BGKey: Group Key Generation for Backscatter Communications among Multiple Devices

Jiajun Li, *Student Member, IEEE,* Pu Wang, *Graduate Student Member, IEEE,* Zheng Yan, *Senior Member, IEEE,* Yishan Yang, *Student Member, IEEE,* Kai Zeng, *Member, IEEE,*

***Abstract*—Backscatter communication (BC) is an emerging radio technology for achieving sustainable wireless communications. However, the literature still lacks an effective secret group key generation scheme for safeguarding communications among multiple resource-constrained backscatter devices (BDs). In this paper, we propose a novel physical layer group key generation framework, BGKey, for securing backscatter communications among multiple BDs. BGKey contains three schemes: Centralized Group Key Generation (CGKG), Decentralized Group Key Generation (DGKG), and Decentralized Hierarchical Group Key Generation (DHGKG). Each scheme has its own advantages, applicable in different scenarios. We analyze the performance of BGKey schemes regarding computation and communication complexity and security under eavesdropping and three active attacks. We conduct extensive simulations with different system parameters to evaluate their performance. CGKG is the most efficient and accurate for generating a group key, but it depends on a trusted radio frequency source (RFS) and is the least secure under eavesdropping and three active attacks among three schemes. DGKG exhibits better security and higher key generation rate (KGR) against eavesdropping and three active attacks compared with CGKG. However, the bit disagreement ratio (BDR) of group key increases when the size of BD group increases. DHGKG dramatically enhances the performance of group key generation compared with DGKG and retains its excellent security against eavesdropping and three active attacks.**

***Index Terms*—Backscatter Communication, Group Key Generation, Physical Layer Security, Passive Eavesdropping**

I. Introduction

BACKSCATTER communication (BC), remarkable for its energy harvesting and low energy consumption, is an emerging radio technology that guarantees proper and sustainable operation of Internet of Things (IoT) networks [1]. Backscatter devices (BDs) utilize ultra-low-power communications by backscattering radio frequency (RF) signals from a radio frequency source (RFS) and simultaneously harvest energy to power their circuits, eliminating the need for battery [2]–[5]. The emergence of BDs addresses the limitations of battery-powered traditional IoT devices that require frequent maintenance and battery replacement, thus has received widespread attentions. As a result, BC facilitates the widespread deployment of BDs in various locations, including body implantation, to support a wide range of applications such as environmental monitoring, healthcare, and smart homes [6]–[8].

Compared to traditional wireless devices with power-hungry RF functionalities, BDs operate without active RF components. Instead, they rely on an RFS to transmit RF signals, which are then backscattered and modulated by BDs to transmit data among themselves [9], [10]. However, due to the openness and broadcast nature of backscatter, a BC system may suffer from various security threats and vulnerabilities, i.e., eavesdropping, BD identity impersonation and wireless spoofing attacks [6], [11]–[13]. These security risks can lead to severe data interception and privacy breaches. Therefore, it is crucial to develop a practical security scheme to safeguard backscatter communications among BDs.

Secret key generation plays a crucial role in safeguarding private communications among different BDs. The utilization of a secret key for encryption ensures the confidentiality and security of BD communications within the BC system. However, most current studies primarily concentrate on key generation between two legitimate BDs in BC systems, employing lightweight cryptography [14], [15] and physical layer (PHY) mechanisms [6]. Nevertheless, there are numerous scenarios in which the generation of a shared secret key becomes necessary for the secure exchange of confidential information among multiple legitimate BDs. For instance, in a body sensor network where human health information needs to be exchanged among a group of implanted devices, it is essential to protect the transmitted information within the group using a secret group key to ensure the confidentiality of data in an e-health on-body network. Similarly, in the context of crowd-sourcing, when the content of a data collection task needs to be shared with a number of BDs, maintaining confidentiality becomes imperative. Moreover, for efficiently routing in a BC system containing multiple BDs, the confidentiality of information transmitted among multiple hops should be ensured with a secret group key [16]. However, the application of existing key generation scheme for two BDs to multiple BDs is exhausted on both computation and time since it needs to operate $N(N-1)/2$ times for N BDs. Hence, an efficient group key generation scheme in the BC system becomes a practical necessity.

There are two main approaches of group key generation in BC systems. One is traditional lightweight group key generation mechanisms based on cryptography, the other is PHY key generation mechanisms. In terms of the cryptographic approach, a family of Group Diffie–Hellman (GDH) protocols have been proposed [17], [18], which are straightforward extensions of the two-party Diffie-Hellman exchange (D-H) protocol [19]. Kim et al. improved GDH and proposed an efficient tree-based GDH (TGDH) scheme [20], [21]. However, both GDH and TGDH require each group member to perform D-H exchange with every other group member, resulting in significant computational overhead for the devices. This poses a challenge for BDs as they have limited energy supply and low computational capabilities [22]. Alternatively, for resource-constrained BC systems, PHY key generation is regarded as a promising technology for generating a group key that can achieve information-theoretic security. Owing to the channel reciprocity inherent in time-division duplex (TDD) systems, one can readily attain the inherent reciprocal insights through the estimation of wireless fading channels between legitimate users. This highlights the benefits of harnessing source model-based key generation techniques to bolster secure communication services [23]–[25]. The intrinsic randomness within wireless channels, arising from their reciprocity, can be harnessed to create a shared key with minimal computational overhead.

Nonetheless, existing PHY group key generation schemes, despite their wide application, prove unsuitable for BC systems, even though various schemes have been proposed based on diverse channel characteristics. Typically, these existing approaches necessitate the exchange of pilot signals between

J.J. Li, P. Wang, Z. Yan (corresponding author), and Y.S. Yang are with the State Key Lab of ISN, School of Cyber Engineering, Xidian University, Xi'an, Shaanxi, 710026 China. (email: jiajunli1204@stu.xidian.edu.cn, wangpu@stu.xidian.edu.cn, zyan@xidian.edu.cn, ysyangxd@stu.xidian.edu.cn)

K. Zeng is with the Department of Electrical and Computer Engineering, Cyber Security Engineering, and the Department of Computer Science, George Mason University, Fairfax, VA, 22030 USA (email: kzeng2@gmu.edu)

devices to gauge correlated channel attributes such as received signal strength (RSS) [26], [27] and channel state information (CSI) [28]–[33] at both communication endpoints to generate keys. However, BDs lack the capability to autonomously generate channel probing signals, conduct channel estimations for CSI, or measure RSS. Consequently, the practical application of these existing schemes in BC systems for safeguarding group BD communications remains infeasible.

Although the existing literature has studied pairwise key agreement between two BDs, the generation of a group key among multiple BDs has yet to be studied. Wang et al. [6] proposed a lightweight PHY pairwise key generation scheme to secure a two-BD system by utilizing the inherent reciprocal randomness of a triangle channel formed between two BDs and an RFS. However, authors did not address the issue of generating a shared session key for a group of BDs. Generating a key among multiple BDs is exceptionally challenging due to the random channels associated with each BD. Simply extending the pairwise PHY key generation scheme [6] to multiple devices raises concerns about security and performance, as broadcasting or forwarding channel information among BDs may expose vulnerabilities to eavesdropping and other active attacks. As such, there is a distinct lack of an effective PHY group key generation scheme in the current literature to ensure secure communications among multiple BDs.

In this paper, we propose a novel PHY group key generation framework, called BGKey, for securing backscatter communications among multiple BDs. BGKey consists of three distinct schemes: Centralized Group Key Generation (CGKG), Decentralized Group Key Generation (DGKG), and Decentralized Hierarchical Group Key Generation (DHGKG). Each scheme offers unique benefits and performance, making them suitable for different scenarios. CGKG is based on round-trip channel measurements, which involve calculating the product of the downlink and uplink channels between the RFS and BDs. In this scheme, a trusted RFS plays a crucial role by participating in key generation by broadcasting the differences of round-trip channel measurements from a reference BD's. This allows each BD to obtain all the necessary round-trip channel measurements for group key generation. Conversely, DGKG enables the establishment of a group key without requiring the involvement of the RFS, ensuring that RFS remains oblivious to the actual key generated. In DGKG, each pair of BDs operates in a backscattering/listening mode with time division, allowing them to obtain the relevant triangle channel measurements. These measurements are obtained by multiplying the downlink channels of the two BDs with their respective inward channels. Subsequently, a set of combinations of these triangle channel measurements is broadcasted within the group, enabling each BD to acquire all the necessary measurements for group key generation. To address the issue of high group key bit disagreement ratio (BDR) in DGKG when the number of BDs is large, DHGKG is introduced. This scheme organizes the BDs into multiple hierarchical sub-groups and performs DGKG within each sub-group to generate sub-group keys. The sub-groups then exchange their keys through secure channels established between the border BDs and a central group. Eventually, all the sub-group keys are combined to create a final group key, which is securely distributed to the remaining sub-groups. This paper's main contributions can be summarized as follows:

- We propose BGKey, a novel PHY group key generation framework for BC, which is the first work to generate a group key for multiple BDs in a BC system.
- We propose an efficient and accurate group key generation scheme CGKG, which utilizes the round-trip channels and depends on a trusted RFS for group key generation.
- We propose a secure group key generation scheme DGKG to establish a group key without letting RFS know group key information in case that the RFS cannot be fully trusted.
- We further propose a scalable and adaptable group key generation scheme DHGKG to improve the group key BDR of DGKG when the number of BDs is big. How to divide a large group of BDs into multiple sub-groups is also discussed.
- We analyze the security of BGKey schemes under eavesdropping and other three active attacks, and evaluate and compare their performance through theoretical analysis and simulations. The results show their advantages and effectiveness.

II. Background and Related Works

This section introduces BDs, specifies its difference from reconfigurable intelligent surfaces (RISs), and reviews related works.

A. BDs and RISs

BDs, ultra-low power consumption devices, utilize backscattering of ambient RF signals broadcasted by RFS to facilitate communication without the need for active RF transmission. Unlike conventional radio-based devices, BDs lack RF modules and are unable to transmit RF signals actively. Moreover, by incorporating simultaneous wireless information and power transfer (SWIPT) technology, BDs can operate efficiently with limited-capacity batteries or even without batteries. However, the computational capabilities of BDs are constrained by the energy harvesting rate, which limits battery consumption in their computation modules and renders them less powerful than traditional radio-based devices.

RISs [34]–[36], similar to BDs operating in the backscatter mode, possess the ability to reflect and phase-shift RF signals. The potential of RISs in facilitating PHY key generation has attracted significant attention. However, the roles of BDs and RIS in the key generation process differ. BDs actively participate in key generation by exchanging information using ambient RF signals, and the resulting keys are stored in the BDs. On the other hand, RIS-assisted networks typically involve RIS assuming a passive or intermediate role in communication or key generation. RISs are commonly utilized to ensure communication quality or enhance shared randomness in key generation among authorized devices [36]. Specifically, BDs initiate key generation or communication processes while RISs actively support and assist without having access to the communication information or generated keys among devices. Unlike BDs, RISs do not possess communication information or the generated keys among legitimate devices.

B. Existing Related Works

This subsection offers a succinct overview of existing PHY group key generation schemes. A range of secret group key generation schemes have been proposed, capitalizing on distinct channel information properties and accommodating various multi-device system topologies [26], [28]. One straightforward approach to group key generation involves each device performing channel estimation to measure CSI and subsequently disseminating combinations of its CSI measurements to other devices [28]–[33]. In [27] and [26], secret group key generation schemes were devised for star and chain topologies, respectively, leveraging the received signal strength (RSS) of probing signals sent by multiple devices to collaboratively create group keys.

Tang et al. [37] introduced an efficient multiple-input–multiple-output (MIMO) scheme, utilizing the indices of non-activated antennas among all devices within the group as shared randomness. Notably, this scheme eliminates the need for devices to engage in channel estimation and exhibits potential applicability in group key generation for BC systems. However, it is important to note that a direct analysis or validation of its compatibility with BC systems is currently lacking. **In a BC system, BDs encounter inherent limitations, as they are unable to transmit probing (pilot) signals to acquire channel properties, and their hardware or**

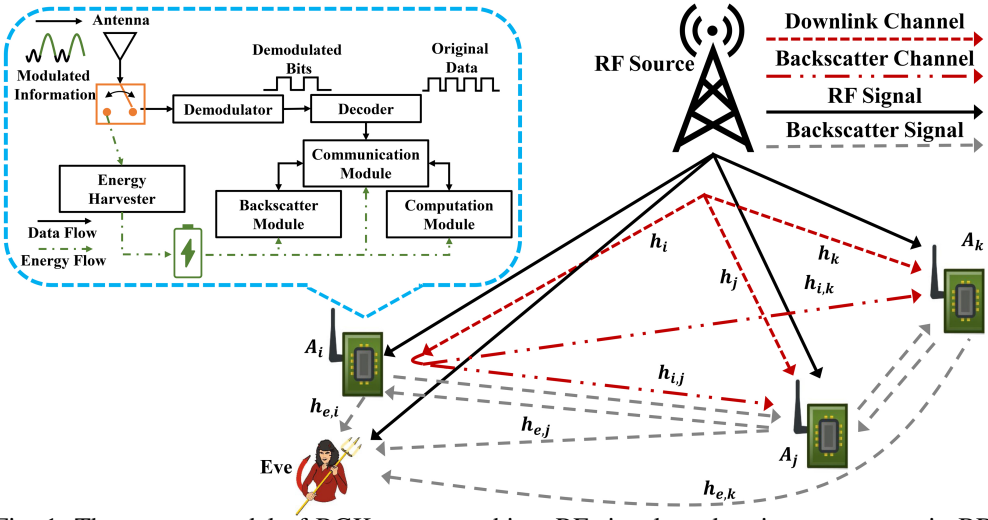


Fig. 1: The system model of BGKey over ambient RF signals and main components in BDs.

circuits do not facilitate RSS measurement. Consequently, the existing schemes, as they stand, prove unsuitable for BDs in BC contexts.

Wang et al. [6] proposed an innovative triangle-channel based scheme for establishing a pairwise key between two BDs, eliminating the necessity for transmitting probing signals or performing channel estimation. Notably, this pioneering scheme represents the initial and solitary effort in the realm of physical layer key generation for BDs. It is the first and only work we can find in the literature. Nonetheless, it is worth noting that the authors did not delve into the generation of group keys within the broader context of BC systems. An approach that presents itself as straightforward yet inefficient for generating a group key within a fully meshed topology involving N BDs is to repeatedly apply the triangle-channel based pairwise key generation method from [6], a process we refer to as Multi-D2D. In the Multi-D2D approach, each BD establishes N-1 pairwise keys with other BDs within the group. Subsequently, a secure channel is employed to exchange the remaining $(N-1)(N-2)/2$ pairwise keys among the group members. The group key is then synthesized by combining all $N(N-1)/2$ pairwise keys. However, it is worth noting that Multi-D2D exhibits a relatively low efficiency in the group key generation process, as it necessitates each device to establish a pairwise key with every other device and exchange these keys within the group to derive the final group key. More precisely, in a mesh topology, each BD is required to exchange pairwise key information with the remaining $N-1$ BDs in order to achieve the group key, which incurs a significant overhead. Based on our knowledge, the literature still lacks an effective framework for group key generation to secure communications among multiple BDs in a BC system.

III. Problem Statement

A. System Model

Fig. 1 illustrates the system model of BGKey. The system comprises two types of entities: RFS (e.g., WIFI Access Point) and BD. The RFS serves two main purposes: transmitting RF signals to its dedicated BDs and processing information related to key generation. All BDs within the coverage of the RFS alternately perform backscatter transmission in a time-division manner [2], [38]. However, when the RFS operates in full-duplex (FD) mode, its receiver also receives signals from its transmitter, resulting in self-interference (SI) [39]. To overcome this problem, the RFS needs to conduct an active suppression and use its estimated SI channel to generate a SI cancellation signal, and deducts it from a receive signal to obtain an SI free signal [40].

We make the assumption that only one RFS performs as intended, while multiple passive BDs can simultaneously receive and backscatter the RFS signals using a single antenna. Fig. 1 illustrates the main components of a BD, which include a modulator (load impedance), an information receiver, a communication module, a backscatter module, a computation module, an energy harvester, a battery, and other modules (such as sensors) [41], [42]. It should be noted that the demodulator and decoder primarily consist of passive RF chains and do not require external power to operate [34], [43]. The BD also contains several active modules that require batteries for power supply, such as the backscatter module, communication module, and computation module. The information receiver consists of three main parts: an antenna, a demodulator, and a decoder. The demodulator first converts the modulated received signals to demodulated bits, and the decoder can decode demodulated bits to derive original data. BDs can work in an energy harvesting mode or an information handling mode by switching the modulator. When BDs operate in the information handling mode, they can further operate in a listening mode and a backscatter mode. In the listening mode, BDs can demodulate and decode the information from received signals to obtain original data. While in the backscatter mode, BDs can intentionally alter the phase or amplitude of the received signals and reflect them to other devices using their backscatter module. In both modes, BDs can receive information.

Regarding the operational environment of the BC system, it is commonly deployed in densely populated spaces such as warehouses, where BDs are affixed to various objects. Due to deep shadowing, there is no direct link between a signal source and its destination. In such situations, the main channels (downlink or inbound channels) are frequently obstructed by obstacles. Hence, we represent the intricate relationship between the RFS and BDs as independent Rayleigh fading channels [6], [44]. Let h_i denote the downlink channel between the RFS and BD A_i, or Eve e, $h_{i,j}$ denotes the inward channel between A_i and A_j $(i,j \in \{1,2,...,N,e\}$, $i \neq j)$, and $h_i h_{i,j}$ denotes the cascade backscatter channel that concatenates the h_i and $h_{i,j}$. Herein, N is the total number of BDs. Table I summarizes the notations used throughout the paper.

In our scheme, we have opted for the TDD mode due to the robust channel reciprocity observed between uplink and downlink channels within TDD systems. While it is certainly feasible to implement the Frequency Division Duplex (FDD) mode in a BC system, it poses challenges for BDs in extracting reciprocal channel information. In most existing PHY key generation schemes operating under FDD, keys are generated by extracting frequency-independent reciprocal channel parameters or constructing reciprocal channel gains. This, however, can introduce substantial computational overhead and raise security concerns. Zhang et al. [45] introduced a

key generation method based on a channel-to-channel mapping function, enabling devices to extract genuine reciprocal channel features rather than relying on frequency-independent reciprocal channel parameters or constructing reciprocal channel gains. Nevertheless, as pointed out by Zhang and colleagues, acquiring the channel mapping function poses challenges, as truly learning it demands a high computational cost. Consequently, it remains impractical for resource-constrained BD devices to acquire the feature mapping for key generation purposes. We foresee the emergence of more efficient methods for extracting authentic reciprocal channel features that can be readily implemented on resource-limited BDs operating within FDD systems. This will be a key focus of our future research endeavors.

Our BGKey framework consists of three distinct schemes: CGKG, DGKG, and DHGKG. These schemes leverage the dynamic nature of wireless channels, characterized by rapid variation or fast fading, which provides a high degree of randomness that can be utilized for key generation. However, it is important to note that in static environments with slow-changing channels, the key generation process may considerably slow down. Building upon the insights from our prior research [46], we expound upon the utilization of time-variant backscatter coefficients to address the challenge of limited randomness within static environments in ***Appendix-D***. To introduce additional randomness into the static channel, we employ a combination of time-variant amplitude backscatter coefficients and a time-variant phase shifting coefficient matrix. Subsequently, in the subsequent sections, we place a heightened emphasis on the generation of keys by harnessing the dynamic nature of wireless channels characterized by rapid variations and fast fading.

B. Security Model

1) Eavesdropping Attack: Our adversary model considers a passive eavesdropper, named Eve, who attempts to intercept information about the generated keys without engaging in active attacks. Eve can overhear the RF signals broadcasted from the RFS, the backscattered signals from BDs, or the unencrypted information from any entities. Nevertheless, we make the assumption that Eve cannot be too close (less than a few wavelengths) to any BDs so as to avoid easy detection. This assumption ensures that Eve measures uncorrelated channels with BDs [23]. Note that the RFS may have an interest in knowing the group key and may disclose the key to a malicious party, leading to different design for group key generation.

2) Active Attacks: Three active attacks could be launched: channel control attack (CCA) [47], [48], signal manipulative attack (SMA) [49], and untrusted RFS attack (URSA) [46]. In the CCA, also commonly referred to as a man-in-the-middle attack, an active attacker (say Mallory) can manipulate the channel between a sender and receiver by strategically moving intermediate objects to manipulate the channel's gain. Concretely, the attacker can insert reciprocal information to the main channels. When launching an SMA, Mallory injects similar signals into legitimate devices and manipulates the agreement on valid key bits. The URSA exploits the non-authentication characteristic of the BDs during key generation. Mallory can impersonate an RFS and send a signal to a BD to manipulate the outcome of the key generation process.

C. Models in BGKey Framework

1) Signal Model: The signal model is constructed based on our previous work [6]. Denote a bandpass signal transmitted from the RFS during a symbol interval as

$$\tilde{s}(t) = \Re\{\sqrt{p}s(t)e^{j2\pi f_c t}\}, \tag{1}$$

where $s(t)$ is a unit baseband signal with transmission power p, and f_c represents carrier frequency. $\Re\{\cdot\}$ represents the operation of getting the real part of a complex number. The downlink signal received at A_i can be represented as [50]

$$\tilde{d}_i(t) = \Re\{[\sqrt{p}h_i(t)s(t)]e^{j2\pi f_c t}\} + \tilde{\omega}_i(t), \tag{2}$$

where $d_i(t) = \sqrt{p}h_i(t)s(t)$ is the baseband representation of $\tilde{d}_i(t)$ and $\tilde{\omega}_i(t)$ is the received passband additive white Gaussian noise (AWGN) at A_i. Let $b_i(t)$ be the signal of A_i to be transmitted, the signal backscattered from A_i can be expressed as $\alpha_i\tilde{d}_i(t)b_i(t)$, where α_i denotes the backscatter coefficient at BD A_i. This modulation technique exempts BDs from generating RF signals locally and thus significantly reduces their power consumption [10]. The received signal at A_j, superposed by the downlink signal from RFS and the backscattered signal from A_i, can be expressed as,

$$\tilde{y}_j(t) = \alpha h_{i,j}(t)\tilde{d}_i(t)b_i(t) + h_j(t)\tilde{s}(t) + \tilde{\omega}_j(t), \tag{3}$$

and the baseband representation of (3) can be expressed as

$$y_j(t) = y_j^b(t) + y_j^d(t) + \omega_j(t), \tag{4}$$

where $y_j^b(t) = \alpha h_{i,j}(t)d_i(t)b_i(t)$ is the backscatter signal reflected from A_i, $y_j^d(t) = \sqrt{p}h_j(t)s(t)$ is the original ambient signal directly from the RFS, and $\omega_j(t)$ is the baseband representation of $\tilde{\omega}_j(t)$ with power σ_j^2, i.e., $\omega_j(t)\sim\mathcal{CN}(0,\sigma_j^2)$ and $\mathcal{CN}(\mu,\sigma^2)$ means the circularly symmetric complex Gaussian distribution (CSCG) with mean μ and variance σ^2.

Let $h_i[n]$, $h_j[n]$, and $h_{i,j}[n]$ denote the discrete-time representation of $h_i(t)$, $h_j(t)$, and $h_{i,j}(t)$, respectively. For convenience, the discrete-time representation of the received signal at A_j can be expressed as

$$y_j[n] = y_j^b[n] + y_j^d[n] + w_j[n], \tag{5}$$

where $y_j^b[n] = \alpha h_{i,j}[n]d_i[n]b_i[n]$, $y_j^d[n] = \sqrt{p}h_j[n]s[n]$, and $w_j[n]\sim\mathcal{CN}(0,\sigma_j^2)$.

2) Delay estimation and cancellation: It is noteworthy that y_j^b and y_j^d do not arrives at A_j at the same time. Since y_j^b passes through the cascade-channel $h_ih_{i,j}$ and y_j^d passes through the downlink channel h_j before arrives at A_j, the distance travelled by the signals between y_j^b and y_j^d may be different. Therefore, $y_j(t)$ can be expressed as:

$$y_j(t) = y_j^b(t+\tau) + y_j^d(t) + \omega_j(t), \tag{6}$$

where τ represents the delay. This situation arises due to the varying arrival delay of the signal across different links. The problem of disparate delays in direct-link (through downlink channel) and backscatter-link (through cascade channel) propagation delays is widely recognized in ambient backscatter

TABLE I: Notations

Notation	Definition
A_i	The i-th BD
N	The number of BDs
α_i	The backscatter coefficient at BD A_i
h_i	The downlink channel between RFS and A_i
$h_{i,j}$	The inward channel between A_i and A_j
β_i	The difference between the measurement of A_i and a reference BD (say A_1) regarding round-trip channel information.
$\mathcal{D}$	The round-trip channel difference set broadcasted by RFS
R_i	The round-trip channel measurement set of A_i
$T_{i,j}$	The triangle channel measurement constructed between A_i and A_j
$\mathcal{Q}_i^1$	The set of $N-1$ triangle channel measurements constructed by A_i in Phase 1 (Backscattering/Listening Phase)
$\mathcal{Q}_i^2$	The set of $(N-1)(N-2)/2$ triangle channel measurements constructed by A_i in Phase 2 (Broadcasting Phase)
$\mathcal{Q}_i$	The set of $N(N-1)/2$ triangle channel measurements constructed by A_i in the Construction Phase
$\mathcal{Q}_e^1$	The set of N triangle channel measurements eavesdropped by Eve in Phase 1 (Backscattering/Listening Phase)
x_j	The triangle channel measurements combinations broadcasted by A_j
g_j	The randomly generated weight matrix that multiplied by $\mathcal{Q}_i^1$ to obtain x_j
r_j	The number of combinations broadcasted by r_j

communication systems. Hence, before sampling the modulated signal and extracting the original data from it, BDs must estimate these two delays in order to obtain accurate information from the received signals. One approach to achieve this is by employing a cross-correlation based method at the BDs to precisely estimate the direct-link propagation delay and the backscatter-link propagation delay [51], [52]. This phase, which utilizes the cross-correlation based method mentioned above, is referred to as the **delay estimation and cancellation phase**. If there is a multi-link path to the terminals (i.e., BDs and RFS), the delay estimation and cancellation phase needs to be conducted immediately after devices receive signals. To simplify the subsequent equations and enhance readability, we use only t to represent the arrival time of y_j^b and y_j^d, rather than using $t+\tau$ and t, as this eliminates the impact of different propagation delays on the received information [6], [52].

3) Time Synchronization in TDD OFDM system: In a TDD system, BDs or RFS need to operate independently during each time slot. Specifically, equation (4) mandates that A_i operates in the backscattering mode while A_j operates in the listening mode. If synchronization is not considered, it could lead to the situation where both BDs are either in the listening phase or the backscattering phase simultaneously within the same time slot.

OFDM symbols commonly utilize preambles for time synchronization. In WLAN systems following the 802.11a standard, a synchronization preamble is created utilizing fixed frequency-domain OFDM symbols [52], which comprises multiple identical training symbols. In order to attain synchronization, both BDs and RFS need to be aware of the system's synchronization preamble.If BDs and RFS are unaware of the synchronization preamble, they can utilize the previously mentioned propagation delay estimation method to adjust their local time and achieve synchronization [52].

IV. BGKey Design

This section describes the design of CGKG, DGKG and DHGKG in BGKey.

A. CGKG Scheme

CGKG is designed based on round-trip channel measurements between RFS and BDs. It relies on the RFS to broadcast the difference in round-trip channel measurements from a reference BD, enabling each BD to acquire all the necessary round-trip channel measurements for group key generation. With its minimal computational demands on BDs and the high accuracy of the generated group key, CGKG can serve as a performance benchmark for group key generation in BC systems. In this subsection, we first demonstrate the key generation scheme between the RFS and a single BD. We then expand this scheme to facilitate group key generation among a set of N BDs. Finally, we perform security analysis on CGKG with our assumed eavesdropping model.

1) Key Generation between BD and RFS: In CGKG, RFS receives backscattered signals from BDs while continuously sending signal $s(t)$. This requests that the RFS is capable of estimating SI channels since SI signals drastically disrupt received signals of the RFS. The key generation process between a BD and the RFS involves two distinct phases.

Phase 1: In time slot t, the RFS transmits constant signal $s(t)$ and performs SI channel estimation. A_i works in the listening mode. The RFS can use an active suppression technology like [39] to obtain the SI channel estimation, denoted as H_{est}. Then, the RFS can generate a cancellation signal $s(t)H_{est}$ using H_{est} and $s(t)$.

Phase 2: In time slot $t^{'}$, the RFS keeps transmitting signal $s(t^{'})$ and BD works in the backscattering mode. By deducting the received signal with the cancellation signal $s(t)H_{est}$, the RFS can obtain SI-free signals. The received signals at A_i and the RFS are:

$$y_i(t^{'}) = s(t^{'})h_i(t^{'}) + \omega_i, \tag{7a}$$

$$\begin{aligned} y^i_{RFS}(t^{'}) &= \alpha_i s(t^{'})h_i(t^{'})h_i^b(t^{'}) + s(t^{'})(h_{SI}(t) - s(t)H_{est}) + \omega_{RFS} \\ &\approx \alpha s(t^{'})h_i(t^{'})h_i^b(t^{'}) + \omega_{RFS}, \end{aligned} \tag{7b}$$

where ω_i and ω_{RFS} are AWGN at A_i and RFS, $\alpha s(t^{'})h_i(t^{'})h_i^b(t^{'})$ is the signal sent from the RFS through downlink $h_i(t^{'})$ to BD A_i and finally backscattered by A_i through backscatter channel $h_i^b(t)$ to the RFS. $h_{SI}(t^{'})$ is the SI channel in time slot $t^{'}$ and $H_{est}(t)$ is the estimation of the SI channel in time slot t. Due to channel reciprocity, we have $h_i^b(t) \approx h_i(t)$ and we define $\alpha h_i(t)h_i^b(t)$ as an round-trip channel between RFS and BD. If the channel reciprocity holds, that is $t^{'} - t < T_{ch}$ and T_{ch} is the coherence time [53], we have $H_{est}(t) \approx h_{SI}(t^{'})$. Then, the most of the SI can be cancelled. Although there remains some residual SI (the remaining SI after cancellation) in the system that corrupts channel symmetric, most of the existing works show that SI can be suppressed below receiver noise or ambient co-channel interference [54]. We analysis the influence of residual SI on the key consistency in ***Appendix-E***. After the above two phases, if we assume there is no noise in the system and there is no residual SI, the received signals at A_i and RFS can be respectively expressed as follows:

$$y_i = h_i s, \tag{8a}$$

$$y^i_{RFS} = \alpha_i h_i h_i^b s, \tag{8b}$$

where y_i is the received downlink signal at A_i and y^i_{RFS} is the received signal from round-trip channel at RFS. Observing (8a) and (8b) we can find that y^i_{RFS} contains the multiplication of a downlink channel information h_i and a backscattered channel information h_i^b. Since the downlink signal received at A_i is immediately reflected back to the RFS, the time lag between RFS receives backscattered signal and BD receives downlink signal is within coherence time. Therefore, $h_i \approx h_i^b$ holds. To transform y_i to y^i_{RFS}, we can square y_i and use the known backscatter coefficient α_i to obtain $y_i^{'} = \alpha_i h_i h_i s^2$. Obviously, we can multiply the RF signal s to y^i_{RFS} to obtain $y^{'i}_{RFS} = \alpha_i h_i h_i^b s^2$. Then, we have $y_i^{'} \approx y^{'i}_{RFS}$. Therefore, the RFS and A_i can exploit the identical round-trip information as shared randomness to generate a secret key between them. To be specific, by using discrete-time representation of $y_i^{'}$ and $y^{'i}_{RFS}$ we have:

$$y_i^{'}[n] = \alpha_i h_i[n] h_i[n] s^2[n], \tag{9a}$$

$$y^{'i}_{RFS}[n] = \alpha_i h_i[n] h_i^b[n] s^2[n]. \tag{9b}$$

Clearly, from [6], RFS and A_i can average the constructed information of $J = N_{cp} - L + 1$ samples and **can obtain round-trip channel information without conducting channel estimation**, where N_{cp} is the CP length of OFDM symbols, L is the maximum channel spread in the BC system.

$$\begin{aligned} v_i &= \frac{1}{J}\sum_{n=L-1}^{N_{cp}-1} y_i^{'}[n] = \frac{1}{J}\sum_{n=L-1}^{N_{CP}-1} |\alpha h_i[n] h_i[n] s^2[n]| \\ &= |\alpha h_i h_i|, \end{aligned} \tag{10a}$$

$$\begin{aligned} v^i_{RFS} &= \frac{1}{J}\sum_{n=L-1}^{N_{cp}-1} y^{'i}_{RFS}[n] = \frac{1}{J}\sum_{n=L-1}^{N_{cp}-1} |\alpha h_i[n] h_i^b[n] s^2[n]| \\ &= |\alpha h_i h_i^b|. \end{aligned} \tag{10b}$$

Where $|\cdot|$ denotes the operation of taking an absolute value and $\frac{1}{J}\sum_{n=L-1}^{N_{cp}-1} |s^2[n]| = 1$ because the baseband OFDM signal $s[n]$ holds unit power. Because the channel reciprocity holds within T_{ch}, i.e., $h_i \approx h_i^b$, the identical round-trip channel information ($v_i \approx v^i_{RFS}$) can be used as a shared random source between A_i and RFS. However, in practice,

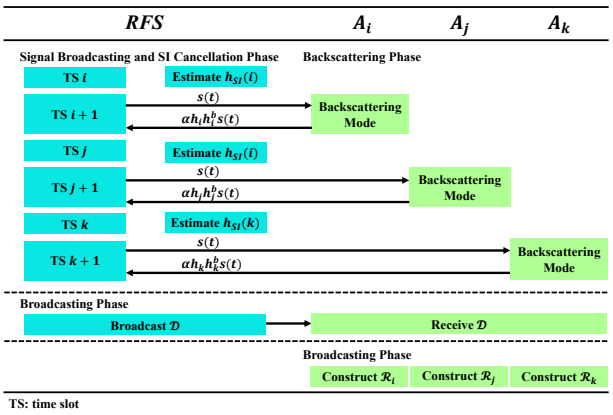


Fig. 2: The process of CGKG.

hardware impairments lead to the reduction in reciprocity of measurements between the same channels even within the coherence time. In order to represent the decreased reciprocity of channel measurements caused by hardware imperfections, a cross-correlation coefficient can be employed to model the correlation between channels [55]. Specifically, when constructing round-trip channel information, RFS uses $h_i \cdot cor$ instead of h_i if A_i uses h_i. It is important to note that '$\cdot$' here represents a more intricate relationship between h_i and the cross-correlation coefficient cor, rather than a simple multiplication [55]. Subsection V-C analyzes the actual impact of the cross-correlation coefficient on the performance of key generation.

2) CGKG Design: In CGKG, we exploit the key generation scheme between BD and RFS with SI cancellation to further generate a group key with four phases, shown in Fig. 2, where A_i, A_j, and A_k represent different BDs that communicate with RFS.

Signal Broadcasting and SI Cancellation Phase: The RFS continuously transmits RF signals to all BDs to power their operation and communication. According to Sethi et al. [56], a SI channel is a multi-path channel, and in [39], [40], it is further shown that the SI channel can be modelled as a slowly varying Rician fading channel. Therefore, the SI channel characteristic changes when the RFS receives backscattered signals from different BDs in time division manner. To obtain accurate SI-free signals, it is necessary to perform SI estimation and cancellation for each BD during this phase, despite the slow variation of the SI channel characteristics over time.

Backscattering Phase: Each BD takes its turn to operate in the backscattering mode to construct a round-trip channel with the RFS as mentioned in Subsection IV-A1. Each BD A_i ($i \in \{1, 2, ..., N\}$) can obtain the round-trip channel measurement$\alpha h_i h_i^b$ between the RFS and itself. Meanwhile, the RFS holds the measurements of all the N round-trip channels between all BDs and itself.

Broadcasting Phase: The RFS selects a BD as a reference BD (denoted as A_1). The RFS calculates the difference between the measurement of A_i and A_1 regarding the round-trip channel information:

$$\beta_i = v_{RFS}^i - v_{RFS}^1 = \alpha h_i h_i^b - \alpha h_1 h_1^b, \quad i \in \{2, 3, ..., N\}. \quad (11)$$

Then, the RFS broadcasts a round-trip channel difference set $\mathcal{D} = \{\beta_2, ..., \beta_N\}$ to all BDs.

Construction Phase: Each BD first constructs $\alpha h_1 h_1^b$ based on the difference between its round-trip channel measurement and that of A_1. Then, all BDs can calculate all round-trip channel measurement $\{\alpha h_1 h_1^b, ..., \alpha h_N h_N^b\}$ with the received $\mathcal{D} = \{\beta_2, ..., \beta_N\}$, respectively. For example, BD A_1 can calculate all v_{RFS}^i ($i \in \{2, 3, ..., N\}$) because it knows $v_{RFS}^1 = \alpha h_1 h_1^b$; BD A_2 can first calculate the v_{RFS}^1 from $\beta_2 = v_{RFS}^1$ with $v_{RFS}^2 = \alpha h_2 h_2^b$ and then calculate all other v_{RFS}^i ($i \in \{3, 4, ..., N\}$), and so on for other BDs.

One notable advantage of CGKG is its immunity to the scaling problem that arises when the time interval between successive measurements of the same channel exceeds the coherence duration. For the BD, its downlink channel information is measured as soon as the downlink signal arrives. Additionally, since the BD can process measurements while backscattering, the backscattered signal is backscattered almost immediately. Therefore, the RFS measures the round-way channel information simultaneously with the BD measurement of the downlink channel. However, in DGKG, when the number of BDs is large, the time interval for different BDs to measure the same triangle channel information may surpass the coherence duration. A detailed analysis of this scheme will be presented later.

Therefore, the RFS and all BDs can exploit their respective round-trip channel measurement set $R_i = \{\alpha h_1 h_1^b, ..., \alpha h_N h_N^b\}$ as the identical shared randomness to generate a secret group key with a sequence of measurements $\boldsymbol{\mathcal{R}}_i = \{R_i(1), ..., R_i(M)\}$ ($i \in \{1, 2, ..., N\}$), where M is the length of sequence and $R_i(j)$ ($j \in \{1, 2, ..., M\}$) are the measured values of the set of N round-trip channels in different independent measurements. With the shared randomness measurement, three more steps are processed to generate a secret key: quantization, information reconciliation and privacy amplification.

① Quantization

In the key generation process, quantization is used to convert analog measurements into binary bits. Quantization level and threshold are two parameters that influence a quantizer. The quantization level refers to the number of key bits quantified in each measurement. A high quantization level increases the key generation rate while deteriorating the BDR between keys. The threshold serves as the reference level that categorizes measurements into distinct groups. A distribution-based threshold method relies on estimated statistical values (such as mean value or standard deviation) of the shared randomness measurement [57]. In Section V-C we adopt a distribution-based threshold method in quantization and examine the impact of different quantization levels on the performance of group key generation.

② Information Reconciliation

Following quantization, there is still a possibility of key disagreements among the measurements obtained from legitimate BDs, despite the enhancement of correlation through preprocessing. o address this, various information reconciliation techniques can be employed for mismatch correction, such as low-density parity-check (LDPC) [24] and Golay code [58], etc. In our study, we use Cascade for information reconciliation since it leaks less information [59] and with lower complexity than LDPC [60].

③ Privacy Amplification

The purpose of privacy amplification is to minimize the potential information that an eavesdropper could access regarding the derived key [61]. Consequently, any exposed information is removed from the shared key sequence between the communicating parties. Reliance on universal hash functions enables the transformation of the reconciled bit stream into a highly randomized bit stream, thereby improving its security properties.

3) Eavesdropping Analysis: In our adversary model, Eve has the capability to intercept both the RF signal $s(t)$ and the round-trip channel difference set $\mathcal{D}$ that is broadcasted from the RFS. Furthermore, Eve can estimate its own downlink channel and construct the round-trip channel between the RFS and itself using the same procedure as described in equations (9) and (10):

$$v_e = \alpha h_e h_e^b = \alpha h_e h_e. \quad (12)$$

The reference BD, denoted as A_1, selected during the *Broadcasting Phase* of CGKG, is known to all parties in the system. Thus, in order to gather as much key information as possible, Eve needs to be in close proximity to the reference BD.Furthermore, a number of N round-trip values $R_e = \{\alpha h_e h_e^b, \alpha h_2 h_2^b - (\alpha h_1 h_1^b - \alpha h_e h_e^b), ..., \alpha h_N h_N^b -$

$(\alpha h_1 h_1^b - \alpha h_e h_e^b)\}$ calculated by Eve by solving the round-trip difference set $\mathcal{D}$ and its round-trip channel is uncorrelated to $R_i = \{\alpha h_1 h_1^b, ..., \alpha h_N h_N^b\}$, since $\alpha h_1 h_1^b - \alpha h_e h_e^b \neq 0$. Nevertheless, if Eve is very close to A_1, Eve can intercept some information of the round-trip channel of $\alpha h_1 h_1^b$, which can be modeled with a **cross correlation coefficient** between h_e and h_1 (denoted as cor_{h_1,h_e}) [55]. The actual effect of the cross correlation coefficient on information leakage is analyzed in Subsection V-C.

4) Achievable Rates of CGKG: In this subsection, we analyze the achievable secret key rate of CGKG under the presence of an eavesdropper. In the GGKG, the objective of RFS is to provide information for BDs to ensure that each BD obtains $R = \{\alpha h_1 h_1^b, ..., \alpha h_N h_N^b\}$, which is the common information among the group. And this case can be viewed as an instance of Slepian-Wolf coding [62]. Thus, the lossless coding rates can be defined as:

$$\begin{aligned} R_{Key}(M) &= \frac{1}{M} H(R(1), R(2), ..., R(M)) \\ &= \frac{1}{M} H(r_1(M), r_2(M), ..., r_N(M)), \end{aligned} \tag{13}$$

where $H(\cdot)$ represents the calculation of the entropy of a certain sequence, (i) represents the measurement of the i-th time of key generation, $R(i)$ represents the measurement of round-trip channel information set at the i-th time of key generation and $r_i(M) = \{\alpha h_i h_i^b(1), ..., \alpha h_i h_i^b(M)\}$. We denote R_{Pro} as the rate of the information that RFS needs to provide to the BDs in the group with the 'worst' observation for the generation of the group key [62]. And R_{pro} can be expressed as:

$$R_{Pro}(M) = \max_{1 \leq i \leq N} \frac{1}{M} I([r_1(M), ..., r_N(M)] | \boldsymbol{\beta}(M), r_i(M)), \tag{14}$$

where $\boldsymbol{\beta}(M) = [\beta_1(M), \beta_2(M), ..., \beta_N(M)]$. With the Slepian-Wolf source encoding with the above random binning structure, the key mismatch probability goes to 0 as $M \to \infty$ and the upper bound of key rate can be obtained. With the above analysis, we can now define the asymptotic group information rate [26], [62]:

$$\begin{aligned} R_{CGKG}(M) &= \lim_{M \to \infty} R_{Key}(M) - R_{Pro}(M) \\ &= \min_{1 \leq i \leq N} \lim_{M \to \infty} \frac{1}{M} I([r_1(M), r_2(M), ..., r_N(M)]; [\boldsymbol{\beta}(M), r_i(M)]). \end{aligned} \tag{15}$$

In the broadcasting phase of RFS, an eavesdropper can obtain the information $[\boldsymbol{\beta}(M), V_e(M)]$ as analyzed in the previous subsection, where $V_e(M) = \{v_e(1), ..., v_e(M)\}$. And we define the key rate between the common information of the BD group and the eavesdropped information by Eve as:

$$R_{Eve}(M) = \lim_{M \to \infty} \frac{1}{M} I([r_1(M), ..., r_N(M)]; [\boldsymbol{\beta}(M), V_e(M)]). \tag{16}$$

Now, we can write the achievable secret key rate (SKR) as:

$$R_{Sec}^{CGKG}(M) = R_{CGKG}(M) - R_{Eve}(M). \tag{17}$$

In what follows, we evaluate R_{Sec}^{CGKG} for the case that all channels are symmetric. And for simplicity, noise and the observation of reciprocal part are assumed to be independent and identically distributed. Since the observation of each BD is identically distributed, R_{Sec}^{CGKG} can be obtained by analyzing the SKR for an arbitrary BD (say A_i). The SKR of CGKG given in equation (17) can be written after dropping key length indices for simplicity as follows:

$$\begin{aligned} R_{Sec}^{CGKG} &= I([r_1, r_2, ..., r_N]; [\boldsymbol{\beta}, r_i]) - I([r_1, r_2, ..., r_N]; [\boldsymbol{\beta}, V_e]) \\ &= I([r_1, r_2, ..., r_N]; r_i | \boldsymbol{\beta}) + I([r_1, r_2, ..., r_N]; \boldsymbol{\beta}) \\ &\quad - I([r_1, r_2, ..., r_N]; V_e | \boldsymbol{\beta}) - I([r_1, r_2, ..., r_N]; \boldsymbol{\beta}) \\ &= I([r_1, r_2, ..., r_N]; r_i | \boldsymbol{\beta}) - I([r_1, r_2, ..., r_N]; V_e | \boldsymbol{\beta}) \end{aligned} \tag{18}$$

Therefore, the achievable SKR of CGKG is $I([r_1, r_2, ..., r_N]; r_i | \boldsymbol{\beta}) - I([r_1, r_2, ..., r_N]; V_e | \boldsymbol{\beta})$. Since it is difficult to measure how much information can be obtained in V_e using the correlation coefficient, we can focus on analyzing $I([r_1, r_2, ..., r_N]; r_i | \boldsymbol{\beta})$ to obtain the achievable key rate. We can use the chaining rule of mutual information to perform the following transformations to $I([r_1, r_2, ..., r_N]; r_i | \boldsymbol{\beta})$:

$$\begin{aligned} R_{Gen}^{CGKG} &= I([r_1, r_2, ..., r_N]; r_i | \boldsymbol{\beta}) \\ &= I(r_j; r_i | \boldsymbol{\beta}) + I([r_1, ..., r_{j-1}, r_{j+1}, ..., r_N]; r_i | r_j, \boldsymbol{\beta}) \\ &\geq I(r_j; r_i | \boldsymbol{\beta}). \end{aligned} \tag{19}$$

Thus, the achievable KGR of CGKG is $I(r_j; r_i | \boldsymbol{\beta})$. However, we can not derive the closed-form expression of achievable KGR by analyzing the distribution of variables in $I(r_j; r_i | \boldsymbol{\beta})$. This is because $r_i = \alpha h_i h_i^b$ is obtained by multiplying two Rayleigh-distributed variables and therefore r_i does not obey Gaussian distribution. Based on some assumptions (i.e., $h_i h_i^b$ obeys Gaussian distribution and correlated noise variables are independent from all channels, etc.), the closed-form expression can be obtained, but we do not elaborate it in detail herein. To analyze the SKR in CGKG, we provide numerical simulations to analyze SKR in different settings (i.e., different SNRs or different cross-correlation coefficients between the main channel and the wiretap channel) in Subsection V-C.

B. DGKG Scheme

DGKG is designed to address the scenario where the RFS cannot be fully trusted by utilizing pairwise key generation between two BDs. In DGKG, a set of triangle channel measurements obtained by a BD is broadcasted to enable each BD to obtain the complete triangle channel information required for group key generation. **Notably, an important advantage of DGKG is that the RFS remains unknown to the generated group key.** While DGKG shares a similar concept with other techniques that utilize device combinations for key generation, such as the approach described by Thai et al. [28], our scheme leverages energy-saving ambient backscattering rather than sending probing signals for forming device-to-device channel information. Additionally, we establish lower and upper bounds for the broadcasted combinations to minimize the risk of information leakage. In this subsection, we introduce a fundamental pairwise key generation scheme based on a triangle channel [6], which serves as the basis for the design of DGKG. Then, we describe the detail design of DGKG. Finally, we conduct a security analysis of DGKG using our assumed eavesdropping model.

1) Pairwise Key Generation between Two BDs: Two phases are needed to generate a pairwise key between two BDs.

Phase 1: In time slot t, RFS transmits constant signal $s(t)$ with unit power [52], [63] and A_i operates in the backscattering mode. A_i receives signals from downlink channel h_i and backscatters the downlink signals to A_j. While A_j is in the listening mode, it receives signals, including the cascade backscattered signal from A_i and the signal $s(t)$ from the RFS. The received signals at A_i and A_j at t are:

$$y_i(t) = s(t) h_i(t) + \omega_i(t), \tag{20a}$$

$$y_j(t) = \alpha_i s(t) h_i(t) h_{i,j}(t) + s(t) h_j(t) + \omega_j(t), \tag{20b}$$

where $\omega_i(t)$ and $\omega_j(t)$ are AWGN, and $\forall i, j \in \{1, 2, ..., N\}$, $\alpha_i \neq 0$.

Phase 2: Similarly, in time slot $t^{'}$, A_j operates in the backscattering mode when A_i is listening. The received signals at A_i and A_j are

$$y_i(t^{'}) = \alpha_j s(t^{'}) h_j(t^{'}) h_{j,i}(t^{'}) + s(t^{'}) h_i(t^{'}) + \omega_i(t^{'}), \tag{21a}$$

$$y_j(t^{'}) = s(t^{'}) h_j(t^{'}) + \omega_j(t^{'}). \tag{21b}$$

After the above two phases, A_i and A_j can combine the received signals in above two phases to construct a *triangle channel* with three sides h_i, h_j and $h_{i,j}$ or $h_{j,i}$.

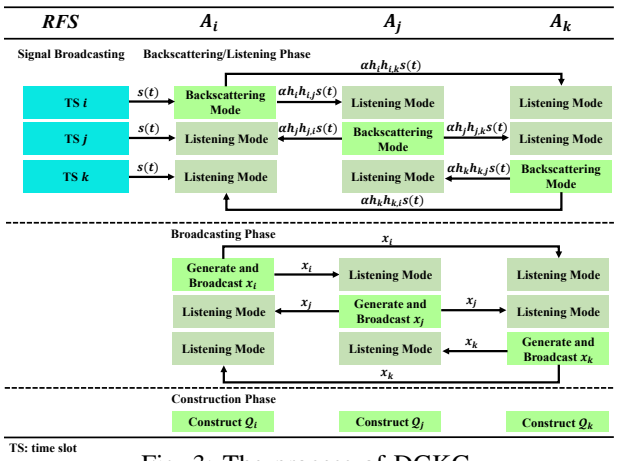


Fig. 3: The process of DGKG.

$$Y_i = (y_i(t^{'}) - y_i(t))y_i(t) = [\alpha_j s(t^{'})h_j(t^{'})h_{j,i}(t^{'}) + (s(t^{'})h_i(t^{'}) - s(t)h_i(t)) + (\omega_i(t^{'}) - \omega_i(t))](s(t)h_i(t) + \omega_i(t)), \tag{22a}$$

$$Y_j = (y_j(t) - y_j(t^{'}))y_j(t^{'}) = [\alpha_i s(t)h_i(t)h_{i,j}(t) + (s(t)h_j(t) - s(t^{'})h_j(t^{'})) + (\omega_j(t) - \omega_j(t^{'}))](s(t^{'})h_j(t^{'}) + \omega_j(t^{'})). \tag{22b}$$

If $(t^{'} - t) < T_{ch}$, where T_{ch} is coherence time [51], then due to channel reciprocity, we have $h_i(t) \approx h_i(t^{'})$, $h_{i,j}(t) \approx h_{j,i}(t^{'})$. After two slots above, if we assume there is no noise in the system, the constructed triangle channel information Y_i and Y_j of A_i and A_j can be presented in time-discrete manner as follows:

$$Y_i[n] = \alpha_j h_j[n]h_{j,i}[n]h_i[n]s^2[n], \tag{23a}$$

$$Y_j[n] = \alpha_i h_i[n]h_{i,j}[n]h_j[n]s^2[n]. \tag{23b}$$

Similar to CGKG, A_i and A_j can average the constructed information of $J = N_{cp} + N - 1$ samples and **can obtain triangle channel information without conducting channel estimation**:

$$v_i = \frac{1}{J}\sum_{n=L-1}^{N_{cp}-1} Y_i[n] = \frac{1}{J}\sum_{n=L-1}^{N_{CP}-1} |\alpha_j h_j[n]h_{j,i}[n]h_i[n]s^2[n]| = |\alpha_j h_j h_{j,i} h_i|, \tag{24a}$$

$$v_j = \frac{1}{J}\sum_{n=L-1}^{N_{cp}-1} Y_j[n] = \frac{1}{J}\sum_{n=L-1}^{N_{cp}-1} |\alpha_i h_i[n]h_{i,j}[n]h_j[n]s^2[n]| = |\alpha_i h_i h_{i,j} h_j|. \tag{24b}$$

Where $\frac{1}{J}\sum_{n=L-1}^{N_{cp}-1} |s^2[n]| = 1$. Note that even though α_i and α_j are different, the correlation coefficient of v_i and v_j between A_i and A_j consistently equals to 1 based on our analysis in [46]. Thus, for simplicity, we use a uniform symbol α to express the backscattering coefficient of all BDs. Obviously, the two triangle channels constructed in (36) can serve as the shared randomness between A_i and A_j for pairwise key generation since they are identical if channel reciprocity holds, i.e., $h_{i,j} \approx h_{j,i}$ and therefore $v_i \approx v_j$.

2) DGKG Design: DGKG consists of three phases: backscattering/listening, broadcasting and construction. DGKG combines all $N(N-1)/2$ triangle channel measurement among N BDs as a shared randomness. In the backscattering/listening phase, each BD A_i can obtain $N-1$ triangle channel measurements. During the broadcasting phase, the other BDs transmit combinations of their own $N-1$ triangle channel measurements to A_i. These combinations include the remaining $(N-1)(N-2)/2$ triangle channel information and can be utilized in the subsequent construction phase. In the construction phase, A_i utilizes its own $N-1$ measurements obtained in the backscattering/listening phase to determine the $(N-1)(N-2)/2$ triangle channel measurements from the combinations shared by the other BDs. Fig. 3 shows the process of DGKG. We use A_i, A_j, and A_k to represent any BDs.

Backscattering/Listening Phase: In the backscattering/listening phase, the RFS continuously broadcasts RF signals in each frame, each BD takes its turn to operate in the backscattering mode while the others work in the listening mode. For example, there are N frames in this phase and in the i-th frame, BD A_i operates in the backscattering mode while the others work in the listening mode. Herein, each of the rest of the $N-1$ BDs A_j $(j \in \{1,2,...,N\}, i \neq j)$ can construct a triangle channel with A_i by utilizing the pairwise key generation scheme described in Subsection IV-B1. And we let the triangle channel measurement constructed between A_i and A_j be $T_{i,j} = \alpha h_i h_{i,j} h_j = v_i = v_j$ $(i \neq j)$.

After the N frames, each BD can construct $N-1$ triangle channels with other $N-1$ BDs. The triangle channel measurement set constructed by A_i in this phase is $\mathcal{Q}_i^1 = \{T_{1,i}, T_{2,i}, ..., T_{i-1,i}, T_{i+1,i}, ..., T_{N,i}\} \subseteq \mathcal{Q}_i$. $\mathcal{Q}_i = \{T_{1,2}, ..., T_{1,N}, ..., T_{i,i+1}, ..., T_{i,N}, ..., T_{N-1,N}\}$ denotes the set of all triangle channel measurements used for group key generation. We have $|\mathcal{Q}_i^1| = N-1$ and $|\mathcal{Q}_i| = N(N-1)/2$.

Broadcasting Phase: In this phase, each BD in each frame broadcasts combinations of its constructed triangle channel measurements with a weight vector to all other BDs. For example, in the j-th frame of broadcasting, BD A_j broadcasts combinations $x_j = g_j \mathcal{Q}_j^1$, where $g_j = \mathcal{B}^{r_j \times (N-1)}$ is a weight matrix, in which $\mathcal{B}$ is a Boolean variable matrix with only 1 and 0, and r_j is the number of combinations broadcasted by A_j. A weight matrix set $g = [g_1; ...; g_j; ...; g_N]$ is formed with $g \in \mathcal{B}^{\sum_{j=1}^{N} r_j \times (N-1)}$.

Construction Phase: In this phase, each BD A_i obtains all triangle channel measurements $T_{i,j}$ in set $\mathcal{Q}_i$ by calculating the triangle channel measurements they have not obtained in the backscattering/listening phase (i.e., $\mathcal{Q}_i^2 = \mathcal{Q}_i - \mathcal{Q}_i^1$ and $|\mathcal{Q}_i^2| = (N-1)(N-2)/2$) from the received combinations from other $N-1$ BDs (i.e., $x_1, ..., x_{i-1}, x_{i+1}, ..., x_N$). This is because the weight matrix g is also known to all BDs in the system and A_i can further utilize the combinations and the weight matrices to calculate Q_i^2 by solving a number of linear equations.

Then, each BD A_i can generate the set of all triangle channel measurements $\mathcal{Q}_i$ by combining $\mathcal{Q}_i^1$ and $\mathcal{Q}_i^2$. Thus, all BDs, respectively, can exploit their triangle channel measurement set $\mathcal{Q}_i$ as an identical shared randomness to generate the secret group key with a sequence of measurements $V_i = \{\mathcal{Q}_i(1), ..., \mathcal{Q}_i(M)\}$ $(i \in \{1,2,...,N\})$, where M is the length of the key and $\mathcal{Q}_i(j)$ $(j \in \{1,2,...,M\})$ are the measured values of the set of $N(N-1)/2$ triangle channels in different independent measurements.

In comparison to existing schemes, DGKG utilizes $N(N-1)/2$ triangle channel information to form shared randomness, whereas the existing schemes only utilize $N(N-1)/2$ inward channel information to form shared randomness. Notably, our scheme introduces an additional N downlink channel information, greatly enhancing the randomness and security of the shared information.

3) Eavesdropping Analysis: In this subsection, we analyze the leaked information in DGKG under the eavesdropping attack. We establish lower-bound and upper-bound restrictions on the broadcasted combinations in the broadcasting phase for DGKG to ensure that DGKG can generate a secure group key under the presence of eavesdroppers.

In the backscattering phase, Eve can construct N triangle channels with all BDs (i.e., $\mathcal{Q}_e^1 = \{T_{1,e}, T_{2,e}, ..., T_{N,e}\}$). If Eve is positioned at a sufficient distance from any BDs, let's consider the case where Eve intends to compromise the group key information of A_1. The triangle channels constructed by Eve between itself and other A_j $(j \in \{2,3,...,N\})$, and the

triangle channels constructed by A_1 between A_1 and A_j, are uncorrelated.

In the broadcasting phase, Eve can intercept all the combinations sent from every BD. The number of intercepted combinations is $\sum_{j=1}^{N} r_j$. To prevent Eve from deducing all channel information through these intercepted combinations, it is necessary to satisfy the following upper-bound restriction:

$$\sum_{j=1}^{N} r_j < N(N-1)/2, \quad j \in \{1, 2..., N\}. \tag{25}$$

In addition, to ensure a robust estimation of the unknown variable BD A_i for solving a set of linear equations, it is necessary to have more received combinations than the number of unknown triangle channel measurements. Consequently, a lower-bound restriction can be formulated as follows:

$$\sum_{j=1}^{N-1} r_j \geq (N-1)(N-2)/2, \quad i \in \{1, 2, ..., N\}, j \neq i. \tag{26}$$

If the number of combinations satisfies both the lower-bound restriction (25) and the upper-bound restriction (26) simultaneously, it ensures that the generated group key is reliable. In this case, Eve cannot compromise the group key information by only intercepting the combinations sent by BDs during the broadcasting phase.

If Eve aims to compromise more information, it should position itself very close to a BD like A_1 in order to intercept $N-1$ measurements from the other $N-1$ BDs, excluding A_i. This enables Eve to obtain measurements that are similar to those obtained by A_1 and allows it to solve the $N(N-1)/2$ triangle channel measurements using the intercepted information. However, if Eve is located at a sufficient distance from A_1, the triangle channels constructed by Eve between itself and other A_j ($j \in \{2, 3, ..., N\}$) are not correlated with the triangle channels constructed by A_1 between A_1 and A_j.

Nevertheless, if Eve is very close to A_1, Eve can intercept some information of $N-1$ triangle channels information that are constructed between A_1 and A_i ($i \in \{2, 3, ..., N\}$). In pairwise key generation between two BDs, the eavesdropping model can be represented using the **cross-correlation coefficient**. Specifically, we use cross-correlation coefficients cor_{h_1,h_e}, cor_{h_i,h_e}, $cor_{h_{1,i},h_{1,e}}$, and $cor_{h_{i,1},h_{i,e}}$ to measure the correlations between h_1 and h_e , h_i and h_e, $h_{1,i}$ and $h_{1,e}$, and $h_{i,1}$ and $h_{i,e}$, respectively. When Eve is close to A_1 (away from A_i), we have $cor_{h_{1,i},h_{1,e}} = 0$ and $cor_{h_i,h_e} = 0$, $cor_{h_{e,1},h_{i,1}} > 0$ and $cor_{h_1,h_e} > 0$. Therefore, the cross correlation coefficient of two triangle channels information sets, $\mathcal{Q}_1^1 = \{T_{2,1}, T_{3,1}, ..., T_{N,1}\}$ measured by A_1 and $\mathcal{Q}_e^1 - T_{1,e} = \{T_{2,e}, T_{3,e}, ..., T_{N,e}\}$ measured by Eve, can be denoted as cor_{Eve,A_1} [55]. For simplicity and uniformity, the cross correlation coefficient cor_{Eve,A_1} is used to represent the value of non-zero cross correlation coefficient in the pairwise key generation between two BDs, i.e., $cor_{Eve,A_1} = cor_{h_1,h_e} = cor_{h_{e,1},h_{i,1}}$. The actual effect of the cross-correlation coefficient on leaked information is analyzed in Subsection V-C.

4) Achievable Rates of DGKG: In this subsection, we analyze the achievable secret key rate of DGKG under the presence of an eavesdropper. Similar to the analyzing analysis in Subsection IV-A4, we can first obtain lossless coding rate of DGKG as follows:

$$\begin{aligned} R_{Key}(M) &= \frac{1}{M} H(\mathcal{Q}(1), \mathcal{Q}(2), ..., \mathcal{Q}(M)) \\ &= \frac{1}{M} H(T_{1,2}(M), T_{1,3}(M)..., T_{N,N-1}(M)), \end{aligned} \tag{27}$$

where $T_{i,j}(M) = \{\alpha h_i h_{i,j} h_j(1), ..., \alpha h_i h_{i,j} h_j(M)\} (i \neq j)$. And R_{pro} can be expressed as:

$$\begin{aligned} R_{Pro}(M) = &\max_{1 \leq i \leq N} \frac{1}{M} I([T_{1,2}(M), T_{1,3}(M)..., T_{N,N-1}(M)] \\ &|[x_1, x_2, ..., x_N], [T_{i,1}, T_{i,2}, ..., T_{i,N}]) \\ = &\max_{1 \leq i \leq N} \frac{1}{M} I([\boldsymbol{T}_1(M), ..., \boldsymbol{T}_N(M)] | \boldsymbol{x}(M), \boldsymbol{T}_i(M)), \end{aligned} \tag{28}$$

In the broadcasting phase of each BDs, Eve can obtain information $[\boldsymbol{x}(M), \boldsymbol{T}_e(M)]$ as analyzed in previous subsection. The key rate between the common information of the BD group and the eavesdropped information of Eve can be defined as:

$$R_{Eve}(M) = \lim_{M \to \infty} \frac{1}{M} I([\boldsymbol{T}_1(M), ..., \boldsymbol{T}_N(M)]; [\boldsymbol{x}(M), \boldsymbol{T}_e(M)]). \tag{29}$$

By dropping the key length indices for simplicity, the SKR of DGKG at A_i can be expressed as:

$$\begin{aligned} R_{Sec}^{DGKG} &= R_{Key} - R_{Pro} - R_{Eve} \\ &= I([\boldsymbol{T}_1, ..., \boldsymbol{T}_N]; [\boldsymbol{x}, \boldsymbol{T}_i]) - I([\boldsymbol{T}_1, ..., \boldsymbol{T}_N]; [\boldsymbol{x}, \boldsymbol{T}_e]) \\ &= I([\boldsymbol{T}_1, ..., \boldsymbol{T}_N]; \boldsymbol{T}_i | \boldsymbol{x}) - I([\boldsymbol{T}_1, ..., \boldsymbol{T}_N]; \boldsymbol{T}_e | \boldsymbol{x}) \end{aligned} \tag{30}$$

Therefore, the achievable SKR of DGKG is $I([\boldsymbol{T}_1, ..., \boldsymbol{T}_N]; \boldsymbol{T}_i | \boldsymbol{x}) - I([\boldsymbol{T}_1, ..., \boldsymbol{T}_N]; \boldsymbol{T}_e | \boldsymbol{x})$. Since it is difficult to measure how much information can be obtained in $\boldsymbol{T}_e$ using the correlation coefficient, we can focus on analyzing $I([\boldsymbol{T}_1, ..., \boldsymbol{T}_N]; \boldsymbol{T}_i | \boldsymbol{x})$ to obtain the achievable key rate. We can apply the chaining rule of mutual information to perform the following transformations over $I([\boldsymbol{T}_1, ..., \boldsymbol{T}_N]; \boldsymbol{T}_i | \boldsymbol{x})$:

$$\begin{aligned} R_{Gen}^{DGKG} &= I([\boldsymbol{T}_1, ..., \boldsymbol{T}_N]; \boldsymbol{T}_i | \boldsymbol{x}) \\ &= I(r_j; r_i | \boldsymbol{\beta}) + I([\boldsymbol{T}_1, ..., \boldsymbol{T}_{j-1}, \boldsymbol{T}_{j+1}, ..., \boldsymbol{T}_N]; \boldsymbol{T}_i | \boldsymbol{T}_j, \boldsymbol{x}) \\ &\geq I(\boldsymbol{T}_j; \boldsymbol{T}_i | \boldsymbol{x}). \end{aligned} \tag{31}$$

Thus, the achievable KGR of DGKG is $I(\boldsymbol{T}_j; \boldsymbol{T}_i | \boldsymbol{x})$. However, we can not derive the closed-form expression of achievable KGR by analyzing the distribution of variables in $I(\boldsymbol{T}_j; \boldsymbol{T}_i | \boldsymbol{x})$. This is because $\boldsymbol{T}_i = \{\alpha h_i h_{i,1} h_1, \alpha h_i h_{i,2} h_2, ..., \alpha h_i h_{i,N} h_N\}$ is obtained by multiplying three Rayleigh-distributed variables and therefore $\boldsymbol{T}_i$ does not obey Gaussian distribution. Based on some assumptions (i.e., $h_i h_{i,j} h_j$ obeys Gaussian distribution and correlated noise variables are independent from all channels, etc.), the closed-form form expression can be obtained, but we do not elaborate it in detail herein. To analyze the SKR in DGKG, we provide numerical simulations to analyze SKR in different settings (i.e., different SNRs or the different cross-correlation coefficients between the main channel and the wiretap channel) in Subsection V-C.

C. DHGKG Scheme

In this subsection, we present the design of DHGKG, which **aims to address the limitations of DGKG when dealing with a large number of BDs.** When the group size becomes large, DGKG exhibits two negative effects. Firstly, the inward channel measurement in the later frames of the backscattering phase may deviate from the measurement in the early frames, resulting in a non-negligible residue when extracting cascade channel measurements [6]. Secondly, the reciprocity error is likely to increase when there is a large time interval between two triangle measurements on the same bidirectional link. To counter these issues, we propose DHGKG, where the BDs in a large group are first divided into smaller independent sub-groups based on their geographic locations. Sub-group keys are then used to generate the entire group key. In the following sections, we outline the process of dividing a large group into independent sub-groups and explain how these sub-groups communicate with each other to generate the group key.

1) BD Grouping: The division of BDs into multiple sub-groups is a challenging task known as the maximally diverse grouping problem (MDGP), which has been proven to be NP-hard [64]. In the case of NP-hard problems, heuristic algorithms are commonly employed to search for solutions [65]. However, the optimization problem is hard for our scenario since channel length is not the only factor that affects the BDR of generated group key. The number of BDs in each group also exerts an impact, as demonstrated by the simulation results in Fig. 4(a) and 4(c).

We propose a feasible grouping method for DHGKG based on the locations of BDs. However, achieving an optimal grouping outcome is an area that requires further investigation in our future work. To determine the positions of BDs, the RFS utilizes backscattered signals and employs robust techniques such as probing received signal strength (RSS) and angle of arrival (AoA), or measuring the time of arrival (ToA) and AoA of the signals. Previous studies have extensively explored these reliable methods [44], [66]. In this paper, we illustrate the calculation of BD locations using RSS and AoA. The grouping of BDs involves three phases, which are described in detail below.

Backscattering Phase: Each BD performs in the backscattering mode in turns and backscatters its downlink signal. The RFS uses the RSS and AoA of the backscattered signal to locate the position of each BD. To achieve accurate positioning estimated from the RSS and AoA, the RFS needs to perform SI cancellation (refer to Subsection IV-A1).

Grouping Phase: The RFS employs an unsupervised classification algorithm to group the BDs into multiple sub-groups based on their location matrix $\mathcal{L}$. The literature offers various robust classifiers to handle situations where the number of classes is unknown, such as unsupervised decision tree [67], complete-linkage hierarchical clustering [68], and mean-shift [69]. Based on our experiments, all three mentioned methods can be utilized for grouping BDs as long as their parameters are appropriately configured. For instance, the bandwidth parameter in mean-shift or the maximum distance d_τ between adjacent sub-groups in complete-linkage hierarchical clustering can be adjusted to achieve consistent grouping outcomes. Herein, we select complete-linkage hierarchical clustering as our preferred grouping algorithm for several reasons. Firstly, it produces a clustering dendrogram that provides valuable insights into the impact of d_τ on the clustering results, enabling further refinement. Additionally, we can impose a maximum limit N_τ of devices in each group during the clustering process, as the size of sub-groups has a significant influence on the BDR of the group key (refer to Fig. 4(c)). If a particular sub-group exceeds this maximum threshold, it will be further subdivided until the number of devices in each sub-group is within the allowed limit N_τ.

Grouping Result Broadcasting Phase: Once the group is divided into multiple sub-groups, the RFS proceeds to select a central group among these sub-groups. The central group collects key information from each sub-group and combines their sub-group keys to generate the complete group key. In order to facilitate the exchange of key information between the central group and the remaining sub-groups, it is crucial to select a central group that minimizes the average distance to all other sub-groups among the available candidates.

After selecting the optimal central group, the RFS can proceed with the selection of gateway BDs based on the distances between different sub-groups and the central group. If the distance between a BD A_i in one sub-group (denoted as sub-group I) and the nearest BD A_j in another sub-group (denoted as sub-group II) is the shortest among all the BDs in sub-group I, BDs A_i and A_j are chosen as the gateway BDs to facilitate information exchange between sub-group I and sub-group II. It is worth noting that the selection process can also take into consideration factors such as BD energy availability or computation capability, in addition to the distance criterion.

2) Group Key Generation in DHGKG: Upon receiving the grouping result from the RFS, each sub-group of BDs executes the DGKG to generate a sub-group key. Subsequently, the gateway BD A_{BD_k} of sub-group k ($k \in \{1, 2, ..., K\}$) and the gateway BD $A_{BD_{cen}}$ of the central group establish a pairwise key using the pairwise key generation scheme. After that, the sub-group k shares its secret sub-group key V_{sub}^k with the central group through the established secure channel between gateway A_{BD_k} and $A_{BD_{cen,k}}$. Subsequently, $A_{BD_{cen,k}}$ utilizes the central group key V_{cen} to encrypt the acquired V_{sub}^k and broadcasts it within the central group. Finally, the gateway BDs of the central group can construct the entire group key by *XOR* all collected sub-group keys and send this group key through the secure inward channels between gateway BDs to the sub-groups.

TABLE II: Default simulation parameters

Environment and key generation parameters:	
Wireless Channel	Rayleigh fading channel
Path-loss exponent	$\lambda = 5$
Backscatter ratio	$\alpha = 0.5$
Signal-to-noise ratio (SNR)	30dB
Switching interval of BD between two modes	10^{-4}
Bits of key	$5,000$
Number of simulation runs	100
Modulation method	OFDM modulation
Quantization approach	1 bit quantization

3) Eavesdropping Analysis: We conducted an analysis of the leaked information in DHGKG under eavesdropping. The eavesdropping model assumes the presence of a single passive eavesdropper in the BC system. However, in DHGKG, there are multiple sub-groups and the keys generated by each sub-group are independent of one another. Eve can only select one sub-group to conduct eavesdropping, and it possibly obtains a certain sub-group key. One possible method for Eve to acquire the group key is by compromising the pairwise keys of the gateway BDs and recovering the exchanged sub-group key information between the gateway BDs. However, it has been analyzed in [6] that it is difficult for Eve to compromise the pairwise key information simply by being physically close to the gateway BDs. Therefore, the only way to obtain the group key would be to have multiple eavesdroppers work together, each eavesdropping on a different sub-group, and then combining all the eavesdropped information to obtain the group key. As a result, we can conclude that the security of DHGKG against eavesdropping attacks is higher than that of DGKG, as the cost of Eve to compromise DHGKG is significantly higher.

V. Performance Evaluation

In this section, we analyze and compare the computation and communication complexity of the three proposed schemes and Multi-D2D (mentioned in Section II), as shown in Table III. Additionally, we conduct Monte Carlo simulations to study the effects of various parameters on the performance of the three proposed schemes and the leaked information under eavesdropping attacks.

A. Computation and Communication Complexity Analysis

Multi-D2D requires any two BDs to generate a pairwise key. Subsequently, A_i needs to communicate with A_j in order to obtain the remaining triangle channel measurements that it cannot directly construct (i.e., $T_{j,k}$ where $k \neq i, j$). Therefore, the computation complexity and communication complexity of each BD are both $O(N^2)$ since each BD needs to encrypt transmitted information $N-1$ times and decrypt received information $(N-1)(N-2)/2$ times in order to decode messages sent by other BDs. In CGKG, the RFS just needs to deduct each round-trip channel with $\alpha h_1 h_1^b$ to obtain the difference set $\mathcal{D}$ and further broadcast $\mathcal{D}$ to all BD. As a result, the computation complexity of CGKG is $O(N)$ and its communication complexity is $O(1)$. Additionally, the computation complexity of each BD in CGKG is also $O(N)$ as it only needs to subtract one time to acquire $\alpha h_1 h_1^b$ and add $\alpha h_1 h_1^b$ $N-1$ times with the remaining round-trip differences to obtain its round-trip measurement set R_i.

In DGKG or DHGKG, each BD needs to solve the linear equation using Gauss-Jordan elimination [70] to obtain the group or sub-group key. This results in a computation complexity of $O(N^3)$ and $O(N_\tau^3)$, respectively, for the generation of one (sub-)group key. In DHGKG, where there are multiple sub-groups (referred to as K sub-groups), the computation complexity can be expressed as $O(K \cdot N_\tau^3)$.Furthermore, each BD (except gateway BDs) needs to operate broadcasting once

TABLE III: Comparison of the direct expansion scheme with our proposed three schemes regarding efficiency, security, advantage, disadvantage, and suitable application scenarios

Scheme	Multi-D2D	CGKG	DGKG	DHGKG
CompC of RFS	-	$O(N)$	-	$O(N^3)$
CompC of BD	$O(N^2)$	$O(N)$	$O(N^3)$	$O(K \cdot N_\tau^3)$
CommC of BD	$O(N)$	$O(1)$	$O(1)$	$O(1)$
Advantage	Resistant to untrusted RFS; High security performance.	High efficiency and accuracy; Good adaptability in a large-scale group of BDs and in fast fading scenario.	Resistant to untrusted RFS; High security performance.	Resistant to untrusted RFS; Good adaptability in a large-scale group of BDs.
Disadvantage	Require substantial computation, communication, and time.	Overreliance on RFS.	Low performance in a large group of BDs and in fast fading scenario; Relatively high computation needs of BDs.	High computation and communication requirement of RFS.
Application Scenario	-	A trusted RFS is in the system; The scale of BD group is huge, in a low or fast fading channel scenario.	When RFS is untrusted but serviceable, the number of BD small (within 5), in a low fading channel scenario.	When RFS is untrusted but serviceable, the number of BD group is huge, a low fading channel scenario.
Malicious BD Resistance	◓	●	◓	◓
CCA Resistance	●	●	●	●
SMA Resistance	●	◓	●	●
URSA Resistance	●	○	●	◓

1. CompC: Computation complexity; CommC: Communication complexity; N: The number of BDs; Settings: It is noteworthy that the values of CompC and CommC in the four schemes are obtained for the same number of BDs N.
2. ○: Have no resistance; ◓: Conditionally have resistance; ●: Have resistance.

in DGKG and need to conduct an additional key reception with the gateway BDs in its sub-group in DHGKG, and therefore its communication complexity is $O(1)$. The RFS in DHGKG employs unsupervised classification, specifically the standard hierarchical clustering, for grouping, with a computation complexity of $O(N^3)$ [71]. The algorithm to select the central group and gateway BDs involves iterating over an adjacency matrix with BDs, resulting in a computation complexity of $O(N^2)$. Finally, the computation complexity of the RFS in DHGKG is $O(N^3 + N^2) = O(N^3)$. It is worth noting that N_τ is smaller than 6 (as analyzed in Subsection V-C3), while N has no limitation. Therefore, the computation complexity of DHGKG is much smaller than that of DGKG when the number of BDs is relatively large (e.g., $O(3 \times 4^3) \ll O(12^3)$).

B. Experimental Settings and Evaluation Metrics

We model each channel tap as an independent complex Gaussian random variable (Rayleigh fading) with its average power that follows an exponentially decaying power delay profile by referring to the system model specified in Section III. The channel is given in the form of $h = \vartheta d^{-\frac{\lambda}{2}}$ [51], where ϑ is a circularly symmetric complex Gaussian (CSCG) variable, d is the distance between a considered transmitter and a receiver, and λ is a path-loss exponent. Some basic simulation parameters are listed in Table II. Since BDs are often used in storage systems and medical diagnoses, the value range of the path-loss exponent in a building can be obtained according to an empirical formula, which is $\lambda \in [4, 6]$ [51], we take its mean value as the value of the path-loss exponent and let $\lambda = 5$. In our experimental simulations, the default number of BDs is 7 if there is no any special explanation. The distances between the BDs are randomly generated within a fixed range, resulting in a fully meshed topology with a complete graph for their formed BC system. The metrics used to evaluate the performance of the generated group key are as follows:

1) Average Mutual Information (AMI): Mutual Information (MI) is a general measure of dependence between two random variables. It helps verify the feasibility of the constructed channels as shared randomness. AMI is the average value of a number of MIs. AMI measures the average mutual information between multiple pairs of legitimate devices, which is defined as:

$$AMI = \frac{1}{P} \sum_{p=1}^{P} I_p \left(V_i, V_j \right),$$

where $I(\cdot, \cdot)$ represents the mutual information between two sequences, P is the total number of pairs of BDs, where $I_p(V_i, V_j)$ represents the MI between V_i and V_j when A_i and A_j is the p-th pair of BDs among all pairs.

2) Average Leaked Information (ALI): Leaked information (LI) measures the mutual information between A_i's generated group key sequence and Eve's overheard group key sequence. It represents the information that could be eavesdropped by Eve. While ALI is the average value of a number of LIs. Concretely, in our paper, ALI measures the average leakage information of group key information within the group with N BDs. Therefore, ALI can be defined as:

$$ALI = \frac{1}{N} \sum_{i=1}^{N} I \left(V_i, V_e \right),$$

where N is the number of BDs in the group, and V_e is the key sequence at Eve.

3) Bit Disagreement Ratio (BDR): BDR is the number of disagreed bits divided by the total number of bits in a generated key. Concretely, a lower BDR represents a higher consistency of the generated group key.

4) Key Generation Rate (KGR): KGR describes the amount of key bits produced in one second/measurement. It indicates how much the key information/bits that is shared among legitimate devices.

5) Secret Key Rate (SKR): SKR describes the amount of secret key bits produced in one second/measurement in the absence of an attacker. The SKR of keys between two legitimate devices when under eavesdropping can be defined as:

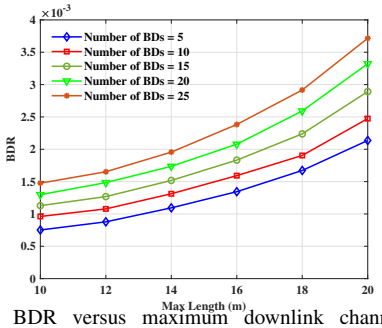
$$SKR = I(V_i; V_j | V_e).$$

C. Simulation Results

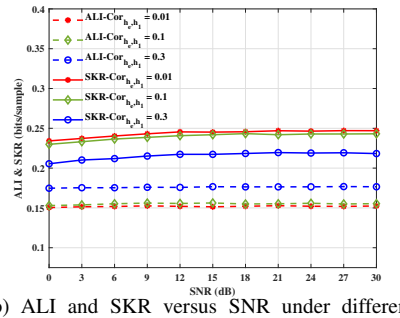
We study the impact of external parameters (e.g., number of BDs and environmental parameters) on the evaluation metrics (i.e., AMI, ALI, and BDR).

1) Performance of CGKG: We test the generated group key performance and security of CGKG in both indoor and outdoor environments.

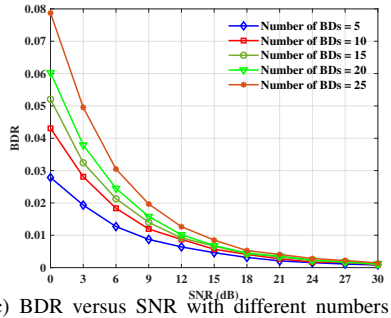
Fig. 4(a) illustrates the BDR of the generated group keys versus the maximum downlink channel length (i.e., the maximum distance between the RFS and BDs) for different numbers of BDs in the BC system. It can be observed that in all parameter settings within an indoor environment, the BDR increases as the maximum downlink channel length or the number of BDs grows. Increasing the downlink channel length results in increased RFS signal fading, which in turn decreases the received signal strength at the BD. A weaker received signal power increases the likelihood of noise interference, leading to degraded round-trip channel information and consequently higher BDR for the group key. Additionally, as the number of BDs increases, the received noise for both the BDs and RFS also increases. Consequently, the round-trip



(a) BDR versus maximum downlink channel length with different numbers of BDs.

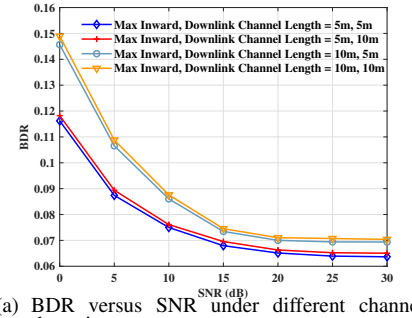


(b) ALI and SKR versus SNR under different correlation coefficients.

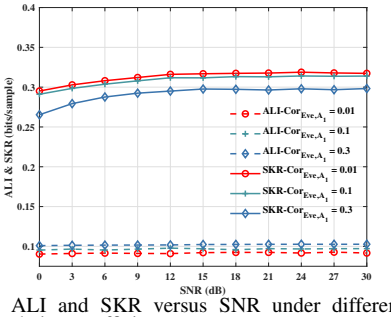


(c) BDR versus SNR with different numbers of BDs.

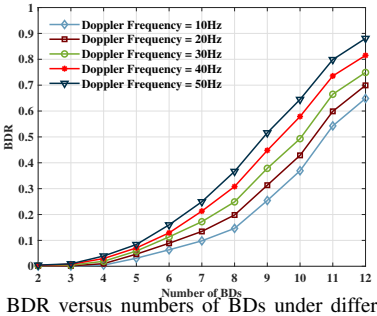
Fig. 4: Simulation results of CGKG.



(a) BDR versus SNR under different channel length settings.



(b) ALI and SKR versus SNR under different correlation coefficients.



(c) BDR versus numbers of BDs under different Doppler frequencies.

Fig. 5: Simulation results of DGKG.

channel measurement set $\mathcal{R}$ contains more AWGN, leading to higher BDR.

Eve is a passive attacker, therefore, its communication environment and eavesdropping location determine eavesdropped information. The change of the eavesdropping position causes a change in the correlation coefficient between Eve's eavesdropping channel and the downlink channel. Therefore, we discuss the variation of ALI and SKR when SNR and the correlation coefficient between an eavesdropping channel and a downlink channel change simultaneously in an indoor environment. The correlation coefficient between Eve's eavesdropping channel and the channel between legitimate devices is investigated. It is shown by experiments that even when Eve and the legitimate device are physically remarkably close, the correlation coefficient between Eve's eavesdropping channel and the channel between legitimate devices is generally about 0.01 and does not exceed 0.1 at the highest [72]. That is, we have $cor_{h_1,h_e} \leq 0.1$. Fig. 4(b) gives the ALI when the cross-correlation coefficient between the wiretapped channels and the main-channels is 0.01, 0.1, and 0.3, respectively, in different SNR settings. In all cross-correlation coefficient settings, SNR has a slight influence on the trifling growth of ALI because noises hinder Eve from obtaining accurate channel measurements. Since the SNR improves the consistency of keys across devices when the ALI remains unchanged, the SKR increases with the gradual increase of SNR. This figure also gives the ALI and SKR that Eve can achieve if it can make the correlation coefficient reach $cor_{h_1,h_e} = 0.3$ by some means.

Fig. 4(c) presents BDR versus signal-to-noise ratio (SNR), with different numbers of BDs in a group, in an outdoor environment with a fixed Doppler frequency of 25Hz. The channels are modeled using the Jakes model [51], which is a commonly used model for outdoor channels [73]. We selected an average Doppler frequency of 25Hz for consistency in our later experiments analyzing the impact of different Doppler frequencies ranging from 0Hz to 50Hz. In general, the BDR decreases as the SNR increases, as the negative influence of noise that causes errors in channel observation is mitigated. Particularly, at high SNR levels, the BDR approaches zero, and the effect of the number of BDs on the BDR becomes less significant. It is important to note that the increase in BDR shows a nearly linear relationship with the increase in the number of BDs, and the BDR performance remains acceptable even at relatively low SNR levels. However, a drawback of CGKG is that the RFS possesses knowledge of the group key, which poses a security risk: once the RFS is compromised, the entire group key is subsequently compromised.

2) Performance of DGKG: In this subsection, we discuss DGKG's key generation BDR and its security performance under eavesdropping in both indoor and outdoor environments.

Maintaining a short average distance between each BD in a group helps reduce the BDR. This simulation result provides a valid justification for the design of DHGKG.

As presented in Fig.5(a), different maximum inward channel length (i.e., the distance between two BDs) and maximum downlink channel length bring distinction to BDR in an indoor environment. We observe that the BDR is more sensitive to the maximum length of the inward channel. This sensitivity arises from our parameter setting, where we have chosen a backscattering coefficient of $\alpha = 0.5$. In situations where the received backscattered signal strength is relatively low, the accuracy of the cascaded channel measurement of the BDs can be compromised, leading to increased errors in the generation of the group key. Therefore, it is important to ensure that the distance between any two BDs within a group is not excessively long. Maintaining a short average distance between each BD in a group helps reduce the BDR. This simulation result provides a valid justification for the design of DHGKG.

Upon comparing Fig. 4(b) and 5(b), it becomes apparent that the security of DGKG is superior to that of CGKG in the context of eavesdropping attacks in an indoor environment, whereas the reverse is true in an outdoor environment. From the perspective of noises, DGKG incurs more AWGN than CGKG, bringing more randomness to Eve and thus causing Eve to compromise less information. In CGKG, multiple N BDs operate in a backscattering mode alternately. Consequently, the RFS's received signals are affected by N uncorrelated AWGN variables, while the BDs' signals encounter N AWGN variables. On the other hand, DGKG employs

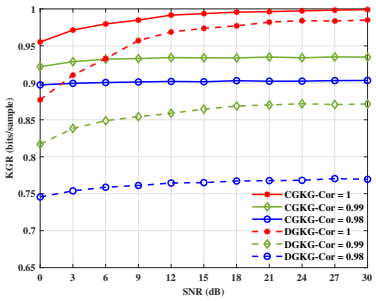


Fig. 6: The KGR of CGKG and DGKG under different correlation coefficient setting.

$N-1$ BDs in the listening mode, leading to $N \cdot (N-1)$ uncorrelated noise variables in the received signals of all BDs after N frames. Given the same correlation coefficient, $cor_{h_1,h_e} = cor_{Eve,A_1}$, CGKG yields a lower BDR than DGKG. Consequently, the bit generation rate (the number of key bits generated after reconciliation and before privacy amplification) of DGKG is slightly lower. Nevertheless, since the ALI of CGKG is greater than that of DGKG, the final SKR of DGKG is higher than that of CGKG. Thus, despite CGKG requiring less computational resources in BDs and producing highly accurate generated keys, it proves to be less resilient to eavesdropping attacks when compared to DGKG.

Fig. 5(c) shows that BDR increases with the growth of the number of BDs and the Doppler Frequency in an outdoor environment. This finding is derived from the constraints of the inward channel length, which is restricted from 1 meter (m) to 5m, and the downlink channel length, which is limited from 1m to 10m. As mentioned in the previous section, we have mentioned that the inward channel measurement in the later frames of backscattering phase is likely to deviate from the one in the early frames of backscattering, which could lead to non-negligible residue when solving the linear equations in the construction phase of DGKG. Accordingly, when the number of BDs in a group increases, BDR tends to grow exponentially due to this reason. Therefore, we need to limit the number of BDs within a group to ensure the BDR is within an acceptable range. A comparison between Fig. 4(a) and Fig. 5(c) reveals that the BDR in CGKG is approximately 10^{-3} when ($N = 10$), while the BDR in DGKG reaches 0.45. When the number of BDs is relatively large, the BDR becomes unacceptable. In such cases, DHGKG can be adopted as it significantly reduces the BDR and can still function in scenarios where the RFS cannot be fully trusted.

Fig. 6 depicts the impact of the decrease in correlation of measurement values on the same channel between two devices due to hardware imperfections on the KGR. It is evident that imperfect channel estimation caused by hardware imperfection has a greater impact on CGKG compared to DGKG. This is because in CGKG, the group-shared information only includes N sets of uplink and downlink channel information, while in DGKG , it includes N sets of downlink channel information as well as N sets of inward channel information. Furthermore, in CGKG, the round-trip channel information is obtained by a single multiplication of the uplink and downlink channel information, whereas in DGKG, the triangle channel information is obtained by three multiplications of two downlink channel information and one inward channel information. Three times multiplication on the channel information mentioned above intensifies the deterioration of the consistency of unrelated channel measurement values for both parties.

3) Performance of DHGKG: Table IV provides the simulation results of CGKG, DGKG and DHGKG under distinct Doppler frequencies in an outdoor environment. The BDR and AMI of these schemes are obtained with same distance data (between any BDs or between BD and RFS). It is evident that CGKG performs optimally across all Doppler frequencies. However, DGKG yields an unacceptably high BDR of 80%.

TABLE IV: AMI and BDR of three schemes at different Doppler frequencies.

Schemes	Metrics	10Hz	20Hz	30Hz	40Hz	50Hz	Group sizes
CGKG	BDR	0.10%	0.12%	0.17%	0.20%	0.28%	12
DGKG	BDR	80.20%	81.74%	82.97%	83.17%	83.65%	12
DHGKG-1	BDR	0.14%	0.32%	0.46%	0.65%	0.79%	4,2,2,2,2
DHGKG-2	BDR	0.15%	0.38%	0.56%	0.83%	1.10%	4,4,2,2
DHGKG-3	BDR	0.16%	0.46%	1.03%	1.77%	2.49%	4,4,4
DHGKG-1	AMI	0.9245	0.9141	0.9046	0.8968	0.8863	4,2,2,2,2
DHGKG-2	AMI	0.9246	0.9151	0.9074	0.8997	0.8908	4,4,2,2
DHGKG-3	AMI	0.9247	0.9164	0.9086	0.9015	0.8939	4,4,4

On the other hand, adopting DHGKG results in a significant decrease in BDR when the total number of BDs is 12. Notably, the performance of DHGKG is very similar to that of CGKG.

We present three group key performance results in the table based on three different grouping with DHGKG. According to Fig. 5(c), when $N = 6$ and Doppler frequency is greater than 20Hz, BDR exceeds 0.1. While when $N = 5$, However, when $N = 5$, the BDR is below 0.1, leading us to conclude that $N \leq 5$ is an acceptable range. Therefore, the maximum number of BDs in a sub-group is $N_\tau = 5$. The aforementioned three different grouping results are obtained by employing the complete-linkage clustering algorithm with varying maximum distance values, denoted as d_τ. As shown in Fig. 5(a), the length of the downlink channel has much less impact on the BDR than the length of the inward channel. Consequently, during clustering, we do not consider the downlink channel length, but only use the adjacency matrix between BDs as clustering input. In DHGKG-1, the threshold $d_\tau = 6m$, while in DHGKG-2 and DHGKG-3, $d_\tau = 7m$ and $d_\tau = 8m$, respectively. A smaller threshold value implies that only BDs close to each other can be clustered into one sub-group. Therefore, when d_τ is small, the number of BDs in a sub-group should be small. As depicted in Table IV, DHGKG-1 possesses the smallest BDR. Conversely, as d_τ increases, the distance between BDs within the sub-group becomes more considerable, leading to an increase in the number of BDs and subsequently, an increase in BDR. Therefore, DHGKG-3 exhibits the highest BDR among the three results.

Although reducing the number of BDs in each sub-group results in a lower BDR, it also increases the time, computation, and communication costs associated with generating a group key. However, when the number of sub-groups is large and the sub-groups are located farther apart, the MI of the pairwise key generated by the gateway BDs tends to be low [6]. We provide the AMI of the pairwise key generated between all gateway BDs in the bottom of Table IV. As examples, in DHGKG-2 and DHGKG-3, two non-central sub-groups each consisting of two BDs are merged into a single sub-group with four BDs in DHGKG-3. In DHGKG-2, these two sub-groups with two BDs need to establish two secure inward channels with a central group. Conversely, in DHGKG-3, only one channel needs to be established, and its length is the shortest among the two inward channels established in DHGKG-2. Consequently, merging the sub-groups leads to an increase in MI and an improvement in AMI due to the reduced length of the secure channels.

4) Active Attacks Analysis: This subsection analyzes the performance of CGKG, DGKG and DHGKG under four active attacks: insider attack (malicious BD or RFS), channel control attack (CCA) [47], [48], signal manipulative attack (SMA) [49], and untrusted RFS attack (URSA) [46]. Since malicious RFS attack is functionally equivalent to URSA, we combine these two types of attacks. Moreover, we specialize internal device attack to malicious BD attack. We assume that the purpose of malicious BDs is to pretend that they are legitimate BDs and participate in the group key generation process. Since key generation is normally conducted after BD authentication [44], the adversary cannot participate in the key generation and obtain the group key. However, since it is not easy for BDs to authenticate RFS, the possibility of the URSA still exists. Table III shows the resistance of three schemes under **malicious BD, CCA, SMA and URSA**.

Herein, "have resistance" means that the generated group key has no correlation with the manipulated information by an attacker when under CCA, SMA, and URSA. And "have no resistance" is opposite to "have resistance", i.e., the generated group key has correlation with the manipulated information of the attack. "Conditionally have resistance" means that the BC system is resistant to attacks when certain conditions are satisfied, e.g., the system can apply certain techniques or the system entities are cooperative.

In CGKG, to resist malicious BD attacks, the RFS authenticates the BDs before key generation [44], which allows for identification and exclusion of malicious BDs from the group of BDs. The RFS does not generate round-trip channel information $\alpha h_m h_m s^2$ with A_m, and the round-trip channel difference set broadcasted by the RFS in the RFS broadcast phase does not include the aforementioned round-trip channel information. As a result, the generated group key remains unaffected by the malicious BD. When under CCA, Mallory only needs to control the main channel between RFS and reference BD A_1, most of the group key information is composed by controlled channel information. Mallory can manipulate some bits of the group key (showed in ***Appendix-A***). When under SMA, the SMA turns to jamming attack (JA) in CGKG (showed in ***Appendix-B***). The JA can be alleviated by using frequency hopping (FH) technique [74]. However, there is no existing works show that the FH technique can be applied into the BC system. Therefore, if BC system can use FH technique, CGKG can resist SMA and JA. When under URSA, since CGKG relies on RFS to broadcast the group key, the group key information is exposed to RFS entirely. Therefore, the CGKG can be compromised entirely.

In Multi-D2D, when experiencing a malicious BD attack, each legitimate BD in the backscattering phase generates triangle channel information with a malicious BD A_m if no additional precautions are taken. As a result, the group key inevitably includes $N-1$ triangle channel information that A_m is aware of. Furthermore, during the BD broadcasting phase, a malicious BD can intentionally broadcast personalized information to gain control over the group key information. Above mentioned circumstance holds true for DGKG and DHGKG as well. Based on our analysis, there are two potential solutions to the problem at hand: 1) Mutual authentication among BDs. 2) Using a trusted RFS to authenticate in the first place: In this method, individual BDs are first authenticated to obtain a whitelist by a trusted RFS. For the detail process of the above two solutions, please refer to Subsection VI. If either of the preceding two solutions can be implemented, Multi-D2D, DGKG, and DHGKG are capable of resisting malicious BD attacks. The resistance of Multi-D2D to CCA, SMA, and URSA can be referred to [46].

DGKG and DHGKG are proposed based on the scheme of generating symmetric keys using triangle channel information (named Tri-Channel) [6], [74]. Therefore, DGKG and DHGKG are equipped with similar features to Tri-Channel when suffering from CCA and SMA. In Tri-Channel when under CCA, its resistance degrades only when three channels are controlled simultaneously [46] (analyzed and validated in ***Appendix-A***). However, in DGKG and DHGKG, there are many triangle channels and it becomes very difficult to control all of the triangle channels. Hence, compared with Tri-Channel, DGKG and DHGKG can resist CCA in a better way. When under SMA, unlike CGKG, manipulated signals injected into any two BDs do not cause jamming in DGKG and DHGKG. According to our analysis and validation through simulation in ***Appendix-B***, we found that DGKG and DHGKG inherit the feature of Tri-Channel with high resistance under SMA. The difference between DGKG and DHGKG is that a group of BDs needs to be firstly divided into multiple sub-groups by RFS in DHGKG. A particular case is if an untrusted RFS refuses to offer grouping results to BDs, i.e., converts to a deny of service attack, but continues to broadcast RF signals, DHGKG transfers into DGKG, which has reduced consistency and performance regarding key generation. If the untrusted RFS ensures the regular operation of the DHGKG scheme, the situation of DHGKG when under URSA becomes similar to DGKG. The untrusted RFS can control a part of the triangle channels and estimate the group key. Similar to Tri-Channel, DGKG and DHGKG have resistance to URSA if the untrusted RFS operates normally during key generation (proved in ***Appendix-C***).

VI. Further Discussion

A. Resistance of Malicious BD Attack

Based on our analysis, only CGKG provides full resistance to malicious BD Attack. It is important to discuss methods to ensure that both DGKG and DHGKG can also withstand such attacks. Based on our analysis, there are two potential solutions to the problem at hand: 1) Mutual authentication among BDs: This approach involves implementing a mutual authentication mechanism among the BDs. While the specific details of this method remain an open question, if the BDs can achieve mutual authentication, several techniques such as Multi-D2D, DGKG, and DHGKG can be employed to resist malicious BD attacks. 2) Using a trusted RFS to authenticate in the first place: In this method, individual BDs are first authenticated to obtain a whitelist by a trusted RFS. The process begins with the use of CGKG for key generation in the initial setup, in conjunction with the trusted RFS. Once the whitelist is obtained, practical applications can utilize such techniques as DGKG and DHGKG. To elaborate on the second method, a trusted RFS first authenticates the BDs and records their fingerprints to create the whitelist. The whitelist is then encrypted using the group key generated by CGKG and broadcasted. In practical applications, when BDs generate triangle channel information, they must cross-check with each other to ensure that they exist in the whitelist. However, the adaptability of this method to dynamically adding new BDs to the group needs further discussion. This entails techniques related to key management. Both of the above mentioned approaches offer potential solutions for tackling the problem, each with its implementation details and considerations.

B. A Hybrid Method

Our paper proposes three schemes that each have their own strengths and weaknesses, making them suitable for different scenarios. In a large-scale network, combining these three schemes (referred to as a hybrid method) may offer better performance compared to using a single scheme alone. Similar to the second approach mentioned in Subsection VI-A, one way to address the issue of malicious BD attacks is to integrate both CGKG and DGKG, rather than relying solely on DGKG. Through further discussion of the hybrid method, the advantages and disadvantages of each scheme can be balanced, leading to a more efficient, adaptive, and secure group key generation scheme in the BC system.

VII. Conclusions

This paper proposed BGKey, the first physical layer group key generation framework that can secure communications among a group of BDs. It contains three different schemes (i.e., CGKG, DGKG and DHGKG) that can be applied into different scenarios due to different advantages. The CGKG is based on the round-trip channel measurements between RFS and BDs and requires a trusted RFS to support group key generation. The DGKG, on the other hand, establishes the group key based on pairwise triangle channel measurements between BDs, thus the RFS has no knowledge of their generated group key. The DHGKG is designed to solve the problem of the DGKG, where its BDR increases dramatically when the scale of BD group enlarges. It hierarchically decomposes a single big group into multiple sub-groups to generate sub-group keys first and then create a final group key by aggregating the

sub-group keys. We analyzed their security under different attacks and compared their computation and communication complexity. We further evaluated their performance with numerical simulations under different parameter settings. The results show that the CGKG is the most efficient and accurate with low computation and communication complexity due to relying on a trusted RFS, but the least secure under eavesdropping. It can afford CCA, but performs badly under other active attacks like SMA and URSA. The DGKG establishes a group key without disclosing it to the RFS and holds better security than CGKG under eavesdropping and all analyzed active attacks, but the computation complexity of BD is high. While the DHGKG greatly improves the performance of DGKG with lower computation complexity at BD and also maintains excellent security under eavesdropping. It can also resist CCA and SMA, but performs worse than DGKG under URSA.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grant 62072351; in part by the Key Research Project of Shaanxi Natural Science Foundation under Grant 2023-JC-ZD-35, in part by the Open Project of Zhejiang Lab under Grant 2021PD0AB01; and in part by the 111 Project under Grant B16037; and in part by the U.S. National Science of Foundation through the Networking Technology and Systems (NeTS) Program under Award 2131507.

REFERENCES

[1] W. Liu, K. Huang, X. Zhou, and S. Durrani, "Next generation backscatter communication: Systems, techniques, and applications," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, p. 69, 2019. [Online]. Available: https://doi.org/10.1186/s13638-019-1391-7

[2] P. Wang, N. Wang, M. Dabaghchian, K. Zeng, and Z. Yan, "Optimal resource allocation for secure multi-user wireless powered backscatter communication with artificial noise," in *IEEE Infocom-IEEE International Conference on Computer Communications*, 2019.

[3] V. Talla, J. Smith, and S. Gollakota, "Advances and open problems in backscatter networking," *GetMobile: Mobile Computing and Communications*, vol. 24, no. 4, pp. 32–38, 2021.

[4] D. Ma, G. Lan, M. Hassan, W. Hu, and S. K. Das, "Sensing, computing, and communications for energy harvesting IoTs: A survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 1222–1250, 2019.

[5] Y. Xu, G. Gui, H. Gacanin, and F. Adachi, "A survey on resource allocation for 5G heterogeneous networks: Current research, future trends, and challenges," *IEEE Communications Surveys and Tutorials*, vol. 23, no. 2, pp. 668–695, 2021.

[6] P. Wang, L. Jiao, K. Zeng, and Z. Yan, "Physical layer key generation between backscatter devices over ambient RF signals," in *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications*, 2021, pp. 1–10.

[7] A. Mitrokotsa, M. R. Rieback, and A. S. Tanenbaum, "Classifying RFID attacks and defenses," *Information Systems Frontiers*, vol. 12, no. 5, pp. 491–505, 2010. [Online]. Available: Go to ISI://WOS:000284428000003

[8] X. Li, M. Zhao, M. Zeng, S. Mumtaz, V. G. Menon, Z. Ding, and O. A. Dobre, "Hardware impaired ambient backscatter NOMA systems: Reliability and security," *IEEE Transactions on Communications*, vol. 69, no. 4, pp. 2723–2736, 2021.

[9] N. Van Huynh, D. T. Hoang, X. Lu, D. Niyato, P. Wang, and D. I. Kim, "Ambient backscatter communications: A contemporary survey," *IEEE Communications surveys & tutorials*, vol. 20, no. 4, pp. 2889–2922, 2018.

[10] G. Yang, Y.-C. Liang, R. Zhang, and Y. Pei, "Modulation in the air: Backscatter communication over ambient OFDM carrier," *IEEE Transactions on Communications*, vol. 66, no. 3, pp. 1219–1233, 2018.

[11] D. Zanetti, B. Danev, and S. Apkun, "Physical-layer identification of UHF RFID tags," in *Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking*, 2010, pp. 353–364.

[12] K. Zeng, K. Govindan, and P. Mohapatra, "Non-cryptographic authentication and identification in wireless networks [security and privacy in emerging wireless networks]," *IEEE Wireless Communications*, vol. 17, no. 5, pp. 56–62, 2010.

[13] Q. Wang, "A novel physical layer assisted authentication scheme for mobile wireless sensor networks," *Sensors*, vol. 17, no. 2, p. 289, 2017.

[14] A. Juels, "RFID security and privacy: A research survey," *IEEE journal on selected areas in communications*, vol. 24, no. 2, pp. 381–394, 2006.

[15] T. Eisenbarth, S. Kumar, C. Paar, A. Poschmann, and L. Uhsadel, "A survey of lightweight-cryptography implementations," *IEEE Design & Test of Computers*, vol. 24, no. 6, pp. 522–533, 2007.

[16] C. Liu, Z. J. Haas, and Z. Tian, "On the design of multi-hop tag-to-tag routing protocol for large-scale networks of passive tags," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1035–1055, 2020.

[17] G. Ateniese, M. Steiner, and G. Tsudik, "Authenticated group key agreement and friends," in *Proceedings of the 5th ACM Conference on Computer and Communications Security*, 1998, pp. 17–26.

[18] M. Steiner, G. Tsudik, and M. Waidner, "Key agreement in dynamic peer groups," *IEEE Transactions on Parallel and Distributed Systems*, vol. 11, no. 8, pp. 769–780, 2000.

[19] W. Diffie and M. E. Hellman, "New directions in cryptography," in *Secure communications and asymmetric cryptosystems*. Routledge, 2019, pp. 143–180.

[20] Y. Kim, A. Perrig, and G. Tsudik, "Tree-based group key agreement," *ACM Transactions on Information and System Security (TISSEC)*, vol. 7, no. 1, pp. 60–96, 2004.

[21] S. Hong and N. Lopez-Benitez, "Enhanced group key generation algorithm," in *2006 IEEE/IFIP Network Operations and Management Symposium NOMS 2006*, 2006, pp. 1–4.

[22] M. G. Samaila, M. Neto, D. A. Fernandes, M. M. Freire, and P. R. Inácio, "Challenges of securing internet of things devices: A survey," *Security and Privacy*, vol. 1, no. 2, p. e20, 2018.

[23] K. Zeng, "Physical layer key generation in wireless networks: challenges and opportunities," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 33–39, 2015.

[24] M. Bloch, J. Barros, M. R. Rodrigues, and S. W. McLaughlin, "Wireless information-theoretic security," *IEEE Transactions on Information Theory*, vol. 54, no. 6, pp. 2515–2534, 2008.

[25] A. A. Gohari and V. Anantharam, "Information-theoretic key agreement of multiple terminals—Part II: Channel model," *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3997–4010, 2010.

[26] H. Liu, J. Yang, Y. Wang, Y. Chen, and C. E. Koksal, "Group secret key generation via received signal strength: Protocols, achievable rates, and implementation," *IEEE Transactions on Mobile Computing*, vol. 13, no. 12, pp. 2820–2835, 2014.

[27] G. Li, L. Hu, and A. Hu, "Lightweight group secret key generation leveraging non-reconciled received signal strength in mobile wireless networks," in *2019 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2019, pp. 1–6.

[28] C. D. T. Thai, J. Lee, J. Prakash, and T. Q. Quek, "Secret group-key generation at physical layer for multi-antenna mesh topology," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 1, pp. 18–33, 2018.

[29] P. Xu, K. Cumanan, Z. Ding, X. Dai, and K. K. Leung, "Group secret key generation in wireless networks: algorithms and rate optimization," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 8, pp. 1831–1846, 2016.

[30] L. Jiao, P. Wang, N. Wang, S. Chen, A. Alipour-Fanid, J. Le, and K. Zeng, "Efficient physical layer group key generation in 5G wireless networks," in *2020 IEEE Conference on Communications and Network Security (CNS)*. IEEE, 2020, pp. 1–9.

[31] Q. Wang, H. Su, K. Ren, and K. Kim, "Fast and scalable secret key generation exploiting channel phase randomness in wireless networks," in *2011 Proceedings IEEE INFOCOM*. IEEE, 2011, pp. 1422–1430.

[32] S. Xiao, Y. Guo, K. Huang, and L. Jin, "Cooperative group secret key generation based on secure network coding," *IEEE Communications Letters*, vol. 22, no. 7, pp. 1466–1469, 2018.

[33] J. Zhang, M. Ding, D. López-Pérez, A. Marshall, and L. Hanzo, "Design of an efficient OFDMA-based multi-user key generation protocol," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 9, pp. 8842–8852, 2019.

[34] Z. Wei, B. Li, and W. Guo, "Adversarial reconfigurable intelligent surface against physical layer key generation," *IEEE Transactions on Information Forensics and Security*, 2023.

[35] L. Jiao, G. Sun, J. Le, and K. Zeng, "Machine learning-assisted wireless PHY key generation with reconfigurable intelligent surfaces," in *Proceedings of the 3rd ACM Workshop on Wireless Security and Machine Learning*, 2021, pp. 61–66.

[36] Y. Liu, M. Wang, J. Xu, S. Gong, D. T. Hoang, and D. Niyato, "Boosting secret key generation for IRS-assisted symbiotic radio communications," in *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*. IEEE, 2021, pp. 1–6.

[37] J. Tang, H. Wen, H.-H. Song, L. Jiao, and K. Zeng, "Sharing secrets via wireless broadcasting: A new efficient physical layer group secret key generation for multiple IoT devices," *IEEE Internet of Things Journal*, vol. 9, no. 16, pp. 15 228–15 239, 2022.

[38] Y. Xu, B. Gu, and D. Li, "Robust energy-efficient optimization for secure wireless-powered backscatter communications with a non-linear EH model," *IEEE Communications Letters*, vol. 25, no. 10, pp. 3209–3213, 2021.

[39] T. Kim, K. Min, and S. Park, "Self-interference channel training for full-duplex massive MIMO systems," *Sensors*, vol. 21, no. 9, p. 3250, 2021.

[40] Y. Jang, K. Min, S. Park, M. Jung, K. Ko, and S. Choi, "Dispersed signal transmission and reception scheme for full-duplex systems," *IEEE Access*, vol. 7, pp. 138 771–138 778, 2019.

[41] P. X. Nguyen, D.-H. Tran, O. Onireti, P. T. Tin, S. Q. Nguyen, S. Chatzinotas, and H. V. Poor, "Backscatter-assisted data offloading in OFDMA-based wireless-powered mobile edge computing for IoT networks," *IEEE Internet of Things Journal*, vol. 8, no. 11, pp. 9233–9243, 2021.

[42] N. Van Huynh, D. T. Hoang, X. Lu, D. Niyato, P. Wang, and D. I. Kim, "Ambient backscatter communications: A contemporary survey," *IEEE Communications surveys & tutorials*, vol. 20, no. 4, pp. 2889–2922, 2018.

[43] T. Gong, N. Shlezinger, S. S. Ioushua, M. Namer, Z. Yang, and Y. C. Eldar, "Rf chain reduction for MIMO systems: A hardware prototype," *IEEE Systems Journal*, vol. 14, no. 4, pp. 5296–5307, 2020.

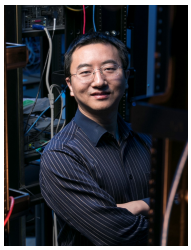
[44] P. Wang, Z. Yan, and K. Zeng, "BCAuth: Physical layer enhanced authentication and attack tracing for backscatter communications," in *IEEE Transactions on Information Forensics and Security*, 2022.
[45] X. Zhang, G. Li, J. Zhang, A. Hu, Z. Hou, and B. Xiao, "Deep-learning-based physical-layer secret key generation for fdd systems," *IEEE Internet of Things Journal*, vol. 9, no. 8, pp. 6081–6094, 2021.
[46] J. Li, P. Wang, L. Jiao, Z. Yan, K. Zeng, and Y. Yang, "Security analysis of triangle channel-based physical layer key generation in wireless backscatter communications," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 948–964, 2023.
[47] S. N. Premnath, S. Jana, J. Croft, P. L. Gowda, M. Clark, S. K. Kasera, N. Patwari, and S. V. Krishnamurthy, "Secret key extraction from wireless signal strength in real environments," *IEEE Transactions on Mobile Computing*, vol. 12, no. 5, pp. 917–930, 2013.
[48] Z. Wei, B. Li, and W. Guo, "Adversarial reconfigurable intelligent surface against physical layer key generation," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 2368–2381, 2023.
[49] R. Jin and K. Zeng, "Manipulative attack against physical layer key agreement and countermeasure," *IEEE Transactions on Dependable and Secure Computing*, 2019.
[50] S. Haykin, *Digital communications*. Wiley New York, 1988.
[51] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang, *MIMO-OFDM wireless communications with MATLAB*. John Wiley & Sons, 2010.
[52] G. Yang, Y.-C. Liang, R. Zhang, and Y. Pei, "Modulation in the air: Backscatter communication over ambient OFDM carrier," *IEEE Transactions on Communications*, vol. 66, no. 3, pp. 1219–1233, 2017.
[53] C. Boyer and S. Roy, "Space time coding for backscatter RFID," *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 2272–2280, 2013.
[54] M. Heino, D. Korpi, T. Huusari, E. Antonio-Rodriguez, S. Venkatasubramanian, T. Riihonen, L. Anttila, C. Icheln, K. Haneda, R. Wichman, and M. Valkama, "Recent advances in antenna design and interference cancellation algorithms for in-band full duplex relays," *IEEE Communications Magazine*, vol. 53, no. 5, pp. 91–101, 2015.
[55] J. Zhang, B. He, T. Q. Duong, and R. Woods, "On the key generation from correlated wireless channels," *IEEE Communications Letters*, vol. 21, no. 4, pp. 961–964, 2017.
[56] A. Sethi, V. Tapio, and M. Juntti, "Self-interference channel for full duplex transceivers," in *2014 IEEE wireless communications and networking conference (WCNC)*. IEEE, 2014, pp. 781–785.
[57] S. Mathur, W. Trappe, N. Mandayam, C. Ye, and A. Reznik, "Radio-telepathy: extracting a secret key from an unauthenticated wireless channel," in *Proceedings of the 14th ACM international conference on Mobile computing and networking*, 2008, pp. 128–139.
[58] S. Mathur, R. Miller, A. Varshavsky, W. Trappe, and N. Mandayam, "Proximate: proximity-based secure pairing using ambient wireless signals," in *Proceedings of the 9th international conference on Mobile systems, applications, and services*, 2011, pp. 211–224.
[59] G. Brassard and L. Salvail, "Secret-key reconciliation by public discussion," in *Workshop on the Theory and Application of of Cryptographic Techniques*. Springer, 1993, pp. 410–423.
[60] M. Bloch and J. Barros, *Physical-layer security: from information theory to security engineering*. Cambridge University Press, 2011.
[61] J. Zhang, T. Q. Duong, A. Marshall, and R. Woods, "Key generation from wireless channels: A review," *IEEE Access*, vol. 4, pp. 614–626, 2016.
[62] T. M. Cover and J. A. Thomas, "Elements of information theory. wiley, new-york," 2006.
[63] V. Liu, A. Parks, V. Talla, S. Gollakota, D. Wetherall, and J. R. Smith, "Ambient backscatter: Wireless communication out of thin air," *ACM SIGCOMM computer communication review*, vol. 43, no. 4, pp. 39–50, 2013.
[64] T. A. Feo and M. Khellaf, "A class of bounded approximation algorithms for graph partitioning," *Networks*, vol. 20, no. 2, pp. 181–195, 1990.
[65] Z. Fan, Y. Chen, J. Ma, and S. Zeng, "Erratum: A hybrid genetic algorithmic approach to the maximally diverse grouping problem," *Journal of the Operational Research Society*, vol. 62, no. 7, pp. 1423–1430, 2011.
[66] Y. Li, S. Liu, Z. Yan, and R. H. Deng, "Secure 5G positioning with truth discovery, attack detection and tracing," *IEEE Internet of Things Journal*, vol. PP, no. 99, pp. 1–10, 2021.
[67] J. Basak and R. Krishnapuram, "Interpretable hierarchical clustering by constructing an unsupervised decision tree," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 1, pp. 121–132, 2005.
[68] S. C. Johnson, "Hierarchical clustering schemes," *Psychometrika*, vol. 32, no. 3, pp. 241–254, 1967.
[69] L. Kaufman and P. J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons, 2009, vol. 344.
[70] S. J. Leon, L. De Pillis, and L. G. De Pillis, *Linear algebra with applications*. Pearson Prentice Hall Upper Saddle River, NJ, 2006.
[71] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009, vol. 2.
[72] K. Zeng, D. Wu, A. Chan, and P. Mohapatra, "Exploiting multiple-antenna diversity for shared secret key generation in wireless networks," in *2010 Proceedings IEEE INFOCOM*, 2010, pp. 1–9.
[73] W. C. Jakes and D. C. Cox, *Microwave mobile communications*. Wiley-IEEE press, 1994.
[74] M. Letafati, A. Kuhestani, H. Behroozi, and D. W. K. Ng, "Jamming-resilient frequency hopping-aided secure communication for internet-of-things in the presence of an untrusted relay," *IEEE Transactions on Wireless Communications*, vol. 19, no. 10, pp. 6771–6785, 2020.
[75] L. Jiménez Rodríguez, N. H. Tran, and T. Le-Ngoc, "Performance of full-duplex af relaying in the presence of residual self-interference," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 9, pp. 1752–1764, 2014.
[76] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Transactions on Wireless Communications*, vol. 11, no. 12, pp. 4296–4307, 2012.

Jiajun Li (Student Member, IEEE) received the bachelor degree in computer science from Xidian University in 2022. He is currently pursuing the master degree in cyberspace security Xidian University. His research interests are in wireless backscatter communication, physical layer security and key generation.

Pu Wang (Graduate Student Member, IEEE) received the Ph.D. degree in cyberspace security from Xidian University in 2021. His research interests are in backscatter communication, wireless information and power transfer, physical layer security, and information security in the Internet of Things.

Zheng Yan (Senior Member, IEEE) received the doctor of science in technology in electrical engineering from Helsinki University of Technology, Helsinki, Finland, in 2007. She is currently a Professor in the School of Cyber Engineering, Xidian University, Xi'an, China. Her research interests are in trust, security, privacy, and data analytics. Dr. Yan is an area editor or an associate Editor of IEEE INTERNET OF THINGS JOURNAL, Information Fusion, Information Sciences, IEEE NETWORK MAGAZINE, etc. She served as a General Chair or Program Chair for numerous international conferences, including IEEE TrustCom 2015 and IFIP Networking 2021. She is a Founding Steering Committee co-chair of IEEE Blockchain conference. Her recent achieved awards include 2021 N²Women: Stars in Computer Networking and Communications, Nokia Distinguished Inventor Award, Aalto ELEC Impact Award, the Best Journal Paper Award issued by IEEE Communication Society Technical Committee on Big Data and the Outstanding Associate Editor of 2017 and 2018 for IEEE Access.

Yishan Yang (Student Member, IEEE) received a master's degree in cyberspace security from Xidian University in 2021. She is currently pursuing the Ph.D degree with Xidian University. Her research interests include backscatter communication and physical layer security.



Kai Zeng (Member, IEEE) received the PhD degree in electrical and computer engineering from Worcester Polytechnic Institute (WPI), in 2008. He is an associate professor with the Department of Electrical and Computer Engineering, Department of Computer Science, and Center for Secure Information Systems, George Mason University. He was a postdoctoral scholar with the Department of Computer Science, University of California, Davis (UCD) from 2008 to 2011. He worked with the Department of Computer and Information Science, University of Michigan - Dearborn as an assistant professor from 2011 to 2014. He was a recipient of the U.S. National Science Foundation Faculty Early Career Development (CAREER) award in 2012. He won Excellence in Postdoctoral Research Award at UCD in 2011 and Sigma Xi Outstanding PhD Dissertation Award at WPI in 2008. He is an editor of the IEEE Transactions on Wireless Communications. His current research interests include cyber-physical system security and privacy, physical layer security, network forensics, and cognitive radio networks.

APPENDICES

APPENDIX A-CCA

Under the control channel attack (CCA), the main channel between A_i, A_j and RFS changes from h to $h+H$. We denote H_i as a controlled channel of channel h_i, and the absolute value $|H_i|$ is the controlled channel strength (CCS).Since the maximum value of CCS is related to the communication distance, the maximum CCS is distinct for different channels [46], [51]. And the CCS ratio in Fig. 7,8 is unified as $\frac{H}{H_{max}}$ (ranged from 0 to 1).Considering that there is no noise, it is easy to understand the impact of CCA on CGKG, DGKG and DHGKG.

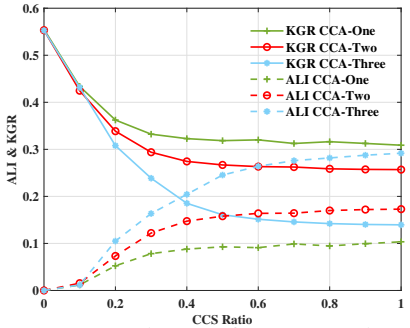


Fig. 7: KGR and ALI of CGKG under CCA.

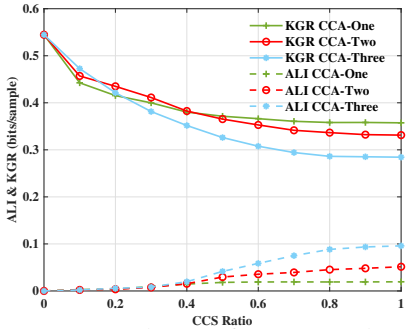


Fig. 8: KGR and ALI of DGKG under CCA.

A. CGKG

When under CCA, the round-trip channel between A_1 and RFS changes, which leads to the fact that the round-trip channel measurement R_i of A_i contains the controlled channel. Therefore, β_i and R_i an be expressed as follow:

$$\beta_i = v^i_{RFS} - v^1_{RFS} = \alpha(h_i + H_i)(h^b_i + H_i) - \alpha(h_1 + H_1)(h^b_1 + H_1),$$

$$R_i = \{\alpha(h_1 + H_1)(h^b_1 + H_1), ..., \alpha(h_N + H_N)(h^b_N + H_N)\},$$

if Mallory only wants to attack channel h_i, other controlled channels except for H_i (like H_j, $j \neq i$) are zero.

According to the expression of R_i and validation of the simulation result in Fig. 7, the correlation between the generated group key and the controlled channel rises with the increase of the controlled number of downlink channels in the system. Therefore, when the number of BDs is small, since the number of downlink channels is reduced, the resistance of CGKG on CCA is reduced. Fig. 7 shows that when Mallory controls three downlink channels simultaneously, the KGR eventually stops decreasing and stabilizes when the CCS ratio reaches 0.7. Therefore, CGKG is resistant to CCA when Mallory does not control all downlink channels.

B. DGKG & DHGKG

Since DGKG and DHGKG work similarly, we can analyze their resistance on CCA together. When under CCA, the triangle channel measurement between A_i and A_j can be expressed as follows:

$$T_{i,j} = [(h_{i,j} + H_{i,j})(h_i + H_j)] \cdot (h_i + H_j),$$

if Mallory only wants to attack channel h_i, other controlled channels except for H_i (like H_j and $H_{i,j}$) are zero.

In Tri-Channel [46], there are two downlink channels and one inward channel. If Mallory wants to compromise additional bits of keys when it increases CCS, it needs to control all three channels simultaneously In DGKG, there are N downlink channels and $N(N-1)/2$ inward channels. Therefore, although Mallory controls all three channels in one triangle channel simultaneously, there are $N(N+1)/2-3$ channels it does not know. Therefore, DGKG increase the resistance under CCA, compared with Tri-Channel. However, compared with DGKG, in DHGKG, one group is divided into multiple sub-groups. Therefore, DHGKG reduces the number of channels available for key generation in the system, thus weakening its resistance to CCA. Fig. 8 shows the KGR and ALI of DGKG under CCA and SMA. Its KGR and ALI becomes stable when CCS ratio nearly reaches 0.8, which implies that Mallory cannot compromise additional bits of keys when it increases CCS.

APPENDIX B-SMA

Under the signal manipulative attack (SMA), Mallory injects similar signals to those of the two BDs and is engaged in agreeing on some valid but manipulated key bits. When Mallory imposes SMA, two legitimate devices (RFS and A_i in CGKG, A_i and A_j in DGKG and DHGKG) receive a manipulated signal $P(t)$ from Mallory.

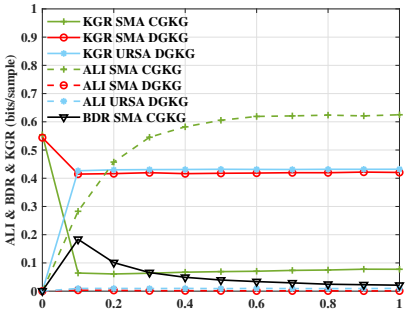


Fig. 9: KGR, ALI and BDR of CGKG and DGKG under SMA and URSA.

C. CGKG

The received signal at RFS and A_i in CGKG can be expressed as:

$$y_i = h_i s + h_{m,i} P,$$

$$y^i_{RFS} = \alpha_i h_i h^b_i s + h_{m,RFS} P,$$

where $h_{m,i}$ and $h_{m,RFS}$ are the attack channel of Mallory between A_i and RFS, respectively and Mallory can estimate the channels by utilizing the backscatter characteristic of BD before key generation. From the definition of SMA [46], [49] , we have $h_{m,i} \approx h_{m,RFS}$ and for simplicity, we use $I = h_{m,RFS}P = h_{m,i}P$ to denote the manipulated signal. Then, the constructed round-trip channel information of A_i and RFS can be expressed as:

$$y^{'}_i = \alpha(h_i s + I)(h_i s + I),$$

$$y^{'i}_{RFS} = \alpha_i h_i h^b_i s^2 + I.$$

Obviously, $y^{'}_i \neq y^{'i}_{RFS}$ and the round-trip channel information constructed by RFS and A_i cannot act as a shared

randomness for RFS and A_i anymore. Therefore, when CGKG is suffering from SMA, it turns to jamming attack (JA). From Fig. 9 we can observe that when SMA is just occurring, that is when the MSS reaches 0.1, BDR rises and KGR drops significantly. This shows a JA on the system, thus validating our analysis.

D. DGKG & DHGKG

Since DGKG and DHGKG work similarly, we can analyze their resistance on SMA together. When under SMA in DGKG, the triangle channel measurement between A_i and A_j can be expressed as follows:

$$T_{i,j} = [\alpha h_{i,j}(h_j + I)] \cdot (h_i + I) = \alpha[h_i h_j h_{i,j} + I(h_i h_{i,j} + h_j h_{i,j}) + I^2 h_{i,j}].$$

From the above expression of $T_{i,j}$, since Mallory does not know h_i, h_j and $h_{i,j}$, $T_{i,j}$ is uncorrelated with I. Therefore, even though the signal strength (MSS) of $P(t)$ is manipulated, $T_{i,j}$ remains uncorrelated with I and Mallory cannot manipulate additional bits of keys when it increases MSS. Fig. 9 shows that ALI and KGR eventually becomes steady after a slight variation, which validates our analysis.

When the number of BDs in a group becomes smaller, the proportion of triangle channel information containing the manipulated signal to all triangle channel information rises. As a result, the correlation of the group key with the manipulated signal elevates. Therefore, DHGKG reduces the number of BDs in a sub-group, which weakens its resistance to SMA in the sub-group. However, since the group key is generated from all sub-group keys through XOR operations and Mallory does not have the group key information of any other sub-groups. Therefore, the correlation between the manipulated signal and the group key is reduced in DHGKG compared with DGKG.

Appendix C-URSA

Due to the hardness of BD authentication on RFS, it is possible for Mallory to disguise itself as a fake RFS and transmit a signal to BDs to manipulate group key generation. In this circumstance, CGKG can not work any more. Herein, we focus on analyzing the impact of URSA on DGKG and DHGKG.

E. DGKG & DHGKG

When under URSA, the triangle channel measurement between A_i and A_j can be expressed as follows:

$$T_{i,j} = (\alpha h_{i,j} h_{m,j} P) \cdot (h_{m,i} P) = \alpha h_{i,j} h_{m,j} h_{m,i} P^2.$$

Similar to SMA, $h_{m,i}$ and $h_{m,j}$ are the attack channel of Mallory between A_i and A_j, respectively. From the above expression of $T_{i,j}$, even though Mallory knows $h_{m,i}$, $h_{m,j}$ and P, it still dose not know $h_{i,j}$. And therefore, $T_{i,j}$ is uncorrelated with $h_{m,i}P$,$h_{m,j}P$ or $h_{m,i}h_{m,j}P^2$. Fig. 9 shows that ALI and KGR eventually becomes steady after a slight variation in DGKG, which validates our analysis.

And if the untrusted RFS ensures the regular operation of the DHGKG scheme, the situation of DHGKG when under URSA becomes similar to DGKG. Unlike SMA, when under URSA, the untrusted RFS can get to know all downlink channel information of all sub-groups. Therefore, all the sub-group keys contain the manipulated signal. Compared with DGKG, DHGKG reduces the number of inward channels available for key generation, which weakens its resistance to URSA.

Appendix D-Static Environment

To introduce additional randomness to the system, BDs can use a time-variant backscatter coefficient at the backscatter phase. BDs possess the capability to perform amplitude or phase modulation on the downlink channel signals during the process of backscattering. Additionally, we denote $\alpha_i(t)$ and $\alpha_j(t)$ as the time-variant (random) amplitude backscatter coefficient of BD A_i and A_j. We denote $\phi_i(t)$ and $\phi_j(t)$ as the time-variant (random) phase shifting coefficient matrix of BD A_i and A_j. Therefore, in CGKG, the round-trip channel information r measured by A_i in two different time slow t_1 and t_2 can be expressed as follows:

$$r_{t_1} = \alpha_i(t_1) \cdot h_i \phi_i(t_1) h_i \quad (35a)$$

$$r_{t_2} = \alpha_i(t_2) \cdot h_i \phi_i(t_2) h_i \quad (35b)$$

By leveraging BDs' time-variant phase shift coefficient matrix and amplitude backscatter coefficient, we can create additional randomness in the round-trip channels. The extracted triangle channel information by BDs can be expressed as:

$$T_{t_1} = \alpha_i(t_1) \cdot h_i h_j \phi_i(t_1) h_{i,j} \quad (36a)$$

$$T_{t_2} = \alpha_i(t_2) \cdot h_i h_j \phi_i(t_2) h_{i,j} \quad (36b)$$

Since RFS is reliable and trusted in CGKG, we still recommend utilizing RFS to create an artificial multi-path effect in CGKG to provide additional randomness in static environments. And in DGKG and DHGKG, we recommend to use time-variant amplitude backscatter coefficient and time-variant phase shifting coefficient matrix together to introduce extra randomness to the static channel.

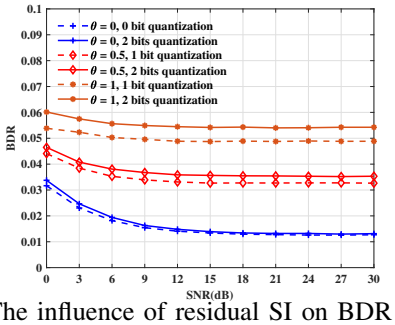


Fig. 10: The influence of residual SI on BDR in CGKG.

Appendix E-Residual SI

Both CGKG and DGKG require a full-duplex RFS in the system. In CGKG, the residual SI affects the symmetric of channel and corrupts the consistency of round-trip channel information measured by RFS and BDs. While in DGKG, residual SI does not influence the consistency of triangle channel measurements among BDs but hinders the accurate estimation of the location of BDs by RFS. Since reducing the deterioration of RSI for estimating an accurate RSS is not our focus, we only discuss RSI in CGKG. Herein, we assume that the residual SI is zero-mean and additive, obeying Gaussian distribution similar [75], [76]. Melissa et al. [76] first reported a statistical characterization of the self-interference based on extensive measurements. Leonardo et al. [75] summarized the experiments conducted by Melissa et al. and modeled the variance of residual SI as $P = \gamma P_r^{\theta}$, where P_r is the average power transmitted by the RFS, θ ($0 \leq \theta \leq 1$) is constant, which represents the quality of cancellation technique, γ is a correction factor that guarantees the unit of V is still the unit of power (W,Watts). Therefore, we focus on the effect of different cancellation quality θ on the performance of CGKG in this paper. Fig. 10 shows the influence of residual SI on BDR in CGKG under different settings. When $\theta = 1$, a pessimistic condition in which variance increases linearly with transmitted power, a higher quantization level also increases BDR since more information in the key measurements is considered and amplifies the influence of residual SI on key consistency. Therefore, using a low quantization level when the cancellation quality is poor can reduce the error caused by residual SI. When $\theta = 0$, an optimistic circumstance in which the variance is a constant, the impact of residual SI on CGKG becomes relatively small.