A Novel Algorithm of Coset Coding for Wiretap Channel in Physical Layer Security

Hui SHI
Institute of Electronic Science and Engineering
National University of Defense Technology
Changsha, Hunan Province, China
e-mail: shihui11@nudt.edu.mtn

Jing LEI
Institute of Electronic Science and Engineering
National University of Defense Technology
Changsha, Hunan Province, China
e-mail: leijing@nudt.edu.cn

Erbao LING
Institute of Electronic Science and Engineering
National University of Defense Technology
Changsha, Hunan Province, China
e-mail: lierbaonudt@gmail.com

Abstract—Coset coding is a typical method of secure coding in physical layer security. This paper focuses on coset coding based on the random linear codes in binary symmetric channel. While analyzing system security, we derive the formula of eavesdropper’s equivocation and proposed a novel code extending algorithm based on the optimization of minimum distance. On the basis of ensuring minimum distance of the code reaching the maximum, the algorithm searches the achievable extended column to ensure the equivocation of eavesdropper up to the most as well. The simulations show, compared with the best known code (BKC) and best equivocation code (BEC) of the same code length and rate, the proposed algorithm has better security and higher efficiency.

Keywords—physical layer security; coset coding; equivocation; minimum distance

I. INTRODUCTION

Physics security coding is an important composition of Physical layer security, the basic idea of which is increasing the security constraint by using channel differences in the tradition channel coding and modulation technique and making the legitimate receiver decoding correctly while the eavesdropper meets the high bit error rate. Therefore, this technique is a method of secure communication independent of the eavesdropper’s limit on his computing power.

Part of the research on secure coding thought it possible to achieve appropriately noiseless of the main channel by power control, man-made noise and et al. Hence, no need to consider the reliability constraint, we should give priority to the construction of the method of secure communication under the security constraint, among which the typical one is coset coding. Traditional channel coding increases redundancy by increasing the length of parity bits, so part of the code-word space is used. However, coset coding divides the whole space into multiple cosets and randomly choose one of corresponding coset to replace the secret information when transmits, which ensures the secure transmission in wiretap channel. Obviously, compared with traditional channel coding, both the randomness and security of coset coding are better.

In [5], the concept of coset coding was firstly proposed. Cohen and Zemor [6] provided a brief analysis on information leakage of coset coding in wiretap channel. Chen and Vinck [7] proved that, coset coding with random linear codes can achieve the secrecy capacity. In the paper above, only theoretical derivation on the performance of coset coding was made, the implement method in detail was not underway yet. In [8-9], Klinc, Andersson and et al. added randomness to increase the uncertainty of the eavesdropper. However, these methods decreased the reliability of the legitimate receiver in reverse. Zhang, Tomlinson and et al. [10] proposed a syndrome coding algorithm based on the iteration of hash function, which efficiently hid secrete message in the cosets. The hash function searches fastly but occupies plenty of memory at the same time. YI Ming, JI Xin-sheng and et al. [11] constructed a coset coding algorithm under the Gaussian channel, which greatly reduced the bit error rate of the legitimate receiver, but the system security increased unobviously. Zhang, Tomlinson and et al. [12] extended the parity check matrix and makes the security of syndrome coding better by searching the extended column with the maximum equivocation. However, along with the increasing of the length of information bits, the search scope would increase at a exponential function, which brought a high complexity.

In this paper, we mainly analyze and improve the algorithm proposed in [12], combined merits and demerits of the recent works. We propose a novel coset coding algorithm, which consider the effect of the minimum distance on the performance of the code. Good security codes are designed in combination with the method of matrix extending. We give a detail theoretical derivation and simulation proof later in the article. The simulation proves that, compared with the results in [12], the algorithm we proposed has higher search efficiency as well as further improved security.
II. SYSTEM MODEL AND RELATED CONCEPT

A. System Model

We use the system model in Fig.1. Assume that the main channel is noiseless and the wiretap channel is binary symmetric channel (BSC), where Alice and Bob are both legitimate communication sides and Eve is the eavesdropper.

![Wiretap channel model](image)

Theoretical proof shows, using random linear codes for coset coding in binary symmetric channel can reach the ideal secrecy capacity of the channel [7]. On the basis of the formula for the secrecy capacity proposed by Wyner, the secrecy capacity in this channel is

\[ C_s = -p_e \times \log_2(p_e) - (1 - p_e) \times \log_2(1 - p_e) \]  

where \( p_e \) denotes the error transition probability of the wiretap channel.

B. Related Concept

Assume that the length of binary linear block code \( C \) is \( n \), the length of information bits is \( k \), the length of parity bits is \( m = n - k \), minimum distance of the code is \( d \). Based on the code-word set \( F_2^n \), we divide n-dimensional space in binary field into \( 2^k \) subspaces which are mutually disjointed. Each of the subspaces is called a coset and the code-word set \( F_2^n \) is called the mother code of the coset divided. Assume that \( H \) is the parity check matrix of code \( C \), when \( c \in F_2^n \), the vector \( s \in F_2^m \) which satisfy

\[ s = c \cdot H^T \]  

is called the syndrome of code \( C \). If \( c \in C \), then \( s = 0 \). There is a one-to-one relationship between the correctable error pattern and the syndrome. Assume \( \sigma \) is the syndrome function of code \( C \) and it satisfy the formula below

\[ \sigma : \{0,1\}^k \rightarrow \{0,1\}^m \]  

\[ x \rightarrow H^T x \]  

If the sender Alice wants to communicate with the legitimate receiver Bob, secrete message satisfies \( s \in \{0,1\}^m \), then Alice only needs to choose randomly one of vectors satisfying \( \sigma(x) = s \) as the transmitted information. The process is called coset coding. Because of the noise, the eavesdropper Eve receives scrambling information \( y = x + e \), where \( e \) is the noise vector.

III. NOVEL CODE CONSTRUCTION ALGORITHM PROPOSED

A. Principle of Encoding and Decoding

Any binary linear block code \( (n,k,d) \) can be denoted by its parity check matrix \( H \) and the structure of parity check matrix has an important effect on the performance of the code. In general, parity check matrix of coset coding can be expressed as systematic form as

\[ H = [I_m \mid P] \]  

where \( I \) denotes \( m \) order identity matrix and \( P \) denotes a non-identity matrix. Based on the structure of the parity check matrix, we continue the following encoding and decoding steps:

1) Coset coding

To transmit \( m \) bits sequence \( S \) to Bob, Alice encodes \( S \) as \( n \) bits security code-word \( X \). On the basis of the principle of coset coding,

\[ S = X \cdot H^T \]  

\[ X = R \cdot G + E \]  

where \( R \) denotes the binary random vector whose length is \( k \), \( E \) denotes the error pattern. Therefore, \( R \cdot G \) generates all possible code-word.

The same syndrome can correspond to many error patterns. However, if we choose the lightest one as the coset leader, then all elements in the same coset will correspond to the same syndrome. In traditional coset coding, the relationship between error patterns and syndromes is got by look-up table. For long code, the look-up table occupies too much memory and has a high complexity. Because the syndrome of any code-word is zero, then any code-word plus one of the error patterns will produce the same syndrome. We consider to put \( k \) zeros after the transmit sequence \( S \) and treat it as an error pattern \( E \) in the calculation, that is

\[ E = [S \mid 00 \cdots 0]_k \]  

2) Syndrome decoding

While the main channel is noiseless, the legitimate receiver Bob receives information below.

\[ Y = X = R \cdot G + E \]
When we decode, multiply (9) on both sides by $H^T \cdot$.

$$\hat{S} = Y \cdot H^T = X \cdot H^T = R \cdot G \cdot H^T + E \cdot H^T = S$$ (10)

Because of $G \cdot H^T = 0$, (10) can be simplified to

$$\hat{S} = E \cdot H^T = S$$ (11)

which shows that Bob can successfully recover the original message. However, for Eve, on account of the noise when decode,

$$\hat{S} = E \cdot H^T + E_c \cdot H^T = S + S_c$$ (12)

Considering that the error is difficult to eliminate, so Eve cannot decode successfully.

**B. Derivation and Calculation of Equivocation**

The security of coset coding is measured by the equivocation of the eavesdropper's decoding output $H(S | \hat{S})$, specific process is below.

$$H(S | \hat{S}) = H(S) + H(S_c | S) - H(\hat{S})$$ (13)

Because $S_c$ and $S$ are mutually independent, (13) can be further simplified to

$$H(S | \hat{S}) = H(S_c) = -\sum_{j=0}^{2^n-1} p(S_j) \cdot \log_2 p(S_j)$$ (14)

Normalization of equivocation is

$$\text{Eqv. rate} = H(S | \hat{S}) / m = -\frac{1}{m} \sum_{j=0}^{2^n-1} p(S_j) \cdot \log_2 p(S_j)$$ (15)

Where $H(S_c)$ is the entropy of $S_c$, $S_j$ is any syndrome, $m$ is the length of secret messages. Where the equivocation is larger, the uncertain message of eavesdropper is larger and the security of coset coding is better.

For Eve, there are $2^m$ possible error pattern occurring with the probability below

$$p(E_{BSC}(i)) = p_{e}^w(i) \cdot (1 - p_{e})^{n - w(i)}$$ (16)

where $w(i)$ is the weight of $E_{BSC}(i)$. Therefore, for the same syndrome, its probability is

$$p(S_j) = \sum_{i=0}^{2^n-1} p(E_{BSC}(i)) \cdot \delta(S_C(i) - S_j)$$ (17)

where $\delta(\cdot)$ is the Dirac function. Bringing (17) into (15), we can get the equivocation. However, as the length of the code increases, the number of error patterns becomes an exponential function and the complexity of calculation improves continually. Use the derivation below to simplify.

Assume that the probability mass function of $S_j(j = 0 \sim 2^n - 1)$ can be denoted to $p(S_j) = \beta(j)$, where $\beta(j)$ is the factor of the probability mass function after Z conversion. Denote the function after Z conversion as $p_z(S)$, $p_z(S)$ is decided only by the column of parity check matrix and the error probability of BSC channel $p_e$.

$$p_z(S) = \sum_{j=0}^{2^n-1} \beta(j)Z^j = \prod_{i=0}^{n-1} ((1 - p_e) + p_e \cdot Z^b_i)$$ (18)

where $b_i$ is the integer representation of the column of $H$. Bit-wise exclusive-or the exponent of index of $Z$ separately, that is $Z^{b_0 \oplus b_i}$. Extend one column after matrix $H$, the length of the code changes from $r$ to $r + 1$, then the probability mass function of the extended code is

$$p_z(S_{r+1}) = \prod_{i=0}^{r} ((1 - p_e) + p_e \cdot Z^{b_i}) = p_z(S_r)((1 - p_e) + p_e \cdot Z^b)$$ (19)

Denote the factor of original code with $\beta_r(j)$

$$p_z(S_r) = \sum_{j=0}^{2^n-1} \beta_r(j)Z^j$$ (20)

then for the extended code,

$$p_z(S_{r+1}) = (1 - p_e)\sum_{j=0}^{2^n-1} \beta_r(j)Z^j + p_e \sum_{j=0}^{2^n-1} \beta_r(j)Z^{b_i}$$ (21)

Combine terms of the same index, then (21) can be simplified as

$$p_z(S_{r+1}) = \sum_{j=0}^{2^n-1} \beta_{r+1}(j)Z^j$$ (22)

Obviously, the probability mass function of the code can be got from the recursion process above that $p_z(S_r)$
decides \( p_r(S_r) \), \( p_r(S_r) \) decides \( p_r(S_r) \), and so on. Therefore, when calculate the probability mass function of the extended code, we only need the original parameter of the code which has not been extended. The process greatly improves the efficiency of the algorithm.

C. Code Design Algorithm

For good code that has high equivocation, its probability mass function of syndrome should be as uniform as possible. Because the wiretap channel is BSC channel, when the error probability of channel is very low, the equivocation will be affected more by the lighter error pattern. Therefore, when we design the code, we should consider the following situation:

1. If the weight of the error pattern is lighter, the probability of error is higher.
2. If the code can reach the largest minimum distance \( d \), the number of the non-zero terms in each column of matrix \( P \) should not be less than \( d - 1 \), otherwise the minimum distance of the code will be less than \( d \).
3. Avoid repetitive columns in parity check matrix, otherwise it will produce an all-zero syndrome, which adverse to the homogenization of the distribution of syndrome.

For linear code, the minimum distance is larger, the capacity of correction and inspection of the code is stronger. If \( C \) is a \((n,k,d)\) linear code, whose parity check matrix is \( H \), then the minimum distance \( d \) equals the minimum number of linearly independent columns in \( H \). In all known \((n,k)\) linear block code, \((n,k)\) best known code (BKC) has the largest minimum distance \( d_m \). Considering the effect of minimum distance on the performance of security code, we will construct the code which has largest minimum distance and equivocation, compared with the BKC code. The full detail of algorithm is expanded as follow. When increasing the minimum distance, there is a certain promotion but the amplitude is moderate. We analyze the reason as follow. When increasing the minimum distance, the capacity of correction and inspection of the code has been strengthened, also is reliability being strengthened. However, reliability and security are a couple of contradictory propositions by nature. When we strengthen the reliability, it has an inevitable effect on the promotion of security. Therefore, the amplitude of equivocation just increases slightly. Nonetheless, as a whole, overall performance of the code has been improved.

Step 5: output the maximum equivocation \( \text{Eqv. rate}_{\text{max}} \), the corresponding extended column \( c_{r, \text{max}} \) and the minimum distance of extended code \( d \).

IV. SIMULATION

The simulation is realized in BSC channel. We compare performance of new code with the BKC code and the best equivocation code (BEC) proposed in [12] and call it the best minimum distance code (BMDC).

Firstly, we compare the complexity between different code design algorithms, as is shown in table 1. Considering the binary linear codes have a large amount, we take \( n = 50, m = 11 \) as an example. Obviously, the search scope has been greatly reduced while ensuring the maximum minimum distance, which makes up almost half of the BKC code. Therefore, the algorithm improves the efficiency and applicability of the algorithm.

<table>
<thead>
<tr>
<th>Code Design Algorithm</th>
<th>Minimum Distance</th>
<th>Search Scope of Column</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMDC code</td>
<td>4</td>
<td>1149</td>
</tr>
<tr>
<td>BEC code</td>
<td>3</td>
<td>1998</td>
</tr>
<tr>
<td>BKC code</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>
Finally, we compare the system performance with the uncoded system from the bit error rate (BER) perspective. Parameter is set the same. As is seen above, after the new algorithm, the minimum distance increases from 3 to 4 compared to BEC code. In fig 3, it is obviously that the curve of eavesdropper’s BER is always higher than BEC code and close to 0.5, which approaches the requirements of secure communication. Because the BKC code does not take the security factors into consideration, its equivocation is much lower than the first two codes and increases slowly. On the other hand, performance of the uncoded system has a close connection with channel quality in BER and it is linearity to the channel error probability and therefore, meets a bad security.

The idea of BMDC code can be spreaded to the situation when the main channel is not noiseless, at which moment we should consider reliability and security at the same time. Homogenization of equivocation being higher than BKC code and close to one while \( p_e = 0.15 \), the system has a good security performance. Meanwhile, BMDC code can reach the largest minimum distance that is achievable, so we can foresee that, combining with the concrete system model, the capacity of correction and inspection of the code will not be lower than or even beyond some familiar error correcting code.

V. CONCLUSION

We do research on the coset coding algorithm of binary linear random codes. On the basis of the analysis of the system security, we derive the formula of eavesdropper’s equivocation. Considering the basic idea of coset coding, we construct the parity check matrix of the code and proposed a novel code design algorithm based on the optimization of minimum distance. The algorithm chooses the achievable extended column as required and ensures the minimum distance of the code being the largest. Simulation results show, compared with BKC code and BEC code of the same code length and rate, the newly proposed algorithm can bring more uncertainty to the eavesdropper and has better security.

REFERENCES