Physical Layer Security—From Theory to Practice

Abstract—In this article, we present the evolutionary journey of physical layer security (PLS), starting from its theoretical foundations in wiretap coding and secret key generation followed by the revolutionary advancements in authentication using physical unclonable functions, localization, and RF fingerprinting. Since the middle of the last century, the theory behind PLS has been constantly evolving, however, practical solutions are still scarce. Through a critical review, we identify the key hurdles hindering the widespread adoption of PLS. Moreover, we highlight important research directions and possible solutions that can bridge the gap between theory and practice.

Introduction

The genesis of information theoretic security dates back to 1949 and Shannon’s pioneering work [1]. In his paper, Shannon looked at the security of cipher systems considering a three-terminal scenario, i.e., a legitimate source (Alice), a legitimate receiver (Bob), and an eavesdropper (Eve). An important assumption in this model is the absence of noise, which makes all transmissions perfectly available to Bob and Eve. However, Alice and Bob have an advantage over Eve in the form of a shared secret key $k$. In this setup, Shannon was interested in identifying the conditions under which the legitimate users can share a message $m$, while keeping it completely secret from Eve.

Shannon came to both positive and negative conclusions. An encouraging result was that perfect secrecy can be achieved. However, it is possible only if the secret key is of the same length as the message. In detail, it was demonstrated that xor-ing a message $m$ with a uniform random key $k$ of the same length (a scheme known as one-time pad) gives a ciphertext $c = m \oplus k$ that leaks no information about the plaintext, i.e., $H(c|e) = H(m)$, where $H(\cdot)$ denotes entropy. Although this result was relevant for the communication systems at that time, one-time pad schemes are impractical in the case of modern wireless systems, as the key itself would require secure distribution.

Shannon’s work on information theoretic security, while not directly applicable, highlighted the importance of randomness in keeping messages secure from eavesdroppers, i.e., having enough confusion at the eavesdropper allows the development of provably secure crypto systems. His ideas laid the foundation for physical layer security (PLS), which involves securing communication at the physical layer, as well as wiretap coding, a method of encoding messages to make them harder to intercept. In 1975, another researcher named A. Wyner looked at a similar scenario but focused on wireless channels, which are not perfect and transmissions are impaired by noise [2]. Wyner found that both receivers would receive different, noisy versions of the transmitted signal, which presented additional challenges for secure communication.

Similar to Shannon’s goal, Wyner was interested in identifying the conditions under which the message $m$ can be conveyed reliably to Bob, while keeping it secret from Eve. He presented a new interesting view on the problem by looking at the reliable rate to Bob versus the equivocation at Eve (perfect secrecy can be achieved if the former is less than the latter). Equivocation at Eve refers to the uncertainty or lack of information that she has about the original message after she intercepts the communication. To give a better perspective on this problem, Wyner [2] introduced a new metric known as secrecy capacity. Secrecy capacity gives the maximum reliable rate equal to the equivocation. Most importantly, Wyner [2] demonstrated that noise in wireless
In essence, the fundamental concept behind PLS is to leverage the unique properties of the physical world. This is in contrast to traditional crypto algorithms, where security goals are achieved in the digital domain through solving complex modulo-arithmetic operations. As a consequence, while both aim at achieving similar goals, PLS and traditional crypto algorithms have major differences.

For example, the lack of complex operations required by PLS could minimize latency. For instance, a typical crypto-based authentication procedure on a 1.73-GHz processor tablet requires approximately 165.5 ms in static scenarios and 336.7 ms in high mobility environments [9]. Reported experimental data reveal that a PUF-based authentication can be indeed faster (roughly 5.6 ms [10]).

Another difference comes from the attack model. An attack targeting crypto algorithms can be performed directly at the digital world, while an attack directed toward PLS would require some sort of physical presence. In this sense, access to quantum computers would give the chance to break various cryptographic algorithms that are widely used today. This is due to their ability to solve certain mathematical problems, which form the basis of security today, (e.g., search through an unsorted database or factoring large numbers) exponentially faster than classical computers. Due to that, the NIST is actively investigating the topic of quantum resistance and postquantum cryptography. While some algorithms have been already approved, they are based on increased latency and complexity.

Differently, PLS does not require solving complex operations, instead, it relies on the uniqueness of physical features, e.g., transmission medium, hardware properties, location, or biometrics. This provides natural resistance to quantum threats, however, it makes it vulnerable to attackers which have access to the corresponding PHY feature. For instance, an attacker at very close proximity may successfully impersonate the location of legitimate device.

Given the above, it is clear that integrating PLS into communication systems (e.g., to complement traditional crypto), comes with the promise of a new breed of lightweight, quantum-resilient, low latency, and low-footprint security schemes. However, after decades of research, the deployment of practical PLS solutions is still in its infancy and has met significant resistance. In this article, we first discuss whether a fundamental change in security is actually necessary in order to ensure trustworthy future generations. Then, we discuss some of the key reasons behind the lag between theory and practice in PLS, and propose a roadmap to bridge the gap between theoretical analysis to products. Finally, we present our more general vision for an intelligent, context-aware 6G security, incorporating the physical layer for the first time.

Consequently, the idea of exploiting the properties of the physical layer (PHY) to achieve specific security goals has been extensively researched. Apart from confidentiality using wiretap coding, opportunities for key generation and distribution, user and device authentication and resilience to PHY denial-of-service attacks have been identified.

A mature research direction is that of secret key generation (SKG) from remote observations of a common random source. U. Maurer acknowledged that the propagation channel in wireless systems can be such a random source, allowing devices to distill secret keys and use them for pairing and/or encryption [4]. The corresponding procedures are well studied and concrete countermeasures in the case of active attacks have been proposed [5], [6].

In addition to confidentiality, physical features can also provide a basis for authentication. There are several approaches in this area, including physical unclonable functions (PUFs), localization-based authentication, and RF fingerprinting. The concept of PUFs was first proposed by R. Pappu in his work on physical one-way functions [7]. He developed a challenge-response authentication scheme using a laser beam that passes through a complex 3-D micro-structure, with the challenge being the transmission angle and the response the resulting pattern at the output of the system. Later, the authors of [8] brought this idea to integrated circuits (IC) and coined the term PUF. In that paper, they proposed to exploit the unclonable variability in hardware manufacturing processes. Since then, other physical parameters, such as localization and RF fingerprinting, have also been proposed as soft authentication factors.
The rest of the article is organized as follows. In Section II, we discuss the reasons why PLS is pertinent to 6G and present the current state-of-the-art along with open research issues. Section III presents future perspectives and concludes the article.

PLS—State-of-the-art and Open Issues

As discussed earlier, the theory behind PLS is well established, yet there are still only a few standardized and practical solutions. Our opinion is that this may be attributed to two main reasons: 1) there was no clear need for PLS, and, 2) we lacked some essential tools to bridge the gap between theory and practice.

In further detail, higher layer cryptography solutions have been ensuring security for decades and certainly have their advantages, i.e., resistance after years of scrutiny from the crypto community, along with years of experience in their deployment. Nevertheless, these techniques come with certain drawbacks, e.g., they require infrastructure, they rely on the assumption that the communication connection is already established (note that even in 5G there exist numerous vulnerabilities during the network entry phases), they do not scale easily and, as discussed earlier, depend on the complexity of specific mathematical problems that are thought to be hard.\(^1\) It is clear now with the emergence of new technologies that innovative solutions are necessary.

Second, it is only in the beyond-5G era that we can confidently make the case that certain conditions for the deployment of PLS solutions are met. A few examples:

- **Online learning and engineering of propagation conditions**: Through sensing, intelligent reflecting surfaces and ML algorithms, allows us to identify relevant PLS solutions [11].

- **Massive multiple input multiple output (mMIMO) systems**: Give the opportunity to form pencil sharp beams.\(^2\) This creates a plausible scenario for wiretap coding, as it becomes possible to guarantee an advantage in terms of SNR for certain scenarios, e.g., when considering environments hardly reachable by malicious nodes such as the case of industrial IoT.

  - Localization will be a default service in 6G, offering a lightweight second authentication factor.
  - The evolution in network design provides the required flexibility for widespread deployment of PLS technologies such as PUFs.
  - Different demonstrators have appeared showing that the wireless channel is a viable source of entropy and can be used for key generation.
  - Aspects related to distributed anomaly detection in wireless sensor networks [12] have shown that it is possible to identify compromised sensors by monitoring hardware behavior.

The above showcases that for the first time in 6G the physical world and infrastructure across the board can be used to enhance trustworthiness. Indeed, the advancements in technology continue to reshape our lives and make conventional security approaches insufficient. As discussed earlier, cybersecurity has been focused predominantly on digital defenses. However, with the emergence of new sophisticated security threats, it becomes clear that securing the PHY is equally vital in ensuring trustworthiness.

By incorporating PLS into security mechanisms, it will become possible to protect a wider range of applications and counter a broader spectrum of threats. For example, an insider attack, e.g., an attack launched by a user within the network who has access to usernames and passwords, can be easily countered by including location-based authentication. In this case, the attacker should not only enter the correct credentials, but also be present at a possibly well-secured and camera-monitored area. Similarly, as mentioned in Section I, scenarios where ultra low latency is required are hard to address through traditional cryptographic solutions. To comply with such stringent requirements, a device could benefit from PLS. Particularly, it could use the unique fingerprint of its hardware to perform fast authentication (e.g., using PUF or RF fingerprinting) or/and use the properties of the wireless channel to achieve confidentiality (e.g., through wiretap coding or SKG). Following from above, the need for reshaping the future security landscape is evident. Specifically, it is essential to recognize PLS as an important aspect, fostering a more robust and adaptive security.

A glimpse toward some of the features that PLS can bring into the 6G world is given in Figure 1. The figure illustrates how key physical aspects, e.g., hardware, location, channel variability or guaranteed quality advantage, sensing, etc., can bring forth additional (and important) set of opportunities to deploy PLS. In this section, we delve deeper in PLS and provide an overview of the state-of-the-art. In particular, we

\(^1\)Recently we have witnessed attacks on isogeny-based elliptic cryptography, that was considered among the candidates for postquantum algorithms.

\(^2\)The term pencil sharp beams refers to the fact that mMIMO systems allow to shape the transmission beam and concentrate the energy of the signal in the direction of the intended user. Recalling the example of generalized broadcast wiretap channel from Section I, this would mean that Alice can concentrate the energy of the signal that bears the secret message towards Bob. This would increase the SNR at his end and decrease it at Eve, unless she positions herself on the way between Alice and Bob, where she can also benefit from the increased energy.
explain how current limitations can be overcome to fulfill the need, as well as the promise, for security controls at all layers, including at the physical layer.

**Keyless Transmission of Confidential Messages**

*Theory:* Since Wyner’s pioneering work in [2], the secrecy capacity region has been characterized for different setups [13]. The majority of results, however, rely on the assumption that the transmitter has perfect knowledge of the channel state information (CSI), i.e., it has a precise estimate of the path loss, fading characteristics, noise and interference levels at any given time. However, in real-world wireless systems, due to various factors, such as signal propagation complexities, environmental changes or hardware limitations it becomes challenging to obtain a good estimate of the CSI (far from perfect). Hence, perfect CSI knowledge is an ideal scenario and while theoretical results obtained under such an assumption have given us better intuition, a more practical view is required.

This has been addressed in [14], where a new metric was introduced, namely secrecy degrees of freedom (SDoF). The use of SDoF, led to a significant conclusion: achieving perfect secrecy in the case of imperfect CSI is possible only when asymmetric statistical properties are present for the channels towards both receivers. In this sense, when the channels have symmetrical properties, positive SDoF can be ensured by paying the cost of additional overhead in terms of side information that is used to introduce asymmetry at the encoder [13]. Having this result, it is clear that the quality of the CSI can play a vital role on the achievable secrecy.

In this regard, an important result has been published in [15]. It shows that even an outdated CSI at the transmitter can be used towards increasing the SDoF. The general idea is that, delayed CSI can be successfully incorporated towards interference alignment between users. While these are encouraging findings, further research is still needed to render such secrecy mechanisms possible in a more general context. We note in passing that the idea of artificial noise injection has attracted a lot of attention. However, it seems unlikely that such approaches will be used in practice, at least in the near future, due to strict regulations for the levels of electromagnetic radiation and the need for lowering energy consumption.

Another critical aspect, concerning the practicality of wiretap schemes, is the assumption that the eavesdropper’s CSI is also available at the transmitter. This is highly unlikely in many actual scenarios. To overcome such difficulties, one possible metric is the secrecy outage probability (SOP), which is given by

$$P_{\text{out}}(R) = P(C_S < R)$$  \hspace{1cm} (1)

where $C_S$ denotes the secrecy capacity and $R$ denotes a target secrecy rate. Closely related is the probability of nonzero secrecy capacity, defined as

$$P_{NZ} = P(C_S > 0) = 1 - P_{\text{SOP}}(R = 0).$$  \hspace{1cm} (2)
With the emergence of new communication standards and advanced beamforming techniques, such metrics can play vital role in ensuring security. In detail, focusing a beam toward a legitimate user can give advantage over eavesdroppers in particular scenarios. Depending on the environment and capabilities of devices, it may become possible to evaluate (1) and (2) in an online manner. This makes wiretap codes a viable alternative to traditional mechanisms in cases where the values are high.

Road to practice: While theory has been advancing for decades, the deployment of wiretap codes should proceed with caution. In [16] it was shown that even in THz systems, weak directivity results in large insecure areas, and while such areas can be minimized they cannot be fully eliminated. Therefore, at the moment even with ultramassive MIMO systems and pencil sharp beamforming, it remains an open question how to guarantee zero information leakage without any assumptions regarding the adversarial position, the numbers of antennas, cooperation between distributed adversarial actors, etc. Partially controllable channels, e.g., using intelligent reflective surfaces, could be worth investigated in this aspect to facilitate channel engineering [11].

As discussed earlier, efficient CSI estimation is key for wiretap coding; an example using ray-tracing tools is depicted in Figure 2. The figure shows signal strength evaluation in an indoor premise. Obtaining such a map at the transmitter could be propelled in 6G by online learning together with location-based channel estimation. In a controlled environment (e.g., in geofencing), where potential eavesdroppers are unable to reach all parts of the grid, such a map provides an estimate of the maximum achievable SNR that a malicious node could experience. When this information is available together with the information on the SNR levels at a legitimate node, it becomes trivial to evaluate (1) and (2) at any given time.

Another practical issue, which attracts researchers’ attention, concerns the security guarantees in the finite blocklength. In [17], the achievable secrecy rate was shown to be a function of 1) the blocklength, 2) the error rate and 3) the information leakage, i.e., at finite blocklengths, it is impossible to guarantee zero information leakage. Figure 3 shows the impact of blocklength on the achievable secrecy rate. A comparison is provided between the lower bound on the achievable secrecy rates of Reed–Muller and polar codes for a semideterministic wiretap channel, where the main channel is noiseless and the wiretap channel is a binary-erasure channel with erasure probability $p = 0.4$ and information leakage $\delta = 0.001$, with the second-order approximation secrecy rate [18].

The figure demonstrates that 1) the theoretical limits (secrecy capacity line) might be far from what is achievable in reality, and, 2) the choice of code and blocklength strongly affects the performance. Therefore, making wiretap codes a viable approach for practical systems would require further over-the-air tests and exhaustive performance evaluation in different environments.

Secret Key Generation

Theory: The SKG from channel randomness is a well defined three step procedure [4]. A sketch of the protocol is depicted in Figure 4. In the first phase, referred to as shared randomness distillation, Alice and Bob observe a common random source. Due to channel reciprocity their observations, denoted by $Y_A, Y_B$, respectively, are dependent random variables. An eavesdropper, referred to as Eve, observes $Y_E$, which may be correlated or not with $Y_A$ and $Y_B$. In wireless channels, a readily available source of shared randomness is the multipath fading, which is caused by reflections, diffraction, and scattering from a random environment. In case the same frequency is used, then the equivalent baseband channel between two nodes is reciprocal during the coherence time. These
where $H(K)$ denotes the entropy of the key $K$ and $I(K; V)$ denotes the mutual information between $K$ and $V$.

The first inequality demonstrates that the SKG process can be made error free; (4) ensures that the exchange of side information through public discussion does not leak any information about $K$ to eavesdroppers; while (5) establishes that the generated keys attain maximum entropy (i.e., are uniform). Under the three conditions, an upper bound on the rate for the generation of secret keys is given by [4]

$$\min\{I(Y_A; Y_B), I(Y_A; Y_B|Y_E)\}.$$  \hspace{1cm} (6)

Assuming rich multipath environments, the decorrelation properties of the wireless channel over short distances can be exploited to ensure that Eve’s observation $Y_B$ is unrelated with $Y_A$ and $Y_B$; in this case, the SKG capacity is given by [19, Sec. II] $C_K = I(Y_A; Y_B)$.

Road to practice: Unfortunately, the conditions discussed above are rarely met in real life. In particular, correlations and dependencies in four domains, space, time, frequency, and antenna between Alice’s, Bob’s and the Eve’s observations have to be taken explicitly into account. While subsampling in the time, frequency and antenna domains can constitute simple approaches to recreate a memoryless channel, so that the observations between Alice and Bob are independent from the observations of Eve along these domains, correlations and dependencies in space need on the other hand to be taken explicitly into account. Preprocessing steps to address these issues have recently been reported in [20].

SKG is a mature technique, but one of its major challenges is that the achievable key generation rate depends on the channel statistics. However, upper layers will require a minimum or at least a known rate. Understanding how the achievable rate depends on the channel parameters was the subject of several papers [21], but is still an open issue. Furthermore, parts of the SKG algorithm itself, like the sampling rate in time and frequency, and the CSI quantizer should also be optimized according to the channel properties. It is important to note that availability of reliable CSI could also be an issue. Existing wireless chipsets usually do not provide this information, and, even if they did, they would have to be trusted to provide the correct information. An alternative, using a separate encryption box was proposed in [22].

Finally, active attacks have been addressed in [5] and [6] and hybrid designs of authenticated encryption leveraging SKG along with symmetric block ciphers have appeared in [23]. As a result, SKG emerges as one of the most mature and promising PLS technologies for 6G. Clearly, SKG will be helpful in use cases where key distribution is a major issue, such as massive IoT, addressing scalability in constrained devices that cannot run public key encryption handshakes (or their corresponding postquantum counterparts).

At the end of the SKG process, a common key $K \in \mathcal{K}$ is extracted at Alice and Bob, such that, for any $\epsilon > 0$, the following statements hold [19]:

$$P_t(K = f_A(Y_A, V) = f_B(Y_B, V)) \geq 1 - \epsilon$$  \hspace{1cm} (3)

$$I(K; V) \leq \epsilon$$  \hspace{1cm} (4)

$$H(K) \geq \log |\mathcal{K}| - \epsilon$$  \hspace{1cm} (5)
PUFs and Biometrics

Theory: Some of the most prominent authentication techniques that come from the physical layer are PUFs and biometrics. The idea of PUFs is to authenticate devices using the unique properties of ICs. Such properties appear due to unpredictable variations during their fabrication process. To build a protocol, such variations are typically used in a challenge-response manner. Depending on the PUF architecture, a challenge could refer to measuring gate delays, power-on state or other variable features.

A popular architecture, illustrated in Figure 5, is the arbiter PUF. The scheme is based on the transmission of rising edge signals through two "identical" delay paths, each composed of a series of switching elements. Due to variation properties, the delay required for each signal to pass through the trace will be different. A challenge to this scheme, as illustrated in Figure 5, is a bit sequence that defines the configuration of the switching elements; and a response is a single bit output that defines which signal arrives first at the end. Depending on the number of challenge-response pairs (CRPs) that a PUF can support, architectures are divided into two groups: weak and strong PUFs. The number of CRPs of a weak PUF increases linearly or polynomially with the component blocks (some architectures support only a single CRP) and the number of CRPs of a strong PUF increases exponentially with the component blocks. In this sense, arbiter PUF is considered to be a strong PUF.

Following from the discussion above, biometrics can be seen as a weak PUF structure that measures unique birthmarks of human users (as opposed to devices). Such features include voice, palm vein, iris, behavioral biometrics, and more. Each of these features can produce a CRP for user authentication. In this sense, building a PUF-based or a biometric-based authentication protocol requires identical steps.

Enrollment—This step is carried out offline on a secure channel. During enrollment, a set of responses $R_1, \ldots, R_t \in \mathcal{R}$ (biometric or PUF) are collected by running a set of challenges $C_1, \ldots, C_t \in \mathcal{C}$. Additionally, the measurement noise of the process is characterized in order to generate helper data, $hd$. An authenticator creates a database where CRPs and helper data are associated with a particular user/device.

Authentication—During the online authentication step, the authenticator sends a random challenge $C_i$ from its database to the corresponding user requesting to reproduce the response $R_i$. The user then replies with its PUF or biometric measurement $R'_i$. Due to the presence of noise the newly generated response will differ from the one generated during enrollment, i.e., $R'_i \neq R_i$, therefore the helper data are used in a reconciliation decoder to regenerate $R_i$, in which case authentication is successful. To prevent replay attacks a CRP pair should never be reused, or other measures should be taken, e.g., time stamps. Next, some key issues in the application of such authentication approaches are discussed.

Road to practice: First, a topic that is seeing growing interest is the privacy of biometric data. To perform biometrics-based authentication, the collected measurements are normally passed through third-party authentication servers. This may lead to privacy leakage, i.e., users are clueless about how and where their data are stored or used. Furthermore, as biometrics are permanent features, if adversaries get access to the collected data they could use it to build a human-digital twin. Therefore, it is important that biometric protection techniques are employed. One approach that can be used to avoid storing biometric data is through the use of homomorphic encryption [24]. In such a scheme, performing an operation on the encrypted data is equivalent to performing the same operation on the plain text. Hence, users can provide only encrypted biometric data to authenticate themselves without revealing sensitive content. However, homomorphic encryption requires complex and slow operations, i.e., it is not suitable for constrained devices and low-latency scenarios. In this sense, further research on lightweight and secure biometric protection is required.
RF-based authentication. Before transmission, information passes through different system layers and a number of processing steps. The RF chain is typically the last stage before transmission. After digital-to-analog conversion (DAC) most of the operations become nonlinear and device dependent. Such nonlinearities affect the transmitted signal in a way that is hard to replicate and can uniquely be linked to a particular device. A receiver can extract these features from the arriving signal and confirms the identity of the sender.

Another important topic that has to be addressed concerns the unclonability and randomness of PUFs. First, due to the low number of CRPs supported by weak PUFs, they are susceptible to exhaustive search attacks. Strong PUFs, on the other hand, have large CRP space, which makes exhaustive search attacks impractical. However, the interactive fashion of executing the authentication protocol described above can leak numerous CPRs and in specific cases helper data streams. It has been shown that an attacker can use the leaked information in machine-learning (ML) algorithms to successfully model a PUF [25]. Some of the directions that can help solving this issue are the introduction of more complex structures, e.g., XOR-ing the outputs of multiple PUFs, as opposed to using their individual outputs, can already prevent multiple ML modeling attacks [26]. This gives a shorter but unpredictable sequence. Combining multiple PLS schemes, e.g., PUFs and SKG, can also be used as a preventive measure against ML attacks. PHY generated keys can be used to encrypt and hide the transmission of CRPs and/or helper data, minimizing the leakage to adversaries [27].

Aging is another problem that affects the performance of PUFs. Due to factors such as temperature, humidity, and voltage variations, the physical properties of an electronic circuit can be altered over time. As a result, PUFs become less reliable in generating unclonable, random, and reproducible responses. This creates difficulty in comparing responses obtained at different times, hence, PUF-based authentication protocols that rely on long-term storage become impractical. Possible solutions to overcome aging effects include the use of error correction mechanisms, frequent update on CRP databases [27], and the development of less susceptible to aging PUF architectures.

Finally, it is important to identify appropriate use cases for both, biometrics and PUFs. There are already a variety of commercial products on the market [25], however, as noted above, when either of the techniques is used as a single authentication factor there might be serious concerns. Therefore, the combination of PUF, biometrics, and other authentication factors can be used toward building a secure and reliable multifactor authentication. Scenarios where such approach might be beneficial include eHealth (e.g., for accessing medical records), smart factories (e.g., for access control), and commercial applications (e.g., online banking). What is important to mention is that these schemes are not here to replace authentication handshakes, but to contribute for their efficient and lightweight implementation.

**Location-Based Authentication and RF Fingerprinting**

**Theory:** Apart from PUF authentication, there are other PHY-based authentication techniques, which can be categorized into two types, i.e., RF based and location based.

RF fingerprinting is the process of measuring the unique, stable, and long-term imperfections of analog front-ends in wireless transceivers and wireless communication links [28]. Unlike PUFs, however, there is no guarantee of unclonability. Some of the typically considered imperfections include in-phase quadrature-phase (IQ) imbalances, oscillator drifts, digital-to-analog conversion, power-amplifier nonlinear characteristics, carrier frequency offset, etc. The general idea of RF-based authentication is illustrated in Figure 6, it is also explained briefly below.

An RF-fingerprint-based authentication protocol consists of two phases. First, an offline processing is carried out, where an authenticator captures a set of signals, extracts representative features and creates a classification function that maps features to a particular class, e.g., legitimate and not (optimally, the estimated features would perfectly describe all RF-imperfections of the transmitter). Next, during the online authentication phase, features of the received signals are measured and subsequently passed through the classifier (typically implemented as a hypothesis test).

Location-based authentication relies on relating more specifically a node to a particular location. In detail, an
authenticator should first obtain reliable information concerning the position of other nodes (e.g., a map that contains coordinates of other users). Next, the authentication process is based on online localization of users and comparing their estimated location to earlier stored coordinates. Authentication is successful if the estimated position passes a hypothesis test.

A major advantage of both RF and location-based authentication is that they enhance trust. Naturally, there are still challenges that must be addressed. Some of which are listed below.

Road to practice: A major challenge comes from the increasing complexity in wireless systems and the difficulty in describing the parameters over the chain transmitter, channel, and receiver. The two main approaches currently used are: 1) model-based approach (e.g., using communication theory), where end-to-end communication is modeled as a set of blocks and each block can be parameterized and optimized independently, and, 2) model-free based approach (e.g., using machine learning techniques), where the whole system can be modeled and optimized as a single block [29]. The former approach is typically static, and, hence, performs well in stable environments. However, it could hardly capture the changes in dynamic environments. The latter approach have shown more success in complex environments, but it usually requires great amount of training data and more computational power, hence, it is not well-suited for lightweight devices.

The observations above indicate that a tradeoff must be identified. The complexity of the environment (including number of devices, mobility, etc.) and the approach itself must be taken into account. RF-based authentication would typically require high sensitivity at the receiver (e.g., spectrum analyzer) to identify the unique imperfections of the transmitter. Hence, RF-based authentication might be more suitable for unilateral authentication—access points to identify users. On the other hand, location-based authentication could be well suited for low-end devices, hence, when available could easily provide mutual authentication. Overall, devices should be able to adaptively switch between static and dynamic approaches. To reduce complexity, recent practice leverages an initial approximate model and only uses training data to fine tune the representation.

The accuracy of location and RF information is vital for ensuring trust in these approaches. The accuracy can be affected by variety of factors, including: choice of classification and loss functions, channel quality, choice of metrics, and mobility of users. As discussed earlier both of the authentication approaches rely on prefiled database (e.g., a channel model, a trained neural network or a downloaded map), however, during the authentication phase devices would observe noisy and time-varying features. In this sense, it is important that devices must consider a combination of features, as opposed to a single one. Hence, a promising research direction is online learning of the channel and feature selection aided by dimensionality reduction, to enable real-time analysis of multidimensional data [30].

While there are issues to be addressed both approaches, i.e., RF and location-based authentication can be of great benefit in numerous use cases. One of the main advantages is that, both can provide per packet authentication, hence, apart from authentication, both could contribute to methods such as anomaly detection and trust building. In terms of possible uses cases, some obvious examples are: 1) false base station identification and 2) handover notifications. Access points and base stations are static and as a result, location could be easily introduced as an additional authentication factor to counteract on false base station attacks by using inverse localization (user locating the BS) [27]. On the other hand, BSs could track devices and easily predict the time when a device will leave a cell and enter a neighboring one. This information could be transferred between BSs to allow speeding up authentication for the device handover [28].

Another possible approach for identifying adversarial users could be through the uniqueness of antenna arrays in mMIMO communication, e.g., the beam patterns from different devices will differ even if they are colocated. This can be used by authenticators to identify users in close proximity. In fact, it has already been shown that the sweeping beam patterns could be used as unique and reliable source to counter spoofing attacks [31].

The discussion above gives some initial ideas on how RF and location information could contribute to the system's security. However, there are still open issues that need to be addressed before their full integration into the standards. Depending on the application, different problems may arise. For example, if considering a human device the authentication would typically be end-to-end; if considering fully autonomous system authentication would be device-to-device [28]. An important research topic, for both cases, is the development of cross-layer security protocols. In particular, how should upper layers access, process, and use PHY information. Answering this question could pave the way for a new lightweight and cross-layer security solutions.

General Overview

With the advent of 6G, we are entering a new era of massive connectivity of autonomous cyberphysical agents, equipped with enhanced sensing, processing, and learning capabilities. In the past, static security solutions were introduced as add-ons to earlier network design choices. A break from this paradigm is needed as static security
solutions cannot scale efficiently while meeting latency, computation, and power constraints.

In this framework, exploiting the characteristics of physical phenomena to provide security and ensure privacy becomes pertinent. As discussed earlier, there are number of open issues on the road toward practical PLS. However, it is clear that security at the PHY can both complement conventional upper-layer security schemes and strengthen the overall trust and resilience of 6G. Different security solutions are attainable by exploiting novel opportunities such as in sub-GHz to THz frequency bands, intelligent reflective surfaces, joint communications and sensing, localization and RF fingerprinting.

However, the novel opportunities contribute also to the increase of the attack surface. Therefore, one of the utmost important issues to be solved before adopting PLS approaches is related to the threat model. While, all cryptographic based authentication protocols have a unified threat model, i.e., the well-known Dolev–Yao model,\(^3\) PHY-based techniques rely on different assumptions for the adversary.

For example, location-based authentication might falsely identify an attacker as a legitimate user if the former is in the close vicinity to the latter. Similarly, an attacker positioned at a favorable location can break the confidentiality of SKG or wiretap codes. In this sense, studies should not focus only on a single attack, e.g., spoofing (multiple devices same ID), Sybil attacks (1 device multiple IDs), jamming or injection attacks, but propose a unified model that captures all threats present in wireless communications. This will unlock the full potential of PLS, as defences will involve multiple PHY features working simultaneously toward ensuring trustworthiness.

Following from the above, sensor fusion can become a vital aspect toward enabling practical PLS solutions. By combining hardware and software capabilities, i.e., by integrating data of multiple modalities and processing it in real time, multimodal fusion can provide more reliable and accurate representation of the PHY.

In particular, combining information from multiple antennas can improve channel state estimation and can potentially give an estimate of direction and distance toward jamming devices (which in turn can help in mitigating such attacks). Other sensors can provide a “sense” of mobility, e.g., through GPS coordinates or speed measures. Cameras can be used to increase the vision by observing whether LoS or NLoS condition exists between two points and/or monitoring for eavesdropping attempts in the vicinity. From this perspective, sensor fusion can minimize the attack surface and be an important stepstone toward practical PLS implementations.

With this in mind, Figure 7 showcases our vision toward a context-aware PLS. As discussed throughout this article, we do not believe that there exists a single PLS scheme that can be used in all possible scenarios. Instead, we think that a context-driven approach, which takes different PHY aspects into account.

\(^3\)A Dolev-Yao type of adversary 1) has control over the legitimate channel and can send any type of queries using knowledge that has been gained through observation of previous protocol executions; 2) all functions and operations used for authentication between the legitimate users are assumed public; 3) the adversary can perform denial of service (DoS) attacks to block parts of the authentication procedure and desynchronize the legitimate connection.

Figure 7
Context-aware PLS. PHY features like channel quality, mobility, presence of LoS/NLoS, can be used to improve trust.
account, should be utilized. Depending on the available contextual information, PLS schemes could be used as lightweight security solutions, toward ensuring trust.

It is clear that PLS is a set of useful tools, which can greatly contribute toward the security of future networks. This article has presented our vision and some concrete examples on what PLS can do for the future generation of wireless networks. However, while there is a vast theory behind all PLS schemes, a generalized practical perspective is still missing. Along with all pros behind PLS, we have also highlighted some of the major gaps in the area and hope that this will stimulate further research.

Acknowledgment
This work was supported by the Saxon State government out of the State budget approved by the Saxon State Parliament.

References
[1] C. E. Shannon, “Communication theory of secrecy systems,” Bell Syst. Tech. J., vol. 28, no. 4, pp. 656–715, 1949.
[2] A. D. Wyner, “The wire-tap channel,” Bell Syst. Tech. J., vol. 54, no. 8, pp. 1355–1387, 1975.
[3] I. Csiszar and J. Korner, “Broadcast channels with confidential messages,” IEEE Trans. Inf. Theory, vol. 24, no. 3, pp. 339–348, May 1978.
[4] U. M. Maurer, “Secret key agreement by public discussion from common information,” IEEE Trans. Inf. Theory, vol. 39, no. 3, pp. 733–742, May 1993.
[5] E. V. Belmega and A. Chorti, “Protecting secret key generation systems against jamming: Energy harvesting and channel hopping approaches,” IEEE Trans. Inf. Forensics Secur., vol. 12, no. 11, pp. 2611–2626, Nov. 2017.
[6] M. Mitev, A. Chorti, E. V. Belmega, and H. V. Poor, “Protecting physical layer secret key generation from active attacks,” Entropy, vol. 23, 2021, Art. no. 960.
[7] R. Pappu, B. Recht, J. Taylor, and N. Gershenfeld, “Physical one-way functions,” Science, vol. 297, no. 5589, pp. 2026–2030, 2002.
[8] B. Gassend, D. Clarke, M. van Dijk, and S. Devadas, “Silicon physical random functions,” in Proc. 9th ACM Conf. Comput. Commun. Secur., 2002, pp. 148–160.
[9] P. Bachan and B. Singh, “Performance evaluation of authentication protocols for IEEE 802.11 standard,” in Proc. Int. Conf. Comput. Commun. Technol., 2010, pp. 792–799.
[10] R. Maes, A. Van Herreweghe, and I. Verbauwhede, “PUFKY: A fully functional PUF-based cryptographic key generator,” in Cryptographic Hardware and Embedded Systems–CHES. Berlin, Germany: Springer, 2012, pp. 302–319.
[11] S. Tomasin, H. Zhang, A. Chorti, and H. V. Poor, “Challenge-response physical layer authentication over partially controllable channels,” IEEE Commun. Mag., vol. 60, no. 12, pp. 138–144, Dec. 2022.
[12] G. A. N. Segura, A. Chorti, and C. B. Margi, “Centralized and distributed intrusion detection for resource-constrained wireless SDN networks,” IEEE Internet Things J., vol. 9, no. 10, pp. 7746–7758, May 2022.
[13] S. Yang, M. Kobayashi, P. Plantanidia, and S. Shamai, “Secrecy degrees of freedom of MIMO broadcast channels with delayed CSIT,” IEEE Trans. Inf. Theory, vol. 59, no. 9, pp. 5244–5256, Sep. 2013.
[14] Y. Liang, G. Kramer, H. V. Poor, and S. Shamai, “Compound wiretap channels,” EURASIP J. Wireless Commun. Netw., vol. 2009, 2009, Art. no. 142374.
[15] M. A. Maddah-Ali and D. Tse, “On the degrees of freedom of MISO broadcast channels with delayed feedback,” Dept. Elect. Eng. Comput. Sci., Univ. California, Berkeley, CA, USA, Tech. Rep. UCB-EECS-2010-122, Sep. 2010.
[16] R. Schulz et al., “Semantic security for indoor THz-wireless communication,” in Proc. 17th Int. Symp. Wireless Commun. Syst., 2021, pp. 1–6.
[17] W. Yang, R. F. Schaefer, and H. V. Poor, “Secrecy-reliability tradeoff for semi-deterministic wiretap channels at finite blocklength,” in Proc. IEEE Int. Symp. Inf. Theory, 2017, pp. 2133–2137.
[18] M. Shakiba-Herfah, L.uzzi, and A. Chorti, “Finite blocklength secrecy analysis of polar and Reed–Muller codes in BEC semi-deterministic wiretap channels,” in Proc. IEEE Inf. Theory Workshop, 2021, pp. 1–6.
[19] R. Ahlswede and I. Csiszar, “Common randomness in information theory and cryptography. I. Secret sharing,” IEEE Trans. Inf. Theory, vol. 39, no. 4, pp. 1121–1132, Jul. 1993.
[20] M. Srinivasan, S. Skaperas, M. S. Herfeh, and A. Chorti, “Joint localization-based node authentication and secret key generation,” in Proc. IEEE Int. Conf. Commun., 2022, pp. 32–37.
[21] M. Mitev, A. N. Barreto, T. M. Pham, and G. Fettweis, “Secret key generation rates over frequency selective channels,” in Proc. IEEE Veh. Technol. Conf., 2022, pp. 1–5.
[22] M. Zoli, A. N. Barreto, S. Köpsell, P. Sen, and G. Fettweis, “Physical-layer-security box: A concept for time-frequency channel-reciprocity key generation,” EURASIP J. Wireless Commun. Netw., vol. 2020, 2020, Art. no. 114.
[23] M. Mitev, A. Chorti, M. Reed, and L. Musavian, “Authenticated secret key generation in delay-constrained wireless systems,” EURASIP J. Wireless Commun. Netw., vol. 2020, 2020, Art. no. 122.
[24] M. Morampudi, M. Prasad, and U. Raju, “Privacy-preserving Iris authentication using fully homomorphic encryption,” Multimedia Tools Appl., vol. 79, pp. 19215–19237, 2020.
[25] P. Gope and B. Sikdar, “A comparative study of design paradigms for PUF-based security protocols for IoT devices: Current progress, challenges, and future expectation,” *Computer*, vol. 54, no. 11, pp. 36–46, 2021.

[26] U. Rührmair, F. Sehnke, J. Sölter, G. Dror, S. Devadas, and J. Schmidhuber, “Modeling attacks on physical unclonable functions,” in *Proc. 17th ACM Conf. Comput. Commun. Secur.*, New York, NY, USA, 2010, pp. 237–249.

[27] M. Mitev, M. Shakiba-Herfeh, A. Chorti, M. Reed, and S. Baghaee, “A physical layer, zero-round-trip-time, multifactor authentication protocol,” *IEEE Access*, vol. 10, pp. 74555–74571, 2022.

[28] X. Wang, P. Hao, and L. Hanzo, “Physical-layer authentication for wireless security enhancement: Current challenges and future developments,” *IEEE Commun. Mag.*, vol. 54, no. 6, pp. 152–158, Jun. 2016.

[29] Z. Sun, H. Wu, C. Zhao, and G. Yue, “End-to-end learning of secure wireless communications: Confidential transmission and authentication,” *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 88–95, Oct. 2020.

[30] L. Senigagliesi, M. Baldi, and E. Gambi, “Comparison of statistical and machine learning techniques for physical layer authentication,” *IEEE Trans. Inf. Forensics Secur.*, vol. 16, pp. 1506–1521, 2021.

[31] N. Wang, L. Jiao, P. Wang, W. Li, and K. Zeng, “Machine learning-based spoofing attack detection in mmWave 60 GHz IEEE 802.11ad networks,” in *Proc. IEEE Conf. Comput. Commun.*, 2020, pp. 2579–2588.

Thuy M. Pham (Member, IEEE) received the Ph.D. degree in wireless communications from Maynooth University, Maynooth, Ireland, in 2020.

In 2018, he worked for Airrays GmbH, Germany, as a part-time Research Engineer on an LTE Project. He is with Barkhausen Institut, Germany, since 2020. His research interests include ad hoc wireless routing protocols, wireless communications, and physical layer security.

Arsenia Chorti (Senior Member, IEEE) received the Ph.D. degree in communications and signal processing from Imperial College London, U.K., in 2005.

She is a Professor with the École Nationale Supérieure de l’Électronique et de ses Applications (ENSEA), the Joint Head of the Information, Communications and Imaging (ICI) Group of the ETIS Lab UMR 8051, a Leader of the Wireless Connectivity group of the Barkhausen Institut gGmbH, and a Visiting Scholar at Princeton University. Her research spans the areas of wireless communications and wireless systems security for 5G and 6G, with a particular focus on physical layer security. Her current research topics include context aware security, multifactor authentication protocols, 5G/6G and IoT, anomaly detection, machine learning for communications, new multiple access techniques and scheduling.

Dr. Chorti is a member of IEEE INGR on Security and is the Chair of IEEE Focus Group on Physical Layer Security. Recently, she has participated in the reduction of the ITU report M.2516-0 on Future Technology Trends of Terrestrial International Mobile Telecommunications Systems Towards 2030 and Beyond (sections on trustworthiness).
André Noll Barreto (Senior Member, IEEE) received the M.Sc. degree from the Catholic University (PUC-Rio), Rio de Janeiro, Brazil, in 1996, and the Ph.D. degree from the Technische Universität Dresden, Dresden, Germany, in 2001, both in electrical engineering. He held several positions with academia and industry in Switzerland (IBM Research) and Brazil (Claro, Nokia Technology Institute/INDT, Universidade de Brasília, and Ektrum). From 2018 to 2022, he worked with the Barkhausen Institut, researching wireless communications for the reliable, resilient, and secure Internet of Things. He was the Chair of the Centro Norte Brasil Section of IEEE in 2013 and 2014, and the General Co-Chair of the Brazilian Telecommunications Symposium in 2012. Since September 2022, he has been with Nokia Standards in Paris, working on the evolution towards 6G.

Gerhard Fettweis (Fellow, IEEE) received the Ph.D. degree in electrical engineering under H. Meyr at RWTH Aachen, Aachen, Germany, in 1990. After a postdoctoral at IBM Research, San Jose, he joined TCSI, Berkeley, USA. Since 1994, he has been the Vodafone Chair Professor with TU Dresden, Germany. Since 2018, he has been the founding Director/CEO of the Barkhausen Institute. His research works include wireless transmission and chip design, coordinates 5G++Lab Germany, spun-out 17 tech and 3 nontech startups. He is a member of the German Academy of Sciences (Leopoldina), and German Academy of Engineering (acatech).