Noname manuscript No.
(will be inserted by the editor)

Authenticated Secret Key Generation In Delay Constrained Wireless Systems

Miroslav Mitev* · **Arsenia Chorti** · **Martin Reed** · **Leila Musavian**

Received: date / Accepted: date

Abstract With the emergence of 5G low latency applications, such as haptics and V2X, low complexity and low latency security mechanisms are sought. 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 novel authenticated encryption using SKG; ii) a combined PUF / SKG authentication to reduce computational overhead; iii) a zero-round-trip-time (0-RTT) resumption authentication protocol; iv) pipelining of the SKG and the encrypted data transfer. With respect to the latter, 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 resource allocation is identified under security, power and delay constraints, by formulating the subcarrier allocation as a subset-sum $0-1$ knapsack optimization problem. A heuristic approach 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.

*Correspondence: M. Mitev
School of CSEE, University of Essex, Colchester, UK
E-mail: mm17217@essex.ac.uk

A. Chorti
ETIS, Université Paris Seine, Université Cergy-Pontoise, ENSEA, CNRS, Cergy-Pontoise, France
E-mail: arsenia.chorti@ensea.fr

M. Reed
School of CSEE, University of Essex, Colchester, UK
E-mail: mjreed@essex.ac.uk

L. Musavian
School of CSEE, University of Essex, Colchester, UK
E-mail: leila.musavian@essex.ac.uk

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]. For example, a 3GPP report on the security of ultra reliable low latency communication (URLLC) systems notes that "for a URLLC service with higher speed than 65 kbps, the 3GPP Release 15 radio access network (RAN) cannot fulfill the quality of service (QoS) requirement while enforcing user plane integrity protection" [3]. 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 public key encryption schemes – unless the key sizes increase to impractical lengths.

In the past years, physical layer security (PLS) [4–6] has been studied as a possible alternative to classic, complexity based, security approaches. Notably, it is explicitly mentioned as a 6G enabling technology in the first white paper on 6G: "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 (KA) protocol between two communicating parties. The principle of secret key generation (SKG) from correlated observations was first studied in [7] and [8]. A straightforward SKG approach can be built by exploiting the reciprocity of the wireless fading coefficients between two terminals within the channel coherence time [9] and this paper builds upon this mechanism. This is pertinent to many forthcoming B5G applications that will require a strong, but nevertheless, lightweight security KA; in this direction, PLS may offer such a solution, or, complement existing algorithms. With respect to authentication, physical unclonable functions (PUFs) [10] could also enhance authentication and key agreement (AKA) in demanding scenarios, including (but not limited to) device to device (D2D) and tactile Internet. We note that others also point to using physical layer security to reduce the resource overhead in URLLC [11].

A further advantage of PLS is that it is information-theoretic secure [12], *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 [13] is a promising alternative to PKE for KA.

However, unauthenticated key generation is vulnerable to man in the middle (MiM) attacks. In this sense, PUFs, firstly introduced in [14], are seen as a low complex authentication mechanism that can be used in conjunction with the SKG. PUFs are promising lightweight alternative to the currently used authentication mechanisms which require high computational capabilities. As summarised in [10] the employment of PUFs can decrease the computational cost and have a high impact on the time complexity.

In this study, we first discuss how standard SKG schemes can be used to develop authenticated encryption (AE) primitives [15–17]. Subsequently, we introduce the joint use of PUF authentication and SKG in a zero-round-trip-time (0-RTT) [18, 19] approach, allowing to build quick authentication mechanisms with forward security. We further investigate the possibility of implementing the AE SKG scheme via the pipelining of SKG and encrypted data transmission by effective scheduling of the PHY resources (*i.e.*, by optimal allocation of the subcarriers in 5G resource blocks). This analysis is then extended to account for statistical delay quality of service (QoS) guarantees, a pertinent scenario in B5G.

With respect to the latter aspect, to perform an analysis accounting for QoS, the metric to be studied needs to be carefully considered. 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 [20], 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.

We formulate two combinatorial optimization scheduling problems to study the performance of the proposed approach. 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, a subset of which is used for SKG (in the sense of side information) and the rest for encrypted data transfer. The optimal subcarrier allocation under security and power constraints is identified, i) to optimize the long-term average rate and ii) to optimize the effective rate in a delay constrained scenario. We formulate a subset-sum $0-1$ knapsack problem [21], which is solved using dynamic programming techniques [22] and a proposed heuristic approach of linear complexity. We show that the heuristic approach – according to which the strongest subcarriers in terms of signal-to-noise ratio (SNR) should be used for encrypted data transfer and the weakest for SKG (in the sense of side information) – only induces a negligible penalty in terms of performance for any realistic set of parameters. Our findings are supported by numerical results, while the efficiency of the proposed 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, depending on the exact values of the system parameters.

The paper is organized as follows: a brief summary of employed methods within our study is given in Section 2, the general system model is introduced in Section 3.1. The joint use of PUF authentication and SKG is illustrated in Section 3.2, next, in Sections 3.3 and 3.4 we present an AE scheme using SKG and a resumption protocol, respectively. Subsequently, we evaluate the optimal power and subcarrier allocation policy considering both long term average rate in Section 4 and effective capacity in Section 5. In Section 6, the efficiency of the proposed hybrid approach is evaluated against that of an alternative sequential approach, while conclusions are presented in Section 7.

2 Methods

In the following, we begin by revisiting the standard SKG scheme and present an AE primitive based on it. Subsequently, we discuss the possibility of using PUF

authentication and move on to propose an authentication approach that exploits the use of resumption secrets as used in 0-RTT protocols. Then, we investigate a possible implementation of the AE SKG in which SKG side information and encrypted data transfer are pipelined; we refer to this as the *parallel* transmission approach. In further detail, in our proposal the key generation is pipelined with encrypted data transfer, *i.e.*, key generation side information (such as syndromes in block codes) and data encrypted with the key that corresponds to the side information are transmitted over the same 5G resource block(s), *i.e.*, in (multiple) frames of 12 orthogonal frequency division multiplexing (OFDM) subcarriers. To take into consideration practical wireless aspects, we further account for the impact of imperfect CSI measurements in the evaluation of the optimal subcarrier allocation to maximize the data rate. This is formulated as a subset-sum 0–1 knapsack problem, that is known to be solvable optimally in pseudo-polynomial time using dynamic programming techniques.

Analyzing the results, a trend in the allocation was found, leading to the proposal of a lightweight heuristic scheduling approach of linear complexity. This heuristic approach is based on ordering the subcarriers – in terms of SNR. To evaluate the impact of this ordering, the theory of order statistics is used. To show the benefits of using the *parallel* approach, its efficiency is compared with a *sequential* approach, where the encrypted data transfer takes place in a subsequent frame after the key generation process has been concluded at both parties. The efficiency of both methods are compared by simulations in Matlab.

In addition, using the theory of the effective capacity, a further step in our study is taken where the system is also constrained by a statistical delay limit. We introduce the concept of effective rate for the particular system and we find the optimal power and subcarrier allocations while satisfying a delay-outage probability constraint. By using combinatorial optimization tools and dynamic programming, we found that the same trend appears in the optimal subcarrier allocation as in the previous, non-delay constrained, case. Furthermore, the achievable effective rates using the proposed optimal dynamic programming solution or the simple heuristic approach are compared through numerical evaluation.

In brief, in our study, methods that cut across multiple disciplines have been employed, e.g., encryption and network security including AE and resumption secrets, combinatorial optimization and dynamic programming, Shannon capacity and effective rate under statistical delay QoS constraints, order statistics and convex optimization.

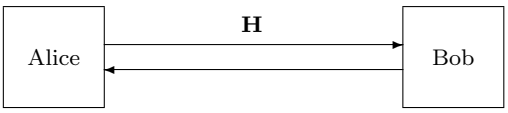

Fig. 1 Alice and Bob exchange pilot signals over a Rayleigh fading channel with realization $\mathbf{H} = [H_1, \ldots, H_N]$ in order to distill a shared secret key.

3 Authenticated SKG and Node Authentication Using PUFs

3.1 SKG System Model

The SKG system model is shown in Fig. 1. 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 [9]. Furthermore, Alice and Bob communicate over a BF-AWGN channel that comprises N orthogonal subcarriers. The fading coefficients, denoted by $H_j, j = 1, \ldots, N$, are assumed to be independent and identically distributed (i.i.d), $H_j \sim \mathcal{CN}(0, \sigma^2)$. Although in actual multicarrier systems neighbouring 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 [23–25] and will not enhance the problem formulation in the following Sections.

The SKG procedure encompasses three phases: *advantage distillation*, *information reconciliation*, and *privacy amplification* [7], [8] as described below:

1) Advantage distillation: This phase takes place over two periods. The legitimate nodes sequentially exchange constant probe signals with power P on all subcarriers[1], 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 [26,27]. 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 [28]. At the end of this phase, Alice and Bob obtain observations $X_{A,j}, X_{B,j}$, respectively, on the j-th subcarrier that can be expressed as:

$$X_{A,j} = \sqrt{P} H_j + Z_{A,j}, \tag{1}$$

$$X_{B,j} = \sqrt{P} H_j + Z_{B,j}, \tag{2}$$

[1] An explanation of the optimality of this choice under different attack scenarios is discussed in [13].

$j = 1, \ldots, N$, where by $Z_{A,j}, Z_{B,j}$ we denote zero-mean, unit variance circularly-symmetric complex AWGN random variables, $(Z_{A,j}, Z_{B,j}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_2)$. At the end of this phase the observations $X_{A,j}, X_{B,j}, j = 1 \ldots, N$ are quantized [29], so that Alice and Bob distill binary vectors $\mathbf{r}_{A,j}, \mathbf{r}_{B,j}, j = 1, \ldots, N$ respectively.

2) Information reconciliation: Due to the presence of noise, $\mathbf{r}_{A,j}, \mathbf{r}_{B,j}, j = 1, \ldots, N$ 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 encoding, the distilled binary vectors can be reconciled to corresponding codewords $\mathbf{c}_j, j = 1, \ldots, N$, with

$$\mathbf{r}_{A,j} = \mathbf{c}_j + \mathbf{e}_{A,j}, \tag{3}$$

$$\mathbf{r}_{B,j} = \mathbf{c}_j + \mathbf{e}_{B,j}. \tag{4}$$

Numerous practical information reconciliation approaches using standard forward error correction codes (e.g., LDPC, BCH, etc.,) have been proposed [9], [28]. As an example, if a block encoder with parity check matrix $\mathbf{Q}$ is used, then for the errors in the local observations the following hold [28]:

$$\mathbf{Q}\mathbf{e}_{A,j}^T = \mathbf{S}_{A,j}, \tag{5}$$

$$\mathbf{Q}\mathbf{e}_{B,j}^T = \mathbf{S}_{B,j}, \tag{6}$$

where $\mathbf{S}_{A,j}, \mathbf{S}_{B,j}$ denote the syndromes of $\mathbf{r}_{A,j}, \mathbf{r}_{B,j}$ with respect to the codeword $\mathbf{c}_j$ for $j = 1, \ldots, N$. To perform reconciliation, Alice (or Bob) transmit their corresponding syndrome $\mathbf{S}_{A,j}$ ($\mathbf{S}_{B,j}$), so that both parties can reconcile $\mathbf{r}_{A,j}, \mathbf{r}_{B,j}$ to $\mathbf{r}_{A,j}$ or $\mathbf{c}_j, j = 1, \ldots, N$. In this work, we assume that $\mathbf{c}_j, j = 1, \ldots, N$ the reconciliation information (e.g., the transmission of syndromes in the previous example) takes place on the same subcarrier index, *i.e.*, the syndrome $\mathbf{S}_{A,j}$ is sent from Alice to Bob on subcarrier with index j.

3) Privacy amplification: The secret key is generated by hashing $[\mathbf{c}_1 \| \ldots \| \mathbf{c}_N]$, where $[\cdot\|\cdot]$ denotes concatenation of the corresponding binary vectors. To this end, modern hash functions can be employed, e.g., SHA-256. The privacy amplification step ensures that the generated keys are completely unpredictable by an adversary and that they have maximum entropy (*i.e.*, are uniformly distributed). Note that the final step of privacy amplification, is executed locally without any further information exchange.

3.2 Node Authentication Using PUFs

As shown in Sec. 3.1 the SKG procedure requires only a few simple operations such as quantization, syndrome calculation and hashing. However, for security against a spoofing attack 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, as noted above these existing schemes are computationally intensive and can lead to significant latency.

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 concerns PUF. The concept of PUF was first introduced in [14], its idea is to utilize the fact that every integrated circuit differs to others due to manufacturing variability [30,31] and cannot be cloned [32]. 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* [33–37]. 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 (as in the SKG scenario) to be used in a Slepian Wolf decoder for reconciliation of the dithered measurements. 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 send 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 [38].

On the other hand, PUF authentication protocols have very low computational overhead and require overall authentication times that can be less than 10 ms [34, 39]. Furthermore, our key generation scheme, proposed in Section 3.1, requires just a hashing operation and (syndrome) decoding. Hashing mechanisms such as SHA256 performed on an IoT device requires less than 0.3ms [39, 40]. 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 [41].

As a conclusion, by using PLS for key generation and PUF as an authentication mechanism, intensive computations are avoided, thus leading to significant reduction in the authentication and key generation latency compared to traditional mechanisms. Thus, using SKG in conjunction with PUF authentication is a promising EAP-TLS alternative. In future work we will examine the real possibilities of implementing such a mechanism in practical systems.

3.3 AE Using SKG

Under the system model in Fig. 1, the SKG rate on any subcarrier is (note that the noise variances are here normalized to unity for simplicity) [9, 42]:

$$R_k = \log_2\left(1 + \frac{P\sigma^2}{2 + \frac{1}{P\sigma^2}}\right), \tag{7}$$

while the corresponding *minimum* necessary reconciliation rate has been shown to be $h(H_{B,j}|H_{A,j})$ [8].

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 $\mathbf{m}$ 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 $\mathtt{G} : \mathcal{H} \to \mathcal{K} \times \mathcal{S}$, accepting as inputs N-dimensional vectors of complex numbers (the fading coefficients), and generating as outputs N binary vectors of sizes n and $n-k$, respectively, $n, k \in \mathbb{N}$, (in the key and the syndrome spaces), *i.e.*,

$$\mathtt{G}(\mathbf{H}) = (\mathbf{K}, \mathbf{S}_A), \tag{8}$$

  where $\mathbf{K} \in \mathcal{K}$ denotes the key obtained from $\mathbf{H}$ after privacy amplification and $\mathbf{S}_A = [\mathbf{S}_{A,1} \| \dots \| \mathbf{S}_{A,N}] \in \mathcal{S}$ is the concatenation of Alice's syndromes.
- A symmetric encryption algorithm, e.g., AES GCM, denoted by $\mathtt{Es} : \mathcal{K} \times \mathcal{M} \to \mathcal{C}$ where $\mathcal{C}$ denotes the ciphertext space with corresponding decryption $\mathtt{Ds} : \mathcal{K} \times \mathcal{C} \to \mathcal{M}$, such that

$$\mathtt{Es}(\mathbf{K}, \mathbf{m}) = \mathbf{c}, \tag{9}$$

$$\mathtt{Ds}(\mathbf{K}, \mathbf{c}) = \mathbf{m}, \tag{10}$$

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

$$\mathtt{Sign}(\mathbf{K}, \mathbf{m}) = \mathbf{t}, \tag{11}$$

$$\mathtt{Ver}(\mathbf{K}, \mathbf{m}, \mathbf{t}) = \begin{cases} yes, \text{ if integrity verified} \\ no, \text{ if integrity not verified} \end{cases} \tag{12}$$

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 $[\mathbf{S}_A \| \mathbf{c} \| \mathbf{t}]$.
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 $\mathbf{m}=\mathtt{Ds}(\mathbf{K}_e, \mathbf{c})$.

3.4 Resumption Protocol

In Section 3.2 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 transport layer security (TLS) [18]. 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 [19]. 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.

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 [19]. 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 3.2. 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 k. Then referring to the notation and steps in Section 3.1:

1. Advantage distillation phase is carried out as before (See section 3.1), where both parties obtain channel observations and obtain the vectors $\mathbf{r}_A$ and $\mathbf{r}_B$ (channel indicies are dropped here for simplicity).
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 $\mathbf{r}_B$, obtaining 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 $\mathbf{c} \oplus \mathbf{R}_s$.
3. The privacy amplification step in Section 3.1 is carried out as before, but now on the key space $\mathbf{c} \oplus \mathbf{R}_s$ 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 [19].

4 Pipelined SKG and Encrypted Data Transfer

We have discussed in Section 3.1 how Alice and Bob can distill secret keys from estimates of the fading coefficients in their wireless link and in Section 3.3 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[2]. 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

[2] 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.

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 [43]. 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 3.1, 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 [9, 42] with power P. Using the observations (1), Alice estimates the channel coefficients as

$$\hat{H}_j = H_j + \tilde{H}_j, \tag{13}$$

for $j = 1, \ldots, N$ where $\tilde{H}_j$ denotes an estimation error that can be assumed to be Gaussian, $\tilde{H}_j \sim \mathcal{CN}(0, \sigma_e^2)$ [44]. Under this model, the following rate is achievable on the j-th subcarrier from Alice to Bob when the transmit power during data transmission is p_j [44]:

$$R_j = \log_2\left(1 + \frac{g_j p_j}{\sigma_e^2 P + 1}\right) = \log_2(1 + \hat{g}_j p_j), \tag{14}$$

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

$$\sum_{j=1}^{N} p_j \leq NP, \;\; p_j \geq 0, \; \forall j \in \{1, \ldots, N\}, \tag{15}$$

is achieved with the well known waterfilling 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 (15). In the following, the estimated channel gains $\hat{g}_j$ are – without loss of generality – assumed ordered in descending order, so that:

$$\hat{g}_1 \geq \hat{g}_2 \geq \ldots \geq \hat{g}_N. \tag{16}$$

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 waterfilling 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 3.1 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.

4.1 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 3.1) 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 $\bar{\mathcal{D}}$ with cardinality $|\bar{\mathcal{D}}| = N - D$ used for reconciliation such that, $\mathcal{D} \cup \bar{\mathcal{D}} = \{1, \ldots, N\}$.

Over $\mathcal{D}$ the achievable sum data transfer rate, denoted by C_D is given by

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

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

$$C_R = \sum_{i \in \bar{\mathcal{D}}} \log_2(1 + \hat{g}_i p_i). \tag{18}$$

As stated in Section 3.1 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, $i \in \bar{\mathcal{D}}$, can be expressed as [45]:

$$f(g_i) = \frac{N!}{\sigma^2 (N-i)!(i-1)!} \left(1 - e^{-\frac{\hat{g}_i}{\sigma^2}}\right)^{N-i} \left(e^{-\frac{\hat{g}_i}{\sigma^2}}\right)^{i}, \quad (19)$$

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:

$$\sigma_i^2 = \sigma^2 \sum_{q=i}^{N} \frac{1}{q^2}, \quad i \in \{D+1, \ldots, N\}. \quad (20)$$

Thus, we can now write the SKG rate as:

$$C_{SKG} = \sum_{i \in \bar{\mathcal{D}}} R_k = \sum_{i \in \bar{\mathcal{D}}} \log_2 \left(1 + \frac{P\sigma_i^2}{2 + \frac{1}{P\sigma_i^2}}\right). \quad (21)$$

The minimum rate necessary for reconciliation has been theoretically derived in [8]. Here, alternatively, we employ a more practical design approach in which the rate of the employed encoder is explicitly taken into account. Noting that in a rate $\frac{k}{n}$ block encoder (*i.e.*, $n-k$ syndrome bits for a message of k bits) the relative rate of syndrome to the key (of length k) 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 $\kappa = \frac{n-k}{k} + 1$ 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, $\kappa = 2$, for $\frac{k}{n} = \frac{1}{3}$, $\kappa = 3$, while for $\frac{k}{n} = \frac{1}{4}$, $\kappa = 4$. Based on this discussion, we capture the minimum requirement for the reconciliation rate through the following expression:

$$C_R \geq \kappa C_{SKG}. \quad (22)$$

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 [3]. 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

$$C_{SKG} \geq \beta C_D, \;\; 0 < \beta \leq 1, \quad (23)$$

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 (22) and (23) we formulate the following maximization problem:

$$\max_{p_j, j \in \mathcal{D}} \sum_{j \in \mathcal{D}} R_j \quad (24)$$

$$\text{s.t. } (15), (22), (23),$$

$$\sum_{j \in \mathcal{D}} R_j + \sum_{i \in \bar{\mathcal{D}}} R_i \leq C. \quad (25)$$

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

$$\sum_{j \in \mathcal{D}} R_j \leq \frac{\sum_{i \in \bar{\mathcal{D}}} R_i}{\kappa \beta}. \quad (26)$$

The optimization problem at hand is a mixed-integer convex optimization problem with unknowns both the sets $\mathcal{D}, \bar{\mathcal{D}}$, 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.

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 (25) 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 (15) is also satisfied with equality. Thus, if we follow that approach, we determine the power allocation vector uniquely and can combine the remanining constraints (25) and (26) into a single one as:

$$\sum_{j \in \mathcal{D}} R_j \leq \frac{C}{\kappa \beta + 1}. \quad (27)$$

The new optimization problem can be re-written as

$$\max_{x_j \in \{0,1\}} \sum_{j=1}^{N} R_j x_j \quad (28)$$

$$\text{s.t. } \sum_{j=1}^{N} R_j x_j \leq \frac{C}{1 + \kappa \beta}. \quad (29)$$

The problem in (28)-(29) is a subset-sum problem from the family of $0-1$ knapsack problems, that is known to be NP hard [21]. However, these type of problems are solvable optimally using dynamic programming techniques in pseudo-polynomial time [21,22]. Furthermore, it is known that greedy heuristic approaches are bounded away from the optimal solution by half [46].

Algorithm 1 Heuristic Greedy Algorithm for (28)-(29)

procedure HEURISTIC(start, end, R_j)
  $j \leftarrow 1, C_0 \leftarrow 0, R_{N+1} \leftarrow 0$
  while $j \leq N-1$ and $C_j \leq \frac{C}{1+\kappa\beta}$ **do**
    $C_j \leftarrow C_{j-1} + R_j$
    if $C_j \leq \frac{C}{1+\kappa\beta}$ **then**
      $j \leftarrow j+1$
    else do $C_j \leftarrow C_j - R_j; R_j \leftarrow 0; j \leftarrow j+1$
    end if
  end while
end procedure

We propose a simple greedy heuristic algorithm with *linear complexity*, as follows. Let us assume that the estimated channel gains, and, consequently, the rates R_j are ordered in descending order (the ordering is a $\mathcal{O}(N \log N)$ operation, so if the gains are not ordered the overall complexity will be dominated by the sorting operation). The data subcarriers are selected starting from the best – in terms of SNR – until (29) 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 (29) 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:

$$\eta_{\text{parallel}} = \frac{\mathbb{E}\left[\sum_{i\in\mathcal{D}} R_i\right]}{\mathbb{E}[C]}. \tag{30}$$

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 [11].

4.2 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

$$C_{SKG} = N \log_2\left(1 + \frac{P\sigma^2}{2 + \frac{1}{P\sigma^2}}\right). \tag{31}$$

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.*, it's long-term value) $\mathbb{E}[C_R]$. The average number of frames needed for reconciliation is then computed as:

$$M = \left\lceil \frac{\kappa C_{SKG}}{\mathbb{E}[C_R]} \right\rceil, \tag{32}$$

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:

$$L = \left\lfloor \frac{C_{SKG}}{\beta \mathbb{E}[C]} \right\rfloor, \tag{33}$$

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

$$\eta_{\text{sequential}} = \frac{L}{L+M}. \tag{34}$$

5 Effective Data Rate Taking into Account Statistical Delay QoS Requirements

Here, 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 [47] 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:

$$\lim_{x\to\infty} \frac{\ln(\Pr[Q(\infty) > x])}{x} = -\theta, \tag{35}$$

where θ indicates the asymptotic exponential decay-rate of the overflow probability. For a large threshold x, (35) can be represented as $Pr[Q(\infty) > x] \approx e^{-\theta x}$. Furthermore, the delay-outage probability can be approximated by [20] :

$$\Pr^{\text{out}}_{\text{delay}} = \Pr[\text{Delay} > D_{\max}] \approx \Pr[Q(\infty) > 0] e^{-\theta\zeta D_{\max}}, \tag{36}$$

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, 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 [20]:

$$E_C(\theta) = -\lim_{t\to\infty} \frac{1}{\theta} \ln \mathbb{E}[e^{-\theta S[t]}] (\text{bits/s}), \tag{37}$$

where $S[t] = \sum_{i=1}^{t} s[i]$ denotes the time-accumulated service process, and $s[i], i = 1, 2, ...$ denotes the discrete-time stationary and ergodic stochastic service process. From the above it can be seen that 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 i^{th} frame. In the context of the investigated data and reconciliation information transfer, $\tilde{R}_i$, is given by:

$$\tilde{R}_i = \frac{1}{F} \sum_{k \in \mathcal{D}} \log_2(1 + p_k \hat{g}_k), \tag{38}$$

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 [48,49] is satisfied, the *effective data rate*[3] $E_C(\theta)$ is given as:

$$E_{C,\mathcal{D}}(\theta) = -\frac{1}{\theta T_f B} \ln\left(\mathbb{E}\left[e^{-\theta T_f B \tilde{R}_i}\right]\right). \tag{39}$$

We set $\alpha = \frac{\theta T_f B}{\ln(2)}$. By inserting (38) into (39) we get:

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

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

Assuming i.i.d. channel gains, by using the distributive property of the mathematical expectation, (40) becomes [50]:

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

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

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

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

[3] 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 rigour. We note that we assume the information data and reconciliation information are accumulated in separate independent buffers within the transmitter.

Using that, we now reformulate the maximization problem given in (24) by adding a delay constraint. The reformulated problem can be expressed as follows:

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

$$\text{s.t. } (15), (26),$$

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

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.

In the proposed approach, we assume that the constraint (44) is satisfied with equality. Given that, the optimal power allocation can now be evaluated from (15) and (44) using convex optimization tools. First, since $\log(\cdot)$ is monotonically increasing concave function for any $\theta > 0$, solving the optimization problem in (43) which finds the optimal power allocation, *i.e.*, E_C^{opt} is equivalent to solving the following minimization problem:

$$\min_{p_i, i=1,2,...N} \sum_{i=1}^{N} \left(\mathbb{E}\left[(1 + p_i \hat{g}_i)^{-\alpha N^{-1}}\right]\right), \tag{45}$$

$$\text{s.t. } (15).$$

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

$$\mathcal{L} = \left(\mathbb{E}\left[(1 + p_i \hat{g}_i)^{-\alpha N^{-1}}\right]\right) + \lambda \left(\sum_{i=1}^{N} p_i - NP\right). \tag{46}$$

By differentiating (46) w.r.t. p_i and setting the derivative equal to zero [51] we get:

$$\frac{\partial \mathcal{L}}{\partial p_i} = \lambda - \frac{\alpha \hat{g}_i}{N} (\hat{g}_i p_i + 1)^{-\frac{\alpha}{N} - 1} = 0. \tag{47}$$

Solving (47) gives the optimal power allocation policy:

$$p_i^* = \frac{1}{g_0^{\frac{N}{\alpha+N}} \hat{g}_i^{\frac{\alpha}{\alpha+N}}} - \frac{1}{\hat{g}_i}, \tag{48}$$

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 $E_C^{\text{opt}}(\theta)$:

$$E_C^{\text{opt}}(\theta) = -\frac{1}{\alpha} \sum_{i=1}^{N} \log_2\left(\mathbb{E}\left[\left(\frac{\hat{g}_i}{g_0}\right)^{-\frac{\alpha}{\alpha+N}}\right]\right) \tag{49}$$

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 (48) we can easily find the optimal subcarrier allocation that satisfies (26). As in Section 4 to do that we first formulate a

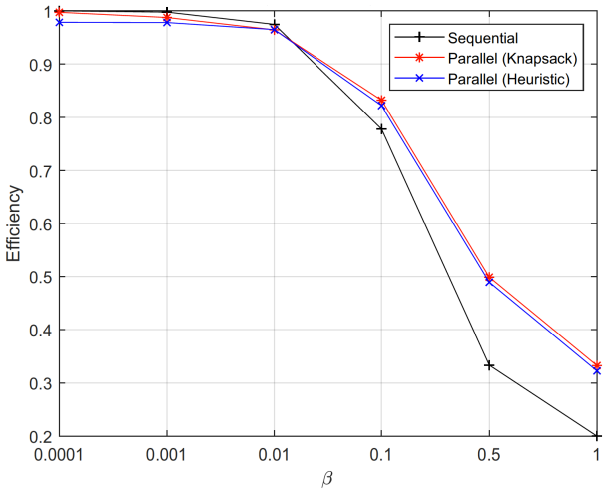

Fig. 2 a) Efficiency comparison for $N = 12$, SNR=10 dB and $\kappa = 2$.

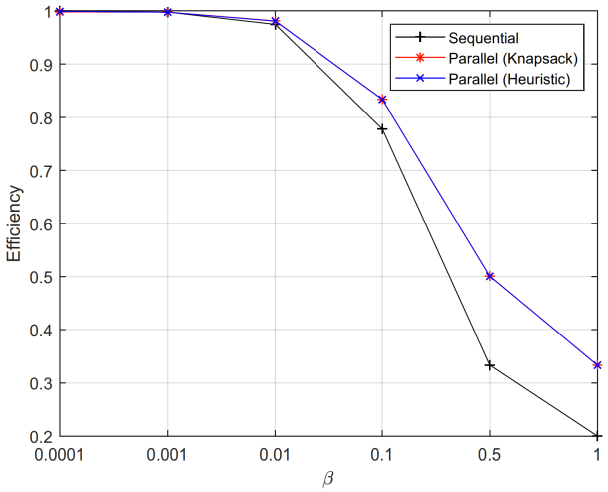

Fig. 2 b) Efficiency comparison for $N = 64$, SNR=10 dB and $\kappa = 2$.

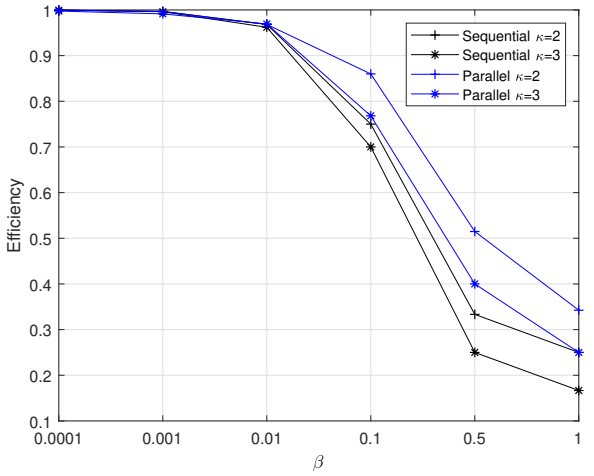

Fig. 3 Efficiency vs κ, for $N = 24$, SNR=10 dB.

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.*

6 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.* In this Section we present numerical results for both long term average data rate C_D given in (17) and *effective data rate* $E_{C,D}(\theta)$ given in (41), however, for better illustration of each case they are separated into different subsections.

6.1 Numerical results for the case long term average C_D

Figures 2a and 2b 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 and the 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.

Next, in Fig. 3 the efficiency of the parallel and the sequential methods are shown for two different values of

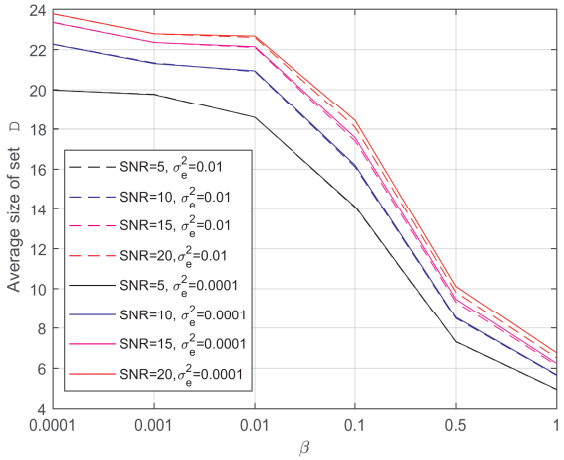

Fig. 4 a) Size of set $\mathcal{D}$ for different SNR levels and σ_e^2 when $N = 24$.

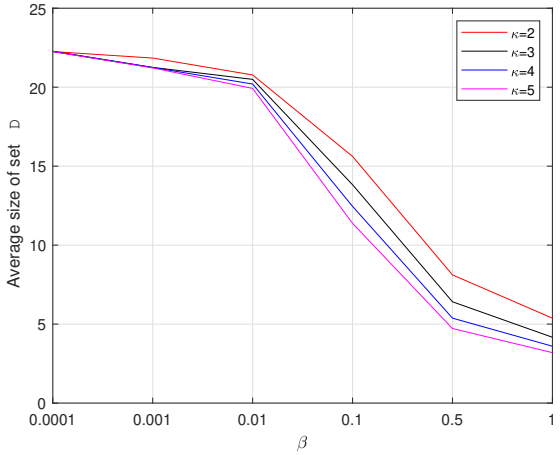

Fig. 4 b) Size of set $\mathcal{D}$ for different values of κ when $N = 24$.

$\kappa \in \{2, 3\}$ 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. 4, 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. 4a) and κ (Fig. 4b), for $N = 24$. As expected, in Fig. 4a 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 reconcilliation 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. 4b is set to 10 dB. The figure shows that when increasing κ the size of set $\mathcal{D}$ decreases. This result can be easily predicted from inequality (22), meaning, when κ increases more reconciliation data has to be sent, hence fewer subcarriers can be used for data. In both Fig. 4a and Fig. 4b when β increases the size of set $\mathcal{D}$ decreases; this effect is a consequence of constraint (29) as the data rate is decreasing with β.

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

In Fig. 5 we see the achieved *effective data rate* $E_{C,D}(\theta)$ given in (41), for different values of N and θ while the SNR=5 dB and $\kappa = 2$. Fig. 5a 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.

Fig. 5b illustrates the case when $N = 12$ and $\theta = 100$ (very stringent delay constraint). For this case we can see that for small values of β the sequential approach performs better than the parallel. As mentioned in Section 5 as θ increases the power allocation transforms from water-filling to total channel inversion. Consequently, the rate achieved on all subcarriers converges to the same value, hence when we have small number of subcarriers (such as $N = 12$) and small values of β[4] then using a single subcarrier for reconciliation data will be more than what is 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. 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 $\bar{\mathcal{D}}$ will be fully or almost fully utilised and the parallel approach shows better performance for these values.

In the next two Fig.: 5c and 5d 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 perfom 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. 5 is that looking at the parallel approach, it can easily be seen that in all

[4] *i.e* that the ratio of reconciliation information to data is small as seen from Eq. (26))

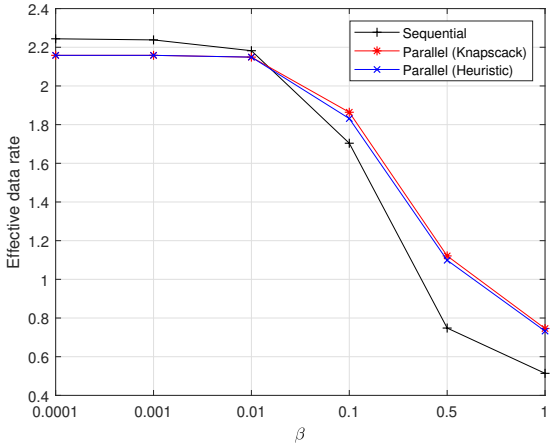

Fig. 5 a) Effective data rate achieved by parallel and sequential approaches when $N = 12$, SNR= 5dB, $\theta = 0.0001, \kappa = 2$.

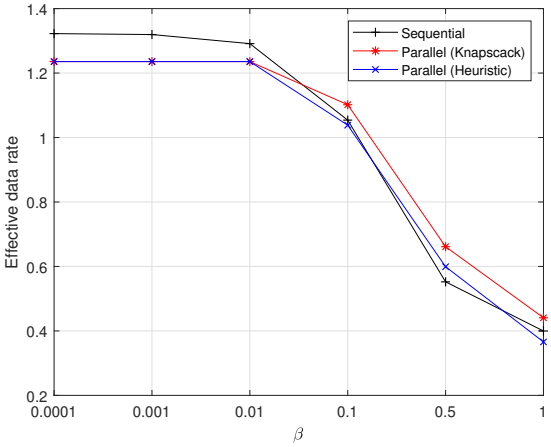

Fig. 5 b) Effective data rate achieved by parallel and sequential approaches when $N = 12$, SNR= 5dB, $\theta = 100, \kappa = 2$.

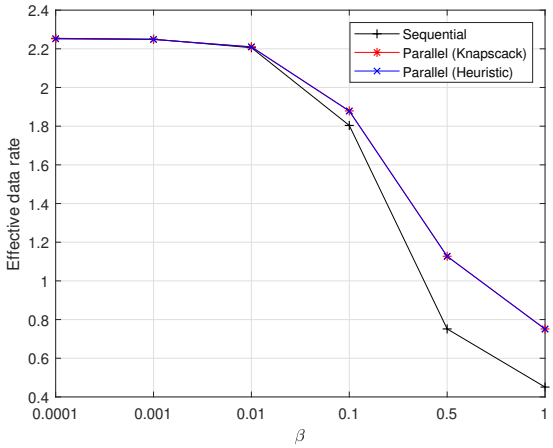

Fig. 5 c) Effective data rate achieved by parallel and sequential approaches when $N = 64$, SNR= 5dB, $\theta = 0.0001, \kappa = 2$.

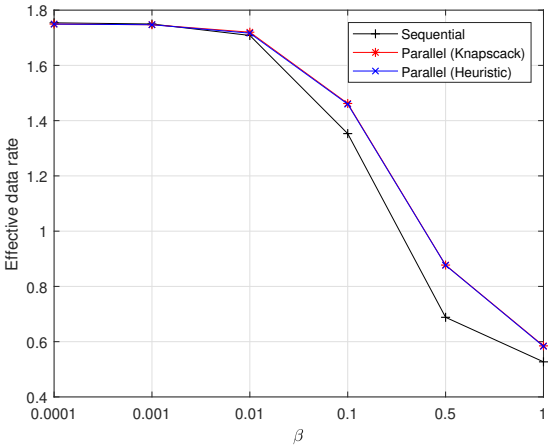

Fig. 5 d) Effective data rate achieved by parallel and sequential approaches when $N = 64$, SNR= 5dB, $\theta = 100, \kappa = 2$.

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. Due to this, for the next figures we exclude the performance of the Knapsack algorithm and use only the heuristic presented in *Algorithm 1*.

In Fig. 6 we give a three-dimensional plot showing the dependence of the achievable *effective data rate* $E_{C,D}(\theta)$ on β and θ. Figures 6a and 6b compare the parallel heuristic approach and the sequential approach for high SNR levels, whereas Fig. 6c and 6d compare their performance for low SNR level. In Fig. 6a and 6c we have $N = 12$ while in Fig. 6b and 6d 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 above, for small values of β and N in the parallel approach the devoted part of the total achievable effective capacity $E_C^{\text{opt}}(\theta)$ to reconciliation (syndrome communication) is more than what is required and this can be seen in Fig. 6a and 6c. When the SNR is high (See Fig. 6a) this effect is mostly noticeable for large values of θ and small values of β, whereas for small values of β and θ both algorithms perform identical. A similar trend can be seen at the low SNR

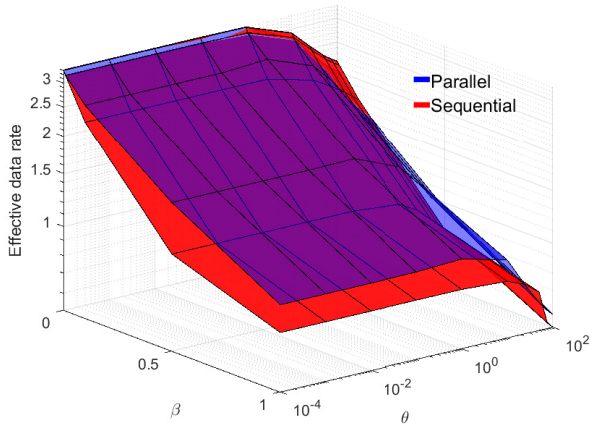

Fig. 6 a) Effective data rate achieved by the parallel heuristic approach and the sequential approach when $N = 12$, SNR= 10 dB and $\kappa = 2$.

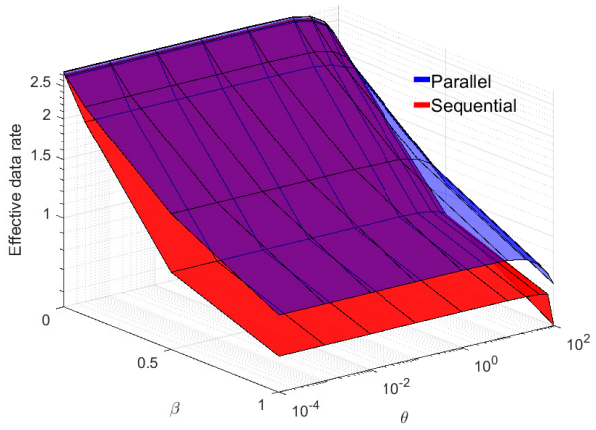

Fig. 6 b) Effective data rate achieved by the parallel heuristic approach and the sequential approach when $N = 64$, SNR= 10 dB and $\kappa = 2$.

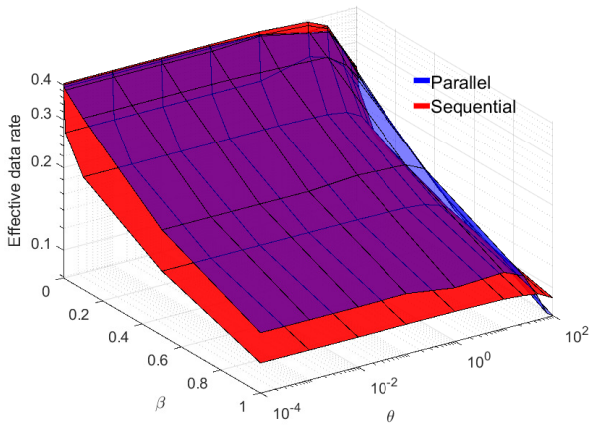

Fig. 6 c) Effective data rate achieved by the parallel heuristic approach and the sequential approach when $N = 12$, SNR= 0.2 dB and $\kappa = 2$.

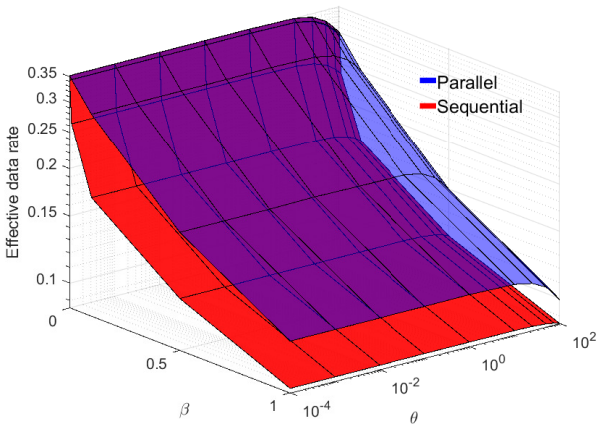

Fig. 6 d) Effective data rate achieved by the parallel heuristic approach and the sequential approach when $N = 64$, SNR= 0.2 dB and $\kappa = 2$.

regime in Fig. 6c. However, having a low SNR affects the sequential approach and its effectiveness decreases. 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 but as mentioned above there are different factors that influence it, which leads to identical performance by both parallel and sequential approaches. However, in both scenarios high SNR Fig. 6a and low SNR Fig. 6c, 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. 6b and 6d, 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. 6b and 6d, respectively.

6.3 Discussion

In this study we have indicated that SKG and PUF technologies can be exploited to build latency aware hybrid crypto-PLS systems, in which encryption schemes

are combined with PLS to generate AE and 0-RTT primitives.

Furthermore it was shown that pipelining the key agreement and the encrypted data transmission in a parallel approach is more efficient than a sequential approach, for most cases. The only instances where the sequential scheme is better is when there are a small number of subcarriers, due to the fact that we chose to transmit only reconciliation information on the optimised subcarriers; this choice can be dropped in future studies. 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. In the future, explicit delay calculations will be performed to further scrutinize the results presented in this study, that concerned only statistical delay guarantees.

7 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; ii) by adding a delay constraint, represented by parameter θ, to the security and power constraint we found the optimal *effective data rate*.

To finalise our study we illustrated through numerical comparisons the efficiency of the proposed parallel method, in which SKG and data transfer are interweaved 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. 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.

List of abbreviations

Declarations

Availability of data and material

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.

7.1 *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, Martin Reed is supported by the project SerIoT which has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 780139, Leila Musavian is supported by the RECENT project which has recevied funding from European Union Horizon 2020, RISE 2018 scheme (H2020-MSCA-RISE-2018) under Marie Sklodowska – Curie grant agreement No. 823903.

7.2 *Authors' contributions*

MM, AC, 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.

7.3 *Acknowledgements*

Not applicable

References

1. A. Mukherjee, "Physical-Layer Security in the Internet of Things: Sensing and Communication Confidentiality under Resource Constraints," *Proc. IEEE*, vol. 103, no 10, Oct. 2015.
2. A. Yener, S. Ulukus, "Wireless Physical-Layer Security: Lessons Learned From Information Theory," *Proc. IEEE*, vol. 103, No. 10, October 2015.
3. 3rd Generation Partnership Project, Technical Specification Group Services and System Aspects, available online https://www.3gpp.org/ftp/Specs/archive/33_series/33.825/, 3GPP TR 33.825 V0.3.0, Study on the Security for 5G URLLC (Release 16)
4. A. Chorti, C. Hollanti, J.-C. Belfiore, H.V. Poor, Physical Layer Security: A Paradigm Shift in Data Confidentiality, Springer, *Lecture Notes in Electrical Engineering - Physical and Data-Link Security Techniques for Future Communication Systems*, vol. 358, pp. 1-15, Sep. 2015.
5. A. Chorti, K. Papadaki, H.V. Poor, "Optimal power allocation in block fading channels with confidential messages", *IEEE Trans. Wireless Commun.*, vol. 14, no 9, pp. 4708-4719, Sep. 2015.
6. A. Chorti, S. Perlaza, Z. Han, H.V. Poor, "On the resilience of wireless multiuser networks to passive and active eavesdroppers", *IEEE J. Sel. Areas Commun.*, Special issue on signal processing techniques for wireless physical layer security, vol. 31 no 9, pp. 1850-1863, Sep. 2013.
7. U. Maurer. "Secret key agreement by public discussion based on common information," *IEEE Trans. Inf. Theory*, Vol. 39, No. 5, pp. 733 – 742, May 1993.
8. R. Ahlswede and I. Csiszar. "Common randomness in information theory and cryptography – part I: Secret sharing," *IEEE Trans. Inf. Theory*, Vol. 39, No. 7, pp. 1121 – 1132, Jul 1993.
9. C. Ye, A. Reznik and Y. Shah. "Extracting secrecy from jointly Gaussian random variables," *Proc. IEEE Int. Symp. Inf. Theory (ISIT)* pp. 2593 – 2597, Jul 2006.
10. R. Maes, I. Verbauwhede (2010). "Physically Unclonable Functions: A Study on the State of the Art and Future Research Directions", *Towards Hardware-Intrinsic Security*, pp 3 – 37, Oct 2010.
11. A. Weinand, M. Karrenbauer, H. Schotten, "Security Solutions for Local Wireless Networks in Control Applications based on Physical Layer Security" *IFAC-PapersOnLine*, vol. 51, pp. 32 – 39, Elsevier B.V., 2018
12. A. Mukherjee, S. A. A. Fakoorian, J. Huang and A. L. Swindlehurst, "Principles of Physical Layer Security in Multiuser Wireless Networks: A Survey," *IEEE Communications Surveys and Tutorials*, vol. 16, no. 3, pp. 1550 – 1573, 2014.
13. A. Chorti, "A study of injection and jamming attacks in wireless secret sharing systems," *in Proc. Workshop on Communication Security (WCS)*, EUROCRYPT, Mar 2017.
14. R. Pappu, B. Recht, J. Taylor, N. Gershenfeld, "Physical One-Way Functions", *Science*, Vol. 297, pp. 2026 – 2030, Sep 2002.
15. M. Bellare, C. Namprempre. "Authenticated encryption: Relations among notions and analysis of the generic composition paradigm". *Advances in Cryptology – ASIACRYPT 2000*, vol. 1976, *Lecture Notes in Computer Science*, pp. 531 – 545, Springer-Verlag, Berlin Germany, Dec 2000.
16. T. Krovetz , P. Rogaway, "The Software Performance of Authenticated-Encryption Modes", *Fast Software Encryption (FSE)*, *Lecture Notes in Computer Science*, vol. 6733, Springer, Berlin, 2011.
17. S. Koteshwara, A. Das, "Comparative Study of Authenticated Encryption Targeting Lightweight IoT Applications," *IEEE Design and Test*, vol. 34, pp. 26 – 33, Aug 2017.
18. Rescorla, E.: The Transport Layer Security (TLS) Protocol Version 1.3. RFC 8446 (2018), https://rfc-editor.org/rfc/rfc8446.txt
19. Aviram N., Gellert K., Jager T. "Session Resumption Protocols and Efficient Forward Security for TLS 1.3 0-RTT", *EUROCRYPT 2019*, Lecture Notes in Computer Science, vol. 11477, Springer.
20. D. Wu and R. Negi, "Effective capacity: A wireless link model for support of quality of service", *IEEE Trans. wireless Commun.*, vol. 2, no. 4, pp. 630 – 643, Jul 2003.
21. S. Martello and Paolo Toth, "*Knapsack problems: algorithms and computer implementations*", John Wiley and Sons, Inc. New York, NY, 1990
22. H. Kellerer and U. Pferschy, D. Pisinger, "*Knapsack Problems*", Springer-Verlag Berlin Heidelberg, 2004
23. C. Chen and M. A. Jensen, "Secret Key Establishment Using Temporally and Spatially Correlated Wireless Channel Coefficients", *IEEE Trans. Mobile Comput.*, vol. 10, pp. 205 – 215, Feb 2011.
24. J. Zhang and A. Marshall and R. Woods and T. Q. Duong, "Efficient Key Generation by Exploiting Randomness From Channel Responses of Individual OFDM Subcarriers", *IEEE Trans. Commun*, vol. 64, pp. 2578 – 2588, Jun 2016.
25. J. Zhang and B. He and T. Q. Duong and R. Woods, "On the Key Generation From Correlated Wireless Channels", *IEEE Commun. Lett.*, vol. 21, pp. 961 – 964, Apr 2017.
26. A. Chorti, "A Study of Injection and Jamming Attacks in Wireless Secret Sharing Systems", Springer, *Lecture Notes in Electrical Engineering*, Sep 2017.
27. M. Mitev, A. Chorti, V. Belmega, M. Reed, "*Man-in-the-Middle and Denial of Service Attacks in Wireless Secret Key Generation*", to appear in IEEE Proc. Global Communications Conference (Globecom), Dec 2019.
28. C. Saiki, A. Chorti, "A novel physical layer authenticated encryption protocol exploiting shared randomness", *Proc. IEEE Conf. Commun. Netw. Security (CNS)*, pp. 113 – 118, Florence, Italy, 2015
29. Qian Wang, Hai Su, Kui Ren and Kwangjo Kim, "Fast and scalable secret key generation exploiting channel phase randomness in wireless networks", *Proc. IEEE Int. Conf. Computer Commun. (INFOCOM)*, pp. 1422 – 1430, 2011.
30. C. Herder, M. Yu, F. Koushanfar and S. Devadas, "Physical Unclonable Functions and Applications: A Tutorial", *Proc. IEEE*, Vol. 102, pp. 1126 – 1141, Aug 2014.
31. G. E. Suh and S. Devadas, "Physical Unclonable Functions for Device Authentication and Secret Key Generation," 44th ACM/IEEE Design Automation Conference, San Diego, pp. 9 – 14, CA, 2007.
32. C. Böhm and M. Hofer. 2012. "*Physical Unclonable Functions in Theory and Practice*", Springer, New York, 2013.
33. U. Chatterjee, R. S. Chakraborty, and D. Mukhopadhyay, "A PUF-Based Secure Communication Protocol for IoT" *ACM Trans. Embed. Comput. Syst.*, Vol. 16, Article 67, Apr 2017.
34. M. N. Aman, M. H. Basheer and B. Sikdar, "Two-Factor Authentication for IoT With Location Information," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3335 – 3351, Apr 2019.
35. M. H. Mahalat, S. Saha, A. Mondal and B. Sen, "A PUF based Light Weight Protocol for Secure WiFi Authentication of IoT devices," *International Symposium on Embedded Computing and System Design (ISED)*, Cochin, India, pp. 183 – 187, 2018.

36. A. Braeken, "PUF based authentication protocol for IoT", Symmetry, vol. 10, 2018.
37. Y. Yilmaz, S. R. Gunn and B. Halak, "Lightweight PUF-Based Authentication Protocol for IoT Devices," *IEEE International Verification and Security Workshop (IVSW)*, pp. 38 – 43, Costa Brava, 2018.
38. S. Ahmad, A. H. Mir and G. R. Beigh, *"Latency evaluation of extensible authentication protocols in WLANs*," 2011 Fifth IEEE International Conference on Advanced Telecommunication Systems and Networks (ANTS), pp. 1 – 5, Bangalore, 2011.
39. P. Gope and B. Sikdar, "Lightweight and Privacy-Preserving Two-Factor Authentication Scheme for IoT Devices," *IEEE Internet Things J.*, Vol. 6, pp. 580 – 589, Feb 2019.
40. A. Ometov, P. Masek, L. Malina, R. Florea, J. Hosek, S. Andreev, J. Hajny, J. Niutanen, Y. Koucheryavy, *"Feasibility characterization of cryptographic primitives for constrained (wearable) IoT devices,"* 2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops), pp. 1 – 6, Sydney, 2016.
41. J. Cho and W. Sung, "Efficient Software-Based Encoding and Decoding of BCH Codes," *IEEE Trans. Comput*, vol. 58, pp. 878 – 889, Jul 2009.
42. E. V. Belmega and A. Chorti, "Protecting secret key generation systems against jamming: Energy harvesting and channel hopping approaches", *IEEE Trans. Inf. Forensics Security*, Vol. 12, Nov. 2017.
43. M. Mitev, A. Chorti and M. Reed, "*Optimal Resource Allocation in Joint Secret Key Generation and Data Transfer Schemes*," 2019 15th International Wireless Communications and Mobile Computing Conference (IWCMC), pp. 360 – 365, Tangier, Morocco, Jun. 2019.
44. M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel", *IEEE Trans. Inf. Theory*, vol. 46, pp. 933 – 946, May 2000.
45. H.-C. Yang and M.-S. Alouini, "*Order Statistics in Wireless Communicatiobs*", Cambridge University Press, NY, 2011.
46. V. Vazirani, "*Approximation Algorithms*", Springer-Verlag Berlin Heidelberg, (2003)
47. C. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks", *IEEE Transactions on Automatic Control*, Vol. 5, pp. 913 – 931, May 1994.
48. J. Gärtner, "On Large Deviation from Invariant Measure", *Theory Prob. Appl.*, Vol. 22, pp. 24 – 39, 1977.
49. R. Ellis, "Large deviations for a general class of random vectors", *Annu. Probab.*, Vol. 12, pp. 1 – 12, 1984.
50. T. Abrão, S. Yang, L. Sampaio, P. Jeszensky and L. Hanzo, "Achieving Maximum Effective Capacity in OFDMA Networks Operating Under Statistical Delay Guarantee", *IEEE Access*, Vol. 5, pp. 14333 – 14346, Jul 2017.
51. S. Boyd, L Vandenberghe, "*Convex Optimization*", Cambridge University Press, NY, USA, 2004.