Proceedings of CONF-SPML 2027 Symposium: Intelligent Network Security and Machine Learning for Communication Systems
DOI: 10.54254/2755-2721/2027.MELB38424

Security Indexing Coding under the Eavesdropper Model: A Review of Capacity Boundaries and Secure Indexing Methods

Xueyao Wang

School of Cyberspace Security, Xi'an University of Posts and Telecommunications, Xi'an, China
2530541387@qq.com

Abstract. With 5G and Internet technology developing quickly, broadcast and multicast communications are widely applied in many scenarios such as the Internet of Vehicles, edge distributed networks and group transmission for the Internet. Wireless broadcast channels are open, but transmitted signals are easy to intercept and pose a risk of data leakage. This study focuses on a survey of secure index encoding. This work defines the secure index encoding model, adversary model, strong security and weak security. Based on the above, it discusses how to use and where to use linear coding and random linear coding. Through analysis, this paper find that secure index encoding also have some troubles, like inaccurate minimum rank, difficulty in solving optimal coding solution and hard to achievement. This paper reviews the theoretical framework of secure index encoding based on existing research, which can provide a reference for more in-depth research in the future.

Keywords: secure index encoding, eavesdropping security, strong security and weak security

1. Introduction

Index coding is a part of network coding. The concept was proposed by Birk and Kol. Index coding is based on network coding, but they have some differences. Index coding is a single-hop broadcast channel, where source nodes combine packet encoding, and the receiver uses edge information stored locally to recover data [1]. Nowadays, index coding has developed quickly. Security frameworks, including secure eavesdropping, strong security and weak security, have long been established to form different security models. Among them, in the linear coding layer, scenarios such as eavesdropping already exist. Designing coding schemes to achieve information-theoretic security also reveals the problem that random keys cannot provide a secure environment under index encoding [2]. There are different research methods for security index encoding with different model structures, conditional derivations, and scheme construction. This study systematically summarizes various theoretical achievements in capacity boundaries and security index encoding construction methods in this field, clarifies the constraints of different boundary conditions, and provides an understanding of current research achievements. This study is aimed at scholars interested in eavesdropping security and provides a comprehensive understanding for scientific research.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2027.MELB38424

2. Definition of model and eavesdropping security

2.1. Index codes and symbols

The foundational model of index coding was first proposed by Birk and others. In single-task broadcast communication, the source node is constructed for multi-user transmission scenarios [3]. In the Standard Model, there are n pieces of information data, m receiving users, and a set of side information x. When users need to cache content, the remaining portion is retrieved from other users via directed graphs--this is called abstract modeling. Subsequent research established corresponding symbolic systems, using the least rank to characterize the optimal code length of linear index codes. This mathematical tool also became an important foundation for secure index code modeling [4]. Secure index coding adds a link to the base model, allowing eavesdroppers to eavesdrop on data. To prevent eavesdroppers, security constraints must be added to the user model. While ensuring legitimate users operate normally, it also prevents eavesdroppers from obtaining valid information. This model is built on top of index encoding with additional security protection [5].

2.2. Parameterization of rival models

The adversary model refers to a link where the listener can fully access all encoded data packets broadcast from the source, clearly understand the encoding algorithm, system parameters, and cryptographic applications, but cannot access cached user data or extract the original packets, only obtaining the broadcast packet [5]. This is just a passive attack scenario for eavesdropping. But in real life, you also need to consider the existence of multiple eavesdropping issues. Multiple eavesdroppers are not a single type; they can be divided into conspiracy and non-collusion based on whether the eavesdroppers share resources with each other. Collusion involves multiple eavesdroppers communicating to share the data packet to steal all user information; non-collusion means multiple eavesdroppers do not communicate and only steal part of the user's information, equivalent to multiple single eavesdropping. However, it can be achieved through the adversary, eavesdropping parameters, and manipulation parameters to fully characterize the attacker's capabilities [6].

2.3. The spectrum of security definitions

The above content introduces the adversary model, and on this basis, security index encoding further defines two types of security defenses: strong and weak security, to determine the security factor under eavesdropping. Weak security means that eavesdroppers can only steal part of the broadcast information during broadcast transmission and cannot steal all the information, meaning they can provide the thief with part of the information but cannot obtain the complete information. Compared to weak security, strong security protects information more strictly and cannot leak any information [5, 6]. Therefore, the conditions for achieving strong security are stricter than those for weak security.

2.4. Safe feasibility

With these models in place, people need to determine whether they can be implemented and create a secure feasibility. Secure implementation means providing conditions such as data pockets, user requirements, eavesdroppers, and security levels in an index coding environment, allowing users to receive cached information and meet security constraints. Therefore, the necessary condition for

DOI: 10.54254/2755-2721/2027.MELB38424

secure implementation is that, when using linear coding for weak security protection, the data packet (t) obtained by the eavesdropper must not exceed the minimum distance of the matrix minus 2 (d-2), and all data cannot be cached by only one person, so that the eavesdropper can obtain all data via broadcasting; if conducted in a strongly secure environment, the information obtained by the eavesdropper must be less than the total amount of information [2]. Therefore, researchers need to verify whether the scheme meets the necessary conditions when designing secure index coding.

3. Capacity boundary

3.1. The outside world and the unreachable

Here, the external is used to push out an unreachable boundary (unreachable), aiming to prove that some transmission rates are unreachable, meaning that no matter which encoding is used, it cannot simultaneously satisfy both normal user access and security constraints [7]. Therefore, today's external elements include (secure polymatroidal) secure multi-simulated array (external layer), (secure MAIS) maximum acyclic induced sub-graph bound, and (secure acyclic chain) acyclic boundary (acyclic chain). Multi-faceted pseudo-arrays are suitable for secure external environments in all scenarios, mainly due to their comprehensive decoding and confidentiality limitations. The maximum acyclic induced subgraph boundary and the acyclic chain boundary can only be used in graphical rule scenarios, mainly because they rely on the structure diagram formed by user cache relationships for calculation. However, there are also opposing conclusions. For example, a few Converse derives the inner bound of safe capacity, so experimenters should consider the opposite conclusion separately when analyzing to avoid confusion.

3.2. Inner realm and accessibility

Unlike external and unreachable elements, intrinsic and reachability mainly demonstrate through specific encoding that they can simultaneously meet users' needs for normal access and security constraints [7]. But here lies the cost of the key: accessibility requires a key approaching zero speed to ensure the legitimacy of the same message as the eavesdropper's request, and the receiver knows at least two pieces of information to achieve information isolation. Secure composite coding is the most typical accessibility solution. In it, the cost of constructing keys is a shared key to compensate for the inability to meet security requirements by meeting at least two pieces of information, but the trade-off is that it consumes storage resources.

4. Construction scheme for secure index encoding

4.1. Linear coding

Linear coding mainly involves a fixed matrix that combines numbers by adding or subtracting to produce corresponding data packets. However, linear coding is not just about writing a matrix, it also depends on whether the matrix is feasible or not. This depends on the transmission speed and whether it prevents eavesdropping, with parameters determining the minimum rank and minimum distance. The smaller the minimum rank value, the fewer broadcasts it receives, and the higher the transmission efficiency. The minimum distance determines how capable the linear code can withstand listeners--the larger the code, the safer it is. Besides these two key values, MDS codes are a special subclass of linear encoding, which can increase the minimum distance to the theoretical maximum, effectively improving overall safety performance. However, linear coding still faces the

DOI: 10.54254/2755-2721/2027.MELB38424

problem of NP difficulty. In one-to-one scenarios, linear coding is fully feasible, but when users and data are stacked, it is impossible to quickly obtain approximate data [8]. Therefore, although linear coding is simple, it still has loopholes that can lead to other coding structures.

4.2. Random linear coding

Random linear coding adds other codes or symbols to linear coding to achieve the highest level of security, preventing eavesdroppers from tampering with data. Its main principle is to divide all the code into many groups, with the raw data determined only by which group it belongs to. Then, a random number is used to select a random item from the group and send it out . Even if an eavesdropper acquires the information, they cannot extract the complete information, thus achieving strong security. The most critical aspect of strong security is the optimal code length, which is an additional security protection compared to ordinary encoding. The optimal encoding also includes the lowest rank. Besides minimum rank, coding design also needs to consider the number of messages that adversaries can eavesdrop on or tamper with. The more packets a listener can receive, the more security symbols need to be added, which directly lengthens the overall transmission length and increases resource overhead. Thus, the total encoding length = length required for normal transmission + increased length for anti-eavesdropping + added length for counter-serial modification. This formula provides the condition that the total number of codes must not be less than the sum of these values [6].

Stochastic linear coding and secure composite coding are both random, but they are completely different. Secure composite coding is only theoretically fundamental and requires keys to implement, whereas stochastic linear coding can be easily implemented using existing mathematical models. The main differences are shown in Table 1.

Table 1. The difference between secure composite coding and random linear coding

	Secure composite coding	Random linear coding
Theory	Nonlinear, matrix-free	Linear, adding based on the matrix framework
Core Role	Proof of existence, broadcast rate reachable	Enables protection against eavesdropping and code tampering
Safety overhead	There are shared keys close to zero, but they do not consume instant symbols	No keys exist, but additional random symbols are consumed
Meaning	It only proves the existence of the code, not enforcement	Clear thinking, precise implementation

5. Conclusion

This study first introduces the basic model and background of index coding, introduces adversary models by addressing the risk of eavesdropping on Sing Island, and classifies data stolen by listeners into strong and weak eavesdropping. This paper introduces the capacity boundary corresponding to eavesdropping scenarios and clarifies under what conditions encoding can achieve secure and reliable transmission. Through analysis, this paper finds that secure index encoding cannot guarantee security in everyday broadcast transmission, highlighting the importance of studying security boundary characteristics and encoding construction schemes. By reviewing the constraint capability

DOI: 10.54254/2755-2721/2027.MELB38424

and construction schemes of the security boundary, this paper clearly clarifies the limitations of existing schemes, providing a reference for deeper research and possessing theoretical value. This paper does not delve into the construction schemes of index coding, only presenting two relatively simple construction methods. Subsequent research on secure index coding can focus on complex adversaries, non-ideal channels, computational complexity, key overhead optimization, and multi-hop scenario expansion.

References

[1] Lee, N., Dimakis, A. G., & Heath, R. W. (2015). Index coding with coded side-information. *IEEE Communication Letters*, 319-322.
[2] Ong, L., Vellambi, B. N., Yeoh, P. L., et al. (2016). Secure index coding: Existence and construction. *IEEE International Symposium on Information Theory (ISIT)*, 2157-8117.
[3] Birk, Y., & Kol, T. (1998). Informed-source coding-on-demand (ISCOD) over broadcast channels. *IEEE INFOCOM'98, the Conference on Computer Communication*, 1257-1264.
[4] Arbabjolfaei, F., & Kim, Y.-H. (2018). Fundamentals of index coding. *Foundations and Trends in Communications and Information Theory, 14*(3-4), 163-346.
[5] Mojahedian, M. M., Gohari, A., & Aref, M. R. (2015). Perfectly secure index coding. *2015 IEEE International Symposium on Information Theory (ISIT)*, 2157–8095.
[6] Dau, S. H., Skachek, V., & Chee, Y. M. (2012). On the security of index coding with side information. *IEEE Transactions on Information Theory*, 3975–3988.
[7] Liu, Y., & Sadeghi, P. (2021). On converse results for secure index coding. *2021 IEEE Information Theory Workshop (ITW)*, 5245–5260.
[8] Dau, S. H., Skachek, V., & Chee, Y. M. (2011). On secure index coding with side information. *2011 IEEE International Symposium on Information Theory*, 2315–2319.