[Mandeep Singh, Albert Carlson. (Jan. 06, 2025). Using the y-intercept in Collision Attacks for Encryption. 2025 IEEE 15th Annual Computing and Communication Workshop and Conference (CCWC), IEEE Xplore. Source: https://ieeexplore.ieee.org/document/10903877]

Using the y-intercept in Collision Attacks for Encryption

Mandeep Singh[*], **Albert Carlson**[†], **Christopher Briscoe**[‡].
[*]Independent Researcher
[†]Chair for Entropy and Encryption, Quantum Security Alliance, Kyle, TX, USA
[‡]Department Chair of Mathematics and Physics, Emet Classical Academy, New York, NY, USA
Email: [*]msingh5@scu.edu, [†]Ltzap1@gmail.com. [‡]Cbriscoe@emetclassicalacademy.org

***Abstract*—Patterns are key to decrypting messages. Language is composed of repeat groupings of symbols encoded in words and the syntax of a language. Patterns bleed through encryption since all encryption algorithms can be replaced by a substitution (S) cipher. This makes it possible to analyze the patterns in cipher text using the collisions in the cipher text message. Each language has its own unique lexicon and rules, making it possible to identify both the original plain text language and the size of the block used for encryption. The ability to identify this data supports both Kerckhoffs' and Shannon's assumption that the attacker knows all of the important data about the encryption algorithm, minus the secret key. In practice, this means that an attacker can quickly focus on the data needed to prosecute an attack on an encrypted message just by analyzing collisions in the message. This paper demonstrates that such an approach is feasible and lays the foundation to explore how to derive the information needed to attack encrypted messages using data derived from the collisions in the cipher text.**

Keywords: Collision attack, cryptography, information theory, heuristic methods, language statistics, birthday paradox, Shannon Theory, set theory, Set Theoretic Estimation

I. Introduction

Cryptographic attacks are a key method in extending the listening (sniffing) [1] techniques and recovering usable data for attacking a target. Many algorithms for breaking ciphers have been explored since electronic ciphers were deployed for civilian use and in military theaters of operation. Commercial entities have also employed encryption to protect their designs and operations. Protection of electronic information has taken on increased importance in the last twenty years as malicious actors have monetized attacks on companies and tried to impose their agendas on society and the economic infrastructure. This paper introduces the practice of using collisions to evaluate messages, discovers the original language of a cipher text message, focuses on the possible block size used to encrypt the message, and demonstrates the relative security of the encryption employed for the message.

The remainder of the paper is broken into several sections. Section II holds essential background information for understanding the approach to the problem. Section III addresses the base of Kerckhoff's and Shannon's assumption of *a priori* knowledge of the encryption algorithm used to obscure a message. In Section IV, the collision attack is described. The mathematics of the collision attack are presented in Section V, with the analysis of the technique given in Section VI. The final section, Section VII, presents conclusions and possible future work in the field.

II. Background

A. Collisions

The basis of the work undertaken relies on the concept of data collision. A "collision" is defined as two instances of information that are identical but occur at different points in the transmission or information stream. For any two symbols (s_x located at positions i and j where $i \neq j$

$$s_i = s_j \tag{1}$$

Collisions can also be viewed as a repetition of the same value, symbol, or information. In this study, collisions are a key metric that indicates the redundancy of symbols in a language and are related to the probability mass function (pmf) of the alphabet for the language.

B. Random Number Generators

Randomness is key to the practice of encryption. Secrecy is assured by selecting and applying random keys to encrypt a message. Random numbers are numbers in a set comprised of all numbers in a range from a low to a high number, either as discrete numbers or a continuous range, depending on the random number generator selected for use [2]. A random number function with some probability mass function has the property that if the random function is run sequentially over time for any two locations i and j in the sequence, the outcome of the function will have the same probability of producing the same output as shown in the pmf for the function. That is $\forall i, j$ in s:

$$pr(x_i = x_j) = pr(x) \tag{2}$$

For a pmf that is uniform

$$pr(x_i = x) = \frac{1}{|s|} \tag{3}$$

Functions having this property are known as "true" random numbers (TRNGs). Functions that exhibit TRNG properties include white noise [3], pink noise [4], fair dice, cosmic particles hitting a particular area, and roulette wheels. Unfortunately, randomness cannot be calculated using a digital computer [5].

Computers can be used to create streams of numbers that can "appear" to be random for relatively long runs, known as pseudorandom number generators (PRNGs). These sequences repeat in a period (λ). If λ is almost as large as the size of numbers represented by the bits in the PRNG, ie. $\lambda \approx 2^{|b|}$, then the period is said to be "maximal." During this period, each number in the sequence is unique, This is because the PRNG is a function such that

$$r_{t+1} = f(r_t) \tag{4}$$

Thus, the period repeats when $r_i = r_j$, $j = i + \lambda$, and the numbers between r_i and r_j are unique. The use of the PRNG by independent applications does require synchronization of the PRNGs between those applications. Since the PRNG sequence is deterministic, all that is required is to select a number in the sequence, known as the "seed," and sharing that number with all legitimate users. There are many types of PRNGs, though high quality PRNGs are always being sought for use. Characterizing RNGs can be accomplished using the Birthday Paradox.

C. Birthday Paradox

The Birthday Paradox [6] is the name given to the mathematics that describes why collisions take place more often than would otherwise be expected. Arising from the problem of predicting how many people in a group have the same month and day of birth, the problem is one of statistics. The statement of the problem says if there are n people in a room, find the probability that at least two people share the same birthday. Only $n = 23$ people are required to have a 50% chance of at least two people with matching birthdays. This is because the number of matches (m) that can produce a matching birthday is given by

$$m = \sum_{i=1}^{n-1} i = \frac{n(n-1)}{2} \tag{5}$$

For k possible outcomes and n in the group, the proportion of birthdays without repetition is

$$V_{nr} = \frac{k!}{(k-n)!} \tag{6}$$

With repetition

$$V_t = k^n \tag{7}$$

and the probability of matching birthdays is

$$P(B) = 1 - P(A) = 1 - \frac{V_{nr}}{V_t} \tag{8}$$

This same problem describes random collision problems, but there is one limitation: each outcome must be equally likely.

D. Shannon Theory

The Birthday Paradox addresses randomness, which directly impacts the predictability of observed data, or applications of entropy. Key measures in Shannon theory are entropy, which is a measure of randomness, known as the "surprise" in newly observed data, given by [7]

$$H(x) = -\sum_{i=1}^{n} pr(x_i) lg\big(pr(x_i)\big) \tag{9}$$

where $pr(x_i)$ is the probability of the symbol x_i appearing next in the information stream.

From entropy, the tendency of symbols in a language to be repeated (known as "redundancy") can be calculated. Redundancy is defined as [7]

$$R_{lambda} = 1 - \frac{H(x_i)}{H_{max}(x_i)} \tag{10}$$

From the patterns associated with redundancy and the accumulation of information from a stream of data and information, it is possible to calculate the "unicity distance" (n) of a message. Unicity distance is the number of characters needed to unambiguously eliminate incorrect (spurious) keys in an encryption. Unicity distance is given by [7]

$$n = \frac{log|K|}{R_{\lambda} log|A|} \tag{11}$$

where $|K|$ is the size of the keyspace for the cipher and $|A|$ is the size of the alphabet for the encrypted message. Shannon Theory is contained in Information Theory.

E. Information Theory

Information Theory (IT) is the study of the quantification, methods of storing, and means of communicating information [8]. This area of mathematics was pioneered by Claude Shannon in the 1940's and 1950's as he applied different statistical techniques to the information content of both data communications and encryption (data obscuring). A major result of the study was that the information in encryption does not decrease, but only changes in representation. IT brings in a number of different areas of information and signal processing, including set theory, encryption, redundancy, error checking, compression, and information transmission. Artificial intelligence (AI) and Machine Learning (ML) are also thought to be included in this field through the application of the Asymptotic Equipartition Property (AEP) [8].

F. Isomorphic Cipher Reduction

Abstract algebra [9] can also be applied to encryption. A major result of abstract algebra is the concept that if two problems are solved using the same mathematics, then the problems must be related. Horst Feistel from IBM used this approach when he wrote that at their heart every cipher is a substitution (S) cipher [10]. Every cipher can be described by using the mapping function ($\mapsto$)

$$E_k(PT_i) = CT_i \rightarrow PT_i \mapsto CT_i \tag{12}$$

Taking advantage of this property, all ciphers may be compared using the 1:1 principle. Even complex product ciphers [11] can be reduced to a single cipher using this technique [12], [13]. Once a cipher is converted into its equivalent S cipher, product ciphers may also be further converted into a single equivalent S cipher using the property of idempotence [14], [15]. By applying these techniques and principles any two ciphers can be compared to each other. The reduction of ciphers allows the use of language statistics for recovering encrypted messages.

G. Language Statistics

Breaking and protecting encrypted messages involves working with language patterns. The reason that the use of such statistics so is valuable is that language use is not random - patterns are replete in communications. If language was truly random, there would be no way to follow totally random symbols.

One of the main properties of S ciphers is that they do not obscure patterns but rather allow that information to be transferred from plain text to cipher text [15]. Patterns are inherent in language both for individuals and the language as a whole. On the average [16], patterns arise from the habits of the person doing the communications and from the syntax and lexicon of that particular language. Among those statistics are the pmf of the symbols, Shannon measures for the language, and the repetition of words in typical language use [17]. These statistics provide valuable clues and information for a cryptographer.

III. Assuming Knowledge of Encryption Characteristics

Both Kerckhoffs [18] and Shannon [7] assumed that the attacker knows the encryption algorithm, block size, and other key characteristics of the encryption process. As early as the late 1880's, Kerckhoff noted that encryption algorithms must be mathematically secure by having a shared secret

among legitimate parties to the message. Termed the "secret key," it was assumed that the attacker is also skilled at the mathematics involved in the encryption process. Carlson has argued that this is correct and that isomorphic cipher reduction further demonstrates that only a rudimentary knowledge is necessary to successfully prosecute attacks on ciphers [12], [13], [15]. The assumption that an attacker has an understanding of important cipher characteristics has not been proven, but has been relied upon for cryptanalysis. Applications of the techniques related to this paper will show that the assumption is provable and that the analysis is easy to apply. Therefore, knowledge of key characteristics is warranted and is not merely an academic constraint.

IV. Visualizing the Collision Attack

A critical measure related to the collision attack is the number of collisions that have occurred to a particular location in the file/message. Graphing the number of collisions seen at any point results in a visual indication of the redundancy in the file. Collisions are often perceived to be a rare event since PRNG functions do not collide during their periodic cycle. However, random numbers produced by truly random number generators often collide since the probability of any number reappearing in a sequence is directly dependent on their pmf. A random number has the same probability of occurring for each number in a sequence [19]. Assume that a number has a probability of occurring given by the pmf as C. Then $\forall i, j$ positions in a sequence where $i \neq j$:

$$pr(x_i = C) = pr(x_j = C) \tag{13}$$

The probability remains the same for any two characters. The mathematics of this event are governed by the Birthday paradox [6], [20]. However, language is not random as it involves habits and syntactic rules [16] and has a characteristic redundancy of symbols [7] and words [17]. That redundancy will be reflected in repeated symbols and patterns.

The first step in the process of investigating the plain text language of the message is to find the limit of the anticipated redundancy in a language. Since each language has a defined number of characters in the alphabet, the maximum entropy for the language takes place when all of the characters in a message are randomly selected. This "encryption" represents the case where the fewest collisions will occur. It is not really necessary to do encryption since all ciphers can be replaced by S ciphers [13] and S ciphers leak patterns. The results for those selections of random characters are shown in Figure 5 through Figure 7. Each curve in a figure represents that random selection of characters for a given alphabet size. Each figure shows the results for a particular m-gram size from $m = 1$ to $m = 5$ where an m-gram is m consecutive letters. Since more alphabet characters in the metalanguage form larger values of m, curves flatten out for larger values of m because collisions come more slowly.

With the lower limit of collisions being related to randomness, the PRNG used to create the alphabet character stream should be as close to truly random as is possible. The data stream has no constraints from the language that would impose bias on the resulting pmf. Data is then only dependent on the PRNG and the size of the alphabet. Statistical significance for the information is reached by running many tests (at least 47 cases [19]) and using the result as the representation of the average by applying the Law of Large Numbers (LLN) [21].

The next step is to take the count for actual examples in the language. A sufficient number of representative texts is selected from a source, such as Project Gutenberg [22], and the collisions are counted. If the number of texts is sufficient, the resulting average curve will represent the average for the language through the LLN.

Collision attacks are those which rely on collisions of data of some type in order to affect and attack an encrypted message. Consider the plot of the number of collisions for a language. Of special interest is the slope of the plot at a particular point x where y is the number of collisions at symbol number x in the file or message sequence.

$$y = mx + b \tag{14}$$

The value of the slope at the point x in the message is m. For comparison between different curves (files/messages), the slope of interest occurs when $m = .8$ due to the fact a significant corpus has been seen and sufficient collisions present. Although this slope was chosen to be statistically significant, it

does not require that all alphabetic characters be present. Eventually, if each character in the language is found in the file, then $m \to 1$ and every $s \in \{A\}$ has been seen at least one time. At this point, the curve can produce no new information thus there is no reason to continue further analysis.

A characteristic of the collision plot is that the y-intercept will always be less than or equal to 0 and individual messages/files will vary

$$y_{max} \leq b \leq y_{min} \tag{15}$$

On average, the first collision should occur at the redundancy rate of the language. This rate is bounded by the Birthday paradox. The maximum number of symbols needed to produce a collision is $|A|$, but thereafter each new symbol seen will result in a collision.

Collisions are also used to directly attack ciphers. Both McGrew [23] and Carlson [24] used collisions to break modes protecting AES and other ciphers. Using collisions to characterize the encryption method demonstrates the importance of collision in the encryption process.

V. The Mathematics

Graphing the collision data allows for analysis of the information conveyed by those collisions. The graph is directly dependent on the size of the alphabet used in the language of the message. The cumulative number of collisions for the location in the message characterizes that message in that language. The curve will be bounded by pure random data which never occurs on average. The difference between the curve for a language [16] and the curve for the random data is defined as Δ, and is defined as

$$\Delta = r_i - \lambda_i \tag{16}$$

where Δ is the relative strength of the message versus pure randomness at the point where $m_\lambda = .8$ (see Figure 1). λ_i is the point on the language curve and r_i on the random curve is the point on the random curve.

Two languages can be compared in the same way. The amount Δ should be seen as the cost of insecurity from random imposed by the syntax and symbols used for the language. It would be good to think of this in terms of Chomsky's observations on syntax and lexicons [25].

Fig. 1. Relative Strength (Delta) Between Curves

The inherent security of a language is measured by the application of the average curve for messages of a language λ. This curve will be defined by $|A|$ and the syntax (including words) for the language. Each language will have a unique curve and a unique range of b values. In fact, each author has his/her own unique language, and therefore, each author can eventually be identified by this value.

Taking the same message and encrypting it with the various ciphers will give relative security for each cipher. Note that Feistel said that all ciphers are equivalent to the S cipher [10]. If this is correct, then the curves are symbol assignment independent and should be identical. Again, a statistically significant corpus is needed to create the average curve for the messages.

Using the slope of $m = .8$ and calculating the y-intercept gives a number that describes the message curve. A range of values for b is possible for a block size of encryption and the language of the message. Each value on the y-axis of the graph is associated with different language/block size pairs. One of the pairs is correct for the message, but which pair is correct is not immediately identifiable just from b. However, knowing the possible alphabet sizes for the source language can greatly reduce the exact pair combinations. The more that is known about the message source, the fewer pairs that must be investigated. Therefore, the message may be represented by the y-intercept and use that to map to the possible language/block size pairs.

VI. Analysis

There are two generally accepted divisions of language [26]: natural and formal languages. Natural languages are composed of those languages which arise from the evolution of human beings in groups that result in an agreed upon set of symbols and rules of syntax that are mapped to phonemes (sounds) [27]. These languages change with response to language use over time, although the syntax (rules) tend to remain relatively static. The base mappings of sounds to characters, called the alphabet, may change over time, but the evolution of the alphabet is often much slower than the change in the lexicon.

Formal languages are those languages that are "designed" [28]. The language is created according to rules. Changes to the language, as necessary, also conform to the rules. An example of such a type of language is mathematics, another is music. Symbols and lexicon in the language are very well defined and agreed upon. The syntax is similarly well-defined and enforced.

Every language has a characteristic alphabet (see Table I) which places it in a smaller set of languages with the same alphabet size. The question of which language is used for the message can still be discerned by the data from the lexicon of the language. Comparing the message curve to the random curves can eliminate languages with smaller alphabet sizes. Additionally, the average curve for a particular alphabet size will overlay the message curve within the variance and can identify the set of languages with the correct alphabet size.

One of the interesting features of these curves is that they can be used to compare the strength of two (or more) ciphers. The strongest cipher is one in which there are no patterns to find, ie., totally random data. A curve of totally random numbers will be on the far right of any set of collision curves for a cipher, message, and key. The offset from this curve to a curve representing a particular key and message represents a loss of security from perfect encryption for that message, algorithm, and key. If the curve is composed of the average data, then the performance of two different ciphers can be compared. Curves made up of the data for the same message using different keys and encryption algorithms can be compared. If these two curves are identical, then the two encryptions are equivalent in strength. Furthermore, comparing a cipher and key pair for one algorithm and the S cipher, it is possible to gather data proving isomorphic cipher reduction for a cipher and the S cipher. Ciphers without randomization all reduce to the S cipher [13], so all ciphers should be identical with the curve for the S cipher. This requires testing to empirically prove the assertion, preferably with at least two languages with the same alphabet size.

Alphabet	Size
Hangul [29]	24
English [30]	26
French [31]	26
Dutch [32]	26
Spanish [33]	27
German [34]	30
Serbian [35]	30
Bulgarian [36]	30
Russian [37]	32
Ukrainian [38]	33
Hindi [39]	46
Katakana [40]	71
Hiragana [40]	71

TABLE I
Size of Various Alphabets

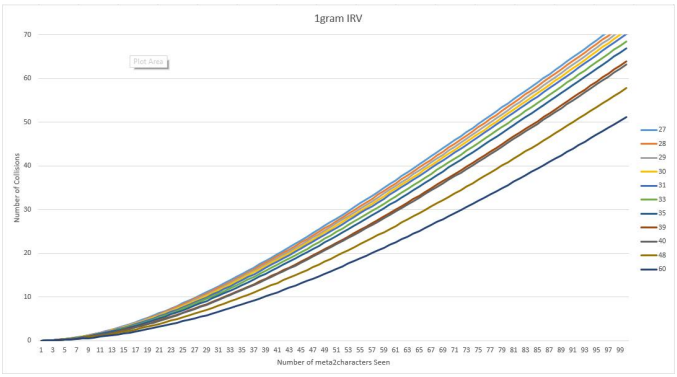

Fig. 2. Results for Single Character Block

VII. Conclusions and Future Work

This paper has focused on collisions as a metric for encrypted messages, using them to determine the possible block sizes of the encryption algorithm employed and the original language used for the message. Deriving this knowledge supports Kerckhoffs' and Shannon's assertion that such knowledge is available to the attacker when they attempt to

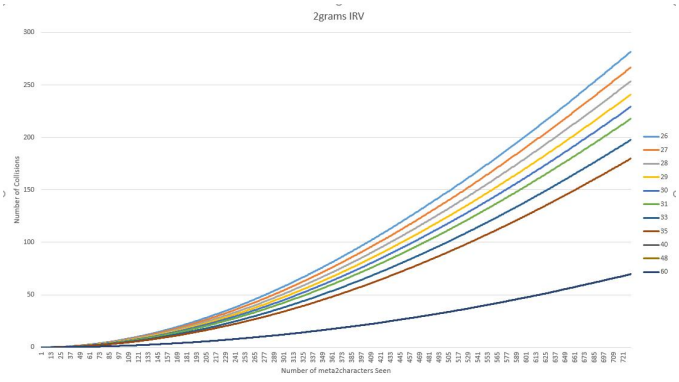

Fig. 3. Results for Two Character Block

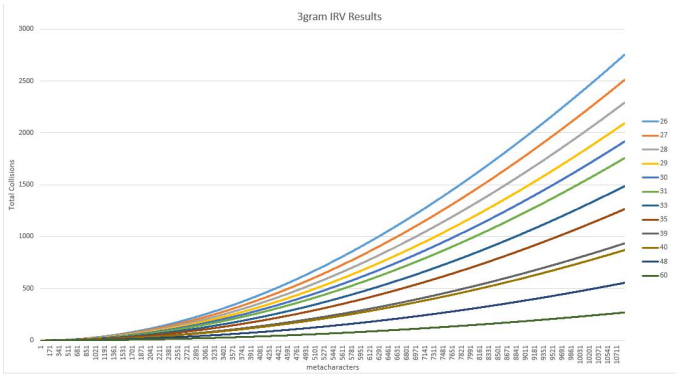

Fig. 4. Results for Three Character Block

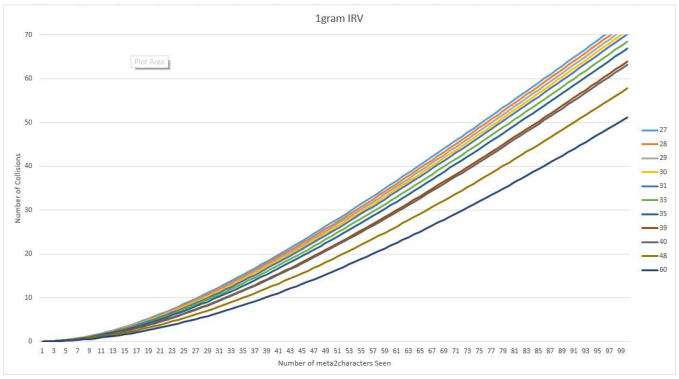

Fig. 5. Results for Single Character Block

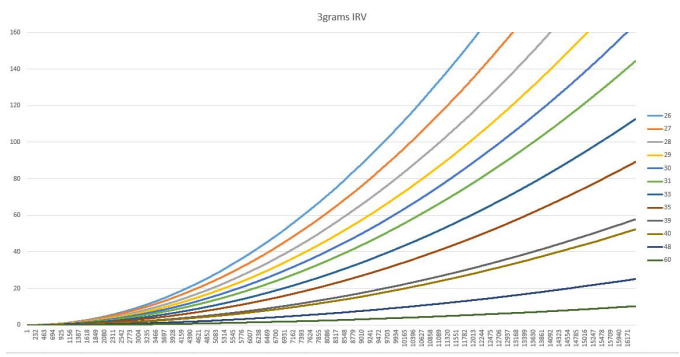

Fig. 6. Results for Four Character Block

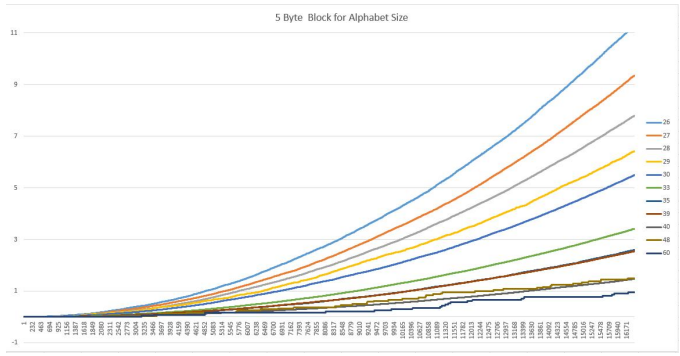

Fig. 7. Results for Five Character Block

decrypt the hidden message. Figures 5 - 7 show the random number graphs for languages that have different alphabet sizes. This indicates that there is a difference in security for each size.

Messages can now be described by the y-intercept for their collision curve. Encoded in that number are the possible block size and language for the message. This information focuses on possible approaches to breaking the message and restricts the number of possible encryption methods to investigate. More information can be derived from the curves and y-intercept. More research is indicated on this metric.

This study centered on the English language, due to the background of the investigators. The same experiment should be conducted on other languages to verify the assertion that this effect is language-independent. Such experiments are planned for the future and will be reported as they are completed.

There is a question about whether different languages with different syntax rules and lexicons have a different curve. This result is very characteristic, since each language has its own pmf. An interesting experiment would be to test two languages with the same number of characters in their alphabet to see if the two languages are distinguishable. The two languages will have different lexicons and pmfs. Different pmfs should result in different collision rates, and therefore, different curves for the average

collision rate.

The same experiment should be run for two different authors or language users to determine if they can be differentiated using the collision curves. Messages encode the pmf and habits of each user. If the result is two distinct curves, then the outcome of the experiment supports the concept of habits in language suggested by stylometry [41] and the Theory of the Vastness of Natural Languages [16]. Such a result indicates that this method can be used in security as a signature to uniquely identify the author of a message.

Implied in the curves is the concept of the inherent security of a language and the ability to measure the security of an encrypted message. The methodology also allows verification of isomorphic cipher reduction. If the curve for an encrypted message using a particular cipher is overlaid with the same message encrypted with the S cipher and the two are identical, then cipher reduction is confirmed.

REFERENCES

[1] P. Anu and S. Vimala. A survey on sniffing attacks on computer networks. *2017 International Conference on Intelligent Computing and Control (I2C2)*, pages 1–5, 06 2017.

[2] P. L'Ecuyer. Random numbers for simulation. *Communications of the ACM*, pages 85 – 98, 1990.

[3] Hui-Hsiung Kuo. *White Noise Distribution Theory*. CRC Press, Boca Raton, 1996.

[4] Allen Downey. *Think Complexity*. O'Reilly Media.

[5] Rod Downey and Denis R. Hirschfeldt. Algorithmic randomness. *Communications of the ACM*, 62(5):70 – 80, 2019.

[6] E. H. McKinney. Generalized birthday problem. *The American Mathematical Monthly*, 73(4):385–387, 1966.

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

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

[9] John B. Fraleigh. *A First Course in Abstract Algebra*. Addison-Wesley, 7^{th} edition, 2003.

[10] Horst Feistel. Cryptography and computer privacy. *Scientific American*, 228(5):15 – 20, 1973.

[11] Uli Maurer and James Massey. Cascade ciphers: The importance of being first. *Journal of Cryptology*, 6(1):55 – 61, 1993.

[12] Petteri Kaski and Pateric R. J. Ostergård. *Classification Algorithms for Codes and Designs*. Springer, 2006.

[13] Bhaskar Ghosh, Indira Dutta, Shivanjali Khare, Albert Carlson, and Michael Totaro. Isomorphic cipher reduction.

[14] Richard Wells. *Applied Coding and Information Theory*. Prentice Hall, Upper Saddle River, 1999.

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

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

[17] M. Lucks. A constraint satisfaction algorithm for the automated decryption of simple substitution ciphers. In *CRYPTO 1988*. CRYPTO, 1988.

[18] Auguste Kerckhoffs. La cryptographie militaire. *Journal des sciences militaires*, IX:5 – 83, 161 – 191, 1883.

[19] R. Lyman Ott and Michael T. Longnecker. *An Introduction to Statistical Methods and Data Analysis 7_{th} edition*. Cengage Learning, 2016.

[20] Kyle Siegrist. The birthday problem, https://www.randomservices.org /random/urn/birthday.html, 2018.

[21] Sheldon Ross. *A First Course in Probability*. MacMillan Publishing, Inc, New York, 1976.

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

[23] David McGrew. Impossible plaintext cryptanalysis and probable-plaintext collision attacks of 64-bit block cipher modes. In *Proceedings of the Fast Software Encryption Workshop*, 2013.

[24] Albert Carlson, Bhaskar Ghosh, and India K. Dutta. Using the collision attack for breaking cryptographic modes.

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

[26] Stefano Crespi Reghizzi. *Formal Languages and Compilation*. Springer, London, UK, 2009.

[27] Michael Garman. *Psycholinguistics*. Cambridge University Press, 1990.

[28] James Allen. *Natural Language Understanding*. Benjamine/Cummings, Redwood City, 1987.

[29] Jennie Lee. *Learn Korean – The Language Workbook for Beginners: An Easy, Step-by-Step Study Book and Writing Practice Guide for Learning How to Read, Write, and Speaking Korean*. Polyscholar, 2021.

[30] Angus Stevenson, editor. *The Oxford Dictionary of English*. Oxford University Press, 3^{rd} edition.

[31] S. Hélène Deacon, Alain Desrochers, and Kyle Levesque. *The Cross-linguistic Study of Reading and Spelling Acquisition: The case of French*, chapter Reading Acquisition across languages and writing systems: An international handbook. Cambridge-UniversityPress, 2015.

[32] Jane Fenoulhet. *Dutch in 3 Months with Free Audio App: Your Essential Guide to Understanding and Speaking Dutch*. DK.

[33] Ralph Penny. *A History of the Spanish Language*. Cambridge University Press.

[34] Langenscheidt KG. *Langenscheidt's German-English English-German Dictionary*. Pocket Books, 2007.

[35] Lena Dragovic. *Learn to Read Serbian in 5 Days*. Wolfdale Press.

[36] Ivan G. Iliev. A short history of the cyrillic alphabet. (2).

[37] Nicholas J. Brown. *The New Penguin Russian Course: A Complete Course for Beginners*. Penguin Books, 1996.

[38] Volodymyr Kubijovyč. *Ukraine: A Concise Encyclopædia*. University of Toronto Press.

[39] Natura Lingua. *Learn Hindi in 100 Days: The 100% Natural Method to Finally Get Results with Hindi! (For Beginners)*.

[40] Yasu-Hiko Tohsaku. *Yookoso An Invitation to Contemporary Japanese*. McGraw-Hill.

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