




Article

# Securing the Insecure: A First-Line-of-Defense for Body-Centric Nanoscale Communication Systems Operating in THz Band

Waqas Aman <sup>1,2</sup>, Muhammad Mahboob Ur Rahman <sup>1</sup>, Hasan T. Abbas <sup>2</sup>,  
Muhammad Arslan Khalid <sup>3</sup>, Muhammad A. Imran <sup>2,4</sup>, Akram Alomainy <sup>5</sup> and Qammer H. Abbasi <sup>2,\*</sup>

<sup>1</sup> Electrical Engineering Department, Information Technology University, Lahore 54000, Pakistan; waqas.aman@itu.edu.pk (W.A.); mahboob.rahman@itu.edu.pk (M.M.U.R.)

<sup>2</sup> Department of Electronics and Nano Engineering, University of Glasgow, Glasgow G12 8QQ, UK; Hasan.Abbas@glasgow.ac.uk (H.T.A.); Muhammad.Imran@glasgow.ac.uk (M.A.I.)

<sup>3</sup> Division of Biomedical Engineering, School of Engineering, University of Glasgow, Glasgow G12 8QQ, UK; arslan.k@live.com

<sup>4</sup> Artificial Intelligence Research Center (AIRC), Ajman University, Ajman 00000, United Arab Emirates

<sup>5</sup> School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, UK; a.alomainy@qmul.ac.uk

\* Correspondence: qammer.abbasi@glasgow.ac.uk

**Abstract:** This manuscript presents a novel mechanism (at the physical layer) for authentication and transmitter identification in a body-centric nanoscale communication system operating in the terahertz (THz) band. The unique characteristics of the propagation medium in the THz band renders the existing techniques (say for impersonation detection in cellular networks) not applicable. In this work, we considered a body-centric network with multiple on-body nano-sensor nodes (of which some nano-sensors have been compromised) who communicate their sensed data to a nearby gateway node. We proposed to protect the transmissions on the link between the legitimate nano-sensor nodes and the gateway by exploiting the path loss of the THz propagation medium as the fingerprint/feature of the sender node to carry out authentication at the gateway. Specifically, we proposed a two-step hypothesis testing mechanism at the gateway to counter the impersonation (false data injection) attacks by malicious nano-sensors. To this end, we computed the path loss of the THz link under consideration using the high-resolution transmission molecular absorption (HITRAN) database. Furthermore, to refine the outcome of the two-step hypothesis testing device, we modeled the impersonation attack detection problem as a hidden Markov model (HMM), which was then solved by the classical Viterbi algorithm. As a by-product of the authentication problem, we performed transmitter identification (when the two-step hypothesis testing device decides no impersonation) using (i) the maximum likelihood (ML) method and (ii) the Gaussian mixture model (GMM), whose parameters are learned via the expectation–maximization algorithm. Our simulation results showed that the two error probabilities (missed detection and false alarm) were decreasing functions of the signal-to-noise ratio (SNR). Specifically, at an SNR of 10 dB with a pre-specified false alarm rate of 0.2, the probability of correct detection was almost one. We further noticed that the HMM method outperformed the two-step hypothesis testing method at low SNRs (e.g., a 10% increase in accuracy was recorded at SNR = −5 dB), as expected. Finally, it was observed that the GMM method was useful when the ground truths (the true path loss values for all the legitimate THz links) were noisy.

**Keywords:** body-centric sensor networks; nanoscale communication; terahertz communication; nano sensors; security; authentication; outlier detection; sensor networks; healthcare systems



**Citation:** Aman, W.; Rahman, M.M.U.; Abbas, H.T.; Khalid, M.A.; Imran, A.; Alomainy, A.; Abbasi, Q.H. Securing the Insecure: A First-Line-of-Defense for Body-Centric Nanoscale Communication Systems Operating in THz Band. *Sensors* **2021**, *21*, 3534. <https://doi.org/10.3390/s21103534>

Academic Editor: Anna Maria Vegni

Received: 21 April 2021

Accepted: 17 May 2021

Published: 19 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

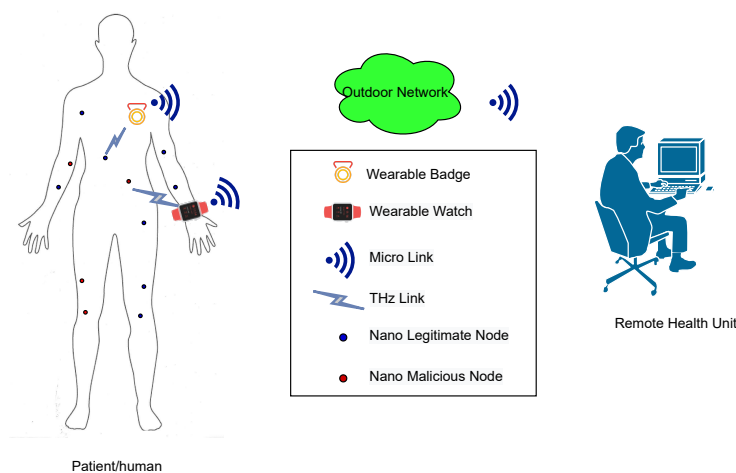
## 1. Introduction

Nanoscale communication systems have attracted researchers due to their promising applications in healthcare, manufacturing industries, environmental control, etc. [1]. On

the other hand, body-centric communication has potential applications in healthcare, entertainment, etc. [2]. Generally, body-centric communication is classified as “off”-, “on”-, and “in”-body communication based on the communication among implanted or wearable electronic devices. In this work, we focused on the body-centric communication systems where nano sensors/devices operating in the THz band are deployed on the body of a human being.

Due to the small size of nano devices, the existing frameworks, techniques, and methods proposed for communication networks such as WiFi, 4G, etc., are not suitable for exchanging information amongst the nano devices [3]. For instance, nano devices are unable to operate at microwave bands due to their small size. They would require molecular communication and the terahertz (THz) band for operation. Additionally, in IoT devices, due to the small energy sources, the computational processing capability is limited. Therefore, it is necessary to meet the requirements for new protocols of nano devices at all layers of the protocol stack. Operating in the THz band (0.1–10 THz) is a promising solution at the physical layer (PL) [4], which makes the antenna size very small and thus suitable for exchanging information between nano devices.

Like other communication networks, the body-centric nanoscale communication networks are also prone to a wide range of active and passive attacks by adversaries [5]. Some of the common attacks include eavesdropping, impersonation, denial of service (DoS), etc. Here, we investigated an impersonation attack in body-centric nanoscale communication networks. Figure 1 shows an illustration of an impersonation attack on a smart healthcare system scenario. The nano nodes are deployed on the body of a person/patient for disease diagnostics or to remotely monitor his/her health parameters. These nano devices are connected to a wearable device, which communicates the data to an outdoor network via a nano-to-micro interface. Assuming an enemy of the person secretly deployed its own nano machines nearby with the aim of impersonating the person’s legitimate nodes to report false measurements to the remote health unit, an incorrect response through the nano machines or nearby doctors could result in devastating consequences. Therefore, we need an authentication mechanism at the nano-to-micro interface device (wearable device) to allow data transmission (reported measurements, i.e., glucose, blood pressure, etc., of nano nodes/sensors) from legitimate nano nodes only, blocking all malicious nodes.



**Figure 1.** An envisioned future body-centric nanoscale healthcare system with possible malicious nodes.

In traditional communication systems, the countermeasures for such attacks are performed at the higher layer using cryptography. Despite the wide work in the field of cryptography, the mechanism can be compromised because of its sole dependency on the predefined shared secret among the legitimate users. With recent advances in quantum computing, traditional encryption has become vulnerable to being easily decoded, and

existing crypto-based measures are not quantum secure unless the size of secret keys increases to impractical lengths [6]. In this regard, physical layer (PL) security finds itself as a promising mechanism in future communication systems. PL security exploits the random nature of the physical medium/layer for security purposes [7].

Authentication is one of the pillars required for the security of any communication system. PL authentication is a systematic procedure that uses PL's features to provide authentication. In conventional systems, asymmetric key encryption (AKE) is typically used in the authentication phase, which is the realm of public key encryption (a crypto-based approach). Such schemes are quantum insecure and incur overhead or high computations, which not only increase the size of the device, but also consume much power. The devices fabricated for nanoscale communication are energy constrained as they incorporate a small source of energy (a battery). PL authentication has a low overhead (a simple procedure that typically includes feature estimation and testing) and is almost impossible to clone unless the devices lie on each other. Various fingerprints including RSS [8], CIR [9,10], CFR [11,12], carrier frequency offset [13], and I/Q imbalance [14] have been reported for PL authentication in conventional communication systems.

**Related Work:** The authors in [15] for the first time studied authentication using path loss (S21 parameter) in body-centric communication using millimeter waves. Regarding the security of systems operating in the THz band, we found some works [5,16–18] in the literature. The work [16] provided the first study on the security challenges faced by nanoscale communication systems, while the work [17] presented some possible promising applications along with the security challenges in the Internet of Nano-Things. Further, the experimental work of Jianjun et al. [5] for the first time rejected the claim about security in the THz band. The claim was that the inherit narrow beamwidth of the THz link makes it secure and thus impossible for a malicious node to accomplish an eavesdropping attack. The authors in [5] in their experiments used reflectors of different shapes between the THz transmitter and receiver. Then, with the help of secrecy capacity and blockage as performance metrics, they clearly demonstrated that eavesdropping attacks in the THz band can be easily performed.

The differences between our work and previous work are as follows: The first work [15], which studied the authentication problem in body-centric communication systems, considered millimeter-wave communication with a three-node setup. In contrast, our work considered multiple legitimate and malicious nodes operating in the THz band. The work [5] considered an eavesdropping attack in a system operating in the THz band, which was a different problem/attack than the attack we considered in our work. Next, in our previous work [18], we studied PL authentication for an in vivo nanoscale communication system whereby we utilized the path loss as the device fingerprint for a three-node system (i.e., Alice, Eve, and Bob). The difference between our previous work [18] and this work was twofold. First, the previous work was limited to the three-node system only, while in this work, the system model was comprised of multiple legitimate and malicious nodes. Second, the previous work was for an in vivo nanoscale communication system where authentication occurs at a nano node (Bob).

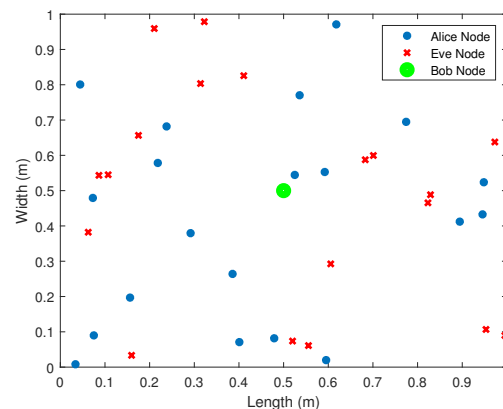
**Contributions:** For the first time, this work studied authentication at a nano-to-micro interface device (wearable device) in an on-body-centric communication system where we exploited the high-resolution transmission molecular absorption (HITRAN) database [19] for computing the path loss. For the first time, impersonation attack detection at the wearable device/receiver/Bob in multiple legitimate and malicious nano nodes operating in the THz band is performed via different mechanisms. We performed authentication by two-step hypothesis testing. We refined the output of the hypothesis testing via the hidden Markov model (HMM) with the Viterbi algorithm. We also performed transmitter identification via the maximum likelihood and Gaussian mixture model (GMM) with the expectation–maximization algorithm.

**Outline:** The rest of this paper is organized as follows. Section 2 provides the system model. Section 3 discusses authentication via two-step hypothesis testing. Section 4

presents the hidden Markov model to refine the output of hypothesis testing. Section 5 provides transmitter identification schemes. Section 6 presents simulation results with discussions, and Section 7 concludes the paper.

## 2. System Model

For the purposes of the simulation, we considered a square 2D map/layout of size  $(1 \text{ m} \times 1 \text{ m})$  where  $M + N$  nano transmission (Tx) nodes,  $M$  Alice (legitimate) nodes  $\{A_i\}_{i=1}^M$ , and  $N$  Eve (malicious) nodes  $\{E_j\}_{j=1}^N$  are deployed according to the uniform distribution model, whilst a nano-to-micro interface device/receiver node, Bob, is placed at the origin, as shown in Figure 2. We assumed that the Tx nodes transmitted with a fixed/pre-specified transmit power so that the path loss can be computed by Bob.



**Figure 2.** System model for the simulation purposes: Bob is placed at the origin. Alice's and Eve's node locations are modeled as uniformly distributed random variables. In this case,  $M = 10$  and  $N = 10$ .

The path loss is given as [20,21]:

$$L(f, d)[dB] = L_a(f, d)[dB] + L_s(f, d)[dB], \quad (1)$$

where  $f$  is the frequency,  $d$  is the distance,  $L_a(f, d)[dB]$  is the absorption loss, and  $L_s(f, d)[dB]$  is the spreading loss. More details of spreading and absorption losses are given in Appendix A.

In the next section, we discuss the two-step mechanism for impersonation detection.

## 3. Authentication via Two-Step Hypothesis Testing

We assumed that the shared channel is time-slotted, whilst the transmit nodes perform channel sensing before transmitting; hence, there are no collisions. Without loss of generality, it can be assumed that  $A_i$  is the legitimate node for slot  $k$ , but if  $A_i$  does not transmit during this time slot,  $E_j$  could transmit to Bob pretending to be an Alice node. Therefore, Bob needs to authenticate each message received on the shared channel and verify the transmitter identity (if no impersonation has been declared) in a systematic manner.

Assume that the noisy measurement  $z(k) = L + n(k)$  has been obtained at time  $k$  (for instance, by using the pulse-based method as discussed in [22]), where  $n(k) \sim N(0, \sigma^2)$  and  $L$  is the path loss. Furthermore, in line with previous studies [18,23], we assumed that Bob has already learned the ground truth via prior training on a secure channel. The ground truth vector can be denoted by  $\mathbf{l} = \{L_1, \dots, L_M\}^T$ . The two-step hypothesis testing or maximum likelihood (ML) hypothesis test can be explained by the following equations:

$$(T^*, i^*) = \min_i |z - L_i|. \quad (2)$$

Next, the binary hypothesis test works as follows:

$$\begin{cases} H_0(\text{no impersonation}) : & T^* = \min_i |z(k) - L_i| < \epsilon \\ H_1(\text{impersonation}) : & T^* = \min_i |z(k) - L_i| > \epsilon \end{cases} \quad (3)$$

Equivalently, we have:

$$T^* \underset{H_0}{\geq} \underset{H_1}{\epsilon}, \quad (4)$$

where  $\epsilon$  is a small threshold—a design parameter. This work followed the Neyman–Pearson theorem [24], which states that, for a pre-specified  $P_{fa}$ ,  $\epsilon$  can be chosen such that  $P_{md}$  is minimized.

The error probabilities for the above hypothesis tests are:

$$\begin{aligned} P_{fa} &= P(H_1|H_0) = \sum_{i=1}^M P(T^* > \epsilon|A_i)\pi(i) \\ &= \sum_{i=1}^M 2Q\left(\frac{\epsilon}{\sigma}\right)\pi(i) = 2Q\left(\frac{\epsilon}{\sigma}\right) \sum_{i=1}^M \pi(i) = 2Q\left(\frac{\epsilon}{\sigma}\right), \end{aligned} \quad (5)$$

where  $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}} dt$  is the complementary cumulative distribution function (ccdf) of a standard normal distribution, and  $\pi(i)$  is the prior probability of  $A_i$ . Thus, the threshold could be computed as follows:

$$\epsilon = \sigma Q^{-1}\left(\frac{P_{fa}}{2}\right). \quad (6)$$

Then,  $P_{md}$  is given as:

$$\begin{aligned} P_{md} &= P(H_0|H_1) = P(T^* < \epsilon|H_1) \\ &= \sum_{j=1}^N \sum_{i=1}^M \left[ Q\left(\frac{L_i - L_j - \epsilon}{\sigma}\right) - Q\left(\frac{L_i - L_j + \epsilon}{\sigma}\right) \right] \pi(j), \end{aligned} \quad (7)$$

where  $\pi(j) = \sum_{i=1}^M \alpha_{ij}\pi(i)$  is the prior probability of  $E_j$ .  $0 < \alpha_{ij} < 1$  is the fraction of slots that were originally dedicated to  $A_i$ , but were found idle and thus utilized by  $E_j$ .

Since  $P_{md}$  is an R.V., the expected value  $\bar{P}_{md} := \mathbb{E}(P_{md})$  is as follows:

$$\begin{aligned} \bar{P}_{md} &= \sum_{j=1}^N \frac{1}{\Delta} \pi(j) \\ &\left( \int_{L_{min}}^{L_{max}} \sum_{i=1}^M Q\left(\frac{L_i - L_j^{(E)} - \epsilon}{\sigma}\right) - Q\left(\frac{L_i - L_j^{(E)} + \epsilon}{\sigma}\right) dL_j^{(E)} \right) \\ &= \sum_{j=1}^N \frac{1}{\Delta} \pi(j). \\ &\left( \int_{L_{min}}^{L_{max}} \sum_{i=1}^M Q\left(\frac{L_i - L^{(E)} - \epsilon}{\sigma}\right) - Q\left(\frac{L_i - L^{(E)} + \epsilon}{\sigma}\right) dL^{(E)} \right), \end{aligned} \quad (8)$$

where we assumed that the unknown path loss  $L_j \sim U(L_{min}, L_{max}) \forall j$  and  $\Delta = L_{max} - L_{min}$ .

Next, we discuss the HMM for refining the outcomes/results of the two-step hypothesis testing.

#### 4. Hidden Markov Model-Based Approach

To refine the output of the two-step hypothesis testing, we used the HMM-based approach. More specifically, at a given time instant  $k$ , the system is in one of the two states with the state-space:  $\mathcal{S} = \{s_0, s_1\}$ . The states  $s_0$  and  $s_1$  imply that there is no impersonation, impersonation respectively, at time  $k$ . However, the true state of the system is hidden; therefore, what we observe through the hypothesis test is another observable Markov chain. The connection between the true/hidden state and the observable state is given by the emission probability matrix:

$$\mathbf{R} = \begin{bmatrix} r_{0,0} & r_{0,1} \\ r_{1,0} & r_{1,1} \end{bmatrix}, \quad (9)$$

where  $r_{i,j} = Pr(x[k] = i | s[k] = j)$ ,  $i, j \in \{0, 1\}$ . The off-diagonal elements in the  $i$ -th row of  $\mathbf{R}$  represents the errors made by the ML test, i.e., deciding the state as  $s[k] = j$ ,  $j \in \{0, 1\} \setminus i$  while the system was actually in state  $s[k] = i$ .

The transition from state  $i$  to state  $j$  occurs after a fixed interval of  $T = t_k - t_{k-1}$  seconds where  $1/T$  is the measurement rate. Assume that the system was in state  $s_0$  at time  $k = 0$ , i.e.,  $\mathbf{x}[0] = [1, 0]^T$ , we are in time  $k - 1$  and want to predict the probability vector  $\mathbf{x}[k]$  at time  $k$ , and the system is in state  $s_i$ ,  $i \in \{0, 1\}$ . To this end, we have the following transition probability matrix:

$$\mathbf{P} = \begin{bmatrix} p_{0,0} & p_{0,1} \\ p_{1,0} & p_{1,1} \end{bmatrix}, \quad (10)$$

where  $p_{i,j} = P(x[k] = j | x[k-1] = i)$ ,  $i, j \in \{0, 1\}$ . Then, we have the following relation:  $\mathbf{x}[k] = \mathbf{P}^k \mathbf{x}[0]$ . Alternatively, we can write:  $\mathbf{x}[k] = \mathbf{P} \mathbf{x}[k-1]$ .

##### ML Estimation of a Hidden Markov Sequence Using the Viterbi Algorithm

The Viterbi algorithm is used for the ML sequence estimation (MLSE) of  $\{s[k]\}_{k=1}^K$ , given  $\{x[k]\}_{k=1}^K$  as:

$$\{s[k]\} = \arg \max_{\{s'[k]\}} p(x[k] | s'[k]). \quad (11)$$

At this stage, we are done with impersonation detection mechanisms. Next, we discuss the transmitter identification mechanisms.

#### 5. Transmitter Identification

The transmitter identification is accomplished via two approaches: ML- and GMM-based transmitter identification.

##### 5.1. ML-Based Approach

In the ML-based approach, the probability of the misclassification error resulting from Equation (2) is given as:

$$P_{mc} = \sum_{i=1}^M P_{mc|i} \pi(i), \quad (12)$$

where  $P_{mc|i} = P(\text{Bob decides } A_j | A_i \text{ was the sender})$ . For the hypothesis test of (4),  $P_{mc|i}$  is given as:

$$P_{mc|i} = 1 - \left( Q\left(\frac{\tilde{L}_{l,i} - \tilde{L}_i}{\sigma}\right) - Q\left(\frac{\tilde{L}_{u,i} - \tilde{L}_i}{\sigma}\right) \right), \quad (13)$$

where  $\tilde{L}_{l,i} = \frac{\tilde{L}_{i-1} + \tilde{L}_i}{2}$ ,  $\tilde{L}_{u,i} = \frac{\tilde{L}_i + \tilde{L}_{i+1}}{2}$ . Additionally,  $\tilde{\mathbf{I}} = \{\tilde{L}_1, \dots, \tilde{L}_M\} = \text{sort}(\mathbf{I})$  where the sort operation ( $\cdot$ ) sorts a vector in increasing order. For the boundary cases, e.g.,  $i = 1, i = M$ ,  $\tilde{L}_{l,1} = L_{min}$ ,  $\tilde{L}_{l,M} = L_{max}$ , respectively.

##### 5.2. Transmitter Identification Using Gaussian Mixture Modeling

The GMM consisted of  $Q = M + N$  component densities where only the  $Q = M$  densities could be trained. The 3Q GMM parameter was learned by running the expectation-

maximization (EM) algorithm on the training data. The GMM, in its standard form, is perfectly suited for transmitter identification. Under the GMM, the probability density function (pdf) of the (observed) mixture random variable  $X$  is the convex/weighted sum of the component pdfs:

$$f_X(x) = \sum_{q=1}^Q \pi_q \phi_q(x), \quad (14)$$

where each  $\phi_q(x)$  is a Gaussian pdf that satisfies:  $\phi_q(x) \geq 0$ ,  $\int_{x \in \mathbb{R}} \phi_q(x) dx = 1$ . The weights/priors satisfy:  $\pi_q(x) \geq 0$ ,  $\sum_{q=1}^Q \pi_q = 1$ .

The GMM has  $3Q$  unknown parameters, which were learned by applying the iterative expectation–maximization algorithm on the training data  $\{x_m\}_{m=1}^M$ . The posterior probability for each point  $x_m$  in the training data (i.e., the likelihood of  $x_m$  belonging to component  $q$  of the mixture) was computed as follows ( $j$  is the iteration number):

$$p_{m,q}^{(j)} = \frac{\pi_q^{(j)} \phi_q(x_m, \mu_q^{(j)}, \Sigma_q^{(j)})}{\sum_{\hat{q}=1}^Q \pi_{\hat{q}}^{(j)} \phi_{\hat{q}}(x_m, \mu_{\hat{q}}^{(j)}, \Sigma_{\hat{q}}^{(j)})}. \quad (15)$$

The  $Q$  number of priors were updated as follows:

$$\pi_q^{(j+1)} = \frac{1}{M} \sum_{m=1}^M p_{m,q}^{(j)}. \quad (16)$$

The  $Q$  number of means were updated as follows:

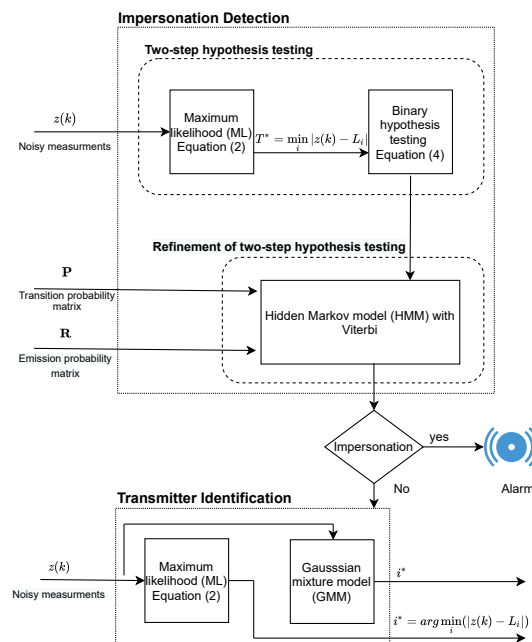
$$\mu_q^{(j+1)} = \frac{\sum_{m=1}^M p_{m,q}^{(j)} x_m}{\sum_{m=1}^M p_{m,q}^{(j)}}. \quad (17)$$

The  $Q$  number of (co-)variances were updated as follows:

$$\Sigma_q^{(j+1)} = \frac{\sum_{m=1}^M p_{m,q}^{(j)} (x_m - \mu_q^{(j)}) (x_m - \mu_q^{(j)})^T}{\sum_{m=1}^M p_{m,q}^{(j)}}. \quad (18)$$

The iterative EM algorithm monotonically increased the objective (likelihood) function value and converged when the increase in the likelihood function value between two successive iterations became less than the threshold  $\epsilon$ .

Figure 3 shows a flow graph of the proposed methodology. The noisy estimated measurement/path loss  $z(k)$  at slot  $k$  was fed to a two-step mechanism for impersonation detection, and the HMM was used to refine the outcomes of the two-step mechanism with the help of transition and emission probability matrices (i.e.,  $\mathbf{P}$  and  $\mathbf{R}$ ) and the Viterbi algorithm. Transmitter identification was done via the ML and GMM approaches when no impersonation was decided.



**Figure 3.** Proposed methodology for impersonation detection and transmitter identification in body-centric nanoscale communication systems operating in the THz band.

## 6. Simulations

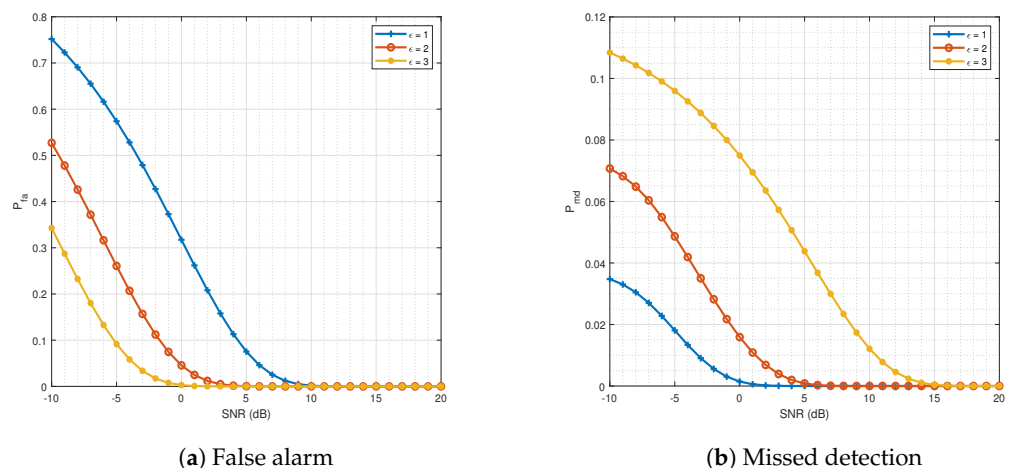
### 6.1. Setup

We kept  $M = N = 10$ ,  $\alpha_{ij} = 0.5 \forall j$ ,  $f = 1$  THz,  $T = 285$  k, and  $p = 1$  atm. Both the Alice and Eve nodes were deployed according to the uniform distribution in a  $1 \text{ m} \times 1 \text{ m}$  area. A total of  $10^5$  random realizations (independent of the Alice and Eve nodes) of the nodes' deployment were taken, and then, errors were averaged over the realizations.

$P_{fa}$  and  $P_{md}$  are two well-known probabilities resulting in hypothesis testing.  $P_{fa}$  was defined as the probability that any  $i$ -th Alice node can be considered as any of the Eve nodes  $P_{md}$  is the probability of the event that any  $j$ -th Eve node can be considered as any of the Alice nodes.

### 6.2. Results

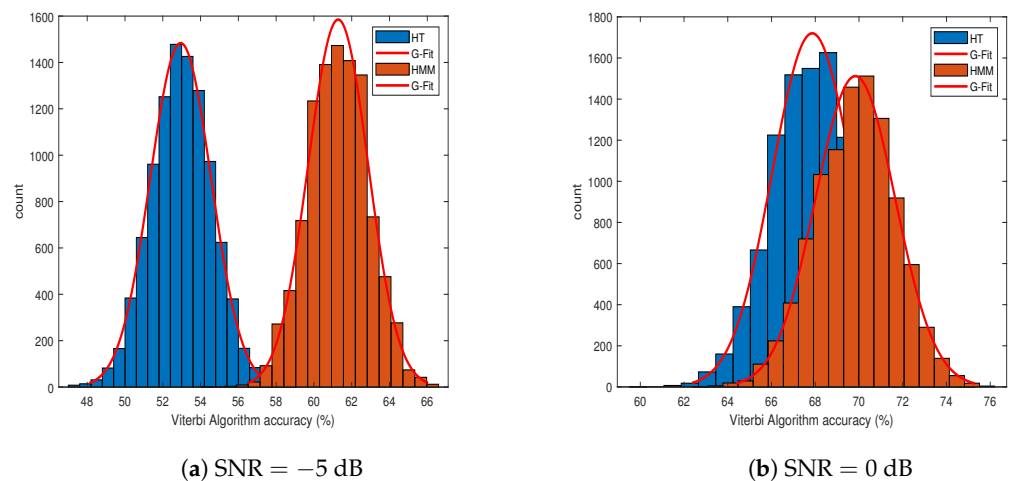
Figure 4 represents the two probabilities against  $\text{SNR} = \frac{1}{\sigma^2}$  where the improvement in error probabilities with an increasing SNR can be seen clearly. The designed parameter  $\epsilon$  decreased  $P_{md}$ , but increased  $P_{fa}$ .



**Figure 4.** The error probabilities against  $\text{SNR} = \frac{1}{\sigma^2}$  with different values for threshold  $\epsilon$ . (a) Probability of false alarm. (b) Probability of missed detection. Both probabilities are decreasing functions of SNR.

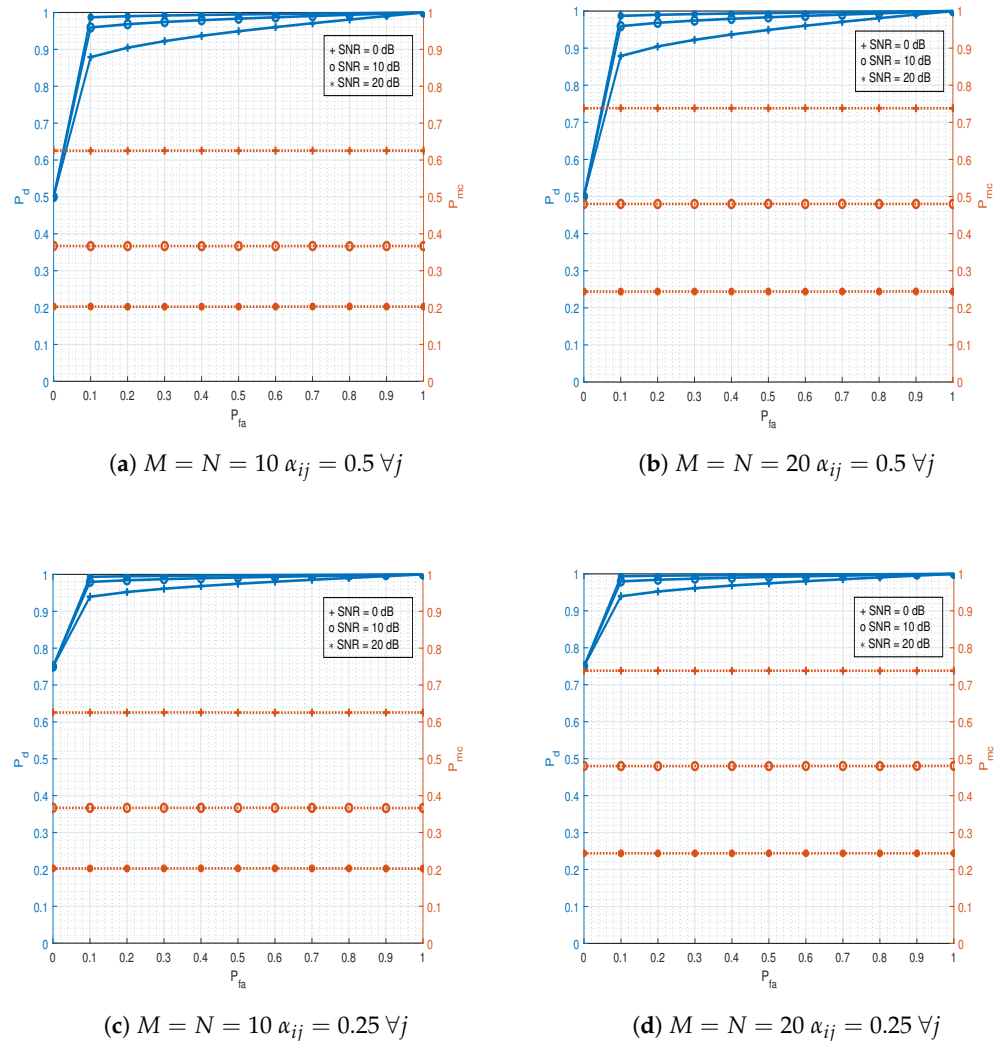


Figure 5 shows the efficacy of the HMM. At a low SNR, the performance of the HMM was far better than HT, and at a high SNR, HT was close to the HMM. The results were produced after the Monte Carlo-based simulation. The total number of transmissions was kept to  $10^5$  (more specifically,  $10^5$  binary states  $(s_0, s_1)$  were generated),  $\epsilon = 1$ ,  $\mathbf{P} = 0.5\mathbf{I}_{2 \times 2}$ , where  $\mathbf{I}$  is the identity matrix and  $K = 10^3$ . The errors resulting from the HT and HMM methods were calculated as the number of times the predicted/estimated state was not equal to the actual state divided by the total transmissions. The accuracy was then computed accordingly. The entries of  $\mathbf{R}$  were calculated according to  $P_{fa}$  and  $P_{md}$ . Figure 6 shows the receiver operating characteristic (ROC) curves for different configurations of the nodes and transmissions from Eve nodes (i.e.,  $\alpha_{ij}$ ). Typically, the ROC contains two error probabilities ( $P_d$  and  $P_{fa}$ ), but due to multiple nodes in this study, we had three probabilities. For any  $P_{fa}$  value,  $P_{mc}$  was constant, which is obvious from Equation (13). Increasing the SNR not only improved  $P_d$ , but also improved  $P_{mc}$  as well.  $P_{fa}$  was chosen as an independent variable and swept in the range from zero to one. Using Equation (6), the threshold was calculated for a given SNR value. Further,  $P_d = 1 - P_{md}$  (the detection probability) and  $P_{mc}$  were computed as the average after doing  $10^5$  uniform realizations of the nodes' deployment. We observed that increasing the number of nodes did not affect  $P_d$ , but  $P_{mc}$  increased with an increase in the number of Alice nodes ( $M$ ). We further observed that when fewer nodes (Alice nodes) remained idle during their allocated slot, the more  $P_d$  we had.



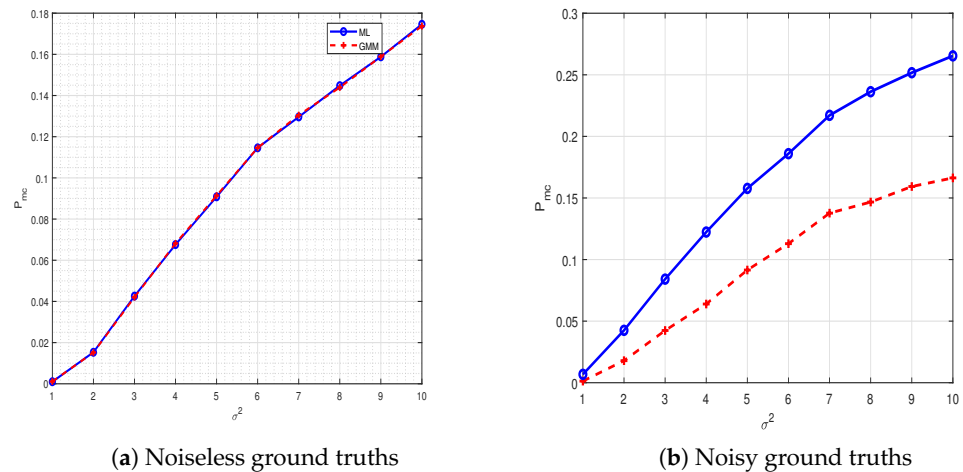
**Figure 5.** Performance comparison of two-step hypothesis testing and the hidden Markov model (HMM) with Viterbi algorithm. (a) Histogram comparison for a highly degraded channel, i.e., SNR =  $-5$  dB. (b) Histogram comparison for a moderately degraded channel, i.e., SNR =  $0$  dB. Performances of both approaches get closer and closer when SNR increases.

$P_{mc}$  is the probability of deciding the  $i$ -th Alice node, as any Alice node without  $i$ .  $P_{mc}$  becomes an important metric when dealing with multiple nodes' identification. Here,  $P_{mc}$  resulted from both transmitter identification algorithms (ML, which is a bi-product of two-step HT-based authentication and the GMM). As the GMM is a learning approach, it requires training data to learn its parameters. That is the reason that we only performed transmitter identification using the GMM. We assumed no data were available for Eve nodes.



**Figure 6.** Receiver operating characteristic (ROC) curves: Three probabilities (false alarm, detection, and misclassification) are considered in the ROC. To study the impact of nodes, subfigures (a,b) are plotted. (a) Ten numbers of legitimate and malicious nodes are considered with 0.5 prior probability for a  $j$ -th malicious node. (b) Twenty numbers of legitimate and malicious nodes are considered with 0.5 prior probability for a  $j$ -th malicious node. Further, subfigures (c,d) are plotted to see the impact of transmissions /prior probabilities of malicious nodes. (c) Ten numbers of legitimate and malicious nodes are considered with 0.25 prior probability for a  $j$ -th malicious node. (d) Twenty numbers of legitimate and malicious nodes are considered with 0.25 prior probability for a  $j$ -th malicious node.

Figure 7a was generated by assuming actual ground truths (noiseless ( $L_i \forall i$ )) of Alice nodes available for performing ML-based transmitter identification. The ML was implemented using Equation (2) having noiseless ground truths. Figure 7a shows that the two approaches performed equally. To test the efficacy of the GMM approach, we performed another experiment and plotted the results in Figure 7b. This time, we assumed that the ground truths of the Alice nodes were noisy  $L_i + n \forall i$  (i.e., when the ground truths were obtained on a secure channel, it also included noise or an error). This time, the ML-based approach was implemented using Equation (2) to include noisy ground truths. The GMM parameters were estimated on  $10^4$  training data generated from the legitimate nodes and then tested on  $10^5$ . The error was calculated as the number of times the estimated state was not equal to the actual value divided by the total transmissions for both approaches and for both cases. We observed from Figure 7b that the overall performance of GMM was improved. The performance improved even further for lower SNR or higher  $\sigma^2$ .



**Figure 7.** Misclassification error  $P_{mc}$  against estimation error  $\sigma^2$  for two-step hypothesis testing and GMM. (a) A scenario is considered where the perfect ground truth vector, i.e.,  $\mathbf{1}$  is obtained via a secure channel. (b) A scenario is considered where acquired ground truths are noisy. GMM has an advantage over ML when ground truths are noisy.

### 6.3. Discussions

- From Figures 4 and 6, we learned that the path loss could be exploited as a fingerprint to carry out authentication in body-centric nanoscale communication systems operating in the THz band. In other words, the proposed mechanisms can be used as a first line of defense against impersonation attacks.
- The results of the proposed two-step mechanism can be improved by using an additional approach (i.e., HMM). In particular, at a low SNR, the improvement was quite significant.
- The results in Figures 4 and 6 indicate that, under the impersonation detection problem, it is not possible to minimize both  $P_{md}$  and  $P_{fa}$  at the same time because of their conflicting nature. In other words, one could minimize one error type only by compromising the other error type.
- GMM (Learning-based scheme) performed the same as our proposed two-step mechanism in transmitter identification. However, we learned that slightly complex nature of the GMM could produce improvements when the ground truths of legitimate nodes are noisy.

## 7. Conclusions

This paper provided an authentication mechanism using path loss as a fingerprint at the physical layer in body-centric nanoscale communication systems operating in the terahertz band. The work's importance was advocated by illustrating envisioned smart healthcare application of body-centric nanoscale communication systems. The complex and quantum insecure crypto measures can be complemented using this approach, which is simple and quantum secure (i.e., no encryption or shared secret key is involved). This was observed from ROC curves after doing the Monte Carlo-based simulation for nodes' deployment under a uniform distribution that with a 20% false rate, the detection probability was almost one when operating with SNR = 10 dB. For simulation purpose, nodes were deployed in a  $1\text{ m} \times 1\text{ m}$  area under a uniform distribution, and air was considered as a medium among the nodes, while the path loss was calculated using the HITRAN database.

**Author Contributions:** Conceptualization, W.A. and M.M.U.R.; methodology, M.M.U.R.; software, W.A.; validation, H.T.A. and M.A.K.; writing—original draft preparation, W.A. and M.M.U.R.; writing—review and editing, H.T.A., M.A.K.; supervision, M.A.I., A.A., and Q.H.A.; project administration, M.A.I. and Q.H.A.; funding acquisition, A.A. All authors read and agreed to the published version of the manuscript.

**Funding:** This work was made possible by EPSRC Grant Number EP/T021063/1. The statements made herein are solely the responsibility of the authors.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The HITRAN database can be found here: <https://hitran.org>.

**Acknowledgments:** Waqas Aman would like to thank the Higher Education Commission of Pakistan (HEC) for providing him the IRSIP scholarship to travel to the University of Glasgow for his research studies.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

THz	Terahertz
HMM	Hidden Markov model
GMM	Gaussian mixture model
SNR	Signal-to-noise ratio
IoT	Internet of things
DoS	Denial of service
PL	Physical layer
AKE	Asymmetric key encryption
RSS	Received signal strength
CIR	Channel impulse response
CFR	Channel frequency response
2D	Two dimensional
ML	Maximum likelihood
CCDF	Cumulative complementary distribution function
MLSE	Maximum likelihood sequence estimation
EM	Expectation maximization
ROC	Receiver operating characteristics
HITRAN	High-resolution transmission molecular absorption

## Appendix A

The spreading loss is given as:

$$L_s(f, d)[dB] = 20 \log_{10}\left(\frac{4\pi fd}{c}\right), \quad (A1)$$

where  $c$  is the speed of light. The absorption loss is given as:

$$L_a(f, d) = \frac{1}{\tau(f, d)}, \quad (A2)$$

where  $\tau$  represents the transmittance of the signal and is given by the Beer–Lambert law:

$$\tau(f, d) = e^{-k(f)d}, \quad (A3)$$

where  $k$  is the medium absorption coefficient, given as:

$$k(f) = \sum_{i,g} k_{i,g}(f), \quad (A4)$$

where:

$$k_{i,g}(f) = \frac{p}{p_0} \frac{T_0}{T} Q_{i,g} \sigma_{i,g}(f) \quad (A5)$$

where  $i$  is the isotopologue (a molecule that differs in isotropic composition),  $g$  is gas,  $p_0(T_0)$  is standard pressure (temperature),  $\sigma_{i,g}(f)$  is the absorption cross-section, and  $Q_{i,g}$  is the molecular density given by:

$$Q_{i,g} = \frac{n}{V} q_{i,g} N_A = \frac{p}{RT} q_{i,g} N_A, \quad (\text{A6})$$

where  $R$  is the gas constant,  $N_A$  is the Avogadro constant, and  $q_{i,g}$  is the mixing ratio for  $i$  of  $g$ . The absorption cross-section can be expressed as:

$$\sigma_{i,g} = S_{i,g} G_{i,g}(f), \quad (\text{A7})$$

where the line intensity  $S_{i,g}$  and line shape  $G_{i,g}(f)$  parameters can be computed using data from the HITRAN database [19].

## References

- Mohammad, H.; Shubair, R.M. Nanoscale Communication: State-of-Art and Recent Advances. *arXiv* **2019**, arXiv:1905.07722. [[CrossRef](#)]
- Abbasi, Q.H. (Ed.) *Advances in Body-Centric Wireless Communication: Applications and State-of-the-Art*; Telecommunications; Institution of Engineering and Technology: London, UK, 2016.
- Akyildiz, I.F.; Brunetti, F.; Blázquez, C. Nanonetworks: A new communication paradigm. *Comput. Netw.* **2008**, *52*, 2260–2279. [[CrossRef](#)]
- Lemic, F.; Abadal, S.; Tavernier, W.; Stroobant, P.; Colle, D.; Alarcón, E.; Marquez-Barja, J.; Famaey, J. Survey on Terahertz Nanocommunication and Networking: A Top-Down Perspective. *IEEE J. Sel. Areas Commun.* **2019**. [[CrossRef](#)]
- Ma, J.; Shrestha, R.; Adelberg, J.; Yeh, C.Y.; Hossain, Z.; Knightly, E.; Jornet, J.M.; Mittleman, D.M. Security and eavesdropping in terahertz wireless links. *Nature* **2018**, *563*, 89–93. [[CrossRef](#)] [[PubMed](#)]
- Gidney, C.; Ekerå, M. How to factor 2048 bit RSA integers in 8 h using 20 million noisy qubits. *Quantum* **2019**, *5*, 433. [[CrossRef](#)]
- Shakiba-Herfeh, M.; Chorti, A.; Poor, H.V. Physical Layer Security: Authentication, Integrity and Confidentiality. In *Physical Layer Security*; Springer: Cham, Switzerland, 2020.
- Yang, J.; Chen, Y.; Trappe, W.; Cheng, J. Detection and Localization of Multiple Spoofing Attackers in Wireless Networks. *IEEE Trans. Parallel Distrib. Syst.* **2013**, *24*, 44–58. [[CrossRef](#)]
- Zafar, S.; Aman, W.; Rahman, M.M.U.; Alomainy, A.; Abbasi, Q.H. Channel Impulse Response-based Physical Layer Authentication in a Diffusion-based Molecular Communication System. In Proceedings of the 2019 UK/China Emerging Technologies (UCET), Glasgow, UK, 21–22 August 2019; pp. 1–2.
- Mahmood, A.; Aman, W.; Iqbal, M.O.; Rahman, M.M.U.; Abbasi, Q.H. Channel Impulse Response-Based Distributed Physical Layer Authentication. In Proceedings of the 2017 IEEE 85th Vehicular Technology Conference (VTC Spring), Sydney, Australia, 4–7 June 2017; pp. 1–5.
- Xiao, L.; Greenstein, L.J.; Mandayam, N.B.; Trappe, W. Using the physical layer for wireless authentication in time-variant channels. *IEEE Trans. Wirel. Commun.* **2008**, *7*, 2571–2579. [[CrossRef](#)]
- Baracca, P.; Laurenti, N.; Tomasin, S. Physical Layer Authentication over MIMO Fading Wiretap Channels. *IEEE Trans. Wirel. Commun.* **2012**, *11*, 2564–2573. [[CrossRef](#)]
- Hou, W.; Wang, X.; Chouinard, J. Physical layer authentication in OFDM systems based on hypothesis testing of CFO estimates. In Proceedings of the 2012 IEEE International Conference on Communications (ICC), Ottawa, ON, Canada, 10–15 June 2012; pp. 3559–3563.
- Hao, P.; Wang, X.; Behnad, A. Performance enhancement of I/Q imbalance based wireless device authentication through collaboration of multiple receivers. In Proceedings of the 2014 IEEE International Conference on Communications (ICC), Sydney, Australia, 10–14 June 2014; pp. 939–944.
- Zhao, N.; Zhang, Z.; Ur-Rehman, M.; Ren, A.; Yang, X.; Zhao, J.; Zhao, W.; Dong, B. Authentication in Millimeter-Wave Body-Centric Networks Through Wireless Channel Characterization. *IEEE Trans. Antennas Propag.* **2017**. [[CrossRef](#)]
- Dressler, F.; Kargl, F. Security in nano communication: Challenges and open research issues. In Proceedings of the 2012 IEEE International Conference on Communications (ICC), Ottawa, ON, Canada, 10–15 June 2012; pp. 6183–6187.
- Atlam, H.F.; Walters, R.J.; Wills, G.B. Internet of Nano Things: Security issues and applications. In Proceedings of the 2018 International Conference on Cloud and Big Data Computing, FuZhou, China, 15–17 November 2018; ACM: New York, NY, USA, 2018.
- Rahman, M.M.U.; Abbasi, Q.H.; Chopra, N.; Qaraqe, K.; Alomainy, A. Physical Layer Authentication in Nano Networks at Terahertz Frequencies for Biomedical Applications. *IEEE Access* **2017**, *5*, 7808–7815. [[CrossRef](#)]
- Gordon, I.; Rothman, L.; Hill, C.; Kochanov, R.V.; Tan, Y.; Bernath, P.F. The HITRAN2016 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.* **2017**, *203*, 3–69. [[CrossRef](#)]
- Jornet, J.M.; Akyildiz, I.F. Channel Capacity of Electromagnetic Nanonetworks in the Terahertz Band. In Proceedings of the 2010 IEEE International Conference on Communications, Cape Town, South Africa, 23–27 May 2010; pp. 1–6.

21. Jornet, J.M.; Akyildiz, I.F. Channel Modeling and Capacity Analysis for Electromagnetic Wireless Nanonetworks in the Terahertz Band. *IEEE Trans. Wirel. Commun.* **2011**, *10*, 3211–3221. [[CrossRef](#)]
22. Cid-Fuentes, R.G.; Jornet, J.M.; Akyildiz, I.F.; Alarcon, E. A receiver architecture for pulse-based electromagnetic nanonetworks in the Terahertz Band. In Proceedings of the 2012 IEEE International Conference on Communications (ICC), Ottawa, ON, Canada, 10–15 June 2012; pp. 4937–4942.
23. Rahman, M.M.U.; Yasmeen, A.; Gross, J. PHY layer authentication via drifting oscillators. In Proceedings of the 2014 IEEE Global Communications Conference, Austin, TX, USA, 8–12 December 2014; pp. 716–721.
24. Yan, Q.; Blum, R.S. Distributed signal detection under the Neyman-Pearson criterion. *IEEE Trans. Inf. Theory* **2001**, *47*, 1368–1377. [[CrossRef](#)]

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.