Method Article

Enhancing Data Access Security through the Utilization of [2048,512,4] Linear Code

Abstract

Linear code is a type of error-correcting code that satisfies the property of closure under addition. Linear codes have been a major component in data security and cryptography and has been proven as an important tool in maintaining privacy. Linear code,[*n,k,d*] is a type of linear code with specific parameters that describe its properties and capabilities with *n* being the length of codewords, *k*, the dimension of the code and *d* denotes the minimum Hamming distance. Code [2048,512,4] is an exampleof [*n,k,d*] Linear code with large codeword length, high dimension and a moderate minimum hamming distance. It was generated by a combination of two of [1024,256,4] linear code and its validity has been verified and confirmed using Gilbert Varshamov bound.This paper proposes a technique to enhance data access security by utilizing linear code [2048,512,4]. The proposed algorithm utilizes a sophisticated linear code equivalence test methodology to create a robust and secure system that generates and validates unique [2048, 512, 4] linear codes. These codes serve as highly effective authentication mechanisms, granting individualized access privileges to different users within the system. Also, by employing cutting-edge techniques in linear code equivalence testing, the algorithm ensures the generation of distinct and non-repeating [2048, 512, 4] linear codes, uniquely tailored to be each user's credentials. The analysis of the results reveals that to breach such a security measure, an intruder would need to contend with a vast number of permutations and combinations. In particular, there are **2048 P 2048** patterns and an astounding 131,328 distinct combinations of codewords that remain unknown to any potential breacher. These findings underscore the robustness and invulnerability of the proposedtechnique for enhancing data access security.The proposed technique also demonstrates a potential for significantly enhancing the protection of sensitive datafor organisations in both pre-quantum and post quantum computing within the digital ecosystem.

Keywords:*Linear Code, [n,k,d] Code, Data Access, Cryptography, Sensitive Data,Gilbert Varshamov Bound.*

1. Introduction

Linear codes have played a major role in data security and cryptosystem over the years as an important tool in maintaining privacy. Linear code, [*n*,*k*,*d*] is a type of linear code with specific parameters that describe its properties and capabilities with *n* being the length of codewords, *k*, the dimension of the code and *d* denoting the minimum Hamming distance. It follows that the code [2048,512,4] is an example of a [*n*,*k*,*d*] Linear code with large codeword length, high dimension and a moderate minimum hamming distance. It was generated by a combination of two of [1024,256,4] linear code using concept by (AA, PB, and NM, 2018) and its validity has been verified and confirmed using Gilbert Varshamov bound in a research conducted by Olaewe (2021).Tô&SafaviNaini (2004), showed that linear coded Kurasawa-Desmedt scheme, can be modified to enable revocation of users, this further indicates that revocation scheme can be derivedfrom a linear code. The researchers further proved that modified scheme is semanticallysecure against passive adversary.Qi, Tang, and Huang (2016) indicated that linear codes can be applied in secret sharing, authentication codes, association schemes, and strongly regular graphs suggesting the effectiveness of linear code in secrecy in communication systems.With the current trend of digitization, it is important to preserve privacy of various sensitivedata available around us. These data can be represented as numerical data, graph data,categorical data, etc. To prevent these data from being illegally used, it is necessary toapply an efficient privacy model (Kar, 2017).

The traditional username and password methodology, once a cornerstone of digital security, is now marred by critical vulnerabilities that undermine its effectiveness. One of the most glaring issues is its susceptibility to data breaches. Moreover, the password's limitations in providing strong security are further magnified by users' behaviours. Human psychology often leads to predictable password choices, like birthdays or common words, that can be easily cracked.The username and password model is a stagnant in the face of modern threats, such as advanced malware and social engineering tactics. Its static nature lacks adaptability to evolving attack methods, rendering it woefully inadequate for safeguarding sensitive information in today's complex digital landscape.

The [*n, k,d*] linear code [2048, 512, 4] exhibits robustness when applied to security systems through the integration of linear code equivalency. Linear code equivalency is a powerful concept that allows researchers to compare different representations of the same code without compromising essential properties. Applying linear code equivalency to the [2048, 512, 4] linear code enhances security by introducing a layer of complexity to potential attackers. The equivalency concept enables multiple valid codeword representations, making it challenging for malicious actors to determine the original data and its representation. This complexity adds a level of obscurity to the code, enhancing the security of sensitive information. Indeed, the versatility nature of the [2048, 512, 4] linear code makes it suitable for various security applications. Its combination of long codeword length and high dimension enables it to handle substantial amounts of data, while its error correction properties ensure accurate data transmission. This versatility allows the code to be integrated into diverse security systems without compromising its robustness.

Since users cannot control data security to ensure confidentiality and integrity of data, there is the need to store data using the traditional computing model, that is saving on personal computer which in turn will require optimal security measures to maintain its integrity, (Sun et al.,

2014).Bertino (2016) opined that data as an asset is more critical for all organization. Recent advances and happenings such as sensor systems, IoT and data analytics are making possible to efficiently and effectively collect data. However, for data to be fully utilized, data security and privacy are critical.Kohnert (2013) carried out research creating optimal linear code [47,15,16] which is relevant to the applied construction. The researcher used high minimum distanced codes as these allow the correction of [(d-1)/2] codes.Linear code was used by (O'Connor & Kleijn, 2019) who investigated the concept of dataprivacy in unbounded public networks where linear codes were used to create hard limitson the number of nodes contributing to a distributed task. Linear codes as an important toolin maintaining privacy was used in wrapping. It is worthy of not that secret sharing schemes were first deployed by Blakly and Shamur in 1979. Since then, many constructions have been proposed. The relationship between Shamir's secret sharing scheme and Reed-Solomon codes was pointed out by McEliece and Sarwate in 1981, after which, several other authors have contributed to secret sharing schemes using linear error correcting codes including Massey who used linear codes for secret sharing schemes and pin pointed the connection between the access structure and the minimal codewords for the dual code of the underlying codes (Tang, Gao, & Zhang, 2013).

Data security in communication and storage systems is very essential in this digital age, (Sun et al. 2014; Bertino, 2016). Some other issues arise because of emergence of new data collection and processing devices such as those utilized in IoT systems, increase the data attack surface.Due to increasing number of companies storing individual and business information on computers and the fact that this information is very sensitive and not meant for the public, block cipher-based cryptographic algorithmsbelonging to both the symmetric and asymmetric could be employed. This paper presents a technique for enhancing data access security by utilizing the Linear code [2048,512,4] which, utilises a sophisticated linear code equivalence test methodology to create a robust and secure system that generates and validates uniqueness of the [2048, 512, 4] linear codes. The rest of the paper is structured as follows: Section 2 presents the proposed technique/scheme which entails the general formulae and algorithm. The test results of the proposed scheme are present in Section 3; Section 4 is the concluding part of the paper.

2. Methodology

2.1 Gilbert Varshamov bound

To determine if the generated linear codes exist, Gilbert Varshamov bound as implemented by (Hoffman et al., 1991) was used in this study.

Theorem: There exist a linear code of length *n*, dimension *k* and distance *d* if

$$\binom{n-1}{0} + \binom{n-1}{1} + \cdots + \binom{n-1}{d-2} < 2^{n-k} \quad \textbf{(1)}$$

Corollary: $if\ n \neq 1, d \neq 1$ then there exists an (*n*,*k*,*d*) linear code with

$$|c| \geq \frac{2^{n-1}}{\binom{n-1}{0} + \binom{n-1}{1} + \cdots + \binom{n-1}{d-2}} \quad (\mathbf{2})$$

The corollary is a lower bound for the number of words in a linear code of length n and distance d.

2.2 Hamming bound

Theorem: Let C be a code of length n and $d = 2t + 1$ then the number of words in the code is given as

$$|c| \leq \frac{2^n}{\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{t}} \quad (3)$$

The Hamming bound is an upper bound for the number of words in a linear code of length n and distance $d = 2t + 1$.

2.3 Equivalent Test Algorithm

The Optimized Linear Code Equivalent Test Algorithm by (Olaewe,2021) is used for authenticating linear code as follows:

1. Check if the two codes, C and C ' are linear codes by verifying that the two codes satisfy the properties of a linear code.
2. Locate the positions of 1's in the codewords of C and C '.
3. Find $d = d_1 - d_2$, where d_1 and d_2 are the positions of 1's in codewords of C and C' respectively.
4. Compute sum of d's denoted by $\sum d$
5. If $\sum d = 0$ then $C \cong C'$ otherwise they not equivalent.

2.4 The (U | U+ V)Construction

The work in (Ibrahim, Chun, & Kamoh, 2018) established thata new linear code can be formed by combining two codes of the same length, with this approach, the newly formed code would be twice the length in a way similar to the direct sum of the codes construction. This is achieved as follows:

Let C_i be an [n,k,d] code for $i \in \{1,2\}$, both over the same finite field F_q. The $(u\,|u + v)$ construction produces a $[2n, k_1 + k_2, min(2d_1, d_2)]$ linear code.

$$C = \{(u, u + v) | u \in C_1, v \in C_2\} \quad (4)$$

3. Results&discussion :

The proposed data security algorithm is derived from the optimised linear equivalency test algorithm. The objective of the scheme is achieved by enabling a system to provide more than one authentication codes for different users to access the same system. This algorithm enhances data security as it is impossible to get or manipulate an equivalent code of a linear code with higher [n,k,d] parameters, hence making it an efficient way of ensuring data integrity in systems. In the case of a system using linear code [2048,512,4] for its security, a breacher would need to

guess **2048P2048** entries as well as **512C2** codewords to gain access such system. The difficulty of an attacker breaching proposed data security can be expressed mathematically as follows:

$$Total\ Guesses\ = P(2048,2048)\ \times C(512,2) \quad \textbf{(5)}$$

The proposed data security schema is predicated on an adversary's requirement to conjecture a precise sequence from a permutation of 2048 distinct entries, denoted by $P(2048,2048)$ and to identify a correct pair from a combination of 512 codewords, denoted by $C(512,2)$. In computational complexity theory, the permutation problem is a factorial-time problem with 2048! Representing the factorial of 2048, delineating the total number of possible arrangements of the set. The combination problem is less computationally intensive than the permutation counterpart, still presents a significant challenge due to the substantial number of possible pairs.

From the perspective of nondeterministic polynomial (NP) complexity class, the crux of the scheme lies in the verification of potential solutions. For the permutation scenario, confirming the validity of a proposed sequence of 2048 unique entries against an established set is an operation that can be executed within polynomial time bounds. Similarly, the affirmation of a pair as a legitimate combination drawn from 512 codewords is also a polynomial-time endeavor. Nonetheless, the discovery of the correct permutation or combination via exhaustive enumeration is presumed to be NP-hard, as it may necessitate a number of guesses that grows exponentially with the size of the element set.

It is imperative to underscore that while NP-hardness traditionally pertains to decision problems—those yielding binary outcomes—the computation of permutations and combinations inherently constitutes a counting problem. However, an associated decision problem can be formulated, for instance, determining the existence of a permutation or combination that fulfills a set of predefined constraints. Should these constraints be verifiable in an expedient manner, the problem could be classified within the NP domain. This subtle yet profound distinction underscores the theoretical complexity of the authentication algorithm and illuminates the daunting task faced by potential unauthorized entities attempting to breach the scheme.

For authentication, the proposed algorithm uses the optimised linear equivalent algorithm to verify the provided authentication code and grant access only if the code passes the equivalence test.

Steps

1. Request for authentication code (a linear code)
2. Check if provided code is equivalent to original code using (Olaewe, 2021) optimized algorithm.
3. Grant access to data if step 2 shows that it is equivalent otherwise revoke access

List of Linear Codesused for the generation of [2048,512,4]

i. [8,2,4]
ii. [16,4,4]
iii. [32,8,4]
iv. [64,16,4]
v. [128,32,4]

vi. [256,64,4]
vii. [512,128,4]
viii. [1024,256,4]
ix. [2048,512,4]

Testing existence of [2048,512,4] used in this research

$$\boldsymbol{n = 2048, k = 512, d = 4}$$

$$= \binom{2047}{0} + \binom{2047}{1} + \binom{2047}{2}$$
$$= Y$$
$$Also, 2^{2048-512}$$
$$= 2.290 \times 10^{463}$$
$$2.290 \times 10^{463} < Y \{\text{true}\}$$
This implies linear code [2048,512,4]exists

Hamming bound computation for some of the Linear codes used.

i. $n = 8, k = 2, d = 4$

$$t = \left\lfloor \frac{4-1}{2} \right\rfloor = 1$$

$$|c| \leq \frac{2^8}{\binom{8}{0} + \binom{8}{1}} = \frac{256}{1+8} = \frac{256}{9}$$

$$\leq 28.44$$

$|c| \leq 2^4 = 16$ codewords

Computing the lower bound

$$|c| \geq \frac{2^7}{\binom{7}{0} + \binom{7}{1} + \binom{7}{2}}$$

$$\geq \frac{132}{1+7+21} = \frac{132}{29}$$

$$\geq 4.5517$$

$\geq 2^3 = 8$ codewords

ii. $n = 16, k = 4, d = 4$

$$t = \left\lfloor \frac{4-1}{2} \right\rfloor = 1$$

$$|c| \leq \frac{2^{16}}{\binom{16}{0} + \binom{16}{1}}$$

$$= \frac{65536}{1+16} = \frac{65536}{17}$$

$$\leq 3855.05$$

$|c| \leq 2^{11} = 2048$ codewords

Computing the lower bound

$$|c| \geq \frac{2^{15}}{\binom{15}{0} + \binom{15}{1} + \binom{15}{2}}$$

$$\geq \frac{32768}{1+15+105}$$

$$= \frac{32768}{121}$$

$$\geq 270.8099$$

$\geq 2^9 = 512$ codewords

iii. $n = 32, k = 8, \ n = 4$

$$t = \left\lfloor \frac{4-1}{2} \right\rfloor = 1$$

$$|c| \leq \frac{2^{32}}{\binom{32}{0} + \binom{32}{1}}$$

$$= \frac{4{,}294{,}967{,}296}{1+32} = \frac{4{,}294{,}967{,}296}{33}$$

$$\leq 130{,}150{,}524.12$$

$|c| \leq 2^{26} = 67{,}108{,}864$ codewords

Computing the lower bound

$$|c| \geq \frac{2^{31}}{\binom{31}{0} + \binom{31}{1} + \binom{31}{2}}$$

$$\geq \frac{32768}{1+31+496}$$

$$= \frac{32768}{528}$$

$$\geq 62.0606$$

$$\geq 2^6 = 64 \text{ codewords}$$

iv. $n = 64, k = 16, d = 4$

$$t = \left\lfloor \frac{4-1}{2} \right\rfloor = 1$$

$$|c| \leq \frac{2^{64}}{\binom{64}{0} + \binom{64}{1}}$$

$$= \frac{18,446,744,073,709,551,616}{1+64} = \frac{18,446,744,073,709,551,616}{65}$$

$$\leq 283,796,062,672,454,640.25$$

$|c| \leq 2^{57} = 144,115,188,075,855,872$ codewords

Computing the lower bound

$$|c| \geq \frac{2^{63}}{\binom{63}{0} + \binom{63}{1} + \binom{63}{2}}$$

$$\geq \frac{9,223,372,036,854,775,808}{1+63+1953}$$

$$= \frac{9,223,372,036,854,775,808}{2017}$$

$$\geq 4,572,817,073,304,301.3426$$

$$\geq 2^{53} = 9,007,199,254,740,992 \text{ codewords}$$

4. CONCLUSION

The utilization of the [2048, 512, 4] linear code marks a significant advancement in data access security. Characterized by its extensive codeword length, high dimensionality, and moderate Hamming distance, this code serves as a robust mechanism for authentication within security systems. The innovative algorithm that leverages linear code equivalency generates unique and non-repeating codes, enabling personalized access privileges for users. This paper has underscored the wide-ranging applications of linear codes in enhancing data privacy and security across domains such as cryptography, secret sharing, and communication systems. Theoretical foundations, validated through mathematical theorems like the Gilbert-Varshamov bound and Hamming bound, confirm the robustness and existence of the [2048, 512, 4] linear code.

As we look to the future, further research is encouraged in the optimization and implementation of this linear code in emerging technologies like quantum computing and Internet of Things (IoT) devices. There is also potential for exploring the integration of these codes with artificial intelligence algorithms to further enhance security protocols. Additionally, adapting this linear code for more efficient error correction in high-speed data transmission remains a promising avenue. As organizations continue to face sophisticated cyber threats, the integration of the [2048, 512, 4] linear code represents a forward-looking solution, offering a level of security that surpasses traditional methods. Its intricate layer of authentication, combined with adaptability across various security systems, positions it as an essential safeguard for sensitive data in the increasingly digital and interconnected world.

References

Bertino, E. (2016). Data Security and Privacy: Concepts, Approaches, and Research Directions. *Proceedings - International Computer Software and Applications Conference*, pp. 400–407. https://doi.org/10.1109/COMPSAC.2016.89

Ding, C., Heng, Z., & Zhou, Z. (2018). Minimal Binary Linear Codes. *IEEE Transactions on Information Theory*, 6536–6545. https://doi.org/10.1109/TIT.2018.2819196

Ibrahim, A., Chun, P., & Kamoh, N. (2018). A New [14 8 3]-Linear Code From the Aunu Generated [7 4 2] -Linear Code and the Known [7 4 3] Hamming Code Using the (U|U+V) Construction. *Journal of Applied & Computational Mathematics*, *07*(01), pp.1–3. https://doi.org/10.4172/2168-9679.1000379

Kar, T. S. (2017). *A Study on Privacy Preserving Data Publishing with Differential Privacy*. University of Saskatchewan.

Kohnert, A. (2013). Linear binary block code with prescribed minimum and maximum weight. *Electronics Letters*, pp. 541–543. https://doi.org/10.1049/el.2012.4232

Olufemi Olaewe (2021). An Optimized Hoffman Algorithm for Testing Linear Code Equivalency, Unpublished Thesis, University for Development Studies, Tamale, Ghana.

Qi, Y., Tang, C., & Huang, D. (2016). Binary linear codes with few weights. *IEEE Communications Letters*, pp.208–211. https://doi.org/10.1109/LCOMM.2015.2506576

Silva, D., & Kschischang, F. R. (2009). On metrics for error correction in network coding. *IEEE Transactions on Information Theory*, pp.5479–5490. https://doi.org/10.1109/TIT.2009.2032817

Tang, C., Gao, S., & Zhang, C. (2013). The optimal linear secret sharing scheme for any given access structure. *Journal of Systems Science and Complexity*, pp.634–649. https://doi.org/10.1007/s11424-013-2131-4

Tô, V. D., & Safavi-Naini, R. (2004). Linear code implies public-key traitor tracing with revocation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp.24–35. https://doi.org/10.1007/978-3-540-27800-9_3