EURASIP Journal on Wireless Communications and Networking

RESEARCH Open Access

Authenticated secret key generation in delay-constrained wireless systems

Miroslav Mitev[1]*, Arsenia Chorti[2], Martin Reed[1] and Leila Musavian[1]

*Correspondence:
mm17217@essex.ac.uk
[1]School of CSEE, University of Essex, Colchester, UK
Full list of author information is available at the end of the article

Abstract
With the emergence of 5G low-latency applications, such as haptics and V2X, low-complexity and low-latency security mechanisms are needed. Promising lightweight mechanisms include physical unclonable functions (PUF) and secret key generation (SKG) at the physical layer, as considered in this paper. In this framework, we propose (i) a zero round trip time (0-RTT) resumption authentication protocol combining PUF and SKG processes, (ii) a novel authenticated encryption (AE) using SKG, and (iii) pipelining of the AE SKG and the encrypted data transfer in order to reduce latency. Implementing the pipelining at PHY, we investigate a *parallel* SKG approach for multi-carrier systems, where a subset of the subcarriers are used for SKG and the rest for data transmission. The optimal solution to this PHY resource allocation problem is identified under security, power, and delay constraints, by formulating the subcarrier scheduling as a subset-sum $0-1$ knapsack optimization. A heuristic algorithm of linear complexity is proposed and shown to incur negligible loss with respect to the optimal dynamic programming solution. All of the proposed mechanisms have the potential to pave the way for a new breed of latency aware security protocols.

Keywords: Physical layer security, Secret key generation, Physical unclonable functions, Resumption protocols, Effective capacity, QoS, Wireless communications, 5G applications

1 Introduction

Many standard cryptographic schemes, particularly those in the realm of public key encryption (PKE), are computationally intensive, incurring considerable overheads and can rapidly drain the battery of power-constrained devices [1, 2], notably in Internet of Things (IoT) applications [3]. For example, a 3GPP report on the security of ultra-reliable low-latency communication (URLLC) systems notes that authentication for URLLC is still an open problem [4]. Additionally, traditional public key generation schemes are not *quantum secure*—in that when sufficiently capable quantum computers will be available, they will be able to break current known PKE schemes—unless the key sizes increase to impractical lengths.

In the past years, physical layer security (PLS) [5–9] has been studied as a possible alternative to classic, complexity-based, cryptography. As an example, signal properties as in [10] can be exploited to generate opportunities for confidential data transmission

[11, 12]. Notably, PLS is explicitly mentioned as a 6G enabling technology in the first white paper on 6G [13]: "The strongest security protection may be achieved at the physical layer." In this work, we propose to move some of the security core functions down to the physical layer, exploiting both the communication radio channel and the hardware, as unique entropy sources.

Since the wireless channel is reciprocal, time-variant and random in nature, it offers a valid, inherently secure source that may be used in a key agreement protocol between two communicating parties. The principle of secret key generation (SKG) from correlated observations was first studied in [14] and [15]. A straightforward SKG approach can be built by exploiting the reciprocity of the wireless fading coefficients between two terminals within the channel coherence time[16], and this paper builds upon this mechanism. This is pertinent to many forthcoming B5G applications that will require a strong, but nevertheless, lightweight security key agreement; in this direction, PLS may offer such a solution or complement existing algorithms. With respect to authentication, physical unclonable functions (PUFs), firstly introduced in [17] (based on the idea of physical one-way functions [18]), [19] could also enhance authentication and key agreement in demanding scenarios, including (but not limited to) device to device and tactile Internet. We note that others also point to using physical layer security to reduce the resource overhead in URLLC [20].

A further advantage of PLS is that it is information-theoretic secure [21], i.e., it is not open to attack by future quantum computers and it requires lower computation costs as will be explored later in this paper. In this work, we will discuss how SKG from shared randomness [22] is a promising alternative to PKE for key agreement. However, unauthenticated key generation is vulnerable to man in the middle (MiM) attacks. In this sense, PUFs can be used in *conjunction* with SKG to provide authenticated encryption (AE). As summarized in [19], the employment of PUFs can decrease the computational cost and have a high impact on reducing the authentication latency in constrained devices.

In this study, we introduce the joint use of PUF authentication and SKG in a zero round trip time (0-RTT) [23, 24] approach, allowing to build quick authentication mechanisms with forward security. Further, we develop an AE primitive [25–27] based on standard SKG schemes. To investigate a fast implementation of the AE SKG, we propose a pipelined (*parallel*) scheduling method for optimal resource allocation at the physical layer (PHY) (i.e., by optimal allocation of the subcarriers in 5G resource blocks).

Next, we extend the analysis to account for statistical delay quality of service (QoS) guarantees, a pertinent scenario in B5G. The support of different QoS guarantee levels is a challenging task. In fact, in time-varying channels, such as in wireless networks, determining the exact delay bound depending on the users' requirements is impossible. However, a practical approach, namely the effective capacity[28], can provide statistical QoS guarantees and can give delay bounds with a small violation probability. In our work, we employ the effective capacity as the metric of interest and investigate how the proposed pipelined AE SKG scheme performs in a delay-constrained scenario.

The system model introduced in this work assumes that a block fading additive white Gaussian noise (BF-AWGN) channel is used with multiple orthogonal subcarriers. In our *parallel* scheme, a subset of the subcarriers is used for SKG and the rest for encrypted data transfer. The findings of this paper are supported by numerical results, and the efficiency of the proposed *parallel* scheme is shown to be greater or similar to the efficiency of an alternative approach in which SKG and encrypted data transfer are sequentially performed.

The contributions of this paper are as follows:

1. We combine an initial PUF authentication and SKG for resumption key derivation in a single 0-RTT protocol.
2. We develop an AE SKG scheme.
3. We propose a fast implementation of the AE SKG based on pipelining of key generation and encrypted data transfer. This *parallel* approach is achieved by allocation of the PHY resources, i.e., by optimal scheduling of the subcarriers in BF-AWGN channels.
4. We propose a heuristic algorithm of linear complexity that finds the optimal subcarrier allocation with negligible loss in terms of efficiency.
5. We numerically compare the efficiency of our *parallel* approach with a *sequential* approach where SKG and data transfer are performed sequentially. This comparison is performed in two delay scenarios:
   - When a relaxed QoS delay constraint is in place;
   - When a stringent QoS delay constraint is in place.

A roadmap of the paper's contributions is shown in Fig. 1.

The paper is organized as follows: related work is discussed in Section 2 followed by a brief summary of the methods used within this paper in Section 3, and then the general system model is introduced in Section 4. The use of PUF authentication is illustrated in Section 4.1, the baseline SKG in Section 4.2; next, in Sections 4.3 and 4.4, we present an AE scheme using SKG and a resumption scheme to build a 0-RTT protocol. Subsequently, we evaluate the optimal power and subcarrier allocation at PHY considering both the long-term average rate in Section 5 and the effective rate in Section 6. In Section 7, the efficiency of the proposed approach is evaluated against that of a sequential approach, while conclusions are presented in Section 8.

2 Related work

This paper assumes the use of PUF-based authentication with SKG. PUFs are hardware entities based on the physically unclonable variations that occur during the production process of silicon. These unique and unpredictable variations allow the extraction of uniformly distributed binary sequences. Due to their unclonability and simplicity, PUFs are

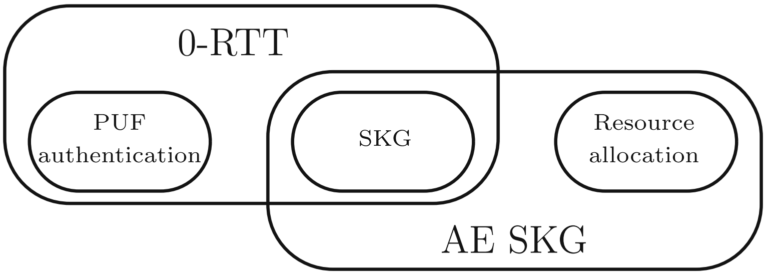


Fig. 1 Roadmap of contributions

seen as lightweight security primitives that can offer alternatives to today's authentication mechanisms. Furthermore, employing PUFs can eliminate the need of non-volatile memory, which reduces cost and complexity [29]. Common ways of extracting secret bit sequences are through measuring delays on wires and gates or observing the power up behavior of a silicon.

Focusing on that, numerous PUF architectures have been proposed for IoT applications in the literature. A few of these architectures are as follows: arbiter PUF [30], ring oscillator PUF [17], transient effect ring oscillator PUF [31], static random access memory PUF [32], hardware-embedded delay PUF [33], and more [34]. Utilizing these basic properties, many PUF-based authentication protocols have been proposed, both for unilateral authentication [35, 36] and mutual authentication [29, 36–38]. A comprehensive survey on lightweight PUF authentication schemes is presented by Delvaux et al. [39].

On the other hand, due to the nature of propagation in a shared free-space, wireless communication remains vulnerable to different types of attacks. Passive attacks such as eavesdropping or traffic analysis can be performed by anyone in the vicinity of the communicating parties; to ensure confidentiality, data encryption is vital for communication security. The required keys can be agreed at PHY using SKG. In this case, all pilot exchanges need to take place over the coherence time of the channel[1], during which Alice and Bob can observe highly correlated channel states that can be used to generate a shared secret key between them. SKG has been implemented and studied for different applications such as vehicular communications [42, 43], underwater communications [44], optical fiber [45], visible light communication [46], and more as summarized in [47]. The key conclusion from these studies is that SKG shows promise as an important alternative to current key agreement schemes.

Widely used sources of shared randomness used for SKG are the received signal strength (RSS) and the full channel state information (CSI) [48]. In either case, it is important to build a suitable pre-processing unit to decorrelate the signals in the time/frequency and space domains. As an example, some recent works have shown that the widely adopted assumption [49] that a distance equal to *half* of the wavelength (which at 2.4 GHz is approximately 6 cm [50]) is enough for two channels to decorrelate may not hold in reality [40]. Other works show that the mobility can highly increase the entropy of the generated key [51, 52] while an important issue with the RSS-based schemes is that they are open to predictable channel attacks[40, 53]. These important issues need to be explicitly accounted for in actual implementations, but fall outside the scope of this paper.

3 Methods

The methods used and introduced in this paper rely upon a range of basic primitives, which in combination provide the full PUF and AE SKG solution. Each of these primitives is introduced below together with a summary of the methods used to analyze and optimize the solution.

***Authentication*:** Before establishing a shared secret key, Alice and Bob must be sure they are communicating with a trusted party. To achieve this, we assume the usage of

[1] The coherence time corresponds to the interval during which the multipath properties of wireless channels (channel gains, signal phase, delay) remain stable [40–42]. It is inversely proportional to the Doppler spread, which, on the other hand, is a dispersion metric that accounts for the spectral broadening caused by the user's mobility (for more details and derivation, please see [41]).

a PLS method, more specifically PUF authentication. As discussed in Section 2 by eliminating the need of non-volatile memory, the usage of PUFs could greatly reduce the complexity compared to existing authentication alternatives.

***Secret key generation*:** To ensure that their communication is private, after authenticating each other, Alice and Bob have to encrypt/decrypt the data. For this work, we assume the use of symmetric encryption where the same key is used for both operations. In order to obtain a shared key, we propose to use SKG which consists of three standard steps: (i) advantage distillation, (ii) information reconciliation, and (iii) privacy amplification; each of these steps is explained in more detail in Section 4.2.

***Re-authentication*:** We present a re-authentication approach that exploits the use of resumption secrets as used in 0-RTT protocols. Instead of performing full authentication before sending data encrypted with a new key, we propose a new method which allows Alice (Bob) to authenticate subsequent keys using a lightweight scheme anchored by the initial authentication process.

***Authenticated encryption SKG*:** To eliminate the possibility of tampering attacks, we build on the SKG process to introduce a new AE SKG method. AE can simultaneously guarantee confidentiality and message integrity. In our AE SKG method, side information and encrypted data transfer are pipelined.

***Pipelined transmission*:** In our proposal, the key generation is pipelined with the encrypted data transfer, i.e., side information and data encrypted with the key that corresponds to the side information are transmitted over the same 5G resource block(s).

***Joint PHY/MAC delay analysis*:** To analyze the system under statistical QoS delay constraints, we use the theory of *effective capacity* [28] and analyze the scheme's *effective rate*.

***Optimization methods*:** Finally, to optimize the pipelined transmission, we take into consideration practical wireless aspects such as the impact of imperfect CSI measurements and formulate two optimization problems to find the optimal resource allocation for Alice and Bob. To solve these problems, we employ tools such as combinatorial optimization, dynamic programming, order statistics, and convex optimization.

3.1 Threat model

In this paper, we assume a commonly used adversarial model with an active man-in-the-middle attacker (Eve) and a pair of legitimate users (Alice and Bob). For simplicity, we assume a rich Rayleigh multipath environment where the adversary is more than a few wavelengths away from each of the legitimate parties. This forms the basis of our hypothesis that the measurements of Alice and Bob are uncorrelated to the Eve's measurements.

3.2 Notation

Random variables are denoted in italic font, e.g., x, and vectors and matrices are denoted with lower and upper case bold characters, e.g., $\mathbf{x}$ and $\mathbf{X}$, respectively. Functions are printed in a fixed-width teletype font, e.g., $\texttt{F}$. All sets of vectors are given with calligraphic font $\mathcal{X}$, and the elements within a set are given in curly brackets, e.g., $\{\mathbf{x}, \mathbf{y}\}$, the cardinality of a vector or set is defined by vertical lines, e.g., $|\mathbf{x}|$ or $|\mathcal{X}|$. Concatenation and bit-wise

XORing are represented as $[\mathbf{x}||\mathbf{y}]$ and $\mathbf{x} \oplus \mathbf{y}$, respectively. We use H to denote entropy, I mutual information, $\mathbb{E}$ expectation, and $\mathbb{C}$ the set of complex numbers.

4 Node authentication using PUFs and SKG

In this section, we present a joint physical layer SKG and PUF authentication scheme. To the best of our knowledge, this is the first work that proposes the utilization of the two schemes in conjunction. As discussed in Section 2, many PUF authentication protocols have been proposed in the literature, with even a few commercially available [54, 55]. We do not look into developing a new PUF architecture or a new PUF authentication protocol; instead, we look at combining existing PUF mechanisms with SKG. In addition, we develop an AE scheme that can prevent tampering attacks. To further develop our hybrid cryptosystem, we propose a resumption type of authentication protocol, inspired by the 0-RTT authentication mode in the transport layer security (TLS) 1.3 protocol. The resumption protocol is important as it significantly reduces the use of the PUF to the initial authentication, thus, overcoming the limitation of a PUFs' challenge response space [34, 56].

4.1 Node authentication using PUFs

As discussed in Section 3, for security against MiM attacks, the SKG needs to be protected through authentication. While existing techniques, such as the extensible authentication protocol-transport layer security (EAP-TLS), could be used as the authentication mechanism, these are computationally intensive and can lead to significant latency [57, 58].

This leads to the motivation to seek lightweight authentication mechanisms that can be used in conjunction with SKG. Such a mechanism that is achieving note within the research community uses a PUF. The concept of a PUF was first introduced in [17]; its idea is to utilize the fact that every integrated circuit differs to others due to manufacturing variability [59, 60] and cannot be cloned [61]. Having these characteristics, a PUF can be used in a challenge-response scheme, where a challenge can refer to a delay at a specific gate, power-on state, etc.

A typical PUF-based authentication protocol consists of two main phases, namely *enrolment phase* and *authentication phase* [62–66]. During the *enrolment phase*, each node runs a set of challenges on its PUF and characterizes the variance of the measurement noise in order to generate side information. Next, a verifier creates and stores a database of all challenge-response pairs (CRPs) for each node's PUF within its network. A CRP pair in essence consists of an authentication key and related side information. Within the database, each CRP is associated with the ID of the corresponding node.

Later, during the *authentication phase*, a node sends its ID to the verifier requesting to start a communication. Receiving the request, the verifier checks if the received ID exists in its database. If it does, the verifier chooses a random challenge that corresponds to this ID and sends it to the node. The node computes the response by running the challenge on its PUF and sends it to the verifier. However, the PUF measurements at the node are never exactly the same due to measurement noise; therefore, the verifier uses the new PUF measurement and the side information stored during the enrollment to re-generate the authentication key. Finally, the verifier compares the re-generated key to the one in the CRP, and if they are identical, the authentication of the node is successful. In order to prevent replay attacks, once used, a CRP is deleted from the verifier database.

In summary, the motivation for using a PUF authentication scheme in conjunction with SKG is to exclude all of the computationally intensive operations required by EAP-TLS, which use modulo arithmetic in large fields. Measurements performed on current public key operations within EAP-TLS on common devices (such as IoT) give average authentication and key generation times of approximately 160 ms in static environments and this can reach up to 336 ms in high mobility conditions [67].

On the other hand, PUF authentication protocols have very low computational overhead and require overall authentication times that can be less than 10 ms [63, 68]. Furthermore, our key generation scheme, proposed in Section 4.2, requires just a hashing operation and (syndrome) decoding. Hashing mechanisms such as SHA256 performed on an IoT device require less than 0.3ms [68, 69]. Regarding the decoding, if we assume the usage of standard LDPC or BCH error correcting mechanisms, even in the worst-case scenario with calculations carried out as software operations, the computation is trivial compared to the hashing and requires less computational overhead [70].

4.2 SKG procedure

The SKG system model is shown in Fig. 2. This assumes that two legitimate parties, Alice and Bob, wish to establish a symmetric secret key using the wireless fading coefficients as a source of shared randomness. Throughout our work, a rich Rayleigh multipath environment is assumed, such that the fading coefficients rapidly decorrelate over short distances [16]. Furthermore, Alice and Bob communicate over a BF-AWGN channel that comprises N orthogonal subcarriers. The fading coefficients $\mathbf{h} = [h_1, \ldots, h_N]$ are assumed to be independent and identically distributed (i.i.d), complex circularly symmetric zero-mean Gaussian random variables $h_j \sim \mathcal{CN}(0, \sigma^2), j = 1, \ldots, N$. Although in actual multicarrier systems neighboring subcarriers will typically experience correlated fading, in the present work, this effect is neglected as its impact on SKG has been treated in numerous contributions in the past [71–73] and will not enhance the problem formulation in the following sections.

The SKG procedure encompasses three phases: *advantage distillation*, *information reconciliation*, and *privacy amplification* [14, 15] as described below:

1. Advantage distillation: This phase takes place during the coherence time of the channel. The legitimate nodes sequentially exchange constant probe signals with power P on all subcarriers[2], to obtain estimates of their reciprocal CSI. We note in passing that the pilot exchange phase can be made robust with respect to injection type of attacks (that fall in the general category of MiM) as analyzed in [22, 74]. Commonly, the received signal strength (RSS) has been used as the source of shared randomness for generating the shared key, but it is possible to use the full CSI [75]. At the end of this phase, Alice and Bob obtain observation vectors $\mathbf{x}_A = [x_{A,1}, \ldots, x_{A,N}]$, $\mathbf{x}_B = [x_{B,1}, \ldots, x_{B,N}]$, respectively, so that:

$$\mathbf{x}_A = \sqrt{P}\mathbf{h} + \mathbf{z}_A, \tag{1}$$

$$\mathbf{x}_B = \sqrt{P}\mathbf{h} + \mathbf{z}_B, \tag{2}$$

[2] An explanation of the optimality of this choice under different attack scenarios is discussed in [22].

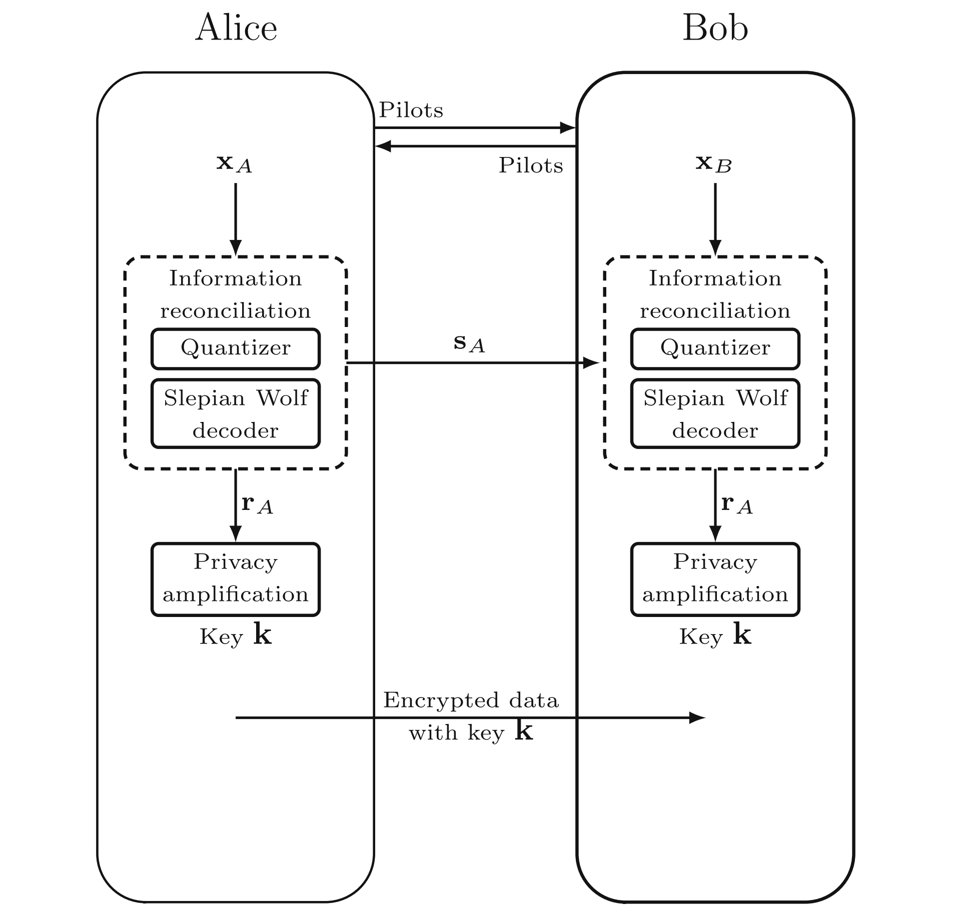


Fig. 2 Secret key generation between Alice and Bob

where $\mathbf{z}_A$ and $\mathbf{z}_B$ denote zero-mean, unit variance circularly symmetric complex AWGN random vectors, such that $(\mathbf{z}_A, \mathbf{z}_B) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{2N})$. On the other hand, Eve observes $\mathbf{x}_E = [x_{E,1}, \ldots, x_{E,N}]$ with:

$$\mathbf{x}_E = \sqrt{P}\mathbf{h}_E + \mathbf{z}_E. \tag{3}$$

Due to the rich Rayleigh multipath environment, Eve's channel measurement $\mathbf{h}_E$ is assumed uncorrelated to $\mathbf{h}$ and $\mathbf{z}_E$ denotes a zero-mean, unit variance circularly symmetric complex AWGN random vector $\mathbf{z}_E \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$.

2. Information reconciliation: At the beginning of this phase, the observations $x_{A,j}, x_{B,j}$ are quantized to binary vectors[3] $\mathbf{r}_{A,j}$, $\mathbf{r}_{B,j}$ $j = 1, \ldots, N$[76–78], so that Alice and Bob distill $\mathbf{r}_A = [\mathbf{r}_{A,1}|| \ldots ||\mathbf{r}_{A,N}]$ and $\mathbf{r}_B = [\mathbf{r}_{B,1}|| \ldots ||\mathbf{r}_{B,N}]$, respectively. Due to the presence of noise, $\mathbf{r}_A$ and $\mathbf{r}_B$ will differ. To reconcile discrepancies in the quantizer local outputs, side information needs to be exchanged via a public channel. Using the principles of Slepian-Wolf decoding, the distilled binary vectors can be expressed as

$$\mathbf{r}_A = \mathbf{d} + \mathbf{e}_A, \tag{4}$$

$$\mathbf{r}_B = \mathbf{d} + \mathbf{e}_B, \tag{5}$$

[3] Note that each observation can generate a multi-bit vector at the output of the quantizer.

where $\mathbf{e}_A, \mathbf{e}_B$ are error vectors that represent the distance from the common observed (codeword) vector $\mathbf{d}$ at Alice and Bob, respectively.

Numerous practical information reconciliation approaches using standard forward error correction codes (e.g., LDPC, BCH) have been proposed [16, 75]. As an example, if a block encoder is used, then the error vectors can be recovered from the syndromes $\mathbf{s}_A$ and $\mathbf{s}_B$ of $\mathbf{r}_A$ and $\mathbf{r}_B$, respectively. Alice transmits her corresponding syndrome to Bob so that he can reconcile $\mathbf{r}_B$ to $\mathbf{r}_A$. It has been shown that the length of the syndrome $|\mathbf{s}_A|$ is lower bounded by $|\mathbf{s}_A| \geq H(\mathbf{x}_A|\mathbf{x}_B) = H(\mathbf{x}_A, \mathbf{x}_B) - H(\mathbf{x}_B)$ [15]. This has been numerically evaluated for different scenarios and coding techniques [77, 79–81]. Following that, the achievable SKG rate is upper bounded by $I(\mathbf{x}_A; \mathbf{x}_B|\mathbf{x}_E)$.

3. Privacy amplification: The secret key is generated by passing $\mathbf{r}_A$ through a one-way collision resistant *compression* function i.e., by hashing. Note that this final step of privacy amplification is executed locally without any further information exchange. The need for privacy amplification arises in order to suppress the entropy revealed due to the public transmission of the syndrome $\mathbf{s}_A$. Privacy amplification produces a key of length strictly shorter than $|\mathbf{r}_A|$, at least by $|\mathbf{s}_A|$. At the same time, the goal is for the key to be uniform, i.e., to have maximum entropy. In brief, privacy amplification *reduces the overall output entropy* while at the same time *increases the entropy per bit*—compared to the input.

The privacy amplification is typically performed by applying either cryptographic hash functions such as those built using the Merkle-Damgard construction or universal hash functions and has been proven to be secure, in an information theoretic sense, through the leftover hash lemma [82]. As an example [40, 83] use a 2-universal hash family to achieve privacy amplification. Summarizing, the maximum key size after privacy amplification is:

$$|\mathbf{k}| \leq H(\mathbf{x}_A) - I(\mathbf{x}_A; \mathbf{x}_\mathbf{E}) - H(\mathbf{x}_A|\mathbf{x}_B) - r_0, \tag{6}$$

where $H(\mathbf{x}_A)$ represents the entropy of the measurement, $I(\mathbf{x}_A; \mathbf{x}_\mathbf{E})$ represents the mutual information between Alice's and Eve's observations, $H(\mathbf{x}_A|\mathbf{x}_B)$ represents the entropy revealed during information reconciliation, and $r_0 > 0$ is an extra security parameter that ensures uncertainty on the key at Eve's side. For details and estimation of these parameters in a practical scenario, please see [84].

As shown in this section, the SKG procedure requires only a few simple operations such as quantization, syndrome calculation, and hashing. In future work, we will examine the real possibilities of implementing such a mechanism in practical systems.

4.3 AE using SKG

To develop a hybrid cryptosystem that can withstand tampering attacks, SKG can be introduced in standard AE schemes in conjunction with standard block ciphers in counter mode (to reduce latency), e.g., AES GCM. As a sketch of such a primitive, let us assume a system with three parties: Alice who wishes to transmit a secret message $\mathbf{m}$ with size $|\mathbf{m}|$, to Bob with confidentiality and integrity, and Eve that can act as a passive and active attacker. The following algorithms are employed:

- The SKG scheme denoted by $\mathtt{G} : \mathbb{C} \rightarrow \mathcal{K} \times \mathcal{S}$, accepting as input the fading coefficients (modeled as complex numbers),and generating as outputs binary vectors $\mathbf{k}$ and $\mathbf{s}_A$ in the key and syndrome spaces, of sizes $|\mathbf{k}|$ and $|\mathbf{s}_A|$, respectively,

$$\mathtt{G}(\mathbf{h}) = (\mathbf{k}, \mathbf{s}_A)\,, \tag{7}$$

where $\mathbf{k} \in \mathcal{K}$ denotes the key obtained from $\mathbf{h}$ after privacy amplification and $\mathbf{s}_A$ is Alice's syndrome.

- A symmetric encryption algorithm, e.g., AES GCM, denoted by $\mathtt{Es} : \mathcal{K} \times \mathcal{M} \rightarrow \mathcal{C}_\mathcal{T}$ where $\mathcal{C}_\mathcal{T}$ denotes the ciphertext space with corresponding decryption $\mathtt{Ds} : \mathcal{K} \times \mathcal{C}_\mathcal{T} \rightarrow \mathcal{M}$, such that

$$\mathtt{Es}(\mathbf{k}, \mathbf{m}) = \mathbf{c}, \tag{8}$$

$$\mathtt{Ds}(\mathbf{k}, \mathbf{c}) = \mathbf{m}, \tag{9}$$

for $\mathbf{m} \in \mathcal{M}$, $\mathbf{c} \in \mathcal{C}_\mathcal{T}$.

- A pair of message authentication code (MAC) algorithms, e.g., in HMAC mode, denoted by $\mathtt{Sign} : \mathcal{K} \times \mathcal{M} \rightarrow \mathcal{T}$, with a corresponding verification algorithm $\mathtt{Ver} : \mathcal{K} \times \mathcal{M} \times \mathcal{T} \rightarrow (yes, no)$, such that

$$\mathtt{Sign}(\mathbf{k}, \mathbf{m}) = \mathbf{t}, \tag{10}$$

$$\mathtt{Ver}(\mathbf{k}, \mathbf{m}, \mathbf{t}) = \begin{cases} yes, & \text{if integrity verified} \\ no, & \text{if integrity not verified} \end{cases} \tag{11}$$

A hybrid crypto-PLS system for AE SKG can be built as follows:

1. The SKG procedure is launched between Alice and Bob generating a key and a syndrome $\mathtt{G}(\mathbf{h}) = (\mathbf{k}, \mathbf{s}_A)$.
2. Alice breaks her key into two parts $\mathbf{k} = \{\mathbf{k}_e, \mathbf{k}_i\}$ and uses the first to encrypt the message as $\mathbf{c} = \mathtt{Es}(\mathbf{k}_e, \mathbf{m})$. Subsequently, using the second part of the key, she signs the ciphertext using the signing algorithm $\mathbf{t} = \mathtt{Sign}(\mathbf{k}_i, \mathbf{c})$ and transmits to Bob the extended ciphertext $[\mathbf{s}_A \| \mathbf{c} \| \mathbf{t}]$, as it is depicted in Fig. 3.
3. Bob checks first the integrity of the received ciphertext as follows: from $\mathbf{s}_A$ and his own observation he evaluates $\mathbf{k} = \{\mathbf{k}_e, \mathbf{k}_i\}$ and computes $\mathtt{Ver}(\mathbf{k}_i, \mathbf{c}, \mathbf{t})$. The integrity test will fail if any part of the extended ciphertext was modified, including the syndrome (that is sent as plaintext); for example, if the syndrome was modified during the transmission, then Bob would not have evaluated the correct key and the integrity test would have failed.
4. If the integrity test is successful then Bob decrypts $\mathbf{m}$=$\mathtt{Ds}(\mathbf{k}_e, \mathbf{c})$.

4.4 Resumption protocol

In Section 4.1 we discussed that using PUF authentication can greatly reduce the computational overhead of a system. Authentication of new keys is required at the start of communication and at each key renegotiation. However, the number of challenges that can be applied to a single PUF is limited. Due to that, we present a solution that is inspired by the 0-RTT authentication mode introduced in the 1.3 version of the TLS [23]. The use of 0-RTT obviates the need of performing a challenge for every re-authentication through the use of a resumption secret $\mathbf{r}_s$, thus reducing latency. Another strong motivation for using this mechanism is that it is forward secure in the scenario we are using here [24]. We first briefly describe the TLS 0-RTT mechanism before describing a similarly inspired 0-RTT mechanism applied to the information reconciliation phase of our SKG mechanism.

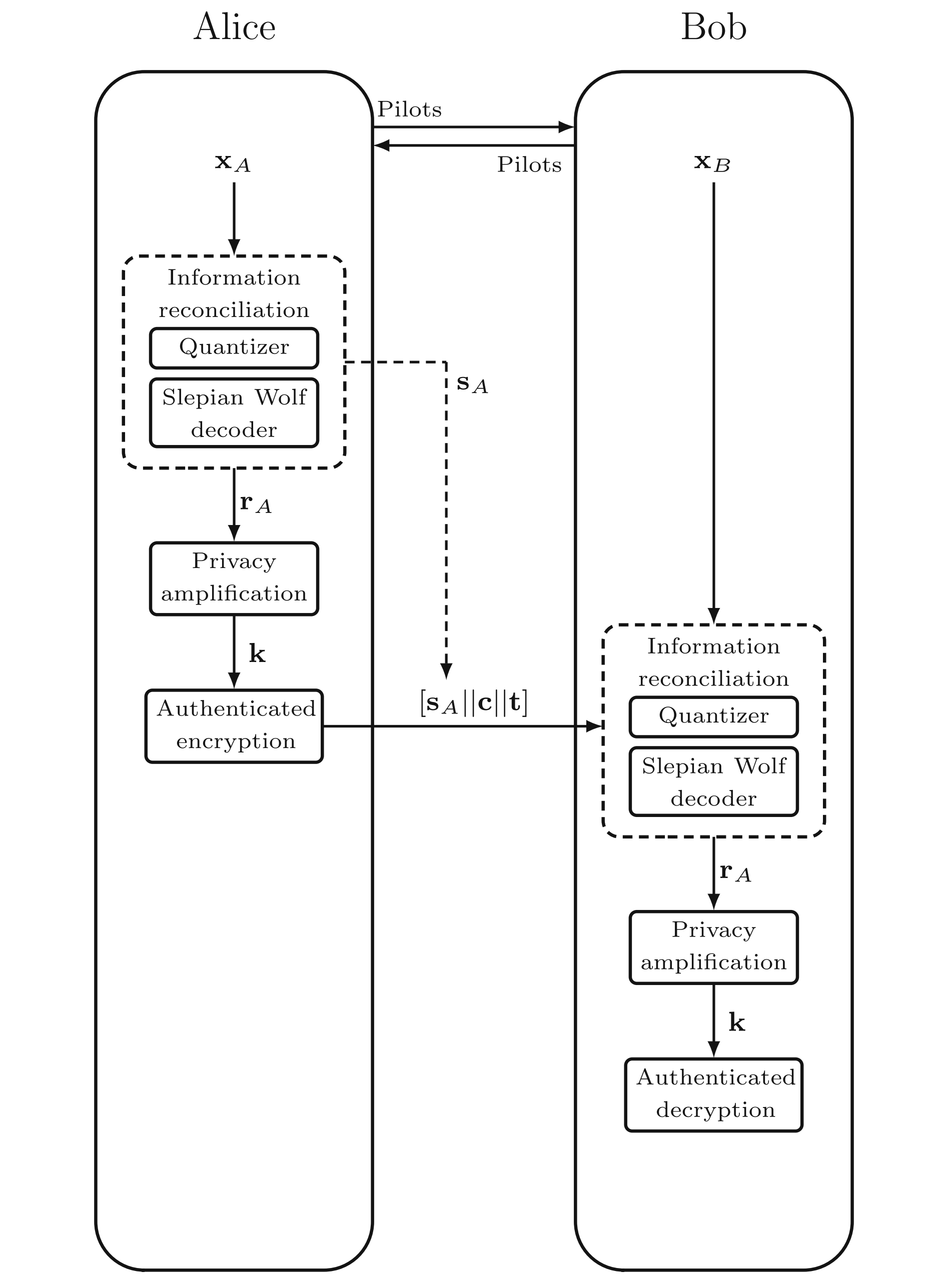


Fig. 3 Pipelined SKG and encrypted data transfer between Alice and Bob

The TLS 1.3 0−RTT handshake works as follows: in the very first connection between client and server, a regular TLS handshake is used. During this step, the server sends to the client a look-up identifier $\mathbf{k}_l$ for a corresponding entry in session caches or it sends a session ticket. Then, both parties derive a resumption secret $\mathbf{r}_s$ using their shared key and the parameters of the session. Finally, the client stores the resumption secret $\mathbf{r}_s$ and uses it when reconnecting to the same server which also retrieves it during the re-connection.

If session tickets are used, the server encrypts the resumption secret using long-term symmetric encryption key, called a session ticket encryption key (STEK), resulting in a session ticket. The session ticket is then stored by the client and included in subsequent connections, allowing the server to retrieve the resumption secret. Using this approach, the same STEK is used for many sessions and clients. On one hand, this property highly reduces the required storage of the server; however, on the other hand, it makes it vulnerable to replay attacks and not forward secure. Due to these vulnerabilities, in this work, we focus on the session cache mechanism described next.

When using session caches, the server stores all resumption secrets and issues a unique look-up identifier $\mathbf{k}_l$ for each client. When a client tries to reconnect to that server, it includes its look-up identifier $\mathbf{k}_l$ in the 0-RTT message, which allows the server to retrieve the resumption secret $\mathbf{r}_s$. Storing a unique resumption secret $\mathbf{r}_s$ for each client requires server storage for each client but it provides forward security and resilience against replay attacks, when combined with a key generation mechanisms such as Diffie Hellman (or the SKG used in this paper) which are important goals for security protocols [24]. In our physical layer 0-RTT, given that a node identifier state would be required for link-layer purposes, the session cache places little comparative load and thus is the mechanism proposed here for (re-)authentication.

The physical layer resumption protocol modifies the information reconciliation phase of Section 3.1 following initial authentication to provide a re-authentication mechanism between Alice and Bob. At the first establishment of communication, we assume initial authentication is established, such as the mechanism shown in Section 4. During that, Alice sends to Bob a look-up identifier $\mathbf{k}_l$. Then, both derive a resumption secret $\mathbf{r}_s$ that is identified by $\mathbf{k}_l$. Note, $\mathbf{r}_s$ and the session key have the same length $|\mathbf{k}|$. Then, referring to the notation and steps in Sections 4, 4.2, and 4.3:

1. Advantage distillation phase is carried out as before (see Section 4.2), where both parties obtain channel observations and obtain the vectors $\mathbf{r}_A$ and $\mathbf{r}_B$.
2. During the information reconciliation phase, both Alice and Bob exclusive or the resumption secret $\mathbf{r}_s$ with their observations $\mathbf{r}_A$ and $\mathbf{r}_B$ and obtain syndromes $\mathbf{s}'_A$ and $\mathbf{s}'_B$ with which both parties can carry out reconciliation to obtain the same shared value which is now $\mathbf{r}_A \oplus \mathbf{r}_s$.
3. The privacy amplification step in Section 4.2 is carried out as before, but now,0 the hashing takes place on $\mathbf{r}_A \oplus \mathbf{r}_s$ to produce the final shared key $\mathbf{k}'$ that is a result of both the shared wireless randomness and the resumption secret.

Note that the key $\mathbf{k}'$ can only be obtained if both the physical layer generated key and the resumption key are valid and this method can be shown to be forward secure [24].

5 Pipelined SKG and encrypted data transfer

As explained in Section 4, if Alice and Bob follow the standard sequential SKG process, they can exchange encrypted data only after both of them have distilled the key at the end of the privacy amplification step. In this Section, we propose a method to pipeline the SKG and encrypted data transfer. Alice can unilaterally extract the secret key from her observation and use it to encrypt data transmitted in the same “extended” message that contains the side information (see Fig. 3). Subsequently, using the side information, Bob can distill the same key $\mathbf{k}$ and decrypt the received data in one single step.

We have discussed in Section 4.2 how Alice and Bob can distill secret keys from estimates of the fading coefficients in their wireless link and in Section 4.3 how these can be used to develop an AE SKG primitive. At the same time CSI estimates are prerequisite in order to optimally allocate power across the subcarriers and achieve high data rates[4]. As a result, a question that naturally arises is whether the CSI estimates (obtained at the end of the pilot exchange phase) should be used towards the generation of secret keys or towards the reliable data transfer and, furthermore, whether the SKG and the data transfer can be inter-woven using the AE SKG principle.

In this paper, we are interested in answering this question and shed light into whether following the exchange of pilots Alice should transmit reconciliation information on all subcarriers, so that she and Bob can generate (potentially) a long sequence of key bits or, alternatively, perform information reconciliation only over a subset of the subcarriers and transmit encrypted data over the rest, exploiting the idea of the AE SKG primitive. Note here that the data can be already encrypted with the key generated at Alice, the sender of the side information, so that the proposed pipelining does not require storing keys for future use. We will call the former approach a *sequential* scheme, while we will refer to the latter as a *parallel* scheme. The two will be compared in terms of their efficiency with respect to the achievable data rates.

A simplified version of this problem, where the reconciliation rate is roughly approximated to the SKG rate, was investigated in [85]. In this study, it was shown that in order to maximize the data rates in the *parallel* approach, Alice and Bob should use the strongest subcarriers—in terms of SNR—for data transmission and the worst for SKG. Under this simplified formulation, the optimal power allocation for the data transfer has been shown to be a modified water-filling solution.

Here, we explicitly account for the rate of transmitting reconciliation information and differentiate it from the SKG rate. We confirm whether the policy of using the strongest subcarriers for data transmission and not for reconciliation is still optimal when the full optimization problem is considered, including the communication cost for reconciliation.

As discussed in Section 4.2, our physical layer system model assumes Alice and Bob exchange data over a Rayleigh BF-AWGN channel with N orthogonal subcarriers. Without loss of generality, the variance of the AWGN in all links is assumed to be unity. During channel probing, constant pilots are sent across all subcarriers [16, 86] with power P. Using the observations (1), Alice estimates the channel coefficients as

$$\hat{h}_j = h_j + \tilde{h}_j, \tag{12}$$

for $j = 1, \ldots, N$ where $\tilde{h}_j$ denotes an estimation error that can be assumed to be Gaussian, $\tilde{h}_j \sim \mathcal{CN}\left(0, \sigma_e^2\right)$ [87]. Under this model, the following rate is achievable on the jth subcarrier from Alice to Bob when the transmit power during data transmission is p_j [87]:

$$R_j = \log_2\left(1 + \frac{g_j p_j}{\sigma_e^2 P + 1}\right) = \log_2(1 + \hat{g}_j p_j), \tag{13}$$

where we use $\hat{g}_i = \frac{g_i}{\sigma_{i,e}^2 P + 1}$, to denote the estimated channel gains. As a result, the channel capacity $C = \sum_{j=1}^{N} R_j$ under the short-term power constraint:

$$\sum_{j=1}^{N} p_j \leq NP, \quad p_j \geq 0, \ \forall j \in \{1, \ldots, N\}, \tag{14}$$

[4] As an example, despite the extra overhead, in URLLC systems advanced CSI estimation techniques are employed in order to be able to satisfy the strict reliability requirements.

is achieved with the well known water-filling power allocation policy $p_j = \left[\frac{1}{\lambda} - \frac{1}{\hat{g}_j}\right]^+$, where the water level λ is estimated from the constraint (14). In the following, the estimated channel gains $\hat{g}_j$ are—without loss of generality—assumed ordered in descending order, so that:

$$\hat{g}_1 \geq \hat{g}_2 \geq \ldots \geq \hat{g}_N. \tag{15}$$

As mentioned above, the advantage distillation phase of the SKG process consists of the two-way exchange of pilot signals during the coherence time of the channel to obtain $\mathbf{r}_{A,j}, \mathbf{r}_{B,j}, j = 1, \ldots, N$. On the other hand, the CSI estimation phase can be used to estimate the reciprocal channel gains in order to optimize data transmission using the water-filling algorithm. In the former case, the shared parameter is used for generating symmetric keys, in the latter for deriving the optimal power allocation. In the parallel approach, the idea is to inter-weave the two procedures and investigate whether a joint encrypted data transfer and key generation scheme as in the AE SKG in Section 4.3 could bear any advantages with respect to the system efficiency. While in the sequential approach the CSI across all subcarriers will be treated as a source of shared randomness between Alice and Bob, in the parallel approach, it plays a dual role.

5.1 Parallel approach

In the parallel approach, after the channel estimation phase, the legitimate users decide on which subcarrier to send the reconciliation information (e.g., the syndromes as discussed in Section 4.2) and on which data (i.e., the SKG process here is not performed on all of the subcarriers). The total capacity has now to be distributed between data and reconciliation information bearing subcarriers. As a result, the overall set of orthogonal subcarriers comprises two subsets: a subset $\mathcal{D}$ that is used for encrypted data transmission with cardinality $|\mathcal{D}| = D$ and a subset $\breve{\mathcal{D}}$ with cardinality $|\breve{\mathcal{D}}| = N - D$ used for reconciliation such that $\mathcal{D} \cup \breve{\mathcal{D}} = \{1, \ldots, N\}$. Over $\mathcal{D}$, the achievable sum data transfer rate, denoted by C_D, is given by

$$C_D = \sum_{j \in \mathcal{D}} \log_2(1 + \hat{g}_j p_j), \tag{16}$$

while on the subset $\breve{\mathcal{D}}$, Alice and Bob exchange reconciliation information at rate

$$C_R = \sum_{j \in \breve{\mathcal{D}}} \log_2(1 + \hat{g}_j p_j). \tag{17}$$

As stated in Section 4.2, the fading coefficients are assumed to be zero-mean circularly symmetric complex Gaussian random variables. Using the theory of order statistics, the distribution of the ordered channel gains of the SKG subcarriers, $j \in \breve{\mathcal{D}}$, can be expressed as [88]:

$$f(g_j) = \frac{N!}{\sigma^2 (N-j)!\,(j-1)!} \left(1 - e^{-\frac{\hat{g}_j}{\sigma^2}}\right)^{N-j} \left(e^{-\frac{\hat{g}_j}{\sigma^2}}\right)^{j}, \tag{18}$$

where σ^2 is the variance of the channel gains. As a result of ordering the subcarriers, the variance of each of the subcarriers is now given by:

$$\sigma_j^2 = \sigma^2 \sum_{q=j}^{N} \frac{1}{q^2}, \quad j \in \{D+1, \ldots, N\}. \tag{19}$$

Thus, we can now write the SKG rate as (note that the noise variances are here normalized to unity for simplicity) [16, 86]:

$$C_{SKG} = \sum_{j \in \check{\mathcal{D}}} \log_2 \left(1 + \frac{P\sigma_j^2}{2 + \frac{1}{P\sigma_j^2}} \right). \tag{20}$$

The minimum rate necessary for reconciliation was discussed in Section 4.2. Here, alternatively, we employ a practical design approach in which the rate of the encoder used is explicitly taken into account. Note that in a rate $\frac{k}{n}$ block encoder, the side information is $n - k$ bits long, i.e., the rate of syndrome to output key bits after privacy amplification is $\frac{n-k}{k}$. However, in each key session, a 0-RTT look-up identifier of length k is also sent. Therefore, we define the parameter $\kappa = \frac{n-k}{k} + 1 = \frac{n}{k}$, i.e., the inverse of the encoder rate, that reflects the ratio of the reconciliation and 0-RTT transmission rate to the SKG rate. For example, for a rate $\frac{k}{n} = \frac{1}{2}$ encoder, $\kappa = 2$, etc. Based on this discussion, we capture the minimum requirement for the reconciliation rate through the following expression:

$$C_R \geq \kappa C_{SKG}. \tag{21}$$

Furthermore, to identify the necessary key rate, we note that depending on the exact choices of the cryptographic suites to be employed, it is possible to reuse the same key for the encryption of multiple blocks of data, e.g., as in the AES GCM, that is being considered for employment in the security protocols for URLLC systems[4]. In practical systems, a single key of length 128 to 256 bits can be used to encrypt up to gigabytes of data. As a result, we will assume that for a particular application it is possible to identify the ratio of key to data bits, which in the following we will denote by β. Specifically, we assume that the following security constraint should be met

$$C_{SKG} \geq \beta C_D, \quad 0 < \beta \leq 1, \tag{22}$$

where, depending on the application, the necessary minimum value of β can be identified. We note in passing that the case $\beta = 1$ would correspond to a one-time-pad, i.e., the generated keys could be simply x-ored with the data to achieve perfect secrecy without the need of any cryptographic suites.

Accounting for the reconciliation rate and security constraints in (21) and (22), we formulate the following maximization problem:

$$\max_{p_j, j \in \mathcal{D}} \sum_{j \in \mathcal{D}} R_j \tag{23}$$

$$\text{s.t.}(14), (21), (22),$$

$$\sum_{j \in \mathcal{D}} R_j + \sum_{j \in \check{\mathcal{D}}} R_j \leq C. \tag{24}$$

(22) can be integrated with (21) to the combined constraint

$$\sum_{j \in \mathcal{D}} R_j \leq \frac{\sum_{j \in \check{\mathcal{D}}} R_j}{\kappa \beta}. \tag{25}$$

The optimization problem at hand is a mixed-integer convex optimization problem with unknowns both the sets $\mathcal{D}, \check{\mathcal{D}}$, as well as the power allocation policy $p_j, j \in \{1, \ldots, N\}$. These problems are typically NP hard and addressed with the use of branch and bound algorithms and heuristics.

Algorithm 1 Heuristic Greedy Algorithm for (27)-(28)

```
procedure HEURISTIC(start, end, $R_j$)
    $j \leftarrow 1, R_0 \leftarrow 0, R_{N+1} \leftarrow 0$
    while $j \leq N - 1$ and $\sum_{j=1}^{N} R_j x_j \leq \frac{C}{1+\kappa\beta}$ do
        $\sum_{j=1}^{N} R_j x_j \leftarrow \sum_{j=1}^{N} R_{j-1} x_{j-1} + R_j x_j$
        if $\sum_{j=1}^{N} R_j x_j \leq \frac{C}{1+\kappa\beta}$ then
            $x_j \leftarrow 1; j \leftarrow j + 1$
        else do $x_j \leftarrow 0; j \leftarrow j + 1$
        end if
    end while
end procedure
```

In this work, we propose a simple heuristic to make the problem more tractable by reducing the number of free variables. In the proposed approach, we assume that the constraint (24) is satisfied with equality. The only power allocation that allows this is the water-filling approach that uniquely determines the power allocation p_j and also requires that the constraint (14) is also satisfied with equality. Thus, if we follow that approach, we determine the power allocation vector uniquely and can combine the remaining constraints (24) and (25) into a single one as:

$$\sum_{j \in \mathcal{D}} R_j \leq \frac{C}{\kappa\beta + 1}. \tag{26}$$

The new optimization problem can be re-written as

$$\max_{x_j \in \{0,1\}} \sum_{j=1}^{N} R_j x_j \tag{27}$$

$$\text{s.t.} \sum_{j=1}^{N} R_j x_j \leq \frac{C}{1 + \kappa\beta}. \tag{28}$$

The problem in (27, 28) is a subset-sum problem from the family of $0 - 1$ knapsack problems that is known to be NP hard [89]. However, these type of problems are solvable optimally using dynamic programming techniques in pseudo-polynomial time [89, 90]. Furthermore, it is known that greedy heuristic approaches are bounded away from the optimal solution by half [91].

We propose a simple greedy heuristic algorithm of *linear complexity*, as follows[5]. The data subcarriers are selected starting from the best—in terms of SNR—until (28) is not satisfied. Once this situation occurs, the last subcarrier added to set $\mathcal{D}$ is removed and the next one is added. This continues either to the last index N or until (28) is satisfied with equality. The algorithm is described in *Algorithm 1.*

The efficiency of the proposed parallel method—measured as the ratio of the long-term data rate versus the average capacity—is evaluated as:

$$\eta_{\text{parallel}} = \frac{\mathbb{E}\left[\sum_{j \in \mathcal{D}} R_j\right]}{\mathbb{E}[\,C]}. \tag{29}$$

[5] Without loss of generality, the algorithm assumes that the channel gains are ordered in decreasing order as in (15), and, consequently, the rates R_j are also ordered in descending order. The ordering is a $\mathcal{O}(N \log N)$ operation and required in common power allocation schemes such as the waterfilling, and, therefore does not come at any additional cost.

This efficiency quantifies the expected back-off in terms of data rates when part of the resources (power and frequency) are used to enable the generation of secret keys at the physical layer. In future work, we will compare the efficiency achieved to that of actual approaches currently used in 5G by accounting for the actual delays incurred due to the PKE key agreement operations [20].

5.2 Sequential approach

In the sequential approach, encrypted data transfer and secret key generation are two separate events; first, the secret keys are generated over the whole set of subcarriers, leading to a sum SKG rate given as

$$C_{SKG} = N \log_2 \left(1 + \frac{P\sigma^2}{2 + \frac{1}{P\sigma^2}} \right). \tag{30}$$

To estimate the efficiency of the scheme, we further need to identify the necessary resources for the exchange of the reconciliation information. We can obtain an estimate of the number of transmission frames that will be required for the transmission of the syndromes, as the expected value of the reconciliation rate (i.e., its long-term value) $\mathbb{E}[\, C_R]$. The average number of frames needed for reconciliation is then computed as:

$$M = \left\lceil \frac{\kappa C_{SKG}}{\mathbb{E}[\, C_R]} \right\rceil, \tag{31}$$

where $\lceil x \rceil$ denotes the smallest integer that is larger than x.

The average number of the frames that can be sent while respecting the secrecy constraint is:

$$L = \left\lfloor \frac{C_{SKG}}{\beta \mathbb{E}[\, C]} \right\rfloor, \tag{32}$$

where $\lfloor x \rfloor$ denotes the largest integer that is smaller than x. The efficiency of the sequential method is then calculated as:

$$\eta_{\text{sequential}} = \frac{L}{L + M}. \tag{33}$$

6 Effective data rate taking into account statistical delay QoS requirements

In the previous section, we investigated the optimal power and subcarrier allocations strategy of Alice and Bob in order to maximize their long-term average data rate and proposed a greedy heuristic algorithm of linear complexity. Here, we extend our work from Section 5 by taking into account delay requirements. In detail, we investigate the optimal resource allocation for Alice and Bob, when their communication has to satisfy specific delay constraints. To this end, we use the theory of *effective capacity* [28] which gives a limit for the maximum arrival rate under delay-bounds with a specified violation probability.

We study the *effective data rate* for the proposed pipelined SKG and encrypted data transfer scheme; the effective rate is a data-link layer metric that captures the impact of statistical delay QoS constraints on the transmission rates. As background, we refer to [92] which showed that the probability of a steady-state queue length process $Q(t)$ exceeding a certain queue-overflow threshold x converges to a random variable $Q(\infty)$ as:

$$\lim_{x \to \infty} \frac{\ln(\Pr[\, Q(\infty) > x]\,)}{x} = -\theta, \tag{34}$$

where θ indicates the asymptotic exponential decay rate of the overflow probability. For a large threshold x, (34) can be represented as $Pr[\,Q(\infty) > x] \approx e^{-\theta x}$. Furthermore, the delay-outage probability can be approximated by [28]:

$$\Pr^{\text{out}}_{\text{delay}} = \Pr[\,\text{Delay} > D_{\max}] \approx \Pr[\,Q(\infty) > 0]\, e^{-\theta \zeta D_{\max}}, \tag{35}$$

where $D_{\max}$ is the maximum tolerable delay, $\Pr[\,Q(\infty) > 0]$ is the probability of a non-empty buffer, which can be estimated from the ratio of the constant arrival rate to the averaged service rate, and ζ is the upper bound for the constant arrival rate when the statistical delay metrics are satisfied.

Using the delay exponent (θ) and the probability of non-empty buffer, the effective capacity, that denotes the maximum arrival rate, can be formulated as [28]:

$$E_C(\theta) = -\lim_{t\to\infty} \frac{1}{\theta} \ln \mathbb{E}[\,e^{-\theta S[t]}]\ \text{(bits/s)}, \tag{36}$$

where $S[\,t] = \sum_{i=1}^{t} s[\,i]$ denotes the time-accumulated service process, and $s[\,i]\,, i = 1, 2, \ldots$ denotes the discrete-time stationary and ergodic stochastic service process. Therefore, the delay exponent θ indicates how strict the delay requirements are, i.e., $\theta \to 0$ corresponds to looser delay requirements, while $\theta \to \infty$ implies exceptionally stringent delay constraints. Assuming a Rayleigh block fading system, with frame duration T_f and total bandwidth B, we have $s[\,i] = T_f B \tilde{R}_i$, with $\tilde{R}_i$ representing the instantaneous service rate achieved during the duration of the ith frame. In the context of the investigated data and reconciliation information transfer, $\tilde{R}_i$, is given by:

$$\tilde{R}_i = \frac{1}{F} \sum_{i\in\mathcal{D}} \log_2(1 + p_i \hat{g}_i), \tag{37}$$

where F is the equivalent frame duration, i.e., the total number of subcarriers used for data transmission, so that for the parallel approach, we have $F = |D|$ while for the sequential approach, $F = N(L+M)L^{-1}$.

Under this formulation and assuming that Gärtner-Ellis theorem [93, 94] is satisfied, the *effective data rate*[6] $E_C(\theta)$ is given as:

$$E_{C,\mathcal{D}}(\theta) = -\frac{1}{\theta T_f B} \ln \left(\mathbb{E}\left[e^{-\theta T_f B \tilde{R}_i} \right] \right). \tag{38}$$

We set $\alpha = \frac{\theta T_f B}{\ln(2)}$. By inserting (37) into (38), we get:

$$E_{C,\mathcal{D}}(\theta) = -\frac{1}{\ln(2)\alpha} \ln\left(\mathbb{E}\left[e^{-\ln(2)\alpha F^{-1} \sum_{i\in\mathcal{D}} \log_2(1+p_i\hat{g}_i)} \right] \right),$$

$$E_{C,\mathcal{D}}(\theta) = -\frac{1}{\alpha} \log_2 \left(\mathbb{E}\left[\prod_{i\in\mathcal{D}} (1 + p_i \hat{g}_i)^{-\alpha F^{-1}} \right] \right). \tag{39}$$

Assuming i.i.d. channel gains, by using the distributive property of the mathematical expectation, (39) becomes [95]:

$$E_{C,\mathcal{D}}(\theta) = -\frac{1}{\alpha} \log_2 \left(\prod_{i\in\mathcal{D}} \mathbb{E}\left[(1 + p_i \hat{g}_i)^{-\alpha F^{-1}} \right] \right). \tag{40}$$

[6] Since part of the transmission rate is used for reconciliation information and part for data transmission, the terms "*effective syndrome rate*" and "*effective data rate*" are introduced instead of the term "effective capacity", for rigor. We note that we assume the information data and reconciliation information are accumulated in separate independent buffers within the transmitter.

We further manipulate by using the log-product rule to obtain:

$$E_{C,D}(\theta) = -\frac{1}{\alpha}\sum_{i\in\mathcal{D}}\log_2\left(\mathbb{E}\left[\left(1+p_i\hat{g}_i\right)^{-\alpha F^{-1}}\right]\right). \tag{41}$$

Similarly, the *effective syndrome rate* can be written as:

$$E_{C,\check{D}}(\theta) = -\frac{1}{\alpha}\sum_{i\in\check{\mathcal{D}}}\log_2\left(\mathbb{E}\left[\left(1+p_i\hat{g}_i\right)^{-\alpha\check{F}^{-1}}\right]\right), \tag{42}$$

where the size of $\check{F}$ here is $|N-D|$.

Using that, we now reformulate the maximization problem given in (23) by adding a delay constraint. The reformulated problem can be expressed as follows:

$$\max_{p_j, j\in\mathcal{D}} E_{C,\mathcal{D}}(\theta), \tag{43}$$

$$\text{s.t.}(14), (25),$$

$$E_{C,\mathcal{D}}(\theta) + E_{C,\check{\mathcal{D}}}(\theta) \leq E_C^{\text{opt}}(\theta), \tag{44}$$

where $E_C^{\text{opt}}(\theta)$ represents the maximum achievable effective capacity for both key and data transmission for a given value of θ over N subcarriers:

$$E_C^{\text{opt}}(\theta) = \max_{p_i, i=1,2,\ldots N}\left\{-\frac{1}{\alpha}\log_2\left(\mathbb{E}\left[\prod_{i=1}^{N}\left(1+p_i\hat{g}_i\right)^{-\alpha N^{-1}}\right]\right)\right\}. \tag{45}$$

In the proposed approach, we assume that the constraint (44) is satisfied with equality. The optimization problem in (43) can be evaluated as two sub-optimization problems: (i) finding the optimal long-term power allocation from (14) and (45) and (ii) finding the optimal subcarrier allocation that satisfies (25). We solve the first problem that gives the optimal power allocation using convex optimization tools. Next, as in Section 5. we use two methods to solve subcarrier allocation problem, i.e., by formulating a subset-sum 0 – 1 knapsack optimization problem or through a variation of *Algorithm 1*. The efficiency of both methods is compared numerically to the sequential method in Section 7.

Now, following the same steps as in (39)–(41) and using the fact that maximizing $E_C(\theta)$ is equivalent to minimizing $-E_C(\theta)$ (this is due to $\log(\cdot)$ being a monotonically increasing concave function for any $\theta > 0$), we formulate the following minimization problem:

$$\min_{p_i, i=1,2,\ldots N}\sum_{i=1}^{N}\left(\mathbb{E}\left[\left(1+p_i\hat{g}_i\right)^{-\alpha N^{-1}}\right]\right), \tag{46}$$

$$\text{s.t. } (14).$$

where $F = N$ in this case as the full set of subcarriers is concerned. We form the Lagrangian function $\mathcal{L}$ as:

$$\mathcal{L} = \left(\mathbb{E}\left[\left(1+p_i\hat{g}_i\right)^{-\alpha N^{-1}}\right]\right) + \lambda\left(\sum_{i=1}^{N}p_i - NP\right). \tag{47}$$

By differentiating (47) w.r.t. p_i and setting the derivative equal to zero [96], we get:

$$\frac{\partial\mathcal{L}}{\partial p_i} = \lambda - \frac{\alpha\hat{g}_i}{N}\left(\hat{g}_i p_i + 1\right)^{-\frac{\alpha}{N}-1} = 0. \tag{48}$$

Solving (48) gives the optimal power allocation policy:

$$p_i^* = \frac{1}{g_0^{\frac{N}{\alpha+N}} \hat{g}_i^{\frac{\alpha}{\alpha+N}}} - \frac{1}{\hat{g}_i}, \tag{49}$$

where $g_0 = \frac{N\lambda}{\alpha}$ is the cutoff value which can be found from the power constraint. By inserting p_i^* in $E_C(\theta)$, we obtain the expression for $E_C^{\text{opt}}(\theta)$:

$$E_C^{\text{opt}}(\theta) = -\frac{1}{\alpha} \sum_{i=1}^{N} \log_2 \left(\mathbb{E} \left[\left(\frac{\hat{g}_i}{g_0} \right)^{-\frac{\alpha}{\alpha+N}} \right] \right) \tag{50}$$

When $\theta \to 0$, the optimal power allocation is equivalent to water-filling, and when $\theta \to \infty$, the optimal power allocation transforms to total channel inversion.

Now, fixing the power allocation as in (49), we can easily find the optimal subcarrier allocation that satisfies (25). As in Section 5, to do that, we first formulate a subset-sum $0-1$ knapsack optimization problem that we solve using the standard dynamic programming approach. Furthermore, we evaluate the performance of the heuristic algorithm presented in *Algorithm 1*.

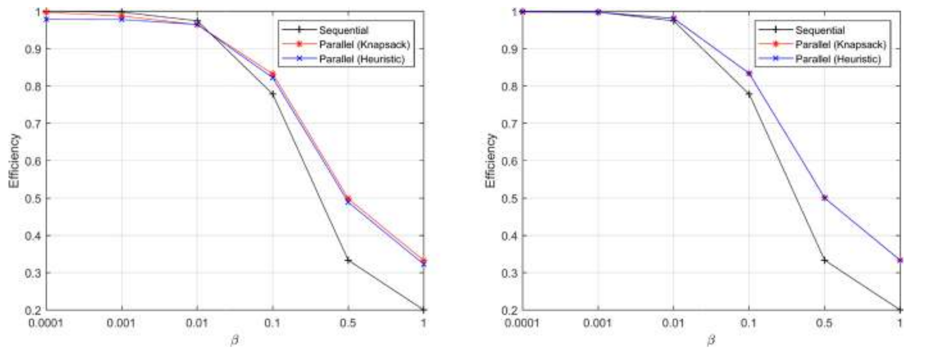
7 Results and discussion

In this section, we provide numerical evaluations of the efficiency that can be achieved with the presented methods (i.e., sequential and parallel) for different values of the main parameters. With respect to the parallel approach, we provide numerical results of the optimal dynamic programming solution of the subset-sum $0 - 1$ knapsack problem, as well as of the greedy heuristic approach presented in *Algorithm 1*. For the case of the long-term average data rate C_D (16), we compare the two methods through their efficiencies, i.e., $\eta_{\text{sequential}}$ and η_{parallel} given in (33) and (29), respectively. Next, to compare the two methods in the case of *effective data rate*, we evaluate $E_{C,D}(\theta)$ given in (41). For better illustration of each case, they are separated into different subsections.

7.1 Numerical results for the case long-term average C_D

Figure 4a and b show the efficiency of the methods for $N = 12$ and $N = 64$, respectively, while $\kappa = 2$ and $P = 10$. We note that the proposed heuristic algorithm has a near-optimal performance (almost indistinguishable from the red curves achieved with dynamic programming). Due to this fact (which was tested across all scenarios that follow), only the heuristic approach is shown in subsequent figures for clarity in the graphs.

We see that when there are a small number of subcarriers (N=12, typical for NB-IoT) and small β, the efficiency of both the parallel η_{parallel} and the sequential $\eta_{\text{sequential}}$ approaches are very close to unity, a trend that holds for increasing N. With increasing β, due to the fact that more frames are needed for reconciliation in the sequential approach (i.e., M increases), regardless of the total number of subcarriers, the parallel method proves more efficient than the sequential. While the efficiency of the sequential and parallel methods coincide almost until around $\beta = 0.01$ for $N = 12$, for $N = 64$, the crossing point of the curves moves to the left and the efficiency of the two methods coincide until around $\beta = 0.001$. This trend was found to be consistent across many values of N, only two of which are shown here for compactness of presentation.



a) Efficiency comparison for $N = 12$, SNR=10 dB and $\kappa = 2$.

b) Efficiency comparison for $N = 64$, SNR=10 dB and $\kappa = 2$.

Fig. 4 **a** Efficiency comparison for $N = 12$, SNR=10 dB, and $\kappa = 2$. **b** Efficiency comparison for $N = 64$, SNR=10 dB, and $\kappa = 2$

Next, in Fig. 5, the efficiency of the parallel η_{parallel} and the sequential $\eta_{\text{sequential}}$ methods are shown for two different values of $\kappa \in \{2, 3\}$ for SNR $=$ 10 dB and $N = 24$. It is straightforward to see that they both follow similar trends and when κ increases the efficiency decreases. On the other hand, regardless of the value of κ, they both perform identically until around $\beta = 0.001$.

Finally, in Fig. 6, focusing on the parallel method, the average size of set $\mathcal{D}$ is shown for different values of σ_e^2 and SNR levels (Fig. 6a) and κ (Fig. 6b), for $N = 24$. As expected, in Fig. 6a, we see when the SNR increases the size of the set increases, too. This is due to the fact that more power is used on any single subcarrier and consequently a higher reconciliation rate can be sustained. Regarding the estimation error σ_e^2 of the CSI, it only slightly affects the performance at high SNR levels. Hence, more subcarriers have to be used for reconciliation, and fewer for data. The SNR level in Fig. 6b is set to 10 dB. The

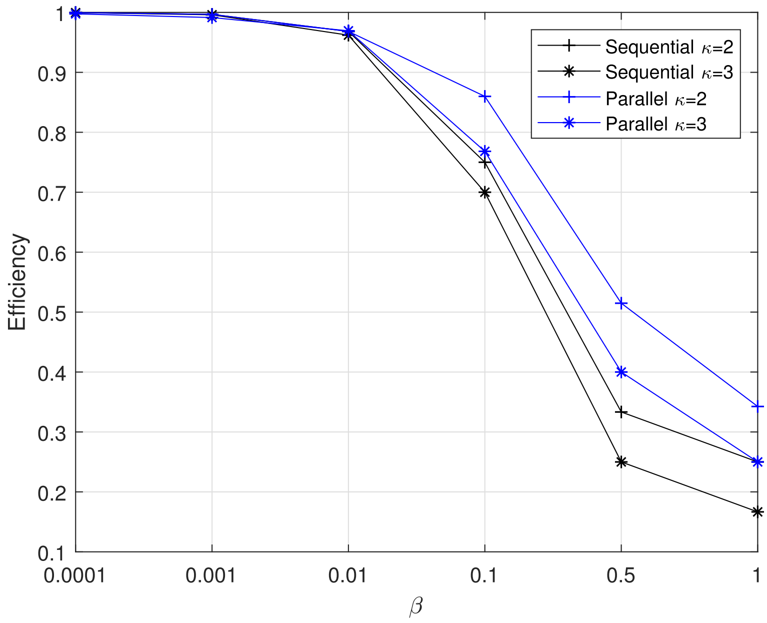
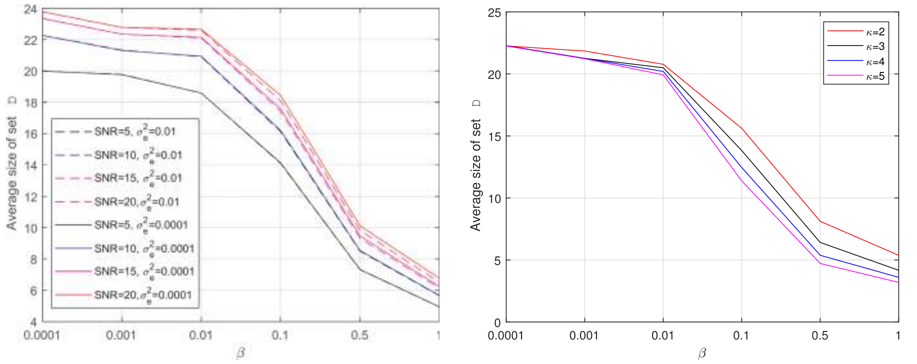


Fig. 5 Efficiency vs κ, for $N = 24$, SNR=10 dB



a) Size of set $\mathcal{D}$ for different SNR levels and σ_e^2 when $N = 24$.

b) Size of set $\mathcal{D}$ for different values of κ when $N = 24$.

Fig. 6 **a** Size of set $\boldsymbol{\mathcal{D}}$ for different SNR levels and σ_e^2 when $N = 24$. **b** Size of set $\boldsymbol{\mathcal{D}}$ for different values of $\boldsymbol{\kappa}$ when $N = 24$

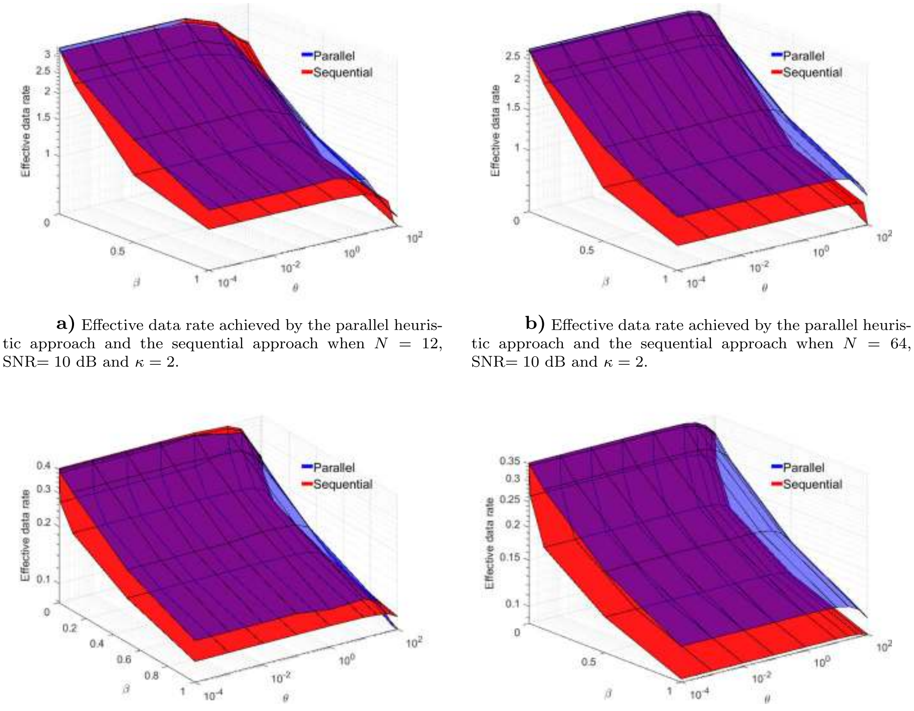
figure shows that when increasing κ, the size of set $\mathcal{D}$ decreases. This result can be easily predicted from inequality (21), meaning, when κ increases, more reconciliation data has to be sent; hence, fewer subcarriers can be used for data. In both Fig. 6a and b when β increases, the size of set $\mathcal{D}$ decreases; this effect is a consequence of constraint (28) as the data rate is decreasing with β.

7.2 Numerical results for the case of *effective data rate*

Inspired by the good performance of *Algorithm 1*, in the case where long-term average rate is the metric of interest, here, we continue our investigation with a variation of *Algorithm 1*, with the following differences: at lines 3 and 5 instead of (26), we use the constraint (25); the power allocation is fixed as in (49). The performance of our system is again compared with a sequential method, and the metric of interest here is the *effective data rate*. The comparison is performed by taking into account the following parameters: signal-to-noise ratio (SNR), number of subcarriers N, ratio of the reconciliation and 0−RTT transmission rate to the SKG rate κ, delay exponent θ, and the ratio of key bits to data bits β.

In Fig. 7, we give a three-dimensional plot showing the dependence of the achievable *effective data rate* $E_{C,D}(\theta)$ on β and θ. Figure 7a and b compare the parallel heuristic approach and the sequential approach for high SNR levels, whereas Fig. 7c and d compare their performance for low SNR level. In Fig. 7a and c, we have $N = 12$ while in Fig. 7b and d, the total number of subcarriers is $N = 64$. All graphs compare the performance of the heuristic parallel approach and the sequential approach for $\kappa = 2$.

As discussed in Section 6, when the delay exponent θ increases, the optimal power allocation transforms from waterfilling to total channel inversion. Consequently, the rate achieved on all subcarriers converges to the same value; hence, when we a have small number of subcarriers (such as $N = 12$) and small values of β, then using a single subcarrier for reconciliation data will use more capacity than needed and most of the rate on this subcarrier is wasted. Devoting a whole subcarrier for sending the reconciliation data for the case of $N = 12$ and $\beta = 0.0001$ is almost equivalent of losing $1/12$ of the achievable rate.



a) Effective data rate achieved by the parallel heuristic approach and the sequential approach when $N = 12$, SNR= 10 dB and $\kappa = 2$.

b) Effective data rate achieved by the parallel heuristic approach and the sequential approach when $N = 64$, SNR= 10 dB and $\kappa = 2$.

c) Effective data rate achieved by the parallel heuristic approach and the sequential approach when $N = 12$, SNR= 0.2 dB and $\kappa = 2$.

d) Effective data rate achieved by the parallel heuristic approach and the sequential approach when $N = 64$, SNR= 0.2 dB and $\kappa = 2$.

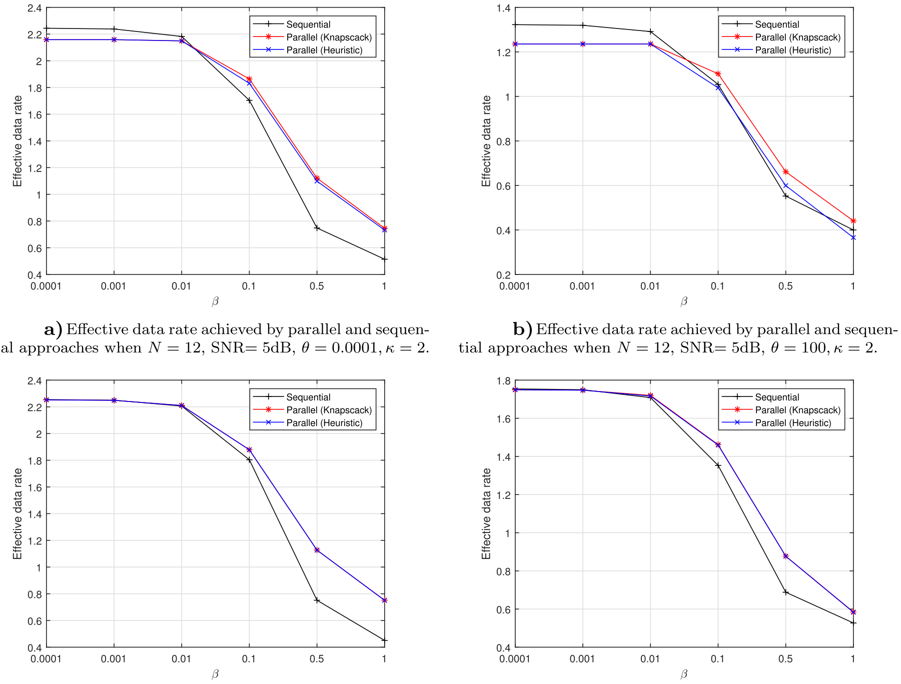
Fig. 7 **a** Effective data rate achieved by the parallel heuristic approach and the sequential approach when $N = 12$, SNR= 10 dB, and $\kappa = 2$. **b** Effective data rate achieved by the parallel heuristic approach and the sequential approach when $N = 64$, SNR= 10 dB, and $\kappa = 2$. **c** Effective data rate achieved by the parallel heuristic approach and the sequential approach when $N = 12$, SNR= 0.2 dB, and $\kappa = 2$. **d** Effective data rate achieved by the parallel heuristic approach and the sequential approach when $N = 64$, SNR= 0.2 dB, and $\kappa = 2$

This can be seen for in Fig. 7a and c, where $N = 12$. When the SNR is high (See Fig. 7a), as discussed, this effect is mostly noticeable for large values of θ and small values of β[7], whereas for small values of β and θ, both algorithms perform nearly identically. A similar trend can be seen at the low SNR regime in Fig. 7c. However, at a low SNR, the sequential approach has a lower effective data rate. This happens because at high SNR levels, each reconciliation frame will contain more information and hence more data frames will follow. Therefore, at the low SNR regime, the reconciliation information received will decrease; hence, less data can be sent afterwards. This does not affect the parallel approach. However, in both scenarios high SNR Fig. 7a and low SNR Fig. 7c, when β increases regardless of the value of θ, the parallel approach always achieves higher *effective data rate* $E_{C,D}(\theta)$.

In the next case, when the total number of subcarriers is $N = 64$, illustrated in Fig. 7b and d, we see that the penalty of devoting a high part of the achievable effective capacity $E_C^{\text{opt}}(\theta)$ to reconciliation disappears and the heuristic parallel approach always achieves higher or identical *effective data rate* $E_{C,D}(\theta)$ compared to the sequential approach. This trend repeats for high and low SNR levels as given in Fig. 7b and d, respectively.

Now, we take a closer look and transform some specific cases from the 3D plots to two-dimensional graphs. In Fig. 8, we see the achieved *effective data rate* $E_{C,D}(\theta)$ given in

[7] That is, that the ratio of reconciliation information to data is small as seen from Eq. (25))



a) Effective data rate achieved by parallel and sequential approaches when $N = 12$, SNR= 5dB, $\theta = 0.0001, \kappa = 2$.

b) Effective data rate achieved by parallel and sequential approaches when $N = 12$, SNR= 5dB, $\theta = 100, \kappa = 2$.

c) Effective data rate achieved by parallel and sequential approaches when $N = 64$, SNR= 5dB, $\theta = 0.0001, \kappa = 2$.

d) Effective data rate achieved by parallel and sequential approaches when $N = 64$, SNR= 5dB, $\theta = 100, \kappa = 2$.

Fig. 8 **a** Effective data rate achieved by parallel and sequential approaches when $N = 12$, SNR= 5dB, $\theta = 0.0001, \kappa = 2$. **b** Effective data rate achieved by parallel and sequential approaches when $N = 12$, SNR= 5dB, $\theta = 100, \kappa = 2$. **c** Effective data rate achieved by parallel and sequential approaches when $N = 64$, SNR= 5dB, $\theta = 0.0001, \kappa = 2$. **d** Effective data rate achieved by parallel and sequential approaches when $N = 64$, SNR= 5dB, $\theta = 100, \kappa = 2$

(41), for different values of N and θ while the SNR =5 dB and $\kappa = 2$. Figure 8a gives the achieved effective rate on set $\mathcal{D}$ for $N = 12$ and $\theta = 0.0001$ (relaxed delay constraint). Similarly to the case of long-term average value of C_D, we see that for small values of β, the sequential approach achieves slightly higher effective data rate. As before, the increase of β results in more reconciliation frames M required in the sequential case. This effect is not seen in the parallel case and for high values of β it performs better.

Figure 8b illustrates the case when $N = 12$ and $\theta = 100$ (very stringent delay constraint). Similarly, as in Fig. 7, we can see that for small values of β, the sequential approach performs better than the parallel. As discussed, the efficiency loss is caused by the fact that the devoted part of the total achievable effective capacity $E_C^{\text{opt}}(\theta)$ to reconciliation (syndrome communication) is more than what is required. However, a higher β leads to an increase in the reconciliation information that needs to be sent, and the rate of the subcarriers in set $\breve{\mathcal{D}}$ will be fully or almost fully utilized and the parallel approach shows better performance for these values.

In the next two, Fig. 8c and d, we show the performance of the two algorithms for higher value of $N = 64$. It is easy to see that regardless of the value of θ and β, both algorithms perform identical or the parallel is better. In the previous case of $N = 12$, increasing θ might reduce the effectiveness of the parallel approach; however, when $N = 64$, increasing θ does not incur such a penalty and the parallel is either identical to the sequential or outperforms it.

Another interesting fact from Fig. 8 is that looking at the parallel approach, it can easily be seen that in all cases, the heuristic approach almost always performs as well as the optimal knapsack solution. The case of small values of θ is similar to the one when we work with long-term average rate and choosing the best subcarriers for data transmission works as well as the optimal Knapsack solution. Interestingly, *Algorithm 1* works well for high values of θ, too. This can be explained by the fact that when θ increases, the rate on all of the subcarriers becomes similar, and switching the subcarriers in set $\mathcal{D}$ does not incur high penalty.

8 Conclusions

In this work, we discussed the possibility of using SKG in conjunction with PUF authentication protocols, illustrating this can greatly reduce the authentication and key generation latency compared to traditional mechanisms. Furthermore, we presented an AE scheme using SKG and a resumption protocol which further contribute to the system's security and latency reduction, respectively.

In addition, we explored the possibility of pipelining encrypted data transfer and SKG in a Rayleigh BF-AWGN environment. We investigated the maximization of the data transfer rate in parallel to performing SKG. We took into account imperfect CSI measurements and the effect of order statistics on the channel variance. Two scenarios were differentiated in our study: (i) the optimal data transfer rate was found under power and security constraints, represented by the system parameters β and κ, which represent the minimum ratio of SKG rate to data rate and the maximum ratio of SKG rate to reconciliation rate and (ii) by adding a delay constraint, represented by parameter θ, to the security and power constraint, we found the optimal *effective data rate*.

To finalize our study, we illustrated through numerical comparisons the efficiency of the proposed parallel method, in which SKG and data transfer are inter-weaved to a sequential method where the two operations are done separately. The results of the two scenarios showed that in most of the cases, the performance of both methods, parallel and sequential, is either equal or the parallel performs better. As the possible advantage of using the sequential is small and only applies in particular scenarios, we recommend the parallel scheme as a universal mechanism for general protocol design, when latency is an issue. Furthermore, a significant result is that although the optimal subcarrier scheduling is an NP hard $0-1$ knapsack problem, it can be solved in linear time using a simple heuristic algorithm with virtually no loss in performance.

Abbreviations

AE: Authenticated encryption; BF-AWGN: Block fading additive white Gaussian noise; B5G: Beyond 5G; CRP: Challenge-response pair; CSI: Channel state information; EAP-TLS: Extensible authentication protocol-transport layer security; IoT: Internet of Things; MAC: Message authentication code; MiM: Man in the middle; NB-IoT: Narrowband IoT; OFDM: Orthogonal frequency division multiplexing; PHY: Physical layer; PKE: Public key encryption; PLS: Physical layer security; PUF: Physical unclonable function; QoS: Quality of service; RAN: Radio access network; RSS: Received signal strength; SKG: Secret key generation; SNR: Signal-to-noise ratio; STEK: Session ticket encryption key; TLS: Transport layer security; URLLC: Ultra-reliable low-latency communication; V2X: Vehicle-to-everything communication; 0-RTT: Zero round trip time; 3GPP: The 3rd Generation Partnership Project

Authors' contributions

MM, AC, and MR conceived this study. LM contributed on the use of effective capacity within this framework. MM carried out the simulations and prepared the graphs. All authors contributed and edited the manuscript. All authors read and approved the final manuscript.

Funding
Miroslav Mitev is supported by the Doctoral Training Programme of CSEE, University of Essex. Arsenia Chorti is supported by the ELIOT ANR-18-CE40-0030 and FAPESP 2018/12579-7 and the INEX Project eNiGMA. Martin Reed is supported by the project SerIoT which has received funding from the European Union's Horizon 2020 Research and Innovation program under grant agreement No 780139. Leila Musavian is supported by the RECENT project which has received funding from European Union Horizon 2020, RISE 2018 scheme (H2020-MSCA-RISE-2018) under Marie Sklodowska – Curie grant agreement No. 823903.

Availability of data and materials
No data sets were used in the production of the results shown in this paper. All the results can be regenerated from first principals using the formulations derived within the paper.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]School of CSEE, University of Essex, Colchester, UK. [2]ETIS UMR8051, CY University, ENSEA, CNRS, F-95000 Cergy, France.

Received: 7 November 2019 Accepted: 21 May 2020
Published online: 09 June 2020

References

1. A. Mukherjee, Physical-layer security in the Internet of Things: sensing and communication confidentiality under resource constraints. Proc. IEEE. **103**(10), 1747–1761 (2015). https://doi.org/10.1109/JPROC.2015.2466548
2. A. Yener, S. Ulukus, Wireless physical-layer security: lessons learned from information theory. Proc. IEEE. **103**(10), 1814–1825 (2015). https://doi.org/10.1109/JPROC.2015.2459592
3. D. Karatzas, A. Chorti, N. M. White, C. J. Harris, Teaching old sensors new tricks: archetypes of intelligence. IEEE Sensors J. **7**(5), 868–881 (2007). https://doi.org/10.1109/JSEN.2007.893986
4. 3GPP TR 33.825 V0.3.0, Study on the Security for 5G URLLC (Release 16). 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects. https://www.3gpp.org/ftp/Specs/archive/33_series/33.825/. Accessed 1 2019
5. A. Chorti, C. Hollanti, J.-C. Belfiore, H. V. Poor, Physical layer security: a paradigm shift in data confidentiality. Lect. Notes Electr. Eng. **358** (2016). https://doi.org/10.1007/978-3-319-23609-4_1
6. A. Chorti, K. Papadaki, H. V. Poor, Optimal power allocation in block fading channels with confidential messages. IEEE Trans. Wirel. Commun. **14**(9), 4708–4719 (2015). https://doi.org/10.1109/TWC.2015.2424964
7. A. Chorti, S. M. Perlaza, Z. Han, H. V. Poor, On the resilience of wireless multiuser networks to passive and active eavesdroppers. IEEE J. Sel. Areas Commun. **31**(9), 1850–1863 (2013). https://doi.org/10.1109/JSAC.2013.130917
8. A. Chorti, H. V. Poor, in *2012 International Conference on Computing, Networking and Communications (ICNC)*, Achievable secrecy rates in physical layer secure systems with a helping interferer, (2012), pp. 18–22. https://doi.org/10.1109/ICCNC.2012.6167408
9. M. Mitev, A. Chorti, M. Reed, in *2019 IEEE Global Communications Conference (GLOBECOM)*, Subcarrier scheduling for joint data transfer and key generation schemes in multicarrier systems, (2019), pp. 1–6. https://doi.org/10.1109/GLOBECOM38437.2019.9013809
10. Y. Kanaras, A. Chorti, M. Rodrigues, I. Darwazeh, in *Proc. 13th Int. OFDM WS*, An optimum detection for a spectrally efficient non orthogonal FDM system, (2008), pp. 65–68
11. A. Chorti, H. V. Poor, in *2011 Asilomar Conf. Signals, Systems and Computers (ASILOMAR)*, Faster than Nyquist interference assisted secret communication for OFDM systems, (2011), pp. 183–187. https://doi.org/10.1109/ACSSC.2011.6189981
12. A. Chorti, in *2012 46th Annual Conference on Information Sciences and Systems (CISS)*, Helping interferer physical layer security strategies for M-QAM and M-PSK systems, (2012), pp. 1–6. https://doi.org/10.1109/CISS.2012.6310861
13. M. Latvaaho, K. Leppänen, Key drivers and research challenges for 6G ubiquitous wireless intelligence (2019). http://urn.fi/urn:isbn:9789526223544
14. U. M. Maurer, Secret key agreement by public discussion from common information. IEEE Trans. Inf. Theory. **39**(3), 733–742 (1993). https://doi.org/10.1109/18.256484
15. R. Ahlswede, I. Csiszar, Common randomness in information theory and cryptography. i. secret sharing. IEEE Trans. Inf. Theory. **39**(4), 1121–1132 (1993). https://doi.org/10.1109/18.243431
16. C. Ye, A. Reznik, Y. Shah, in *2006 IEEE International Symposium on Information Theory*, Extracting secrecy from jointly gaussian random variables, (2006), pp. 2593–2597. https://doi.org/10.1109/ISIT.2006.262101
17. B. Gassend, D. Clarke, M. van Dijk, S. Devadas, in *Proceedings of the 9th ACM Conference on Computer and Communications Security, CCS '02*, Silicon physical random functions (Association for Computing Machinery, New York, 2002), pp. 148–160. https://doi.org/10.1145/586110.586132
18. R. Pappu, B. Recht, J. Taylor, N. Gershenfeld, Physical one-way functions. Science. **297**(5589), 2026–2030 (2002). https://doi.org/10.1126/science.1074376
19. R. Maes, I. Verbauwhede, *Physically unclonable functions: a study on the state of the art and future research directions*, (2010), pp. 3–37. https://doi.org/10.1007/978-3-642-14452-3_1
20. A. Weinand, M. Karrenbauer, H. Schotten, Security solutions for local wireless networks in control applications based on physical layer security. IFAC-PapersOnLine. **51**, 32-39 (2018)
21. A. Mukherjee, S. A. A. Fakoorian, J. Huang, A. L. Swindlehurst, Principles of physical layer security in multiuser wireless networks: a survey. IEEE Commun. Surv. Tutor. **16**(3), 1550–1573 (2014). https://doi.org/10.1109/SURV.2014.012314.00178

22. A. Chorti, in *in Proc. Workshop on Communication Security (WCS)*, A study of injection and jamming attacks in wireless secret sharing systems (Springer, Cham, 2017)
23. E. Rescorla, The transport layer security (TLS) protocol version 1.3. RFC 8446 (2018). https://rfc-editor.org/rfc/rfc8446.txt. Accessed 8 2018
24. N. Aviram, K. Gellert, T. Jager, Session resumption protocols and efficient forward security for TLS 1.3 0-RTT. Cryptology ePrint Archive, Report 2019/228 (2019). https://eprint.iacr.org/2019/228. Accessed 2 2019
25. M. Bellare, C. Namprempre, Authenticated encryption: relations among notions and- analysis of the generic composition paradigm. J. Cryptol. **21**(4), 469–491 (2008). https://doi.org/10.1007/s00145-008-9026-x
26. T. Krovetz, P. Rogaway, in *FSE, Lecture Notes in Computer Science*, The software performance of authenticated-encryption modes (Springer, Berlin, 2011)
27. S. Koteshwara, A. Das, Comparative study of authenticated encryption targeting lightweight IoT applications. IEEE Design Test. **34**(4), 26–33 (2017). https://doi.org/10.1109/MDAT.2017.2682234
28. D. Wu, R. Negi, Effective capacity: a wireless link model for support of quality of service. IEEE Trans. Wirel. Commun. **2**(4), 630–643 (2003). https://doi.org/10.1109/TWC.2003.814353
29. W. Che, M. Martin, G. Pocklassery, V. K. Kajuluri, F. Saqib, J. F. Plusquellic, A privacy-preserving, mutual puf-based authentication protocol. Cryptography. **1**, 3 (2016)
30. B. Gassend, D. Clarke, M. van Dijk, S. Devadas, in *Proceedings of the 9th ACM Conference on Computer and Communications Security (CCS '02)*, Silicon physical random functions (Association for Computing Machinery, New York, 2002), pp. 148–160. https://doi.org/10.1145/586110.586132
31. C. Marchand, L. Bossuet, U. Mureddu, N. Bochard, A. Cherkaoui, V. Fischer, Implementation and characterization of a physical unclonable function for IoT: a case study with the TERO-PUF. IEEE Trans. Comput.-Aided Des. Integr. Circ. Syst. **37**(1), 97–109 (2018). https://doi.org/10.1109/TCAD.2017.2702607
32. J. Guajardo, S. S. Kumar, G.-J. Schrijen, P. Tuyls, in *Proceedings of the 9th International Workshop on Cryptographic Hardware and Embedded Systems (CHES '07)*, FPGA intrinsic PUFs and their use for IP protection (Springer, Berlin, 2007), pp. 63–80. https://doi.org/10.1007/978-3-540-74735-2_5
33. J. Aarestad, P. Ortiz, D. Acharyya, J. Plusquellic, Help: a hardware-embedded delay PUF. IEEE Des. Test. **30**(2), 17–25 (2013). https://doi.org/10.1109/MDT.2013.2247459
34. A. Babaei, G. Schiele, in *Sensors*, Physical unclonable functions in the internet of things: State of the art and open challenges (MDPI, Basel, 2019)
35. P. Maurya, S. Bagchi, A secure PUF-based unilateral authentication scheme for RFID system. Wirel. Pers. Commun. **103** (2018). https://doi.org/10.1007/s11277-018-5875-2
36. M. Yu, M. Hiller, J. Delvaux, R. Sowell, S. Devadas, I. Verbauwhede, A lockdown technique to prevent machine learning on PUFs for lightweight authentication. IEEE Trans. Multi-Scale Comput. Syst. **2**(3), 146–159 (2016). https://doi.org/10.1109/TMSCS.2016.2553027
37. J. Calhoun, C. Minwalla, C. Helmich, F. Saqib, W. Che, J. Plusquellic, Physical unclonable function (PUF)-based e-cash transaction protocol (PUF-Cash). Cryptography. **3**, 18 (2019). https://doi.org/10.3390/cryptography3030018
38. M. N. Aman, K. C. Chua, B. Sikdar, Mutual authentication in IoT systems using physical unclonable functions. IEEE Internet Things J. **4**(5), 1327–1340 (2017). https://doi.org/10.1109/JIOT.2017.2703088
39. J. Delvaux, R. Peeters, D. Gu, I. Verbauwhede, A survey on lightweight entity authentication with strong PUFs. ACM Comput. Surv. **48**(2) (2015). https://doi.org/10.1145/2818186
40. S. Jana, S. N. Premnath, M. Clark, S. K. Kasera, N. Patwari, S. V. Krishnamurthy, in *Proceedings of the 15th Annual International Conference on Mobile Computing and Networking (MobiCom '09)*, On the effectiveness of secret key extraction from wireless signal strength in real environments (Association for Computing Machinery, New York, 2009), pp. 321–332. URL https://doi.org/10.1145/1614320.1614356
41. T. Rappaport, *Wireless communications: principles and practice, 2nd edn.* (Prentice Hall PTR, USA, 2001)
42. J. Wan, A. B. Lopez, M. A. Al Faruque, in *2016 ACM/IEEE 7th International Conference on Cyber-Physical Systems (ICCPS)*, Exploiting wireless channel randomness to generate keys for automotive cyber-physical system security, (2016), pp. 1–10. https://doi.org/10.1109/ICCPS.2016.7479103
43. B. Zan, M. Gruteser, F. Hu, Key agreement algorithms for vehicular communication networks based on reciprocity and diversity theorems. IEEE Trans. Veh. Technol. **62**(8), 4020–4027 (2013). https://doi.org/10.1109/TVT.2013.2254507
44. Y. Liu, J. Jing, J. Yang, in *2008 9th International Conference on Signal Processing*, Secure underwater acoustic communication based on a robust key generation scheme, (2008), pp. 1838–1841. https://doi.org/10.1109/ICOSP.2008.4697498
45. I. U. Zaman, A. B. Lopez, M. A. A. Faruque, O. Boyraz, Physical layer cryptographic key generation by exploiting PMD of an optical fiber link. J. Light. Technol. **36**(24), 5903–5911 (2018). https://doi.org/10.1109/JLT.2018.2880957
46. D. Tian, W. Zhang, J. Sun, C. Wang, in *2019 IEEE/CIC International Conference on Communications in China (ICCC)*, Physical-layer security of visible light communications with jamming, (2019), pp. 512–517. https://doi.org/10.1109/ICCChina.2019.8855859
47. J. Zhang, T. Q. Duong, A. Marshall, R. Woods, Key generation from wireless channels: a review. IEEE Access. **4**, 614–626 (2016). https://doi.org/10.1109/ACCESS.2016.2521718
48. J. K. Tugnait, L. Tong, Z. Ding, Single-user channel estimation and equalization. IEEE Signal Proc. Mag. **17**(3), 17–28 (2000). https://doi.org/10.1109/MSP.2000.841720
49. W. C. Jakes, D. C. Cox, *Microwave mobile communications*. (Wiley-IEEE Press, New York, 1994)
50. H. Liu, Y. Wang, J. Yang, Y. Chen, in *2013 Proceedings IEEE INFOCOM*, Fast and practical secret key extraction by exploiting channel response, (2013), pp. 3048–3056. https://doi.org/10.1109/INFCOM.2013.6567117
51. S. Mathur, W. Trappe, N. Mandayam, C. Ye, A. Reznik, in *Proceedings of the 14th ACM International Conference on Mobile Computing and Networking (MobiCom '08)*, Radio-telepathy: extracting a secret key from an unauthenticated wireless channel (Association for Computing Machinery, New York, 2008), pp. 128–139. https://doi.org/10.1145/1409944.1409960
52. S. T. Ali, V. Sivaraman, D. Ostry, Eliminating reconciliation cost in secret key generation for body-worn health monitoring devices. IEEE Trans. Mobile Comput. **13**(12), 2763–2776 (2014). https://doi.org/10.1109/TMC.2013.71

53. S. Mathur, R. Miller, A. Varshavsky, W. Trappe, N. Mandayam, in *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services (MobiSys '11)*, Proximate: proximity-based secure pairing using ambient wireless signals (Association for Computing Machinery, New York, 2011), pp. 211–224. https://doi.org/10.1145/1999995.2000016
54. Intrinsic-id company. https://www.intrinsic-id.com/sram-puf
55. ICTK holdings corporation. https://ictk-puf.com/puf-technology
56. A. Maiti, I. Kim, P. Schaumont, A robust physical unclonable function with enhanced challenge-response set. IEEE Trans. Inf. Forensic Secur. **7**(1), 333–345 (2012). https://doi.org/10.1109/TIFS.2011.2165540
57. M. Akhlaq, B. Aslam, M. A. Khan, M. N. Jafri, in *Proceedings of the 11th Conference on 11th WSEAS International Conference on Communications - Volume 11(ICCOM'07)*, Comparative analysis of IEEE 802.1x authentication methods (World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, 2007), pp. 1–6
58. A. Chiornită, L. Gheorghe, D. Rosner, in *9th RoEduNet IEEE International Conference*, A practical analysis of EAP authentication methods (IEEE, Sibiu, 2010), pp. 31–35
59. C. Herder, M. Yu, F. Koushanfar, S. Devadas, Physical unclonable functions and applications: a tutorial. Proc. IEEE. **102**(8), 1126–1141 (2014). https://doi.org/10.1109/JPROC.2014.2320516
60. G. E. Suh, S. Devadas, in *2007 44th ACM/IEEE Design Automation Conference*, Physical unclonable functions for device authentication and secret key generation (IEEE, San Diego, 2007), pp. 9–14
61. C. Bhm, M. Hofer, *Physical unclonable functions in theory and practice*. (Springer, New York, 2012)
62. U. Chatterjee, R. Chakraborty, D. Mukhopadhyay, A PUF-based secure communication protocol for IoT. ACM Trans. Embedded Comput. Syst. **16**, 1–25 (2017). https://doi.org/10.1145/3005715
63. M. N. Aman, M. H. Basheer, B. Sikdar, Two-factor authentication for IoT with location information. IEEE Internet Things J. **6**(2), 3335–3351 (2019). https://doi.org/10.1109/JIOT.2018.2882610
64. M. H. Mahalat, S. Saha, A. Mondal, B. Sen, in *2018 8th International Symposium on Embedded Computing and System Design (ISED)*, A PUF based light weight protocol for secure WiFi authentication of IoT devices, (2018), pp. 183–187. https://doi.org/10.1109/ISED.2018.8703993
65. A. Braeken, PUF based authentication protocol for IoT. Symmetry. **10**, 352 (2018). https://doi.org/10.3390/sym10080352
66. Y. Yilmaz, S. R. Gunn, B. Halak, in *2018 IEEE 3rd International Verification and Security Workshop (IVSW)*, Lightweight PUF-based authentication protocol for IoT devices, (2018), pp. 38–43. https://doi.org/10.1109/IVSW.2018.8494884
67. S. Ahmad, A. H. Mir, G. R. Beigh, in *2011 Fifth IEEE International Conference on Advanced Telecommunication Systems and Networks (ANTS)*, Latency evaluation of extensible authentication protocols in WLANs, (2011), pp. 1–5. https://doi.org/10.1109/ANTS.2011.6163654
68. P. Gope, B. Sikdar, Lightweight and privacy-preserving two-factor authentication scheme for IoT devices. IEEE Internet Things J. **6**(1), 580–589 (2019). https://doi.org/10.1109/JIOT.2018.2846299
69. A. Ometov, P. Masek, L. Malina, R. Florea, J. Hosek, S. Andreev, J. Hajny, J. Niutanen, Y. Koucheryavy, in *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, Feasibility characterization of cryptographic primitives for constrained (wearable) IoT devices, (2016), pp. 1–6. https://doi.org/10.1109/PERCOMW.2016.7457161
70. J. Cho, W. Sung, Efficient software-based encoding and decoding of BCH codes. IEEE Trans. Comput. **58**(7), 878–889 (2009). https://doi.org/10.1109/TC.2009.27
71. C. Chen, M. A. Jensen, Secret key establishment using temporally and spatially correlated wireless channel coefficients. IEEE Trans. Mob. Comput. **10**(2), 205–215 (2011). https://doi.org/10.1109/TMC.2010.114
72. J. Zhang, A. Marshall, R. Woods, T. Q. Duong, Efficient key generation by exploiting randomness from channel responses of individual ofdm subcarriers. IEEE Trans. Commun. **64**(6), 2578–2588 (2016). https://doi.org/10.1109/TCOMM.2016.2552165
73. J. Zhang, B. He, T. Q. Duong, R. Woods, On the key generation from correlated wireless channels. IEEE Commun. Lett. **21**(4), 961–964 (2017). https://doi.org/10.1109/LCOMM.2017.2649496
74. M. Mitev, A. Chorti, E. V. Belmega, M. Reed, in *2019 IEEE Global Communications Conference (GLOBECOM)*, Man-in-the-middle and denial of service attacks in wireless secret key generation, (2019), pp. 1–6. https://doi.org/10.1109/GLOBECOM38437.2019.9013816
75. C. Saiki, A. Chorti, in *2015 IEEE Conference on Communications and Network Security (CNS)*, A novel physical layer authenticated encryption protocol exploiting shared randomness, (2015), pp. 113–118. https://doi.org/10.1109/CNS.2015.7346818
76. Q. Wang, H. Su, K. Ren, K. Kim, in *2011 Proceedings IEEE INFOCOM*, Fast and scalable secret key generation exploiting channel phase randomness in wireless networks, (2011), pp. 1422–1430. https://doi.org/10.1109/INFCOM.2011.5934929
77. C. Ye, S. Mathur, A. Reznik, Y. Shah, W. Trappe, N. B. Mandayam, Information-theoretically secret key generation for fading wireless channels. IEEE Trans. Inf. Forensics Secur. **5**(2), 240–254 (2010). https://doi.org/10.1109/TIFS.2010.2043187
78. C. Huth, R. Guillaume, T. Strohm, P. Duplys, I. A. Samuel, T. Gneysu, Information reconciliation schemes in physical-layer security. Comput. Netw. **109**(P1), 84–104 (2016). https://doi.org/10.1016/j.comnet.2016.06.014
79. L. Guyue, Z. Zhang, Y. Yu, A. Hu, A hybrid information reconciliation method for physical layer key generation. Entropy. **21**, 688 (2019). https://doi.org/10.3390/e21070688
80. P. Treeviriyanupab, P. Sangwongngam, K. Sripimanwat, O. Sangaroon, in *2012 9th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, BCH-based Slepian-Wolf coding with feedback syndrome decoding for quantum key reconciliation, (2012), pp. 1–4. https://doi.org/10.1109/ECTICon.2012.6254266
81. J. Etesami, W. Henkel, in *2012 1st IEEE International Conference on Communications in China (ICCC)*, LDPC code construction for wireless physical-layer key reconciliation, (2012), pp. 208–213. https://doi.org/10.1109/ICCChina.2012.6356879

82. C. H. Bennett, G. Brassard, C. Crepeau, U. M. Maurer, Generalized privacy amplification. IEEE Trans. Inf. Theory. **41**(6), 1915–1923 (1995). https://doi.org/10.1109/18.476316
83. F. Zhan, N. Yao, On the using of discrete wavelet transform for physical layer key generation. Ad Hoc Netw. **64**, 22–31 (2017). https://doi.org/10.1016/j.adhoc.2017.06.003
84. M. Bloch, J. Barros, M. R. D. Rodrigues, S. W. McLaughlin, Wireless information-theoretic security. IEEE Trans. Inf. Theory. **54**(6), 2515–2534 (2008). https://doi.org/10.1109/TIT.2008.921908
85. M. Mitev, A. Chorti, M. Reed, in *2019 15th International Wireless Communications Mobile Computing Conference (IWCMC)*, Optimal resource allocation in joint secret key generation and data transfer schemes, (2019), pp. 360–365. https://doi.org/10.1109/IWCMC.2019.8766766
86. E. V. Belmega, A. Chorti, Protecting secret key generation systems against jamming: energy harvesting and channel hopping approaches. IEEE Trans. Inf. Forensic Secur. **12**(11), 2611–2626 (2017). https://doi.org/10.1109/TIFS.2017.2713342
87. M. Medard, The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel. IEEE Trans. Inf. Theory. **46**(3), 933–946 (2000). https://doi.org/10.1109/18.841172
88. H.-C. Yang, M.-S. Alouini, *Order statistics in wireless communications: diversity, adaptation, and scheduling in MIMO and OFDM systems, 1st edn*. (Cambridge University Press, USA, 2011)
89. S. Martello, P. Toth, *Knapsack problems: algorithms and computer implementations*. (Wiley, USA, 1990)
90. H. Kellerer, U. Pferschy, D. Pisinger, *Knapsack problems*. (Springer, Boston, 2004)
91. V. V. Vazirani, *Approximation algorithms*. (Springer, Berlin, 2001)
92. C.-S. Chang, Stability, queue length, and delay of deterministic and stochastic queueing networks. IEEE Trans. Autom. Control. **39**(5), 913–931 (1994). https://doi.org/10.1109/9.284868
93. J. Gärtner, On large deviation from invariant measure. Theory Prob. Appl. **22**, 24–39 (1977)
94. R. Ellis, Large deviations for a general class of random vectors. Ann. Probab. **12** (1984). https://doi.org/10.1214/aop/1176993370
95. T. Abrao, S. Yang, L. D. H. Sampaio, P. J. E. Jeszensky, L. Hanzo, Achieving maximum effective capacity in ofdma networks operating under statistical delay guarantee. IEEE Access. **5**, 14333–14346 (2017). https://doi.org/10.1109/ACCESS.2017.2731851
96. S. Boyd, L. Vandenberghe, *Convex optimization*. (Cambridge University Press, USA, 2004)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.