Fundamentals of Arabic cryptology and covert communication networks

Adam Miles
Santa Clara University

Follow this and additional works at: http://scholarcommons.scu.edu/amth_mstr

Recommended Citation
SANTA CLARA UNIVERSITY
DEPARTMENT OF APPLIED MATHEMATICS

Date: July 13, 2012

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY
SUPERVISION BY

Adam Miles

ENTITLED

Fundamentals of Arabic Cryptology and
Covert Communication Networks

 أساسيات التشفير العربي و شبكات الاتصالات السرية

BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF

MASTER OF SCIENCE IN APPLIED MATHEMATICS

[Signatures]
Thesis Advisor
Department Chair
Fundamentals of Arabic Cryptology and Covert Communication Networks

 أساسيات التشفير العربي و نetworks السرية

by

Adam Miles

Submitted in partial fulfillment of the requirements
for the degree of
Master of Science in Applied Mathematics
School of Engineering
Santa Clara University

Santa Clara, California
July 13, 2012
Dedication

I would like to dedicate this thesis to my family and friends, who have always supported my professional endeavors. Specifically, I would like to recognize my friend Saad and his family in Rabat, Morocco, who took me into their home and who taught me a great deal about Arab Muslim culture and the Moroccan Arabic dialect.
Acknowledgments

First, I would like to acknowledge Stephen A. Chiappari, Ph.D., who has provided me with support and encouragement as an advisor and department chair. In particular, I appreciate his guidance in helping me accurately define mathematical and computer programming concepts related to covert networks. I would also like to acknowledge Ahmed F. Yassin, Ph.D., who served as my Spoken Iraqi Dialect instructor at Georgetown University and who helped verify the Modern Standard Arabic grammar, spelling, and translation. I would also like to thank my good friend and fellow graduate computer engineer Nick Treat, M.S., for his assistance in running the steganography computer programs and for evaluating the correctness of the output. Finally, I would like to thank my sister Jessica Miles, M.S., for her assistance in editing and formatting the final document.
Fundamentals of Arabic Cryptology and Covert Communication Networks

 أساسيات التشفير العربي و شبكات الاتصالات السرية

Adam Miles

Department of Applied Mathematics
Santa Clara University
July 13, 2012

ABSTRACT

The need for accurate intelligence concerning possible terrorist attacks, spies, and other hostile military type actions, whether it be at home or abroad, remains of critical importance to the U.S. Intelligence Community. In this context, this paper focuses directly on the foundational aspects of covert communication networks and how they may be formed by groups or organizations such as al-Qaeda, jihadists, insurgents, etc. using spy tradecraft, cryptography and the language of Modern Standard Arabic.

The paper itself is divided into two parts, one that focuses upon communicating covertly through methods without the use of electronics, and the other with electronics. The reason for this division is to highlight the fact that a group of individuals or an organization may purposefully choose to use one style of covert communication over the other in order to achieve a particular objective. For instance, electronic surveillance may be a concern. The examples that will be given in each part are centered around
theoretical covert network operations conducted in the countries of Afghanistan, Pakistan, Iraq, Saudi Arabia, and Yemen. Accompanying these examples, a useful system for network organization is presented in order to help the reader, in this case U.S. intelligence officials, to analyze the flow of covert communications in general. Not only does this system demonstrate the differences between electronic and non-electronic networks, but it also shows that with proper planning and training, the groups formerly mentioned could possibly create intensely abstract and complicated networks in order to adapt to certain battlefield conditions. With this said, it is the author’s suggestion to the reader to think ‘outside of the box’ when approaching the paper and to know that everything may not always be as it appears.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEDICATION</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>x</td>
</tr>
<tr>
<td>DISCLAIMER</td>
<td>xii</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td><strong>PART I. COVERT NON-ELECTRONIC NETWORKS AND ARABIC CRYPTOLOGY</strong></td>
<td>2</td>
</tr>
<tr>
<td>APPLICATIONS</td>
<td>2</td>
</tr>
<tr>
<td>Chapter 1: Covert Non-Electronic Networks (CNN)</td>
<td>6</td>
</tr>
<tr>
<td>1.1 Basic Graph Theory</td>
<td>6</td>
</tr>
<tr>
<td>1.2 Covert Networks and Data Fields</td>
<td>9</td>
</tr>
<tr>
<td>1.3 Covert Non-Electronic Communication Methods</td>
<td>12</td>
</tr>
<tr>
<td>1.4 Examples of CNN</td>
<td>14</td>
</tr>
<tr>
<td>Courier: The Messenger Boy</td>
<td>14</td>
</tr>
<tr>
<td>Courier: The Hotel Manager</td>
<td>15</td>
</tr>
<tr>
<td>Contact: Downtown Henna</td>
<td>16</td>
</tr>
<tr>
<td>Contact: The Doorman</td>
<td>16</td>
</tr>
<tr>
<td>Dead Drop: The Purple Plant</td>
<td>17</td>
</tr>
<tr>
<td>Live Drop: The Suitcase</td>
<td>17</td>
</tr>
<tr>
<td>Timed Drop: The Cave</td>
<td>18</td>
</tr>
</tbody>
</table>
Chapter 2: Covert Non-Electronic Networks Combined with Cryptography

[CNN(C)] .........................................................................................19

2.1 An Example of a CNN(C) ..............................................................20

Shift Cipher: Blackmail .................................................................20

2.2 Complete Communication for CNN(C) ........................................23

2.3 Modern Standard Arabic Language Statistics ...............................25

2.4 Additional Examples of CNN(C) ..................................................29

Affine Cipher: The Foreign Students ..............................................29

Playfair Cipher: The Pakistani Government ....................................32

The Vigenère Cipher: The Ghost ....................................................36

The Vigenère Cipher: Cryptanalysis ..............................................40

Kasiski’s Method ..........................................................................41

The Frequency Distribution of Single Ciphertext Characters ..........42

The Measure of Roughness ..........................................................42

The Index of Coincidence ..............................................................44

The Monoalphabetic Indices of Coincidence for

a Set Value of t ............................................................................46

Applying Cryptanalysis to The Ghost ..........................................46

The Expected Value of the Index of Coincidence ............................53

2.5 Hidden Objects ..........................................................................53

Hidden Objects: Flashlight ..........................................................53
PART II. COVERT ELECTRONIC NETWORKS AND ARABIC CRYPTOLOGY

APPLICATIONS

Chapter 3: Covert Electronic Networks (CEN)

3.1 Covert Electronic Communication Methods

Chapter 4: Covert Electronic Networks Combined with Cryptography [CEN(C)]

4.1 Real Life Examples of al-Qaeda Encryption Software: Mujahideen

Secrets

Global Islamic Media Front (GIMF)

Islamic Faithful/Devoted Network or Al-Ekhlaas Islamic Network

4.2 Examples of CEN(C)

Hill Cipher: Discussion

Transposition Cipher: The Yemeni Camp

One-Time Pad: A Sensitive Conversation

Audio Steganography: The Call to Prayer

RSA: The Harbor

Applying Cryptanalysis to the Harbor

Part II Summary

REFERENCES

APPENDIX

A. Seberry Correspondence
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Two Nodes with an Undirected Edge</td>
</tr>
<tr>
<td>2</td>
<td>Two Nodes with a Directed Edge</td>
</tr>
<tr>
<td>3</td>
<td>An Undirected Network</td>
</tr>
<tr>
<td>4</td>
<td>A Directed Network</td>
</tr>
<tr>
<td>5</td>
<td>A Mixed Network</td>
</tr>
<tr>
<td>6</td>
<td>The Data Field of Spy Tradecraft</td>
</tr>
<tr>
<td>7</td>
<td>The Messenger Boy Network</td>
</tr>
<tr>
<td>8</td>
<td>Modern Standard Arabic Alphabet (Labeled with Integers 0 - 27)</td>
</tr>
<tr>
<td>9</td>
<td>MSA Special Characters/Modified Letters (Labeled with Integers 28 - 35)</td>
</tr>
<tr>
<td>10</td>
<td>Shift Cipher Encryption</td>
</tr>
<tr>
<td>11</td>
<td>Letter Frequencies and Relative Frequencies of the Modern Standard Arabic Alphabet and its Special Characters</td>
</tr>
<tr>
<td>12</td>
<td>MSA Character Probability Mass Function</td>
</tr>
<tr>
<td>13</td>
<td>Affine Cipher Encryption</td>
</tr>
<tr>
<td>14</td>
<td>Playfair Cipher Encryption</td>
</tr>
<tr>
<td>15</td>
<td>The Pakistani Government Network</td>
</tr>
<tr>
<td>16</td>
<td>Vigenère Cipher Encryption</td>
</tr>
<tr>
<td>17</td>
<td>Ciphertext Frequency Distribution</td>
</tr>
<tr>
<td>18</td>
<td>Monoalphabetic Indices of Coincidence</td>
</tr>
<tr>
<td>19</td>
<td>First Step of Hiding Process</td>
</tr>
</tbody>
</table>

Page

7 7 8 8 8 11 15 21 21 22 27 28 30 33 35 38 46 49 53
Figure 20 Second Step of Hiding Process...............................................................54
Figure 21 Third Step of Hiding Process...............................................................54
Figure 22 Hill Cipher Encryption.......................................................................67
Figure 23 The Discussion Network....................................................................68
Figure 24 Transposition Cipher Encryption.......................................................71
Figure 25 Random Binary Representations.......................................................73
Figure 26 One-Time Pad Encryption.................................................................75
Figure 27 Hex Editor: Song Comparison...........................................................79
Figure 28 RSA Encryption..................................................................................86
Disclaimer

The examples provided in this thesis are entirely fictitious and do not contain real people, events, or cities. This thesis also does not encourage the formation of covert network ideas or elements or conducting Arabic cryptography/steganography. However, to maintain the life-like quality of the examples, the situations depicted will take place in the countries of Afghanistan, Pakistan, Iraq, Saudi Arabia, and Yemen. This is more of an attempt by the author to stay true to the terrain, infrastructure, languages, and cultures that are exhibited in these countries. These five countries have been chosen since they are at the top of the U.S. government’s list of countries that have played a key role in the War on Terrorism and the war in Iraq. Most importantly, the author would like to express his support to the U.S. government and other countries in their endeavors to fight terrorism by showing how terrorists could possibly create covert networks in an effort to deceive their enemies.
**Introduction**

Effective intelligence depends heavily upon the study of how people interact with one another and how well one can predict future events resulting from those interactions. Of course, every interaction requires some degree of communication, whether it be in the form of a phone call, courier system, email, cryptography, or just plain silence in order to achieve a certain goal. In the context of intelligence, the purpose of this paper is to provide a general system for the analysis and categorization of covert communication networks by presenting the reader with the basic concepts of graph theory, simple algorithms or procedures for cryptography and steganography, and hypothetical scenarios of people attempting to communicate secretly in real time. All concepts and examples presented in this paper are meant to be seen through the eyes of United States (U.S.) intelligence officials, and for this reason, a few sections regarding cryptanalysis are included.

The examples presented in the first part of this paper demonstrate fictional covert non-electronic cross-border operations conducted by groups such as the Taliban, al-Qaeda, etc. between the countries of Afghanistan and Pakistan. That is, communication does not rely upon electronic devices such as cell-phones, walkie-talkies, and computers. In contrast, the second part presents examples of long distance covert electronic networks created by al-Qaeda and other militants in the countries of Iraq, Saudi Arabia, and Yemen. All examples incorporate cultural and geographical attributes pertaining to these selected countries and/or the language of Modern Standard Arabic.
Part I

Covert Non-Electronic Networks and Arabic Cryptology Applications

The Pentagon’s creation of an ultra-secret surveillance organization known as the Intelligence Support Activity or, more simply, ‘The Activity’… [and] its pursuit of enemy big fish failed repeatedly in Somalia, Bosnia, and the pursuit of bin Laden as the targets learned to avoid using vulnerable cellphones, relying on human couriers instead. (Geraghty, 2010, pp. 52 & 66-67)

Imagine that an American spy is living in Iran and that he or she has been tasked with the assignment to gather intelligence on Iran’s nuclear and/or long-range missile research and development programs. Furthermore, suppose that the American spy is posing as a Swiss banker - since he or she is in deep cover and the Swiss are neutral - and that he or she has been given a list of the names of some Iranians who are known to have ties to these programs. The spy has also been told that these Iranians are not particularly fond of the Iranian regime and that they might be willing to give out information for a price or for a chance to leave Iran with their families. Bottom line is, the spy knows that he or she wants to contact these individuals, but he or she also knows that doing so is extremely dangerous. That is, the American spy knows that the Iranian regime is intent on finding traitors or Western spies by monitoring electronic communications and through intimidation from the secret police. Some of the questions the spy might ask are: How is he or she going to make contact with these people without getting caught? What are some of the different ways the spy can try to maneuver around the barrier of being
electronically monitored to gain the right information? What are some of the specific steps the spy can take to safeguard his or her communications, and possibly his or her life? All of these questions relate to the field of spy tradecraft, which includes the fundamental operative techniques used by spies to conduct successful intelligence operations (International Spy Museum, 2011).

In the context of spy tradecraft, this part will describe a variety of ways in which two or more people can secretly communicate with one another without using electronic devices such as cell phones, computers, two-way radios, etc. These covert non-electronic communication methods can include the use of couriers or messengers, specific courier routes and destinations, drivers, dead drops, cryptography, and steganography, to name a few. (Note that the word covert in this instance denotes an activity that is “secret” to the controllers of communication, but in fact may be exposed to outsiders.) These communication tactics can also rely entirely upon human physical encounters or they can depend upon a combination of both physical encounters and written cryptographic schemes. When covert non-electronic communication methods are systematically combined together, they ultimately form covert non-electronic networks (CNN) or covert non-electronic networks combined with cryptography (CNN(C)). The main advantage of such clandestine networks is that any information exchanged between participating parties is not subject to electronic interception as a result of poorly secured transmissions or open electronic channels. One may also benefit strategically by adopting these primitive types of networks when access to modern or sophisticated technology is limited or nonexistent. However, these communication networks are not
without flaw. For one thing, they can be slow - and in a world of spies where network speed might be the least of one’s worries, they can be infiltrated or observed by outsiders, corrupted by insiders, and potentially deadly for the players of the game.

We include some examples that demonstrate how one’s enemies, such as al-Qaeda, the Taliban, jihadists, and insurgents, could possibly construct CNN and CNN(C) in an attempt to adapt to certain battlefield conditions. These battlefield conditions primarily involve, but are not limited to, these groups’ inability to access or afford advanced technology, in addition to these groups’ awareness of U.S. military prowess concerning electronic surveillance. These chapters will also present how covert non-electronic methods can be manipulated, expanded, and creatively altered by other factors (e.g., multiple couriers or cryptosystems) to perform an objective that is specific to the particular covert network at hand. In the future, it is these structural modifications that will be of interest when categorizing more complicated CNN/CNN(C), performing network analysis, and/or conducting cryptanalysis. Going from the chapter of CNN to CNN(C), it will be shown why these particular groups would want to fundamentally modify or upgrade CNN with cryptography in order to make it harder for authorities to gain timely and significant intelligence.

It is important to mention at this point that intelligence officials may still have the opportunity to observe or intercept non-electronic enemy communications. For instance, officials may use wiretapping or eavesdropping, Predator drones, and/or on-the-ground human intelligence, to collect vital information from relevant non-electronic sources. However, these types of networks may also impose hardships on those conducting
surveillance. For example, authorities observing such networks generally must be in
closer proximity to the communications, thereby increasing costs, hazards, and
manpower, associated with surveillance. The use of these types of communications by an
enemy also reduces the number of ways intelligence personnel may gather information
across vulnerable and easy-access places like the Internet.
Chapter 1: Covert Non-Electronic Networks (CNN)

What we’re dealing with here is potentially a global conflagration that requires constant diligence in order to suppress. Now you see because our enemy has realized that they are fighting guys from the future. Now, it is brilliant as it is infuriating. If you live like it’s the past, and you behave like it’s the past, then guys from the future find it very hard to see you. If you throw away your cell phone, shut down your e-mail, pass all your instructions face-to-face, hand-to-hand, turn your back on technology, [you can] just disappear into the crowd. So what [has] changed is that our allegedly unsophisticated enemy, has cotten [sic] on to the factually unsophisticated truth, we’re an easy target.

(Scott, 2008)

1.1 Basic graph theory.

According to Chen (1997) in basic graph theory, an undirected graph or network $G_u(V, E_u)$, is formed by combining a set of elements $V$, called nodes, with a set of elements $E_u$ in $V$, called undirected edges. The term graph is synonymous with network. Visually, a node is represented by a point or junction, while an undirected edge is represented by a simple undirected line or branch (a line without an arrow). In practice, the nodes $i$ and $j$ are usually connected by an undirected edge $(i, j)$ or $(j, i)$; here, nodes $i$ and $j$ are called the endpoints of the undirected edge $(i, j)$ or $(j, i)$. Moreover, the undirected edge, $(i, j)$, is an unordered pair and it is often denoted by a letter such as $e$. Note an edge may be represented by a straight line or a curved line. This option may or
may not hold meaning to network diagramming and must be considered by the analyst of
the network.

*Figure 1.* Two nodes with an undirected edge.

\[ i \quad e \quad j \]

Similarly, a directed graph or network \( G_d(V, E_d) \), is formed by combining a set of
elements \( V \), called nodes, with a set of elements \( E_d \) in \( V \), called directed edges. Visually,
the nodes \( i \) and \( j \) are usually connected by a directed line (a line with an arrow) or
directed edge \( (i, j) \); here, the node \( i \) is called the initial node and the node \( j \) is called the
terminal node. Again, the nodes \( i \) and \( j \) are called the endpoints of the directed edge \( (i, j) \).
Moreover, the directed edge, \( (i, j) \), is an *ordered pair* and it is often denoted by a letter
such as \( e \). This edge may also be represented by a straight or curved line.

*Figure 2.* Two nodes with a directed edge.

\[ i \quad e \quad j \]

Finally, a mixed graph or network \( G_m(V, E_m) \) contains both undirected and
directed edges. Some examples of these three different types of networks are presented
below.
Figure 3. An undirected network.

Figure 4. A directed network.

Figure 5. A mixed network.
1.2 Covert networks and data fields.

In this paper, we define a *covert network* as an *undirected network* \( G_u(V, E_u) \), or more simply, \( G(V, E) \). For simplicity, we define nodes as physical locations and edges as “connections” between locations. For example, the locations of a house and a post-office in a particular city may be considered nodes. If a person living in that house receives a letter from the post-office, then there is a “connection” between the two nodes. We also say that each node contains zero or more **data fields**, which uniquely determine the properties of the node. Examples of data fields include: addresses, telephone numbers, persons of interest, descriptions of connections to other nodes, other data fields, etc.

These data fields will be helpful in terms of network organization and analysis since they can be easily managed and written in the form of code with a computer language such as C. This code could then be used for the development of a computer program and/or graphical interface. For instance in a hypothetical graphical interface setting, clicking the mouse on a particular node would reveal the data fields for edit. Newly edited node data could then be synthesized with other node data via another computer program to produce network data. It might also be possible to integrate the use of directed edges to show network flow for a particular data set such as money. To save time we have not included code, a graphical interface, or a computer program for this paper, but instead we provide the basic framework for general covert network organization and analysis.

A particular data field may encompass other data fields. For example, the data field of address includes the data fields of street location, zip code, country, etc. With this said, we say that the data field of spy tradecraft includes the data fields of: type of
espionage (e.g., military, corporate, etc.), covert communication methods, location (e.g., domestic or overseas), time period, etc. Of course, these spy tradecraft data fields contain other data fields. The data fields located under the data field of covert communication methods are the primary focus of this paper. Although we state these data fields explicitly, the node data fields of an efficient computer program and/or graphical interface should not be set in stone, and therefore should be left open to customization by the software developers, analysts, and computer programmers.
**Figure 6. The data field of spy tradecraft.**

<table>
<thead>
<tr>
<th>Spy tradecraft</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Type of espionage</td>
</tr>
<tr>
<td>• Military</td>
</tr>
<tr>
<td>• Intelligence service</td>
</tr>
<tr>
<td>• Corporate</td>
</tr>
<tr>
<td>• Covert communication methods</td>
</tr>
<tr>
<td>• Covert non-electronic communication</td>
</tr>
<tr>
<td>• Addresses or locations of interest, descriptions</td>
</tr>
<tr>
<td>• Persons of interest, couriers and contacts (time and place of contact,</td>
</tr>
<tr>
<td>profiles, courier routes, etc.)</td>
</tr>
<tr>
<td>• Drops (dead, live, timed), brush pass</td>
</tr>
<tr>
<td>• Connections to other locations</td>
</tr>
<tr>
<td>• Steganography (non-electronic)</td>
</tr>
<tr>
<td>• Cryptography (non-electronic)</td>
</tr>
<tr>
<td>• Cryptosystems</td>
</tr>
<tr>
<td>• Keys</td>
</tr>
<tr>
<td>• Ciphertexts</td>
</tr>
<tr>
<td>• Plaintexts</td>
</tr>
<tr>
<td>• Objects/items (may be hidden or used to hide other objects/messages)</td>
</tr>
<tr>
<td>• Written messages, verbal messages</td>
</tr>
<tr>
<td>• Covert electronic communication</td>
</tr>
<tr>
<td>• Addresses or locations of interest, descriptions</td>
</tr>
<tr>
<td>• Electronic couriers and contacts (time and place of contact, profiles,</td>
</tr>
<tr>
<td>electronic routes, etc.)</td>
</tr>
<tr>
<td>• Electronic drops (dead, live, timed)</td>
</tr>
<tr>
<td>• Connections to other locations</td>
</tr>
<tr>
<td>• Steganography (electronic)</td>
</tr>
<tr>
<td>• Cryptography (includes cryptosystems, keys, ciphertexts, plaintexts)</td>
</tr>
<tr>
<td>• General files (e.g., Microsoft Word or Excel)</td>
</tr>
<tr>
<td>• Picture/image files, audio files</td>
</tr>
<tr>
<td>• Messages (email, text message, verbal, etc.)</td>
</tr>
<tr>
<td>• The Internet, websites</td>
</tr>
<tr>
<td>• Hardware (computers, two-way radios, etc.)</td>
</tr>
<tr>
<td>• Hybrid communication (combination of non-electronic and electronic</td>
</tr>
<tr>
<td>communication)</td>
</tr>
<tr>
<td>• Location</td>
</tr>
<tr>
<td>• Time period</td>
</tr>
</tbody>
</table>

11
1.3 Covert non-electronic communication methods.

Some descriptions of covert non-electronic communication methods and how they relate to the field of spy tradecraft are provided below.

A contact is a clandestine meeting between two parties or more at a particular location. A covert non-electronic network can contain multiple contacts. According to the International Spy Museum (2011), a cut-out is a situation in which two parties agree to exchange information by using a specific depository or a third party contact/intermediary, under the condition that the two parties do not meet directly during the transaction. Examples of cut-outs include various types of drops as well as brush passes. A dead drop is a secret depository or physical location where two parties agree to exchange information (International Spy Museum, 2011). Both parties can choose to interact with the dead drop site directly or they can choose to send couriers. It is also possible to have multiple dead drop sites and couriers in order to relay a message. A live drop is similar to a dead drop in the sense that instead of a secret depository an actual person, such as a third party/intermediary, knowingly or unknowingly passes information between two parties (Spytrainer, 2009). According to Antonio and Jonna Mendez (2002), former C.I.A. operatives, a timed drop is a dead drop with a set time limit, in which the deposited information must be retrieved before the time limit expires. Finally, a brush pass is a timed and coordinated maneuver between two parties, in which one party quickly and subtly hands a message to the second party in an attempt to project a nonexistent relationship to outside observers; this maneuver is usually conducted in a crowd or public space (Spytrainer, 2009).
A **courier** or **messenger** is a trusted individual who is given a written message and delivery instructions by one party, and is expected to follow through by delivering the message to a second party or specified location. A covert non-electronic network can consist of multiple couriers, destinations, and vested parties. Remember, couriers can be identified and followed by outsiders, rebel against their handlers, or be an intelligence service in disguise.

According to Cole (2003), **steganography** is “on the simplest level...hidden writing, whether it consists of invisible ink on paper or copyright information hidden in an audio file” (p. 5). We have decided to take Cole’s definition of steganography and divide it into two subcategories, non-electronic and electronic. **Non-electronic steganography** is the non-electronic component of steganography, in which writing or information is hidden in a physical setting, such as a tattoo that can only be seen under black-light or a message hidden in an elaborate drawing (Cole, 2003). Non-electronic steganography may also involve transporting an item, which may seem trivial or meaningless to an interceptor but in fact has a predetermined meaning to the receiver. Conversely, **electronic steganography** is the electronic component of steganography, in which writing or information is hidden electronically or digitally in image files, audio files, word documents, and other types of mediums (Cole, 2003). *Only* non-electronic steganography is allowed in the definition of CNN. A hybrid type of network involving CNN and electronics, such as the physical transport of electronic and digital steganography on a USB flash drive, is a topic for further study and classification. The data field of cryptography will be discussed in the next chapter.
1.4 Examples of CNN.

**Courier: The messenger boy.**

Afghanistan - Suppose that Karim and his friends are renting an inexpensive farmhouse for their summer vacation on the outskirts of a small mountain town called Henna near the Afghanistan/Pakistan border.

For now, Karim tells his friends that he would like to send a secret written message to his cousin Sami, who he says lives in a small village in Pakistan just a few miles from the border. Because the farmhouse has no internet or land-line telephone, and none of them has brought a cell phone, computer, or two-radio, communication will be difficult, but not impossible. To overcome this problem, Karim suggests to his friends that they pay a young boy in town who knows the area well to deliver the written message. Karim also says that they must instruct the boy not to tell anyone about the message and not to reveal where the message came from if he is caught. Karim’s friends add that they could use the same boy for all future messages, or that they could find new and different couriers for each message; and if worse comes to worst, one of them could deliver the message.
Figure 7. The messenger boy network.

<table>
<thead>
<tr>
<th>Location 1 Data Fields</th>
<th>Location 2 Data Fields</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Location (Country - Afghanistan, town - Henna, description - near Afghanistan/Pakistan border, place - farmhouse)</td>
<td>• Location (Country - Pakistan, town/place - a small town in Pakistan, description - near Afghanistan/Pakistan border)</td>
</tr>
<tr>
<td>• Persons of interest (Karim and friends, messenger boy)</td>
<td>• Persons of interest (Karim’s cousin Sami)</td>
</tr>
<tr>
<td>• Covert non-electronic communication methods</td>
<td>• Covert non-electronic communication methods (receiver of Karim’s message from messenger boy)</td>
</tr>
<tr>
<td>(description - no electronics available, messenger or courier - messenger boy with written message is sent to Karim’s cousin Sami)</td>
<td></td>
</tr>
</tbody>
</table>

**Courier: The hotel manager.**

Pakistan - Meanwhile just across the border, Karim’s cousin Sami is busily preparing to receive and host a few out-of-town business associates at his family-owned hotel called the Fahkir Hotel. Before his business associates arrive, Sami asks his brother-in-law and hotel manager Abdul, to deliver a secret package to his cousin Karim, who he says is vacationing in a farmhouse not too far from the Hotel. Sami knows that by sending the package on foot, he does not risk the chance of the package’s information
being electronically intercepted. Sami also assumes that the chance of the package being physically intercepted is very low. Abdul, who is very trustworthy, quickly agrees to take the assignment because he will get a few days off from work.

Note on couriers: These two scenarios illustrate that couriers can be random strangers, friends, family members and/or business associates. They also show that in a covert non-electronic network, a particular courier can be used more than once, or that there can be a system of multiple couriers in place. In addition, these examples also demonstrate that the secrecy of the message depends upon the trustworthiness of the courier. This can be problematic for both sender and receiver.

**Contact: Downtown Henna.**

Afghanistan - Karim’s friends are leisurely sitting around the farmhouse one afternoon when Karim enters. Karim tells his friends that he is going to visit a man named Farooq in downtown Henna for the day to discuss some business affairs which might be of interest to them. When he returns, Karim says that they should all have a meeting to go over Farooq’s information.

**Contact: The doorman.**

Pakistan - The well-known doorman of the Fahkir Hotel, Mohammed, has just finished unloading the luggage of some new guests, when his boss Sami approaches him and pulls him aside. Sami asks Mohammed if he would take the afternoon off to go into the city of Dora (which is about five minutes away) to have a secret exchange of words with his friend Maria. Sami tells Mohammed that he is to return immediately once the conversation has ended so that he can hear about the meeting’s details. He also tells
Mohammed that once they have finished their conversation, he should go back into Dora to a place nicknamed the Stonehouse to deliver Maria’s information to a woman named Fatima.

**Dead drop: The purple plant.**

Afghanistan - While sitting on the porch of the farmhouse, Karim asks his friends if anyone has heard of a man named Alim, and if so, how does one go about contacting him. Karim’s friend, Basil, says that he has heard of this man and that he knows how to contact him. Basil tells Karim that he should go into the town of Henna to a place called the Rikshire café. Here, Basil instructs Karim, he must place his message in an envelope underneath the soil of a fake purple plant located in the back of the café near some tables. Basil says that Alim checks the pot for new envelopes every couple of days. Basil also adds that if Alim is told to respond to the message, Alim will place a similar looking envelope beneath the actual pot of the same plant in exactly three days.

**Live drop: The suitcase.**

Pakistan - One day Sami decides that he wants to reconnect with a man named Hanif and his associates. From past experience, Sami knows that he must place his message inside a suitcase full of clothes and bring it to a hotel called the Imad. There, Sami must pretend to be a guest checking out of the Hotel and he must ask the valet to hold onto his bag for his cousin to pick up in a few hours. Sami also knows that he must place two blue luggage tags on the suitcase so that one of Hanif’s men can recognize it when scouting out new deliveries. The valet does not know that he is passing information along to Hanif.
**Timed drop: The cave.**

Afghanistan - Before going to bed one night, Karim orders Basil to wake up early the following morning to deliver a small package. Furthermore, he instructs Basil to grab the map on the kitchen table and to use it to hike up to a particular cave in the nearby mountains. Karim says the ‘red x’ on the map signifies the cave’s location. Once he has arrived at the cave, Basil should place the small package underneath some loose rocks just outside the cave’s entrance. Karim also tells Basil that he must return to the cave at noon to see if the package has been taken. In the event that it’s still there, he should bring it back to the farmhouse.

**Brush pass: The man in the black thawb.**

Pakistan - Standing in the lobby of the Hotel, Sami calls for his doorman Mohammed to come in. Sami gives a letter to Mohammed and tells him that he must discreetly pass the letter over to a man wearing a black thawb (a black robe) in the suuq (market) downtown. Sami emphasizes to Mohammed that he cannot stop walking when he enters the suuq, and that the man will approach from the East in front of the Jala Patisserie at precisely three o’clock. Failure to do so in an inconspicuous manner, Sami adds, could be potentially dangerous.
Chapter 2: Covert Non-Electronic Networks Combined with Cryptography [CNN(C)]

The Vietnamese – contrary to the peasant image projected by some Western journalists – often proved themselves superior to the U.S. even on the arcane battlefield of electronic warfare. “The Vietcong cryptographers learned their lessons well. While throwing an electronic fishing net into the ether, they regularly reeled it back in bulging with American communications, but they seldom used radios themselves. While they listened to broadcasts from Hanoi on inexpensive transistor radios, they sent messages back to their commands with couriers, except in dire emergencies.” (Geraghty, 2010, p. 3)

Unlike ordinary CNN, Covert Non-Electronic Networks combined with Cryptography (CNN(C)) look to the field of cryptography to cleverly disguise all communications. For example, an organization might choose to encrypt the text of a written message by using a designated cryptosystem and key before he or she gives it to a courier. As for the cryptosystems involved, an organization could choose from a variety of classical cryptosystems such as a shift cipher, affine cipher, or simple Vigenère cipher. Although classical cryptosystems are not secure in general, they may be suitable for situations in which sensitive information expires quickly and/or persons exposed to the ciphertext are not familiar with cryptography.
2.1 An example of a CNN(C).

*Shift cipher: Blackmail.*

The shift cipher is a well-known cryptosystem within the field of cryptography, and the following example parallels the same shift cipher technique found in Trappe and Washington (2006). However, this particular shift cipher uses Modern Standard Arabic (MSA) as well as the author’s own set of numerical values.

Afghanistan - Karim and his friends are becoming increasingly paranoid that some of the locals they are paying to deliver messages are passing information to outside groups for cash rewards. This came to light when Basil said that one of the locals, who was not one of their messengers, approached him and asked him for money for something relating to one of their previous letters. Basil said that it essentially amounted to blackmail. This situation, Karim notes, poses a serious threat to the privacy of their communications, as well as to the internal workings of their organization. To overcome this problem, Karim suggests that they start encrypting their messages. Karim also points out that they can continue to employ the locals to deliver messages as long as their messages are encrypted.

Karim wishes to send a message to Sami before he becomes overwhelmed with hotel reservations during the high season. Karim’s message will notify Sami that their friend Ibrahim has returned to Iraq. Before Karim gives his message to one of the local messengers, he encrypts it using a shift cipher. (Note that the message will be written down on a piece of paper and placed in an envelope.)
For the shift cipher, Karim labels the letters of the Modern Standard Arabic (MSA) alphabet and its special characters or modified letters with the ordered set of integers \( \{0, 1, 2, 3, \ldots, 34, 35\} \), starting with \( 0 = \text{ﺍ} \). Special characters are taken from non-voweled MSA script, and we label them with integers as given in Figure 9 below. Since the MSA alphabet is read right to left, the integers will also go right to left. The MSA alphabet can be permuted, but for now it will be left in its original form.

Analogously, the English alphabet can be labeled with the ordered set of integers \( \{0, 1, 2, 3\ldots, 25\} \), starting with \( A = 0 \). Adding punctuation to the character set only containing the English alphabet, one might label the question mark character ‘ ‘ as integer 26, thereby augmenting the set of considered integers to \( \{0, 1, 2, 3\ldots, 26\} \). For simplicity, MSA vowelization and punctuation marks are not considered.

*Figure 8. Modern Standard Arabic alphabet (labeled with integers 0 - 27).*

| ا | ب | ت | ث | ج | ح | خ | ض | ط | ظ | ف | ق | ك | ل | م | ن | ه | و | ي |
| 27 | 26 | 25 | 24 | 23 | 22 | 21 | 20 | 19 | 18 | 17 | 16 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 |

*Figure 9. MSA Special characters/modified letters (labeled with integers 28 - 35).*

<table>
<thead>
<tr>
<th>ء</th>
<th>ئ</th>
<th>ش</th>
<th>ص</th>
<th>ث</th>
<th>م</th>
<th>ر</th>
<th>إ</th>
<th>أ</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>34</td>
<td>33</td>
<td>32</td>
<td>31</td>
<td>30</td>
<td>29</td>
<td>28</td>
<td></td>
</tr>
</tbody>
</table>

Karim’s message \( m \) (or plaintext) to Sami is,

راجع سليم إلى العراق
message translation:

Ibrahim returned to Iraq.

For the shift cipher, Karim uses the encryption formula,

\[ f(x) \equiv (x + k) \mod 36 \]

where \( 0 \leq k \leq 35 \). The integer \( k \) is called the key.

Karim selects a secret integer \( k = 6 \) as the key, so \( f(x) \equiv (x + 6) \mod 36 \).

Figure 10. Shift cipher encryption.

| ٣٣ ٢٢ ٢٩ ٢٣ ٢٧ ٢٢ ١١ ١٧ ٤ ٩ | plaintext |
| ٣ ٢٨ ٣٥ ٢٩ ٣٣ ٢٨ ١٧ ٢٣ ١٠ ١٥ | f(x) |
| ﺛ ﺧ ﺖ ﺦ ﻎ ﻉ ﻊ ﺖ ﺧ ﺖ | ciphertext |
| ٢٠ ٠ ٩ ١٧ ٢٢ ٠ | plaintext |
| ٢٦ ٦ ١٥ ٢٣ ٢٨ ٦ | f(x) |
| ﺛ ﺥ ﺖ ﺖ ﺖ ﺖ | ciphertext |

Karim’s ciphertext and encrypted message is,

طزمعئىإأكختمخنو

Note: The encrypted message in connected letter form is طزمعئىإأكختمخنو, while the same encrypted message in disconnected letter form is طزمعئىإأكختمخنو. These versions are identical except that there are spaces between the letters in the disconnected form. Letters with naturally occurring spaces in the connected letter form cannot be joined together, since they are not defined in MSA. These extra spaces, whether they occur in the connected or disconnected forms, are ignored for simplicity.
throughout the paper during both the encryption and decryption processes. *Sentence form* is identical to the connected and disconnected forms, except that sentence form combines correctly ordered letter combinations to form words. As another example, take the plaintext sentence form message above and give all 3 forms.

*sentence form*: رجع سليم إلى العراق

*connected letter form*: رجع سليم إلى العراق

*disconnected letter form*: رجع سليم إلى العراق

Of course, ciphertext sentence form is probably meaningless and nonexistent as the above example demonstrates (see طمزعمإبأناشامطخو).

Karim puts his message in an envelope and gives it to the messenger.

decryption formula derivation:

\[
\begin{align*}
f(x) &\equiv (x + 6) \mod 36 \\
f(x) - 6 &\equiv x \mod 36 \\
x &\equiv (f(x) - 6) \mod 36
\end{align*}
\]

2.2 Complete communication for CNN(C).

An important question is, how will Sami know how to decrypt Karim’s message m using the formula \(x \equiv (f(x) - k) \mod 36\) and the key \(k = 6\)? With this question in mind, if Sami later receives only the formula from Karim (delivered perhaps by a different courier through a separate message), he will be able to search through all the possible keys to obtain Karim’s original message, though this may take a while. If on the other hand, Sami receives both the formula and the key (again delivered by a separate courier), then decryption will not be difficult. And if Sami just receives the key, then he will not
know what cryptosystem is being used for the ciphertext, which will make decryption somewhat difficult, especially if the ciphertext is short and/or a modern day cryptosystem is involved. This is where the art of ciphertext/cryptosystem sharing and key passing techniques comes into play. The term art is used here because there are an infinite number of creative ways in which one can covertly deliver ciphers, cryptosystems and keys to a receiving party. The term complete communication in a CNN(C) is used to describe a situation, in which all corresponding ciphertext(s), cryptosystem(s) and/or key(s) for a given plaintext are delivered on time to their correct destinations, under the condition that the delivery of these items to their destinations is synchronized in a way by the controllers of the network to allow for the successful decryption of the ciphertext(s).

For this reason, we divide the data field of cryptography into four new fields: keys, plaintexts, ciphertexts, and cryptosystems.

Ciphertext/cryptosystem passing or sharing is a method in which one delivers a ciphertext/cryptosystem to another party through the use of couriers, contacts, cell phones, computers, etc. Like ciphertext/cryptosystem sharing, key passing or sharing is a method in which one delivers the keys of a ciphertext/cryptosystem to another party using one of the methods mentioned above. For now, we will only focus on ciphertext and key passing techniques in a non-electronic setting.

Ciphertext sharing and key passing are equally important for the receiver, since they depend upon one another. For example, a ciphertext without a key is useless to a receiving party, since the ciphertext is meant to be viewed anyway (though this ciphertext
might be extremely valuable to an outside observer). With this said, it is best to keep the ciphertext delivery completely separated from the delivery of the key.

2.3 Modern Standard Arabic language statistics.

Frequency analysis.

The following statistical data on the frequencies of the Modern Standard Arabic alphabet and its special characters was computed by a Canadian software company called Intellaren. Intellaren (2012) specializes in multilingual software solutions and has produced software products such as Intellyze, which is an Arabic letter and word frequency analyzer.

In order to calculate the letter frequencies of MSA, Intellaren (2012) used its software to conduct an experiment on four different input sources (see below). For the experiment, their software counted the number of times each character of the MSA alphabet and its special characters appeared in each source, while omitting spaces, punctuation, and vowelization. The frequency count data is separated into two categories; one category for Source 1 (The Quran القرآن), the second, a combination of Sources 2-4.

Four original input sources (Intellaren, 2012).

Source 1: The Quran القرآن (total number of letters: 330,709)

Source 2: The first seven volumes of the series البداية والنتيجة (The Beginning and The End) by Ibn Katheer (total number of letters: 4,326,031)
Source 3: The Sealed Nectar: Biography of the Noble Prophet الرحيق المختوم by Ar-Raheeq Al-Makhtum (total number of letters: 553,740)

Source 4: The Masterpiece of the Brides تحفة العروسين by Al-shuri (total number of letters: 242,361 letters)

Combined Total: 5,452,841 letters

Combining the data for the four original input sources.

There are a variety of valid arguments for using different data combinations of the four original input sources (or other input sources), or perhaps leaving each data input source on its own, for the use of particular topics. Some of these arguments hold more value than others in topics such as language, religion, culture, or statistics. For instance, if one is analyzing Middle Eastern religious texts, one would likely want to stay focused on The Quran data or other dialects pertaining to archaeological sites. Furthermore, since The Quran is a religious text it may skew the statistical properties of other less religious text data. This is due to the fact that The Quran is more prone to repeating certain religious words and phrases, and consequently, its data could potentially statistically alter the data of purely nonreligious material. Religious and specific dialect data may also hold value if one is analyzing codes from ideological or religious websites in a particular region or country. As for statistical purposes, an organization or company may want to collect data on all Arabic dialects, words, and characters from the vast sea of the Internet for marketing research, business intelligence, and government intelligence. Another perhaps very subtle reason for combining religious and multiple dialect data, is that the
Religious influence of Islam is almost a constant in the Arab world in terms of language and can be seen and heard in most daily life activities.

The data for the four original input sources have been combined by the author in order to reflect the combination of the topics of religious influence in the Arab world and the specific dialect of Modern Standard Arabic. Therefore, the author states that these data are statistically relevant to this combination of topics and to this specific combination of topics only. In terms of cryptanalysis, one could potentially benefit from this statistical data if the background of ciphertext warrants this reflection.

**Figure 11.** Letter frequencies and relative frequencies of the Modern Standard Arabic alphabet and its special characters.

<table>
<thead>
<tr>
<th>$x$</th>
<th>Frequency</th>
<th>Relative Frequency</th>
<th>$x$</th>
<th>Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>ا</td>
<td>683,904</td>
<td>.125422</td>
<td>غ</td>
<td>17,914</td>
<td>.003285</td>
</tr>
<tr>
<td>ب</td>
<td>250,553</td>
<td>.045949</td>
<td>ف</td>
<td>154,034</td>
<td>.028248</td>
</tr>
<tr>
<td>ت</td>
<td>144,017</td>
<td>.026411</td>
<td>ق</td>
<td>144,837</td>
<td>.026562</td>
</tr>
<tr>
<td>ث</td>
<td>46,097</td>
<td>.008454</td>
<td>ك</td>
<td>114,900</td>
<td>.021072</td>
</tr>
<tr>
<td>ج</td>
<td>66,217</td>
<td>.012144</td>
<td>ل</td>
<td>656,656</td>
<td>.120425</td>
</tr>
<tr>
<td>ح</td>
<td>99,614</td>
<td>.018268</td>
<td>م</td>
<td>360,593</td>
<td>.066129</td>
</tr>
<tr>
<td>خ</td>
<td>43,032</td>
<td>.007892</td>
<td>ن</td>
<td>365,916</td>
<td>.067106</td>
</tr>
<tr>
<td>د</td>
<td>142,488</td>
<td>.026131</td>
<td>ه</td>
<td>274,960</td>
<td>.050425</td>
</tr>
<tr>
<td>ز</td>
<td>53,998</td>
<td>.009903</td>
<td>و</td>
<td>321,877</td>
<td>.059029</td>
</tr>
<tr>
<td>ر</td>
<td>227,689</td>
<td>.041756</td>
<td>ي</td>
<td>347,804</td>
<td>.063784</td>
</tr>
<tr>
<td>س</td>
<td>28,378</td>
<td>.005204</td>
<td>أ</td>
<td>156,998</td>
<td>.028792</td>
</tr>
<tr>
<td>ش</td>
<td>132,282</td>
<td>.024259</td>
<td>ء</td>
<td>56,141</td>
<td>.010296</td>
</tr>
<tr>
<td></td>
<td>39,730</td>
<td>.007286</td>
<td>أ</td>
<td>9,003</td>
<td>.001651</td>
</tr>
</tbody>
</table>
Note: Relative Frequency = Frequency/Total. For example, 683,904/5,452,841 = .125422

Interpretation of MSA letter frequency data.

It is quite obvious from the data that the two most frequent occurring letters in the
MSA alphabet are “ﺍ” and “ﺩ”. This is due to the fact that the particular digram
combination of the two letters, “ﺍﺩ”, is the definite article for MSA, which is the U.S.
English equivalent of the word “the”, such as “the apple” or “the scooter.” The indefinite articles for U.S. English are represented by the words “a” and “an”, for example, “an apple” or “a book.” In MSA, indefiniteness is signified by the removal of “ﺕ” followed by nunation, which is given by the nunation case markers ً،ٍ،ٌ (Abboud & McCarus, 1983, pp. 110-112). This nunation was not considered in the four original input sources since it is a part of vowelization.

2.4 Additional examples of CNN(C).

**Affine cipher: The foreign students.**

The affine cipher is a well-known cryptosystem within the field of cryptography and the following example parallels the same affine cipher found in Trappe and Washington (2006). Again, this particular affine cipher uses MSA as well as the author’s own set of numerical values.

Pakistan - Receiving news that some of Karim’s local messengers might be corrupt, Sami issues an order to his staff that all messages leaving the Hotel must be encrypted.

Sami would like to send another message to his friend Maria in Dora telling her about some of the foreign students he has met over the last few weeks at one of the private universities. Sami encrypts his message using an affine cipher and gives it to his doorman Mohammed for delivery.

For the affine cipher, Sami labels the letters of the MSA alphabet and special characters with the same integer values as in Figures 8 and 9.
Sami’s message \( m \) to Maria is,

message \( m \):

I met a number of Syrian and Egyptian students. As for the Kuwaiti students, I haven’t met one of them so far” (Abboud & McCarus, 1983, p. 206).

For the affine cipher, Sami uses the encryption formula (affine function),

\[
f(x) \equiv (ax + \beta) \mod 36
\]

where \( \gcd(a, 36) = 1 \) and \( \beta \) is an integer from the set \( \{0, 1, \ldots, 35\} \).

Sami selects \( a = 29 \) and \( \beta = 35 \) as the key and writes it as an ordered pair \((a, \beta)\).

So in this case, the key is \((29, 35)\) and \( f(x) \equiv (29x + 35) \mod 36 \). Sami also makes sure that the greatest common divisor of 29 and 36 is equal to one or \( \gcd(29, 36) = 1 \). This property means that 29 and 36 are relatively prime and it guarantees that there exists an inverse for 29 modulo 36, and hence an inverse for the affine function.

**Figure 13.** Affine cipher encryption.

<table>
<thead>
<tr>
<th>plaintext</th>
<th>ciphertext</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>1 0 22 15 22 0 24 23 0 7 7 17 2 22 1 0 20</td>
</tr>
<tr>
<td>((29x + 35) \mod 36)</td>
<td>28 35 25 2 25 35 11 18 35 22 22 24 21 25 28 35 3</td>
</tr>
<tr>
<td>ciphertext</td>
<td>تُؤَهَّسُ هُؤَامُلُ لُمُؤَغُ لُهُكُنُ أُمُؤَائُ</td>
</tr>
<tr>
<td>plaintext</td>
<td>الــسُورِيُّ يُن و الـلـم صـرـيـيـن</td>
</tr>
<tr>
<td>x</td>
<td>24 27 27 9 13 23 22 0 26 24 27 27 9 26 11 22 0</td>
</tr>
</tbody>
</table>
This message is given to Mohammed for delivery. In order for Maria to decrypt the ciphertext, she will need the corresponding key and cryptosystem decryption formula,

\[
f(x) \equiv (29x + 35) \mod 36
\]

\[
f(x) - 35 \equiv (29x) \mod 36
\]

\[
29^{-1}(f(x) - 35) \equiv x \mod 36
\]

\[
5(f(x) - 35) \equiv x \mod 36
\]

\[
(5f(x) - 31) \mod 36 \equiv x
\]
which are \((29, 35)\) and \((5y - 31) \mod 36 \equiv x\), respectively. In this case, Sami will assign a new courier to deliver these items to Maria in three days.

**Playfair cipher: The Pakistani government.**

The Playfair cipher is a classical cryptosystem within the field of cryptography and the following example parallels the same Playfair cipher found in Trappe and Washington (2006). Again, this particular Playfair cipher uses MSA.

Afghanistan - Karim receives a letter from Sami asking if it is safe for their group to conduct operations in one particular area of the region without interference from the Pakistani government. Karim decides to encrypt his message using the Playfair cipher. For the Playfair cipher, Karim chooses a word for the key.

Karim chooses the word, نقود translation: 

\[\text{cash, money; coins}\]

Having selected a key, Karim places his word in the first row of a \(6 \times 6\) matrix and fills in the remaining spaces with the remaining letters of the alphabet (and special characters). If a letter is repeated in the key, one of the letters is omitted. Karim’s key is in purple.

Karim’s matrix is therefore,

\[
\begin{array}{cccccc}
\text{ن} & \text{ق} & \text{و} & \text{د} & \text{اب} \\
\text{ت} & \text{ش} & \text{ج} & \text{خ} & \text{ذ}
\end{array}
\]

\[
\begin{array}{cccccc}
\text{ر} & \text{ز} & \text{س} & \text{ش} & \text{ض} \\
\text{ط} & \text{ظ} & \text{ع} & \text{غ} & \text{ف}
\end{array}
\]
Karim’s message is,

لا تسمح الحكومة لأصدقائنا بالإقامة هناك طويلا


*The government does not permit our friends to stay there long.*

Karim breaks his message into blocks of size 2 and adds an extra ب in the last block. This extra letter is chosen at random from the alphabet in order to complete a block of size 2.

*Figure 14. Playfair cipher encryption.*

Using his matrix, Karim derives the ciphertext from the following rules taken from Trappe and Washington (2006):

1. Letters that are not in the same row or column are replaced by the letters that are in the row and the column of the other letter.
2. Letters that are in the same row are replaced by the letters to the right with the row continuously wrapping around the matrix.

3. Letters that are in the same column are replaced with the letters below with the column continuously wrapping around the matrix.

ciphertext:

Karim then sends the key and the ciphertext to Sami separately. Karim sends the key with a messenger to a particular dead drop site, whereby another messenger picks up the key and delivers it to Sami. This dead drop method is repeated with a different location and with different messengers for the ciphertext. The ciphertext is always delivered three days after the delivery of the key.
Figure 15. The Pakistani government network.

<table>
<thead>
<tr>
<th>Location 1 Data Fields</th>
<th>Location 2 Data Fields</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Location (Country - Afghanistan, town - Henna, description - near Afghanistan/Pakistan border, place - farmhouse)</td>
<td>• Location (Town/place - unknown, description - near Afghanistan/Pakistan border)</td>
</tr>
<tr>
<td>• Persons of interest (Karim and friends, two messengers - unknown)</td>
<td>• Persons of interest (two messengers - unknown)</td>
</tr>
<tr>
<td>• Covert non-electronic communication methods (description - no electronics available, two different messengers deliver key and ciphertext to two different dead drop sites, cryptography - Playfair cipher - key and ciphertext)</td>
<td>• Covert non-electronic communication methods (ciphertext exchange at dead drop site)</td>
</tr>
</tbody>
</table>
The Vigenère cipher: The ghost.

The Vigenère cipher is a classical cryptosystem and the following example parallels the same Vigenère cipher method for encryption found in Trappe and Washtington (2006). Again, this particular cipher uses MSA as well as the author’s own set of numerical values.

Pakistan - Sami would like to inform Karim about his recent trip to the university. Moreover, he would like to notify Karim that he had a meeting with the university’s president and that the president received a letter from a very important man who goes by the alias of the Ghost. Sami decides to encrypt his message using the Vigenère cipher (ignoring spaces and punctuation).

Sami’s message $m$ is,

وصل رئيس الجامعة و هو يحمل رسالة هامة من الشيخ
The president of the university arrived while carrying an important letter from the
Ghost. If the Ghost leaves Damascus in the morning he will be here before the evening.

The letter:

Pakistan tries to follow a policy of non-alignment; that means it is not aligned
with either the East or the West (Abboud & Mccarus, 1983).

For the Vigenère cipher, Sami chooses a vector \( v \) of length \( k \) for the encryption
key. In this case, Sami chooses an Arabic keyword for the vector, since the word can be
broken down numerically and it can be easily remembered. Sami chooses the Arabic
keyword,

\[ انتخاب \]

which means,

\[ election \]

The length of the vector is 6, since the Arabic word has 6 letters. Figures 8 and 9 give the
corresponding vector \( v = (1, 0, 6, 2, 24, 0) \). Remember, the entries of the vector are read
right to left, as are the letters in the Arabic word.

\[ انتخاب → 1 0 6 2 24 0 \]
To encrypt his message, Sami repeatedly writes the vector below the plaintext and shifts each letter of the plaintext by the amount given in the corresponding vector entry (mod 36), hence the Vigenère cipher contains multiple shift ciphers.

*Figure 16. Vigenère cipher encryption.*

<table>
<thead>
<tr>
<th>plaintext</th>
<th>ciphertext</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
</tr>
<tr>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
<tr>
<td>plaintext</td>
<td>ciphertext</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
<tr>
<td>plaintext</td>
<td>ciphertext</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
<tr>
<td>plaintext</td>
<td>ciphertext</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>plaintext</th>
<th>ciphertext</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
</tr>
<tr>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
<tr>
<td>plaintext</td>
<td>ciphertext</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
<tr>
<td>plaintext</td>
<td>ciphertext</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
<tr>
<td>plaintext</td>
<td>ciphertext</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>plaintext</th>
<th>ciphertext</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
</tr>
<tr>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
<tr>
<td>plaintext</td>
<td>ciphertext</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
<tr>
<td>plaintext</td>
<td>ciphertext</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
<tr>
<td>plaintext</td>
<td>ciphertext</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>plaintext</th>
<th>ciphertext</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
</tr>
<tr>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
<tr>
<td>plaintext</td>
<td>ciphertext</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
<tr>
<td>plaintext</td>
<td>ciphertext</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
<tr>
<td>plaintext</td>
<td>ciphertext</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>plaintext</th>
<th>ciphertext</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
</tr>
<tr>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
<tr>
<td>plaintext</td>
<td>ciphertext</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
<tr>
<td>plaintext</td>
<td>ciphertext</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
<tr>
<td>plaintext</td>
<td>ciphertext</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>plaintext</th>
<th>ciphertext</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
</tr>
<tr>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
<tr>
<td>plaintext</td>
<td>ciphertext</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
<tr>
<td>plaintext</td>
<td>ciphertext</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
<tr>
<td>plaintext</td>
<td>ciphertext</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>(x + c) mod 36</td>
<td>(x + d) mod 36</td>
</tr>
<tr>
<td>plaintext</td>
<td>ciphertext</td>
</tr>
<tr>
<td>-----------</td>
<td>------------</td>
</tr>
<tr>
<td>1 0 6 2 24 0</td>
<td>1 0 6 2 24 0</td>
</tr>
<tr>
<td>24 22 6 24 25 20</td>
<td>6 0 7 15 10 0</td>
</tr>
</tbody>
</table>

(\(x + c\)) mod 36
Sami puts his ciphertext in an envelope and gives it to a courier. He also sends the secret keyword and Arabic Vigenère square with another courier, who is told to take a different route.

**The Vigenère cipher: Cryptanalysis.**

In order to analyze the Vigenère Cipher we need five things:

1. Kasiski’s method
2. The frequency distribution of single ciphertext characters
3. The measure of roughness
4. The index of coincidence
5. The monoalphabetic indices of coincidence for a set value of t

*Kasiski’s method.*

Kasiski’s method is a cryptanalysis technique that helps determine the number of alphabets $t$ of a polyalphabetic substitution cipher. Moreover, the value $t$ represents the length of the unknown keyword, which is sometimes described as the number of monoalphabetic substitutions or shifts resulting from each individual letter of the keyword, hence the term polyalphabetic substitution. Kasiski’s method is based on the idea that identical portions of plaintext encrypted with identical portions of the keyword produce identical portions of ciphertext (Menezes, van Oorschot, & Vanstone, 1997).

With this said, it is then hypothesized that the number of characters between the initial starting points of identical portions of ciphertext is a multiple of the number of the alphabets or keyword length $t$. Once a list of the possible number of alphabets or keyword length $t$ is established, a statistical comparison of all individual ciphertext alphabets for each set value of $t$ is made to the inherent properties of one particular source language alphabet. Note, if all individual ciphertext alphabets contain identical characters to this particular source language alphabet, then the cipher is classified as a simple Vigenère cipher. Furthermore, each ciphertext alphabet is a shift cipher of this original source language alphabet. If the characters of the individual ciphertext alphabets differ from the characters of the original source language alphabet, the statistical properties of each ciphertext alphabet will still reflect the original properties of the source.
language. This is due to simple substitution, however; the cipher is not classified as a simple Vigenère cipher.

*The frequency distribution of single ciphertext characters.*

If the ciphertext is short, it is best to count the frequency of each character by hand. If the ciphertext is long, it may be better to use software to count the frequency of each character.

*The measure of roughness.*

The **measure of roughness** is a quantity that measures the deviation of plaintext (author’s emphasis) character relative frequencies from a uniform frequency distribution (UFD) for a given *source language* (Menezes et al., 1997). The measure of roughness is defined as follows,

$$MR = \sum_{i=0}^{n-1} p_i^2 - \left(\frac{1}{n}\right)$$

where \(n\) equals the number of distinct characters of the *plaintext* alphabet \(\{a_{n-1}, a_{n-2}, \ldots, a_1, a_0\}\) from the *source language*, and \(p_i\) are the corresponding character relative frequencies. Hence, the measure of roughness can be quantified for any *source language* with *plaintext* alphabet \(\{a_{n-1}, a_{n-2}, \ldots, a_1, a_0\}\). The term *plaintext* alphabet may be substituted by the term *plaintext* character set. In practice, it would be wise to use software such as Intellyze (see Figure 11), to estimate the character relative frequencies \(p_i\) of the *plaintext* alphabet from a large data set containing the *source language*. This is due to the fact that character relative frequencies from a small source, say a handwritten
message, do not accurately reflect the overall structure of character relative frequencies for an entire source language.

To derive the MR we expand the expression below,

\[ MR = \sum_{i=0}^{n-1} (p_i - \frac{1}{n})^2 \]

\[ = \sum_{i=0}^{n-1} p_i^2 - 2p_i(\frac{1}{n}) + (\frac{1}{n})^2 \]

\[ = \sum_{i=0}^{n-1} p_i^2 - \sum_{i=0}^{n-1} 2p_i(\frac{1}{n}) + \sum_{i=0}^{n-1} (\frac{1}{n})^2 \]

\[ = \sum_{i=0}^{n-1} p_i^2 - 2(\frac{1}{n})\sum_{i=0}^{n-1} p_i + n(\frac{1}{n})^2 \]

\[ = \sum_{i=0}^{n-1} p_i^2 - 2(\frac{1}{n}) + \left(\frac{1}{n}\right) \]

\[ = \sum_{i=0}^{n-1} p_i^2 - \left(\frac{1}{n}\right) \]

We can find the minimum value for the MR of the source language by looking at its value for a UFD. For a UFD, every character \(a_i\) of the plaintext alphabet occurs with equal probability \(p_i = (1/n)\), therefore \(\min(MR) = 0\). A maximum value for the MR occurs when the plaintext character relative frequencies \(p_i\) of the source language are present (Menezes et al., 1997). Note, however, that this maximum value is an approximation, since it is based on previously estimated character relative frequencies \(p_i\). This is given by \(\max(MR) = \sum p_i^2 - (1/n) = k_p - (1/n)\), where \(k_p = \sum p_i^2\) is called the roughness constant of the source language. The roughness constant \(k_p = \sum p_i^2\) is the probability that two randomly chosen characters from the plaintext alphabet of source language are in fact equal. This is quite obvious since encryption of the source language is not present.
The index of coincidence.

The **index of coincidence (IC)** is a quantity that measures the variation in ciphertext (author’s emphasis) character frequencies for a given ciphertext (Pieprzyk, Hardjono, & Seberry, 2003). The formula for the index of coincidence is,

\[
IC = \frac{\sum_{i=0}^{n-1} \binom{f_i}{2}}{\binom{L}{2}} = \frac{\sum_{i=0}^{n-1} f_i (f_i - 1)}{L (L - 1)}
\]

where \( L \) is the length of the ciphertext, and \( f_i \) is the frequency of letter \( b_i \) in the ciphertext alphabet \( \{b_{n-1}, b_{n-2}, \ldots, b_1, b_0\} \) (Menezes et al., 1997). The term ciphertext alphabet may be substituted by the term ciphertext character set. Also note, that unlike the MR, character frequencies are used instead of character relative frequencies.

Given a ciphertext of length \( L \), the binomial coefficient \( \binom{L}{2} \) represents the number of distinct character groups of size 2 that can be selected from the ciphertext when order of selection is not considered relevant (Ross, 2010). Similarly, given a particular ciphertext character \( b_i \) with frequency \( f_i \), the binomial coefficient \( \binom{f_i}{2} \) represents the number of distinct groups of size 2 that can be selected from the character group \( b_i \) when order of selection is not considered relevant. Therefore, the IC is the probability that a pair of identical ciphertext characters \( (b_i, b_i) \) from the ciphertext alphabet \( \{b_{n-1}, b_{n-2}, \ldots, b_1, b_0\} \) with corresponding frequency \( f_i \), are the same for a given ciphertext of length \( L \).
The IC is a direct estimate of the source language roughness constant, $k_p = \sum p_i^2$ (Menezes et al., 1997), given that the ciphertext alphabet $\{b_{n-1}, b_{n-2}, \ldots, b_1, b_0\}$ and plaintext alphabet $\{a_{n-1}, a_{n-2}, \ldots, a_1, a_0\}$ are statistically identical. For instance, in simple substitution, an augmented version of an English ciphertext alphabet $\{b_{n-1}, b_{n-2}, \ldots, b_1, b_0\} = \{1, 2, 3, \ldots, c, b, a\}$ may represent the MSA plaintext alphabet $\{a_{n-1}, a_{n-2}, \ldots, a_2, a_1, a_0\} = \{\mathbf{漳州}, \mathbf{汉语}, \mathbf{阿拉伯语}, \mathbf{土耳其语}, \mathbf{阿拉伯语}\}$. This would be accomplished by substituting the first letter of the MSA alphabet with the first letter of the English alphabet, the second letter of MSA with the second letter of English, and so on. English numbers would replace the remaining MSA plaintext letters. Although, a random ciphertext sample of length $L$ with character frequencies $f_i$ would appear as English, it would statistically exhibit the MSA source language with plaintext character relative frequencies $p_i$. This is due to the fact that as $L$ increases for the sample English ciphertext, $f_i/L \approx p_i$. For now, we will assume that the character set of the ciphertext alphabet is identical to the character set of the plaintext alphabet. As mentioned earlier, it would be wise to estimate the plaintext character relative frequencies $p_i$ of the source language with software.

Proof (Pieprzyk et al., 2003).

Substituting $f_i \approx L \cdot p_i$ into the formula for the IC above we get,

$$\text{IC} \approx \sum_{i=0}^{n-1} \frac{p_i^2(L - \frac{1}{k})}{L - 1}$$

Letting $L$ go to infinity, we get $\text{IC} \approx \sum p_i^2$. Therefore, the IC approximates the roughness constant $k_p = \sum p_i^2$ of the source language.
The monoalphabetic indices of coincidence for a set value of t.

For each ciphertext alphabet, compare its corresponding index of coincidence to the roughness constant of the source language.

**Applying cryptanalysis to the ghost.**

Statistically speaking, we assume that the ciphertext reflects the combination of the topics of religion and the MSA dialect.

**Kasiski’s method.**

Looking at the ciphertext on p. 41, we notice that lines 4 and 5 both contain the trigram “ﻥ ﺧﻝ”. The number of characters between the initial starting points “ﻥ” of these two trigrams (making sure not to double count) is 18. Therefore, it is reasonable to conclude that the number of alphabets or keyword length t is either 1, 2, 3, 6, 9, or 18.

**The frequency distribution of single ciphertext characters.**

The frequency distribution of single ciphertext characters is directly computed from the ciphertext and its letter frequencies found on p. 41.

*Figure 17. Ciphertext frequency distribution.*

<table>
<thead>
<tr>
<th></th>
<th>12</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ﺞ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ﻋ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ﺕ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ﻙ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ﺞ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ﻋ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ﺞ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ﻙ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ﺞ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ﻋ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ﺞ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ﻙ</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The measure of roughness.

For the source language of MSA, there exists 36 distinct characters in the plaintext nonvowelized character set. This plaintext character set is given by \( \{a_{n-1}, a_{n-2}, \ldots, a_2, a_1, a_0\} = \{ﺍ،ﺏ،ﺕ،ﺓ،ﺅ،ﻱ،ﺃ،ﺇ،ﺀ،ﺉ،ﻯ،ﺓ\} \), where \( n = 36 \) and \( 35 \geq i \geq 0 \). Using the plaintext character relative frequencies from Figure 1 we can approximate the roughness constant \( k_p \) for MSA. For \( k_p \) we get,

\[
k_p = \sum_{i=0}^{35} p_i^2 = (.000925)^2 + (.013809)^2 + (.012591)^2 + \ldots + (.125422)^2
\]

\[
= 0.0608
\]

Therefore, \( \max(MR) = k_p - (1/n) \approx 0.0608 - (1/36) = 0.0330 \), is an approximate maximum for the MR of MSA. The minimum value of the MR is taken from the UFD.

\[ 0.0 \leq MR \leq 0.0330 \]
The MR relates to the index of coincidence, which is described below.

*The index of coincidence.*

Since \( IC \approx \sum p_i^2 = k_p \), we can find approximate lower and upper bounds for the IC of MSA. Since \( IC \approx \sum p_i^2 = k_p = MR + 1/n \) and \( 0.0 \leq MR \leq 0.0330 \), we know that \( 0 + 1/n \leq MR + 1/n \leq 0.0330 + 1/n \). Here, \( 1/n = 1/36 = .0278 \).

For MSA they are,

\[
0.0278 \leq IC \leq 0.0608
\]

For MSA cryptanalysis involving simple substitution or shifts, we are only concerned with ciphertext alphabets that have an IC value gravitating towards the upper bound value of 0.0608. This is due to the fact that this value represents the roughness constant \( k_p \) of MSA, and is therefore a strong indicator that a particular ciphertext alphabet exhibits the same statistical properties of the previously computed MSA character relative frequencies \( p_i \).

It appears that the Internet only gives one source on the index of coincidence for Arabic; which specific dialect of Arabic this is for is not mentioned (note, English sources only). The value is presently stated as \( IC = 0.075889 \) (3/1/2012). This author contacted J. Seberry, a well-known cryptographer and computer scientist and an author of *Fundamentals of Computer Security* (2003). Her reply confirmed that this value is incorrect and that there was an error in the original data (personal communication, May 8, 2011; see Appendix A).
The monoalphabetic indices of coincidence for a set value of \( t \).

In practice, software should be used to compare the indices of coincidence for all ciphertext alphabets for all values of \( t \) to the roughness constant \( k_p \) of MSA. To save time, we assume \( t = 6 \) for the number of alphabets or keyword length and divide the ciphertext accordingly into 6 monoalphabetic groups (Pieprzyk et al., 2003). These monoalphabetic groups are formed by individual character positions within the ciphertext. The total length of the ciphertext is \( L = 156 \), therefore the first monoalphabetic group is given by character positions \( 6i \), where \( i = 0, 1, 2, \ldots 25 \), since \( L/6 = 156/6 = 26 \). The second group is given by character positions \( 6i + 1 \), where \( i = 0, 1, 2, \ldots 25 \), and so on. We then compose a separate ciphertext frequency distribution for each monoalphabetic group and compute its index of coincidence.

The table below shows the monoalphabetic indices of coincidence for \( t = 6 \).

**Figure 18. Monoalphabetic indices of coincidence.**

<table>
<thead>
<tr>
<th>Group 6</th>
<th>Group 5</th>
<th>Group 4</th>
<th>Group 3</th>
<th>Group 2</th>
<th>Group 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 6i + 5 )</td>
<td>( 6i + 4 )</td>
<td>( 6i + 3 )</td>
<td>( 6i + 2 )</td>
<td>( 6i + 1 )</td>
<td>( 6i )</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>
By comparing the respective index of coincidence of each group to the MSA roughness constant $k_p = \sum p_i^2 = 0.0608$, the IC gives an approximate measure of how close the ciphertext frequency distribution is to the MSA plaintext frequency distribution. With this said, we look at groups 1 and 5 since they both have an IC value of 0.0677, which is close to 0.0608. The most frequently occurring letter in these two groups is “ﺍ”, so we assume “ﺍ” is part of the overall keyword for positions 1 and 5. Why? The nice thing about MSA is that the most frequently occurring letter in the alphabet is the first letter of alphabet, which is “ﺍ”. This means that the most frequently occurring letter in a simple substitution ciphertext should not only reveal the number of shifts, but also the actual
letter of the shift. This property of MSA is called the **alif property of simple substitution**, since the letter “ا” is pronounced alif. This is different from U.S. English, where e is the most common letter. This is demonstrated below for a shift by “ﺙ” for MSA and a shift “i” for American English.

**MSA:**

\[
\text{ﺙ =ﺙ + ١} \\
3 = (3 + 0) \mod 36 \\
3 =ﺙ
\]

**American English:**

\[
e + i = m \\
(4 + 8) \mod 26 = 12 \\
12 = м
\]

Now looking at the IC values for groups 2 and 4, we can see that they are not far off from 0.0608. The most frequent letter in group 2 is “ﻥ”, so by the alif property, the second letter of the keyword is “ﻥ”. Next, for group 4, the most frequent letter is “ﺥ”, so again by the alif property, “ﺥ” is the fourth letter of the keyword. For group 3, the most common letter is “ﻥ”, but in MSA, two consonants such as “ﻥ” cannot occur in a row, and since we have decided “ﻥ” is keyword letter for group 2 we must choose a different letter. The next most common letter in group 3 is “ﺙ”, so this will be our choice for the third letter of the keyword. For groups, 1, 2, 3, 4, 5, our keyword is,

loon تﺥ

Guessing “ﺏ” for the last letter of the key in group 6 we find that the keyword is,

loon غﺏ
or,

\[
\text{انتخاب}
\]

which translates to “elections”.

**The expected value of the index of coincidence.**

Remember, the IC is the probability that a pair of identical ciphertext characters \((b_i, b_i)\) from the ciphertext alphabet \(\{b_{n-1}, b_{n-2}, \ldots, b_1, b_0\}\) with corresponding frequency \(f_i\), are the same for a given ciphertext of length \(L\). Letting \(T\) be the number of alphabets \(t\) of a polyalphabetic substitution cipher, the probability mass function for the index of coincidence with \(T\) alphabets is defined as follows:

\[
P[T = 1] = \text{simple substitution, 1 alphabet, or } k_p\text{ such that } k_p = \sum p_i^2 \text{ is the roughness constant of the source language.}
\]

\[
P[T = 2] = 2 \text{ alphabets}
\]

\[
\vdots
\]

\[
P[T = \infty] = k_r = 1/n, \text{ where } T \text{ is said to have an infinite number of alphabets } t \text{ or an infinite period (Pieprzyk et al., 2003). This statement is equivalent to saying that every character of the source language has equal probability.}
\]

The expected value for the index of coincidence of a polyalphabetic substitution cipher of \(t\) alphabets is given by,

\[
E[IC] = \frac{1}{t} \cdot \frac{L-t}{L-1} \cdot k_p + \frac{t-1}{t} \cdot \frac{L}{L-1} \cdot k_r
\]

where \(L\) is the length of the ciphertext, \(k_p\) is defined as the roughness constant of the source language, and \(k_r = 1/n\), where \(n\) is the size of the source language alphabet (Menezes et al., 1997). Many Internet sources and books claim that by tabulating values
of \( t \) with corresponding IC expected values, one can derive the number of alphabets \( t \) of the ciphertext by comparing these expected values to the IC of the source language (see. Menezes et al., 1997 pp. 249 -250). However, this formula is statistical in nature and can give misleading results (Pieprzyk et al., 2003).

### 2.5 Hidden objects.

*Hidden objects: Flashlight.*

Pakistan - Sami and his friend Mohammed come up with a new way to exchange secret messages or codes in order to plan their next meeting. The new plan involves writing a secret message or code on a small piece of plastic from an index card and then hiding it inside a flashlight. The flashlight is then given to a courier or intermediary, who will deliver it to either Sami or Mohammed, hence resulting in two-way communication.

Sami writes the Arabic word for the market “سوق” with a dry-erase marker on the piece of plastic and inserts it into the flashlight (the plastic insert and dry-erase marker allows for re-use of the cryptosystem). The figures below depict this process:

*Figure 19. First step of hiding process.*
Part I Summary

Although one may quickly jump at the opportunity to label covert non-electronic networks (CNN) and covert non-electronic networks combined with cryptography [CNN(C)] as primitive, obsolete, and perhaps fantasies of the imagination, one should note that covert non-electronic communication methods can often provide for an overall effective battlefield strategy. Secret human networks comprising of unique non-electronic communication methods can be extremely difficult to detect or observe if no prior knowledge is given beforehand regarding these network components. These covert non-electronic methods, which can include drops, courier routes, drivers, cryptography, and steganography, can at times be very abstract and unyielding to those who wish to
observe or disrupt. Therefore, CNN and CNN(C) must be taken very seriously by intelligence services across the globe. Perhaps too much emphasis has been focused on electronic surveillance and the assumption that people will follow normal communication patterns. As for combatting such networks, one must pay close attention to the individual features of a particular network and always be prepared to encounter an unexpected complexity, whether it be in regards to mathematics, intelligence, language, or culture.

For more information on a real life example of a CNN, search how the C.I.A. was able to discover Osama bin Laden’s compound by tracking one of his couriers.
PART II

Covert Electronic Networks and Arabic Cryptology Applications

The reality is that cybercrime is happening internationally among organized criminal groups, corporations, and governments, as well as at the individual hacker level. Besides simply communicating about their various activities, one of the most frequent uses of the Internet by organized crime and terrorists alike is to launder money, which involves secret communication about that money as well as financial transactions that move it from account to account. The Internet provides the perfect climate for criminals for several reasons. It’s fast, it’s anonymous, it’s unregulated. (Cole, 2003, p. 93)

Now that one has seen how the art of hand-to-hand secret communication works, or as more recently stated, covert non-electronic networks, our attention will now shift in this part to more sophisticated covert networks involving modern day electronics.

Returning to the Iranian example in part I, suppose now that the American spy has been given new orders to infiltrate an Iranian telecommunications company, called Telecommunication Company of Iran (TCI). The spy’s new orders state that he or she must infiltrate TCI through electronic means only, since non-electronic methods are just too dangerous at this particular point in time. More specifically, the spy is told that he or she could possibly send an email containing a trojan horse to TCI, that when opened by an unsuspecting employee, uses code to obtain access to pieces of TCI’s mobile services infrastructure. Once access has been granted, the spy will then use his or her computer
skills to locate the phone numbers and other important data of top Iranian government officials. Some of the questions and concerns of this project might be: Is a trojan horse the fastest and most efficient electronic method available to the spy to achieve his or her goals? If not, what are some alternative computer/electronic schemes that could be used instead? For instance, is there an insider at the company who would be willing to encrypt the information and send it to the American spy with the corresponding encryption keys? Could the spy possibly create a fake identity and trick a naïve employee into giving what he or she wants over the phone? Obviously such methods must be compared beforehand in terms of cost, time efficiency, and whether or not they place people or groups at risk or exposure to harm.

In the context of spy tradecraft, this part will describe a variety of ways in which two or more people can secretly communicate with one another with the support of electronic devices and applications such as cell phones, email, Internet forums, instant messaging, etc. These **covert electronic communication methods** can include the use of contacts, electronic dead drops, specific computer terminal locations, cryptography, and steganography. (Again, note that the word *covert* in this instance denotes an activity that is “secret” to the controllers of communication, but in fact may be exposed to outsiders.) These covert electronic communication methods can rely entirely upon electronic devices/applications or they can depend upon both a combination of electronic devices/applications with cryptographic schemes. When covert electronic communication methods are systematically combined together, they ultimately form **covert electronic networks (CEN)** or **covert electronic networks** combined with

cryptography (CEN(C)). The main advantages of these types of secret networks are that they are fast, affordable, and easy to use (sometimes). However, the benefits of such networks can often be short lived. Disadvantages can include eavesdropping, wiretapping, and cryptanalysis, all of which can go undetected or remain silent in the background.

We include a few theoretical examples that demonstrate how in this age of high-tech electronics and other things like easy access to the Internet, such networks can be quickly formed without notice or detection by groups such as al-Qaeda, the Taliban, jihadists, and other insurgents. This part will also present how these covert electronic communication methods can be manipulated, expanded, and creatively altered by other factors (e.g., abstract electronic steganography) to achieve an objective that is specific to an individual covert network. In the future, it is these structural modifications that will be of interest when categorizing CEN/CEN(C), and perhaps performing steganography analysis or cryptanalysis on a particular CEN or CEN(C).
CHAPTER 3: Covert Electronic Networks (CEN)

Today terrorist groups are on the cutting edge of technology. They use computers, the Internet, encryption, and steganography to conduct business. If their cryptography is good, it can take decades to crack. If they use steganography, their transmission of data may go completely undetected. (Cole, 2003, p. 8)

3.1 Covert electronic communication methods.

All data fields located under the data field of covert electronic communication methods can be found in Figure 6. We have provided some descriptions of these data fields below and how they relate to spy tradecraft.

An electronic courier or messenger, is a trusted individual who is sent an electronic message by one party, and is expected to follow through by delivering the message (by electronic means only) to a second party at a specified location. A covert electronic network can consist of multiple digital couriers, destinations, and vested parties. An electronic contact, on the other hand, is a person or clandestine meeting between two parties over an electronic channel. An electronic dead drop is an electronic depository or location where two parties agree to exchange secret information. An example would be a computer file (Cole, 2003). Electronic steganography is the electronic component of steganography, in which writing or text is hidden in an electronic setting such as an image file, audio file, word document, etc. In order to save time we only consider examples of covert electronic networks combined with steganography and cryptography, which are provided in the following chapter.
CHAPTER 4: Covert Electronic Networks Combined with Cryptography [CEN(C)]

Unlike ordinary CEN, CEN(C) look to the field of cryptography to cleverly
disguise all electronic communications. For example, someone might choose to encrypt
a text message by using a designated cryptosystem and key before he or she sends it to a
contact. As for the cryptosystems involved, someone could choose from a variety of
cryptosystems such as: Hill cipher; transposition cipher; one-time pad; the Ron Rivest,
Adi Shamir, and Leonard Adleman algorithm (RSA); and Advanced Encryption Standard
(AES).

4.1 Real life examples of al-Qaeda encryption software: Mujahideen Secrets
أسرار الجهادين

An al-Qaeda affiliated group, called the Global Islamic Media Front (GIMF)
لجبهة الإعلامية الإسلامية العالمية, and an Islamic Web forum named the Islamic Faithful/
Devoted Network or al-Eklaas Islamic Network شبكة الإخلاص الإسلامية, tried to use the
Internet to distribute encryption software for the purpose of sending and receiving secret
messages. Encrypted messages were exchanged using symmetric and asymmetric
encryption methods, such as AES and RSA algorithms, respectively. For more
information on symmetric and asymmetric encryption techniques, see Trappe and

Global Islamic Media Front (GIMF)
الجبهة الإعلامية الإسلامية العالمية

The Global Islamic Media Front (GIMF) is a multi-media outlet, which is part of
al-Qaeda’s “public relations” effort. It spreads jihadi rhetoric and propaganda across
media sources such as the Internet, television, magazines, and text messages. Since the
Internet is a relatively safe place to post controversial information, GIMF uses the Internet as its primary source to distribute all types of jihad material. These materials include short video clips of insurgent attacks in Iraq and Afghanistan, first-person video games, and downloadable software. According to www.globalsecurity.org, GIMF’s sphere of influence across the Internet is primarily centered around the country of Pakistan, which gives a good indication to where GIMF’s creators and maintainers operate. One of GIMF’s most significant technical achievements in terms of computer programming was the release of a free encryption tool called Mujahideen Secrets in 2007 (Danchev, 2008; Vijayan, 2008).

*Mujahideen Secrets 1 (MS1) - أسْرَارُ الْمُجاهِدِينَ ١ - first edition.*

We refer to this version of Mujahideen Secrets as MS1 for clarity purposes. MS1 gives users the option and capability to encrypt email messages using either symmetric or asymmetric encryption. The symmetric, or private-key, encryption process uses AES algorithms with a Rijndael block length of 256 bits, while the asymmetric, or public-key, encryption process uses RSA algorithms with a 2048-bit length (Danchev, 2008). Some other features of MS1 are listed below.

MS1 Features (Danchev, 2008):
- Uses the best five cryptographic algorithms in cryptography (AES finalist algorithms)
- Ultra strong symmetric encryption
- Pressure data ROM (the highest levels of pressure)
- Keys and encryption algorithms with ghost changing technology (stealthy cipher)
- Automatic identification decryption algorithm (cipher auto-detection)
- Program consisting of one facility file - does not need assistance to install and can run from portable memory
• Files can be erased without the possibility of retrieving the files (file shredder)

Islamic Faithful/Devoted Network or Al-Ekhlaas Islamic Network

The Islamic Faithful Network, or al-Ekhlaas, was a popular U.S. based Islamic Web forum created by a man named Peterson Hoffman on January 22, 2006, in Amman, Jordan (Altman, 2008). The website was first located on a server in Rochester, Minnesota and was password protected (Henry (2008c)). In early January 2008, the site released what is known as, Mujahideen Secrets 2 or MS2 (author’s abbreviation), the second version of Mujahideen Secrets (MS1). MS2 was discovered by J.M. Berger, a Cambridge, Massachusetts-based freelance journalist and documentary film-maker who specializes in terrorism (Techworld Staff, 2008). Shortly after MS2’s release, al-Ekhlaas was “arguably the single most important al’Qaeda Web site in operation” with an estimated 17 million total visits (Altman, 2008). The website then moved to a server in Tampa, Florida, which was owned by a company called Noc4hosts, Inc. Once Noc4hosts, Inc. received word of jihad activity from authorities and computer security experts, the company quickly disabled and removed the site. After the site’s second shutdown, it moved again to a server in Phoenix, Arizona where it was removed for a third time (Techworld Staff, 2008). It is unclear as to why GIMF released MS1, but did not release MS2.

Mujahideen Secrets 2 (MS2) - second edition.

Like its predecessor, MS2 could be downloaded for free and created a lot of chatter on the Internet among jihadists. In addition, both MS1 and MS2 allow users to
type Arabic script, as opposed to transliterated Arabic, which is often used on non-Arabic compatible devices and applications such as cell phones and instant messaging. For example, the Arabic word أسْرَار, meaning secrets, is pronounced “asraar”, therefore its English transliteration, keystroke, or typed equivalent is “asraar.”

MS2 is an extension of MS1 and provides a number of new applications and features, which have been integrated into the older software. Some of these new applications and features include increased portability and RSA security, the production of digital signatures, and the ability to encrypt Yahoo and MSN instant messages (Techworld Staff, 2008). The increase of MS2’s portability makes it easier to distribute the software via a USB device for example. Below is a complete list of all new features offered by MS2.

MS2 Features (Danchev, 2008):
• Multicast encrypted via text messages supporting the immediate use forums (Secure Messaging)
• Transfer files of all kinds to be shared across text forums (Files to Text Encoding)
• Production of digital signatures
• Digital signatures of messages and files for authenticity
• Upgraded graphical user interface
• Improved RSA security encryption codes in order to handle functions such as key generation and key management

According to Paul Henry (2008a), MS2 also has, “the ability to take a binary file and encrypt it in such a way that the file can be posted in a pure ASCII or text-only format.” Moreover, “as a result, individuals could use MS2 to encrypt files and post them on sites that aren’t even on the Internet - for instance, on a telephone-accessed bulletin board system” (Techworld Staff, 2008). Additionally, MS2 is well-written,
harder to crack than MS1, and has fewer coding mistakes (Henry, 2008a). The files also have a unique fingerprint, which can be used for tracking (Henry, 2008a).

It seems that all websites offering the encryption tools MS1 and MS2 have been currently disabled. According to Paul Henry (2008b), the codes for MS1 were broken by law enforcement authorities. It is unclear whether MS2 codes have been broken. The purpose behind al-Qaeda’s creation of both MS1 and MS2 is not difficult to explain; al-Qaeda members desire a possible means of communicating with one another without exposing their plans of warfare to authorities. More specifically, the software was created, “in order to support the Mujahideen in general and the Islamic State in Iraq in particular” (Reuters Group, 2008). However, there is also a subtle purpose to the creation of the software from a technical standpoint. One computer security expert states that, “al-Qaeda supporters are building their own encryption tools because they are concerned that industry built and public encryption software contain back doors” (Henry, 2008c). Therefore, it is not far fetched to say that al-Qaeda is trying to adapt to the technological complexities of today’s world in order to stay active and covert.

4.2 Examples of CEN(C).

*Hill cipher: Discussion.*

The Hill cipher is a classical cryptosystem and the following example parallels the same Hill cipher technique found in Trappe and Washington (2006). However, this particular Hill cipher uses Modern Standard Arabic (MSA) as well as the author’s own set of numerical values.
Iraq - Ibrahim would like to text message his friend Yassin (who lives in Yemen) to tell him about some discussions that took place last week among some key members of their political party. Ibrahim knows that Yassin will relay this information to other members in the Yemen faction. Ibrahim, however, is afraid that the text message will be intercepted by some members of an opposing party, so he decides to encrypt his message using the Hill cipher.

Ibrahim’s message m is,

\[
\text{جرت في الأسبوع الماضي محادثات هامة}
\]


*Important discussions took place last week.*

For the Hill cipher, Ibrahim chooses an \( n \times n \) matrix \( M \) for a key whose entries are integers mod 36 and \( \gcd(\det(M), 36) = 1 \). Ibrahim chooses the matrix,

\[
M = \begin{bmatrix}
3 & 14 & 27 \\
16 & 4 & 17 \\
10 & 33 & 3 \\
\end{bmatrix}
\]

such that \( n = 3 \). Ibrahim then checks if \( M \) is invertible modulo 36. If \( M \) is invertible mod 36, then Ibrahim knows that his friend Yassin will be able to decrypt the ciphertext into the original plaintext. In order to do this, Ibrahim computes \( \det(M) \equiv 13237 \pmod{36} \equiv 25 \) and checks if \( \gcd(\det(M), 36) = \gcd(25, 36) = 1 \). Therefore, Ibrahim knows that there exists an inverse \( N \) such that \( MN \equiv I \pmod{36} \), where \( I \) is the identity matrix.

It follows that the inverse of \( M \) is,
Here, \((13237) \mod 36 \equiv 25\). Also, Ibrahim replaces \((1/25)\) or \(25^{-1}\) with 13, since 13 is the inverse of 25 mod 36. Therefore, after multiplying each entry above by the scalar 13, in addition to reducing each entry by modulo 36, Ibrahim gets,

\[
\begin{array}{ccc}
-549 & 849 & 130 \\
(1/13237) & 122 & -261 & 381 \\
488 & 41 & -212 \\
\end{array}
\]

\[
\begin{array}{ccc}
-549 & 849 & 130 \\
(1/25) & 122 & -261 & 381 \\
488 & 41 & -212 \\
\end{array}
\]

Ibrahim now checks that,

\[
N = \begin{bmatrix}
27 & 21 & 34 \\
2 & 27 & 21 \\
8 & 29 & 16 \\
\end{bmatrix}
\]

\[
MN = \begin{bmatrix}
3 & 14 & 27 \\
16 & 4 & 17 \\
10 & 33 & 3 \\
\end{bmatrix} \begin{bmatrix}
27 & 21 & 34 \\
2 & 27 & 21 \\
8 & 29 & 16 \\
\end{bmatrix} = \begin{bmatrix}
27 & 21 & 34 \\
2 & 27 & 21 \\
8 & 29 & 16 \\
\end{bmatrix} \equiv I \pmod{36}
\]

To encrypt his message, Ibrahim decomposes his message \(m\) into a series of row vectors of length \(n\) and multiplies each row vector by matrix \(M\) on the right. The entries of the row vectors are read right to left.
Sitting in a café, Ibrahim sends the matrix key directly to Yassin in Yemen using his computer. Ibrahim then sends the ciphertext separately as a cell phone text message to his friend Alliya who lives in Saudi Arabia. Alliya then relays this text message on his cell phone to his friend Fadil in Saudi Arabia. Once again, Fadil texts this ciphertext to another person in Saudi Arabic who ultimately text messages it to Yassin in Yemen.
Figure 23. The discussion network.

<table>
<thead>
<tr>
<th>Location 1 Data Fields</th>
<th>Location 2 Data Fields</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Location (country - Iraq, town/place - unknown, description - in a café)</td>
<td>• Location (country - Saudi Arabia, town/place - unknown, description - none)</td>
</tr>
<tr>
<td>• Persons of interest (Ibrahim - friend to Alliya and Yassin)</td>
<td>• Persons of interest (Alliya - Ibrahim’s friend)</td>
</tr>
<tr>
<td>• Covert electronic communication methods (description - electronics available, hardware - cell phone and computer, cryptography - Hill cipher - Ibrahim sends matrix key directly to Yassin in Yemen using his computer, Ibrahim then sends the ciphertext separately in a text message to his friend Alliya in Saudi Arabia)</td>
<td>• Covert electronic communication methods (hardware - cell phone, cryptography - receiver of Hill cipher text message ciphertext from Ibrahim, Alliya then text messages the ciphertext to his friend Fadil in Saudi Arabia)</td>
</tr>
<tr>
<td>Location 3 Data Fields</td>
<td>Location 4 Data Fields</td>
</tr>
<tr>
<td>------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>• Location (country - Saudi Arabia, town/place - unknown, description - none)</td>
<td>• Location (country - Saudi Arabia, town/place - unknown, description - none)</td>
</tr>
<tr>
<td>• Persons of interest (Fadil - friend of Alliya)</td>
<td>• Persons of interest (person - unknown, friend to Fadil and Yassin)</td>
</tr>
<tr>
<td>• Covert non-electronic communication methods (description - electronics available, hardware - cell phone, cryptography - receives the Hill cipher ciphertext from Alliya and text messages it to an unknown friend in Saudi Arabia)</td>
<td>• Covert non-electronic communication methods (description - electronics available, hardware - cell phone, cryptography - receives the Hill cipher ciphertext from Fadil and text messages it to Yassin in Yemen)</td>
</tr>
</tbody>
</table>
In practice, the size of \( n \) for Ibrahim’s original \( n \times n \) key matrix \( M \) may in fact be considerably larger, thus requiring more computational power for calculations. A powerful, yet affordable computer program such as MATLAB, may be used as a quick and efficient supplement for tedious hand calculations involving modular arithmetic and inverses of matrices.

**Transposition cipher: The Yemeni camp.**

The transposition cipher is a classical cryptosystem and the following example parallels the same transposition cipher technique described in Menezes et al. (1997). However, this cipher uses MSA as well as this author’s own set of numerical values. Note, the term transposition is synonymous with the term permutation.

Yemen - Yassin has been told by his superiors that he must report back to Ibrahim about the movement of troops in a particular Yemeni camp over the last three months. In addition, he must describe how many troops are attending the camp and where they are coming from. Yassin gathers this information from a few of his sources and composes his message.

Yassin’s message \( m \) is,

\[
\text{يصل المعسكر اليمني كل يوم عدد كبير من المقاتلين من داخل البلاد وخارجها}
\]


*Every day the Yemeni camp receives a large number of fighters from within and from outside the country.*
Yassin encrypts his message using the transposition cipher. For the transposition cipher, Yassin considers a finite set \( S = \{k, \ldots, 3, 2, 1\} \), where \( k \) represents the cardinality of \( S \). The elements of the set \( S \) are listed in decreasing order because Arabic is read right to left. Yassin chooses length \( k = 5 \), therefore \( S = \{5, 4, 3, 2, 1\} \). Yassin then selects a permutation of \( S \) as the key; for example, the permutation given by,

\[
p(5) = 2, \ p(4) = 1, \ p(3) = 5, \ p(2) = 3, \ p(1) = 4
\]

\[
p = \begin{array}{ccccc} 
5 & 4 & 3 & 2 & 1 \\
2 & 1 & 5 & 3 & 4 \\
\end{array}
\]

\[
p^{-1} = \begin{array}{ccccc} 
5 & 4 & 3 & 2 & 1 \\
3 & 1 & 2 & 5 & 4 \\
\end{array}
\]

*Figure 24. Transposition cipher encryption.*

<table>
<thead>
<tr>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>٥</td>
<td>٤</td>
<td>٣</td>
<td>٢</td>
<td>١</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>٥</th>
<th>٤</th>
<th>٣</th>
<th>٢</th>
<th>١</th>
</tr>
</thead>
<tbody>
<tr>
<td>٥</td>
<td>٤</td>
<td>٣</td>
<td>٢</td>
<td>١</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>٥</th>
<th>٤</th>
<th>٣</th>
<th>٢</th>
<th>١</th>
</tr>
</thead>
<tbody>
<tr>
<td>٥</td>
<td>٤</td>
<td>٣</td>
<td>٢</td>
<td>١</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>٥</th>
<th>٤</th>
<th>٣</th>
<th>٢</th>
<th>١</th>
</tr>
</thead>
<tbody>
<tr>
<td>٥</td>
<td>٤</td>
<td>٣</td>
<td>٢</td>
<td>١</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>٥</th>
<th>٤</th>
<th>٣</th>
<th>٢</th>
<th>١</th>
</tr>
</thead>
<tbody>
<tr>
<td>٥</td>
<td>٤</td>
<td>٣</td>
<td>٢</td>
<td>١</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>٥</th>
<th>٤</th>
<th>٣</th>
<th>٢</th>
<th>١</th>
</tr>
</thead>
<tbody>
<tr>
<td>٥</td>
<td>٤</td>
<td>٣</td>
<td>٢</td>
<td>١</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>٥</th>
<th>٤</th>
<th>٣</th>
<th>٢</th>
<th>١</th>
</tr>
</thead>
<tbody>
<tr>
<td>٥</td>
<td>٤</td>
<td>٣</td>
<td>٢</td>
<td>١</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>٥</th>
<th>٤</th>
<th>٣</th>
<th>٢</th>
<th>١</th>
</tr>
</thead>
<tbody>
<tr>
<td>٥</td>
<td>٤</td>
<td>٣</td>
<td>٢</td>
<td>١</td>
</tr>
</tbody>
</table>
Yassin sends his encrypted message over his cell phone and tells his friend to send the key to Ibrahim in a particular Internet chat room.

**One-time pad: A sensitive conversation.**

The one-time pad is a classical cryptosystem and the following example parallels the same one-time pad cipher technique found in Trappe and Washington (2006). However, this cipher uses MSA as well as the author’s own set of numerical values. The term one-time pad is synonymous with the term Vernam cipher.

Iraq - Yassin and Ibrahim are just finishing up a conversation on what type of design they would like to incorporate into the flag of their political party. They know that the design of their new flag could cause unrest and stir hatred among other opposing parties and for this reason, Yassin and Ibrahim decide to encrypt their messages using a series of one-time pads.

Ibrahim’s last message m to Yassin is,

شكرا. أستطيع أن نتابع بحث العلم غدا؟


Thanks. Can we continue the discussion of the flag tomorrow?
For the one-time pad, Ibrahim converts his Arabic message $m$ into binary or a string of 1s and 0s, whereby the binary version of $m$ is also read right to left. In order to do this, Ibrahim assigns a unique sequence of one or more bits, called a bitstring, to each letter of MSA and to any other symbol he wishes to encrypt. In this case, Ibrahim chooses bit string combinations of length seven. This will allow Ibrahim to assign $2^7 = 128$ possible bitstring combinations to different text values. Once this has been done, Ibrahim creates what is known as a look-up table, which is a written or an electronic document used by two or more network parties during the coding and decoding process, to convert all textual characters to their respective numerical values, and vice-versa. In this case, the numerical values are bitstrings. This means that Ibrahim will somehow have to physically send or electronically communicate an identical copy of his look-up table to Yassin before Yassin is able to successfully convert the binary plaintext back into ordinary text. Yassin will also need a key in this case since they are using one-time pads.

Ibrahim and Yassin’s binary look-up table for MSA and a few other symbols is provided below.

*Figure 25. Random binary representations.*

<table>
<thead>
<tr>
<th>Arabic</th>
<th>Binary</th>
<th>Arabic</th>
<th>Binary</th>
</tr>
</thead>
<tbody>
<tr>
<td>ا</td>
<td>0101000</td>
<td>غ</td>
<td>100011</td>
</tr>
<tr>
<td>ب</td>
<td>0110000</td>
<td>ف</td>
<td>0100100</td>
</tr>
<tr>
<td>ت</td>
<td>0111000</td>
<td>ق</td>
<td>0101100</td>
</tr>
<tr>
<td>ث</td>
<td>1000000</td>
<td>ك</td>
<td>0110100</td>
</tr>
<tr>
<td>ج</td>
<td>0100001</td>
<td>ل</td>
<td>0111100</td>
</tr>
<tr>
<td>ح</td>
<td>0101001</td>
<td>م</td>
<td>1000100</td>
</tr>
</tbody>
</table>
Having converted his message $m$ to a binary string, Ibrahim then generates a random string of binary with length equivalent to $m$ for the key. For simplicity concerning hand calculations, Ibrahim assumes that the binary key for his message is random. Ibrahim then performs the bit operation **bitwise exclusive or**, frequently denoted **XOR**, on the two strings. This bit operation is provided in detail in the table below.

<table>
<thead>
<tr>
<th>X</th>
<th>0110001</th>
<th>N</th>
<th>0100101</th>
</tr>
</thead>
<tbody>
<tr>
<td>د</td>
<td>0111001</td>
<td>ه</td>
<td>0101101</td>
</tr>
<tr>
<td>ن</td>
<td>1000001</td>
<td>و</td>
<td>0110101</td>
</tr>
<tr>
<td>ر</td>
<td>0100010</td>
<td>ي</td>
<td>0111101</td>
</tr>
<tr>
<td>ز</td>
<td>0101010</td>
<td>أ</td>
<td>1000101</td>
</tr>
<tr>
<td>س</td>
<td>0110010</td>
<td>إ</td>
<td>0100110</td>
</tr>
<tr>
<td>ش</td>
<td>0111010</td>
<td>أ</td>
<td>0101110</td>
</tr>
<tr>
<td>ص</td>
<td>1000010</td>
<td>ء</td>
<td>0110110</td>
</tr>
<tr>
<td>ض</td>
<td>0100011</td>
<td>ي</td>
<td>0111110</td>
</tr>
<tr>
<td>ط</td>
<td>0101011</td>
<td>ئ</td>
<td>1000110</td>
</tr>
<tr>
<td>ظ</td>
<td>0110011</td>
<td>ة</td>
<td>0100111</td>
</tr>
<tr>
<td>ع</td>
<td>0111011</td>
<td>ﺩ</td>
<td>0101111</td>
</tr>
<tr>
<td>ﺔ</td>
<td>0110111</td>
<td>0111111</td>
<td></td>
</tr>
</tbody>
</table>

**Exclusive Or/XOR**

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>$x \oplus y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Note that XOR is the same as adding corresponding bits mod 2, therefore Ibrahim writes the key directly below the message and adds them together mod 2. The result is a binary string of ciphertext. This bit operation can be performed even faster with a computer.

*Figure 26. One-time pad encryption.*
The task of assigning letters different bitstrings is similar to substitution in the sense that in substitution, one assigns different integer values to the letters of the alphabet. Look-up tables, in general, do not always have to contain numerical values assigned to respective text characters. That is to say, one may also substitute entire words with numerical or text values. Such a table is called a dictionary look-up table.

Given that a particular one-time pad is unbreakable, it may be possible for authorities to record keystrokes or use computer forensics software to retrieve unknown keys.

**Audio steganography: The call to prayer.**

Yemen - Yassin would like to send a one-time pad cipher (on p. 76) using steganography to his friend Ibrahim who is living in Iraq. Yassin will send the key for the one-time pad at a later date in the mail to Ibrahim disguised as a series of barcodes on some new pirated DVDs (see p. 76 for the key; also pirated DVDs and pirated DVD covers are very common in the Middle East and North Africa and can usually be bought
on the street for about one American dollar). Yassin decides to use a freeware
steganography CD-ROM program that accompanies the book *Hiding In Plain Sight*
(Cole, 2003) to embed the cipher into an audio file, or more specifically, a WAV song file.
Steps 1-9 show the process of how Yassin accomplishes this task.

*Step 1.*

Yassin installs the main steganography program, called stego, from the *Hiding In
Plain Sight* CD-ROM onto his computer. Yassin will run routines 9 and 10 from the
stego program on the command-line prompt for the encoding and decoding processes,
respectively. Routine 9 is called WAV Twiddle encode and routine 10 is called WAV
Twiddle decode.

*Step 2.*

Before running routines 9 and 10, Yassin uses the Internet to download a free
MP3 Muslim prayer song, *Call to prayer in Jakarta* (djbacon, 2004), from Masjid
Baiturrahman Bintaro, a mosque located in Jakarta, Indonesia. He then converts his MP3
prayer song to a WAV file, since the steganography program only works with WAV files.
This song is saved as Masjid_Baiturrahman_Bintaro_Jakarta_0.wav (this may be done
with any freeware MP3 WAV converter available on the Internet). To start the encoding
process, Yassin writes and saves his one-time pad into a .txt file called covert.txt.

*Step 3.*

Yassin saves the covert.txt file, the song .wav file
Masjid_Baiturrahman_Bintaro_Jakarta_0.wav, and the stego application or program in
the file called StegoTwiddle.
**Step 4.**

Yassin runs routine 9 to encode or embed his one-time pad message in the covert.txt file into the new .wav file called new_song_twiddle.wav. This new song file will contain Yassin’s covert.txt message and will sound identical to the original song. That is, the tampering of the sound of this new song with the added covert.txt data will be indiscernible to the human ear.

**Step 5.**

The new song (new_song_twiddle) containing the hidden message can now be found in the StegoTwiddle file.

**Step 6.**

Yassin checks that he can extract the covert.txt data from the new song by running routine 10. This will ensure that Ibrahim will be able to successfully extract or decode the covert.txt from the new song. In order to accomplish this task, Yassin must create a new .txt file called covertdecoded.twiddle.txt. This new file will receive the covert.txt data from the new song during the extraction process and not the actual covert.txt file.

**Step 7.**

The covertdecoded.twiddle.txt file can now be found in the StegoTwiddle folder.

**Step 8.**

Opening the new covertdecoded.twiddle.txt file reveals the original message or one-time pad.
**Step 9.**

A comparison of the two songs in a Hex Editor program reveals that the two songs are in fact different. This proves that the new song contains the original covert.txt message.

*Figure 27. Hex editor: Song comparison.*
Step 10.

Yassin sends a box of new pirated DVDs to Ibrahim in the mail with barcodes containing sections of the key of the one-time pad. The last two digits of the barcodes reveal the order in which they must be placed to form the original key.

**RSA: The harbor.**

The following RSA algorithm parallels the same RSA algorithm found in Trappe and Washington (2006). However, this RSA example uses MSA as well as this author’s own set of numerical values.

Iraq - Ibrahim has informed Yassin that he has discovered a new and more secure way to encrypt their messages online using a public key cryptosystem called the RSA algorithm. For the RSA algorithm, Ibrahim follows four simple preliminary steps:

*Preliminary steps (Trappe & Washington, 2006).*

i. Ibrahim chooses two distinct primes, say p = 4789 and q = 2971, and computes their product.

\[ n = pq = 1422819 \]

Ibrahim keeps these two distinct primes secret, since their relationship uniquely determines the procedures of encryption and decryption conducted by Yassin and Ibrahim, respectively.

ii. Next, Ibrahim computes \( \phi(n) = (p - 1)(q - 1) = (4788)(2970) = 14220360 \) and selects an encryption exponent e such that e and \( \phi(n) \) are relatively prime. In order for e and \( \phi(n) \) to be relatively prime, gcd(e, \( \phi(n) \)) = 1. Ibrahim chooses e = 765717 and checks that gcd(765717, 14220360) = 1 using the Euclidean algorithm. For the Euclidean algorithm, Ibrahim divides the larger of the two numbers by the other number
to give a quotient $q_1$ and a remainder $r_1$. Ibrahim assumes $\varphi(n)$ is greater than $e$. This is written as $\varphi(n) = q_1 \cdot e + r_1$. Here, if $e$ perfectly divides $\varphi(n)$, then of course $r_1 = 0$. If $\varphi(n) > 0$ and $e = 0$, then $q_1 = 0$ and $r_1 = 0$. Ibrahim assumes that neither $\varphi(n)$ or $e$ is equal to zero. If $r_1 \neq 0$, then the remainder $r_1$ is clearly able to divide the number $e$ from the original expression $\varphi(n) = q_1 \cdot e + r_1$, such that there exists a new quotient $q_2$ and a new remainder $r_2$. This is due to the fact that $r_1 < e$, whereby $r_1$ is indeed the remainder of $\varphi(n)$ divided by $e$. By continuation of this quotient and remainder process, the expression for the smaller number $e$ is similarly written as $e = q_2 \cdot r_1 + r_2$. The next expression for $r_1$ is given by $r_1 = q_3 \cdot r_2 + r_3$. This process is continued until a remainder of zero is achieved. Ibrahim’s Euclidean algorithm calculations are given below:

\begin{align*}
14220360 &= 1 \cdot 7657117 + 6563243 \\
7657117 &= 1 \cdot 6563243 + 1093874 \\
6563243 &= 5 \cdot 1093874 + 1093873 \\
1093874 &= 1 \cdot 1093873 + 1 \\
1093873 &= 1093873 \cdot 1 + 0
\end{align*}

Note that the last nonzero remainder is the greatest common divisor.

iii. Ibrahim then computes $d$ such that $de \equiv 1 \pmod{(p - 1)(q - 1)} \equiv 1 \pmod{\varphi(n)}$, [$d$ is the inverse of $e$ modulo $\varphi(n)$], using the **Extended Euclidean algorithm**. Since $e = 7657117$ and $\varphi(n) = 14220360$ are relatively prime, there exists integers $x$ and $y$ such that,

\[ ex + \varphi(n)y = \gcd(x, y) = 1 \]

For the Extended Euclidean algorithm, Ibrahim sets $x_0 = 0$, $x_1 = 1$. 

81
and defines $x_j = - q_{j-1} x_{j-1} + x_{j-2}$ for $1 \leq q_i \leq n$, where $q_i$ are the quotients found in the Euclidean algorithm:

$$ex_n + \varphi(n)y_n = \gcd(a, b) = 1$$

The quotients from the Euclidean algorithm in ii. are $q_1 = 1$, $q_2 = 1$, $q_3 = 5$, $q_4 = 1$, $q_5 = 1$ (from top to bottom).

Setting $x_0 = 0$, $x_1 = 1$, Ibrahim gets,

- $x_2 = -x_1 + x_0 = -1$
- $x_3 = -x_2 + x_1 = 2$
- $x_4 = -5x_3 + x_2 = -11$
- $x_5 = -x_4 + x_3 = 13$

Ibrahim now knows that $ex_n + \varphi(n)y_n = 7657117 \cdot x_5 + 14220360 \cdot y_5 = \gcd(x_5, y_5) = 1$, such that $x_5 = 13$. Therefore, Ibrahim shows that $d = 13$ is the inverse of $e$ modulo $\varphi(n)$,

$$7657117 \cdot x_5 + 14220360 \cdot y_5 = 1$$
$$7657117 \cdot 13 + 14220360 \cdot y_5 = 1$$

$$(7657117 \cdot 13 + 14220360 \cdot y_5) \equiv 1(\text{mod } 14220360)$$
$$7657117 \cdot 13 \equiv 1(\text{mod } 14220360)$$

Hence, $7657117d \equiv 1(\text{mod } 14220360)$, where $d \equiv 13 (\text{mod } 14220360)$.

iv. Ibrahim keeps his **private key**, $(p, q, d) = (4789, 2971, 13)$, secret and sends his **public key**, $(e, n) = (7657117, 14228119)$, to Yassin in Yemen.

*Encryption (Trappe & Washington, 2006).*

Yemen - Yassin and his friend Yusef have just returned from a small trip to the docks of one of the city’s many harbors. During their trip, they met with an old man who showed them an inexpensive piece of real estate near the waterfront. Eager to report back to Ibrahim about his trip, Yassin sits down at his computer and downloads Ibrahim’s public key, $(e, n) = (7657117, 14228119)$. He then types the following MSA message.
Yassin’s message $m$,

ذهبنا به إلى المياء في الصباح. لا شيء


*We went with him to the harbor in the morning. There is nothing.*

To encrypt his message $m$ using the RSA algorithm, Yassin uses the formula,

$$c \equiv m^e \pmod{n}$$

$$c = m^{765717} \pmod{1422819}$$

1. First labeling attempt.

Next, in order to represent his message $m$ in numeric form, Yassin labels the MSA alphabet and its special characters using the numeric values given below.

<table>
<thead>
<tr>
<th>Arabic</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>ا</td>
<td>1</td>
</tr>
<tr>
<td>ب</td>
<td>2</td>
</tr>
<tr>
<td>ت</td>
<td>3</td>
</tr>
<tr>
<td>ث</td>
<td>4</td>
</tr>
<tr>
<td>ج</td>
<td>5</td>
</tr>
<tr>
<td>ح</td>
<td>6</td>
</tr>
<tr>
<td>خ</td>
<td>7</td>
</tr>
<tr>
<td>د</td>
<td>8</td>
</tr>
<tr>
<td>ذ</td>
<td>9</td>
</tr>
<tr>
<td>ر</td>
<td>10</td>
</tr>
<tr>
<td>س</td>
<td>11</td>
</tr>
<tr>
<td>ز</td>
<td>12</td>
</tr>
<tr>
<td>ق</td>
<td>13</td>
</tr>
<tr>
<td>ش</td>
<td>14</td>
</tr>
<tr>
<td>ص</td>
<td>15</td>
</tr>
<tr>
<td>ض</td>
<td>16</td>
</tr>
<tr>
<td>ط</td>
<td>17</td>
</tr>
<tr>
<td>ظ</td>
<td>18</td>
</tr>
<tr>
<td>ع</td>
<td>19</td>
</tr>
<tr>
<td>غ</td>
<td>20</td>
</tr>
<tr>
<td>ف</td>
<td>21</td>
</tr>
<tr>
<td>ك</td>
<td>22</td>
</tr>
<tr>
<td>ل</td>
<td>23</td>
</tr>
<tr>
<td>م</td>
<td>24</td>
</tr>
<tr>
<td>ن</td>
<td>25</td>
</tr>
<tr>
<td>ه</td>
<td>26</td>
</tr>
<tr>
<td>و</td>
<td>27</td>
</tr>
<tr>
<td>ي</td>
<td>28</td>
</tr>
</tbody>
</table>

Yassin, however, notices that the word “ذهبنا” from his message loses the final character “ا” during encryption because its value is zero. This is shown below.

<table>
<thead>
<tr>
<th>1st labeling attempt trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>0241258</td>
</tr>
<tr>
<td>$c \equiv 241258^{765717} \pmod{14228119}$</td>
</tr>
</tbody>
</table>
2. Second labeling attempt.

To solve this problem, Yassin relabels the MSA Alphabet and its special characters starting at one.

<table>
<thead>
<tr>
<th>الابن</th>
<th>ثأر</th>
<th>حج</th>
<th>ب،</th>
<th>گ</th>
<th>د</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>27</td>
<td>26</td>
<td>25</td>
<td>24</td>
<td>23</td>
</tr>
</tbody>
</table>

This time, Yassin notices that although he solves the problem of the disappearing “١”, from the word “ذهيناً”, another phrase from his message containing the same number of characters with an “١”, namely the phrase “لا شيء”, appears to have a different number of digits during encryption. Yassin suspects that this will cause future encryption and decryption problems.

<table>
<thead>
<tr>
<th>2nd labeling attempt trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>1252269</td>
</tr>
<tr>
<td>(c = 1252269^{7657117} \pmod{14228119})</td>
</tr>
<tr>
<td>322813123</td>
</tr>
<tr>
<td>(c = 322813123^{7657117} \pmod{14228119})</td>
</tr>
</tbody>
</table>

3. Third labeling attempt.

To solve this problem, Yassin pads the first nine digits with zeros.
This time, however, Yassin notices that he again has the problem of a disappearing digit resulting from the character “١” for the word “ذهينا”. This is due to the fact that Arabic is read right to left. The “١” character remains, but one of its digits is lost during encryption. Likewise, the loss of a zero digit occurs for the next eight padded characters.

<table>
<thead>
<tr>
<th>3rd labeling attempt trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>0125022609</td>
</tr>
<tr>
<td>c = ٢٥٠٢٢٦٠٩(7657117\pmod{14228119})</td>
</tr>
<tr>
<td>3228130123</td>
</tr>
<tr>
<td>c = ٣٢٨١٣٠١٢٣(7657117\pmod{14228119})</td>
</tr>
</tbody>
</table>

4. Fourth labeling attempt.

To solve the problems of disappearing digits and equal digit length, Yassin starts by labeling the alphabet with the number 11.
To simplify his calculations, Yassin groups three consecutive letters at a time from his original message $m$ to form equal number blocks of length six. He then uses the RSA encryption formula to encrypt each block separately. This process is given by the revised RSA block encryption formula,

\[ c_i = m_i^e \pmod{n} \]
\[ c_i = m_i^{765717} \pmod{14228119} \]

such that $m_i$ represents the numeric plaintext form of block $i$ and $c_i$ represents the corresponding ciphertext. The blocks are ordered from right to left.

**Figure 28. RSA encryption.**

<table>
<thead>
<tr>
<th>plaintext</th>
<th>2</th>
<th>334036</th>
<th>6500748</th>
<th>5</th>
<th>304211</th>
<th>13819400</th>
<th>8</th>
<th>113316</th>
<th>11202703</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>121135</td>
<td>1452031</td>
<td>4</td>
<td>353834</td>
<td>12083064</td>
<td>7</td>
<td>111224</td>
<td>2013713</td>
<td>423823</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>123619</td>
<td>4793407</td>
<td>3</td>
<td>331144</td>
<td>12717422</td>
<td>6</td>
<td>331138</td>
<td>9854143</td>
<td>2202791</td>
</tr>
</tbody>
</table>

86
Yassin and Yusef send the ciphertext in an email to Ibrahim, whereby Ibrahim decodes each ciphertext block individually using his private key, \((p, q, d) = (4789, 2971, 13)\), and the RSA block decryption formula,

\[
m_i \equiv c_i^d \pmod{n}
\]

\[
m_i \equiv c_i^{13} \pmod{14228119}
\]

The general RSA decryption formula without block notation is given by \(m \equiv c^d \pmod{n}\).

Ibrahim decrypts the first ciphertext block:

\[
m_0 \equiv c_0^{13} \pmod{14228119}
\]

\[
m_0 \equiv 4793407^{13} \pmod{14228119}
\]

\[
m_0 \equiv 123619
\]

The security of RSA rests in the fact that it is extremely difficult to factor large \(n\).

Consequently, it is hard to correctly determine the decryption exponent \(d\) from the formula \(de \equiv 1 \pmod{(p - 1)(q - 1)}\), where \((p - 1)(q - 1) = \varphi(n)\).

**Applying cryptanalysis to the harbor.**

We are given \(e = 7657117\) and \(n = 14228119\), since \(e\) and \(n\) were made public by Ibrahim. Our goal is to factor \(n\). In order to do accomplish this task, we need two theorems and a definition.
Theorem 1 (from Trappe and Washington (2006)).

Suppose \( p, q \) are primes such that \( q < p < 2q \). Let \( n = pq \) and let \( 1 \leq d, e < \phi(n) \), whereby

\[
\phi(n) = (p - 1)(q - 1)
\]

and \( de \equiv 1 (\text{mod} \ \phi(n)) \). If \( d < \frac{1}{3} n^{1/4} \), then \( d \) can be found by quick calculation in time that is polynomial in \( \log n \).

Proof (Trappe & Washington, 2006).

Multiplying both sides of the inequality \( q < p \) by \( q \) we get,

\[
q^2 < pq
\]

Since \( n = pq \),

\[
q^2 < n
\]

This means, \( q < \sqrt{n} \)

Now subtracting \( \phi(n) \) from both sides of the equality \( n = pq \) we obtain,

\[
n - \phi(n) = pq - \phi(n) = pq - (p - 1)(q - 1) = pq - pq + q + p - 1 = q + p - 1 \quad (1)
\]

Since \( p < 2q \) we know that,

\[
q + p - 1 < q + 2q - 1 < 3q < 3 \sqrt{n} \quad (2)
\]

Now take \( ed = 1 + \phi(n)k \) for some integer \( k \geq 1 \), [there always exists an integer \( k \geq 1 \) for \( ed = 1 + \phi(n)k \) since \( d \) is the inverse \( e \) modulo \( \phi(n) \)], then \( \phi(n)k < 1 + \phi(n)k \).

Since \( e < \phi(n) \) and considering the original assumption that \( d < \frac{1}{3} n^{1/4} \) we get,

\[
\phi(n)k < ed < \phi(n)d < \frac{1}{3} \phi(n) n^{1/4}
\]
This shows that $k < \frac{1}{3} n^{1/4}$.

Taking $kn$ minus both sides of $ed = 1 + \varphi(n)k$ and looking at both (1) and (2) in addition to the result that $k < \frac{1}{3} n^{1/4}$ we get,

$$kn - ed = kn - (1 + \varphi(n)k)$$

$$= k(n - \varphi(n)) - 1 < k(n - \varphi(n)) < 3k \sqrt{n} < \frac{1}{3} n^{1/4} (3 \sqrt{n}) = n^{3/4}$$

We know that $n - \varphi(n) > 1$, since $n - \varphi(n) = q + p - 1$ with $p > 1$ and $q > 1$.

Now, since $k \geq 1$ and $1 \leq d, e < \varphi(n) < n$, we know $1 < k(n - \varphi(n))$, then clearly,

$0 < k(n - \varphi(n)) - 1$ and $0 < kn - ed$.

Therefore, dividing $0 < kn - ed < n^{3/4}$ by $dn$ in (3) and again considering the original assumption that $d < \frac{1}{3} n^{1/4}$ we get,

$$0 < \frac{k}{d} - \frac{e}{n} < \frac{1}{dn^{1/4}} < \frac{1}{3d^2}$$

The theorem states that if $d < \frac{1}{3} n^{1/4}$, then $d$ can be found by quick calculation in time polynomial in log $n$. This statement is explored further below.

**Definition: Continued fractions.**

For every rational number, $\frac{p}{q}$, such that $p$ and $q$ are positive integers, there exists a unique finite continued fraction representation given by,
such that $a_i > 0$ for $i \geq 1$ and $a_n > 1$ [see Khinchin (1997) for restrictions on $a_i$ and uniqueness related to the last element $a_n$]. Here, the set of numbers $\{a_0; a_1, a_2, \ldots, a_n\}$ are called the **partial quotients** of $p/q$ (Weisstein, 2012). In the case that $p$ and $q$ are positive integers, the partial quotients are always nonnegative integers. We will now see how to obtain the finite continued fraction representation of $\frac{p}{q}$.

Using the floor function, $\lfloor x \rfloor$, define the finite continued fraction representation of the rational number $\frac{p}{q}$ by the following recurrence relation,

Set

$$\frac{p}{q} = r_0$$

then

$$r_n = \frac{1}{r_{n-1} - \bar{a}_{n-1}}$$

where $a_n = \lfloor r_n \rfloor$

where $r_n$ are called the **remainders** of $x$ (Weisstein, 2012).

From the preceding recurrence relation we obtain,
Given the finite continued fraction representation above, we can say
\[
\frac{p}{q} = \frac{p_n}{q_n} = c_n ,
\]
where \( c_n \) is called the \textit{nth convergent} (Weisstein, 2012).

Following the recurrence relation above, it can be shown that the rational number
\[
p_k/q_k
\]
with the set of partial quotients \( \{a_0; a_1, a_2, \ldots, a_k\} \) yields a closer
approximation to \( p/q \) than the rational number \( p_j/q_j \) with partial quotients
\( \{a_0; a_1, a_2, \ldots, a_j\} \), where \( 1 \leq j < k \).

A rational number \( p_i/q_i \) such that \( 1 \leq i \leq n \) is called the \textit{ith convergent}.

It can also be shown that if \( \frac{p_n}{q_n} < 1 \), then \( a_0 = 0 \).

\textit{Theorem 2 (from Trappe and Washington (2006)).}

For positive integers \( a, b \), and the positive real number \( x \) if \( \left| \frac{a}{b} - x \right| < \frac{1}{2b^2} \),

then \( \frac{a}{b} = \frac{p_i}{q_i} \).

\textit{Breaking RSA with continued fractions (from Trappe and Washington (2006)).}

Looking at theorem 2, we can replace \( x \) by \( \frac{c}{n} \). We can also replace \( a \) and \( b \) by
\( k \) and \( d \), respectively.

Therefore, we have
Looking at theorem 1 we know that,

\[ 0 < \frac{k}{d} - \frac{e}{n} < \frac{1}{2d^2} \]

Then clearly from theorem 1,

\[ \frac{1}{3d^2} < \frac{1}{2d^2} \]

By inserting the result found in theorem 1 into the conditional statement of theorem 2 we get,

\[ \text{If } 0 < \left| \frac{k}{d} - \frac{e}{n} \right| < \frac{1}{dn^{1/4}} < \frac{1}{3d^2} < \frac{1}{2d^2} \text{, then } \frac{k}{d} = \frac{p_i}{q_i}. \]

This convergent is a direct result of the RSA algorithm’s implementation under the Theorem 1’s conditions. Using this fact, we can follow five simple steps to factor n (from Trappe and Washington (2006)):

1. Start by computing the finite continued fraction expansion of \( e/n \).
2. Calculate convergent \( p_1/q_1 \).
3. Set \( k = p_1 \) and \( d = q_1 \) and compute \( h = (ed - 1)/k \). The value \( h \) is a possible value for \( \varphi(n) \) since \( ed = 1 + \varphi(n)k \).
4. If \( h \) is not an integer, stop, repeat steps 2 and 3 for convergent \( p_2/q_2, p_3/q_3 \), and so on.
5. If \( h \) is an integer, find roots \( \lambda_1, \lambda_2 \) of the equation \( x^2 - (n - h - 1)x + n = 0 \). The equation \( x^2 - (n - \varphi(n) + 1)x + n = (x - p)(x - q) \).
Remember, we have intercepted Ibrahim’s public key which is given by \((n, e) = (14228119, 7657117)\). Furthermore, Ibrahim does not need to protect this key from other people trying to steal/view it, since the security of RSA depends upon the private key and the fact that it is hard to factor large \(n\). We then follow steps 1 through 5 to factor \(n\):

Step 1a.

Using the recurrence relation above we get,

Setting,

\[
\frac{e}{n} = \frac{7657117}{14228119} = r_0
\]

then

\[
a_0 = \lfloor r_0 \rfloor = \lfloor \frac{7657117}{14228119} \rfloor = 0
\]

\[
a_1 = \lfloor r_1 \rfloor = \lfloor \frac{1}{r_0 - a_0} \rfloor = \lfloor \frac{1}{7657117} \rfloor = 1
\]

\[
a_2 = \lfloor r_2 \rfloor = \lfloor \frac{1}{r_1 - a_1} \rfloor = \lfloor \frac{1}{1} \rfloor = 1
\]

\[
a_3 = \lfloor r_3 \rfloor = \lfloor \frac{1}{r_2 - a_2} \rfloor = \lfloor \frac{1}{1} \rfloor = 1
\]

Proceeding in this manner, we obtain the partial quotients \(\{0; 1, 1, 6, 19, 1, 433, 2, 62\}\).

Therefore, the finite continued fraction expansion of \(e/n\) is given by,

\[
\frac{7657117}{14228119} = 0 + \frac{1}{1 + \frac{1}{1 + \frac{1}{\ddots + \frac{1}{62}}}}
\]
Step 2a.

Convergent $p_1/q_1$ is given by $1/1$, so we set $p_1 = k = 1$ and $q_1 = d = 1$ and compute $h = (ed - 1)/k$.

$$h = (ed - 1)/k$$
$$h = (7657117 \cdot 1 - 1)/1$$
$$h = 7657116$$

Step 3a.

Since $h$ is an integer, we go to step 4.

Step 4a.

We must find roots $\lambda_1, \lambda_2$ of equation $x^2 - (n - h + 1)x + n$ where $h = 7657116$.

The equation with $h = 7657116$,

$$x^2 - (14228119 - 7657116 + 1)x + 14228119 = 0$$
$$x^2 - (6571004)x + 14228119 = 0$$

gives roots $\lambda_1 = -\sqrt{10794509163885} + 3285502$ and $\lambda_2 = \sqrt{10794509163885} + 3285502$, which of course are not integers, so we proceed to the next convergent.

Step 2b.

Convergent $p_2/q_2$ is given by $1/2$, so we set $k = 1$ and $d = 2$. In fact, $d$ cannot be equal to an even number, since an even number modulo another even number cannot give a remainder of 1. [This statement comes directly from Ibrahim’s original equation $7657117d \equiv 1 \pmod{14220360}$].

Step 2c.

$$\frac{1}{1 + \frac{1}{1 + \frac{1}{6}}} = \frac{7}{13}$$
Convergent \( p_3/q_3 = 7/13 \), so set \( k = 7 \) and \( d = 13 \). Computing \( h \) we get,

\[
\begin{align*}
h &= (ed - 1)/k \\
h &= (7657117 \cdot 13 - 1)/7 \\
h &= 14220360
\end{align*}
\]

Step 3c.

Since \( h \) is an integer, we proceed to step 4.

Step 4c.

Again, we must find roots \( \lambda_1, \lambda_2 \) of the equation \( x^2 - (n - h + 1)x + n \) where \( h = 14220360 \).

This time we get,

\[
\begin{align*}
x^2 - (14228119 - 14220360 + 1)x + 14228119 &= 0 \\
x^2 - (7760)x + 14228119 &= 0
\end{align*}
\]

where \( \lambda_1 = 4789 \) and \( \lambda_2 = 2971 \).

We have therefore factored \( n \). We check that product of \( \lambda_1 \) and \( \lambda_2 \) is indeed equal to \( n \).

\[
n = (4789)(2971) = 14228119
\]

We can now easily decode Yassin’s and Yusef’s cipher.

**Part II Summary**

For the sake of argument, if one removes the word “covert” from the term covert electronic networks, then perhaps one is just left with “regular” electronic networks. If so, then it is reasonable to conclude that regular electronic networks form the sea of technology interaction which drives today’s modern world. For example, the number of emails sent over the Internet everyday is enormous. Now inserting the word covert back in, it is not hard to see why the total amount of information emanating from a covert
electronic network, would only contribute to a small fraction of all daily electronic information.

If covert electronic networks are truly malicious and criminal on the terrorist level, then the scarcity of these networks, as well as their sometimes complex associated attributes, such as cryptography and steganography, does not bode well for law enforcement agencies. To combat such networks requires a deep understanding of how individuals or organizations operate at the lowest level. For instance, how they relay their data across different locations. This also goes for learning how individuals and organizations use and adapt to today’s technology trends. In this respect and with suggestion, law enforcement agencies must not only have the technical capacity to uncover such networks, but they must also have the correct foresight and creative edge to unlock the secrets of today’s covert electronic networks.
References


Appendix A

Electronic Correspondence with J. Seberry (personal communication, May 8, 2011)

Dear Adam,
The data we gave in our book was incorrect. We tried to have this changed before publication but it didn't happen.

There have been a few Saudi's interested in this but to my knowledge no better or newer data has been published. Of course it may have been published in Arabic (or Hebrew).

I do not know of anything related to specific dialects and extrapolating from English dialects I believe this could be a fruitful field of endeavour.

We originally looked extensively at Bahasa-Malaysia/Indonesia and Pin-yin for Chinese but I don't have any more recent studies. We could distinguish between classically encrypted Nordic languages.

Good researching

Jennifer Seberry

Professor of Computer Science
Foundation Professor, Centre for Computer and Information Security Research,
School of Computer Science and Software Engineering
University of Wollongong
Wollongong
NSW 2522, Australia