A. Carlson, S. R. Mikkilineni, M. W. Totaro, R. E. Hiromoto and R. B. Wells. (Jul. 19-22, 2022). An Introduction to Local Entropy and Local Unicity, *2022 International Symposium on Networks, Computers and Communications (ISNCC)*, Shenzhen, China, 2022, pp. 1-7, doi: https://doi.org/10.1109/ISNCC55209.2022.9851755]

An Introduction to Local Entropy and Local Unicity

Albert Carlson[*], **Sai Ranganath Mikkilineni**[†], **Michael W. Totaro**[‡], and **Robert E. Hiromoto**[§].
[*] Computer Science Department, Austin Community College, Austin, TX, USA.
[†] [‡]Intelligent Systems, Modeling, and Simulation (ISMS) Research Lab, Center for Advanced Computer Studies (CACS), University of Louisiana at Lafayette, Lafayette, LA, USA.
[§] Department of Computer Science, University of Idaho, Moscow, ID, USA.
Email: [*]albert.carlson@austincc.edu, [†]sai-ranganath.mikkilineni1@louisiana.edu, [‡]michael.totaro@louisiana.edu, [§]hiromoto@uidaho.edu.

***Abstract*—As introduced by Shannon in "Communication Theory of Secrecy Systems", entropy and unicity distance are defined at a global level, under the assumption that the properties of symbols resemble that of independent random variables. However, while applying entropy and unicity to language(s), e.g., encryption and decryption, the symbols (letters) of a language are not independent. Thus, we introduce a new measure, $H_L(s)$, called the "local entropy" of a string s. $H_L(s)$ includes *a priori* information about the language and text at the time of application. Since the unicity distance is dependent on the entropy (because entropy is the basis of calculations for the unicity distance), local entropy leads to a local unicity distance for a string. Our local entropy measure explains why some texts are susceptible to decryption using fewer symbols than predicted by Shannon's unicity while other texts require more. We demonstrate local entropy using a substitution cipher along with the results for an algorithm based on the principle, and show that Shannon's unicity is an average measure rather than a lower bound; this motivates us to present a discussion on the implications of local entropy and unicity distance.**

***Keywords*—** *entropy, redundancy, unicity distance, shannon theory, hartley's function, set theoretic estimation*

I. Background

A. Shannon Theory

1) Entropy: Claude Shannon studied communication and information theory during the early-to-mid 20^{th} century; in his works, Shannon explored the elements of cryptography. "A Mathematical Theory of Communication" [1] and "Communication Theory of Secrecy" [2] are the most prominent amongst Shannon's publications in the field of communications; Shannon introduced the key concepts of entropy, language redundancy, and unicity distance in these publications, consequentially leading to his founding work in the field of information theory (IT) [3].

978-1-6654-8544-9/22/$31.00 ©2022 IEEE

Entropy (uncertainty) is considered as the gold standard measurement for the efficacy of a cryptographic system, while redundancy and unicity distance are defined in terms of entropy. Before Shannon, Hartley introduced an uncertainty measure [4] in 1928 that Shannon adapted as his basis for information entropy. Hartley addressed the concept by measuring the amount of information obtained about the set of symbols in a message, after randomly selecting a member of the set [4] (This is referred to as Hartley's function, or entropy.) The Hartley function is defined as:

$$H(x) = log_b(|A|) \tag{1}$$

where x is a particular member of the considered set, $|A|$ is the size of a message's alphabet set, and b is an arbitrary base of the logarithm; if $b = 2$ (i.e., binary) then the measure is in the unit of bits.

In this paper, we focus on the base-2 implementation of Hartley's function. Since Hartley's function assumes that the elements of the set are uniformly distributed, the probability of the distribution function selecting a given member in the set is defined as:

$$\frac{1}{|A|} \tag{2}$$

If the member selected from a set is known *a priori*, then the number of items in the corresponding set, A is said to be singular and Hartley's function is applied to a set with the property $|A| = 1$. Therefore, knowing the identity of an element from the set reduces the uncertainty about the set by

$$H(x) = log_b(1) = 0 \tag{3}$$

Therefore, $H(x) = 0$ for all known data, contributing no uncertainty to the information about the set.

Shannon extended Hartley's function for defining the

Letter	Frequency	Letter	Frequency	Letter	Frequency	Letter	Frequency
e	0.1247	t	0.09692	a	0.082	i	0.07681
n	0.07641	o	0.071409	s	0.070677	r	0.06681
l	0.044831	d	0.036371	h	0.0350386	c	0.03439
u	0.028778	m	0.028178	f	0.023515	p	0.020317
y	0.018918	g	0.019119	w	0.01352	v	0.0104567
b	0.010658	k	0.00393	x	0.002198	j	0.001998
q	0.000935	z	0.000599				

TABLE I: English Letter Frequency

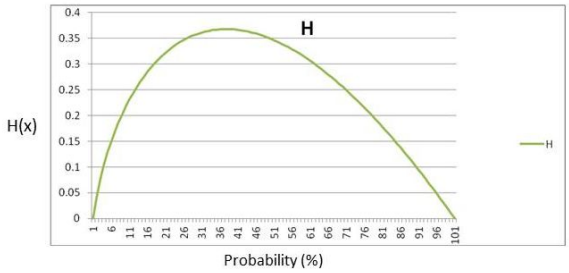

Fig. 1: Entropy Values

entropy for a string, s, drawn from a language as:

$$H(s) \leq nlog_2|A| \tag{4}$$

where, $n = |s|$, and $|A|$ is the size of the alphabet of the language (i.e., the size of a set in Hartley's function). For a string composed of randomly selected independent characters (here, a and b), the entropy of the string has the property of:

$$H(a,b) \leq H(a) + H(b) \tag{5}$$

This inequality holds only for a string composed of statistically independent symbols; however, as shown by the frequency of letters in Table I, a language and its constituent strings are neither uniformly distributed nor statistically independent. For instance, in English, the letter 'e' typically occurs between 12% and 14% of the time, instead of approximately 3.84% if the set of letters were uniformly distributed. This led Shannon to restate Hartley's function in a more general form [2]:

$$H(s) = \sum_{i=1}^{n} P_{x_i} \; log_2 \; \frac{1}{P_{x_i}} \tag{6}$$

where P_{x_i} is the probability of symbol x_i appearing at position i in a string s. Maximum entropy is then achieved if all the symbols appearing in the string are equally likely, as defined for a string purely composed of uniformly distributed letters that are statistically independent of each other [5]. Figure 1 shows a plot of Shannon's entropy for a single character, depicting how the "uncertainty" from a letter contributes to the entropy of the string in which it is found. The contribution of a single letter to the entropy of a string is highest when $P(x_i) \approx .37$ (refer to Figure 1). As a letter's probability varies about $P(x_i) \approx .37$, its entropy is reduced, and the symbol contributes less towards the uncertainty of the string. Similarly, the contribution to the entropy of a string by the constituent letter is reduced when the letter is either eliminated from consideration ($P(x_i) = 0$), or the range of possible substitutions for the letter's probability approaches one.

2) Unicity Distance: The unicity distance [2] is the average number of symbols needed from an encrypted message string to provide enough information to be correctly decrypted. The unicity distance is inversely proportional to entropy, and is defined as:

$$n = \frac{log_2|k|}{R_L log_2|A|} \tag{7}$$

where $|k|$ represents the number of unique keys possible for a specific cipher. R_L is the redundancy of the encrypted text. For a successful decryption of a ciphertext, on average n symbols are required.

3) Redundancy: Redundancy is the tendency of a language to repeat symbols. Mathematically, redundancy [4] is defined as:

$$R = 1 - \frac{H(x)}{H_{max}} \tag{8}$$

Here H_{max} is the maximum entropy of a string s, and is the sum of the maximum entropy of each letter in the string.

$$H(s) \leq n\sum_{i=0}^{n-1} log_2|A| = |s|\sum_{i=0}^{n-1} log_2|A| \tag{9}$$

Since maximum entropy is achieved when each member of the set is uniformly possible and statistically independent. If we consider English, with an alphabet of 26 letters, the H_{max} should be ≈ 4.7 bits per symbol; however, considering the letter probabilities presented in Table-I and treating each symbol to be independent, the entropy for English is $H(x) \approx 4.136$ bits per symbol.

In the redundancy equation, $H(x)$ represents the Shannon entropy for an unknown letter or a combination of letters. Minimizing this $H(x)$ will maximize the redundancy of the considered language. For any language, R varies between two limits,

$$0 \leq R \leq 1. \tag{10}$$

When all of the data is known about a string, then for each letter x_i in the string, $P_{x_i} = 1$, and $H(x_i) = 0$. Thus, for the string, s:

$$H(s) \;=\; \sum_{i=0}^{n-1} H(x_i) \;=\; \sum_{i=0}^{n-1} 0 \;=\; 0 \qquad (11)$$

Then the redundancy for string s is:

$$R \;=\; 1 \;-\; \frac{0}{H_{max}} \;=\; 1 \qquad (12)$$

As shown above, When all the data about string s is known, R is equal to one (the maximum value). Conversely, R is at its minimum value zero, when $H(s)$ is composed of completely random and uniformly distributed letters, i.e., $H(s) \;=\; H_{max}$. For this case, the redundancy is given by:

$$R \;=\; 1 \;-\; \frac{H_{max}}{H_{max}} \;=\; 1 \;-\; 1 \;=\; 0 \qquad (13)$$

In a natural language [6], however, R is unlikely to be 0 because all languages have rules [7] that disallow a random and uniform distribution of letters or words.

For instance, considering an unstructured language, i.e., $R_L \;=\; 0$ (i.e., no symbol would be repeated due to lack of binding rules/grammar), the denominator of the unicity distance equation would become 0. In this case, the unicity distance becomes $n \;=\; \infty$ (which is not viable). Thus, languages with a structure (rules/grammar) must have a non-zero entropy and a non-zero redundancy.

4) M-grams and Style: Variability in expression using a language (natural or formal) is a universal characteristic [6], [8]. An interesting pattern of writing (semiotics) is shown in [9], there are different ways to express an idea [6], [10], all of which are recognizable/comprehensible by a speaker of the language. The different ways of expressing an idea are often informally referred to as "style(s)", and are unique to an individual [8]. The ability to handle the variability in a language is crucial to a system that relies heavily on the properties of the data it receives.

M-grams [2], [11] are employed to capture this variability in the usage of a language: M-grams are a string of continuous letters taken from a text. Each M-gram has a length of $|m|$ letters and is composed solely of alphabetic symbols. Any non-alphabetic character(s) in the considered string are ignored. M-grams may contain a word or span over multiple words (i.e., being composed of two or more words). Several decryption methods have used M-grams [2], [11], [12] as a means to exploit the statistical features of a language (like, redundancy).

In order to capture the variability in the usage of a language, M-grams must be drawn from a variety of distinct texts representing various genre, subject matter, and authors. The collection of M-grams encountered in the collection of texts, or corpora, identifies a set of M-grams allowable in a language. The M-grams that are not possible in a language (for example, 'qwz') are categorized as "forbidden." Forbidden M-grams represent a majority of letter combinations for $|m| \;\geq\; 3$, and they provide an enormous amount of information about a language. M-grams can be assembled from a dictionary by combining two, or more, words. Randomly assembling word combinations will produce all of the possible combinations in the language, but does not necessarily reflect the actual use of the language. Prior statistical studies of language(s) that used M-grams, relied on spaces embedded in written language as a source of *a priori* information. In practice, however, spaces are removed from the plain text before encryption. In the latter case, the M-gram analysis applied to a text string may not detect boundaries between words. The training of allowable M-grams based on a language dictionary is therefore of limited value. We addressed this limitation by applying M-gram analysis to an extensive corpora of text where the usage incorporates a representative cross section of language use, instead of using a dictionary of a language for the analysis. In this work, 21 authors were surveyed, (see Table-III), whose work(s) span from the 16^{th} to the 20^{th} century. At least two works of each author are included in the corpora. Genres of the works are chosen to include various combinations of fiction, non-fiction, poetry and novels. No foreign works or translations are included.

$\|m\|$	Forbidden	Allowed	Total	% Forbidden
1	0	26	26	0.00%
2	15	661	676	2.2189%
3	6261	11315	17576	35.6224%
4	347292	109684	456976	75.9979%
5	11251945	629431	11881376	94.7024%
6	306789115	2126661	308915776	99.3116%

TABLE II: M-gram Summary

Separating M-grams into the forbidden and "allowed" sets [11], [13] (see Table II) is the first step for our set based analysis using set theoretic estimation (STE). STE [14] is a set based technique that requires unambiguous set membership for proper operation. Property sets are sets that embody information about the problem and solution which are composed of sets, like M-grams that are applied to inputs, and are used to focus on possible solutions. Forbidden M-grams constitute a large amount of information and are an efficient way to determine the set membership.

In the following section, we define and state the local entropy and the local unicity distance measures, followed by the experimental conditions and results in the succeeding section, then we conclude the paper with our discussion about the implications and our conclusions.

II. Local Entropy and Local Unicity

In cryptography involving languages, *a priori* knowledge about the probability of letters appearing, and the frequency of symbols in a text, play a cardinal role. It is rare for a cryptographer to have an entire message available for analysis. Even if an entire message is available, it is divided and analyzed as blocks (where the size of each block varies on a case to case basis), with one block at a time. For the analysis of an encrypted string, the information contained in a block of that string, along with any available *a priori* information, is used. Given this nature of the cryptanalysis, our notion of "local entropy" and "local unicity" emerged.

The properties of local entropy are characterized by $H_L(s)$, defined in terms of the Shannon entropy for a block of message. For a message M of length $|message|$, and a string or block within the message that currently needs to be decrypted, s of length $|string|$, it is assumed that the entropy is measured for $|string| << |message|$. The value of this analytical approach occurs when information specific to the string is known. Information such as letter frequency, properties of prior decryption results, or other *a priori* information (like context) in general makes up this knowledge, denoted by S. The combination of $string$ and the size s defines the local entropy in terms of Shannon's entropy as

$$H_L(s) = H(s|S) \tag{14}$$

Since it can be shown that

$$H(s|S) \leq H(s) \tag{15}$$

then

$$H_L(s) \leq H(s) \tag{16}$$

For every block in a message, there is an associated H_L. Calculating H_L for two blocks of the same length provides a baseline for comparison of the uncertainty contained in each block based on the information and symbols present in the respective blocks.

Because redundancy is defined as a function of entropy, thus, a local definition for redundancy (R_L) based on equation (8):

$$R_L(H_L) = 1 - \frac{H_L(s)}{H_{max}} \tag{17}$$

Likewise, the corresponding notion of local unicity (n_L) can be defined as:

$$n_L = \frac{log_2|k|}{R_L(H_L)log_2|A|} \tag{18}$$

Taking the ratio between Shannon's unicity distance (Equation-(7)) and the minimum unicity distance (n_{min}, computed based on the cipher used for a given language) results in the relationship:

$$n = \frac{n_{min}}{R_L} \tag{19}$$

Synonymous to the global unicity distance, n_L is also minimum when the denominator of the unicity distance equation (Equation-(18)) is at its maximum. Since R_L varies between 0 and 1, the denominator of the unicity distance equation (Equation-(18)) is maximized when $R_L = 1$ (i.e., when there is a perfect redundancy that leads to null uncertainty about any character in the message). Therefore, The minimum unicity distance when $R_L = 1$ is:

$$n_{min} = \frac{log_2|k|}{log_2|A|} \tag{20}$$

Some key solution sets, such as the solution set for a substitution cipher, are also affected by a reduced number of symbols in a block. For a block with u unique symbols, the possible number of keys in a substitution cipher, for instance, is given by

$$|k_s| = \binom{|A|}{u}$$

The key solution set for the block are, $|k_s| \leq |k|$ and $n_L \leq n$.

The unicity distance is bounded for values of $0 \leq R_L \leq 1$. When none of the *a priori* information about the block is available, i.e., $R_L = 0$, then $n = \infty$, and when total information about the block is available, i.e., $R_L = 1$, then $n = n_{min}$. Therefore,

$$n_{min} \leq n \leq \infty \tag{21}$$

III. Experimental Conditions and Results

As an example for demonstrating H_L and n_L, we will show how the analysis of local information (i.e., M-grams of size much smaller than the message size) from a ciphertext in conjunctions with *a priori* information can be used and applied to the blocks in the ciphertext to solve a decryption problem. In this case, the language of the message is English ($|A| = 26$) and the cipher used is a substitution cipher (i.e., with $|k| = 26! \approx 4.2 \times 10^{26}$) [15]. Encrypted data for the messages comes from selected works of literature found at the Project Gutenburg's [16] website.

Since our demonstration involves analyzing *a priori* information consisting of M-grams of lengths 2 and 3, identical letters that occur in succession in a string (such

Author	Title	Author	Title
Asimov	Foundation	Dickens	A Christmas Carols
Asimov	Foundation II	Dickens	Great Expectations
Bacon	The Advancement of Learning	Fitzgerald	Anthony Patch
Bacon	The New Atlantis	Fitzgerald	Flappers and Philosophers
Boswell	The Journal of a Tour to the Hebrides	Grey	Riders of the Purple Sage
Boswell	Life of Johnson	Grey	The Plainsmen
Burroughs	Tarzan	Hume	Enq. Concerning the Principles of Morals
Burroughs	The Lost Continent	Hume	Dialogues Concerning Natural Religion
Bronte	The Professor	Milton	Areopagitica
Bronte	Jane Eyre	Milton	Paradise Lost
Bulfinch	Bulfinch's Mythology, the Age of Fable	O'Henry	Cabbages and Kings
Bulfinch	Legends of Charlemagne	O'Henry	Options
Bunyan	An Exhortation to Peace and Unity	Poe	Collected Works, Vol. 1
Bunyan	The Works of John Bunyan	Poe	Collected Works, Vol. 2
Carroll	Alice in Wonderland	Scott	Ivanhoe
Carroll	Through the Looking-Glass	Scott	The Kenilworth
Christie	The Mysterious Affair at Styles	Stevenson	Dr. Jekyll and Mr. Hyde
Christie	The Secret Adversary	Stevenson	Kidnapped
Dafoe	Jane Eyre	Swift	A Modest Proposal
Dafoe	Moll Flanders	Swift	Gulliver's Travels
		Shakespeare	Complete Works

TABLE III: Training Corpora

as 'ee', 'ss', etc.), and the maximum number of times a letter appears in a string of length n. Here, $n = 26$.

M-grams for the property sets are taken from training texts representative of the English langauge. Any M-gram encountered during training is marked as allowed while the remainder are forbidden. Each M-gram found in the ciphertext are analyzed and mapped, resulting in impossible M-grams removed from consideration for the key mapping. The ciphertext is mapped to possible plain text and the set of resulting mappings are compared to the list of forbidden M-grams. Forbidden M-grams are a powerful set because the number of forbidden M-grams increases rapidly as M increases (see Table II) [11], [17].

In a limited length strings, $|string| << |message|$, letter counts are indicative of the distribution of letters. Among the counted symbols, the high and low frequency letters are of high interest. For instance, In a string where $|string| = 26$, only the letters o, l, and e occur 12 or more times, while the letters j, v, and z occur less than 5 times. Therefore, any ciphertext letter that appears 6 or more times in a string of length 26 cannot be mapped to j, v, or z, and any cipher text letter occurring 12, or more, times can only map to o, l, or e.

Intersecting the results from each of the above sets until a single solution is left, narrows the possible mapping solutions. Obtaining that solution completes the process. During testing, the final mapping is compared against the known key to verify accuracy.

Training for the tests was conducted on a corpora of forty one texts representing multiple genres and time periods. Tests were performed on texts curated from Project Guttenburg library [16].

The size of the M-grams and the words are typically much smaller than the length of the entire message. Information about the message is derived from looking at only a portion of the available message. Global information is limited to the key known while processing the M-grams. As previously stated, the focus of our investigation is a small subset of letters from which we attempt to infer as much information as possible using the sets that represent observed language statistics. The *a priori* knowledge about the language, possible key mappings, and input ciphertext are applied to M characters from the message for analysis. Forbidden M-grams, for example, restrict possible key mappings. As the possible mappings are combined through set intersection local data is combined and then shared.

An example of this effect can demonstrate the use of local information. Assume that a part of a message given in the ciphertext as:

rqxxrcq

Further, assume that M-grams allowed are limited to $m = 2$ and $m = 3$ as shown in Table IV. Beginning at the first symbol in the message (i.e., the 'r'), the ciphertext letter 'r' can only map to plaintext letters

'a', 'e', 'g', 'i', 'l', 'm', 'n', 'o', 's', or 't'. All other mappings are not possible, given the restrictions on the allowed M-grams list for $m = 2$. The next ciphertext letter (i.e., 'q') also maps to the same possible letters. Taking the next symbol in the message (i.e., 'x'), it adds no new information, but the addition of the second occurrence of the letter 'x' from the message narrows the range of possible letter mappings for the 'x' to those 2-grams that appear on the list from Table IV as doubled letters. The ciphertext letter 'x' can only map to the plaintext letters 'e', 'l', or 's'. Intersecting the ciphertext's 3-gram 'qxx',and the possible mappings for 'x', the only possible 3-gram mapping is 'ess', which means,

$$q \mapsto e$$

and

$$x \mapsto s$$

Coming back to the 3-gram at the beginning of the message(i.e., 'rqx'), with the limitations on the allowed 3-grams leading to a single mapping of:

$$r \mapsto a$$

The *a priori* knowledge in the form of statistics of the language taken from a representative sample of the language, is employed using the input ciphertext message stream as an inference target to infer more information about the plaintext message. Taking few symbols at a time, the information is inferred locally and then shared as new *a priori* knowledge about the rest of the ciphertext. However, the global information about the full message was not required for decryption.

Previous and contemporary methods used to decrypt substitution cipher messages rely on global information. For instance, the letter frequency method [18] counts the number of occurrences of a symbol in a message, and a guess is made about the mapping based on the *a priori* knowledge about the letter frequency in the language based on global counts of letters encountered in the message. Although the letter frequency method is effective for about 80% of the time [19], the accuracy is highly dependent on the number of symbols in the message, generally requiring several thousand symbols for a successful decryption of the ciphertext. The relaxation method used by Peleg and Rosenfeld [12] requires solving a global constraint equation requiring more than a thousand characters on average, for a reliable decryption [18]. But, using our method results in a successful decryption with just several hundred symbols (i.e., our method shows an increased efficiency in decryption, approximately at an order of magnitude of 10, while simultaneously decreasing the amount of time required for the decryption.). A summary of the experimental results comparing our method to our predecessors' and contemporaries' is shown in Table V.

Allowed 2-grams	Allowed 3-grams
ag	age
am	ame
ee	ent
en	ess
ge	ion
io	sag
ll	ssa
me	
nt	
on	
sa	
ss	
ta	

TABLE IV: M-grams for example

Even though our demonstration example has limited language possibilities, but, since the principles guiding the usage of a language are universal, that makes our method and findings (i.e., the information about a part of the plaintext message can be inferred from a cipher text block that is adjacent to it in the ciphertext message) reported in this paper extrapolatable.

Type of Test	Authors	Average Symbols	Test Time
Relaxation	Peleg and Rosenfeld	$\approx$ 1000	18+ hrs
Letter Frequency	Experimental	2000 - 4000	$\approx$ 3.3 sec
M-grams	Experimental	200 - 400	$\approx$ .25 sec

TABLE V: Previous Experimental Results [12], [18]

IV. Discussion

In this paper, we presented a closer examination of entropy and its relationship to entropy in decryption. Local entropy (i.e., the entropy of a block of the ciphertext) can be exploited using the available *a priori* knowledge about the cipher and the language of the plaintext message. The *a priori* information in terms of language statistics, including the information given by forbidden M-grams, can focus and narrow down the decryption effort on a subset of all possible solutions, increasing the fidelity of decryption. Making use of the local entropy facilitated the reduction of the number of symbols required for decryption of a message using a substitution cipher by the factor of 10 over other commonly used decryption methods. The time required for decryption was also substantially reduced.

Making use of local entropy can be accomplished by implementing a system that calculates $H_L(s)$ and n_L for all blocks in the ciphertext. The purpose of the calculation is to find the local minimums for $H_L(s)$ for various lengths of a block, s, in the message. As evident from our results, some blocks of the encrypted message may have lower local entropy, and provide

opportunities to infer information about the cipher key leading to a faster and better decryption. Identifying the blocks of the message that are susceptible to local analysis allows focusing on that block of the message to capitalize on the available information, and curate the *a priori* information and use it for the decryption of other blocks of the ciphertext. Local entropy and local unicity distance provides an explanation for the success of algorithms that do not decrypt an entire ciphertext at once. Techniques involving both local entropy and local unicity, aid in understanding the message, the decryption process, and show promise in the implementation of efficient decryption algorithms.

Another use of local entropy is to calculate how much information has accumulated in a message fragment, or "shard." Breaking the message into shards and then applying distinct and different key/cipher combinations for each shard results in a much stronger cipher while treating each shard as an orthogonal decryption problem. If the shard is correctly sized in such a way that there is not enough redundancy present in the shard to meet the local unicity distance constraints will help in hindering the decryption of the shard by an adversary. The difficulty of decryption is inversely proportional to the size of the shard. This is the crux for the development of polymorphic ciphers [20], and spurred investigations into polymorphic RNGs to combat Venona style attacks [21].

REFERENCES

[1] Claude Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379 – 423, 623 – 656, 1948.

[2] Claude Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28:656 – 715, 1949.

[3] Thomas Cover and Joy Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc, New York, 2nd edition, 2005.

[4] Paul Garrett. *The Mathematics of Coding Theory*. Pearson/Prentice Hall, Upper Saddle River, 2004.

[5] Claude Shannon. Prediction and entropy of printed english. *Bell System Technical Journal*, 30:50 – 64, 1951.

[6] D. Terence Langendoen and Paul Postal. *The Vastness of Natural Languages*. The Camelot Press, Ltd., Southampton, 1984.

[7] Fernando C. N. Pereira and Stuart M. Shieber. *Prolog and Natural-Language Analysis*. Microtome, 1987.

[8] Andrew Morton. *Literary Detection*. Scribners, New York, 1978.

[9] Daniel Chandler. *Semiotics for beginners*. Routledge, Milton Park, UK., 1994.

[10] Noam Chomsky. *Syntactic Structures*. Mouton, The Hague, 1957.

[11] Albert Carlson. *Set Theoretic Estimation Applied to the Information Content of Ciphers and Decryption*. PhD thesis, University of Idaho, 2012.

[12] Shmuel Peleg and Azriel Rosenfeld. Breaking a substitution cipher using a relaxation algorithm. *Communications of the ACM*, 22:598 – 605, 1979.

[13] Robert E. Hiromoto, Albert H. Carlson, and Richard B. Wells. An information based approach to cryptography. In *The 6th Computer Information Systems and Industrial Management Applications*, 2007.

[14] Patrick Combettes. The foundations of set theoretic estimation. *Proceedings of the IEEE*, 81(2):182 – 208, 1993.

[15] Bruce Schneier. *Applied Cryptography: Protocols, Algorithms, and Source Code in C*. John Wiley and Sons Inc., New York, 2nd edition, 1996.

[16] The Gutenburg Project. Main page-http://www.gutenburg.net. Internet, 2008.

[17] Robert Lewand. *Cryptological Mathematics*. Mathematical Association of America, Washington D.C., 2000.

[18] George Hart. To decode short cryptograms. *Communications of the ACM*, 37(9):102–108, 1994.

[19] Albert H. Carlson, Robert E. Hiromoto, and Richard B. Wells. Breaking block and product ciphers applied across byte boundaries. In *The 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, pages 733–736, 2011.

[20] Albert Carlson and Robert Le Blanc. Polymorphic encryption engine, 2015.

[21] John Earl Haynes and Harvey Klehr. *Venona: Decoding Soviet Espionage in the United States (Yale Nota Bene)*. Yale University Press, 1999.