy capacity of the AVWTC provided the nonsymmetrizability condition is satisfied.

An immediate side result of Property 9.10(a) is the following property.

> **Property 9.10(b):** If the main channel of the discrete memoryless AVWTC S is nonsymmetrizable, then its privacy capacity satisfies
>
> $$C_{\mathrm{p}} \geq \max_{P_X(x)} \left[\min_{\bar{s} \in \bar{\mathcal{S}}} I(X; Y_{\mathrm{Rx},\bar{s}}) - \max_{\bar{s} \in \bar{\mathcal{S}}} I(X; Y_{\mathrm{ES},\bar{s}}) \right] \tag{9.121}$$
>
> provided the AVWTC has a worst channel for intercept.

9.8.6 Examples

The following notations are used in these examples. A binary symmetric channel with crossover probability p is denoted by BSC(p). For laconicalness, the operation $*$ is defined by $a * b \triangleq a(1-b) + b(1-a)$. The binary entropy function is indicated by $H_{\mathrm{b}}(p) = -p \log_2 p - (1-p) \log_2 (1-p)$.

9.8.6.1 Example for $C_{\mathrm{p}} = 0$

In this section, we assume that the deterministic privacy capacity is zero. It illustrates that even a single bad averaged state precludes secure communication over the AVWTC. We also show that the leakage of information to ES is a fact

beyond the ability of ES for correct decoding of the messages because the intercepts have zero channel capacity.

Example 9.8: Suppose the AVWTC has $\mathcal{X} = \mathcal{Y}_{\text{Rx}} = \mathcal{Y}_{\text{ES}} = \mathcal{S} = \{0,1\}$, while the main channel is given by

$$\{BSC(0), BSC(1/3)\}$$

and the intercept channel by

$$\{BSC(0), BSC(1)\}$$

We see that the main channel is nonsymmetrizable with $C_{\text{Rx}} = 1 - H_{\text{b}}(1/3)$, since any symmetrizing matrix

$$\mathbf{U}(s|x) = \begin{bmatrix} a & 1-a \\ b & 1-b \end{bmatrix}$$

with $0 \le a, b \le 1$ for Definition 9.5 requires $a + b = -1$. Further we see that the intercept channel is symmetrizable with $C_{\text{ES}} = 0$. Since both receivers observe the input at the state s = 0, the privacy capacity is zero, that is, $C_{\text{p}} = 0$.

9.8.6.2 Example for $C_{\text{p}} > 0$

In this example secure communication over the AVWTC is possible by using wiretap codes using the special class of strongly degraded AVWTCs with independent states. Therefore, we see that if Rx possesses enough advantage over ES in all averaged states, a positive privacy capacity can be established.

Example 9.9: Suppose the AVWTC has $\mathcal{X} = \mathcal{Y}_{\text{Rx}} = \mathcal{Y}_{\text{ES}} = \mathcal{S}_{\text{y}_{\text{Rx}}} = \mathcal{S}_{\text{y}_{\text{ES}}} = \{0,1\}$, the main AVC is

$$\{BSC(1/6), BSC(1/5)\}$$

and the intercept AVC is

$$\{BSC(1/4), BSC(1/3)\}$$

where the states are selected independently for each AVC. We see that the code capacity of the main AVC is $C_{\mathrm{Rx}} = 1 - H_{\mathrm{b}}(1/5)$, and that of the intercept AVC is $C_{\mathrm{ES}} = 1 - H_{\mathrm{b}}(1/3)$.

The AVWTC is strongly degraded since any averaged main channel is a $BSC(p_{\mathrm{Rx}})$ with $1/6 \le p_{\mathrm{Rx}} \le 1/5$, and any averaged intercept channel is a $BSC(p_{\mathrm{ES}})$ with $1/4 \le p_{\mathrm{ES}} \le 1/3$. In addition, by assumption it has independent states. ES has a worst channel since the degradedness of $BSC(1/3)$ with respect to $BSC(1/4)$ implies that the intercept channel in the state $s_{\mathrm{z}} = 1$ is degraded with respect to that in the state $s_{\mathrm{z}}^* = 0$.

Finding the privacy capacity C_{p} of this AVWTC requires determining the symmetrizability status of the main channel. Since plugging the symmetrizing matrix

$$\mathbf{U}(s|x) = \begin{bmatrix} a & 1-a \\ b & 1-b \end{bmatrix} \qquad \text{with} \qquad 0 \le a, b \le 1$$

into condition (3.6) of Definition 9.5 leads to the contradiction $a + b = -18$, we conclude that the main AVC is nonsymmetrizable. Hence, this AVWTC satisfies the conditions of Proposition 9.4, and its privacy capacity is given by (9.120)

$$C_{\mathrm{p}} = \max_{q \in [0,1]} \left\{ \begin{array}{l} \min\limits_{p \in [0,1]} \left[H_{\mathrm{b}}\left(q * \dfrac{p+5}{30} \right) - H_{\mathrm{b}}\left(\dfrac{p+5}{30} \right) \right] \\ \qquad - \max\limits_{p \in [0,1]} \left[H_{\mathrm{b}}\left(q * \dfrac{p+3}{12} \right) - H_{\mathrm{b}}\left(\dfrac{p+3}{12} \right) \right] \end{array} \right\}$$

Since the closure of $\mathcal{S}$, $\bar{\mathcal{S}} = \{0, 1/2, 1\}$, $\min\limits_{p \in [0,1]} \{p : 0, 1/2, 1\}, = 0$ and $\max\limits_{p \in [0,1]} \{p : 0, 1/2, 1\} = 1$, and we have

$$C_{\mathrm{p}} = \max_{q \in [0,1]} \left\{ \begin{array}{l} \left[H_{\mathrm{b}}\left(q * \dfrac{p+5}{30} \right) - H_{\mathrm{b}}\left(\dfrac{p+5}{30} \right) \right]_{p=0} \\ \qquad - \left[H_{\mathrm{b}}\left(q * \dfrac{p+3}{12} \right) - H_{\mathrm{b}}\left(\dfrac{p+3}{12} \right) \right]_{p=1} \end{array} \right\}$$

$$= \max_{q\in[0,1]} \left\{ \left[H_{\mathrm{b}}\left(q * \frac{1}{6}\right) - H_{\mathrm{b}}\left(q * \frac{1}{3}\right) \right] \right\} + H_{\mathrm{b}}\left(\frac{1}{3}\right) - H_{\mathrm{b}}\left(\frac{1}{6}\right)$$

Also $\max_{P_X(x)} = 1/2$ since the channel is binary, so

$$\begin{aligned} C_{\mathrm{p}} &= \left[H_{\mathrm{b}}\left(q * \frac{1}{6}\right) - H_{\mathrm{b}}\left(q * \frac{1}{3}\right) \right]_{q=1/2} + H_{\mathrm{b}}\left(\frac{1}{3}\right) - H_{\mathrm{b}}\left(\frac{1}{6}\right) \\ &= H_{\mathrm{b}}\left(\frac{1}{2}\right) - H_{\mathrm{b}}\left(\frac{2}{3}\right) \\ &\quad + H_{\mathrm{b}}\left(\frac{1}{3}\right) - H_{\mathrm{b}}\left(\frac{1}{6}\right) = 0.35 \qquad \text{bps/Hz} \end{aligned} \tag{9.122}$$

So the privacy rate for this channel and combination of nodes is $C_{\mathrm{p}} = 0.35$ so Tx can send information to Rx without intercept by ES (the intercepted signals look like random noise) and in the presence of jamming by EA as long as the rate is less than 0.35 bps/Hz.

9.8.7 Summary

In this section we presented jamming and intercept effects when there are both jammer and intercept nodes, but they are noncooperating. We saw that under appropriate conditions either the EA or ES node, or both, could reduce the channel to the point where there is no code that can be used to communicate on the channel.

9.9 Concluding Remarks

In this chapter we examined a few measures for evaluating the performance of EW systems, both in the ES and EA modes, and in both modes simultaneously. The formal basis for these discussions was information theory, and, in particular, the broadcast channel in the wiretap configuration.

We discussed five communication EW scenarios: (1) ES performance over a wiretap channel; (2) EA performance over AWGN channels with BBN and PBN thermal noise jamming; (3) MIMO channels where EW could dynamically allocate its antennas to either ES or EA; (4) EW performance over MIMO channels where Tx had the choice of allocating some of its spatial resources (antennas) to sending artificial noise to EA to minimize intercept, while EW had

the choice of conducting intercept or to jam Rx; and (5) EW performance with physically separated ES and EA assets—EA used a arbitrarily varying channel while ES used a wiretap channel.

While we considered only a few MOEs, there are many more, but the ones presented here are indicative of how well EW systems can be expected to perform in realistic environments. The scenarios presented can be viewed as "typical" use of tactical EW systems countering target communication networks. The configurations of target networks and EW systems are not limited to any given type of formulation; the results apply to ground-to-ground, air-to-air, and air-to-ground scenarios.

We showed that EW systems can be used to effectively preclude communications in a channel, and in those cases where such prevention is not possible, we showed how much degradation can be expected by the use of such systems.

References

[1] Wyner, A. D., "The Wiretap Channel," *The Bell Systems Technical Journal*, Vol. 54, No. 8, 1975, pp. 355–387.

[2] Csiszar, I., and J. Korner, "Broadcast Channels with Confidential Messages," *IEEE Transactions on Information Theory*, Vol. 24, No. 3, 1979, pp. 339–348.

[3] Leung-Yan-Cheong, S., and M. Hellman, "The Gaussian Wiretap Channel," *IEEE Transactions on Information Theory*, Vol. 24, No. 4, 1979, pp. 451–456.

[4] Lai, L., and H. El-Gamal, "The Relay Intercept Channel: Cooperation for Privacy," *IEEE Transactions on Information Theory*, Vol. 54, No. 9, 2009, pp. 4005–4019.

[5] Bloch, M., and J. N. Laneman, "Information Spectrum Methods for Information Theoretic Security," *Proceedings of Information Theory and Applications Workshop* (*ITA'09*), San Diego, CA, 2009, pp. 23–29.

[6] Liang, Y., H. V. Poor, and S. Shamai, "Secure Communication over Fading Channels," *IEEE Transactions on Information Theory*, Vol. 54, No. 6, 2009, pp. 2470–2492.

[7] Khisti, A., and G. Wornell, "Secure Transmission with Multiple Antennas II: The MIMOME Wiretap Channel," *IEEE Transactions on Information Theory*, Vol. 56, No. 11, November 2010, pp. 5515–5532.

[8] Liu, T., and S. Shamai, "A Note on the Privacy Capacity of the Multiple-Antenna Wiretap Channel," *IEEE Transactions on Information Theory*, Vol. 55, No. 6, 2009, pp. 2547–2553.

[9] Bloch, M., *Lecture Notes of the Course: Advanced Topics in Information Theory: Information-Theoretic Security*, Department of Electrical Engineering, University of Notre Dame, 2009.

[10] Basar, T., "The Gaussian Test Channel with an Intelligent Jammer," *IEEE Transactions on Information Theory*, Vol. 29, No. 1, 1983, pp. 152–157.

[11] Weiss, M., and S. Schwartz, "On Optimal Minimax Jamming and Detection of Radar Signals," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 21, No. 3, 1985, pp. 385–393.

[12] McEliece, R. J., "Communication in the Presence of Jamming: An Information Theory Approach," G. Longo, (ed.), *Secure Digital Communications*, CISM Courses and Lectures, New York: Springer-Verlag, 1983.

[13] Medard, M., "Capacity of Correlated Jamming Channels," *Proceedings 35th Annual Allerton Conference on Communications, Control, and Computing*, Monticello, IL, 1997, pp. 1043–1052.

[14] Blackwell, D., L. Breiman, and A. J. Thomasian, "The Capacities of Certain Channel Classes Under Random Coding," *The Annals of Mathematical Statistics*, Vol. 31, No. 3, 1960, pp. 558–567.

[15] Ericson, T., "Exponential Error Bounds for Random Codes in the Arbitrarily Varying Channel," *IEEE Transactions on Information Theory*, Vol. 31, No. 1, 1985, pp. 42–49.

[16] Csiszar, I., and P. Narayan, "The Capacity of the Arbitrarily Varying Channel Revisited: Positivity, Constraints," *IEEE Transactions on Information Theory*, Vol. 34, No. 2, 1989, pp. 181–193.

[17] Poisel, R. A., *Modern Communications Jamming Principles and Techniques*, 2nd ed., Norwood, MA: Artech House, 2011, Ch. 2.

[18] Poisel, R. A., *Introduction to Communication Electronic Warfare Systems*, 2nd ed., 2008, p. 7.

[19] Barros, J., and M. R. D. Rodrigues, "Privacy Capacity of Wireless Channels," *Proceedings IEEE International Symposium on Information Theory* (*ISIT*), Seattle, WA, 2009.

[20] Tse, D., and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005, p. 188.

[21] Bloch, M., and J. N. Laneman, "On the Privacy Capacity of Arbitrary Wiretap Channels," *Proceedings 46th Annual Allerton Conference on Communications, Control, and Computing*, Monticello, IL, 2009, pp. 818–825.

[22] Poisel, R. A., *Modern Communication Jamming Principles and Techniques*, 2nd ed., 2011, Ch. 5.

[23] Poisel, R. A., *Modern Communication Jamming Principles and Techniques*, 2nd ed., 2011, Ch. 9.

[24] Mukherjee, A., and A. L. Swindlehurst, "A Full-Duplex Active Intercept in MIMO Wiretap Channels: Construction and Countermeasures," *Proceedings ASILOMAR Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 2011.

[25] Bustin, R. H. et al., "An MMSE Approach to the Secrecy Capacity of the MIMO Gaussian Wiretap Channel," *EURASIP Journal on Wireless Communication Networking*, November 2009.

[26] Kashyap, A., T. Basar, and R. Srikant, "Correlated Jamming on MIMO Gaussian Fading Channels," *IEEE Transactions on Information Theory*, Vol. 50, No. 9, September 2004, pp. 2119–2123.

[27] Bertsekas, D. P., *Nonlinear Programming*, Belmont, MA: Athena Scientific, 1995.

[28] Armijo, L., "Minimization of Functions Having Lipschitz Continuous First Partial Derivatives," *Pacific Journal of Mathematics*, Vol. 16, No. 1, 1966, pp. 1–4, http://projecteuclid.org/DPubS/Repository/1.0/Disseminate?view=body&id=pdf_1&handle=euclid.pjm/1102995080.

[29] Day, B. P., et al., "Full-duplex MIMO Relaying: Achievable Rates Under Limited Dynamic Range," 2011, http://arxiv.org/abs/1111.2618v1.

[30] Riihonen, T., S. Werner, and R. Wichman, "Mitigation of Loopback Self Interference in Full-Duplex MIMO Relays," *IEEE Transactions on Signal Processing*, Vol. 59, No. 12, December 2011, pp. 5983–5993.

[31] Mukherjee, A., and A. L. Swindlehurst, "Optimal Strategies for Countering Dual-Threat Jamming/Eavesdropping-Capable Adversaries in MIMO Channels," *Proceeding IEEE MILCOM*, San Jose, CA., November 2010.

[32] Goel, S., and R. Negi, "Guaranteeing Privacy Using Artificial Noise," *IEEE Transactions Wireless Communications*, Vol. 7, No. 6, June 2009, pp. 2180–2189.

[33] Mukherjee, A., and A. L. Swindlehurst, "Robust Beamforming for Privacy in MIMO Wiretap Channels with Imperfect CSI," *IEEE Transactions Signal Processing*, Vol. 59, No. 1, January 2011.

[34] Mukherjee, A., and A. L. Swindlehurst, "Fixed-Rate Power Allocation Strategies for Enhanced Privacy in MIMO Wiretap Channels," *Proceedings IEEE SPAWC*, Perugia, June 2009, pp. 344–348.

[35] Zhou, X., and M. R. McKay, "Physical Layer Security with Artificial Noise: Privacy Capacity and Optimal Power Allocation," *Proceedings International Conference Signal Processing and Communication Systems*, Omaha, NE, September 2009.

[36] Philosf, T., and R. Zamir, "The Cost of Uncorrelation and Noncooperation in MIMO Channels," *IEEE Transactions Information Theory*, Vol. 53, No. 11, November 2007, pp. 3904–3920.

[37] Hochwald, B., T. Marzetta, and V. Tarokh, "Multiple-Antenna Channel Hardening and Its Implications for Rate Feedback and Scheduling," *IEEE Transactions Information Theory*, Vol. 50, No. 9, September 2004, pp. 1893–1909.

[38] Mukherjee, A., and A. L. Swindlehurst, "Equilibrium Outcomes of Dynamic Games in MIMO Channels with Active Intercepts," *Proceedings IEEE ICC*, Cape Town, South Africa, May 2010.

[39] Myerson, R., *Game Theory: Analysis of Conflict*, Cambridge, MA: Harvard University Press, 1997.

[40] Fudenberg, D., and J. Tirole, *Game Theory*, Cambridge, MA: MIT Press, 1991.

[41] Poisel, R. A., *Introduction to Communication Electronic Warfare Systems*, 2nd ed., Norwood, MA: Artech House, 2008, Section 7.7.

[42] Zeng, Y., and Y. -C. Liang, "Spectrum-Sensing Algorithms for Cognitive Radio Based on Statistical Covariances," *IEEE Transactions Vehicular Technology*, Vol. 58, No. 4, May 2009, pp. 1804–1815.

[43] Yin, Z, D., et al., "Stackelberg vs. Nash in Security Games: Interchangeability, Equivalence, and Uniqueness," *Proceedings AAMAS*, 2010.

[44] MolavianJazi, E., "Secure Communications over Arbitrarily Varying Wiretap Channels," M. S. Thesis, Notre Dame University, 2004.

[45] Jahn. J.-H., "Coding of Arbitrarily Varying Multiuser Channels," *IEEE Transactions on Information Theory*, Vol. 27, No. 2, 1981, pp. 212–226.

[46] Korner, J., and K. Marton, "Comparison of Two Noisy Channels," in I. Csiszar and P. Elias, (eds.), *Topics in Information Theory*, Amsterdam: North-Holland, 1977, pp. 411–423.

Chapter 10

EW Architecture Simulations

10.1 Introduction

One of the best ways to examine the performance of complex situations is through computer simulation. We present such results in this chapter for EW systems and their impact on some (relatively) realistic scenarios.

We first analyze the engineering performance by simulation. Then we present the results for a simulation in two operational scenarios: (1) Northeast Asia, heavy brigade and (2) urban terrain with a reinforced battalion.

Some of these results have been presented elsewhere [1]; in particular, the operational scenario in an urban terrain setting. They are included here for completeness and comparison.

10.2 Engineering Simulation

A technical (engineering) simulation was used to examine a few jamming architectures. The target simulation was basically the same as that described in [1]. The four jamming architectures considered were:

- A single ground-based thick jammer;
- Two ground-based thick jammers;
- Nine ground-based thin jammers;
- A single airborne UAS-based thin jammer.

The thick jammers were dedicated to the EA mission as there were ES operators controlling the jammers in real time. The thin jammers were assumed to be parasites on platforms whose primary mission was not EA. The thin jammers were

controlled by an EWO located at a central site. Targets were assigned to the thin jammers based on the jammer with the closest proximity to the target receiver(s).

A land component scenario is considered here. The target array consists of adversarial units deployed over a 30 × 50 km region, an area nominally the AOR of a BCT.

The target radios for this analysis consist of *modern combat net radios* (CNRs) that employ SFH. The hop rate is 100 hps.

The jamming method considered here is called *follower jamming*. As a target leaves one frequency and moves to another, the spectrum is measured by a wideband ES receiver colocated with the jammer, and new energy is measured. If the new energy satisfies the sorting criteria, herein consisting of matching the hop phase (time of day) and within some amplitude range, then the new hop is associated with the target of interest and energy is placed at that frequency for the remainder of the hop dwell.

The capability of ES systems to support the EA analysis discussed here was not included in the analysis. It is assumed that the ES systems can perfectly determine whether a transmitting target is one of interest or not. This effectively removes the ES from the analysis and conclusions.

10.2.1 Electronic Attack

Two types of EA assets were considered. The first can be loosely described as a "thick" configuration while the second is referred to as a "thin" configuration. These terms were adopted from the world of computer networks. CONOPS for these configurations are illustrated in Figures 5.12 and 5.13, respectively.

In the jammer, a scanning receiver searches the RF spectrum to determine target activity. If a signal has an adequate SNR at the receiver, that signal is considered detected. In addition, if the selection criteria for that jammer based on hop phase and power level are met, then that frequency (target) is placed on the jammer list for that hop period. Once one sweep of the RF range considered was completed, attempts to jam all the targets on the jam list for that jammer were made. The JSR is computed at each jammed target and if an adequate level of JSR is present then that hop is considered jammed.

The jammers contain a very limited capability for signal acquisition and selection. As mentioned, the only two sorting criteria were hop phase and power level. It is essentially the antenna height that separates the functionality of the ground thin sensors from the ground thick sensors. The former are considered as tenants on the vehicles on which they are placed, whereas the latter assumes a dedicated EW vehicle.

The EW payload for the UAS is also assumed to be a tenant on the platform that normally performs other missions. As such, it is assumed that the space, weight, and power for the EW payload must be minimized. Therefore, it too was

assumed to be a thin jammer, albeit at a considerably higher altitude than the ground jammers.

A thick EA system would possess considerable capability, in particular in the area of organic ES. Signal intercept capability would be included in such a system. The expense of this type of system would dictate that a BCT would have one or two of these at most. The tasking for such EA systems would be more generalized than that for the thin jammers. Perhaps of the form: "jam any artillery network discovered between 1200 and 1300 hrs."

A thin EA system would have very limited capability. The targets would be assigned to these systems by proximity. The ES system supporting the BCT would assign detected targets to the closest EA asset to that target.

The ground-based thin jammers were assumed to be part of tactical vehicles whose primary battlespace function is not EW as opposed to being part of a platform dedicated to the EA mission. As such, the height of the jamming antenna was lower than that for the thick jammers. A height of 3m was assumed.

The opposite was assumed for the thick jammers—these systems have dedicated vehicles compatible with the likely size of the electronics necessary to facilitate the enhanced processing. The antenna height could therefore be higher for the ground-based thick jammers.

An airborne configuration of the thin jammer is included in the analysis. The characteristics of this platform are consistent with the Class IV UAS. The power of the HPA was 100W and the aircraft was centered in the target area shown in Figure 10.1. The class IV UAS may be an unmanned helicopter so could reasonably be expected to fly over a small region or even hover.[1]

The processing in the thin jammers is consistent with the flow diagram shown in Figure 10.2. A thin jammer consists of a fast searching ES receiver, a PA, a synthesizer, a *transmit/receive* (T/R) switch and a control module. The *space, weight, and power* (SWAP) footprint for the thin jammer is assumed to be very limited. The fast tuning receiver (scanning superheterodyne, compressive, or digital receiver) is used only for energy detection.

The block diagram for the thick jammer would be similar except there would be equipment for operators. The SWAP for the thick jammers is such that they exist in their own platform and have considerably more functionality than the thin jammers. The complete ES suite, for example, would be colocated with the jammers. There would be more receiving and sorting capability, with DF functionality (although DF was not included in the modeling here, it would likely be used as a sort parameter as well).

The sort parameters used are hop phase and power level. The *hop phase* is the expected time for the next hop to occur given knowledge of the time of the current hop.

[1] Helicopters can be inherently unstable when they hover so may be inappropriate for unmanned applications.

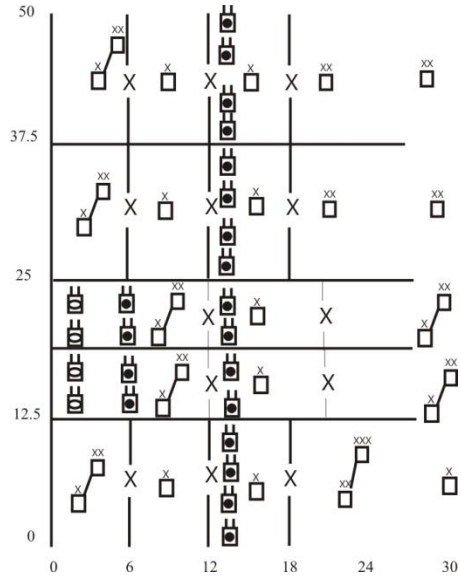


Figure 10.1 Target nodes in the engineering simulation. Distances are in kilometers.

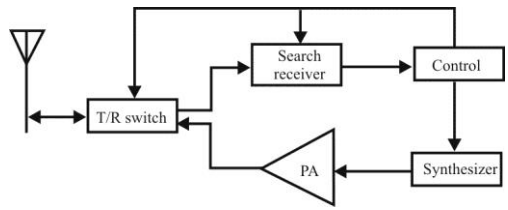


Figure 10.2 Thin jammer process flow diagram.

The power level sort parameter is based on a band of acceptable power levels nominally centered on the power level when the first hop was detected by the ES systems. In this analysis ±10 dBm was used.

The target array is depicted in Figure 10.1. The first 19 target networks are selected from the nodes shown in Figure 10.1. Appendix A contains a listing of what those networks are. For those cases when more than 19 target nodes were included (all of the cases considered herein), the additional networks were randomly placed throughout the target area with a uniform pdf. First the NCS was placed and then the outstations were randomly distributed within a radius of 7 km from the NCS.

The thin jammers had a power level (into the antenna) of 100W. The thick jammers, being housed on their own platform, have more power available—500W. The antenna gain characteristics modified these power levels to establish the ERP.

Likewise, the ground thin jammers, being carried by a nondedicated EW platform, have an effective antenna height of 3m above the ground. For the ground thick jammers, carried on dedicated platforms, higher antenna heights could be accommodated. In this case that height was set at 10m. It is this difference that to a large extent accounts for the difference in performance of these two configurations.

The UAS is flown at an altitude of 1,000m but the jammer is assumed to be a passenger on that aircraft. A thin jammer was therefore assumed for this configuration.

Broadband receivers at each jammer measured the RF spectrum during each time interval. These receivers could be of several varieties; digital or compressive are two of the more common. If a target on the target list of each jammer was detected, then an attempt was made to jam that target by applying some or all of the jammer power on that frequency. The amount of power available at each frequency in such a scheme decreases as the square of the number of simultaneous jammer tones (see Section 8.4.1.2), so if there are too many targets detected, the JSR at the receiver may not have been adequate to accomplish effective jamming. It was assumed that the JSR needed for effective jamming is 0 dB, consistent with most CNRs, and, based on the analysis above, is a reasonable level to ensure effective jamming in most cases.

The receivers dwell on a broadband channel for 200 μs for all the results herein, implying a frequency resolution of 5 kHz. The IBW for all cases here was 4 MHz, which equates to a scan speed of 20 GHz/sec.

An SNR of 15 dB is necessary for a signal to be detected by the jammer. This SNR level would produce a BER in the 10^{-1}–10^{-2} range for a fading channel as illustrated in Figure 10.3. The BER decrease at a linear rate, inversely proportional to the SNR, as opposed to the exponential rate usually encountered for nonfading channels.

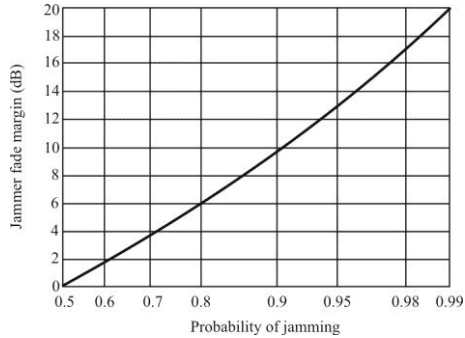


Figure 10.3 Jamming fade margin for barrage jamming.

10.2.1.1 Signal Fading

The Egli model used in this analysis does not take fading into consideration. Signal fading has an effect on both the target communication links as well as the jammer signals [2–4]. For a barrage jammer the jammer fade margin is as shown in Figure 10.3.

The expected jammer coverage radius for follower jamming assuming a 10W jammer, 1W target transmitter, omnidirectional antennas, and a 1 km target link distance has the effect illustrated in Figure 10.4 [2]. λ is the number of target receivers in the coverage area, R^{-4} propagation is assumed, and the targets are assumed to have an 18 dB processing gain. Rayleigh fading is assumed for this example and the signal bandwidth is equal to the data rate (for SINCGARS, the data rate is 16 kbps and the VHF channel width is 25 kHz, so they are close). The results are similar for a pulsed jammer.

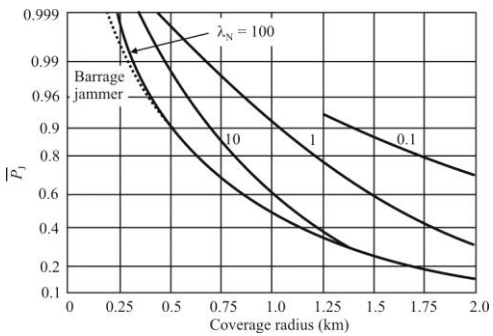


Figure 10.4 Follower jammer reliability degradation due to increased coverage radius example.

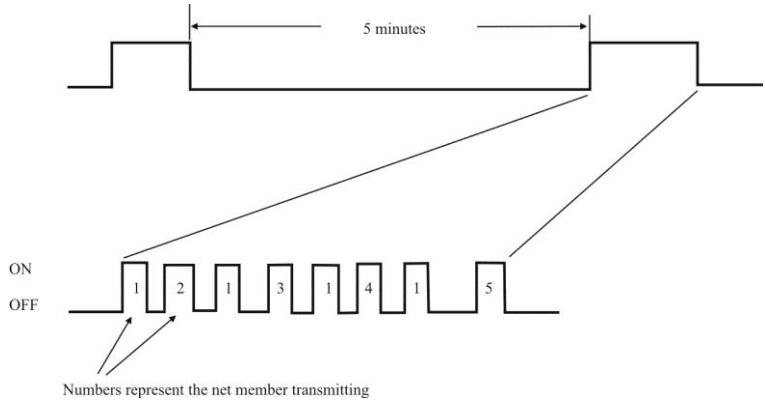


Figure 10.5 Node communication sequence.

10.2.1.2 Noise

Specified levels of SNR at the receivers is required in order to declare detection by the ES receivers. For the results presented here that level is 15 dB. The noise that is added to the signal consists of the receiver noise figure, assumed to be 10 dB here, and rural high external noise levels [5].

10.2.1.3 Dynamic Range Limited

The effects of limited dynamic range were included in the modeling by placing a high powered interfering signal within range of the jammer receivers. For the results presented here, the dynamic range was set at 72 dB, although other values are possible. The interfering signal is a 10 kW radio transmitter placed at 60 MHz, 10 km from the receiver.

10.2.2 Transmission Sequence

The networks in the simulation are assumed to consist of five nodes each. These networks had a *network control station* (NCS) that is the most senior node on the network. The remaining four nodes are called *outstations* (OS).

A precise script for the transmitters is not available. The simulation included the communication pattern illustrated in Figure 10.5 for the networks. All of the time variables followed approximately a normal distribution (approximately because the negative tail was not included). A transmission sequence was initiated by the NCS calling OS 1. These transmissions had a mean value of 5 seconds. OS 1 responded after a short period. That was followed by the NCS calling OS 2, and OS 2 responded. This was continued until all the nodes in the network transmitted. The average time between transmission of the nodes during one such sequence was 300 seconds. After all nodes have communicated the network remained off for a time period with a mean value of 300 seconds.

No value judgment on whether it makes sense to jam every hop. In fact, jamming 1 in 5 will produce a BER of 10^{-1} (assuming scan time and SNR and JSR are adequate to jam half of the hops that are jammed). With the typical coding used for tactical voice or data communications this is probably adequate.

10.2.3 Jammer Placement

When there is a single jammer simulated, it is placed in the center of the target array shown in Figure 10.1. The same is true when the single air jammer is modeled. With two jammers considered, they are placed in the middle, left to right, and evenly spaced north to south. With nine jammers in the simulation, they are placed uniformly throughout the target area shown at coordinates (7, 12), (7, 24), (7, 36), (14, 12), (14, 24), (14, 36), (21, 12), (21, 24), and (21, 36).

The jammers considered herein, in all cases, are intelligent jammers [6, 7]. Such jammers sense and measure the RF environment and produce optimum jamming waveforms depending on the results of the measurements. When it is assumed that the target signals are Gaussian (a somewhat general case), then the optimum jamming waveforms are also Gaussian.

10.2.4 Results

The results of these simulations are summarized in Figures 10.6 and 10.7. Figure 10.6 indicates the probability of jamming a communication transaction from all of the CPs versus the number of target networks for the four jammer configurations considered: (1) a single thick ground-based jammer, (2) two thick, ground-based jammers, (3) nine ground thin jammers, and (4) a single airborne (UAS) thin jammer. The airborne jammer was clearly the best configuration. Not only could it "see" most, if not all of the transmissions in the AOI, but its 100W PA power was sufficient to reach all the receivers in the AOI. The nine thin jammers performed the worst, due to both a signal intercept level inadequacy as well as jammer power inadequacy. The two thick jammer configurations fell in between these two extremes.

Figure 10.7 displays the simulation results when the targets of interest were limited to the division CP communications only. The results are similar to those for all CP communications, with the airborne thin jammer producing the best results and the ground thin jammers the worst.

10.2.4.1 Information Attributes

EW is one of the effector capabilities available to the land component commander. As indicated in Chapter 2, it potentially impacts 23% of the tenents of information according to Fewell. The results obtained herein indicate that there are configurations of EA systems that can deny a considerable amount of information

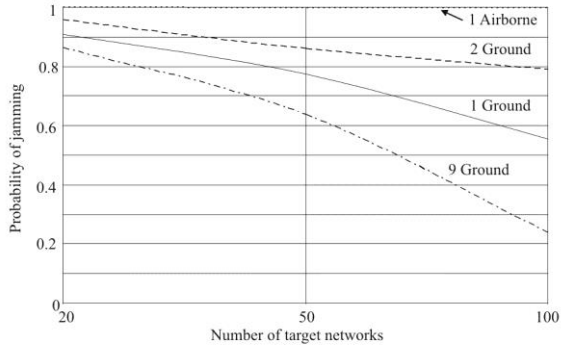


Figure 10.6 Engineering simulation results, all CPs.

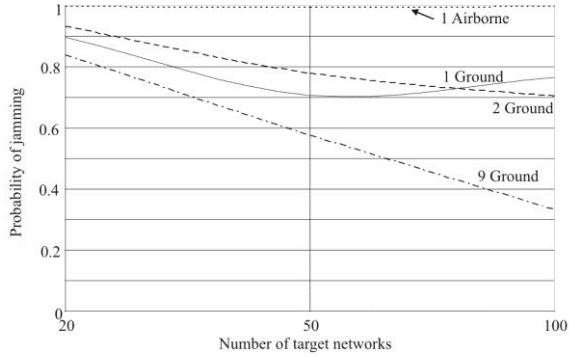


Figure 10.7 Engineering simulation results, division CPs only.

from being exchanged within an adversarial force. The results indicate that an airborne thin jammer is the best suite, followed by ground-based thick jammers.

10.2.4.2 EW Target Analysis

Ideally, most of the EW signal analysis required to successfully conduct EW operations is accomplished prior to hostilities. However, EW intelligence analysis is carried out both in the IPB phase as well as during operations.

10.2.4.3 Cognitive Hierarchy

The OODA loop and the cognitive hierarchy both assume that sensing the environment and exchanging the results with other edge entities occurs unhindered. These results indicate that significant impacts can be imposed by successful EW actions. Information about the environment being faced by an adversary can be denied with reasonable friendly capabilities.

10.2.4.4 Situation Assessment

Since EW against targets can be as effective as Figures 10.6 and 10.7 indicate (remember the ideal circumstances assumed, however), developing an accurate assessment of the situation by an adversary could be very difficult, at least in real time. The timeliness of the information exchanges is significantly affected by the jamming assets.

10.2.4.5 Thin or Thick Jammers

It is clear from Figures 10.6 and 10.7 that the SNR is a significant contributor to success or failure to jam targets. With low antenna heights as assumed for the ground thin jammers, even though there were substantially more of them, they were able to jam considerabley fewer targets.

The SNRs were considerably lower for the ground-based thin jammers, even though there were more jammers available. Note that f_{50} is higher than both the 20 network case and the 100 network case. For the low end, this follows because when there are more targets, there are more placed close to the jammers and the signal levels are higher from closer targets. However, when 100 networks were simulated, the low antenna heights associated with the thin jammers precluded getting enough power to detect the targets.

10.2.4.6 Number of Target Networks

As we showed in Section 8.4.1.2, the power available from a jammer per target decreases as N^{-2} irrespective of the number of jammers and whether they are thin

or thick. Therefore as more target nets are included, it would be expected that fewer targets would be successfully engaged. The SNR required for detection is 15 dB while the JSR required for successful jamming is 0 dB. In this analysis it is clear that if a target were detected, then there was adequate JSR. This can be concluded by comparing the results for the airborne configuration with the ground thin configuration. Even with nine ground thin jammers, the airborne jammer was much more successful.

It is the N^{-2} characteristic that motivates the analysis of jammers that only jam on active targets, as opposed to barrage jamming. Such jamming, on the other hand, requires an ES receiver colocated with the jammer, although this receiver can be simple. A scanning superheterodyne or a fast scanning compressive receiver are adequate. In addition, with barrage jamming, considerable fratricide on friendly communications can result.

10.2.5 Summary and Conclusions from the Engineering Simulation

The results of an engineering simulation for four types of jamming architectures were presented in this section.

It should be noted that the simulation produced results that were displayed in terms of probability of jamming communications. Because it was an engineering simulation where some typical values for target parameters were used, no judgment was concluded nor suggested about the adequacy of the probability of jamming. It can probably be said that producing $P_J \sim 1$ is overkill, and not necessary in most cases. It is more difficult to distinguish between $P_J \sim 0.3$ and $P_J \sim 0.5$, however.

10.3 Operational Simulation

A study on EA architectures via computer simulation was conducted to examine the best configuration to support networked ground-component forces. The three specific cases examined were:

- A set of distributed thin jammers;
- One and two thick jammers;
- One UAS mounted jammer.

In addition to these, simulations were executed for the baseline where there was no jamming and when jamming was complete—no communications were allowed to transpire.

A standard threat model for the Northeast Asia region was used as the first scenario. The second scenario used was *mounted operations in urban terrain* (MOUT).

ES was not modeled for the study principally because the model used does not include the capability for ES. Thus, it was assumed that the locations of the targets were known and the closest jammer to a target was tasked with jamming that target. This in effect removed the ES portion of the problem and concentrated the results strictly on the EA aspect.

10.3.1 Scenario Model

The model used was a force-on-force model. It had high-resolution, with details from brigade to the individual soldier. It modeled individual systems and soldiers and is constructive and event sequenced. It is based on statistics so is stochastic with statistical validity. There is engineering level of rigor at key points in the models. It models the effects of systems and forces capabilities. MOEs included:

- *Loss exchange ratio* (LER)
- Blue/Red kills/losses
- Weapons expended
- Cost effectiveness

The battlefield phenomena represented include:

- Terrain;
- TIREM (ERDEC);
- Okamora/Hata signal propagation models [8];
- Jamming;
- Weather;
- Communications;
- Deployment and Tactics;
- Weapon Characteristics
- Intervisibility;
- Target Acquisition;
- Movement—Ground and air;
- Engagements;
- Direct / Indirect Fire;
- Suppression;
- Smoke and dust;
- COMBIC (ARL-BED).

10.3.2 EW Methodology

The methodology used to model the effects of jamming are summarized in the following three equations:

$$\text{JamPower(dB)} = \text{JamEIRP} - \text{JamLoss} + \text{RcvGain} \quad (10.1)$$

$$\text{RcvPower(dB)} = \text{XmitPwr} - \text{XmitLoss} + \text{RcvGain} \quad (10.2)$$

$$\text{SJR(dB)} = \text{RcvPower} - \text{JamPower} \quad (10.3)$$

The following logic was used to determine whether the jamming was successful:

IF S/J(DB) < THRESHOLD $\rightarrow$ COMMO JAMMED
ELSE $\rightarrow$ COMMO SUCCESSFUL

The attenuation losses in (10.1)–(10.3) are calculated using TIREM (NEA) or the Okumura (MOUT) models. Further logic to ascertain jamming success is based on Figure 10.8. If the target receiver is within the jammer's antenna pattern, as illustrated by the lines emanating from the jammer in Figure 10.8, (10.1)–(10.3) were calculated to determine if the message was received.

10.3.2.1 EW Limitations

The EW capability had some limitations. In particular:

- Only EA is modeled; ES and EP are not currently implemented. Therefore modeling accurate follower jamming is not possible. This results in the lack of ability to explicitly play frequency hopping and the effects of propagation effects due to frequency, multipath fading, and so forth were not examined.
- The jammer only operates on one frequency.

10.3.3 Key Assumptions

The principal assumptions made in the modeling process were:

- Blue maneuver and communications were precoordinated, minimizing negative effects of jamming fratricide.
- Jammer is reactive; it only activates when it receives threat communications and turns off when communications have ceased;
- Red was unaware of Blue jamming resulting in Red tactics not changing due to jamming.

10.3.4 NEA Scenario

The NEA scenario is depicted in Figure 10.9. It consisted of a U.S. heavy modular brigade maneuvering through a heavily defended Red zone to seize the objective

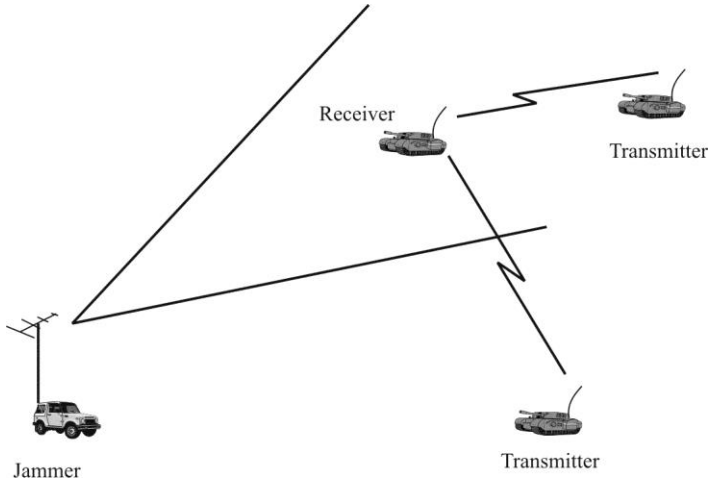


Figure 10.8 Jammer model for the operational simulations.

to the north. The mounted maneuver was through restrictive and mountainous terrain with average visibility. Both mounted and dismounted assaults were conducted. Apache teams supported ground forces in clearing the battlefield during early operations. Artillery provided support based on SA from UAS's, counter battery, and forward scouts.

Threat regiment consisted of conventional armor and mechanized forces. The defense was sectorized and dug in. The Red forces utilized maximum cover, concealment, and camouflage. The *indirect fires* threat was significant consisting of artillery and mortars on reverse slopes or near urban areas. The vehicle exposure was minimal. The reserve Armor battalion maneuvers to block Blue advance, when called forward to do so.

10.3.4.1 Role of EW

The role of the Blue EW in the simulation was to:

- Suppress/prevent threat IF;
- Prevent threat from requesting support from armor reserve;
- General disruption of threat C2/coordination/synchronization.

10.3.4.2 Scenario Timeline

The timeline of the NEA scenario (expressed in minutes) was as follows:

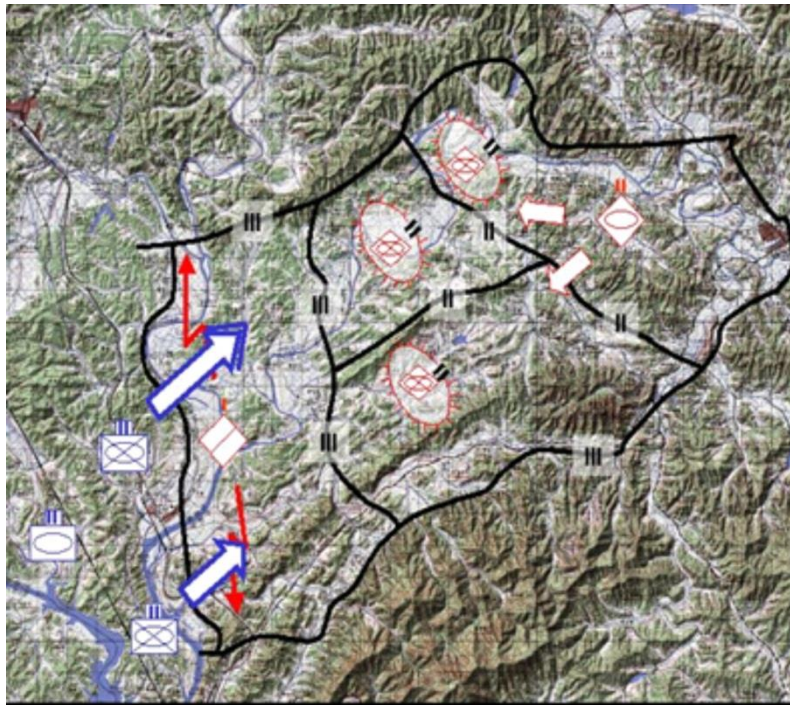


Figure 10.9 NEA operational simulation scenario.

Figure 10.10 Heavy thick jammer.

T = 0 – 110: Blue clearing /recon operations with UASs, Helos, IF, and scout vehicles.
T = 110 – 421: Main Blue maneuver up through valley. Red holds defensive positions and counters with IF. Blue provides support with helos and IF.
T = 370: Red calls for support from armor reserve.
T = 421: Scenario ends.

10.3.4.3 Description of Cases

Thick Jammer Architecture

The thick jammer for the NEA scenario might be housed in an armored tracked vehicle such as the one shown in Figure 10.10. Its radiated power was 27 dBW (500W) and the omnidirectional antenna was placed atop a 5m mast. There were two of these systems in the brigade and jamming was assumed to be the sole mission. The jammers were assumed to travel behind the main combat force and they were leap-frogged to provide persistent jamming.

Ground Thin Jammer Architecture

The ground thin jammer might be placed on a soft vehicle such as the HMMWV shown in Figure 10.11, configured as a command and control vehicle (C2V). The jammer emitted jamming waveforms at 20 dBW (100W). The omnidirectional jamming antenna was mounted on top of a 3m mast. Since there are six C2Vs per Battalion there were 18 in the brigade. The principal mission of the C2V is not jamming so the primary mission of the ground thin jammer vehicle was not jamming either. These vehicles maneuvered with the main combat force and the jamming was active while maneuvering.

Figure 10.11 Thin jammer ground.

Thin UAS Architecture

The UAS-based thin jammer was assumed to follow the flight characteristics of the Fire Scout pictured in Figure 10.12. The jammer emitted 20 dBW (100W) from an omnidirectional antenna placed where it had unobstructed views below the chassis. There was one of these UASs in the brigade and it was assumed that jamming was sole mission of UAS since it was not maneuvered off of its continual orbit above the battlefield. Jamming was active while maneuvering.

Optimal Effects

An optimal case was included in the simulations for comparison purposes. In the optimal case, no Red communication was allowed at all.

10.3.4.4 Base Results

In the complex terrain of NEA, the UAS is able to provide better jamming coverage because it can rise above the clutter that limits ground-based performance. Jamming performance for the base case is illustrated in Figure 10.13.

Loss-Exchange Ratio

- Overall Blue force performance is significantly increased with effective jamming present:
    - 158% increase in Blue performance in UAS case.
    - 231% increase with theoretical “Optimal” coverage.

Figure 10.12 Fire Scout thin jammer.

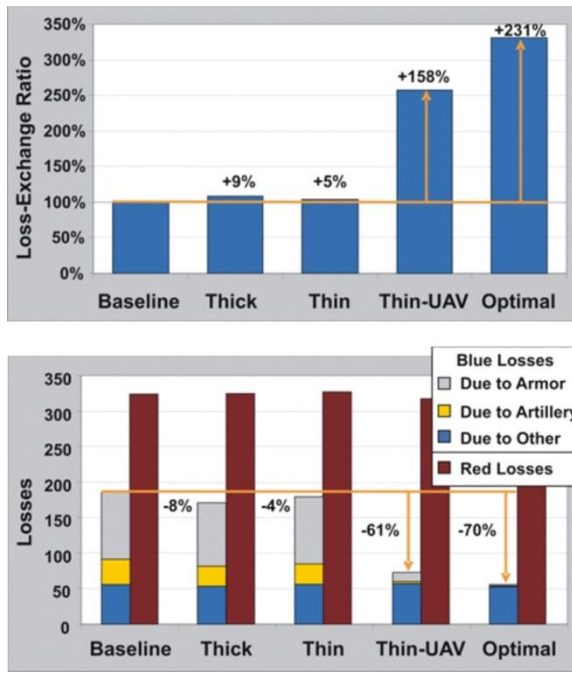


Figure 10.13 NEA 2010 scenario base results.

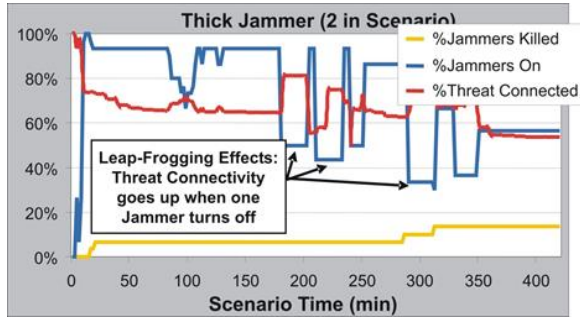


Figure 10.14 NEA scenario thick jammer performance over time.

- UAV system is only feasible architecture that significantly improves LER in this scenario.
- Benefits approach those of unrealistic "optimal" case.
- High vantage point of UAS allows for almost complete saturation of battlefield.
- Disruption of scout reports and call for fires.

Losses

- Red losses remain constant.
- Jamming able to significantly reduce Blue losses due to threat armor and artillery.
- Disruption of call for backup to reserve armor.
- Disruption of scout reports and call for fires.
- 61% reduction in Blue losses in UAS case.
- 70% reduction with theoretical "optimal" coverage.

Ground-based systems did not perform as well as the UAS due to the complex terrain preventing the jammers from affecting distant receivers.

10.3.4.5 Jammer Effectiveness over Time

Thick Jammer Performance

Thick jammer performance over time is shown in Figure 10.14. We can see from this chart that the higher power of the thick jammer allowed for disruption of threat communications but the mountainous terrain limits overall performance. In

Figure 10.15 NEA scenario thin ground jammer performance over time.

addition, leap-frogging maneuvering limits performance. As soon as one system shuts down to maneuver, connectivity rises.

Thin Jammer Performance

Ground-based thin jammer performance in the NEA scenario is shown in Figure 10.15. We can see that thin jammers are weaker in power than the thick jammers and in the first part of the scenario, stand-off considerable distances. They are thus not able to disrupt the majority of the threat communications during this phase. Performance significantly improves as thin jammers maneuver in closer to the threat.

UAS Jammer Performance

The UAS mounted thin jammer performed as illustrated in Figure 10.16. The threat connectivity was low throughout the NEA scenario. The high vantage point

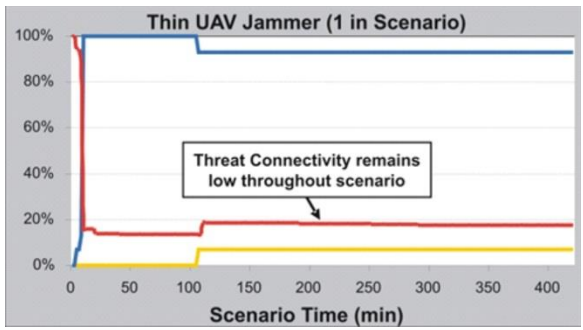


Figure 10.16 NEA scenario UAS thin jammer performance over time.

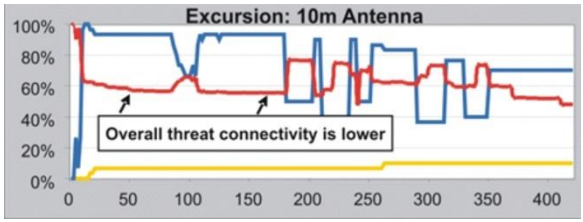


Figure 10.17 NEA scenario higher antenna mast for thick jammer excursion results versus time.

of the UAS allows for nearly complete saturation of battlefield and limits threat connectivity. Threat signal levels were adequate for the UAS to detect the vast majority of them and the 100W power level was sufficient to overpower the threat communications.

10.3.4.6 Thick Jammer Excursions—Performance

There were two excursions modeled for the thick jammer. The first excursion examined the limitations due to the relatively low antenna height of 5m. The omnidirectional antenna was placed atop the 5m mast in the base modeling. Poor jamming performance resulted due to complex terrain limiting the effective range of the thick jammer. For this excursion the antenna was raised to 10m to examine the differences.

The results are shown in Figure 10.17. There was essentially no improvement in performance. A majority of the Red forces were still able to communicate. The terrain in NEA simply limits the performance of stand-off jamming platforms. In addition, the leap-frog maneuvers still had a negative effect. The conclusion was that doubling the height of the antenna did not significantly improve performance in this scenario.

The second excursion modeled for the thick jammer allowed the jammers to be active while moving (the antenna height remained at 5m). The results of this excursion are illustrated in Figure 10.18 (base case) and Figure 10.19 (active while maneuvering). This resulted in leap-frogging no longer being a problem, but the complex terrain still hampers overall performance of the thick jammers. The conclusion was that eliminating leap-frogging by allowing operations on the move improves jamming consistency, but does not significantly lower threat connectivity.

10.3.4.7 Overall Conclusion for the Thick Jammers

Even with double the antenna height, the two thick jammers provided limited battlefield coverage due to complex terrain. Improving jamming consistency has a limited effect on threat connectivity.

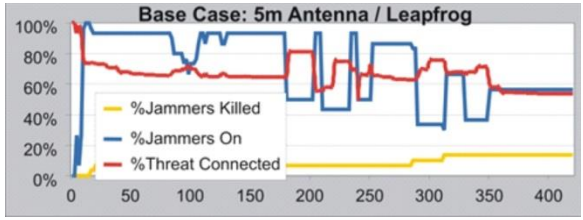


Figure 10.18 NEA scenario base case—5m antenna and inactive while maneuvering.

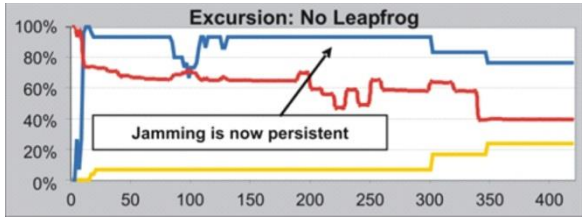


Figure 10.19 NEA scenario excursion—active while maneuvering (5m antenna height).

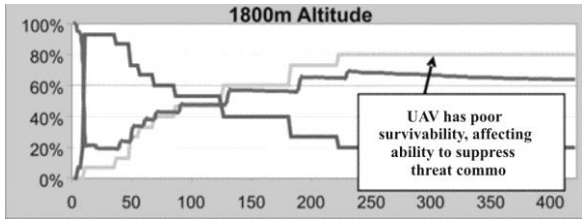


Figure 10.20 NEA scenario base case results for UAS performance at 1,800 m altitude.

10.3.4.8 UAS Excursions – Survivability

The modeling excursion for the UAS thin jammer dealt with survivability of the UAS. In accordance with U.S. Army doctrine, the UASs in the NEA scenario are placed at 1,800 m altitude, but were vulnerable to enemy ADA and consequently lost 80% of the time before the halfway point in scenario. In addition, the best altitude for both good jamming performance and UAS survivability was unknown. Therefore, two altitude excursions were modeled to find a better height for the UASs in the NEA scenario.

The results are shown in Figures 10.20–10.22. Figure 10.20 shows the results for the base case and indicates that the UAS is typically lost at some point during the scenario. (Fractional results are possible even though there was only one jamming UAS because these results are the average of several simulation modeling passes.)

Figure 10.21 shows the results when the altitude was raised to 3 km. Clearly, survivability dramatically improved resulting in suppression of Red communications throughout the simulations. Figure 10.22 shows similar survivability results when the UAS was raised to an altitude of 6 km—in fact shoot-downs were completely eliminated. However, the Red communication connectivity was higher than for 3 km because the jammer was further from the Red communications nodes.

10.3.4.9 Conclusions

The conclusions from these excursions are that the altitude of 3,000m gave the best combination of survivability and communications denial performance for the NEA scenario. As a result an altitude of 3,000m was used for all other UAS cases considered.

The UAS is effective at jamming the battlefield, but is vulnerable to threat air defense artillery. Raising the UAS altitude improved survivability, but degrades jamming performance. As would be expected, the best altitude is scenario dependent.

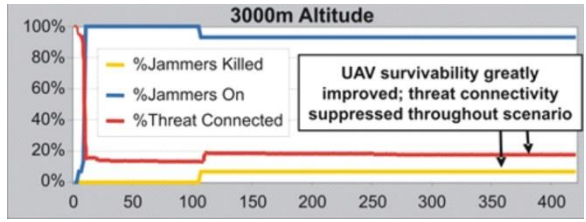


Figure 10.21 NEA scenario UAS excursion, raising altitude to 3,000 m.

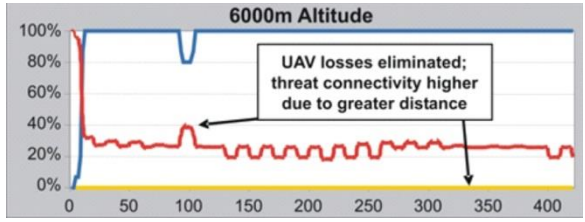


Figure 10.22 NEA scenario UAS excursion, raising altitude to 6,000 m.

10.3.4.10 NEA Overall Conclusions

The NEA scenario is a large-scale engagement with many players. The Red threat communications occur over a large area. As such, jamming must cover a large area to completely disrupt threat communications. A majority of key Red threat players (artillery, reserve units, and so forth) were positioned far away from the advancing Blue forces. The jamming systems must be able to overcome terrain features to affect these key threat players

Both the thin and thick ground-based architectures have degraded range in complex terrain. This prevents coverage of sufficient areas of large-scale battles to provide a substantial benefit.

The UAS-based thin architecture overcomes terrain problems with a high-altitude orbit and is able to cover the entire battlefield. Furthermore, it provides constant suppression of threat communications. In the complex terrain of NEA, the UAS is able to provide better jamming coverage due to it being able to rise above the clutter that limits ground-based jammer performance

10.3.5 MOUT Scenario

10.3.5.1 MOUT Scenario Overview

The MOUT scenario included portions of a U.S. Combined Armored Battalion [denoted by CAB(–)] with a supporting *non-line of sight* (NLOS) Battalion [Bn(–)] conducting MOUT in support of a Joint Task Force (JTF) conducting operations to restore a legitimate government. (See Figure 10.23.)

In the MOUT scenario the urban terrain is restrictive and complex with minimal line of sight (LOS). The Okumura-Hata communication propagation model was developed as the most accurate model for simulating urban terrain. Both mounted and dismounted Blue assaults were conducted. The ground movements were augmented with joint air support available for precision strikes. Artillery/NLOS-LS assets provided on-call support. An augmented threat company [CO(+)] consisting of mixed conventional and unconventional forces attempts to defend government seat of power. Adaptive/evasive urban tactics (OE-compliant). The Red forces reacted dynamically to Blue's advancement. They made good use of coverage provided by urban terrain. There was extensive IED placement. However, they had limited ADA and mortar support and no artillery or air support. Red forces used buildings to their advantage by using pop-up tactics from windows and roofs.

Red forces relied on barriers and IEDs to slow/prevent Blue maneuvers causing Blue to stop at barriers and remove the obstacles, making them vulnerable. There were two types of IEDs in play:

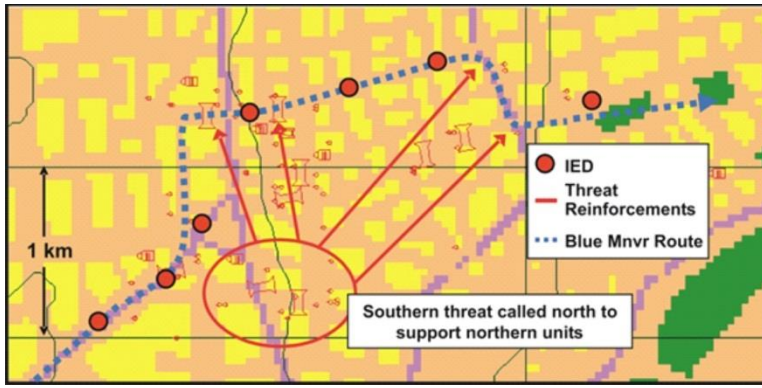


Figure 10.23 MOUT scenario.

- Typical roadside IED, aimed at destroying passing vehicles. (Five of seven IEDs are of this type.)

- IEDs/ bombs placed deliberately near barriers where Blue infantry would be vulnerable. These were aimed at inflicting damage to Blue infantry forces. (Two IEDs near barriers are of this type.)

.

10.3.5.2 Role of Electronic Warfare in the MOUT Scenario

EW was used in the MOUT scenario for three primary purposes:

- Counter Red IED remote triggering (7 in scenario);
- Prevent Red forces from calling for support from southern units;
- General disruption of Red C2 and coordination/synchronization.

10.3.5.3 MOUT Cases Examined

Thick Jammer Architecture

The thick jammer used in the cases examined consists of EW (ES and EA) equipment mounted on a small vehicle, such as the HMMWV shown in Figure 10.24. Such a configuration could house up to two operators, but only one jammer was used in the modeling. The transmitter emitted 27 dBW (500W) and the omnidirectional antenna was mounted atop a 5m mast. It was assumed that jamming was the sole mission of the vehicle. There was one jammer per CAB (–) included. The jamming vehicle travels behind main combat force and it must stop to provide jamming.

Figure 10.24 MOUT scenario thick jammer example.

Thin Jammer Architecture

Two versions of the thin jammer were simulated: (1) a jammer mounted in all of the RSTA vehicles in the CAB(–) for which there were five in the simulation (see Figure 10.25), and (2) a jammer mounted in the Fire Scout UAS (see Figure 10.26). In both cases the thin jammer emitted 20 dBW (100W).

In the ground configuration the omnidirectional antenna was mounted atop a 3m mast. In addition, jamming was not the sole, or even the principal, mission of the vehicles. The vehicles maneuvered ahead of the main combat force and the jammers were active while maneuvering.

UAS Jammer Architecture

There was one thin UAS jammer in the CAB(–). The UAS maneuvered in a continual racetrack above the city. Because of the flight pattern of the UAS, jamming was the sole mission of UAS (the UAS was not retasked to provide coverage of other space than that required for the jamming racetrack). The jammer was active while maneuvering. The flight characteristics of the Fire Scout UAV shown in Figure 10.26 were assumed.

Optimal Effects

Again, in the MOUT scenario an optimum case was included which simulated no Red communications.

10.3.5.4 Base Results

Loss-Exchange Ratio

The base case LER performance is shown in Figure 10.27. Overall blue force performance is increased with effective jamming present. The ground thin

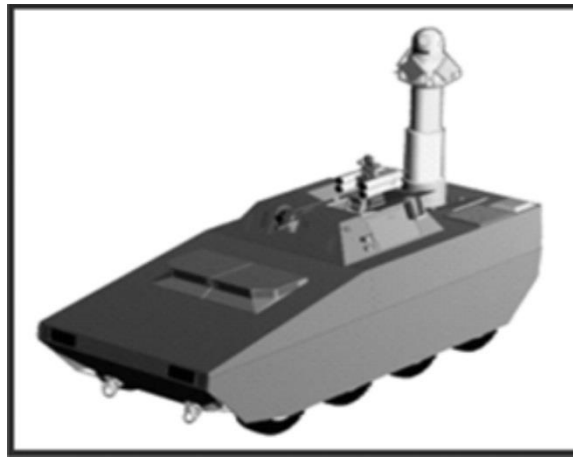


Figure 10.25 MOUT scenario RSTA vehicle.

Figure 10.26 MOUT scenario Fire Scout thin jammer.

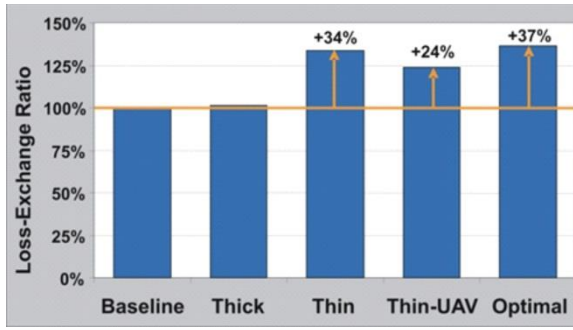


Figure 10.27 MOUT LER base case.

case was close to providing the same benefit as the unrealistic "optimal" case. The first reason for this is the fact that the jamming platforms remain near to Blue forces, protecting them from immediate effects of IEDs. The second reason for this is that the platforms were able to jam while moving, preventing any lapse in protection.

The UAS case does not perform as well due to tall buildings affecting jammer coverage.

Losses

The Red losses remained relatively constant as we can see in Figure 10.28. Jamming was able to significantly reduce IED detonations, decreasing blue losses due to IEDs:

- 92% of IEDs detonated in Baseline;
- 56% of IEDs detonated in Thin case;
- 69% of IEDs detonated in UAS case;
- 23% of IEDs detonated in optimal case.

In contrast, the thick jamming architecture did not perform well. This is primarily due to the fact that the small Blue force had only one thick jammer in the scenario. Furthermore, there were lapses in coverage when the jammer must maneuver to the next position, and when doing so jamming activities ceased leaving no operating jammer in the scenario.

In complex MOUT environments, the thin architecture provides the best jamming coverage due to its proximity to Blue forces and wide distribution across the battlefield.

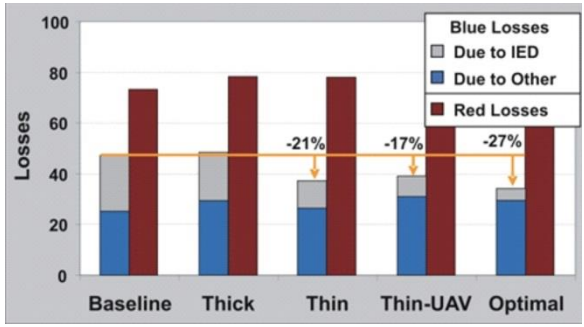


Figure 10.28 MOUT scenario losses for the base case.

10.3.5.5 Jammer Effectiveness over Time

Thick Jammer Performance

The performance versus time for the thick jammer in the MOUT scenario is shown in Figure 10.29. A thick jammer is effective when on but needs to shut-off while maneuvering which allows threat to eventually succeed with communications.

Thin Jammer Performance

The ground thin jammer performance for the MOUT scenario over time is shown in Figure 10.30. The jamming platforms are relatively close to the threat so the weaker power of the thin jammer is adequate, unlike in the NEA scenario. The relatively wide distribution of jammers also provides more widespread disruption of threat communication than a single source. In addition, these jammers could remain operational while maneuvering as opposed to the thick jammer assumption that it could not.

UAS Thin Jammer Performance

These results are shown in Figure 10.31 for the MOUT scenario. Due to the complex terrain, consisting mostly of (relatively) tall buildings, the UAS performance was not very good—especially in the latter stages of the engagement. Throughout the scenario the buildings prevented adequate signal strength from

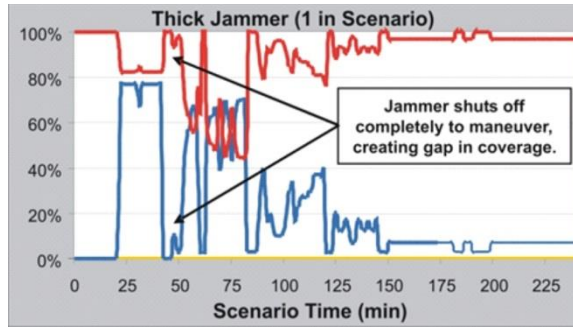


Figure 10.29 MOUT scenario thick jammer performance versus time.

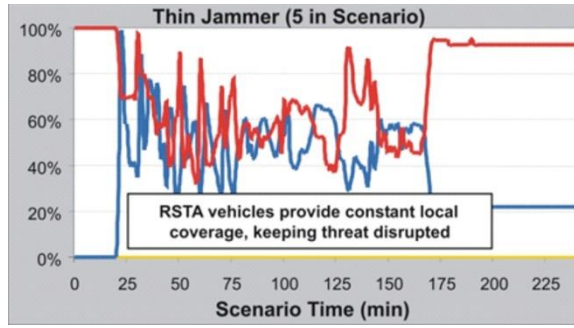


Figure 10.30 MOUT scenario thin jammer performance versus time.

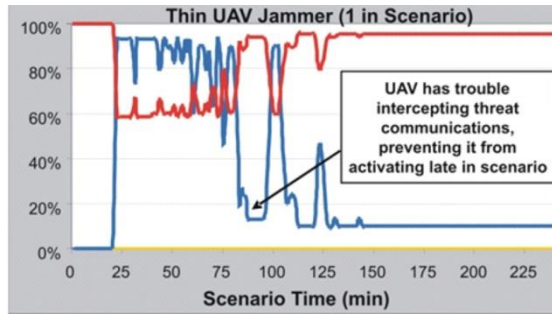


Figure 10.31 MOUT scenario thin UAS jammer performance.

reaching the receiver in the UAS, preventing signal detection, and therefore missing the jamming opportunity.

10.3.5.6 MOUT Conclusions

The MOUT scenario represents a small-scale engagement in an urban environment with few players and a heavy IED presence. Red threat communications travel a relatively small distance. The maximum range of a jamming system is not as crucial as urban penetration. The amount of threat C2 communication/coordination is much less than in a large scale scenario such as for NEA.

A jammer must be in a position to intercept the threat if "reactive" jamming (as opposed to barrage jamming) is used. The jamming platforms must be widely distributed and maneuver with forces to provide constant coverage.

The majority of the key threat players (IEDs, units calling for reinforcement, and so forth) were positioned close to the advancing Blue force. If jammers are sufficiently distributed among the Blue force, the jamming system only needs to influence small local areas which means that low power is required.

The MOUT environment is very complex with many parts of the advancing Blue force being separated from one another. In the complex MOUT environment, the ground-based thin architecture is able to provide the best jamming coverage due to its proximity to protected forces and wide distribution throughout the battlespace.

The thick ground-based architecture is ineffective if only one jammer can be allocated to CAB due to it needing to maneuver often to remain with main force, which requires jamming to be shut-off. Adding another heavy jammer and carefully coordinating maneuvers between the two makes the thick-based architecture a viable option.

Thin ground-based architecture performs well in the MOUT scenario due to it being widely distributed and local to the maneuvering forces, as well as its ability to continue to operate while maneuvering. There was a slight improvement in performance gained when thin jammers added to C2 vehicles as well as the RSTA vehicles due to the RSTA vehicles maneuvering ahead of the main force, leaving some portions unprotected.

The thin UAS-based architecture performed adequately but had difficulty with the complex urban terrain. The UAS did not intercept many of the threat communications due to the scarcity of threat communications and blockage from large buildings, resulting in the UAS jammer being inactive for a large portion of the scenario.

In complex MOUT environments, the ground-based thin architecture is able to provide the best jamming coverage due to its proximity to Blue forces and wide distribution across battlespace. The thin jammer architecture is better suited to the complex MOUT terrain due to the large number of jammers present and the sufficiency of their range.

10.4 Recommendations

For rural operations, one UAS-based thin jammer is the recommended solution. This would be a Class IV UAS due to payload and "high" operating altitude of 3,000m, assuming the latter could be negotiated within the services. For MOUT operations, the ground-based thin jammers is the architecture recommended. The RSTA and C2 ground vehicles, or their equivalents provide the best MOUT protection.

Acquisition costs may be prohibitive for both a dedicated UAS with a thin jammer for rural use plus jammers in five to eight nondedicated ground vehicles for MOUT. If feasible, a multirole UAS with a supplemental jammer payload would be far more cost effective than acquiring a dedicated UAS. Operational costs should be much lower for ground-based architectures.

Multiple UAS-based thin jammers could be distributed in orbits above maneuvering Blue forces instead of distributing among RSTA and C2 vehicles. This avoids purchase of both UAS and ground jamming architectures. Performance should approach, but not equal, that of distributed thin ground jammers.

If only one architecture can be deployed, then a UAS-based thin jammer is the most robust single architecture. A hybrid UAS/ground approach with thin jammers on nondedicated platforms is the most capable and cost-effective architecture. The UAS-based thin jammers offer the most robust performance from a single architecture.

The ideal approach is a hybrid architecture using a UAS-based thin jammer for rural operations and thin jammers on RSTA & C2 ground vehicles for MOUT operations.

10.5 Concluding Remarks

The edge information battlespace was described and considerable justification for the necessity for communications within this battlespace was presented. Future warfare will depend extensively on the ability to communicate. Creating dominant battlespace knowledge, where friendly forces know more about an adversary than that adversary knows about friendly forces will only be possible if communications is facilitated.

10.5.1 Engineering Simulation

Analysis results of four architectures based on simulations were presented. It was shown that the best jammer configuration is airborne. Those results assumed a thin jammer architecture, where little processing was available within the jammer. The

second best configurations were ground thick jammers. The worst performance was ground-based thin jammers. The key discerning factor for these results is the height of the antenna as opposed to the amount of processing in the jammer.

10.5.2 Operational Simulation

The operational scenarios concluded that EW systems can be effective force multipliers, but their effectiveness depends on the architecture selected. Furthermore the best architectures (of those simulated) depend on the particular scenario being examined. The results for the NEA scenario indicated that a UAS configuration was best while that for the MOUT scenario is the distributed thin architecture.

References

[1] Poisel, R. A., *Modern Communications Jamming Principles and Techniques*, 2nd ed., Norwood, MA: Artech House, 2011.

[2] McGuffin, B. F., "Distributed Jammer Performance in Rayleigh Fading," *Proceedings IEEE MILCOM* 2002, Vol. 1, 7–10 October 2002, pp. 669–674.

[3] McGuffin, B. F., "Jammed FH-FSK Performance in Rayleigh and Nakagami-M Fading," *Proceedings IEEE MILCOM* 2003, Vol. 2, 13–16 October 2003, pp. 1077–1082.

[4] Al Hussaini, E. K., "Effects of Nakagami Fading on Antijam Performance Requirements," *Electronic Letters*, Vol. 24, No. 4, February 18, 1988, pp. 208–210.

[5] *Reference Data for Radio Engineers*, New York: Howard W. Sams & Co, Inc, Chapter 29.

[6] Basar, T., "The Gaussian Test Channel with an Intelligent Jammer," *IEEE Transactions on Information Theory*, Vol. IT-29, No. 1, January 1983, pp. 152–157.

[7] Poisel, R. A., *Introduction to Communication Electronic Warfare Systems*, 2nd ed., Norwood, MA: Artech House, 2008, Chapters 10 and 11.

[8] Poisel, R. A., *Modern Communications Jamming Principles and Techniques*, 2nd ed., Norwood, MA: Artech House, 2011, Ch. 16.

Appendix A

Simulated Networks

A.1 Introduction

The specific nets used in the simulations described in the text are shown in this appendix. In those cases where there were only these 12 nets, no other nets were simulated. In those cases when more than 12 nets were simulated, the additional nets were added to the region shown at random, but at tactically significant ranges. All of the numbers in these charts are in kilometers. For readers unfamiliar with the military symbology in the figures, the symbols are defined in Figure A.1.

For example, Net 1 consists of a corps headquarters (×××) as the *network control station* (NCS). The network is comprised of this NCS and three division headquarters (××) along with one independent brigade headquarters (×).

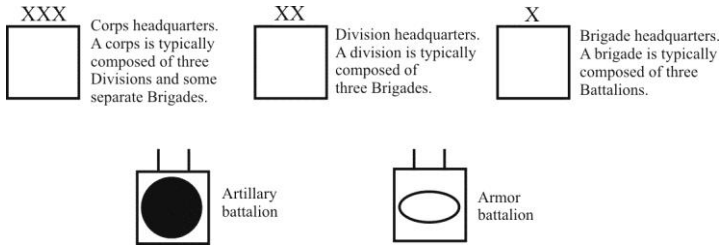
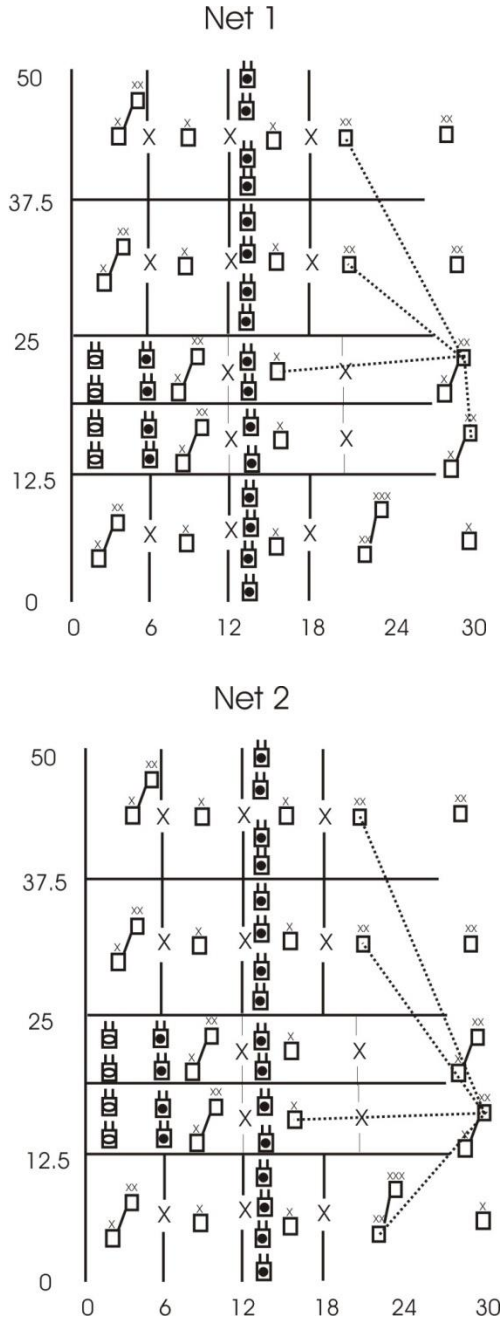
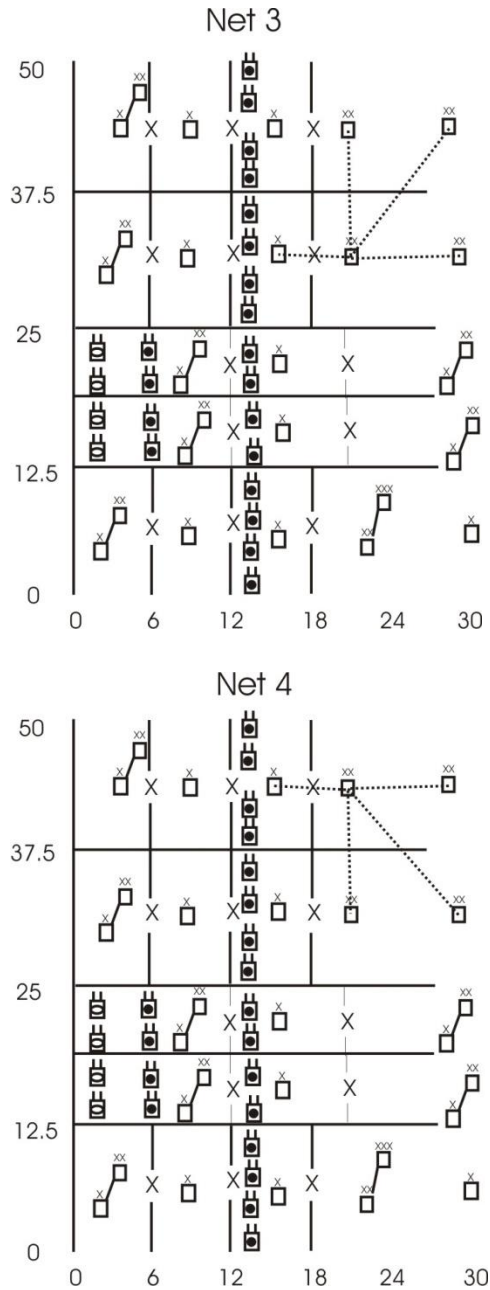
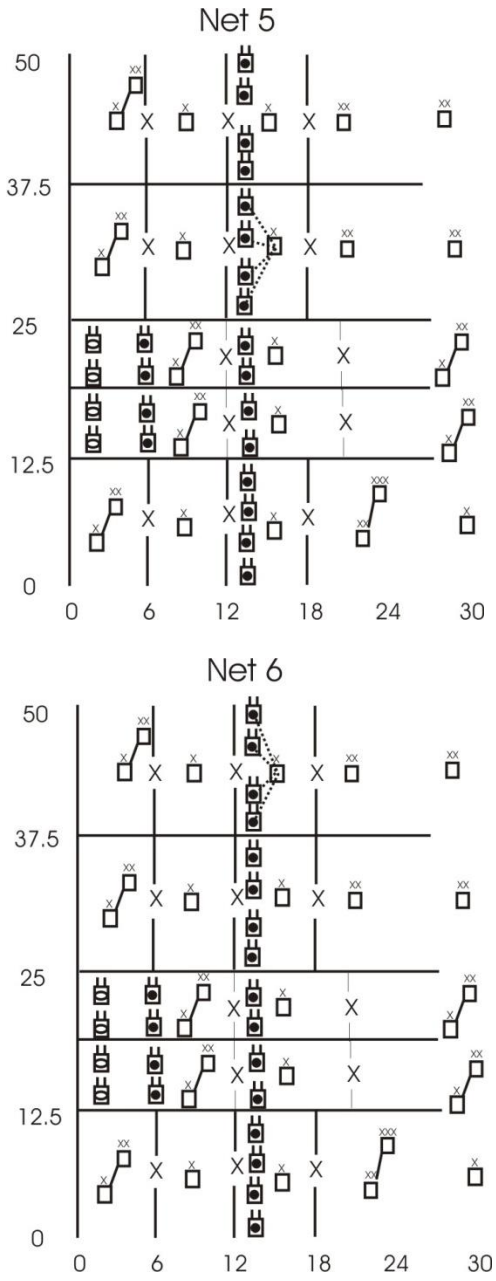
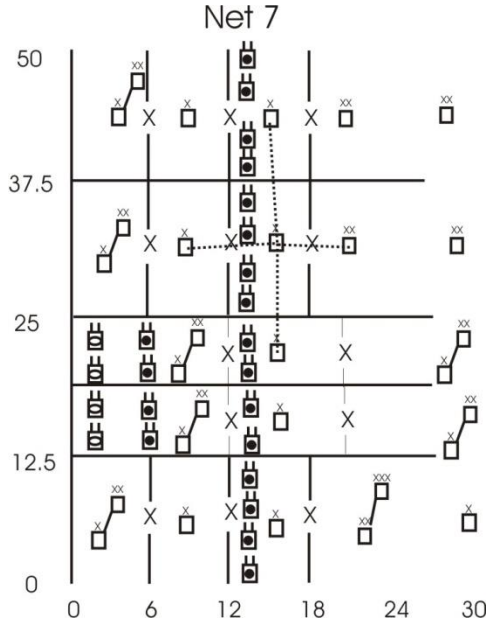


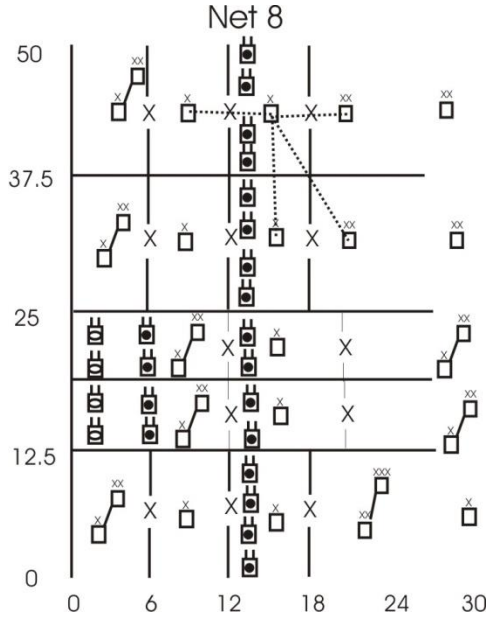
Figure A.1 Symbology used in the network diagrams.

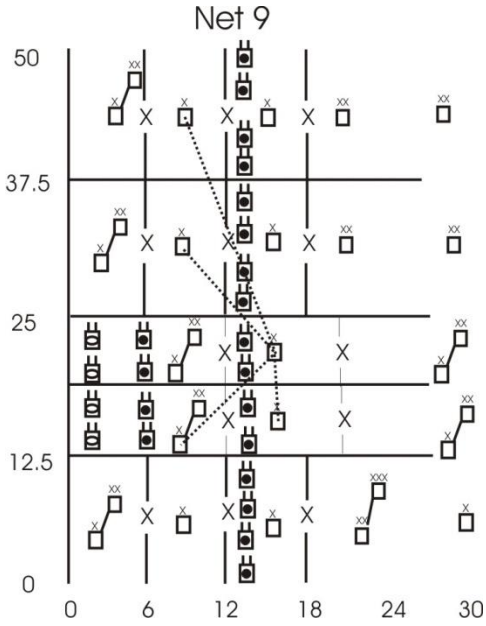

Net 1
50
37.5
25
12.5
0
0
6
12
18
24
30
Net 2
50
37.5
25
12.5
0
0
6
12
18
24
30

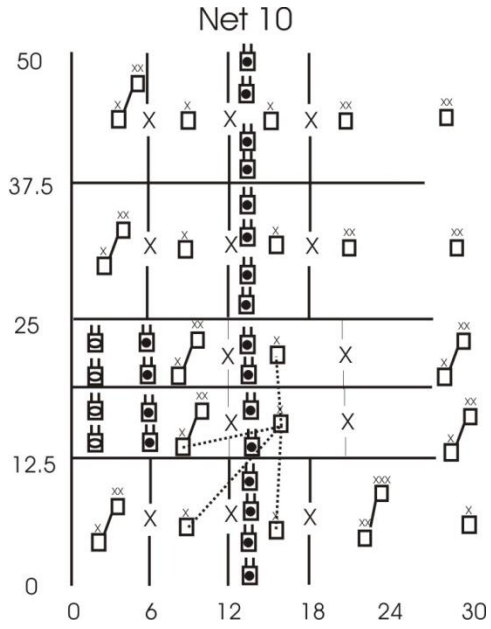

Net 3
50
37.5
25
12.5
0
0
6
12
18
24
30
Net 4
50
37.5
25
12.5
0
0
6
12
18
24
30

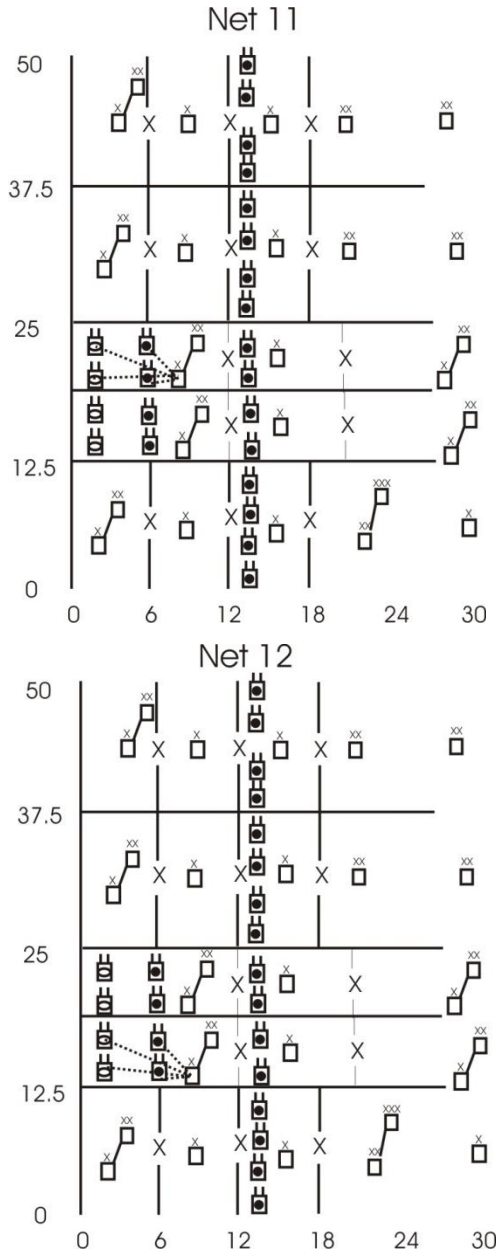

Net 5
50
37.5
25
12.5
0
0
6
12
18
24
30
Net 6
50
37.5
25
12.5
0
0
6
12
18
24
30


Net 7
50
37.5
25
12.5
0
0
6
12
18
24
30


Net 8
50
37.5
25
12.5
0
0
6
12
18
24
30


Net 9
50
37.5
25
12.5
0
0
6
12
18
24
30


Net 10
50
37.5
25
12.5
0
0
6
12
18
24
30


Net 11
50
37.5
25
12.5
0
0
6
12
18
24
30
Net 12
50
37.5
25
12.5
0
0
6
12
18
24
30

Acronyms

μs	microsecond
ABR	associatively based routing
ACK	acknowledgement
ADA	air defense artillery
ALO	air liaison officer
AM	amplitude modulation
AO	area of operations
AODV	ad hoc on-demand distance vector
AOR	area of responsibility
APC	armored personnel carrier
ARQ	automatic repeat request
ASC	all source correlation
AVWTC	arbitrarily varying wiretap channel
AWGN	additive white Gaussian noise
BBN	broadband noise
BC	broadcast channel
BCT	brigade combat team
BDA	battle damage assessment
BEI	background environment information
BER	bit error rate
BFSK	binary frequency shift keying
BSC	binary symmetric channel
C2	command and control
C2W	command and control warfare
CAB	combined arms battalion
CAN	computer network attack
CCRP	Command and Control Research Program
CDMA	code division multiple access
CEOI	communication electronic operational instructions
CEP	circular error probable
CEW	communications electronic warfare
CGSR	cluster-head gateway switch routing
CI	counterintelligence

CMC	convolve multiply convolve
CMOS	complementary metal oxide
CND	computer network defense
CNE	computer network exploitation
CNO	computer network operations
CNR	combat net radio
COMINT	communications intelligence
CONOP	concept of operation
COP	common operating picture
CP	command post
CRL	certificate revocation list
CSI	channel state information
D&D	disruption and destruction
D&M	deception and mimicry
DARPA	Defense Advanced Research Project Agency
dB	decibel
dBi	decibels relative to isotropic
DBK	dominate battlespace knowledge
dBm	decibels relative to a milliwatt
dBW	decibels relative to 1 watt
DDL	dispersive delay line
DEAD	destruction of enemy air defense
DFT	discrete Fourier transform
DIA	Defense Intelligence Agency
DM	decision maker
DMC	discrete memoryless channel
DoD	Department of Defense
DoI	denial of information
DoS	denial of service
DOTMLPF	doctrine, organization, training, materiel, leadership and education, personnel, and facilities
DSDV	destination-sequenced distance-vector
DSR	dynamic source routing
DSSS	direct sequence spread spectrum
EA	electronic attack
EAB	echelons above brigade
EBO	effects-based operations
EC/S	equipment characteristics/space
EEP	elliptical error probable
ELINT	electronic intelligence
EM	electromagnetic
EMP	electromagnetic pulse

EOB	electronic order of battle
EP	electronic protect
ERP	effective radiated power
ES	electronic support
EW	electronic warfare
EWO	electronic warfare officer
EWTA	electronic warfare target analysis
FAAD	forward area air defense
FAC	facility
FDC	fire direction center
FDOA	frequency difference of arrival
FFH	fast frequency hopping
FH	frequency hopping
FIST	fire support team
FM	frequency modulation
FO	forward observer
FRRS	frequency resource record system
FSCOORD	fire support coordinator
FSK	frequency shift keying
G-2	assistant chief of staff, intelligence
G-3	assistant chief of staff, operations
G-5	assistant chief of staff, plans
G-6	assistant chief of staff, signal
G-7	assistant chief of staff, information management
GBS	ground-based sensor
GHz	gigahertz
GIG	global information grid
GP	gradient projection
GPS	global positioning system
GSVD	generalized singular value decomposition
HPA	high power amplifier
HUMINT	human intelligence
Hz	hertz
I&W	indications and warning
IA	information assurance
IBW	instantaneous bandwidth
IED	improvised explosive devices
IO	information operations
IPB	intelligence preparation of the battlefield
IT	information technology
IW	information warfare
JNR	jammer-to-noise ratio

JSC	Joint Spectrum Center
JSR	jam-to-signal ratio
JSTARS	joint surveillance target acquisition, reconnaissance system
JTAC	joint terminal attack controller
JTF	joint task force
KDC	key distribution center
kHz	kilohertz
km	kilometers
kW	kilowatt
LAN	local area network
LER	loss exchange ratio
LMR	lightweight mobile routing
LNO	liaison officer
LOB	line of bearing
LOP	line of position
LOS	line of sight
LPD	low probability of detection
LPE	low probability of exploitation
LPI	low probability of intercept
MAC/MIC	message authentication and integrity codes
MANET	mobile ad hoc network
MASINT	measures and signatures intelligence
MCM	multiply convolve multiply
MD5	message digest 5
MFSK	multiple frequency shift keying
MHz	megahertz
MIMO	multiple input multiple output
MIMOME	multiple input, multiple output, multiple intercept
MISOSE	multiple input, single output, single intercept
MOE	measure of effectiveness
NCO	network centric operations
NCS	network control station
NCW	network centric warfare
NE	Nash equilibrium
NEA	northeast Asia
NLOS	nonline of sight
NPDU	network protocol data units
ns	nanosecond
OODA	observe, orient, decision, act
OPFAC	operational facility
OPSEC	operational security

OS	outstation
OSI	International Standards Organization
OTM	on the move
OV	operational view
PA	power amplifier
PBN	partial-band noise
PCM	pulse code modulation
pdf	probability density function
PF	position fix
PHY	physical
PIR	priority intelligence requirement
PKE	public key encryption
PKI	public key infrastructure
PPI	pulse position indicator
PPM	pulse position modulation
PPP	point to point
psd	power spectral density
PSTN	public switched telephone network
QAM	quadrature amplitude modulation
QoS	quality of service
RAM	random access memory
RAP	radio access point
RCU	receiver control unit
RF	radio frequency
RMS	root-mean-squared
RREP	route reply
RREQ	route request
RSTA	reconnaissance, surveillance, and target acquisition
S-2	intelligence staff officer
S-3	operations staff officer
S-6	signal staff officer
S-7	information engagement staff officer
SA	situation assessment
SAW	surface acoustic wave
SCS	spectrum certification system
SDU	spectral display unit
SEAD	suppression of enemy air defense
SEI	specific emitter identification
SFH	slow frequency hopping
SIGINT	signals intelligence
SIMOSE	single input, multiple output, single intercept
SNR	signal to noise ratio

SOI	signal of interest
SPE	subgame-perfect equilibria
SSR	signal stability routing
STFT	short-term Fourier transform
STP	shielded twisted pair
SV	system view
SWAP	space, weight, and power
T/R	transmit/receive
TACDB	tactical database
TCP/IP	Transmission Control Protocol/Internet Protocol
TDOA	time difference of arrival
TGT	targeting
TOA	time of arrival
TORA	temporally ordered routing algorithm
UAS	unattended aerial system
UAV	unattended aerial vehicle
USN	U.S. Navy
UTP	unshielded twisted pair
WAN	wide area network
WRP	wireless routing protocol
XOR	exclusive OR

About the Author

Richard A. Poisel received a B.S. in electrical engineering from the Milwaukee School of Engineering in 1969 and an M.S. in the same discipline from Purdue University in 1971. He spent three years in the military service from 1971 to 1973. After his service he attended the University of Wisconsin, where he received a Ph.D. in electrical and computer engineering in 1977. From 1977 to 2004, he was with the same government organization, which has had several different names and is currently known as the U.S. Army Research, Development, and Engineering Command, Intelligence and Information Warfare Laboratory. During the 1993–1994 academic year, Dr. Poisel attended the MIT Sloan School of Management as a Sloan Fellow, receiving an M.B.A. Initially a research engineer, Dr. Poisel eventually rose to the role of the director of the laboratory on an acting basis from 1997 to 1999. He was appointed chief scientist in 1999 and was relocated to the Army's Intelligence Center at Ft. Huachuca, Arizona, where he served as a technical advisor to the command group. Retiring from government service in 2004, he served as a Senior Engineering Fellow at Raytheon Missile Systems in Tucson, Arizona, from 2004 to 2011. He is currently a consultant on engineering for electronic warfare applications.

Dr. Poisel holds several patents and is the author of the following books, all published by Artech House: *Introduction to Communications Electronic Warfare Systems*, Second Edition, (2008); *Modern Communications Jamming Principles and Techniques*, Second Edition, (2011); *Target Acquisition in Communications Electronic Warfare Systems*, (2004); *Foundations of Communications Electronic Warfare*, (2008); *Electronic Warfare Target Location Methods*, Second Edition (2012); and *Antenna Systems and Electronic Warfare Applications* (2012).

Index