IEEE Open Journal of the Computer Society

Received 15 July 2020; accepted 19 September 2020. Date of publication 5 October 2020;
date of current version 15 October 2020. The review of this paper was arranged by Associate Editor Dr. S. Guo and Guest Editors Zhou Su, Abderrahim Benslimane, David (Bong Jun) Choi, Jiliang Li, and Dongfeng Fang.

Digital Object Identifier 10.1109/OJCS.2020.3027014

Concatenated Graph Coding on Bandwidth Part for Secure Pilot Authentication in Grant-Free URLLC

DONGYANG XU [1,2] (Member, IEEE), AND PINYI REN [1,2] (Member, IEEE)

[1] School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China
[2] Shaanxi Smart Networks and Ubiquitous Access Research Center, Xi'an, Shaanxi 710049, China

CORRESPONDING AUTHOR: PINYI REN (e-mail: pyren@mail.xjtu.edu.cn)

The research work reported in this paper is supported in part by the National Natural Science Foundation of China under the Grants No. 61941119 and No. 62001368, and Fundamental Research Funds for the Central Universities, China.

ABSTRACT Grant-free multiple access is a critical mechanism introduced in 5G new radio (NR) to support ultra-reliable low-latency communication (URLLC) services. Pilot authentication (PA) is a key security mechanism to guarantee reliable performance of grant-free URLLC. However, PA can be easily paralyzed by pilot-aware attack since pilot signals are usually publicly known, and unprotected. To solve this, we develop the concatenated graph coding (CGC) theory by which time-frequency resources on bandwidth part (BWP) can be encoded flexibly to protect PA securely. Particularly, we use bipartite graph, and multigraph theory to model PA on BWP as transmission, and retrieval of pilot (TRP). Each transmitter in the uplink needs transmit a unique random pilot sequence as subcarrier activation pattern (SAP) on BWP. After observing SAPs from multiple transmitters, the receiver decodes a pilot sequence of interest, and tests its authenticity. The retrievability of authentic pilots is defined, and formulated analytically. We also derive the analytical closed-form expression of system failure probability, and accessibility in the regime of large-scale antenna arrays, and short data packets. Interestingly, we find that four trade-offs exist: retrievability-latency, retrievability-accessibility, reliability-latency, and reliability-accessibility. Simulation results show the security advantage of our proposed theory in grant-free URLLC system.

INDEX TERMS Physical-layer authentication, grant-free URLLC, pilot authentication, pilot-aware attack, security.

I. INTRODUCTION

5G networks are expected to support mission-critical applications demanding ultra-reliable and low-latency communications (URLLC), like industrial automation, remote control, health monitoring and tactile Internet [1], [2]. This vision requires end-to-end (E2E) transmission with ultra-low latency (e.g., ranging from 1 ms to few milliseconds), as well as ultra-high reliability (e.g., 99.999%) [3]. To fulfill that, 5G new radio (NR) introduced grant-free multiple access (GFMA) technology such that transmission without grant avoids the regular handshake delay e.g. sending the scheduling request (SR) and waiting for UL scheduling grant (SG) allocation [4], [5]. In an uplink scenario of grant-free URLLC, user equipment (UE) can transmit data in a pre-configured bandwidth part (BWP) that comprises of time, frequency and pilots, and data transmission is performed without the request/grant procedure. The next generation NodeB (gNB) in GFMA mode needs to detect active UEs using pilots, perform pilot-based channel estimation for those UEs, and decode their data based on estimated channels. Due to the lack of SR and SG, all of these functionalities couple with each other and needs to be accomplished in one-shot access [6]. The pilots or reference symbols become the very important messages that connect them well [7].

If we examine the information security in 5G NR, message authentication in grant-free URLLC plays a critical role since it guarantees the integrity, authenticity, and non-repudiation of messages that flow over the air [8], [9]. However, not

every messages generated or required in current 5G have been protected by reliable message authentication whereas those messages become increasingly important [4], [10], [11]. Signaling radio bearer 0 (SRB0) and pilot sequences are the most typical ones. In 5G NR protocol, SRB0 is a very critical medium distributed between Layer 3 (L3) and Layer 2 (L2) to inform Layer 1 (L1) the control information from L3 [12]. SRB0 carries the pilot-related signaling, such as, the time/frequency location information of pilot sequences on bandwidth part (BWP). However, 5G NR protocol does not confer cryptographic protection on SRB0, which leaves pilot-related information at Ll completely exposed to the public. At L1, pilot sequences are usually stored as a form of database which is deterministic and publicly-known to all parties [13]. In grant-free URLLC, pilot sequence of interest is uniquely selected by one certain UE and then inserted on the BWP physical resources for distinguishing among different UEs and estimating the channel occupied by itself [6]. Non-repeating pilot sequences guarantee the desirable user activity detection (UAD) and channel estimation. With these in preparation, data demodulation can work normally. Therefore in current 5G NR, gNB and UEs have a very clear agreement on the allocation of pilot sequences for each UE and this agreement is also publicly known. Obviously, this setup is a very easy and fragile message authentication mechanism since the authentication performance is guaranteed by assuming that everyone would obey the agreement.

By virtue of those vulnerabilities, pilot-aware attacker can blindly decode the disclosed information, e.g., SRB0 and pilots, and paralyze the agreement in the form of jamming, spoofing and nulling those pilots [14], [15]. This refers to the concept of pilot authentication (PA) which aims to achieve the transmission and retrieval of pilot (TRP) reliably against those attack behaviors. TRP is a complete process from detecting disturbance to authentic maintenance, for example, how to detect any alteration to pilot authenticities and how to protect and further maintain high authenticities. Without precise and secure TRP makes UAD and channel estimation hard to work normally in grant-free URLLC. The challenge in this area embraces how to design and secure TRP under the circumstance of pilot aware attack to accommodate grant-free URLLC services.

A. RELATED WORKS

Protecting PA by encrypting SRB0 or directly pilots under cryptographic framework would cause significant signaling overheads that cannot be endured by grant-free URLLC. Physical layer authentication (PLS) using L1 information is a good choice since it could remove the time-consuming functionalities of security protection for SRB0 and merely focus on inherent security advantages at L1 to bring lightweight overheads and desirable security performance [11]. Authors in [16] have analyzed the advantages of using physical layer information to protect URLLC. Previously, lots of research of PLS in [17], [18] have been done by employing channel state information (CSI) which is usually estimated using pilots. When pilot signals are attacked, those schemes can be destroyed and safeguarding pilots using PLS is an available choice. There are two PLS based research routes to protect pilots, including deterministic-pilot based approach and random-pilot based approach. The former is limited to how to detect pilot-aware attack by exploiting the physical layer information, such as auxiliary training or data sequences [19], [20] and some prior known channel information [21]–[23] whereas the latter aims to avoid it by pilot randomization [15], [24], [25]. The importance of pilot randomization on time-frequency resources to avoid pilot aware attack on pilot subcarriers have been stressed in orthogonal frequency-division multiplexing (OFDM) systems [26]. With pilot randomization, a random pilot sequence x is transmitted from a UE to gNB through a subcarrier channel h and the gNB would observe a signal with the form of $y = hx + gz + w$ where g and z respectively denote the subcarrier channel occupied by attacker and the interfering signal. w represents the noise variable. As we see, since x is random, z should be also random or just zero for attacker. This eliminates the possibilities of spoofing ($z = x$) [27], nulling ($z = -h/gx$) [28] and jamming [28]. However, the gNB cannot recognize x due to the randomness of channels and signals. The challenge for PA becomes a process of designing TRP under pilot randomization.

Recent studies in [29]–[31] have shown the possibilities of transmitting extra messages using subcarrier activation patterns (SAPs), parallel with the normal data transmission on subcarrier channels. Authors in [32] also concluded that the technique of encoding SAPs to carry information in parallel with data transmission can improve the system reliability with less power and complexity, making it a very suitable candidate for 5G URLLC service. This technique was also studied in grant-free URLLC systems in [16]. Those schemes provide a road access to TRP using SAPs. In this vision, UEs can transmit a set of non-repeating pilot sequences as various types of SAPs for pilot transmission. Then in order to retrieve pilots, next generation NodeB (gNB) decodes a pilot sequence of interest according to the observed SAPs and tests whether or not it is a right member of the stored set. However, applying SAPs to TRP directly cannot maintain resilience against the disturbance on subcarriers. When the attacker launches jamming attacks on several subcarriers, the retrieval of SAPs as pilots would be like a random phenomenon and thus no use for gNB.

To solve this, authors in [33] proposed a coding based TRP framework by exploiting pilot randomization and subcarrier-block discriminating coding (SBDC) mechanism. In [15], the authors considered a practical one-ring scattering scenario and proposed an independence-checking coding (ICC) theory to secure TRP under pilot aware attack. When pilot aware attack happens, TRP under those schemes highly rely on the coding diversity and spatial fading correlation difference between UE and the attacker. By relaxing the dependence, authors in [25] proposed a scheme of hierarchical 2-D feature coding (H2DF) coding to encode SAPs to realize TRP reliably.

Those schemes help answer the question of how to design TRP between UEs and gNB under pilot randomization and pilot-aware attack. Under this background, TRP is equivalent to a three-step process, including pilot conveying, separation and identification. For pilot conveying, OFDM subcarriers are selectively activated to create various SAP candidates which are encoded such that those SAPs can carry pilot information in the form of codewords. Those codewords should be optimized such that those codewords, though overlapped with each other and/or even disturbed by Ava, can be separated and identified with high reliability, thus decoded into the original pilots.

However, all of those schemes are expected to work well in grant-based OFDM systems. Our previous conference paper theoretically examined the possibility of applying H2DF coding to achieve TRP in grant-free URLLC system [34]. However, the research ignored practical physical resource limitations on BWP under 5G NR protocol. Redesigning TRP on BWP is a rather different thing for grant free URLLC. Due to the limited capability of radio frequency (RF) chains at UE, the size of operational bandwidth are limited [35], [36], usually at most 100 MHz channel bandwidth for Frequency Range 1 (FR1) and 400 MHz channel bandwidth for Frequency Range 2 (FR2) in 5G. Considering subcarrier spacing of 60 kHz and 240 kHz respectively, at most 512 pilot subcarriers are supported across the full channel bandwidth given a 5G pilot configuration. This is because one regular pilot subcarrier exists every three subcarriers within channel coherence bandwidth to capture the expected frequency-domain variations of the channel to be resolved [13]. With this constraint, the probability of wrongly identifying authentic pilots under H2DF coding can achieve at most $(1+K)$ $(512\sqrt{2K})$ where K denotes the number of access users. It means that the overall reliability of grant-free URLLC is at the level of 10^{-3} in any case as the lower bound of reliability is determined by PA [37]. Besides this, the excessive consumption on one-dimensional resources (i.e., the full occupation of bandwidth) severely limits the flexibility of BWP operations and increases the burden of RF chains of UEs. When bandwidth is reduced, the reliability becomes much lower as well.

B. MOTIVATIONS AND CONTRIBUTIONS

As we can see, on one hand, current PA schemes rely on a SAP-encoded methodology for which "multidimensional resources, parallelism and coding" become their keywords. On the other hand, the allowable resources might be limited and cannot provide a suitable basis for the method. This make us to rethink: *whether or not it is possible to encode grant-free time-frequency physical resources to protect pilot authentication of multiple UEs operating flexibly in BWP with very high reliability and low latency*. We can show the answer is yes and propose a concatenated graph coding (CGC) theory together with comprehensive performance evaluation of it. The details of our contributions are shown as follows:

- *CGC Theory:* We formulate a basic and unified framework of CGC theory to describe coding operations on time-frequency physical resources. In this theory, pilot signals could be modeled as binary codes on bipartite graph, an abstract representation of time-frequency subcarriers. The transmission of pilots from multiple UEs is modeled as the superposition process of multiple binary codes on bipartite graphs. The result constitutes an undirected, completely labeled bipartite (UCLB) multigraph at gNB. We formulate the process of bipartite graph query of gNB on multiple bipartite graphs based on the UCLB multigraph and show how TRP can be modeled as the process of bipartite graph query to realize PA on BWP in grant-free URLLC.
- *CGC on BWP:* We provide a closed-form analytical solution for CGC on BWP and define the retrievability of authentic pilots during TRP. This requires an optimization of the code matrix of CGC with the aim of keeping CGC with high coding diversity. The introduction of CGC on BWP makes possible the redesign of UAD on code domain and therefore we propose a novel UAD algorithm whereby identification of attack modes and UAD can be realized both. With TRP and UAD done, we can thus provide an entire process from attack detection to anti-attack countermeasures while keeping pilots authentic in CGC based grant-free URLLC system.
- *Performance Evaluation and Trade-offs:* In order to characterize the reliability performance of CGC based grant-free URLLC systems, we derive a novel expression of system failure probability under short packet transmission in the regime of large-scale antenna arrays and a matched filter receiver. We also define the α-accessibility of this system as the ratio of the number of active legitimate UEs that could maintain the failure probability less than α to the total number of OFDM symbols consumed during this period. Finally, we prove and characterize four types of trade-offs, including retrievability-latency, retrievability-accessibility, reliability-latency and reliability-accessibility trade-offs. Those results prove the possibility of CGC based grant-free URLLC system.

Organization: In Section II, we introduce the basic system model. The theory of CGC is described in Section III. In Section IV, we show how to make CGC operate on BWP. Performance evaluation and trade-offs of CGC based grant-free URLLC is provided in Section V. Numerical results are presented in Section VI and finally we conclude our work in Section VII.

Notations: We use boldface capital letters $\mathbf{A}$ for matrices, boldface small letters $\mathbf{a}$ for vectors, and small letters a for scalars. $\mathbf{A}^H$ denotes the conjugate transpose of matrix $\mathbf{A}$. $|\mathcal{A}|$ is the cardinality of the set $\mathcal{A}$.

II. SYSTEM MODEL

A. BWP FOR GRANT-FREE URLLC

BWP has been introduced for 5G-NR to provide a means of operating UEs with desirable bandwidth that matches with

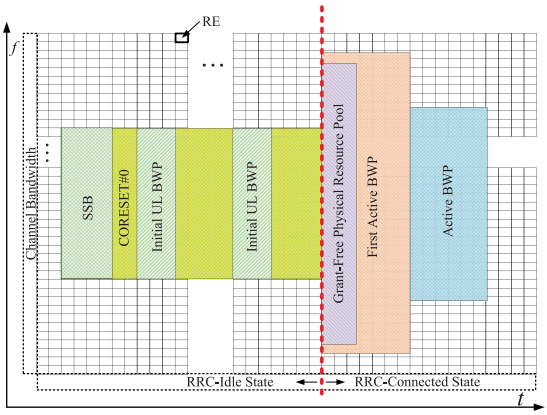

FIGURE 1. Illustration of different types of BWPs in 5G NR.

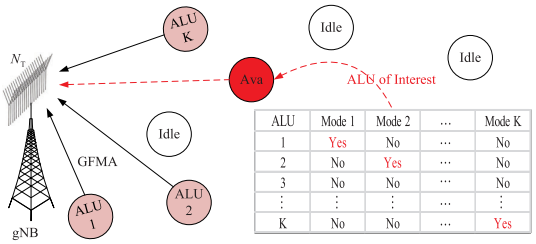

FIGURE 2. Illustration of a hybrid attack on GFMA with random piots.

their RF chain ability [36]. As shown in Fig. 1, a BWP is a contiguous set of physical resource elements (REs) for a given numerology within channel bandwidth. Each BWP defined for a numerology can have three different parameters, including subcarrier spacing, symbol duration and cyclic prefix (CP) length. BWP has many types, depending on the specific scenarios. There is an initial active BWP for a UE during the initial access under radio resource control (RRC)-idle state until the UE is explicitly configured with active BWPs during or after RRC connection establishment. UEs are expected to receive and transmit only within the frequency range (FR) configured for the active BWPs with the associated numerologies.

Fig. 1 shows a basic configuration of various BWPs required from RRC-idle state to RRC-connected state. In RRC-idle state, the initial access occurs after synchronization signal block (SSB) and occupies two types of BWPs, i.e., CORESET#0 serving as an initial DL BWP, and an initial UL active BWP. The configuration information of SSB and those BWPs on the whole channel bandwidth can be decoded in SRB0 in high layer. Those information include but not limited to the location and size of SSB and BWPs. In RRC-connected state, grant-free URLLC occurs and occupies the *grant-free physical resource pool* within the first active BWP. The configuration of the *grant-free physical resource pool* is done within the initial access.

B. SYSTEM DESCRIPTION

As shown in Fig. 2, an uplink single-input multiple-output (SIMO)-OFDM based GFMA system works within a *grant-free physical resource pool*. The system includes a N_{T}-antenna

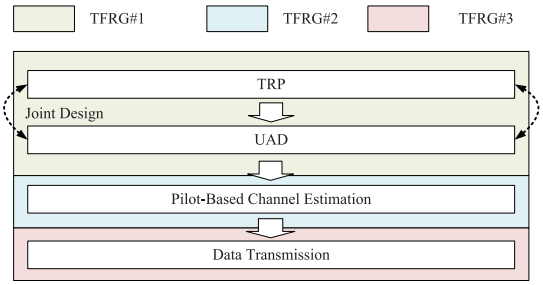

FIGURE 3. A general description of a complete process of grant-free URLLC under a hybrid attack.

gNB named Alice, and G single-antenna legitimate UEs (LUs) indexed by the set $\mathcal{G}$ with $|\mathcal{G}| = G$. LUs consist of $G-K$ non-active LUs (NLUs) and K active LUs (ALUs) indexed by set $\mathcal{K}$ with $|\mathcal{K}| = K$. Upon uplink access, each of ALUs transmits random pilots and data of interest. Alice needs to identify the random pilots from ALUs and this refers to the process of TRP. The model of random pilots and TRP can be respectively seen in Subsection II-C and Subsection II-D. At Alice, TRP and UAD are jointly design to identify ALUs and pilots. Then Alice can estimate multiuser channels based on the identified ALUs and pilots. With these, Alice finally decoded the multiuser data. The relationship among them can be seen in Fig. 3. TRP and UAD work on time-frequency resource grid (TFRG)#1 defined by the set Ψ_{E} with $|\Psi_{\mathrm{E}}| = N_{\mathrm{E}}$. Channel estimation operates on TFRG#2 defined by Ψ with $|\Psi| = \overline{N}_{\mathrm{E}}$. Data transmission works on TFRG#3 defined by the set Ψ_{D} with $|\Psi_{\mathrm{D}}| = N_{\mathrm{D}}$. A single-antenna pilot-aware attacker Ava would like to launch a hybrid attack on the above TRP process. In RRC-connected state, Ava can synchronize with ALU precisely since SRB0 gets no protection in 5G NR and the synchronization information over the air can be decoded by Ava within initial access. The specific model of attack modes can be seen in Subsection II-E.

Each of subcarrier channels from ALUs and Ava to Alice experience the frequency-selective Rayleigh fading. Frequency-domain subcarrier spacing is configured as Δf. Let T_s denote one OFDM symbol time. Δf and T_s both follow the choices in 5G NR numerologies [7]. The total latency T during uplink access satisfies:

$$T_{\mathrm{con}} \geq T = (m_{\mathrm{E}} + m_{\mathrm{D}}) \times T_s + T_{\mathrm{extra}} \tag{1}$$

where T_{extra} denotes the time for operations other than channel estimation and data transmission. $T_{\mathrm{con}} = 1 \times 10^{-3} s$.

In the following subsections, we will detail the models of random pilots, TRP, channel estimation and data transmission.

C. RANDOM PILOT SIGNAL MODEL

We denote $x^i_{\mathrm{L},m}[k]$ and $x^i_{\mathrm{A}}[k]$ respectively as the pilot values for the m-th ALU and Ava at the i-th subcarrier and k-th symbol time. Pilots across subcarriers and OFDM symbols obey the following principle, i.e., $x^i_{\mathrm{L},m}[k] = x_{\mathrm{L},m}[k] = \sqrt{\rho_{\mathrm{L},m}} e^{j\phi_{k,m}}, \forall i, i \in \Psi_{\mathrm{E}} \cup \Psi, m \in \mathcal{G}, \phi_{k,m} \in \mathcal{A}$. $\mathcal{A}$ satisfies $\{\phi : \phi = 2m\pi \ \ C, 0 \leq m \leq C-1, C = |\mathcal{A}|\}$. $x_{\mathrm{L},m}[k]$

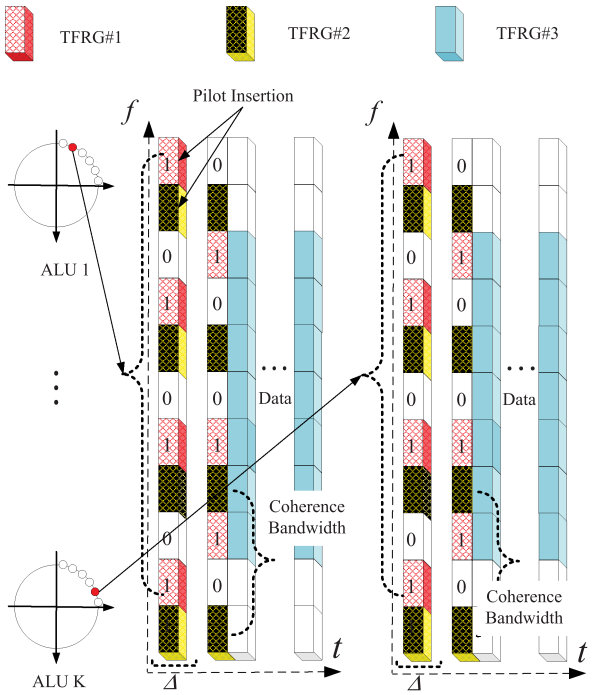

FIGURE 4. Illustration of how to encode pilots as SAPs with different TFRG resources.

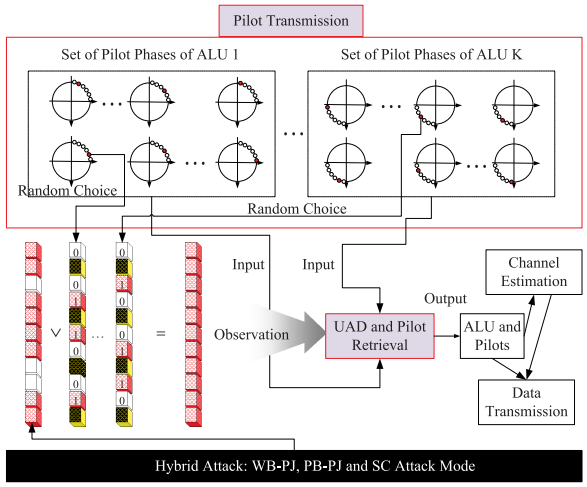

FIGURE 5. Illustration of TRP in the form of encoded SAPs under the hybrid attack.

can be superimposed onto a dedicated pilot sequence having been optimized under a non-security oriented scenario. The new pilot sequence is then utilized for channel estimation. At this point, $\phi_{k,m}$ is an additional phase difference for security consideration. We do not impose constraints on the values of signals inserted by Ava, that is, $x_{\mathrm{A}}^{i}[k] = \sqrt{\rho_{\mathrm{A}}} e^{j\varphi_{k,i}}, i \in \Psi_{\mathrm{E}}$ where $\varphi_{k,i}$ is the random pilot phase of Ava and ρ_{A} denotes the jamming power.

D. MODEL OF TRP

1) MULTI-DIMENSIONAL TIME-FREQUENCY CODING

In order to support flexible BWP under limited RF bandwidth, the system performs multi-dimensional time-frequency coding on BWP. As shown in Fig. 4, Δ OFDM symbols and one pilot subcarrier constitute a basic *resource block*. The optimization of Δ is shown in following sections. In frequency domain, the system, doing as NR protocol specifies, keeps one regular pilot subcarrier every three subcarriers and encode other pilot subcarriers within coherence bandwidth. The spacing configuration is related to the expected coherence bandwidth of the channel, which is in turn related to the channel delay spread and allows the expected frequency-domain variations of the channel to be resolved [13].

2) BASIC PROCEDURES OF TRP

The first thing of TRP is to identify the specific pilot information to be authenticated. The commonly-used is a set of discrete pilot phase candidates, each of which is mapped by default into a unique quantized sample, chosen from the set $\mathcal{A}$. There should be a total of C pilot phase candidates mapped as C elements which can be mapped into C unique codewords. Those codewords constitute a codebook, denoted by a matrix $\mathbf{B}$. Then, the set $\mathcal{A}$ is divided into G subsets of same size $C\ \ G$. Each subset is allocated to a unique LU and corresponds to a unique matrix $\mathbf{B}_{1\le t\le K}$ with $\mathbf{B} = [\,\mathbf{B}_1\ \cdots\ \mathbf{B}_K\,]$. The parameter of C will be optimized later. For pilot transmission, each of ALUs, i.e., the m-th ALU, chooses randomly from its own subset one preferred pilot phase as $\phi_{k,m}$ which is then expressed as a binary codeword denoted by $\mathbf{b}_{j,m}, m \in \mathcal{K}$. For each ALU, the value of j can be any column index of $\mathbf{B}_t$ and represents the random choice. The specific principle of generating SAPs via codeword is that if the i-th digit of the preferred codeword is equal to 1, the pilot signal is inserted on the i-th *resource block* and otherwise this *resource block* will be idle.

As shown in Fig. 5, when pilot transmission proceeds from K ALUs, their pilot phases would be independently expressed in the form of SAPs. Multiple SAPs from multiple ALUs, after undergoing wireless channels, suffer from the superposition interference from each other and the attacker. This superposition is observed by Alice and then decoded as codeword $\mathbf{b}_{\mathrm{I}}$. This rely on the signal energy detection technique. Authors in [38] have provided a function $\gamma(P_f) \triangleq f(N_{\mathrm{T}}, K, P_f)$ for determining on one subcarrier the number of antennas required to achieve a probability P_f of false alarm detection. Alice could always flexibly configure $\gamma(\varepsilon^*)$ to make ε^* approach zero [25] and precisely detect the number of signals on the j-th subcarrier as N_j. Then the binary digit corresponding to the j-th pilot subcarrier is "1" when $N_j \ge 1$ and otherwise when $N_j = 0$, the binary digit is equal to 0. In this way, $\mathbf{b}_{\mathrm{I}}$ can be formulated and further decoded with tolerance of errors as $\mathbf{b}_{j,m}, m \in \mathcal{K}$. This can be seen as a process of pilot retrieval. The above constitutes a complete process of TRP.

E. ATTACK MODEL

Due to the randomness of pilots and SAPs, Ava launches a hybrid attack on SAPs, including wide-band pilot jamming (WB-PJ), partial-band pilot jamming (PB-PJ) and silence cheating (SC) attack [14], [25]. In WB-PJ attack, Ava activates the whole available subcarriers. In PB-PJ attack, Ava arbitrarily activates part of the subcarriers and Ava in SC attack keeps silence to misguide Alice since Alice cannot recognize the non-existence of attacks.

With this attack model, the attack on SAPs observed by Alice can be modeled using the following equations:

$$\mathbf{b}_{j,1} \bigvee \cdots \bigvee \mathbf{b}_{j,K} = \mathbf{b}_{\mathrm{S},K}, \mathbf{b}_{\mathrm{S},K} \bigvee \mathbf{c} = \mathbf{b}_{\mathrm{I}} \tag{2}$$

where $\mathbf{c}$ denotes the attack behaviors on each subcarrier and can be an arbitrary binary codeword with each element in the set $\{0, 1\}$. $\bigvee$ denotes the digit-by-digit Boolean sum operation.

F. RECEIVING SIGNAL MODEL FOR CHANNEL ESTIMATION

As shown in Fig. 4, the pilot tone insertion patten of ALUs and Ava for channel estimation both follow block type. Random pilots are inserted on subcarriers for channel estimation. Consider the basic OFDM procedure. Pilot tone vectors of ALUs and Ava over $\overline{N}_{\mathrm{E}}$ subcarriers are respectively stacked as $\overline{N}_{\mathrm{E}}$ by 1 vector $\mathbf{x}_{\mathrm{L},m}[k] = [x^j_{\mathrm{L},m}[k]]^{\mathrm{T}}_{j\in\Psi}$ and $\mathbf{x}_{\mathrm{A}}[k] = [x^j_{\mathrm{A}}[k]]^{\mathrm{T}}_{j\in\Psi}$. We assume orthogonal pilots with $\mathbf{x}_{\mathrm{L},m}\mathbf{x}^+_{\mathrm{L},n} = 0, \forall m \neq n$. The length of cyclic prefix is assumed to be larger than the maximum number L of channel taps. The parallel streams, i.e., $\mathbf{x}_{\mathrm{L},m}[k]$ and $\mathbf{x}_{\mathrm{A}}[k]$ are modulated with inverse fast Fourier transform (IFFT). Then the time-domain $\overline{N}_{\mathrm{E}}$ by 1 vector $\mathbf{y}^i[k]$, derived by Alice after removing the cyclic prefix at the i-th receiving antenna, can be written as:

$$\mathbf{y}^i[k] = \sum_{m=1}^{K} \mathbf{H}^i_{\mathrm{C},m}\mathbf{F}^{\mathrm{H}}\mathbf{x}_{\mathrm{L},m}[k] + \mathbf{H}^i_{\mathrm{C,A}}\mathbf{F}^{\mathrm{H}}\mathbf{x}_{\mathrm{A}}[k] + \mathbf{v}^i[k] \tag{3}$$

Here, $\mathbf{H}^i_{\mathrm{C},m}$ is the $\overline{N}_{\mathrm{E}} \times \overline{N}_{\mathrm{E}}$ circulant matrices of the m-th ALU, with the first column given by $[\,\mathbf{h}^{i^{\mathrm{T}}}_{\mathrm{L},m}\ \mathbf{0}_{1\times(\overline{N}_{\mathrm{E}}-L)}\,]^{\mathrm{T}}$. $\mathbf{H}^i_{\mathrm{C,A}}$ is a $\overline{N}_{\mathrm{E}} \times \overline{N}_{\mathrm{E}}$ circulant matrix with the first column given by $[\,\mathbf{h}^{i^{\mathrm{T}}}_{\mathrm{A}}\ \mathbf{0}_{1\times(\overline{N}_{\mathrm{E}}-L)}\,]^{\mathrm{T}}$ and $\mathbf{h}^i_{\mathrm{A}}$ is assumed to be independent with $\mathbf{h}^i_{\mathrm{L},m}, \forall m \in \mathcal{K}$. $\mathbf{h}^i_{\mathrm{L},m}$ and $\mathbf{h}^i_{\mathrm{A}}$ respectively denote the channel impulse response (CIR) vectors from the m-th ALU and Ava to the i-th receiving antenna of Alice. $\mathbf{h}^i_{\mathrm{A}}$ is assumed to be independent with $\mathbf{h}^i_{\mathrm{L},m}, \forall m \in \mathcal{K}$. $\mathbf{F} \in \mathbb{C}^{\overline{N}_{\mathrm{E}}\times\overline{N}_{\mathrm{E}}}$ denotes the discrete Fourier transform (DFT) matrix. $\mathbf{v}^i[k] \sim \mathcal{CN}(0, \sigma^2\mathbf{I}_{\overline{N}_{\mathrm{E}}})$ denotes the noise vector on time domain at the i-th antenna of Alice within the k-th symbol time. σ^2 is the average noise power of Alice. Taking FFT, Alice finally derives the frequency-domain N_{E} by 1 signal vector at the i-th receiving antenna as:

$$\widehat{\mathbf{y}}^i[k] = \sum_{j=1}^{K} \mathbf{F}_{\mathrm{L}}\mathbf{h}^i_{\mathrm{L},j}x_{\mathrm{L},j}[k] + \mathbf{I}^i[k] \tag{4}$$

where $\mathbf{F}_{\mathrm{L}} = \sqrt{\overline{N}_{\mathrm{E}}}\mathbf{F}(:, 1:L)$. $\mathbf{F}(:, 1:L)$ denotes the first L columns of $\mathbf{F}$. $\mathbf{I}^i[k]$ satisfies:

$$\mathbf{I}^i[k] = \mathrm{Diag}\{\mathbf{x}_{\mathrm{A}}[k]\}\mathbf{F}_{\mathrm{L}}\mathbf{h}^i_{\mathrm{A}} + \mathbf{w}^i[k] \tag{5}$$

where $\mathbf{w}^i[k] = \mathbf{F}\mathbf{v}^i[k]$. Stacking $\widehat{\mathbf{y}}^i[k]$ within TFRG#2 with $m_{\mathrm{E}} = K$, we can rewrite signals in Eq. (4) as:

$$\mathbf{Y}^i = \sum_{j=1}^{K} \mathbf{F}_{\mathrm{L}}\mathbf{h}^i_{\mathrm{L},j}\mathbf{x}_{\mathrm{L},j} + \mathbf{I}^i \tag{6}$$

where the $\overline{N}_{\mathrm{E}} \times m_{\mathrm{E}}$ matrix $\mathbf{Y}^i$ satisfies $\mathbf{Y}^i = [\widehat{\mathbf{y}}^i[k]]$ and $\mathbf{I}^i$ satisfies $\mathbf{I}^i = [\mathbf{I}^i[k]]$. The $1 \times m_{\mathrm{E}}$ vector $\mathbf{x}_{\mathrm{L},m}$ satisfies $\mathbf{x}_{\mathrm{L},m} = [x_{\mathrm{L},m}[k]]$. Especially, within TFRG#2 we define $x_{\mathrm{L},m}[k+1] - x_{\mathrm{L},m}[k] = \varphi$ where φ is publicly known. This configuration does not break down the randomness of pilots.

Assume that after TRP, each of random pilots of interest, i.e., m-th one, can be known by Alice. Given a least square (LS) estimation of $\mathbf{h}^i_{\mathrm{L},m}$, contaminated by $\mathbf{h}^i_{\mathrm{A}}$ with a noise bias, can be given by:

$$\widehat{\mathbf{h}}^i_{\mathrm{L},m} = \begin{cases} \mathbf{h}^i_{\mathrm{L},1} + (\mathbf{F}_{\mathrm{L}})^+\mathbf{I}^i\big(\mathbf{x}_{\mathrm{L},1}\big)^+ & \text{if } m = 1 \\ \mathbf{h}^i_{\mathrm{L},2} + (\mathbf{F}_{\mathrm{L}})^+\mathbf{I}^i\big(\mathbf{x}_{\mathrm{L},2}\big)^+ & \text{if } m = 2 \\ \quad\vdots & \quad\vdots \\ \mathbf{h}^i_{\mathrm{L},K} + (\mathbf{F}_{\mathrm{L}})^+\mathbf{I}^i\big(\mathbf{x}_{\mathrm{L},K}\big)^+ & \text{if } m = K \end{cases} \tag{7}$$

where $(\cdot)^+$ is the Moore-Penrose pseudoinverse.

G. RECEIVING SIGNAL MODEL FOR DATA TRANSMISSION

Without loss of generality, we assume: 1) The strategy of reusing REs for data transmission is, but not limited to, OFDM, by which K ALUs share the same subcarriers at the same OFDM symbols within TFRG#3; 2) All the ALUs operate at the same rate R calculated as $R = \frac{B}{m_{\mathrm{D}}T_s N_{\mathrm{D}}\Delta f}$; 3) The matched filter is employed on each subcarrier over N_{T} antennas;

The receiving signal model of Alice at each subcarrier, for example, the j-th one, is given as follows:

$$\mathbf{y}_j[k] = \sum_{m=1}^{K} \mathbf{g}_{j,m}d_{\mathrm{L},m}[k] + \mathbf{w}_j[k],\ j \in \Psi_{\mathrm{D}} \tag{8}$$

where $d_{\mathrm{L},m}[k]$ denotes the symbol transmitted by the m-th ALU at the k-th OFDM symbol. There exist $\mathbb{E}[|d_{\mathrm{L},m}[k]|^2] = \gamma$. $\mathbf{w}_j[k] = [w_{j,i}[k]]_{1\le i\le N_{\mathrm{T}}}$ denotes the noise vector at the j-th subcarrier and k-th OFDM symbol, and satisfies $\mathbf{w}_j[k] \sim \mathcal{CN}(0, \mathbf{I}_{N_{\mathrm{D}}})$. $\mathbf{g}_{j,m} = [g_{j,m,i}]_{1\le i\le N_{\mathrm{T}}}, m \in \mathcal{K}$ denotes the j-th subcarrier channel vector stacked by the m-th ALU across N_{T} antennas, and satisfies:

$$\mathbf{g}_{j,m} = \left[\,\mathbf{F}_{\mathrm{L},j}\mathbf{h}^1_{\mathrm{L},m} \cdots\ \mathbf{F}_{\mathrm{L},j}\mathbf{h}^{N_{\mathrm{T}}}_{\mathrm{L},m}\,\right]^{\mathrm{T}}, m \in \mathcal{K} \tag{9}$$

where $\mathbf{F}_{\mathrm{L},j}$ denotes the j-th row of $\mathbf{F}_{\mathrm{L}}$. We denote $\widehat{\mathbf{g}}_{j,m}$ as the stacked vector whose elements are derived from $\widehat{\mathbf{h}}^i_{\mathrm{L},m}$ in Eq. (7) and satisfies:

$$\widehat{\mathbf{g}}_{j,m} = \left[\,\mathbf{F}_{\mathrm{L},j}\widehat{\mathbf{h}}^1_{\mathrm{L},m} \cdots\ \mathbf{F}_{\mathrm{L},j}\widehat{\mathbf{h}}^{N_{\mathrm{T}}}_{\mathrm{L},m}\,\right]^{\mathrm{T}}, m \in \mathcal{K} \tag{10}$$

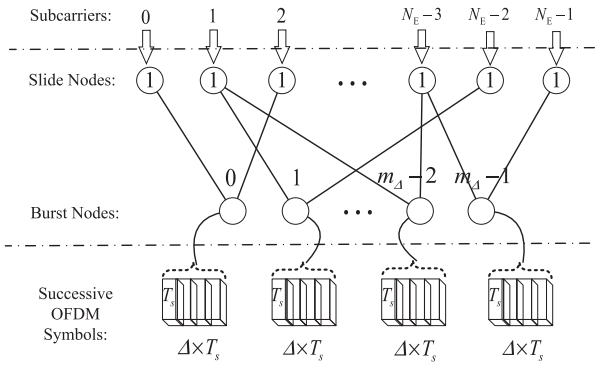

FIGURE 6. Illustration of a bipartite graph with vertices related to physical resources and edges reflecting the coding principle.

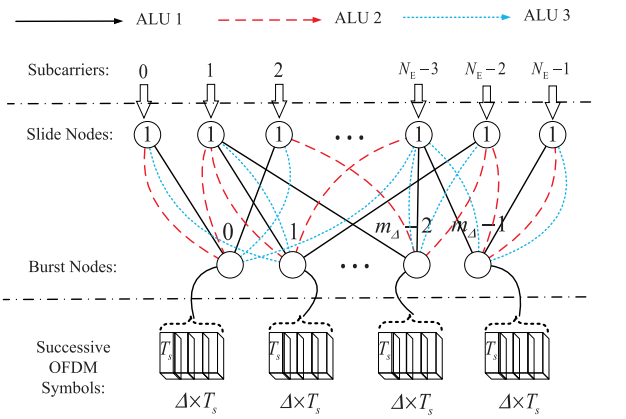

FIGURE 7. Illustration of an UCLB multigraph with its edges superimposed by various bipartite legitimate graphs.

Based on $\widehat{\mathbf{g}}_{j,m}$, Alice generates the matched filter, i.e., $\frac{1}{N_{\mathrm{T}}}\widehat{\mathbf{g}}_{j,m}^{\mathrm{H}}$ on Eq. (8) to decode the symbol of m-th ALU.

III. CGC THEORY

In this section, we focus on the process of TRP and aim to propose a method of encoding multi-dimensional time-frequency resources on TFRG#1, also named CGC theory. To this end, we back to Fig. 4 and Fig. 5.

A. BASIC CONCEPTS

Coding on joint time-frequency domain generates a code on a bipartite graph which can be shown in Fig. 6. Without loss of generality, we consider $m_\Delta \Delta$ OFDM symbols and N_{E} subcarriers. It means N_{E} resource blocks are occupied and the size of each one is equal to $\Delta = m_{\mathrm{E}}\ m_\Delta$. Given an instantaneous population of a total of m_{E} OFDM symbols and N_{E} subcarriers coded by digit 1, the graph representation for the t-th ALU and the binary graph coding can be defined.

Definition 1: A bipartite legitimate graph, $\mathcal{G}_t = (\mathcal{B}, \mathcal{S}, \mathcal{E}_t), t \in \mathcal{C}, \mathcal{C} = \{1, 2, \ldots, C\}$, consists of a set $\mathcal{B}$ of m_Δ burst nodes (one for Δ OFDM symbols with $m_{\mathrm{E}} = m_\Delta \Delta$), a set $\mathcal{S}$ of N_{E} slice nodes (one for each subcarrier), and a set $\mathcal{E}_t$ of edges. The vertex set $\mathcal{V}_t$ of the j-th ALU satisfies $\mathcal{V}_t = \mathcal{B} \cup \mathcal{S}$. An edge $\varepsilon_{i,j,t} = (i, j)_{t \in \mathcal{C}}, i \in \mathcal{B}, j \in \mathcal{S}$ connects a burst node $i \in \mathcal{B}$ to a slice node $j \in \mathcal{S}$ if and only if the j-th subcarrier has been activated within the i-th Δ OFDM symbols. This graph coding generates a codeword $\overline{\mathbf{c}}_{i,t} = [c_{i,j,t}]_{1 \le j \le N_{\mathrm{E}}}, 1 \le i \le m_\Delta$ which corresponds to the i-th burst node of the t-th ALU. The element $c_{i,j,t}$ of the codeword satisfies:

$$c_{i,j,t} = \begin{cases} 1 & \varepsilon_{i,j,t} \neq \emptyset \\ 0 & \text{Otherwise} \end{cases} \tag{11}$$

An example of graph coding is shown in Fig. 6.

Definition 2: A bipartite attack graph is described by $\mathcal{G}_{\mathrm{A}} = (\mathcal{B}, \mathcal{S}, \mathcal{E}_{\mathrm{A}})$, consisting of the same set $\mathcal{B}$ of m_Δ burst nodes and the same set $\mathcal{S}$ of N_{E} slice nodes as the ALUs, and also a random set $\mathcal{E}_{\mathrm{A}}$ satisfying $(i, j) \in \mathcal{E}_{\mathrm{A}}, i \in \mathcal{B}, j \in \mathcal{S}$. There exists $\bigcup_{k \in \mathcal{C}} \mathcal{E}_k \subseteq \mathcal{E}_{\mathrm{A}}$.

Now, let's define the basic operations among different bipartite graphs.

Definition 3: The superposition of arbitrary m bipartite graphs $\mathcal{G}_m = \bigvee_{1 \le j \le m} \mathcal{G}_j$ has the veriex set $\mathcal{V}_m = \bigvee_{1 \le j \le m} \mathcal{V}_j$ and the edge set $\mathcal{E}_{\mathrm{S}}$ containing all the edges of $\mathcal{G}_j, \forall 1 \le j \le m$. Note that the identities of edges of $\mathcal{G}_j, \forall 1 \le j \le m$ are preserved by assigning m different labels to these edges.

Definition 4: The superposition of arbitrary t bipartite graphs $\mathcal{G}_t$ gives an UCLB multigraph, where t edges labelled by their two points and by the number of their respective layers, may join the same pair vertices. Let $\mathcal{M}_{\mathcal{K},\mathcal{C}} = (\mathcal{B}, \mathcal{S}, \mathcal{E}_{\mathcal{K},\mathcal{C}})$ denote an UCLB multigraph with vertices $i \in \mathcal{B}, j \in \mathcal{S}$ and edges $(i, j)_t \in \mathcal{E}_{\mathcal{K},\mathcal{C}}$ for arbitrary $\mathcal{K}, \mathcal{K} \subseteq \mathcal{C}$ ALUs. Every subset $\overline{\varepsilon}_{i,j,t} = \{(i, j)_1 \cdots (i, j)_t\} \cap \mathcal{E}_{\mathcal{K},\mathcal{C}} \neq \emptyset$ is called a connection behween i and j. $\mathcal{M}_{\mathcal{K},\mathcal{C}} = \bigvee_{t \in \mathcal{K}, \mathcal{K} \subseteq \mathcal{C}} \mathcal{G}_t$ holds true.

An example of UCLB multigraph can be shown in Fig. 7. A hybrid attack behavior can be described using an UCLB multigraph which is a superposition of a random bipartite attack graph with multiple bipartite legitimate graphs. Obviously, the structure of UCLB multigraph records all types of behaviors from ALUs and the attacker.

With the aim of defending against attack, each of ALUs has various strategies of resource utilization and thus stores a set of possible bipartite legitimate graphs. We now define the set as follows:

Definition 5: A set of bipartite legitimate graphs of the t-th ALU is defined by $\mathcal{G}_t^{(\mathcal{B},\mathcal{S},\mathcal{E}_t^{\mathrm{T}})}$ in which all graphs occupy same vertexes and their edges constitute the set $\mathcal{E}_t^{\mathrm{T}}$ under parameter $(\mathcal{B}, \mathcal{S}, t)$. $\mathcal{E}_{t,k}$ denotes the k-th edge set of the bipartite legitimate graph for the t-th ALU, satisfying $\mathcal{E}_{t,k} = \bigcup_{i \in \mathcal{B}, j \in \mathcal{S}} \varepsilon_{i,j,t,k}$ where $\varepsilon_{i,j,t,k}$ connects the i-th burst node with the j-th slice node. There exists $\mathcal{E}_t^{\mathrm{T}} = \bigcup_{1 \le k \le |\mathcal{G}_t^{(\mathcal{B},\mathcal{S},\mathcal{E}_t^{\mathrm{T}})}|} \mathcal{E}_{t,k}$ where $|\mathcal{G}_t^{(\mathcal{B},\mathcal{S},\mathcal{E}_t^{\mathrm{T}})}|$ means the number of available codewords each of which uniquely corresponds to one bipartite legitimate graph. A K-bipartite graph set is defined by $(\mathcal{G}_1^{(\mathcal{B},\mathcal{S},\mathcal{E}_1^{\mathrm{T}})}, \ldots, \mathcal{G}_K^{(\mathcal{B},\mathcal{S},\mathcal{E}_K^{\mathrm{T}})})$.

Provided $\mathcal{G}_t^{(\mathcal{B},\mathcal{S},\mathcal{E}_t^{\mathrm{T}})}$, we define the codeword corresponding to the i-th burst node for the k-th graph of t-th ALU as $\mathbf{c}_{i,t,k} = [c_{i,j,t,k}]_{1\le j\le N_{\mathrm{E}}}, 1 \le i \le m_\Delta, 1 \le k \le |\mathcal{G}_t^{(\mathcal{B},\mathcal{S},\mathcal{E}_t^{\mathrm{T}})}|$. The codeword element $c_{i,j,t,k}$ satisfies:

$$c_{i,j,t,k} = \begin{cases} 1 & \varepsilon_{i,j,t,k} \neq \emptyset \\ 0 & \text{Otherwise} \end{cases} \tag{12}$$

With these, the concatenated graph code for the k-th graph of t-th ALU can be denoted by a $N_{\mathrm{E}}m_\Delta \times 1$ codeword $\mathbf{c}_{t,k}$ satisfying

$$\mathbf{c}_{t,k} = \mathbf{P}_{t,k}\bar{\mathbf{c}}_{t,k}, \bar{\mathbf{c}}_{t,k} = \left[\mathbf{c}_{1,t,k}^{\mathrm{T}}\ \mathbf{c}_{2,t,k}^{\mathrm{T}} \cdots\ \mathbf{c}_{N_{\mathrm{E}},t,k}^{\mathrm{T}}\right]^{\mathrm{T}} \tag{13}$$

where $\mathbf{P}_{t,k}$ is an arbitrary permutation matrix with the dimension of $N_{\mathrm{E}}m_\Delta \times N_{\mathrm{E}}m_\Delta$.

Now let us introduce the concept of bipartite graph query. When a hybrid attack happens, it can be modeled as a superposition of UCLB multigraph $\mathcal{G}$ with $\mathcal{G}_{\mathrm{A}}$. After observing the superposition, Alice launches a query $(\mathcal{Q}_1, \ldots, \mathcal{Q}_K, \mathcal{Q}_{\mathrm{A}})$ on $(\mathcal{G}_1^{(\mathcal{B},\mathcal{S},\mathcal{E}_1^{\mathrm{T}})}, \ldots, \mathcal{G}_K^{(\mathcal{B},\mathcal{S},\mathcal{E}_K^{\mathrm{T}})})$ and return a ranking of bipartite graph function $f[(\mathcal{Q}_1, \ldots, \mathcal{Q}_K, \mathcal{Q}_{\mathrm{A}}), \mathcal{G}]$ satisfying:

$$f\left[\left(\mathcal{Q}_1, \ldots, \mathcal{Q}_K, \mathcal{Q}_{\mathrm{A}}\right), \mathcal{G}\right] = \begin{cases} 1 & \bigcup\limits_{1\le i\le K} \mathcal{Q}_i \cup \mathcal{Q}_A = \mathcal{G} \\ 0 & \text{Otherwise} \end{cases} \tag{14}$$

Fig. 8 (a) shows the process of a bipartite graph query.

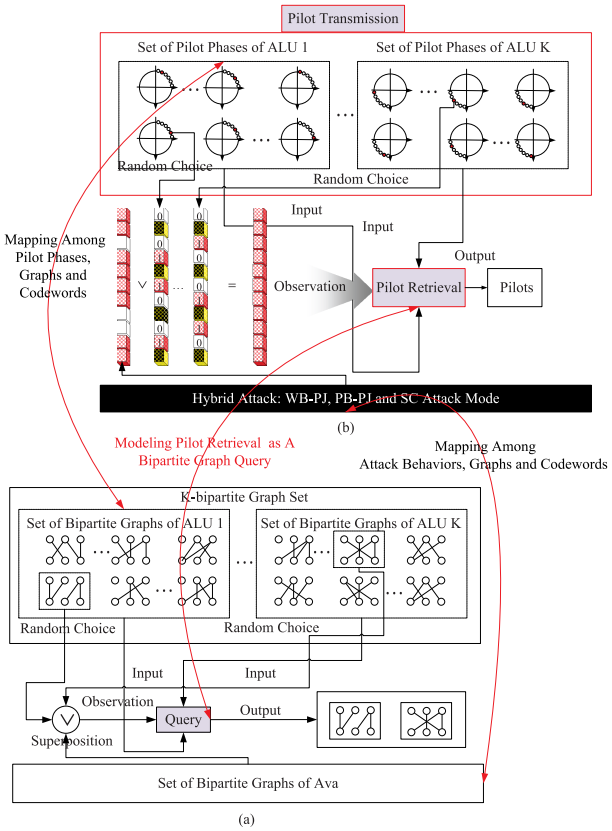

FIGURE 8. Illustration of modeling PA using CSC theory. (a) A bipartite graph query on multiple bipartite graphs; (b) PA Process.

B. MODELING PA AS BIPARTITE GRAPH QUERY

Though with above basic concepts, the relationship between PA and CSC theory is not clear. In this subsection, we would give an explanation of this.

The core of PA lies in TRP which includes pilot transmission and retrieval. As shown between Fig. 8 (a) and Fig. 8 (b), we can model pilot transmission as a superposition of bipartite graphs by formulating a one-to-one mapping among codewords, graphs and pilot phases. In details, the set of discrete pilot phases for the t-th ALU is encoded as the code matrix $\mathbf{B}_t = [\mathbf{c}_{t,k}]_{1\le k\le|\mathcal{G}_t^{(\mathcal{B},\mathcal{S},\mathcal{E}_t^{\mathrm{T}})}|}$. Then we have:

$$\mathbf{B}_t = \mathbf{P}_t \odot \bar{\mathbf{B}}_t, \bar{\mathbf{B}}_t = \left[\bar{\mathbf{c}}_{t,k}\right]_{1\leqslant k\leqslant\left|\mathcal{G}_t^{(\mathcal{B},\mathcal{S},\mathcal{E}_t^{\mathrm{T}})}\right|}, 1 \le t \le K \tag{15}$$

where $\odot$ denotes Hadamard product. $\mathbf{P}_t$ satisfies

$$\mathbf{P}_t = \left[\mathbf{P}_{t,1} \cdots \mathbf{P}_{t,\left|\mathcal{G}_t^{(\mathcal{B},\mathcal{S},\mathcal{E}_t^{\mathrm{T}})}\right|}\right] \tag{16}$$

And we also have:

$$\mathbf{B} = \mathbf{P} \odot \widetilde{\mathbf{B}}, \widetilde{\mathbf{B}} = \left[\bar{\mathbf{B}}_1 \cdots \bar{\mathbf{B}}_K\right] \tag{17}$$

where there exists

$$\mathbf{P} = \left[\mathbf{P}_1 \cdots \mathbf{P}_K\right] \tag{18}$$

For matrix $\mathbf{B}_t$, we define its j-th $N_{\mathrm{E}}m_\Delta \times 1$ column vector as $\mathbf{b}_{j,t} = [b_{i,j,t}]_{1\le i\le N_{\mathrm{E}}m_\Delta}, 1 \le j \le |\mathcal{G}_t^{(\mathcal{B},\mathcal{S},\mathcal{E}_t^{\mathrm{T}})}|$. To keep the distinguishability of codewords, $\mathbf{B}_t$ is exclusively allocated to the t-th ALU. Therefore, each of pilot phases corresponds to one unique codeword in the code matrix. There exist a total of $|\mathcal{G}_t^{(\mathcal{B},\mathcal{S},\mathcal{E}_t^{\mathrm{T}})}|$ pilot phases of the t-th ALU with $\sum_{t\in\mathcal{G}}|\mathcal{G}_t^{(\mathcal{B},\mathcal{S},\mathcal{E}_t^{\mathrm{T}})}| = C$. When pilot transmission proceeds from K ALUs, their pilot phases would be expressed in the form of independent encoded SAPs which are programmed by codewords $\mathbf{b}_{j,t}, t \in \mathcal{K}$. Multiple SAPs from multiple ALUs, after undergoing wireless channels, suffer from the superposition interference from each other. This superposition is observed, separated and identified at Alice. Then pilot retrieval is performed to recover original pilots securely and can be modeled as a bipartite graph query.

With this connection between PA and CGC theory, we can find a way to characterize the performance of PA.

Definition 6: The retrievability of pilot information during TRP for PA is denoted by $R_{(\mathcal{B},\mathcal{S},t)}$, satisfying:

$$R_{(\mathcal{B},\mathcal{S},t)} = \sum_{\mathcal{Q}_i\in\mathcal{G}_i^{(\mathcal{B},\mathcal{S},\mathcal{E}_i^{\mathrm{T}})},\mathcal{Q}_{\mathrm{A}}\in\mathcal{G}_{\mathrm{A}}} \frac{f\left[\left(\mathcal{Q}_1, \ldots, \mathcal{Q}_K, \mathcal{Q}_{\mathrm{A}}\right), \mathcal{G}\right]}{\left(\coprod\limits_{1\le i\le K}\left|\mathcal{G}_t^{(\mathcal{B},\mathcal{S},\mathcal{E}_t^{\mathrm{T}})}\right|\right)|\mathcal{G}_{\mathrm{A}}|} \tag{19}$$

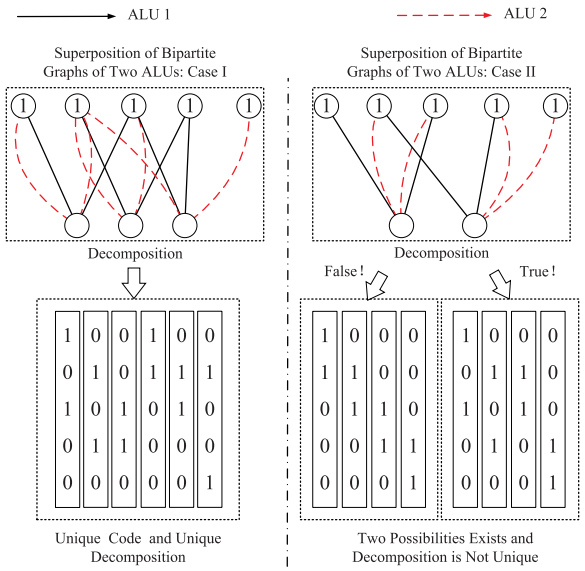

FIGURE 9. Two examples of CGC schemes. The left one satisfies P1, P2, and P3 while the right one does not.

C. GRAPH STRUCTURE REQUIRED FOR TRP

Examining Eq. (19), we hope to find a suitable parameter set $(\mathcal{B}_\xi, \mathcal{S}_\xi, t_\xi)$ such that for arbitrary $1 \le t \le t_\xi$, there exists $\mathcal{B} \subseteq \mathcal{B}_\xi$ and $\mathcal{S} \subseteq \mathcal{S}_\xi$ to guarantee $R_{(\mathcal{B}_\xi,\mathcal{S}_\xi,t_\xi)} = \xi$. $\xi = 1$ means a perfect pilot retrieval. This requires that the graph structure should satisfy the following principles:

- P.1 Every superposition of $K+1$ different bipartite legitimate graphs is distinct from every other sum of $K+1$ or fewer bipartite legitimate graphs, that is, $M_{\mathcal{K}_1,\mathcal{C}} = M_{\mathcal{K}_2,\mathcal{C}}$, holds true iff $\mathcal{K}_1 = \mathcal{K}_2$.
- P.2 Every superposition of $K+1$ different bipartite legitimate graphs can be decomposed by no bipartite legitimate graphs other than those used to form the sum, that is, there exists $\mathcal{K}_2 \subseteq \mathcal{C}$ such that when $M_{\mathcal{K},\mathcal{C}} = \bigcup_{\tau\in\mathcal{K}_2} \mathcal{G}_\tau$, there must exist $\mathcal{K}_2 = \mathcal{K}_1$.
- P.3 A nested structure of graph code holds true, that is, for arbitrary $\mathcal{C}_1 \subseteq \mathcal{C}$, there must exists $M_{\mathcal{C}_1\subseteq\mathcal{C}}$ satisfying both P.1 and P.2. This can support the dynamic access of ALUs of different numbers.
- P.4 A superposition of codes on bipartite legitimate graphs with the code on an arbitrary bipartite attack graph can be retrieved, that is, $M_{\mathcal{K},\mathcal{C}} \cup \mathcal{G}_\mathrm{A} \to \{\mathcal{G}_1, \ldots, \mathcal{G}_t\}$.

P.1 holds true for UCLB multigraphs since otherwise when $\mathcal{K}_1 \neq \mathcal{K}_2$ the degree of each vertex in $M_{\mathcal{K}_1,\mathcal{C}}$ and $M_{\mathcal{K}_2,\mathcal{C}}$ will not match with each other. Obviously, P.2 cannot hold true for all possibilities of parameters of $(\mathcal{B}, \mathcal{S}, t)$ whereas there exists a parameter $(\mathcal{B}_0, \mathcal{S}_0, t_0)$ such that P.2 holds true. P.3 cannot hold true for all possibilities of parameters of $(\mathcal{B}, \mathcal{S}, t)$ whereas there exists parameter $(\mathcal{B}_1, \mathcal{S}_1, t_1)$ such that P.3 holds true. Those examples can be seen in Fig. 9. A fact is that it is impossible to satisfy P.4 for arbitrary parameter $(\mathcal{B}, \mathcal{S}, t)$ and $\xi = 1$ cannot hold true. Due to the uncertainty of attack, the pilot retrieval would inevitably operate with errors.

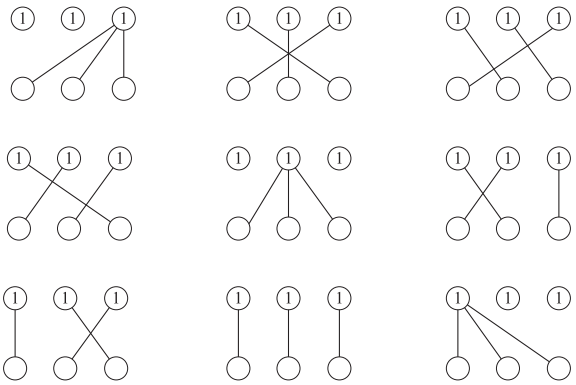

FIGURE 10. Illustration of bipartite graphs, each of which corresponds to a binary superiposed codeword.

IV. CGC ON BANDWIDTH PART

In this section, we first provide a solution of code design under CGC theory and then optimize the code matrix to make it available on BWPs.

A. A SOLUTION

CGC theory provides a unified framework of encoding time-frequency resources to achieve TRP under hybrid attack. In this framework, the graph structure can contribute to various coding methods each of which brings different level of tolerance against errors of pilot retrieval. The key lies in how to design the matrix $\mathbf{B}$. This prompts us to focus on $N_\mathrm{E} m_\Delta \times 1$ vector $\bar{\mathbf{c}}_{t,k}$ according to Eq. (13), Eq. (15) and Eq. (17).

In order to design this vector, let's examine the properties of P.1, P.2 and P.3 in code domain. A coding principle can be derived.

Proposition 1: 1) Every superposition of up to $K+1$ different codewords in $\mathbf{B}$ is distinct from every other sum of $K+1$ or fewer codewords in $\mathbf{B}$. 2) Every superposition of up to different codewords in $\mathbf{B}$ can be decomposed by no codewords other than those used to form the sum.

Proof: Since each bipartite legitimate graph corresponds to a unique codeword, the nested structure determines that up to $K+1$ different codewords should satisfy P.1 and P.2. ■

The first principle guarantees the superposition of codewords can be uniquely decipherable while the second one makes those codewords uniquely identified. In order to satisfy above two principles, we adopt the binary superimposed code [39]. Employing on bipartite graphs the binary superimposed code degenerates the UCLB multigraph into a bipartite graph due to the remove of labels. To further satisfy the property of P4, we update the binary superimposed code as H2DF codeword which combines signal features into superimposed code to make it defend against hybrid attack with desirable error tolerance. The cost for this are extra resource blocks occupied. For example, Δ should be equal to $K+2$ to support precise SAP encoding. An example of H2DF code on bipartite graphs can be seen in Fig. 10. The set of edges and vertexes required are defined as $\mathcal{B}_\mathrm{H}$ and $\mathcal{S}_\mathrm{H}$.

Proposition 2: A CGC codeword can be derived by:

$$\mathbf{B} = \mathbf{P} \odot \widetilde{\mathbf{B}}, \widetilde{\mathbf{B}} = \left[\overline{\mathbf{B}}_1 \cdots \overline{\mathbf{B}}_K \right] \tag{20}$$

where there exists

$$\mathbf{P} = \left[\mathbf{P}_1 \cdots \mathbf{P}_K \right] \tag{21}$$

where

$$\mathbf{P}_t = \left[\mathbf{P}_{t,1} \cdots \mathbf{P}_{t,\left|\mathcal{G}_t^{(B,S,\mathcal{E}_t^{\mathrm{T}})}\right|} \right], 1 \le t \le K \tag{22}$$

where $\widetilde{\mathbf{B}}$ is a H2DF code and $\mathbf{P}_{t,k}$ is an arbitrary permutation matrix with the dimension of $N_{\mathrm{E}} m_{\Delta} \times N_{\mathrm{E}} m_{\Delta}$. Therefore, a parameter set $(\mathcal{B}_{\xi}, \mathcal{S}_{\xi}, t_{\xi})$ exists such that the following condition can be satisfied:

$$\mathcal{B}_{\xi} = \mathcal{B}_{\mathrm{H}}, \mathcal{S}_{\xi} = \mathcal{S}_{\mathrm{H}}, t_{\xi} = \left[\frac{N_{\mathrm{E}} m_{\mathrm{E}}}{[1 + K(k-1)](K+2)} \right]^k \tag{23}$$

where $\xi = 1 - \sqrt{[\frac{[1+K(k-1)](K+2)}{N_{\mathrm{E}} m_{\mathrm{E}}}]^k \frac{1}{2K}}, k = 2, 3$ and $N_{\mathrm{E}} m_{\mathrm{E}} \geqslant K(k-1)[1 + K(k-1)](K+2), K(k-1) \geqslant 3$.

Proof: In H2DF coding, a resource block is with the size of $\Delta = K + 2$ and only one resource block is employed in time domain, that is, $m_{\Delta} = 1$. In this case, the probability of pilot identification is equal to $P = \sqrt{\frac{[1+K(k-1)]^k}{2(N_{\mathrm{E}})^k K}}, k = 2, 3$ with $N_{\mathrm{E}} \geqslant K(k-1)[1 + K(k-1)]$. In our scheme, the use of CGC can relax the constraint. Since CGC bring codewords in multiple OFDM symbols together, the length of each H2DF codeword increases in practice. Or equivalently, the number of available subcarriers increases from N_{E} to $N_{\mathrm{E}} m_{\mathrm{E}}$ $(K + 2)$ despite occupying the same BWP as H2DF coding. At this point, CGC makes H2DF coding suitable for grant-free URLLC. Then the probability of pilot identification is equal to $P = \sqrt{[\frac{[1+K(k-1)](K+2)}{N_{\mathrm{E}} m_{\mathrm{E}}}]^k \frac{1}{2K}}, k = 2, 3$ with $N_{\mathrm{E}} m_{\mathrm{E}} \geqslant K(k-1)[1 + K(k-1)](K+2), K(k-1) \geqslant 3$. The retrievability is then equal to $1 - P$. ■

B. CGC ON BWP: THE OPTIMIZATION OF **P**

Fig. 11 shows how CGC on BWP can help reduce the burdens on bandwidths and support flexible operations on grant-free physical resource pools. However, CGC codebook $\mathbf{B}$ is rather different with H2DF codebook $\widetilde{\mathbf{B}}$ due to the permutation matrix $\mathbf{P}$. The design of $\mathbf{P}$ determines the performance of CGC. Let us focus on each matrix $\mathbf{P}_{t,k}$. The choices of $\mathbf{P}_{t,k}$ under different k for different t are independent with each other. In other words, the value of $\mathbf{P}_{t_1,k_1}$ is independent with $\mathbf{P}_{t_2,k_2}$ for arbitrary $t_1 \neq t_2$ and $k_1 \neq k_2$.

However, extra constraints must be imposed on $\mathbf{P}$. Otherwise, the coding diversity would be very limited. Let us consider a very simple H2DF code based on superimposed

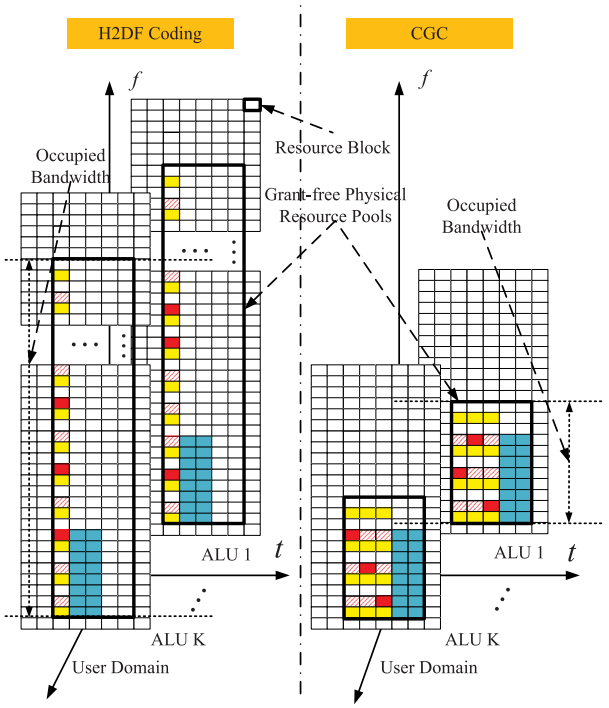

FIGURE 11. Illustration of CGC on BWP.

code shown as follows:

$$\begin{bmatrix} 0&0&0&0&0&0&1&1&1 \\ 0&0&0&1&1&1&0&0&0 \\ 1&1&1&0&0&0&0&0&0 \\ 0&0&1&0&0&1&0&0&1 \\ 0&1&0&0&1&0&0&1&0 \\ 1&0&0&1&0&0&1&0&0 \\ 0&1&0&1&0&0&0&0&1 \\ 0&0&1&0&1&0&1&0&0 \\ 1&0&0&0&0&1&0&1&0 \end{bmatrix} \tag{24}$$

For this code matrix, when performing left multiplication by permutation matrix on each column, we can that the i-th column satisfying $i = 2, 3, 4, 6, 7, 8$ can be transformed into a codeword by find six different permutation matrices. This can be shown in Fig. 12. The cardinality of the CGC code can be significantly reduced. Therefore, we have the following proposition.

Proposition 3: For a CGC code employed on BWP, the following condition should be satisfied:

$$\mathbf{P}_{t_1,k_1} = \mathbf{P}_{t_2,k_2}, \forall t_1 \neq t_2, \forall k_1 \neq k_2 \tag{25}$$

We define the set $\mathcal{P}$ as the set of all possible permutation matrices, then there exists:

$$\mathbf{P}_{t,k} \in \mathcal{P}, \forall t, k, 1 \le t \le K, 1 \leqslant k \leqslant \left|\mathcal{G}_t^{(B,S,\mathcal{E}_t^{\mathrm{T}})}\right| \tag{26}$$

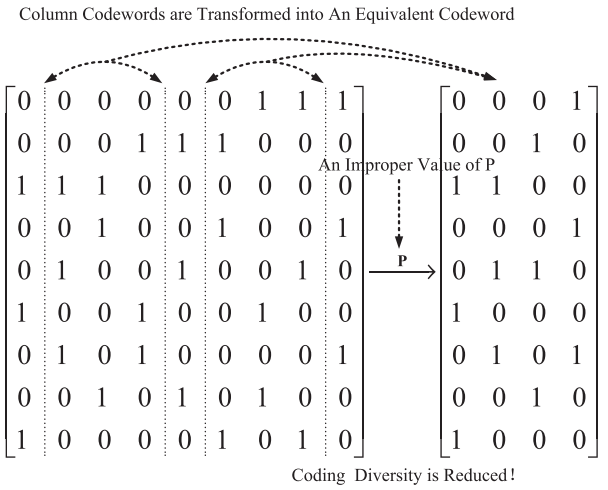

FIGURE 12. Illustration of the effect of improper P on code.

The value of permutation matrix $\mathbf{P}$ is not constrained except that $\mathbf{P}_{t,k}$ should keep identical with each other under different values of t, k.

C. CGC BASED UAD

Despite the fact that we have provided a solution to achieve TRP securely, applying CGC in grant-free URLLC needs also to make clear how UAD, channel estimation and data transmission work together. Since the model of channel estimation and data transmission have been introduced in Subsection II-F and Subsection II-G, we now merely focus on the design of UAD.

UAD refers to the detection of the total number of ALUs and their identities. Due to the lack of grant which contains the temporary cell radio network temporary identity (T-CRNTI) used to identify ALUs, identities of piots employed by ALUs basically determine identities of ALUs in GFMA. Conventionally, UAD is a blind detection process relying on multiple deterministic and distinguishable pilots [6]. *With those pilots being randomized, UAD becomes rather different. Based on Proposition 1, CGC creates many unique temporal identifiers, e.g., unique codewords, for each ALU though only one is employed during uplink access. Those distinguishable identifiers are different with those from other ALUs and their mutual superposition are also distinguishable.* This is the basic of UAD.

To identify the total number of ALUs and their identities by CSC on code domain, Alice detects the number of signals across subcarriers and compares the observed superposition codeword $\mathbf{b}_\mathrm{I}$ with a reference codebook defined by a $N_\mathrm{E} \times \binom{N_\mathrm{L}}{k}$ matrix $\mathbf{C}_k, k = 2, 3, \ldots, G+1$. $\mathbf{C}_k$ is the collection of Boolean sums of all vector codewords from the stacked matrix $[\mathbf{B}_1 \cdots \mathbf{B}_G]$, taken exactly k at a time such that each codeword in it is unique. N_L denotes the total number of columns of $[\mathbf{B}_1 \cdots \mathbf{B}_G]$. The j-th column vector codeword of $\mathbf{C}_k$ is denoted by $\mathbf{d}_{k,j}$. With these preparations, we propose an algorithm of UAD in Algorithm 1.

Algorithm 1: UAD Under Hybrid Attack.

Require: N_j for $j \in N_\mathrm{E}$, $\mathbf{C}_k, k = 2, 3, \ldots, G+1$, $\mathbf{b}_\mathrm{I}$
Ensure: Attack Mode; Number of ALUs; Codewords of ALUs.

if all the elements of $\mathbf{b}_\mathrm{I}$ is equal to 1 **then**
  Indicate WB-PJ attack mode. Encode $N_j - 1$, $j \in \Psi_\mathrm{E}$ as binary codeword $\overline{\mathbf{b}}_\mathrm{I}$.
  while $1 \le j \le G+1$ **do**
    Compare $\overline{\mathbf{b}}_\mathrm{I}$ with each column of matrix $\mathbf{C}_k$.
    ▶ *All possibilities has been stored and just search $\mathbf{C}_k$ for compassions. This is due to the property in Proposition 1.*
    if $\overline{\mathbf{b}}_\mathrm{I}$ belongs to the j_0-th column of $\mathbf{C}_{k_0}$ **then**
      Break
    end if
  end while
  Indicate there exist k_0 ALUs and decompose $\mathbf{d}_{k_0,j_0}$ into k_0 codewords used by ALUs.
  ▶ *Successful detection of WB-PJ attack mode and number of ALUs and their codewords.*
else
  while $1 \le k \le G+1$ **do**
    Compare $\mathbf{b}_\mathrm{I}$ with each column of matrix $\mathbf{C}_k$.
    if $\mathbf{b}_\mathrm{I}$ belongs to the j_1-th column of $\mathbf{C}_{k_1}$ **then**
      Break
    end if
  end while
  Output $\mathbf{d}_{k_1,j_1}$ which is decomposed into k_1 codewords.
  if Number of non-zero elements of all k_1 codewords of ALUs is equal to $\sum_{j=1}^{N_\mathrm{E}} N_j^k$ **then**
    Indicate SC attack mode and k_1 ALUs. Those k_1 codewords are true and authentic.
    ▶ *Successful detection of SC attack mode and number of ALUs and their codewords.*
  else
    Find the set $\mathcal{F}$ satisfying $j \in \mathcal{F}, N_j = 1$.
    while $j \in \mathcal{F}$ **do**
      Make $N_j = 0$. Encode the codeword as $\overline{\overline{\mathbf{b}}}_\mathrm{I}$
      while $1 \le k \le G+1$ **do**
        Compare $\overline{\overline{\mathbf{b}}}_\mathrm{I}$ with each column of $\mathbf{C}_k$
        if $\overline{\overline{\mathbf{b}}}_\mathrm{I}$ belongs to the j_2-th column of $\mathbf{C}_{k_2}$ **then**
          Break.
        end if
      end while
    end while
    Indicate PB-PJ attack mode and k_2 ALUs. Decompose $\mathbf{d}_{k_2,j_2}$ into k_2 original codewords.
    ▶ *Successful detection of WB-PJ attack mode and number of ALUs and their codewords.*
  end if
end if

V. PERFORMANCE EVALUATION AND TRADE-OFFS

In this section, we would evaluate the performance of grant-free URLLC system in the respect of many metrics and further check whether or not it is feasible in URLLC. After this, we will analyze the performance trade-off in this system.

A. FAILURE PROBABILITY OF CGC BASED GRANT-FREE URLLC

Though with the help of TRP, UAD and channel estimation, data transmission also operates with a certain probability of failure which lies in the inevitable decoding error of short packets. In this subsection, we define the failure probability of the overall CGC based GFMA system to characterize the reliability performance. There has been well-known expressions in literature [37] to calculate the decoding error probability P_d of transmissions over fading channels as a function of the average received signal-to-noise-ratio (SNR) γ_0, the transmission rate R and matched filter receiver; that is,

$$P_d = \mathbb{E}\left[Q\left(\frac{C(\gamma)-R}{\sqrt{V(\gamma)/(N_{\mathrm{D}}m_{\mathrm{D}})}}\right)\right] \tag{27}$$

where $R=\frac{B}{m_{\mathrm{D}}T_sN_{\mathrm{D}}\Delta f}$, $Q(x)=\int_x^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$ and $C(\gamma)=\log_2(1+\gamma), V(\gamma)=1-\frac{1}{(1+\gamma)^2}$ with γ defined as the instantaneous SNR on RE on which K_c-1 interfering signals coexist. Without loss of generality, we assume all the interfering signals are Gaussian distributed and received at Alice with the same average SNR γ_0. The distribution $f_{K_c}(x)$ of γ under matched filter receiver can be calculated by:

$$f_{K_c}(x)=\frac{x^{N_T-1}e^{-\frac{x}{\gamma_0}}}{(N_T-1)!\gamma_0^{K_c+1}}\sum_{i=0}^{N_T}\binom{N_T}{i}\frac{\gamma_0^{K_c+i}\Gamma(K_c+i)}{\Gamma(K_c)(x+1)^{K_c+i}} \tag{28}$$

where $\Gamma(\cdot)$ denotes the Gamma function. P_d is calculated by:

$$P_d=\int_0^\infty Q\left(\frac{C(\gamma)-R}{\sqrt{V(\gamma)/(N_{\mathrm{D}}m_{\mathrm{D}})}}\right)f_{K_c}(\gamma)\,d\gamma \tag{29}$$

We assume at most one retransmission can be supported. The first transmission is deemed successful if the intended ALU or its pilot or equivalently estimated channel is correctly identified and its data is decoded successfully. In this case, the failure probability of CGC based grant-free URLLC system is $1-\xi(1-P_d)$. When the intended ALU cannot be identified, or it is identified but its data can not be decoded, the ALU will perform a retransmission over shared resources. The probability of correctly decoding the retransmitted data can be calculated by $[1-\xi(1-P_d)]\xi(1-P_d)$. Finally the failure probability of CGC based grant-free URLLC system, denoted by P_e, is given by:

$$P_e=[1-\xi(1-P_d)]^2 \tag{30}$$

Now let's examine the expressions of failure probability under channel estimation errors. Receiving signal at ALice under matched filtering receiver is expressed in Eq. (31). Without loss of generality and take the m_0-th ALU for example, we have $\widehat{\mathbf{g}}_{j,m_0}=\mathbf{g}_{j,m_0}$ when no estimation error is assumed and otherwise $\widehat{\mathbf{g}}_{j,m_0}=(1-\lambda)\mathbf{g}_{j,m_0}-\lambda\widetilde{\mathbf{g}}_{j,m_0}, 0<\lambda<1$. $\widetilde{\mathbf{g}}_{j,m_0}\sim\mathcal{CN}(0,\mathbf{I}_{N_{\mathrm{T}}})$ is independent with $\widehat{\mathbf{g}}_{j,m_0}$ and larger λ means that channel estimation gets worse.

Theorem 1: With precise channel estimated, the asymptotic expression of received SINR as $N_{\mathrm{T}}\to\infty$ is given by:

$$\gamma_{asy}\triangleq\gamma_{asy}^{\text{perfect}}\xrightarrow[N_{\mathrm{T}}\to\infty]{\text{a.s.}}\frac{N_{\mathrm{T}}\gamma_0}{\gamma_0K_c+1} \tag{32}$$

and the result with estimation error is given by:

$$\gamma_{asy}\triangleq\gamma_{asy}^{\text{error}}\xrightarrow[N_{\mathrm{T}}\to\infty]{\text{a.s.}}\frac{N_{\mathrm{T}}\gamma_0(1-\lambda)}{\gamma_0K_c+\lambda\gamma_0+1},0<\lambda<1 \tag{33}$$

The decoding error probability satisfies:

$$P_{\mathrm{d}}\xrightarrow[N_{\mathrm{T}}\to\infty]{\text{a.s.}}Q\left(\frac{C(\gamma_{asy})-R}{\sqrt{V(\gamma_{asy})/(N_{\mathrm{D}}m_{\mathrm{D}})}}\right) \tag{34}$$

and the failure probability of CGC based grant-free URLLC can be finally expressed as Eq. (35). The specific value of γ_{asy} depends on the estimation assumption above.

Proof: Note that by the strong law of large numbers [40], we have:

$$\frac{1}{N_{\mathrm{T}}}\sum_{i=1}^{N_{\mathrm{T}}}\left|g_{j,m_0,i}\right|^2\xrightarrow[N_{\mathrm{T}}\to\infty]{\text{a.s.}}E\left|g_{j,m_0,i}\right|^2=1 \tag{36}$$

$$\frac{1}{N_{\mathrm{T}}}\sum_{m=1,m\neq m_0}^{K}\sum_{i=1}^{N_{\mathrm{T}}}\widehat{g}_{j,m_0,i}^{*}g_{j,m,i}\xrightarrow[N_{\mathrm{T}}\to\infty]{\text{a.s.}}E\left[\widehat{g}_{j,m_0,i}^{*}g_{j,m,i}\right]=0 \tag{37}$$

$$\frac{1}{N_{\mathrm{T}}}\sum_{i=1}^{N_{\mathrm{T}}}\widehat{g}_{j,m_0,i}^{*}w_{j,i}[k]\xrightarrow[N_{\mathrm{T}}\to\infty]{\text{a.s.}}E\left[\widehat{g}_{j,m_0,i}^{*}w_{j,i}[k]\right]=0 \tag{38}$$

When $\widehat{\mathbf{g}}_{j,m_0}=\mathbf{g}_{j,m_0}$, the received SINR for the i-th ALU can be written as:

$$\gamma_{asy}^{\text{perfect}}=\frac{\frac{1}{N_{\mathrm{T}}}\sum_{i=1}^{N_{\mathrm{T}}}\left|g_{j,m_0,i}\right|^2}{\frac{\left(\sum_{m=1,m\neq m_0}^{K}\sum_{i=1}^{N_{\mathrm{T}}}\widehat{g}_{j,m_0,i}^{*}g_{j,m,i}\right)^2}{N_{\mathrm{T}}\sum_{i=1}^{N_{\mathrm{T}}}\left|g_{j,m_0,i}\right|^2}+\frac{1}{N_{\mathrm{T}}\gamma_0}} \tag{39}$$

According to Eq. (36), the numerator of Eq. (39) is almost surely to be 1. According to Eq. (36), Eq. (37) and Eq. (38), $\left(\sum_{m=1,m\neq m_0}^{K}\sum_{i=1}^{N_{\mathrm{T}}}\widehat{g}_{j,m_0,i}^{*}g_{j,m,i}\right)^2/N_{\mathrm{T}}$ in Eq. (39) almost surely approach to K_c. With these, we can derive Eq. (32). Under estimation errors, the received SINR for the i-th ALU can be written as:

$$\gamma_{asy}^{\text{error}}=\frac{\frac{1}{N_{\mathrm{T}}}\sum_{i=1}^{N_{\mathrm{T}}}\widehat{g}_{j,m_0,i}^{*}g_{j,m_0,i}}{\frac{\left(\sum_{m=1,m\neq m_0}^{K}\sum_{i=1}^{N_{\mathrm{T}}}\widehat{g}_{j,m_0,i}^{*}g_{j,m,i}\right)^2}{N_{\mathrm{T}}\sum_{i=1}^{N_{\mathrm{T}}}\widehat{g}_{j,m_0,i}^{*}g_{j,m_0,i}}+\frac{1}{N_{\mathrm{T}}\gamma_0}} \tag{40}$$

where $\widehat{g}_{j,m_0,i}=(1-\lambda)g_{j,m_0,i}-\lambda\widetilde{g}_{j,m_0,i}, 0<\lambda<1$. After simplification using Eq. (36), Eq. (37), and Eq. (38), we

can derive Eq. (33) based on Eq. (40). Now, we consider the Eq. (29). Under the large number of antennas, the input of Q function is irrelevant with γ. Therefore, we have $P_{\mathrm{d}} \xrightarrow[N_{\mathrm{T}}\to\infty]{\text{a.s.}} Q(\frac{C(\gamma_{asy})-R}{\sqrt{V(\gamma_{asy})\,(N_{\mathrm{D}}m_{\mathrm{D}})}}) \int_0^{\infty} f_{K_c}(\gamma)d\gamma = Q(\frac{C(\gamma_{asy})-R}{\sqrt{V(\gamma_{asy})\,(N_{\mathrm{D}}m_{\mathrm{D}})}})$. By substituting the value of P_{d} and the value ξ of Proposition 2 into Eq. (30), we can derive the Eq. (35). ■

B. TRADE-OFFS IN CGC BASED GRANT-FREE URLLC

We first define the accessibility is the practice of CGC based grant-free URLLC to keep low failure probability within specified tolerances while being usable by as many ALUs as possible.

Definition 7: We define the α-accessibility of CGC based grant-free URLLC as the ratio of the number of multiplexed ALUs that could maintain P_e less than α to the number of OFDM symbols for channel estimation and data decoding during this period, that is,

$$S = \frac{K}{(m_{\mathrm{D}} + m_{\mathrm{E}})\,T_{\mathrm{s}}}, P_{\mathrm{e}} \le \alpha \tag{41}$$

where α denotes the reliability constraint, usually equal to 10^{-5}.

Now the system has four key metrics in hand, that is, retrievability, failure probability of the system, accessibility and latency. Previously, three metrics, including retrievability, failure probability of the system and accessibility, have been derived with closed-form analytical expressions. Based on these metrics, we can see how time-frequency-user resources should and could be controlled to make grant-free URLLC work well. Note that we employ the failure probability of the system to characterize the reliability performance. For example, the lower the failure probability is, the higher the reliability will be.

Note that $N_{\mathrm{E}}m_{\mathrm{E}} \ge K(k-1)[1+K(k-1)](K+2)$ is configured and N_{E} is fixed for the consideration of low overheads on RF chain switching. Based on Eq. (23), the accessibility would increase with the increase of the number of ALUs while the retrievability and reliability would decrease if the latency is fixed. This is because more ALUs with attempt to access gNB would increase the decoding errors of random multiple pilots and data under the uncertainty of attacker. This phenomenon proves the existence of retrievability-accessibility trade-off and reliability-accessibility trade-off.

TABLE 1 Simulation Parameters and Values

Simulation Parameters	Values
Modulation	OFDM with normal CP
Subcarrier spacing	Δf =60kHz, 120kHz
Bandwidth of BWP	$\le$ 100MHz, $\le$400MHz
Coherence bandwidth	$3 \times \Delta f$
OFDM symbol duration	$T_{\mathrm{s}} = 17.86\mu s,\ 8.93\mu s$
Number of OFDM symbols at each burst node	$\Delta = K + 2$
Choices of m_{Δ}	$1, 2, 4, 8$
Number of subcarriers for TRP	$N_{\mathrm{E}} = 512$
Number of subcarriers for channel estimation	$N_{\mathrm{E}} = 128$
Pilot subcarrier arrangement for TRP	CGC
Pilot subcarrier arrangement for channel estimation	Block type
Channel estimator at gNB	LS
Number of subcarriers for data transmission	$N_{\mathrm{D}} = 4$
Combing technique at gNB	Matched filter receiver
Size of data packets	$R = 32$ Bytes
SNR of receiving data at gNB	$\gamma = 10$dB
Time consumed by other processes	$T_{\mathrm{extra}} = 100\ \mu s$
Number of antennas at gNB	$N_{\mathrm{T}} \ge 64$
Channel fading model	Rayleigh
Number of ALUs	$K \le 13$
Number of channel taps	$L = 6$
Failure probability requirement	$P_{\mathrm{e}} \le 10^{-5}$
Latency constraints	$T_{\mathrm{con}} = 1$ms

Let us revisit Eq. (23) to find retrievability-latency trade-off. Given the fixed number of ALUs and fixed N_{E}, the retrievability increases with the increase of time-domain resources. When m_{E} increases, more OFDM symbols can be provided for channel estimation. This also increases the length of codeword of CGC for channel authentication whereas the latency consumed increases as well. Examining Eq. (35), we can find a reliability-latency trade-off, a basic trade-off existing in traditional URLLC systems. But the difference here is that new mathematical characteristic of this trade-off is formulated by considering the reliability of PA, channel estimation and data transmission.

VI. NUMERICAL RESULTS

In this section, we will evaluate the metrics of retrievability, reliability, accessibility and latency under channel estimation errors and show their mutual relationships. FR1 for Sub-6 GHz and FR2 for millimeter wave in 5G NR are considered respectively. It means that the system could be expected to work on BWP of at most 100 MHz channel bandwidth for FR1 and that of at most 400 MHz channel bandwidth for FR2 [36]. Simulation parameters and values on grant free physical resource pool can be seen in Table 1. Note that one pilot every three consecutive subcarriers is inserted to

$$\frac{1}{N_{\mathrm{T}}}\widehat{\mathbf{g}}_{j,m_0}^{\mathrm{H}}\mathbf{y}_j[k] = \frac{d_{\mathrm{L},m_0}[k]}{N_{\mathrm{T}}}\sum_{i=1}^{N_{\mathrm{T}}}\widehat{g}_{j,m_0,i}^{*}g_{j,m_0,i} + \frac{d_{\mathrm{L},m_0}[k]}{N_{\mathrm{T}}}\sum_{m=1,m\neq m_0}^{K}\sum_{i=1}^{N_{\mathrm{T}}}\widehat{g}_{j,m_0,i}^{*}g_{j,m,i} + \frac{1}{N_{\mathrm{T}}}\sum_{i=1}^{N_{\mathrm{T}}}\widehat{g}_{j,m_0,i}^{*}w_{j,i}[k] \tag{31}$$

$$P_{\mathrm{e}} = \left\{1-\left\{1-\sqrt{\left[\frac{[1+K(k-1)](K+2)}{N_{\mathrm{E}}m_{\mathrm{E}}}\right]^{k}\frac{1}{2K}}\right\}\left[1-Q\left(\frac{(1+\gamma_{asy})\left[\log_2(1+\gamma_{asy}) - \frac{R}{m_{\mathrm{D}}T_{\mathrm{s}}N_{\mathrm{D}}\Delta f}\right]\sqrt{N_{\mathrm{D}}m_{\mathrm{D}}T_{\mathrm{s}}}}{\sqrt{\left[(1+\gamma_{asy})^2-1\right]T_{\mathrm{s}}}}\right)\right]\right\}^2 \tag{35}$$

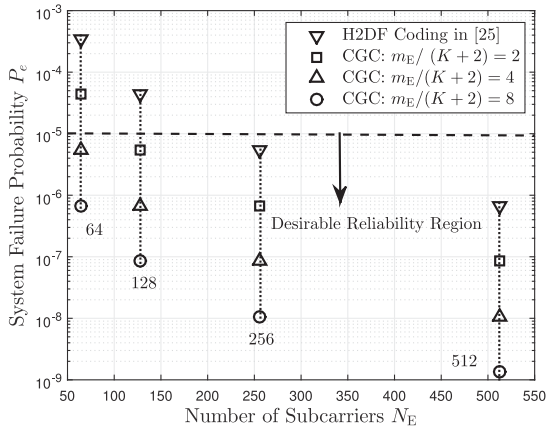

FIGURE 13. Failure probability of CGC based grant-free URLLC versus N_E.

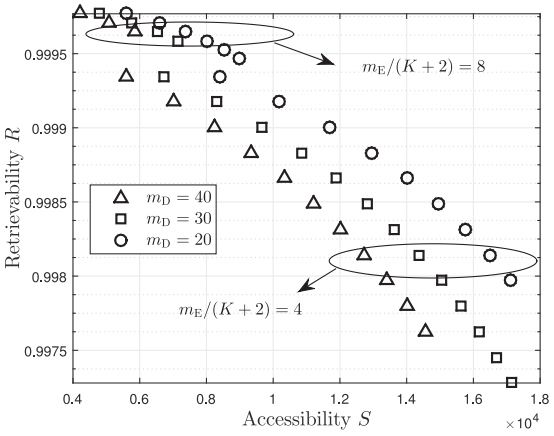

FIGURE 15. Retrievability-accessibility trade-off curves.

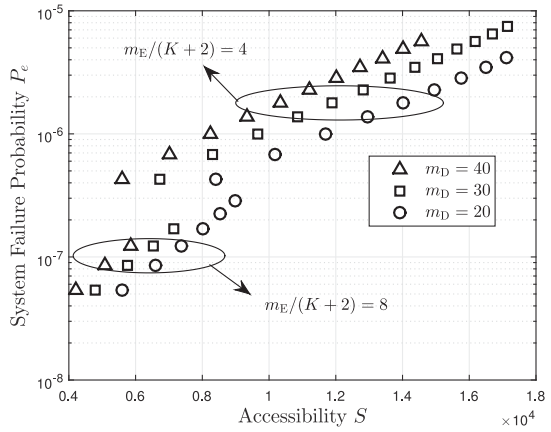

FIGURE 14. Reliability-accessibility trade-off curves.

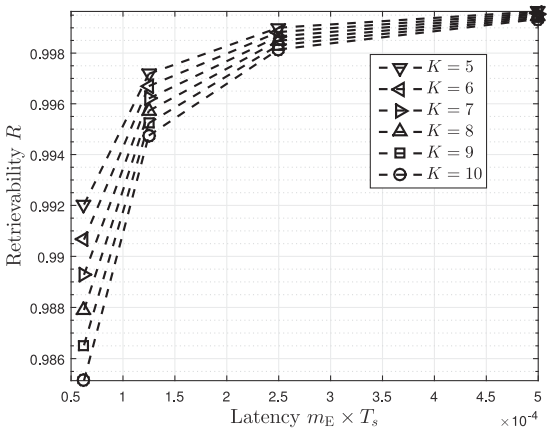

FIGURE 16. Retrievability-latency trade-off curves.

acquire independent frequency-domain variations of the channels. Given the bandwidth constraints, the maximum number of pilot subcarriers for PA is limited to 512. Therefore, at most $512 \times 3 \times 240 = 368.84$ MHz channel bandwidth is occupied for $\Delta f = 120$ kHz and at most $512 \times 3 \times 64 = 98.304$ MHz channel bandwidth is required for $\Delta f = 60$ kHz. In time domain, Δ is configured to be $K+2$ such that the number of signals from at most K ALUs and one attacker can be detected precisely at each subcarrier. Therefore, there exists $m_\Delta = m_E\ (K+2)$.

Fig. 13 presents the curves of the failure probability of CGC based grant-free URLLC versus N_E under various number of antenna groups with antenna interleaving. $\Delta f = 120$ kHz and $T_s = 8.92 \times 10^{-6}$s are configured, supporting 64 antennas at gNB and 4 ALUs. Channel estimation error is configured to be $\lambda = 0.2$. Note that the value of $m_E\ (K+2)$ is selected within a finite discrete set of integers, i.e.,$\{2, 4, 8\}$, and N_E is configured to be 64, 128, 256, and 512 respectively. It can be observed that our proposed CGC method, compared with H2DF coding, can provide better reliability performance given the same frequency-domain overheads. From another perspective, our scheme brings much lower frequency-domain overheads while keeping reliability higher than 99.999%.

In Fig. 14, we simulate the reliability-accessibility trade-off curves. α is configured to be 10^{-5} with $\Delta f = 120$ kHz and $T_s = 8.92 \times 10^{-6}$s. Channel estimation error is configured to be $\lambda = 0.2$ with $N_T = 120$ and $N_E = 128$. $m_E\ (K+2)$ is selected to be 4, 8 and m_D is fixed to be 20, 30, 40 respectively. Thus at most 18 ALUs under $m_D = 20$, $m_E\ (K+2) = 4$ can be supported due to the constraints of $(m_E + m_D)T_s \leqslant T_{con} - T_{extra}$. It can be obvserved that the failure probability mainly lie within the interval of $10^{-5} \sim 10^{-6}$ under $m_E\ (K+2) = 4$ while within the interval of $10^{-6} \sim 10^{-7}$ under $m_E\ (K+2) = 8$. At each interval, there always exist reliability-accessibility trade-off curves indicating that more accessibility will bring less reliability. This is due to the uncertainty of codeword selection from multiple ALUs and attacker.

In order to evaluate the influence of increased number of ALUs on the retrievability of PA, we simulate the retrievability-accessibility trade-off curves in Fig. 15. We also provide the performance curve of retrievability versus the occupied OFDM symbols during the period of PA in Fig. 16. $\Delta f = 120$ kHz and $T_s = 8.92 \times 10^{-6}$s are configured with $N_T = 120$ and $N_E = 128$. More specifically, in Fig. 15, $m_E\ (K+2)$ is selected to be 4 and 8, and m_D is chosen to be

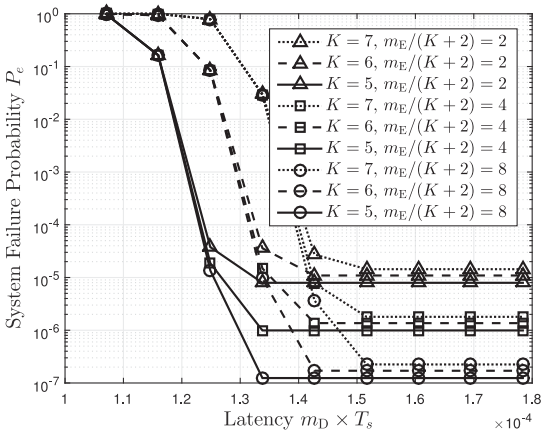

FIGURE 17. Novel reliability-latency trade-off curves.

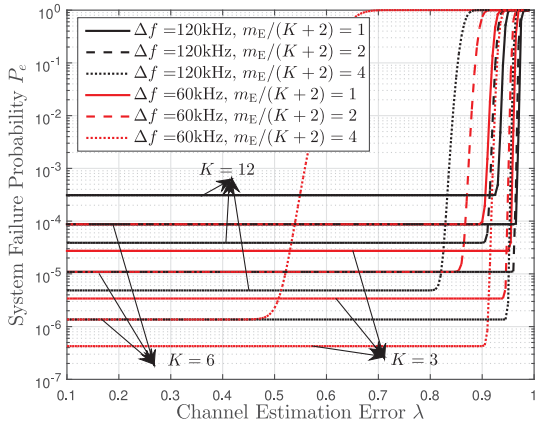

FIGURE 18. Failure probability of CGC based grant-free URLLC versus λ.

20, 30 and 40 respectively. Since the accessibility is related to the threshold of reliability, the simulation configuration is similar with that in Fig. 14. It can be observed that the retrievability mainly lies within the interval of $0.9968 \sim 0.9993$ under $m_E\ (K+2) = 4$ and within the interval of $0.9996 \sim 0.9998$ under $m_E\ (K+2) = 8$. At each interval, there always exist retrievability-accessibility trade-off curves indicating that more accessibility will bring less retrievability. In Fig. 16, $m_E\ (K+2)$ is respectively selected to be 1, 2, 4 and 8. We only focus on the latency caused by PA and leave the remnant OFDM symbols within 1ms for data transmission and other processes. Since N_E is fixed, the increase of the number of occupied OFDM symbols would be the only way to improve the retrievability of PA. In this way, flexible latency control by scheduling OFDM symbols could tune the redundancy among different dimensions and make grant-free URLLC work well even when N_E is constrained. The direct result is that the retrievability of PA would change with the number of occupied OFDM symbols.

Fig. 17 depicts the reliability-latency trade-off curves. $\Delta f = 120\,\text{kHz}$ and $T_s = 8.92 \times 10^{-6}$s are configured with $N_T = 120$ and $N_E = 128$. K is configured from 5 to 7 and m_D is selected from $12 \times T_s$ to $20 \times T_s$. $m_E\ (K+2)$ is chosen to be 2, 4, and 8 respectively. It can be observed that the transmission latency for data processing should not be lower than a threshold if the reliability of 10^{-5} is expected to realize. But increasing the occupied OFDM symbols too much would also bring no further benefits. Compared with the choice of increasing latency and that of decreasing the number of ALUs, improving the number of occupied OFDM symbols to some extent could be much better.

Fig. 18 presents the curve of failure probability of CGC based grant-free URLLC versus channel estimation error λ under $N_T = 120$ and $N_E = 128$. $m_E\ (K+2)$ is configured to be 1, 2, and 4. $T_s = 17.84 \times 10^{-6}$s and $T_s = 8.92 \times 10^{-6}$s are respectively considered with K as 3, 6, 12. In this simulation, m_D is determined by the equation $m_D = (T_{con} - m_E T_s - T_{extra})\ T_s$. As we can see, P_e increases with the increase of λ when λ lies above a certain threshold. This threshold changes with the variations of system parameters, like K and $m_E\ (K+2)$. With the increase of the number of antennas, the fluctuation of thresholds comes to be less sensitive to the variations of system parameters. Moreover, since channel estimation errors can not be eliminated completely in large spatial dimensions, more ALUs would introduce more intra-user interference and disturbance caused by imprecise channel estimation.

VII. CONCLUSION

In this paper, we developed a CGC theory to encode multi-dimensional time-frequency resources within BWP for PA in grant-free URLLC systems. A unified and theoretical framework modeling TRP was formulated to achieve secure PA in current 5G NR protocol. The design created a CGC based grant-free URLLC system operating with UAD, TRP, channel estimation and data transmission. Three main metrics, including retrievability, reliability and accessibility, are formulated respectively with analytical closed-form solutions to characterize the overall system performance. On this basis, we derived four performance trade-offs: retrievability-latency, retrievability-accessibility, reliability-latency and reliability-accessibility. With the help of those hints, we concluded that the design of this system is feasible, with controllable performance benefits.

REFERENCES

[1] 3GPP, TS22.261 v16.1.0, "Servic requirements for the 5G system," Accessed on: Sep. 27, 2019. [Online]. Available: https://www.3gpp.org/ftp/specs/archive/22_series/22.261/

[2] P. Popovski, "Ultra-reliable communication in 5G wireless systems," in *Proc. IEEE Int. Conf. 5G Ubiq. Connect.*, Nov. 2014, pp. 146–151.

[3] P. Popovski *et al.*, "Wireless access in ultra-reliable low-latency communication(URLLC)," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5783–5801. Aug. 2019.

[4] 3GPP, TS38.300 v 15.6.0, "NR; NR and NG-RAN overall desciption," Accessed on: Jan. 8, 2020. [Online]. Available: https://www.3gpp.org/ftp/specs/archive/38_series/300/

[5] C. Wang. Y. Chen, Y. Wu, and L. Zhang, "Performance evaluation of grant-free transmission for uplink URLLC services," in *Proc. IEEE VTC Spring*, Jun. 2017, pp. 1–6.

[6] K. Au *et al.*, "Uplink contention based SCMA for 5G radio access," in *Proc. IEEE Globecom Workshop*, Dec. 2014, pp. 900–905.

[7] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee, and B. Shim, "Ultra-reliabie and low-latency communications in 5g downlink: Physical layer aspects," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 124–130, Jun. 2018.

[8] 3GPP, TS33.501 v15.1.0, "Security architecture and procedures for 5G System," Accessed on: Sep. 25, 2019. [Online]. Available: https://www.3gpp.org/ftp/specs/archive/33_series/33.501/

[9] U. M. Maurer, "Authentication theory and hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1350–1356, Jul. 2000.

[10] Y.-S. Shiu, S.-Y. Cheng, H.-C.Wu, S.-H. Huang, and H.-H. Chen, "Physical layer security in wireless,networks: A tutorial," *IEEE Wireless Commun.*, vol. 18, no. 2, pp. 66–74, Apr. 2011.

[11] P. L. Yu, J. S. Baras, and B. M. Sadler, "Physical-layer authentication," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 1, pp. 38–51, Mar. 2008.

[12] 3GPP, TS 36.331 v15.7.0, "NR; Radio Resource Control (RRC) protocol specification," Accessed on: Apr. 6, 2020. [Online] Available: https://www.3gpp.org/ ftp/specs/archive

[13] S. Sesia, I. Toufik, and M. Baker, *LTE-The UMTS Long Term Evolution: From Theory to Practice.* New York: Hoboken, NJ, USA: Wiley, 2009.

[14] C. Shahriar *et al.*, "PHY-Layer resiliency in OFDM communications: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 292–314, Aug. 2015.

[15] D. Xu, P. Ren, and J. A. Ritsey, "Independence-checking coding for OFDM channcl training authentication: Protocol design, security, stability, and tradeoff analysis," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 22, pp. 387–402, Feb. 2019.

[16] R. Chen, C. Li, S. Yan, R. Malaney, and J. Yuan, "Physical layer security for ultra-reliable and low-latency communications," Accessed on: Jun. 2019. [Online]. Available: https://arxiv.org/abs/1906.08443

[17] L. Xiao, L. J. Greenstein, N. B. Mandayam, and W. Trappe, "Channel-based spoofing detection in frequency-selective Rayleigh channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 5948–5956, Dec. 2009.

[18] Z. Jiang, J. Zhao, X.-Y. Li, J. Han, and W. Xi, "Rejecting the attack: Source authentication for Wi-Fi management frames using CSI Information," in *Proc. IEEE INFOCOM*, Apr. 2013, pp. 2544–2552.

[19] D. Kapetanovic, G. Zheng, K.-K. Wong, and B. Ottersten, "Detection of pilot contamination attack using random training and massive MIMO," in *Proc. IEEE Int. Symp. Pers., Indoor Mobile Radio Commun.*, Sep. 2013, pp. 13–18.

[20] Q. Xiong, Y.-C. Liang, K. H. Li and Y. Gong, "An energy-ratio-based approach for detecting pilot spoofing attack in multiple-antenna systems," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 5, pp. 932–940, May 2015.

[21] D. Kapetanovic, A. Al-Nahari, A. Stojanovic, and F. Rusek, "Detection of active eavesdroppers in massive MIMO," in *Proc. IEEE Int. Symp. Pers. Indoor Mobile Radio Commun.*, Sep. 2014, pp. 585–589.

[22] J. M. Kang, C. In, and H. M. Kim, "Detection of pilot contamination attack for multi-antenna based secrecy systems," in *Proc. IEEE Veh. Technol. Conf.*, May 2015, pp. 1–5.

[23] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing-The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[24] C. Shahriar and T. C. Clancy, "Performance impact of pilot tone randomization to mitigate OFDM jamming attacks," in *Proc. IEEE Consum. Commun. Netw. Conf.*, Jan. 2013, pp. 813–816.

[25] D. Xu, P. Ren, and J. A. Ritcey, "Hierarchical 2-D feature coding for secure pilot authentication in multi-user multi-antenna OFDM systems: A reliability bound contraction perspective," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 3, pp. 592–607, Mar. 2019.

[26] M. Lichtman, J. D. Poston, S. Amuru, C. Shahriar, T. C. Clancy, R. M. Buchrer, and J. H. Reed, "A communications jamming taxonomy," *IEEE Security Privacy*, vol. 14, no. 1, pp. 47–54, Jan. 2016.

[27] T. C. Clancy and N. Georgen, "Security in cognitive radio networks: Threats and mitigations," in *Proc. IEEE 3rd Int., Conf. Cognitive Radio Oriented Wireless Netw. Commun. (CrownCom).*, May 2008, pp. 1–8.

[28] T. C. Clancy, "Efficient OFDM denial: Pilot jamming and pilot nulling," in *Proc. IEEE Int., Conf. Commun.*, Jun. 2011, pp. 1–5.

[29] P. L. Yu, J. S. Baras, and B. M. Sadler, "Multicarrier authentication at the physical layer," in *Proc. Int. Symp. World Wireless, Mobile Mulrimedia Netw.*, 2008, pp. 1–6.

[30] M. Wen, E. Basar, Q. Li, B. Zheng, and M. Zhang, "Multiple-mode orthogonal frequency division multiplexing with index modulation," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3892–3906, Sep. 2017.

[31] X. Cheng, M. Zhang, M. Wen, and L. Yang, "Index modulation for 5G: Striving to do more with less," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 126–132, Apr. 2018.

[32] J. M. Hamamreh, E. Basar, and H. Arslan, "OFDM-subcarrier index selection for enhancing security and reliability of 5G URLLC services," *IEEE Access*, vol. 5, pp. 25863–25875, 2017.

[33] D. Xu, P. Ren, Y. Wang, Q. Du, and L. Sun, "ICA-SBDC: A channel estimation and identification mechanism for MISO-OFDM systems under pilot spoofing attack," in *Proc. IEEE Int. Conf. Commun.*, May 2017, pp. 1–5.

[34] D. Xu, P. Ren, Y. Wang, and J. A. Ritcey, "Fundamental tradeoffs in uplink grant-free multiple access with protected CSI," Accessed on: Sep. 2019. [Online]. Available: https://arxiv.org/abs/1909.01521

[35] J. Jeon, "NR wide bandwidth operations," *IEEE Commun. Mag.*, vol. 56, no. 3. pp. 42–46, Mar. 2018.

[36] 3GPP, TS 38.211 v15.2.0, "Physical channels and modulation," Accessed on: Jan. 11, 2020. [Online]. Available: https://www.3gpp.org/ftp/specs/archive/38_series/38.211/

[37] G. Berardinelli *et al.*, "Reliability analysis of uplink grant-free transmission over shared resources," *IEEE Access*, vol. 6, pp. 23602–23611, 2018.

[38] H. Kobeissi, A. Nafkha, Y. Nasser, O. Bazzi, and Y. Louët, "Simple and accurate closed-form approximation of the standard condition number distribution with application in spectrum sensing," in *Proc. Int. Conf. Cognitive Radio Oriented Wireless Netw.*, May, 2016, pp. 351–362.

[39] W. H. Kautz and R. C. Singleton, "Nonrandom binary superimposed codes," *IEEE Trans. Inf. Theory*, vol. 10, no. 4, pp. 363–377, Oct. 1964.

[40] Z. D. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Berlin, Germany: Springer Series in Statistics, 2009.

DONGYANG XU (Member, IEEE) received the Ph.D. degree in information and communications engineering from Xi'an Jiaotong University in 2019 and B.S. degree in communications engineering from Zhengzhou University in 2013. Currently, he is an assistant professor in School of Information and Communications Engineering in Xi'an Jiaotong University in China. From January 2017 to January 2018, He was a Visiting Student in the Network Security Lab (NSL) at Department of Electrical Engineering, University of Washington, Seattle, USA. His research interests include computer and network security, wireless communication theory, graph signal processing and machine learning. He received the Best Paper Rewards from IEEE CHINA COMMUNICATIONS in 2017. He also served as the Technical Program Committee Member for IEEE/CIC ICCC in 2017. He is a Member of IEEE and IEEE Communications Society.

PINYI REN (Member, IEEE) received the B.S. degree in information and control engineering, the M.S. degree in information and communications engineering, and the Ph.D. degree in electronic and communications system from Xi'an Jiaotong University, Xi'an, China, in 1994, 1997, and 2001, respectively.

He is currently a Professor with the Information and Communications Engineering Department, Xi'an Jiaotong University. He has authored or coauthored over 100 technical papers on international journals and conferences. He has over 30 Patents (First Inventor) authorized by Chinese Government.

Dr. Ren was the recipient of the Best Letter Award of IEICE Communications Society 2010. He serves as an Editor for the *Journal of Xian Jiaotong University* and for the *Journal of Electronics and Information Technology*, and has served as the Leading Guest Editor for the special issue of *Mobile Networks and Applications* on "Distributed Wireless Networks and Services" and the Leading Guest Editor for the special issues of *Journal of Electronics* on "Cognitive Radio." He has served as the General Chair of ICST WICON 2011 and frequently serves as the Technical Program Committee members of IEEE GLOBAL COMMUNICATIONS CONFERENCE (GLOBECOM), IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS (ICC), IEEE CONSUMER COMMUNICATIONS AND NETWORKING CONFERENCE (CCNC), etc. He is a Member of IEEE Communications Society.