University of Nevada, Reno

Permutation Steganography in Many Systems

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Mathematics

by

William G. Rudebusch

Dr. Edward C. Keppelmann/Thesis Advisor

December, 2011

THE GRADUATE SCHOOL

We recommend that the thesis
prepared under our supervision by

William Rudebusch

entitled

Permutation Steganography in Many Systems

be accepted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

Edward C. Keppelmann, Ph.D., Advisor

Thomas Quint, Ph.D., Committee Member

Sushil Louis, Ph.D., Graduate School Representative

Marsha H. Read, Ph. D., Associate Dean, Graduate School

December, 2011

Abstract

The newest method we wish to offer to the world of steganography and cryptology makes use of permutations. The initial idea of permeation steganography was developed independently by Peter Hendrickson (1). Here we extended on his idea by adding composition of permutations into the reply step. We have also included cycle notation; this allows for more freedom and saves bandwidth.

Contents

List of Tables

1

Introduction

1.1 What is Steganography?

Steganography is the art and science of hiding messages in plain sight, "thereby obscuring not just its meaning but its very existence." (10)

Typically the application of steganography makes use of the wasted space inherent in all channels of communication. By wasted space I mean the places in communication where metadata exists. At any instant in time there is more information being shown to an observer than he or she is processing. For example, when reading a sentence the observer typically does not exactly count the white spaces or frequency of the letters. These properties of every-day communication could be used to hide information.

Stenography has just as many legal uses as covert. It can be used to authenticate a copy of a piece of software while not affecting the data in any decreeable way. This is called watermarking. This method is used in the industry to assure authenticity of a product.

Clandestine as well as legal activities require creativity. This is because steganography requires information to unnoticeably mix into existing forms of communication. In this paper we will be focusing on encoding and transmitting plaintext messages. Whether this is cloak-and-dagger or completely innocent is left to the reader.

A historical example of technology and steganography coming together happened in the 1980s. Prime minister Margret Thatcher thought she had a leak in her cabinet. Her idea to fix it was to slightly alter the word processors of government employees. These tampered machines now encoded the information of the author into the word spacing of

the document. Therefore making each piece of internal correspondence secretly linked to its source. Now she could trace a leaked document to is originating word processor. (2)

Due to advancing technology the applications of steganography are growing. One popular technique uses of digital photography. The least significant pixels in a picture can be slightly altered from their original state. This small change in value will not arouse suspicion in a human viewer but can be used to hide information.There are many open source computer programs that can do this operation.

A cryptologist will be quick to point out that these systems are not secure. In the cryptographic sense they are absolutely correct. Steganography relies on the fact that the enemy would first need to know where to look to be able to find anything. In cryptology the code is usually transmitted in the open. This encrypted data is available to anyone who wants to spend the resources to try and decrypt (crack) it.

Whereas it is difficult to quantify exactly how much computing time it would take to crack a steganographic system. The first problem is finding it and then checking every item for hidden data. On a any given website with frequent updates, such as a photo-centric tumblr stream, checking every new photo could become quite burdensome.

Furthermore, two parties speaking in code is typically enough to arouse suspicion. On the other hand, two parties communicating with steganography does not. This ability for encrypted communication to be carried out in an open unsuspicious fashion is yet strength of steganography.

For example, suppose that two parties wish to communicate in a prison. In this situation cryptography is not an option because code will obviously alert the officials to possible illicit activity. Hence, stenography in a letter would be permissible and an effective way for them to still communicate while imprisoned. Another example is if two operatives, Alice and Bob, wish to keep their identities secret but still need to communicate. If Alice were caught transmitting secret data to Bob then Bob would also be caught since he was the obvious recipient. Supposing that their relationship were kept ambiguous this avoids unnecessary risk for all parities involved. In theses situations steganography has many advantages over traditional cryptology.

1.2 Our Proposed System

The newest method we wish to offer to the world of steganography and cryptology makes use of permutations. The initial idea was developed independently by Peter Hendrickson (1). Here we extended on his idea by adding composition of permutations into the reply step. In some cases we use cycle notation which allows for a lot more freedom.

Since we are using permutation composition we are thereby obscuring the nature of the permutations being used between parties. For an eavesdropper to be successful at decoding the message she or he would need to have the entire conversation. This can be quite difficult to assertion in the situations we propose.

Another added feature is that we are using cycle notation instead of tabular. This makes it so fewer characters have to be embedded into the message in order to transmit a permutation. Saving space will be important when using twitter.

The math involved deals mostly with combinatorics, group and number theory. All of our computations are done in Sage; an open source mathematics software system (6).

1.2.1 twitter

The idea of using steganographic techniques in twitter should not come as a shock; many rebel groups have used twitter and Facebook as their method of communication and organization (3). Social networking offers an endless sea of information where one more drop couldn't possibly cause a stir. This idea of information that can exist in an inconspicuous way is the basis of steganography.

Depending on personal creativity there can be quite a few bits of hidden information encoded in the message body of a tweet. One system makes use of the white space between words and sentences. Explicitly a single space for a "0" and two for a "1". This can become a little cumbersome and not very efficient. With a 140 character per post limit, very few bits can be encoded this way.

Another method makes use of special alphabets. Suppose that synonyms for a commonly used word are in fact binary values. This scheme needs some amount of non-twitter communication and a shared codebook but there is a fair amount of data that could be encoded (4). It should be noted that the messages themselves may cause

suspicion depending on the choice of special alphabet.

Finally is the the method that this thesis proposes. Amateur traders talking about stocks, weather, politics and other newsworthy items. Little does the passing observer suspect that these posts are in fact encoding permutations and steganographic data.

1.2.2 Cards

A more involved application our permutation steganography scheme will be demonstrated with a deck of playing cards. There are 52! possible shuffles of a deck of 52 cards. This is a significantly sized number which can be used to encode a substantial amount of information.

What will be interesting to see here is the visible movement of cards. This is a physical way to compute compositions and inverses of permutations in S_{52}. This method could be used as a hands-on teaching tool for elementary group theory.

1.2.3 DNA

Inside of every living organism are billions of nearly identical copies of a genetic code. This information is known as deoxyribonucleic acid (DNA) and has been the subject of intensive scientific study for decades. There are approximately three billion pieces of data in the human genome. However, it has been discovered that not all of this data is significant.

It has been postulated that large parts of any genetic code serve no use. These large pieces of noncoding DNA, also called "junk DNA", compose about 50% of the the human genome (5, 11). Again this is where steganography can be applied to a system which inherently has a lot of wasted information.

Our other aim in this section is to see if, in fact, anything useful can be found in this junk DNA. We are not altering the DNA in a way that would disrupt the organism. As the reader will see in this section, there are many passive ways in which to too encode information into genetics.

Not only can the basic parts of DNA hold extra data but so can the reordering of large pieces. The genome is best thought of as a long piece of code where sections of this strand can be partitioned, duplicated and rearranged. All of these movements could be hiding information.

2

Background Mathematics

2.1 Shannon's Confusion and Diffusion

Claude Shannon is one of the founding fathers of information theory. In his paper *Communication Theory of Secrecy Systems* (1949) he defined confusion and diffusion. These two terms are basic ideas to measure the level of security assured by any given cryptographic system.

In the context of his paper confusion is the relationship between a cipher key and its cipher text. Ideally, this correlation should be difficult to ascertain. For example, if someone were given a large number of ciphertext and plaintext pairs it should still be difficult to figure what the key was that encoded them.

Diffusion, as outlined by Shannon, is the idea that the input of a system should have unpredictable (chaotic) changes to the output. This is also referred to as the avalanche effect. Hash functions and checksums are built around this core idea.

A hash function is commonly used to store and check user passwords. Slightly different passwords will have unpredictably different hashes. A checksum is used to check the integrity of data. If the data was altered in transit, either by error or sabotage, then its associated checksum will return a fail signal. Both of these application rely on what Shannon called diffusion.

For an analyst the initial step in decoding a steganographic message is finding it. The message could be embedded in many different kinds of media. Another problem in detection is a false positive. A analyst is trying to answer the question if there is an obvious message embedded or not. To make matters worse the embedded message

may also be encrypted.

Our system employs both of these layers; hiding the message as well as encryption to some degree. Hence taking some aspects form cryptology and steganography. Our implementation of permutation steganography uses an even blend of confusion, diffusion and intractable mathematics.

2.2 Combinatorial Compositions

A combinatorial composition of a natural number n is a way which n can be represented as a sum of positive integers. Here order does matter. Hence the following are all unique compositions of 3:

$$[3], [2+1], [1+2], [1+1+1]$$

Given any number n the first cut can be in $n-1$ places. This can continue until the interval is exhausted. Therefore, the total number of compositions of a positive integer n is the power set of $(n-1)$ which is 2^{n-1}.

2.3 Symmetric Group

In the group sense of the word a permutation is the set of all bijections of a set onto itself. In the combinatorics sense of the word it is all the possible ways to arrange n items without replacement. There are $n!$ permutations of a set of n elements. Note that n factorial is defined as $n! := (n)(n-1)(n-2)\dots(3)(2)(1)$ and $0! := 1$.

The set of all permutations of $n \in \mathbb{N}$ is called the symmetric group S_n. Sometimes it is also called the permutation group. As mentioned above, for finite n there are $n!$ elements in S_n.

2.3.1 Notation

There are a couple ways to represent a permutation $p \in S_n$. One being cycle notation and the other popular way is tabular notation. Cycle notation gives the cycles of what each element in the permutation maps to. Note that each cycle is disjoint. Tabular lists all of the elements in their new order. Unless otherwise specified this paper will be using tabular notation.

2.3.2 Composition

The group operation in S_n is composition. That is for any $a \in S_n$ and $b \in S_n$ then $fg \in S_n$ and $gf \in S_n$. Note that S_n is non-abelian, meaning that in general $fg \neq gf$. Typically context will define if this operation is meant to be done left-to-right or vice versa. In this paper it will always behave left to right. For example;

$$(ab)(x) = b(a(x))$$

2.3.3 Inverses

Every element $a \in S_n$ has an inverse $a^{-1} \in S_n$ such that $aa^{-1} = e \in S_n$ where e is the trivial permutation. Finding the inverse of a permutation is fairly sight forward. For a single cycle it is simply the cycle written down in reverse order: i.e. $(123450)^{-1} = (054321)$. For a product of cycles, reverse the order of the cycles and invert each cycle: i.e. $((013)(45)(579))^{-1} = (975)(54)(310)$ For products of disjoint cycles, we only need to invert each cycle: i.e. $((013)(45))^{-1} = (310)(45)$. Note that for each cycle of length two there is no reason to switch the order since they are equivalent.

2.4 Factorial Number System

The factorial number system, also called factoradic, is a number system adapted to numbering permutations of any finite size. Factoradic is a mixed radix numeral system. In this setting mixed radix means that the jth digit from the right has base $j!$. Therefore the digit in that place must be strictly less than j. It can be proven that no decimal number can be represented in more than one way. This is because the sum of consecutive factorials, $\{j!+(j-1)!+\ldots 2!+1!+0!\}$, multiplied by their index, $\{j,(j-1),\ldots,2,1,0\}$, is always the next factorial minus one. In symbols:

$$\sum_{j=0}^{n} j \cdot j! = (n+1)! - 1$$

Proof by induction:
Let $n = 0$ then $\sum_{j=0}^{0} j \cdot j! = (0+1)! - 1 = 0$
Let $n = 1$ then $\sum_{j=0}^{1} j \cdot j! = (1+1)! - 1 = 1$

Assume it is true for $k \leq n$. Let $n = k + 1$. Then we want to show that

$$\sum_{j=0}^{k+1} j \cdot j! = (k+2)! - 1$$

It can be seen that

$$\sum_{j=0}^{k+1} j \cdot j! = (k+1)(k+1)! + \sum_{j=0}^{k} j \cdot j! = (k+1)(k+1)! + (k+1)! - 1$$

By properties of factorials,

$$= (k+1)!((k+1)+1) - 1 = (k+1)!(k+2) - 1 = (k+2)! - 1$$

Which is exactly what we wanted since this implies that

$$\sum_{j=0}^{k+1} j \cdot j! = (k+2)! - 1$$

□

Here is an example of a factoradic number being converted to decimal:

$$430110_! = 4 \times 5! + 3 \times 4! + 0 \times 3! + 1 \times 2! + 1 \times 1! + 0 \times 0! = 555$$

Next, we will check this number by converting it back to factoradic:

$555 = 277 \times 2 + 1$
$277 = 92 \times 3 + 1$
$92 = 23 \times 4 + 0$
$23 = 4 \times 5 + 3$
$4 = 0 \times 6 + 4$ stop
Therefore,
$555 = (5 \times 4 + 3)4 + 0)3 + 1)2 + 1 + 0 = 430110_!$

When dealing with bases $n!$ for $n \geq 10$ other symbols or notation must be used so to avoid confusion. The factorial number system will later be used to calculate the rank of a permutation.

2.5 Rank of a Permutation

Consider the list of all permutations in S_n arranged in a lexicographical manner. Naturally the identity permutation $e = \{1, 2, 3, ..., n-1, n\}$ has rank 0. The next permutation $p = \{1, 2, 3, ..., n, n-1\}$ has rank 1 and so on. Up until we get to $\{n, n-1, ..., 3, 2, 1\}$, which would have rank $n! - 1$.

The rank of any permutation can be found by converting to factoradic and then to decimal. The opposite also works; meaning that given any number, $0 \leq r < n!$ a permutation of rank r of n elements can be found. The following is an example of this process.

2.5.1 Example: Rank 4 Permutation of S_3

Consider S_3, we can list all $3! = 6$ permutations quite easily:

$$S_3 = \{(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1)\}$$

Suppose that we want to find the 5th permutation (rank 4) of S_3. For such a small group we could list the elements and start counting from zero. This method soon becomes daunting for larger n. Instead, one could employ the factorial number system to help find the permutation with rank 4. The first step is the convert 4 into factoradic base:

$$4 = (2 \times 2 + 0)1 + 0 = 2 \times 2! + 0 \times 1! + 0 \times 0! = 200_!$$

Next we convert $200_!$ into a permutation in the following manner:

Factoradic:	2	0	0
	|	|	|
	(1, 2, 3)	→ (1, 2)	→ (2)
	|	|	|
Permutation:	(3,	1,	2)

Thusly we found that the 5th permutation of S_3 is $(3, 1, 2)$. Similarly, if given a permutation, p, we can find its rank buy doing the above operation in reverse. Either direction becomes unwieldily as the size of n increases, this is where computers can help.

2.6 Packing Number

Within the scope of this paper the packing number is amount of data that can be encoded into the rank of a permutation $p \in S_n$. Suppose there exists a bijective mapping from the interval $[0, \ldots, k-1]$ to an alphabet of length k. Finding the number of characters that can be encoded in the rank of p is equivalent to finding its maximum base k representation.

Solving the following equation for r helps approximate this number:

$$n! - 1 = (k-1)k^0 + (k-1)k^1 + (k-1)k^2 + \ldots + (k-1)k^{r-2} + (k-1)k^{r-1}$$

$$n! - 1 = k - 1 + k^2 - k + k^3 - k^2 + \ldots + k^{r-1} - k^{r-2} + k^r - k^{r-1}$$

$$n! = k^r$$

Which motivates the following definition:

> Let $n \in \mathbb{N}$ be the number of elements in the permutation p such that $p \in S_n$. Also let $k \in \mathbb{N}$ be the length of the character set of our language (alphabet).
> If $n! \geq k$ then the packing number, r, for any stenographic permutation system is defined as
>
> $$\left\lfloor \frac{\ln(n!)}{\ln(k)} \right\rfloor = \lfloor r \rfloor$$
>
> Where $\lfloor r \rfloor$ is the floor function of r.

This will tell us how many base k digits it will take to encode the rank of any permutation of n elements.

2.6.1 Example: the Packing Number of S_{52}

For example we can find the packing number of a deck of playing cards. Lets use the English alphabet plus all ten digits. Then our constants are $k = 36$ and $n = 52$ then:

$$\left\lfloor \frac{\ln(52!)}{\ln(36)} \right\rfloor = \lfloor 43.633 \rfloor = 43$$

We found $r = 43$, this means that a deck of cards can hold 43 characters. It should be noted that if the alphabet were increased to 37 characters we would get the same r:

$$\left\lfloor \frac{\ln(52!)}{\ln(37)} \right\rfloor = \lfloor 43.3022406174218 \rfloor = 43$$

This 37th character could be white space or something else determined by the user.

2.7 The 36 Character Alphabet

Here is the alphanumeric character set (alphabet) used in all of the steganographic systems proposed in this paper. Essentially it lists all 36 characters being mapped to the integers in the interval $[0, 35]$. This could easily be adapted for a different language with an arbitrary number of letters.

character	**number**	**character**	**number**
a	0	s	18
b	1	t	19
c	2	u	20
d	3	v	21
e	4	w	22
f	5	x	23
g	6	y	24
h	7	z	25
i	8	1	26
j	9	2	27
k	10	3	28
l	11	4	29
m	12	5	30
n	13	6	31
o	14	7	32
p	15	8	33
q	16	9	34
r	17	0	35

Table 2.1: 36 Character Alphabet - correspondence between alphanumeric and $[0, 35]$

3

twitter

3.1 Idea

The internet is a cesspool of largely worthless information and more is being added at an exponential rate. With communication moving at such a rate there are undoubtedly clandestine operations that even the NSA (12) would fail to catch. Hence, this why there is plenty of room for even more.

In addition to plaintext announcements there is the rampant use of poor punctuation, misspelling, abbreviations, internet slang, and URLs; all of which could be used to hide steganographic data.

In this section we are considering twitter. This is due to its popularity, level of activity, minimalistic interface and 140 character limit. As it will be later shown, permutation steganography can flourish in this situation.

3.2 Preliminary Discussion

The weakest point in this system is knowing what twitter account to look for. This is analogous to the problem of key exchange and identification. No matter how the system works, there exists an initial step of determining whom is trustworthy. In the case of steganography there is no cryptographic key exchange. Instead both parties have to be looking in specific places to be able to communicate with each other. Cryptology aims to be impervious to eavesdroppers; whereas steganography aims to look innocuous.

In certain situations it may not be necessary for Alice to know Bob or vice versa. Perhaps the message is simply being broadcasted by Alice. Whomever knows to look

(Bob or Bobs) can find and decode it. These are just a few situations that could possibly arise in modern communication environments.

3.3 Set up

The permutations which we will be studying in this section are from the symmetric group S_{14} where $e = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 0, @, \&, \#, \%\}$. Notice that we are using the symbols @, &, # and % for the 11th, 12th, 13th and 14th characters. These were chosen so that the packing number was sufficiently large.

Embedded inside of a post (or tweet) will be at least one permutation written in cycle notation. By convention the cycles will be separated by periods. Note that the maximum length of a cycle in our system will be 13. The double backslash "\\" denotes a new permutation. Therefore we can define two permutations using at most $13 \times 2 + 2 = 28$ characters of the 140 allowed in a tweet. If done correctly none of these embedded tweets will arouse suspicion within the twitter community.

There are many advantages to using cycle notation. One being that takes up less space than tabular notation. The second is that the encoder is allowed to cycle the elements of the cycles. Finally since the cycles are disjoint Alice is also allowed to move entire cycles as she sees fit.

For example the cycles $(1, 2, \&, @, 3)$, $(\&, @, 3, 1, 2)$ and $(3, 1, 2, \&, @)$ are all equivalent. If reordering gives a less awkward message than this is recommended. Cycling and moving cycles allows more freedom and creative license when embedding a permutation.

Next we compute the packing number, r, for this permutation stenographic system:

$$\left\lfloor \frac{\ln(n!)}{\ln(k)} \right\rfloor = \lfloor r \rfloor$$

In our case, $k = 36$ and $n = 14$. Plugging these numbers in we get:

$$\left\lfloor \frac{\ln(14!)}{\ln(36)} \right\rfloor = \lfloor 7.029744118944 67 \rfloor = 7$$

This is the maximum number of characters from our alphabet that we can fit into the rank of a permutation. By extension we can encode a maximum of seven plaintext characters into one permutation $p \in S_{14}$.

3.4 Encoding

Alice has a message which she wishes to send to Bob. No matter its entire length it will be split up into blocks of seven characters each. Let us consider just one of these blocks. We will name this block of alphanumeric characters m. She now has the task of converting m into its numerical representation. The result will be a list of seven integers such that each is strictly between 0 and 35. In symbols; let i and x_i be integers, then $m = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7 : 0 \leq x_i \leq 35\}$. Therefore m effectively becomes a base 36 number.

This string of base 36 integers is now treated as one one large number: $m = x_1x_2x_3x_4x_5x_6x_7$. After converting m to decimal she is now required to find a permutation of 14 elements that has the mth rank.

After finding the associated permutation $p \in S_{14}$ Alice can shorten its tabular representation into cycle notation. In most cases this will require her to transmit fewer total characters in her tweet while still conveying the same amount of hidden information. Therefore saving more space for words in which to embed the permutation.

Finally, it is up to Alice to mix (embed) the cycle notation of p into a plaintext message and tweet it. There are a few conventions that should be mentioned:

- Any duplicate elements of the permuted set, $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 0, @, \&, \#, \%\}$ are ignored and only the first occurrence is recognized.
- Cycles of the permutation are separated by periods.
- Permutations are separated by either double backslash "\\" or a new post.

Provided Alice had room in her tweet and were able to creatively disguise more cycles, she could embed additional permutations. To do so she would again pick at most seen characters and rerun the encode process as defined above.

3.5 Decoding

Assuming that Bob knows what twitter account to look for, he will find Alice's tweet and tease out the permutation hidden inside. To do this correctly he will use the agreed upon conventions of the system.

After Bob finds the first permutation, p, in cycles he then translates it into the full

14 character tabular notation. Next, Bob finds the rank of that permutation. This an either be carefully done by hand or with the help of a computer. Let r be the rank of the permutation p. Currently, this number is in decimal and we wish to translate it to base 36. After this final conversion Bob simply needs to associate the integers with the 36 member plaintext alphabet.

3.6 Reply

If the situation requires, Bob can reply to the message that Alice sent. This is done with the composition of permutations. Suppose that Alice encoded her message into permutation $a \in S_{14}$ and Bob received it intact. Bob now wishes to respond with his own plain text message m. The following is the process he must execute:

- Find the associated permutation for his plaintext message m by using the encode procedure as outlined above. Lets call Bob's permutation b.
- Compute the inverse of Alice's permutation, a^{-1}.
- Compose $ba^{-1} = x$. This permutation, $x \in S_{14}$, is what is going to be sent via twitter.
- Rewrite x in cycle notion.
- Bob embeds x into his next tweet.

By predefined conventions, Alice knows to compute $xa = ba^{-1}a = b$. Thus revealing the plaintext message m that Bob had intended for her to receive.

If there are multiple permutations present in Alice's transmissions the convention is that Bob will reply to the most recent one.

3.6.1 Round Two Communication

If Alice replies to Bob she will encode her message in the same way Bob did for her. Meaning she will find the associated permutation of her next message, call it a_2, and then compose it with x^{-1}. This will produce $y = a_2x^{-1}$. Bob knows to compute $yx = a_2x^{-1}x = a_2$. This process of replying with compositions can continue indefinitely.

3.6.2 Multiple Permutations Present

There is one extra step if the most prior tweet by Alice had multiple permutations embedded in it. The convention here is that Bob will respond to the most chronologically recent permutation sent by Alice. For example, suppose p_1 and p_2 were both mixed into Alice's most previous tweet. Bob would respond by treating p_2 in the same fashion as a in the first round communication.

3.7 Implementation

Suppose we have two parties Alice and Bob who wish to communicate with each other. Alice makes the first statement. Perhaps communicating something very specific or idiosyncratic. For example the possibly embarrassing phrase "hiboby" could be the initial encoded message from Alice to Bob. What happens next depends what the situation requires or what might have been prearranged. We will discuss a couple possible scenarios.

3.7.1 Case 1: Number Station

In this case there is no dialogue between parties. Therefore Alice is simply a broadcaster of encoded communications. Traditionally speaking, this set up is called a number station. A number station is periodically transmitting code. It is not waiting for inputs or prompts; it simply outputs encrypted information. Historically these stations are shortwave bands that transmit spoken numbers. These messages are only intelligible to operatives who know how to decode it.

For these reasons it has an advantage over traditional two party systems; obscurity. This arrangement keeps the relationship between Bob and Alice secret (or at least ambiguous). Frequent and short updates could be transmitted without arousing any suspicion in the twitter community.

3.7.2 Case 2: Dialogue Between Parties

Alternatively is the case where Bob and Alice need to communicate with each other. In this situation both parties involved make use of the predefined reply operation. Neither party necessarily needs to say "@Bob" or "@Alice" in the usual context of twitter. This

is because they will know where to look beforehand. Furthermore the "@" symbol is an element of our permuted set.

3.8 Dead Drop Example

In this example the cover story (if needed) is that Alice and Bob are teachers sharing (or announcing) administrative information. Again it should be stressed that their relationship is not technically public on twitter.

Numbers and symbols are very common in the language of stock trading, sports, events, and newsworthy statistics. Therefore it is recommended to use some topic along these lines in which to hide a permutation.

3.8.1 Geocache Preliminary

By definition, a dead drop is a specific location where physical items are placed for an operative to pick up at a later time. Suppose that Alice has made a dead drop for Bob. She used the site `geocahce.com` to be specific about the location. Geocache is a website that allows users to physically hide things for other people to find at a later time. This is purely for fun and adventure, except maybe in our case. Every location has a specific webpage generated for it. This address contains details including exact GPS coordinates. This information, the URL specifically, Alice hopes to transmit using the twitter permutation steganographic system outlined above.

Please note that while geocache is useful for this situation, it is also a public website. In the worst case scenario anyone with exact timing could get to the dread drop before the intended recipient (in this case Bob). In the unfortunate event of someone else retrieving the item before Bob, he will communicate that it was lost or stolen. In light of these events, both Alice and Bob would be forced to assume that they are under surveillance. Hence they should regroup at a later time.

3.8.2 Geocache Encode

Suppose that the unique geocode address that Alice generated by her dead drop submission is `http://coord.info/GCQN2M`. It should be noted that every geocahce info URL has the same prefix, specifically: `http://coord.info/`. For the sake of brevity

this leading part is truncated. Leaving the string *GCQN2M* to be encoded and transmitted. Bob will know to latter reattach the leading information in order to find the geocache URL.

The numerical representation of the character sting *GCQN2M* is $[6, 2, 16, 13, 27, 12]$. When this is treated as a base 36 number:

$$6 \times 36^0 + 2 \times 36^1 + 16 \times 36^2 + 13 \times 36^3 + 27 \times 36^4 + 12 \times 36^5 = 771571086$$

Alice is now tasked with finding the 771571086th permutation of 14 elements. Let $p \in S_{14}$, using Sage we found the 771571086th permutation to be

$$p = [1, 3, 10, 6, 11, 5, 4, 9, 12, 2, 8, 7, 13, 14]$$

In cycle notion this permutation is:

$$p = (2, 3, 10), (4, 6, 5, 11, 8, 9, 12, 7)$$

Please note that numbers 10, 11, 12, 13, and 14 have special representations in our character set:

$$p = (2, 3, 0), (4, 6, 5, @, 8, 9, \&, 7)$$

The next step for Alice is to mix the string 230.465@89&7 into her tweet. While making special note to use the predefined conventions of our proposed twitter steganographic system. To avoid confusion of cycles Alice may not use $\{1, \#, \%\}$ anywhere before a "\\". She must also mind periods in her post since they denote cycles. With these things taken into account Alice's tweet might look something like the following:

> Ur Class has 230 cap. This must be 4th,5th or 6th time I tried
> email to jack@89&7com.

3.8.3 Geocache Decode

From Alice's tweet Bob finds the string of special permutation elements in the following arrangement: 230.465@89&7. This string of course represents the permutation $p \in S_{14}$ which Alice carefully encoded into her tweet. Going backwards he converts the string into the following:

$$p = (2, 3, 0), (4, 6, 5, @, 8, 9, \&, 7) = (2, 3, 10), (4, 6, 5, 11, 8, 9, 12, 7)$$

In non-cycle (tabular) notation this becomes:

$$p = [1, 3, 10, 6, 11, 5, 4, 9, 12, 2, 8, 7, 13, 14]$$

Using his preferred method, Bob calculates that $p \in S_{14}$ has rank 771571086. He now converts this decimal number to base 36 to get the ordered set

$$[6, 2, 16, 13, 27, 12]$$

Now he translates these digits into their plaintext alphanumeric representations. This yields:

$$[g, c, q, n, 2, m]$$

Finally, Bob goes to `http://coord.info/gcn2m`. Here he finds the exact GPS coordinates and a description of the dead drop Alice made.

3.9 Longer Example

Here Alice will be splitting up her message across multiple permutations and multiple tweets. This is because what she wants to encode stretches over a the characters per block limit. Her intended plaintext message "the owls are not what they seem" will be split up in the following manner:

$$a = [t, h, e, o, w, l, s],\ b = [a, r, e, n, o, t, w],$$
$$c = [h, a, t, t, h, e, y],\ d = [s, e, e, m]$$

These letters are mapped to their numerical representation as ordered sets:

$$a = [19, 7, 4, 14, 22, 11, 18],\ b = [0, 17, 4, 13, 14, 19, 22]$$
$$c = [7, 0, 19, 19, 7, 4, 24],\ d = [18, 4, 4, 12]$$

Converting these from base 36 to decimal yields the following:

$$a = 39884820175,\ b = 49062195684$$
$$c = 52497309175,\ d = 565218$$

The first tweet will contain the permutations of a and b. Alice's second tweet will not be a reply but a continuation of the same message. This second consecutive tweet will contain the permutations c and d.

She does the encoding process for a, b, c, d and gets the following four permutations:

the ath permutation is $[7, 6, 4, 3, 2, 13, 12, 11, 9, 8, 5, 1, 14, 10]$
the bth permutation is $[8, 13, 6, 2, 4, 5, 7, 11, 9, 12, 10, 1, 3, 14]$
the cth permutation is $[9, 6, 10, 2, 13, 7, 8, 12, 4, 5, 3, 1, 14, 11]$
the dth permutation is $[1, 2, 3, 4, 6, 11, 5, 8, 7, 9, 14, 10, 12, 13]$

In cycle notation:

$a = [(1, 7, 12), (2, 6, 13, 14, 10, 8, 11, 5), (3, 4)]$
$b = [(1, 8, 11, 10, 12), (2, 13, 3, 6, 5, 4)]$
$c = [(1, 9, 4, 2, 6, 7, 8, 12), (3, 10, 5, 13, 14, 11)]$
$d = [(5, 6, 11, 14, 13, 12, 10, 9, 7)]$

This converted to our twitter character set reveals:

$a = [(1, 7, \&), (2, 6, \#, \%, 0, 8, @, 5), (3, 4)]$
$b = [(1, 8, @, 0, 12), (2, \#, 3, 6, 5, 4)]$
$c = [(1, 9, 4, 2, 6, 7, 8, \&), (3, 0, 5, \#, \%, @)]$
$d = [(5, 6, @, \%, \#, \&, 0, 9, 7)]$

Keeping in mind that periods separate cycles and \\ will separate permutations, Alice's first tweet could look something like:

> on feb 17 &. 26 #OWS will grow 11% -got bad 08 trends: DJIA @+5.43 \\ 18 more years? @bob012.2FM #3 in rock i remember when you were 6,5and4!

Her second tweet might look something like:

> 1942=best year ever; 678mm&.305cm #war -2% @obama \\ i think most vets are over 56 @mccain -3% #ows & don't forget 097

In the decoding stage Bob would see these tweets from Alice's account and know how to tease out the plaintext information. To do this he simply does the encoding process in reverse.

3.10 Analysis

3.10.1 Authentication

Guaranteed authentication in a cryptographic (or steganographic) system is an open question. In counterintelligence a common technique to gain access to a system is by impersonation. For the following example Eve will be the enemy. By impersonation Eve acts as a relay of information between Alice and Bob. Alice thinks she is talking to Bob and vice versa. When in fact both are talking to Eve. This allows Eve to see and manipulate the entire conversation as she wishes. This is called a man-in-the-middle (MITM) attack. To be able to carry out a MITM attack in our proposed twitter scheme is a little difficult. Again we make the appeal to obscurity. It boils down to a simple matter of the enemy knowing where to look. If Eve knew this fact then she could become a man in the middle by hacking to Alice's or Bob's twitter accounts.

Another possible impersonation attack would be if someone were to reply to Alice who was not Bob. In this case Alice would compose her message with fake Bob and continue communication in the usual fashion. But real Bob would attempt to decode the communications from Alice and get gibberish. Hence real Bob would know that

the system was compromised at some point and tell Alice.

Ideally the initiation of a dialogue would happen only if Alice trusted Bob and vice versa. Sadly, proof of identity is very difficult to guarantee in a setting as large and anonymous as the internet. One solution to this problem is that there is some small amount of pre-twitter communication between parties. It could be something as simple as a text message or an email containing a twitter account name.

3.10.2 Extra Characters

A larger permutation group would naturally increase r. Hence the number of plaintext characters that could fit into a single permutation would also increase. This step of engineering is left to anyone who wishes to employ this system. One possible idea is to use the more of the alphanumeric set as permutation elements. For example the set $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 0, @, \&, \#, \%, x, q\}$. The possible danger with this idea is that the tweet may begin to look cryptic and nonsensical.

I argue that our plaintext alphabet does not need a special termination character. This would waste encryption time and space. The terminating code could easily be put into the wording of the final tweet. For example the phrases "I'm headed to bed", "bye for now" or any other number of common sign-offs could easily get the point across.

3.10.3 Limitations of twitter

The maximum length of a tweet (including spaces) is 140 characters. This limitation should be kept in mind at all times when using this system. Let us suppose that each tweet is self-contained. Meaning that the cycle notation for a single permutation does not run over into the next tweet. Since the maximum number of characters used to represent a single permutation is $(13 + 2) = 15$, then the maximum number of permutations per post can found by the following equation:

$$\left\lfloor \frac{140}{15} \right\rfloor = \lfloor 9.333 \rfloor = 9$$

It should be noted that fitting $9 \times 15 = 135$ numbers and symbols into a single tweet will not allow much room for a cohesive looking sentence. This will probably look suspicious and very much like code to an observer, even a passive one. Which is why in our examples we use at most two permutations per tweet. Theoretically

it is possible for a single permutation to extend over to tweets if both parties knew beforehand.

4

Cards

4.1 Idea

A deck of playing cards is an everyday item that will likely not cause suspicion. Cards and their particular order will be the mode of communication in this proposed steganographic system. Physical delivery of the encoded deck is left to the imagination of the operatives.

It should be noted that the computations inherent to this system require a computer. The number of all possible shuffles of a deck of cards is mind-boggling

$$52! 80658175170943878571660636856403766975289505440883277824000000000000$$

The largeness of this number plays a big part in our encoding scheme. Much like the twitter example, here we make use of the lexicographical rank of a permutation.

4.2 Preliminary Discussion

Due to the difficulty of sending a physical deck of cards this message will have to be somewhat self-contained. An example situation could be Alice packing a deck of cards into luggage to be later picked up by Bob. The message could be instructions on where to go and what to do next.

A computer is recommended for the encoding and decoding step of this system. This is due to the largeness of the numbers involved and the frequency of human error.

4.3 Set Up

The packing number r in this proposed system is

$$\left\lfloor \frac{\ln(52!)}{\ln(36)} \right\rfloor = \lfloor 43.6333221585942 \rfloor = 43$$

Therefore we can fit 43 characters from our alphabet into any permutation $p \in S_{52}$. This permutation should be thought of as a full deck of cards arranged in a particular way.

Please note that if the number of characters in our alphabet were increased to $k = 37$ then r would still be 43. This 37th letter could be a space character if needed.

Since we are using 52 unique cards we need to define a convention of numbering each card. For face cards let Ace= 1, Jack= 11, Queen= 12 and King= 13. Next is the issue of suits. Let

$$\heartsuit < \diamondsuit < \clubsuit < \spadesuit$$

To achieve a complete list of 52 cards simply use the following function.

Let C be a card. Also let n be the face value as mentioned in the prior convention. Then C is of the form "n suit" (for example King of Hearts abbreviated $K\heartsuit$) and

$$f(C) = \begin{cases} n + 13 \times 0 & : \heartsuit \text{ or H} \\ n + 13 \times 1 & : \diamondsuit \text{ or D} \\ n + 13 \times 2 & : \clubsuit \text{ or C} \\ n + 13 \times 3 & : \spadesuit \text{ or S} \end{cases}$$

4.4 Encoding

Suppose that Alice has a message written in a 36 character alphabet. For her to transmit her message using this system she must do the following procedure:

- Check to see that her message is less than 43 characters in total length.
- Convert her message into its equivalent numerical representation.
- Think of this string as a single base 36 number.
- Change this number into decimal notation. Call this number m.
- Find the mth permutation $p \in S_{52}$.

- Whatever permutation she finds will tell her how to exactly order her deck of cards.
- She stacks her cards in the top-to-bottom order that Bob will read them.
- Finally the arranged deck p is given to Bob in some innocent manner.

Bob then carefully decodes the beck $p \in S_{52}$ in the following way.

4.5 Decoding

The deck of arranged cards is given to Bob. Here it should be noted that this method is not in any way robust. Meaning that moving a single card could ruin the entire message (see: 4.8 Analysis).

Now Bob converts the cards as they appear into their $[1, 52]$ numerical representations. By doing this he has found the unique permutation $p \in S_{52}$. This is permeation that Alice intended for him to find. The next step converts p into a plaintext message.

Bob must now find the rank of $p \in S_{52}$. This can be done carefully by hand, using the factorial number system or by computer. Let m be the rank of p.

After converting m from decimal to base 36 Bob can map the numbers to their alphanumeric characters. This final step will reveal the plaintext message that Alice encoded for Bob.

4.6 Replying

Much like in the twitter example, replies in this system are treated as compositions of permutations. Suppose that Alice sent the message a to Bob in the form of a deck of cards. If Bob wishes to reply to Alice's message with his own message b he will need to compute the permutation $x \in S_{52}$ such that

$$xa = ba^{-1}a = b$$

The permutation $x \in S_{52}$ is the one that gets sent to Alice in the form of an ordered deck of cards.

4.7 Card Example

Suppose that Alice wishes to send the message "To whom it may concern, this deck is stacked" to Bob.

4.7.1 Card Encode

In preparation this message is first rewritten without spaces or punctuation: *towhomitmayconcernthisdeckisstacked*. Then the numerical representation of each charter is then put into a list. Lets call this list m.

m = {19, 14, 22, 7, 14, 12, 8, 19, 12, 0, 24, 2, 14, 13, 2, 4, 17, 13,
19, 7, 8, 18, 3, 4, 2, 10, 8, 18, 18, 19, 0, 2, 10, 4, 3}

This number Alice will treat as a one large number in base 36.

$$m = 19 \times 36^0 + 14 \times 36^1 + 22 \times 36^2 + \ldots + 3 \times 36^{35}$$

Next Alice converts this number into decimal:

$$m = 256024984088891338576199593937852980805296081484065579$$

Then she finds the mth permutation in S_{52}.

[1, 2, 3, 4, 5, 6, 7, 8, 13, 20, 19, 27, 40,
29, 15, 39, 31, 51, 47, 21, 49, 42, 12, 25, 46, 10,
16, 30, 35, 33, 45, 37, 22, 11, 32, 50, 52, 26, 24,
18, 17, 14, 48, 44, 36, 34, 38, 9, 43, 23, 41, 28]

As far as actual cards this is represented as:

a=[1H, 2H, 3H, 4H, 5H, 6H, 7H, 8H, 13H, 7D, 6D, 1C, 1S,
3C, 2D, 13C, 5C, 12S, 8S, 8D, 10S, 3S, 12H, 12D, 7S, 10H,
3D, 4C, 9C, 7C, 6S, 11C, 9D, 11H, 6C, 11S, 13S, 13D, 11D,
5D, 4D, 1D, 9S, 5S, 10C, 8C, 12C, 9H, 4S, 10D, 2S, 2C]

Finally, the encode process is complete. All Alice needs to do now is pack the cards up and give them to Bob.

4.7.2 Card Decode

Bob makes note of the order of the cards as they appear in the deck. Much like the twitter system, he first writes down the permutation of 52 elements. Second, he computes its rank. Third, he converts that rank into base 36. Finally, he converts that string of numbers into its alphanumeric plantext representation. Therefore finding the encoded message:

[t, o, w, h, o, m, i, t, m, a, y, c, o, n, c, e, r, n, t, h, i, s, d, e, c, k,
i, s, s, t, a, c, k, e, d]

4.7.3 Card Reply

Bob wishes to respond with the message "Alice, tell me something I don't know. sincerely Bob".

He will be using the permutation composition method as defined in the twitter example. The first set is to remove all punctuation from his message; *alicetellmesomethingIdontknowsincerelyxxbob*. Here Bob added "xx" before his signature simply to fill out the 43 characters. Doing this is not required but a good idea. The numerical representation of this message is:

[0, 11, 8, 2, 4, 19, 4, 11, 11, 12, 4, 18, 14, 12, 4, 19, 7, 8, 13, 6, 8,
3, 14, 13, 19, 10, 13, 14, 22, 18, 8, 13, 2, 4, 17, 4, 11, 24, 23, 23, 1,
14, 1]

When the message is thought of as a single base 36 number this converts to the following in decimal:

321937955647848906548367828456239941179335683051742444102898566796

The permutation with this rank is:

[1, 12, 32, 15, 35, 26, 28, 38, 13, 14, 25, 30, 36, 2, 24, 5, 33, 19,
41, 46, 29, 20, 27, 7, 9, 48, 34, 39, 23, 47, 18, 50, 51, 16, 10, 49, 3,
21, 37, 8, 22, 17, 4, 6, 11, 31, 44, 42, 45, 40, 43, 52]

In actual cards this translates to,

b=[1H, 12H, 6C, 2D, 9C, 13D, 2C, 12C, 13H, 1D, 12D, 4C, 10C,
2H, 11D, 5H, 7C, 6D, 2S, 7S, 3C, 7D, 1C, 7H, 9H, 9S, 8C, 13C,
10D, 8S, 5D, 11S, 12S, 3D, 10H, 10S, 3H, 8D, 11C, 8H, 9D, 4D,
4H, 6H, 11H, 5C, 5S, 3S, 6S, 1S, 4S, 13S]

Bob now has the task of finding a permutation $x \in S_{52}$ such that $xa = (ba^{-1})a = b$. If what Alice sent was a and what Bob wants to send is b, all that is left to compute is a^{-1} followed by ba^{-1}. The inverse of Alice's message, a^{-1}, can be found mechanically using a couple creative methods. The first method involves making a spread sheet. Column A will have 1 though 52 and Column B will have Alice's permutation a. Then Bob simply sorts by column B and as a result gets the inverse of permutation a in column A.

The other method is similar but has the advantage of involving the cards directly. For Bob the first step to keep a just as Alice sent it. The second step is to find another deck and put it into lexicographical order. Then to lay out the two decks as two large rows:

e=[1H, 2H, 3H, ..., 5S, 6S, 7S, 8S, 9S, 10S, 12S, 13S]
a=[1H, 2H, 3H, ..., 10C, 8C, 12C, 9H, 4S, 10D, 2S, 2C]

The next step is very intuitive; Bob simply has to look at the the bottom row to find $1, 2, 3, \ldots, 50, 51, 52$ and write down whats above each of them. Using this mapping

Bob has found the inverse of permutation a.

$$a^{-1}=[1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 48,\ 26,\ 34,\ 23,\ 9,\ 42,\ 15,\ 27,\ 41,\ 40,\ 11,\ 10,\ 20,\ 33,\ 50,\ 39,\ 24,\ 38,\ 12,\ 52,\ 14,\ 28,\ 17,\ 35,\ 30,\ 46,\ 29,\ 45,\ 32,\ 47,\ 16,\ 13,\ 51,\ 22,\ 49,\ 44,\ 31,\ 25,\ 19,\ 43,\ 21,\ 36,\ 18,\ 37]$$

Remember that Bob is trying to compute x such that $xa = ba^{-1}a = b$. So the next permutation to find is ba^{-1}. This can be done with a computer or with the cards in the following manner:

$$a^{-1}=[1\text{H},\ 2\text{H},\ 3\text{H},\ \ldots,\ 5\text{C},\ 12\text{D},\ 6\text{D},\ 4\text{S},\ 8\text{D},\ 10\text{C},\ 5\text{D},\ 11\text{C}]$$
$$b=[1\text{H},\ 12\text{H},\ 6\text{C},\ \ldots,\ 11\text{H},\ 5\text{C},\ 5\text{S},\ 3\text{S},\ 6\text{S},\ 1\text{S},\ 4\text{S},\ 13\text{S}]$$

This tells Bob that 5C mapped to 11H, 12D to 5C and so on. This is a very good visual aid when computing $x = ba^{-1}$. Now all Bob has to do is pack up the permutation x and send it to Alice. She knows to compose x with a to get Bob's intended message b.

4.8 Analysis

Here we will be discussing what happens to the integrity of the message if the cards were shuffled in some small way. For the following analysis we are using the original message that Alice sent to Bob in the above example, "To whom it may concern, this deck is stacked".

4.8.1 Shuffling

The question naturally arises; what happens to an accidental shuffle? As a start lets see what happens when the last two cards are switched.

> [1, 2, 3, 4, 5, 6, 7, 8, 13, 20, 19, 27, 40, 29, 15, 39, 31, 51, 47, 21,
> 49, 42, 12, 25, 46, 10, 16, 30, 35, 33, 45, 37, 22, 11, 32, 50, 52, 26,
> 24, 18, 17, 14, 48, 44, 36, 34, 38, 9, 43, 23, 28, 41]

This permutation has rank:

$$256024984088891338576199593937852980805296081484065578$$

In base 36 this number becomes the following:

> [18, 14, 22, 7, 14, 12, 8, 19, 12, 0, 24, 2, 14, 13, 2, 4, 17, 13, 19, 7,
> 8, 18, 3, 4, 2, 10, 8, 18, 18, 19, 0, 2, 10, 4, 3]

When translated to out alphabet becomes:

> [s, o, w, h, o, m, i, t, m, a, y, c, o, n, c, e, r, n, t, h, i, s, d, e, c, k,
> i, s, s, t, a, c, k, e, d]

Therefore only one letter changed. The leading t changed to an s. This makes sense because changing the last two digits of a permutation increases the rank by one.

Next lets see what happens when we exchange the first two cards.

[2, 1, 3, 4, 5, 6, 7, 8, 13, 20, 19, 27, 40, 29, 15, 39, 31, 51, 47, 21,
49, 42, 12, 25, 46, 10, 16, 30, 35, 33, 45, 37, 22, 11, 32, 50, 52, 26,
24, 18, 17, 14, 48, 44, 36, 34, 38, 9, 43, 23, 28, 41]

This permutation has rank:

$$155111875328763830520833190780787941065719757299779140808148406557 9$$

Lets name the the difference between the original rank and this new rank:

$$\Delta = 155111875328738228022424301646930321106325972001698611200000000000 0$$

Here is what the new rank when converted to base 36:

[19, 14, 22, 7, 14, 12, 8, 19, 12, 0, 24, 26, 32, 15, 19, 21, 11, 13,
35, 11, 18, 27, 35, 18, 20, 10, 32, 26, 21, 18, 5, 14, 27, 0, 15, 5, 29,
5, 7, 18, 4, 25, 6]

And in our alpha be this becomes:

[t, o, w, h, o, m, i, t, m, a, y, 1, 7, p, t, v, l, h, 0, 1, s, 2, 0, s, u,
k, 7, 1, v, s, f, o, 2, a, p, f, 4, f, h, s, e, z, g]

As the reader can see after the 12th character the message is gibberish. This is due to the fact that changing the leftmost elements of a permutation vastly effects the rank. It should be noted that there are 52 choices for the first slot. Following this reasoning each of those choices should take up $52!/52 = 51!$ of the lexicographical list. This is why the difference Δ between permutation ranks is exactly equal to 51!.

It stands to reason that Bob can visibly tell where a stacked deck got out of its encoded order. This can be done by seeing where the decoded message went from something intelligible into gibberish. Potentially Bob can try some systematic adjustments to try to reproduce the indented message form Alice.

4.8.2 Message Length

A message of one character will only change the position of two cards. Hence, the shortness of the original messages determine how much the cards will move. This may or may not be a problem given the situation at hand. Hence it is recommended to fill up the 43 characters with either relevant data or a long signature.

4.8.3 False Positive

The idea of decoding a message that was not actually put there is a possible problem when dealing with everyday items such as cards. Ideally a random shuffle of cards will not encode something intelligible. A possible solution is to always end the message with a signature of a certain length.

The possibility of a readable message and a unique signature is quite small. In fact the probability of getting Bob's signature *xxBob* in any part of a 43 character sequence is:

$$\frac{1}{36^5} = 1.65381716879202 \times 10^{-8}$$

and simply *Bob* is:

$$\frac{1}{36^3} = 0.0000214334705075446$$

Ergo there is little reason to worry about either party decoding possible messages hidden in random decks from the casino gift shop.

4.8.4 Multiple Decks

If one could identify between different decks even more information could be encoded. For example, given two unique decks this would make 104 cards. Hence, the new packing number would be:

$$\left\lfloor \frac{\ln(104!)}{\ln(36)} \right\rfloor = \lfloor 106.671290242364 \rfloor = 106$$

Note that the packing number for two decks of cards is not comply the same as 52 times two.

And for 3 decks of cards we get:

$$\left\lfloor \frac{\ln((3 \cdot 52)!)}{\ln(36)} \right\rfloor = \lfloor 177.262056867692 \rfloor = 177$$

This can continue indefinitely so long as there is a convention associated with the different decks to make them unique.

5

Genetics

5.1 Motivation

Deoxyribonucleic acid (abbreviated DNA) is the fundamental molecular structure that defines all life on Earth. DNA differs slightly or drastically depending on the organism. What is interesting is that in the genetic code for humans, "only about 1.5% of the human genome consists of protein-coding exons, with over 50% of human DNA consisting of non-coding repetitive sequences."(11).

We will be looking into what permutations and combinatorial composition can offer as far as amounts of hidden data. There are many properties of DNA that make it attractive for steganography. One being that there is a lot of information in the static DNA. Another will be the information associated with DNA's replication process (called transcription).

5.2 Preliminary Genetic Information

The DNA code is make up of our nucleotides: adenine (abbreviated A), cytosine (C), guanine (G) and thymine (T). The chemical structure works so that G only bonds with C and T only with A. Through this chemical paring of four nucleotides the entire structure of DNA is made.

The bonds CG and AT are called base pairs. The human genome has approximately three billion base pairs arranged into 46 chromosomes. The set of chromosomes in a cell makes up its genome. Groups of three base pairs, called codons, are what make up an amino acid.

There are 20 unique amino acids that occur in nature. The number of ways to group three from four is $4^3 = 64$ combinations. This implies that there are in fact many codons that technically represent the same amino acid. This can be seen in the following table:

Codon	**A.A.**	**Codon**	**A.A.**	**Codon**	**A.A.**	**Codon**	**A.A.**
TTT	Pheny	TCT	Ser	TAT	Tyro	TGT	Cys
TTC	Pheny	TCC	Ser	TAC	Tyro	TGC	Cys
TTA	Leu	TCA	Ser	TAA	Stop	TGA	Stop
TTG	Leu	TCG	Ser	TAG	Stop	TGG	Try
CTT	Leu	CCT	Pro	CAT	His	CGT	Arg
CTC	Leu	CCC	Pro	CAC	His	CGC	Arg
CTA	Leu	CCA	Pro	CAA	Glu	CGA	Arg
CTG	Leu	CCG	Pro	CAG	Glu	CGG	Arg
ATT	Iso	ACT	Thr	AAT	Asparagine	AGT	Serine
ATC	Iso	ACC	Thr	AAC	Asparagine	AGC	Serine
ATA	Iso	ACA	Thr	AAA	Lysine	AGA	Arginine
ATG	Met	ACG	Thr	AAG	Lysine	AGG	Arginine
GTT	Valine	GCT	Alanine	GAT	Asp	GGT	Glycine
GTC	Valine	GCC	Alanine	GAC	Asp	GGC	Glycine
GTA	Valine	GCA	Alanine	GAA	Glu.A	GGA	Glycine
GTG	Valine	GCG	Alanine	GAG	Glu.A	GGG	Glycine

Table 5.1: Codon Mapping - Each of the 64 Codons Named

Note that: Pheny=Phenylalanine, Leu=Leucine, Asp=Aspartic acid, Glu.A=Glutamic acid, Met=Methionine, Cys=Cysteine, Ser=Serine, Tyro=Tyrosine, Try=Tryptohan, Pro=Proline, His=Histidine, Arg=Arginine, Glu=Glutamine, Iso=Isoleucine, Thr=Threonine.

5.3 Combinatorial Compositions of DNA

Here we are not trying to encode data but rather decode what might already haven been put there. This idea was birthed from studying the process of genetic transcription. One of the possible explanations as to why there is so much junk DNA is that large portions are repeated in a seemingly unintelligible fashion.

During transcription large portions of the DNA are naturally cut out and not used in the functions of the cell. It is in the location of these cuts that information could reside. We are not saying that information would be encoded in the way we are suggesting but rather we are using our scheme to give a measure of the size of this information set.

The key issue is that as far as our current understanding of biology goes, any message with the prescribed parameters could be encoded and this would have no effect on the underlying organism. This is the very nature of steganography in hiding information in plain sight.

In every system thus far the permuted set was well-defined. For example in the card system the permuted elements were the 52 cards. In twitter it was 14 elements, namely: $\{1, ..., 9, 0, @, \&, \#, \%\}$. Now suppose that we don't explicitly know the number of elements of our permuted set. Meaning that the DNA strand could potentially be cut up in many unique ways.

Here is where combinatorial composition comes in. Suppose that we define our string of DNA to have length m. The first cut of the string can be in any of $m-1$ different spots. The total number of ways that m can be composed as a sum of integers is equal to 2^{m-1}. In our case the length of m can be quite large since we are dealing with at most three billion base pairs.

We can represent the way in which strand m is cut by a binary number. Call this binary number c and let c_i be the digit in the ith place holder. Therefore $i = 1$ for the first place holder, $i = k$ for the kth place and so on. Thusly we can make the following convention:

$$c_i = \begin{cases} 1 & \text{if there is a cut in position } i \\ 0 & \text{if there is no cut in position } i \end{cases}$$

The index i will range from 1 to $m-1$. This will make the binary number c at most $m-1$ digits long. In decimal notation the size of this number will be between 1 and 2^{m-1}.

It should be noted that orientation must be accounted for. That is if starting from one end of the strand and counting the cuts is unique from starting at the opposite end. If orientation cannot be determined, that is if both ends are the same, then there are half as many possible combinatorial compositions:

$$\frac{2^{m-1}}{2} = \frac{2^m}{4}$$

Suppose that we converted c into decimal. Using this information we could associate it with a rank of a permutation $p \in S_n$. The next problem is finding the number of elements which we are permuting. In other words, finding n which defines S_n. Keep in mind that for c to be the rank of a permutation then S_n must have at least 2^{m-1} elements in it.

After the permutation group is agreed upon we can then find the the exact permutation that has rank c. For the sake of argument replying within this system will be assumed to be similar to the card and twitter examples. Suppose that Alice sent Bob message a. Bob wishes to communicate message b to Alice. Therefore he composes $ba^{-1} = x$ and sends x to Alice. Here we are not assuming Alice and Bob are real people but rather unknown agents at work within the organism.

5.4 Possible Process

What could be happening to DNA from a purely steganographic sense could be something along the following lines:

- A permutation $p \in S_n$ needs to be communicated.
- The rank of p is converted from decimal to binary.
- This binary number is represented a combinatorial composition of a string of DNA.

In this system we can still compose permutations. This will make the communication intractable. The composition step will be done outside of the DNA but could involve the nucleotides; any arrangement of $\{G, T, C, A\}$. String orientation is also important as it will determine how many unique combinatorial compositions exist.

The integer length of DNA used to encode this binary number (call it m) must be

at least as big as $n!$. The length of m will always be the sum of the cuts. Finding the smallest $n \in \mathbb{N}$ that satisfies $2^{m-1} \leq n!$ will give the symmetric group S_n. Hence the permutable elements, $\{1, 2, \ldots, n-1, n\}$, will also be found.

5.5 Example

Suppose that Alice and Bob wish to communicate in this manner of combinatorial compositions and permutations. It would look something like the following.

5.5.1 Encode

Suppose that Alice wants to send Bob the permutation $p = [2, 3, 4, 6, 1, 5]$. This permutation happens to be in S_6. Therefore the largest possible rank of any permutation in S_6 will be $6! - 1 = 719$.

When converted to binary 719 is 1011001111_2. Therefore the length of DNA which is needed to encode all permutations in S_6 is $m = 10 + 1 = 11$ since 1011001111 is 10 digits long. We are also assuming that we can distinguish the head from the tail of the DNA.

The rank of p is $154 = 10011010_2$. Therefore the combinatorial composition of $m = 11$ is:

$$0010011010 = 3 + 3 + 1 + 2 + 2 = 11$$

Therefore we have the cut up pieces of DNA that encode permutation $p \in S_6$. The set of parts $[3, 3, 1, 2, 2]$ is what gets sent to Bob.

5.5.2 Decode

Bob knows that he is looking for a combinatorial composition of 11. He finds the parts $[3, 3, 1, 2, 2]$ and knows to work backwards:

$$11 = 3 + 3 + 1 + 2 + 2 = 0010011010$$

The number $0010011010_2 = 154$ he knows is the rank of a permutation $p \in S_6$. Finally, her finds $p = [2, 3, 4, 6, 1, 5]$, which is what Alice has encoded for Bob.

5.5.3 Reply

Bob wants to communicate the permutation $q = [4, 2, 6, 1, 5, 3]$ to Alice. Next Bob must find $r \in S_6$ such that $r = qp^{-1}$. This permutation is what Bob wants to communicate to Alice.

$$r = qp^{-1} = [4, 2, 6, 1, 5, 3][5, 1, 2, 3, 6, 4] = [3, 1, 4, 5, 6, 2]$$

Bob finds the rank of $r \in S_6$ which is 249. In binary this number is $249 = 11111001_2$. In terms of cuts this binary number represents the following:

$$0011111001 = 3 + 1 + 1 + 1 + 1 + 1 + 3 + 1 = 11$$

Finally, this $[3, 1, 1, 1, 1, 1, 3, 1]$ is what is going to be transmitted to Alice.

5.6 Analysis

5.6.1 Ideal Length

Ideally we are finding integer solutions to the following equation:

$$2^{m-1} \leq n!$$

The closer these two values are the fewer unused permutations there are. For example consider the permutations of all 64 codons. If $p \in S_{64}$ then the order of p could be as high as $64! - 1$.

$$64! - 2^{295} \approx 6.32295579254315 \times 10^{88}$$

And one more power of 2 makes this number negative:

$$64! - 2^{296} \approx -4.27816335021216 \times 10^{86}$$

This means that our orientated piece of DNA string should have length 296.

5.6.2 Maximum Length

The largest piece of DNA that we are potentially allowed to manipulate is the junk half of the human genome. If this were the case then we would have the possibility

of making about 500 million different cuts. To encode this number of cuts the binary number would need to be about 500 digits long. In decimal, this number is:

$$2^{500} \approx 3.27339060789614 \times 10^{150}$$

Now it would perhaps be a good idea to find $n \in \mathbb{N}$ such that:

$$3.27339060789614 \times 10^{150} \leq n!$$

It was found exhaustively that:

$$3.27339060789614 \times 10^{150} \leq 96!$$

Suppose that we could pick out a binary number that was 500 digits long. In decimal, this number would then represent a permutation $p \in S_{96}$.

If we were then to compute the packing number of this system using our usual alphanumeric character set it would be:

$$\left\lfloor \frac{\ln(96!)}{\ln(36)} \right\rfloor = \lfloor 96.3799586367100 \rfloor = 96$$

Therefore using this system 96 alphanumeric characters can be encoded.

6

Future Frontiers

The field of steganography is growing in step with technology. There is no need to wait around for a theorem to be put to good use. All that is required is a small amount of creativity to make a steganographic system. There are just as many legal uses for steganography as there are illegal. For these reasons it should come as no surprise that steganography can be found anywhere. Thats the strength and weakness of these systems.

6.1 twitter

A long term goal for the twitter system would be to write some code for people to use in the real world. Something user friendly, taking in arguments of the system and the message they wish to encode. Outputting the permutation plus an idea of how to mix it into a tweet. This would ideally be how it works for the enduser.

The twitter system as it was stated in this paper has a lot of room for personal interpretation. For example, the number of characters in the permuted set could be 10 to 14 depending on user choice. Also the number of permutations per post are left to the reader to decide. These initial parameters need to be communicated between operatives for this system to work.

The twitter program could have a list of phrases to choose from. The user could see a few ideas of how to mix in the permutations that they have generated. Ultimately it would be up to the user to pick through and edit the final post. Of course the things here to avoid are patterns and tweets that look like code.

6.2 Cards

The card permutation system is what I hope to someday use as a teaching aid in the classroom to teach permutations. I believe there is no need to go deep into group theory to introduce an interesting exercise. I would probably do it with fewer cards, maybe only four or six. These cards would come in a couple flavors; numbers, colors and symbols. The idea of *red* being greater than, say, *blue* might come as a shock but thats the point. A natural ordering of the number line is something most students take for granted. Hence this fact in and of itself should be a lecture in K-12 mathematics.

In my lessons the identity would be introduced, then composition and finally inverse. Each of these would come with a group project. This type of math is what is needed in K-12, not route memorization. I digress since the situation of math education is a different discussion entirely.

6.3 Genetics

For my own benefit would I like to learn more about genetic algorithms. I want to know how much data they are dealing with. Whether or not it is even close to the numbers that I came up with. The movement of DNA pieces is probably most of what they are trying to keep track of.

Bioinformatics is something else that has interested me for a long time. The idea that math and computer science can be applied to biological systems. These models are what will undoubtedly be the future of the medical field.

6.4 New Systems

- One could order their purchases at the grocery store as they appear on the receipt in a prescribed order to contain certain information.
- One could assign homework problems in such a way that the numbers of the problems give the order of digits for a permutation of $\{0, 1, 2, \ldots, 9\}$. The numbers chosen in the test questions for a pre calculus final could convey information - each problem would contain a single disjoint cycle.

- One could alter a random number generator to make it less random and available to output digits in a prescribed order.
- One could specify a book in the library of congress whose ISBN number specifies a permutation that contains certain information.
- The ordering of books or magazines in a stack could secretly specify a permutation by comparing the given ordering with the one that is alphabetical by author or title.

I want this information to be free to the public. I want to see what the Internet citizens can do with it. All that is required is a tiny bit of creativity and a work. Permutation steganography can exist anywhere there is a measurable frequency of categories and patterns. These are constructs of human thinking and therefore limitless.

Bibliography

[1] Peter Hendrickson. *Steganography using permutations* retrieved from `http://www.wiredyne.com/software/pstego-detail.html`

[2] Gregory Kipper (2005) *Investigator's guide to steganography* page 34 ISBN 0899324335

[3] John Ribeiro (26 January 2011) *Egypt blocks Twitter during anti-government riots* Protesters demand removal of President Mubarak retrieved from `http://www.computerworlduk.com/news/security/3258124/egypt-blocks-twitter-during-anti-government-riots/`

[4] Luke Hebbes (16 June 2010) *Twitter Steganography* retrieved from `http://blog.rlr-uk.com/2010/06/twitter-steganography.html`

[5] Elizabeth Pennisi (15 June 2007) *DNA Study Forces Rethink of What It Means to Be a Gene* retrieved from `http://www.sciencemag.org/content/316/5831/1556` Science 15 June 2007: Vol. 316 no. 5831 pp. 1556-1557 DOI: 10.1126/science.316.5831.1556

[6] (2011) *sage Permutation module* retrieved from `http://www.sagemath.org/doc/reference/sage/combinat/permutation.html`

[7] Viviana Risca (2000) *Hiding in DNA* retrieved from `http://www.sciencenews.org/view/generic/id/346/title/Hiding_in_DNA`.

[8] Jane Smiley (2005) *Thirteen Ways of Looking at the Novel* NY: Alfred A. Knoph, p. 15

[9] Ono, Ken and Ahlgren, Scott (2011) *Fractal Structure to Partition Function* retrieved from `http://www.aimath.org/news/partition/`

[10] Kristy Westphal (2010) *Steganography Revealed* retrieved from `http://www.symantec.com/connect/articles/steganography-revealed`

[11] Wolfsberg T, McEntyre J, Schuler G *"Guide to the draft human genome"* Nature 409 (6822): 8246 doi:10.1038/35057000 PMID 11236998.]

[12] Jane Maer (May 23, 2011) *The Secret Sharer* The New Yorker retrieved from `http://www.newyorker.com/reporting/2011/05/23/110523fa_fact_mayer` i,
3
2
3
3
4
3

1
4, 37
13