PHY-Layer Authentication with Multiple Landmarks with Reduced Overhead

Liang Xiao, *Senior Member, IEEE*, Xiaoyue Wan, *Student Member, IEEE*, and Zhu Han, *Fellow, IEEE*

***Abstract*—Physical (PHY)-layer authentication systems can exploit channel state information of radio transmitters to detect spoofing attacks in wireless networks. The use of multiple landmarks each with multiple antennas enhances the spatial resolution of radio transmitters, and thus improves the spoofing detection accuracy of PHY-layer authentication. Unlike most existing PHY-layer authentication schemes that apply hypothesis tests and rely on the known radio channel model, we propose a logistic regression based authentication to remove the assumption on the known channel model, and thus to be applicable to more generic wireless networks. The Frank-Wolfe algorithm is used to estimate the parameters of the logistic regression model, in which the convex problem under a ℓ_1-norm constraint is solved for weight sparsity to avoid over-fitting in the learning process. We design a distributed Frank-Wolfe (dFW) based PHY-layer authentication to further reduce the communication overhead between the landmarks and the security agent. Then we construct an incremental aggregated gradient (IAG)-based scheme to provide online authentication with a higher accuracy and lower computation overhead. Simulation and experimental results validate the accuracy of the proposed authentication schemes, and show the reduced communication and computation overheads.**

***Index Terms*—PHY-layer authentication, distributed Frank-Wolfe, incremental aggregated gradient, spoofing, machine learning.**

I. Introduction

Wireless networks are vulnerable to spoofing attacks, in which an attacker (Eve) impersonates another user (Alice) by claiming its higher-layer identity such as its media access control (MAC) address, which cannot be fully addressed by the traditional higher-layer authentication techniques. For example, by spoofing the control information and management frames, a spoofer can further launch Denial of service (DoS) attacks [1]–[6]. Therefore, physical (PHY)-layer authentication techniques have been proposed to exploit the spatial decorrelation property of the PHY-layer information, such as received signal strength indicators (RSSIs) [7]–[9], received signal strength [10], channel phase response [11], channel impulse responses [12], and channel state information [13]–[16] to distinguish radio transmitters, and thus detect spoofing attacks with low overhead.

The system equipped with multiple landmarks at different locations that use multiple antennas to estimate the radio channel information can increase the spatial resolution of the radio channel regarding the radio transmitter. Therefore, a PHY-layer authentication system can utilize multiple landmarks each equipped with multiple antennas to increase the spatial resolution, and thus improve the spoofing detection accuracy.

Most PHY-layer authentication methods use the channel estimation at the receiver (Bob), i.e., the pilots or preambles of the incoming packets are applied to estimate the channel states. A hypothesis test is usually built to evaluate the channel states based on a specific channel model. For instance, the PHY-authentication as developed in [17] assumes the frequency-selective Rayleigh channel model. However, it's challenging to build an accurate channel model in time especially in a dynamic wireless network with multi-path and fast fading. Therefore, logistic regression techniques can be used in the PHY-layer authentication to remove the assumption on the known radio channel distribution. For example, by applying the maximum log-likelihood (ML) estimation to process the training channel data, the logistic regression based authentication in [18] is applicable to a broad range of wireless networks.

The ML-based coefficient estimation of the regression can be formulated as a convex optimization problem, and thus can be efficiently addressed by the Frank-Wolfe (FW) algorithm, which uses linear subproblems to approximate the convex and constrained optimization problem at each iteration [19]. Moreover, the coefficient is estimated under the ℓ_1-norm constraint to restrict the model complexity and make a tradeoff between the training error and the testing error of the regression. However, the FW-based PHY-layer authentication with multiple landmarks requires the transmission of the channel information from each landmark to the receiver Bob or a special security agent, yielding forbidden communication overhead, if the numbers of the landmarks and the antennas are large.

Therefore, the PHY-layer authentication with multi-landmarks can apply the distributed Frank-Wolfe (dFW) algorithm to solve the convex constraint optimization problem of FW in a distributed system. As shown in [20], the dFW-based authentication accelerates the convergence rate with a small amount of computational overhead at each iteration and reduce the communication cost to solve the convex constraint optimization problem in a distributed system. We apply the

This work was supported in part by the National Natural Science Foundation of China under Grant 61671396 and 91638204, the open research fund of National Mobile Communications Research Laboratory，Southeast University (No.2018D08), and US NSF CNS-1717454, CNS-1731424, CNS-1702850, CNS-1646607,ECCS-1547201,CMMI-1434789,CNS-1443917, and ECCS-1405121.

L. Xiao is with the Dept. of Communication Engineering, Xiamen University, Xiamen 361005, China, and also with the National Mobile Communications Research Laboratory, Southeast University, China (e-mail: lxiao@xmu.edu.cn).

X. Wan is with the Dept. of Communication Engineering, Xiamen University, Xiamen 361005, China.

Z. Han is with the University of Houston, Houston, TX 77004 USA (e-mail:zhan2@uh.edu), and also with the Department of Computer Science and Engineering, Kyung Hee University, Seoul, South Korea.

logistic regression model to evaluate the channel information collected at multiple antennas at a number of landmarks at different locations to detect spoofing more accurately. The coefficients of the regression model is estimated with the dFW algorithm to reduce the communication overhead between the landmarks and the security agent.

The dFW-based authentication has to recalculate the gradients of all the training data in each iteration, resulting large computation overheads and long delay. An online method called incremental aggregated gradient (IAG) as illustrated in [21], [22] is introduced to solve the logistic regression problem in authentication, as the objective function in the unconstrained optimization problem can be written as the sum of convex and continuously differentiable component functions. In the IAG-based PHY-layer authentication, the component functions are processed in a deterministic order and the previous gradient values are used to accelerate the convergence. Simulations based on the stochastic channel model and experiments based on universal software radio peripherals (USRPs) are performed to verify the spoofing detection performance of our proposed systems.

The main contributions of our work can be summarized as follows:

1) We investigate the RSSI-based authentication that uses multiple landmarks each with multiple antennas to collect the RSSIs of Alice's signals and applies the FW algorithm based on the logistic regression model to detect spoofing in more generic wireless networks without being aware of the channel model.
2) We propose a dFW-based authentication to reduce the overall communication overhead between the landmarks and the security agent, and an IAG-based authentication to reduce the computation overhead compared with the FW-based authentication. We analyze the convergence performance of the dFW and IAG based authentication schemes and compare their communication and computation overheads.
3) Simulation and experimental results show that the proposed dFW-based PHY-layer authentication can achieve the same spoofing detection accuracy as the FW-based authentication and reduce the communication overhead. The IAG-based PHY-layer authentication reduces the computation overhead and achieves a higher detection accuracy.

The rest of this paper is organized as follows. We review the related work in Section II and describe the system model in Section III. We present the FW-based PHY-layer authentication in Section IV and the dFW-based authentication in Section V. The IAG-based authentication is presented in Section VI. We discuss their convergence performance in Section VII. We present simulation results in Section VIII and conclude the paper in Section IX.

II. Related Work

PHY-layer authentication exploits the PHY-layer information of radio channels to detect spoofing attacks in wireless systems. For instance, the channel response based authentication with a single antenna as developed in [17] uses the generalized likelihood ratio test to detect spoofing under the frequency-selective Rayleigh radio channel model. The PHY-layer authentication proposed in [23] that builds a likelihood ratio test to evaluate the channel responses can achieve fast authentication and minimize the packet transmission overhead in distributed wireless sensor networks. The PHY-layer authentication as proposed in [24] exploits the randomness, reciprocity and location decorrelation characteristics of radio channels to prevent attackers form obtaining the channel state information and thus revealing the authentication key. The time-varying carrier frequency offsets between each transmitter-receiver pair are formulated in [25] as the user signatures in the PHY-layer authentication. The embedded device fingerprint based authentication as proposed in [26] adds low-power authentication tags that is not accessible to adversaries on the message signal to enhance security with low radio bandwidth overhead. The PHY-layer authentication in complex heterogenous networks is investigated in [27], which applies the multi-attribute multi-observation technique to enhance reliability and uses the composite security key to integrate it with existing wireless networks.

The RSS-based authentication system proposed in [10] applies the cluster analysis method and multiple landmarks to improve the spoofing detection accuracy. The energy ratio based authentication developed in [28] explores the asymmetry of the received signal power levels at the transmitter and the legitimate receiver to resist the pilot spoofing and uses multiple antennas to improve the detection accuracy. The PHY-authentication scheme proposed in [29] exploits the wireless link fading to verify whether a received message is indeed sent by the legitimate source node in the cellular Internet of things network, and applies multiple anchor nodes to improve the authentication accuracy.

Machine learning techniques have shown strength to improve the network security [30], [31]. For example, the support vector machine and unsupervised learning methods have been used in [32] to detect the stealthy false data attacks in smart grid. The Q-learning based PHY-layer authentication as proposed in [33] applies Q-leaning to determine the test threshold of the hypothesis test for the single-antenna system without being aware of the radio channel model and this work has been extended to MIMO system in [34]. Compared with [33] and [34], this work can further increase the spoofing detection accuracy.

A PHY-layer authentication with multiple MIMO landmarks as developed in [18] uses the logistic regression model and applies the dFW algorithm to detect spoofing. Compared with our previous work in [18], we propose an IAG-based authentication to reduce the computation overhead and improve the spoofing detection accuracy. We also extend the work by analyzing the convergence of the dFW and IAG based algorithms.

III. System Model

As shown in Fig. 1, we consider the spoofing detection with the Alice-Bob-Eve model, assisted by M landmarks (or radio monitors) each equipped with N antennas. Alice as

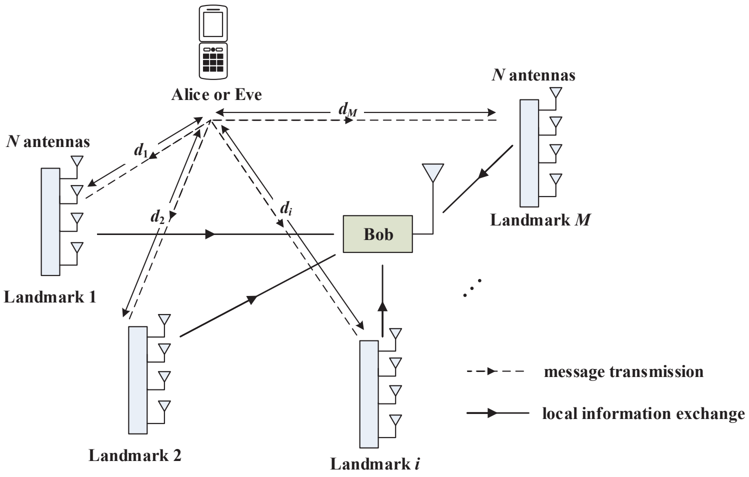


Fig. 1. Illustration of a PHY-layer authentication system consisting of a transmitter that is either Alice or Eve, and a security agent Bob, based on the channel information estimated by M landmarks each with N antennas.

TABLE I
SUMMARY OF SYMBOLS AND NOTATIONS

M	Number of landmarks
N	Number of receive antennas
T	Number of training data in the regression
h^i_{mj}	RSSI at antenna j on landmark m
$\mathbf{H}^i_m$	Local channel vector of message i on landmark m
$\mathbf{H}_i$	Channel vector of message i
L^i_{mj}	Path loss of message i at antenna j on landmark m
$\hat{\boldsymbol{\beta}}_m$	Local estimated coefficient vector of landmark m
$\beta_0/\hat{\beta}_0$	Actual/estimated intercept term
$y_i/\hat{y}_i$	Actual/estimated transmitter of message i
χ	Flag to stop the iteration
ψ_m	Index of the largest local gradient of landmark m
S_m	Partial sum for stopping criterion of landmark m
ρ	Node index with the largest overall gradient
ψ	Index of the largest overall gradient
ϵ	Approximation quality
g	Aggregated gradient

the legal transmitter sends signals with a single antenna. Eve impersonates Alice by sending signals to Bob with her MAC address. Being a legitimate receiver or security agent, Bob has to verify whether the i-th signal is indeed sent by Alice. All the radio nodes are assumed to be static for simplicity.

By claiming the MAC address of Alice, Eve aims to bypass of the access control lists of the serving routers to obtain illegal advantages, and sometimes launch further attacks such as DoS. We assume that Eve cannot compromise the local information exchange from the landmarks to Bob to inject fake information [29], or the training process for simplicity.

Because of the spatial decorrelation of radio channels, the signals sent by Alice infers the PHY-layer radio channel information such as RSSI that are different from that of Eve. The channel states can significantly change if the transmitter moves more than a wavelength away from the original location [35]. Bob authenticates the received signal based on RSSI vector estimated by multiple landmarks according to the previous RSSI vector record. More specifically, upon receiving the i-th message, landmark m estimates the received signal strength indicator of the message at antenna j, denoted by h^i_{mj}, with $1 \leq m \leq M$, and $1 \leq j \leq N$. The MN-dimensional column channel vector of the i-th message is denoted by $\mathbf{H}_i$, which consists of the RSSIs measured by the M landmarks, with $\mathbf{H}_i = [h^i_{mj}]_{M\times N}$. Each landmark can process the channel information and send some information to Bob without interfering with the transmissions of Alice, if needed.

Although our proposed authentication system does not rely on any specific channel model, for evaluation purposes, the simulations in Section VII assume that the path loss of message i between the transmitter and antenna j on landmark m, denoted by L^i_{mj}. Let d_m be the distance between landmark m and the transmitter. According to the generic model as specified in [36], we have

$$L^i_{mj}[dB] = A + 10\gamma \log_{10}(d_m/d_0) + s^i_{mj}, \qquad (1)$$

where s^i_{mj} is the shadow fading (in dB) along that path, intercept A is the decibel path loss at reference distance d_0, γ is path loss exponent and s^i_{mj} is a Gaussian random variable with zero mean and standard deviation σ. The channel estimation error is assumed to be zero-mean, complex Gaussian distributed and independent across the M landmarks and N antennas, with a common variance denoted by σ^2_I. Important symbols and notations are summarized in Table I.

IV. FW-BASED PHY-LAYER AUTHENTICATION

The PHY-layer authentication system exploits the spatial decorrelation of the radio channel information with M landmarks each equipped with N antennas to authenticate a message sent by the radio transmitter under test that is either Alice or Eve. Let y_i denote the actual transmitter of the i-th message. If message i is sent by Alice, $y_i = 1$; otherwise if the message is a spoofing signal sent by Eve, $y_i = 0$. The detection result of message i is denoted by $\hat{y}_i$: if $\hat{y}_i = 1$, the estimated transmitter is Alice; a spoofing alarm is sent if $\hat{y}_i = 0$. The PHY-layer authentication is made based on the training data, which are the authentication of the previous T messages in similar scenarios. The training data are the channel measurement results of the signals from different transmitters and the corresponding MAC address in similar scenarios, which include the channel information from the transmitters at different locations to avoid the bias. Note that the "spoofers" in the training phase are other transmitters in the area that are either legal users or detected attackers. The current authentication results can be included in the training data set to authenticate the messages in the next time slot. For simplicity, we assume the training data are accurate with $\hat{y}_k = y_k, \forall i-T-1 \leq k \leq i-1$.

Let an MN-dimensional row vector $\boldsymbol{\beta}$ indicate the importance of each feature in channel vector $\mathbf{H}_i$ and be the coefficients of the logistic regression model. The term β_0 is the bias in the logistic regression. The logistic regression presents the predicted probability for each class with the softmax function, which is the gradient-log-normalizer of the categorical probability distribution. According to [37], in the authentication of message i, the predicted probability for $y_i = 0$ or $y_i = 1$ based on the channel state information $\mathbf{H}_i$ can be approximately modeled with the softmax function as follows

$$\Pr\left(y_i = 1|\mathbf{H}_i\right) = \frac{e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_i}}{1+e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_i}}, \tag{2}$$

$$\Pr\left(y_i = 0|\mathbf{H}_i\right) = \frac{1}{1+e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_i}}. \tag{3}$$

It's clear from (2) and (3) that $\Pr(y_i = 1|\mathbf{H}_i) > \Pr(y_i = 0|\mathbf{H}_i)$ is equivalent to $\beta_0 + \boldsymbol{\beta}\mathbf{H}_i > 0$ after simplification. Therefore, we have

$$\hat{y}_i = \begin{cases} 1, & \text{if } \beta_0 + \boldsymbol{\beta}\mathbf{H}_i > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Without knowing the distribution of the RSSI vector, Bob can use the logistic regression technique to authenticate message i according to channel vector $\mathbf{H}_i$, and apply the previous T channel vectors, (i.e., $\{\mathbf{H}_k\}_{i-T-1\leq k\leq i-1}$) as the training data to estimate the coefficients of the logistic regression model. According to (4.19) in [37], the minus log-likelihood of T training data in the PHY-layer authentication with multiple MIMO landmarks denoted by $f(\boldsymbol{\beta})$ is defined by definitions

$$f(\boldsymbol{\beta}) = -\ln\left(\prod_{k=i-T-1}^{i-1} \Pr\left(y_k|\mathbf{H}_k\right)\right). \tag{5}$$

The training data includes not only the messages sent by Alice, but also the spoofing messages from Eve.

Lemma 1. *In the FW-based authentication with the logistic regression model, we have*

$$f(\boldsymbol{\beta}) = -\sum_{k=i-1-T}^{i-1} \left(y_k\left(\beta_0 + \boldsymbol{\beta}\boldsymbol{H}_k\right) - \ln\left(1+e^{\beta_0+\boldsymbol{\beta}\boldsymbol{H}_k}\right)\right). \tag{6}$$

Proof. By (2), (3), and (4), we can rewrite $f(\boldsymbol{\beta})$ as follows

$$f(\boldsymbol{\beta}) = -\ln\Bigg(\prod_{\substack{k=i-T-1\\y_k=1}}^{i-1} \Pr(y_k = 1|\mathbf{H}_k) \prod_{\substack{k=i-T-1\\y_k=0}}^{i-1} \Pr(y_k = 0|\mathbf{H}_k)\Bigg) \tag{7}$$

$$= -\sum_{k=i-T-1}^{i-1} \Bigg(y_k \ln \Pr(y_k = 1|\mathbf{H}_k) + (1-y_k)\ln \Pr(y_k = 0|\mathbf{H}_k)\Bigg) \tag{8}$$

$$= -\sum_{k=i-1-T}^{i-1} \Bigg(y_k \ln \frac{e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k}}{1+e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k}} + (1-y_k)\ln\frac{1}{1+e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k}}\Bigg) \tag{9}$$

$$= -\sum_{k=i-1-T}^{i-1} \Bigg(y_k(\beta_0 + \boldsymbol{\beta}\mathbf{H}_k) - \ln\left(1+e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k}\right)\Bigg).$$

□

The coefficient vector $\boldsymbol{\beta}$ is estimated by the maximum log-likelihood method following the ℓ_1-norm regularization to restrict the complexity of the logistic regression model according to [38]. More specifically, the cost function $f(\boldsymbol{\beta})$ in (6) is minimized under the ℓ_1-norm penalty, and the estimated coefficient vector denoted by $\hat{\boldsymbol{\beta}}$ is given by

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{k=i-1-T}^{i-1} \Bigg(\ln(1+e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k}) - y_k(\beta_0 + \boldsymbol{\beta}\mathbf{H}_k)\Bigg) \quad s.t. \quad ||\hat{\boldsymbol{\beta}}||_1 \leq C, \tag{10}$$

where the positive parameter C restricts the model complexity and makes a tradeoff between the training error and the testing error of the regression algorithm.

We do not penalize the intercept term β_0 and standardize the channel vectors to make a meaningful penalty. By [37], β_0 is estimated by

$$\hat{\beta}_0 = \frac{1}{T}\sum_{k=i-T-1}^{i-1} y_k. \tag{11}$$

The Frank-Wolfe algorithm described in [19] is used to estimate $\boldsymbol{\beta}$, as the objective function $f(\boldsymbol{\beta})$ in (6) is convex and continuously differentiable with respect to $\boldsymbol{\beta}$. By (6), the gradient $\nabla f(\boldsymbol{\beta})$ as an MN-dimensional vector consists of the local gradient of landmark m, with the j-th element given by

$$\nabla f(\boldsymbol{\beta})_j = \sum_{k=i-T-1}^{i-1} \left(\frac{e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k} H_{kj}}{1+e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k}} - y_k H_{kj}\right), \tag{12}$$

where H_{kj} is the j-th element of channel vector $\mathbf{H}_k$ and $1 \leq j \leq MN$. The largest absolute value of the overall gradient $\nabla f(\boldsymbol{\beta})$ is chosen as the optimal descent direction of the cost function. The corresponding index of the direction denoted by ψ is given by

$$\psi = \arg \max_{1 \leq j \leq MN} \left| \nabla f(\boldsymbol{\beta})_j \right| . \tag{13}$$

The optimal descent direction of the cost function in terms of $\boldsymbol{\beta}$ is given by $\text{sgn}(-\nabla f(\boldsymbol{\beta})_\psi) C \mathbf{e}^\psi - \boldsymbol{\beta}$. The iteration step size at iteration n denoted by γ is set to be $\gamma = \frac{2}{n+2}$ by [19]. Therefore, the subsequent iteration point after moving step size γ along the optimal the descent direction at the current iteration point is updated with

$$\boldsymbol{\beta} \leftarrow (1-\gamma)\boldsymbol{\beta} + \gamma \text{sgn}\left(-\nabla f(\boldsymbol{\beta})_\psi\right) C \mathbf{e}^\psi , \tag{14}$$

where $\mathbf{e}^\psi$ is the all zeros MN-dimensional vector except 1 at the ψ-th entry. The flag to stop the iteration is denoted by χ and given by

$$\chi = \sum_{j=1}^{MN} \beta_j \nabla f(\boldsymbol{\beta})_j + C \left| \nabla f(\boldsymbol{\beta})_\psi \right| . \tag{15}$$

The approximation quality of the above FW algorithm to solve the convex problem in (10), denoted by ϵ, means that the estimated $\hat{\boldsymbol{\beta}}$ is at least ϵ-close to the optimal $\boldsymbol{\beta}$. If $\chi < \epsilon$, Bob obtains the ϵ-approximate of $\boldsymbol{\beta}$ and then uses (4) to detect message i. By (4), if $\hat{y}_i = 1$ Bob receives message i. Otherwise, Bob sends the spoofing alarm, if $\hat{y}_i = 0$.

In summary, the FW-based authentication calculates the local gradient of each landmark to find the largest overall gradient of the cost function of the training data and determine the subsequent iteration point to estimate $\boldsymbol{\beta}$ in the logistic regression model. Therefore, the resulting communication overhead between the security agent Bob and the M landmarks depends on the landmark topology and can be prohibitive under a large number of landmarks and antennas. The PHY-layer authentication with FW is summarized in Algorithm 1.

Algorithm 1 FW-based PHY-layer Authentication

```
Initialize: $\boldsymbol{\beta} = \mathbf{0}, n = 0, \chi, \epsilon$, and $C$
Channel estimation to form $\mathbf{H}_i$
Calculate $\hat{\beta}_0$ via (11)
While $\chi > \epsilon$ do
    Compute $\psi$ via (13)
    Set $\gamma = \frac{2}{n+2}$
    Update $\boldsymbol{\beta}$ via (14)
    Compute $\chi$ via (15)
    $n \leftarrow n + 1$
End while
Obtain $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$
Calculate $\hat{y}_i$ via (4)
If $\hat{y}_i = 1$ then
    Accept message $i$
Else
    Send spoofing alarm for message $i$
End if
```

V. dFW-based PHY-layer Authentication

In order to reduce the communication overhead of the FW-based authentication, the PHY-layer authentication can apply the dFW algorithm as described in [20] to solve the convex optimization problem with the ℓ_1-norm regularization in (10). As an advanced FW algorithm designed for distributed systems, the dFW-based authentication requires each landmark to send a few channel information instead of the entire channel vector. More specifically, we use $\mathcal{A}_m = \{(m-1)N+1, \cdots, mN\}$, i.e., the set of the column indices in both $\nabla f(\boldsymbol{\beta})$ and $\boldsymbol{\beta}$ that associated with landmark m. Thus, the local gradient of landmark m is given by the elements in $\nabla f(\boldsymbol{\beta})$ with indices belonging to $\mathcal{A}_m$. At iteration n, similar to (13) in FW, landmark m identifies the largest component of its local gradient in absolute value, whose indice in the gradient is denoted by ψ_m and given by

$$\psi_m = \arg \max_{j \in \mathcal{A}_m} |\nabla f(\boldsymbol{\beta})_j|. \tag{16}$$

Landmark m calculates the local flag to stop the iteration denoted by S_m and given by [20] as

$$S_m = \sum_{j \in \mathcal{A}_m} \beta_j \nabla f(\boldsymbol{\beta})_j. \tag{17}$$

Then the information set, $\{\nabla f(\boldsymbol{\beta})_{\psi_m},\ \psi_m,\ S_m\}$ is sent by landmark m to Bob. Upon receiving the information set from all the M landmarks, Bob identifies the landmark with the largest overall gradient, whose index is denoted by ρ and given by

$$\rho = \arg \max_{m \in [1,M]} |\nabla f(\boldsymbol{\beta})_{\psi_m}|, \tag{18}$$

and the index of the largest overall gradient is denoted by ψ_ρ.

Bob computes the flag to stop the iteration χ by

$$\chi = \sum_{m=1}^{M} S_m + C |\nabla f(\boldsymbol{\beta})_{\psi_\rho}|. \tag{19}$$

If $\chi > \epsilon$, Bob broadcasts $\{\psi_\rho, \nabla f(\boldsymbol{\beta})_{\psi_\rho}\}$ to the M landmarks, and then each landmark updates the coefficient vector accordingly by

$$\boldsymbol{\beta} \leftarrow (1-\gamma)\boldsymbol{\beta} + \gamma \text{sgn}(-\nabla f(\boldsymbol{\beta})_{\psi_\rho}) C \mathbf{e}^{\psi_\rho}. \tag{20}$$

Each landmark also determines whether to broadcast its local channel information to other landmarks. For simplicity, let the $T \times MN$ matrix $\mathbf{H} = [\mathbf{H}_k]_{i-T-1 \leq k \leq i-1}$ represent the training channel data matrix of message i. If finding $\psi_m = \psi_\rho$, landmark m broadcasts the ψ_m-th column of $\mathbf{H}$, which is used by the other $M-1$ landmarks to update their local gradients in the next iteration.

Otherwise, if $\chi \leq \epsilon$, Bob stops the iteration and notifies all the landmarks about it. At this time, landmark m obtains the estimated $\hat{\boldsymbol{\beta}}$. Landmark m the determines $\hat{\boldsymbol{\beta}}_m$, which is the locally estimated coefficient vector with $\hat{\boldsymbol{\beta}}_m = \{\hat{\beta}_j\}_{j \in \mathcal{A}_m}$. Let the N-dimensional vector $\mathbf{H}_m^i$ denote the local channel vector of message i at landmark m, which sends $\hat{\boldsymbol{\beta}}_m \mathbf{H}_m^i$ to Bob. Bob

Algorithm 2 Procedure at landmark m

```
Initialize: $\boldsymbol{\beta}=\mathbf{0}$, $n=0$, flag=0, $\chi$, and $C$
Channel estimation to form $\mathbf{H}_m^i$
While flag=0 do
    Compute $\nabla f(\boldsymbol{\beta})_{j\in\mathcal{A}_m}$ via (12)
    Compute $\psi_m$ via (16)
    Compute $S_m$ via (17)
    Send $\{\nabla f(\boldsymbol{\beta})_{\psi_m},\, \psi_m,\, S_m\}$ to Bob
    Receive $\{\psi_\rho,\, \nabla f(\boldsymbol{\beta})_{\psi_\rho}\}$ from Bob
    if $\psi_m=\psi_\rho$ do
        Broadcast the $\psi_m$-th column of $\mathbf{H}$
    End if
    Set $\gamma=\frac{2}{n+2}$
    Update $\boldsymbol{\beta}$ via (20)
    $n\leftarrow n+1$
End while
Send $\hat{\boldsymbol{\beta}}_m\mathbf{H}_m^i$ to Bob
```

Algorithm 3 DFW-based PHY-layer authentication

```
Initialize: flag=0, $\chi,\epsilon$, and $C$
Calculate $\hat{\beta}_0$ via (11)
While $\chi>\varepsilon$ do
    Receive $\{\nabla f(\boldsymbol{\beta})_{\psi_m},\, \psi_m,\, S_m\}$ from landmark $m$.
    Compute $\rho$ via (18)
    Broadcast $\{\psi_\rho,\, \nabla f(\boldsymbol{\beta})_{\psi_\rho}\}$ to each landmark
    Compute $\chi$ via (19)
End while
Set flag=1 and notify the $M$ landmarks
Receive $\hat{\boldsymbol{\beta}}_m\mathbf{H}_m^i$ from the $M$ landmarks
Compute $\hat{\beta}_0+\sum_{m=1}^{M}\hat{\boldsymbol{\beta}}_m\mathbf{H}_m^i$
Estimate $\hat{y}_i$ via (21)
If $\hat{y}_i=1$ then
    Accept message $i$
Else
    Send spoofing alarm for message $i$
End if
```

classifies message i with

$$\hat{y}_k=\begin{cases}1, & \text{if } \hat{\beta}_0+\sum\limits_{m=1}^{M}\hat{\boldsymbol{\beta}}_m\mathbf{H}_m^i>0\\ 0, & \text{otherwise.}\end{cases} \tag{21}$$

The procedure of landmark m and that of Bob are summarized in Algorithms 2 and 3, respectively. Only a small portion of the data, $\{\nabla f(\boldsymbol{\beta})_{\psi_m},\, \psi_m,\, S_m,\, \psi_\rho,\, \nabla f(\boldsymbol{\beta})_{\psi_\rho}\}$ are transmitted in the dFW-based authentication between Bob and the landmarks at each iteration.

VI. IAG-based PHY-layer Authentication

Both the FW-based and dFW-based authentication calculate the gradients of all the training data in each iteration to update $\boldsymbol{\beta}$, and thus result in a huge computation overhead and long delay if the size of the training data is large. We apply the IAG algorithm described in [22] to solve the convex optimization as given by (6) and provide an online PHY-layer authentication method with a reduced computational overhead. More specifically, the objective function $f(\boldsymbol{\beta})$ in (6) can be written as the sum of T component functions

$$f(\boldsymbol{\beta})=-\sum_{k=i-1-T}^{i-1}f_k(\boldsymbol{\beta}), \tag{22}$$

The k-th component function $f_k(\boldsymbol{\beta})$ is given by

$$f_k(\boldsymbol{\beta})=-\left(y_k\left(\beta_0+\boldsymbol{\beta}\mathbf{H}_k\right)-\ln\left(1+e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k}\right)\right), \tag{23}$$

which is clearly convex and continuous differentiable. Thus the coefficient vector $\hat{\boldsymbol{\beta}}$ can be estimated by ML as the following

$$\hat{\boldsymbol{\beta}}=\arg\min_{\boldsymbol{\beta}}\sum_{k=i-1-T}^{i-1}f_k(\boldsymbol{\beta}). \tag{24}$$

In the IAG-based authentication, each component function is evaluated in each iteration. The recent gradients of all the component functions f_k, with $1\leq k\leq T$ are used to approximate the gradient descent direction of the objective function $f(\boldsymbol{\beta})$. As shown in Algorithm 4, each landmark estimates its local RSSI information vector $\mathbf{H}_m^i$ and sends it to Bob to form $\mathbf{H}_i$. The previous T channel vectors, (i.e., $\{\mathbf{H}_k\}_{i-T-1\leq k\leq i-1}$) are used as the training data to solve the optimization problem as given in (24) to authenticate message i according to channel vector $\mathbf{H}_i$.

Let τ_k^n be the gradient sampling time for message k at iteration n, which is given by

$$\tau_k^n=\begin{cases}k, & \text{if } k=\big((n-1)\text{mod } m\big)+1, n>0\\ 0, & \text{if } n=0\\ \tau_k^{n-1}, & \text{else}\end{cases}, \tag{25}$$

with $n=0,1,2,...$ and $1\leq k\leq T$. The aggregated gradient denoted by g as an approximation of the descent at iteration n is given by

$$g=\sum_{k=i-1-T}^{i-1}\nabla f_k\left(\boldsymbol{\beta}^{\tau_k^n}\right). \tag{26}$$

In (25), τ_k^n is the same as the last iteration τ_k^{n-1} if $k\neq\big((n-1)\text{mod } m\big)+1$. In other words, only $\nabla f_k\left(\boldsymbol{\beta}^{\tau_k^n}\right)$ of the message k which satisfies the condition $k=\big((n-1)\text{mod } m\big)+1$ needs to be calculated at iteration n. Compared with FW-based and dFW-based authentication, the computation overhead of IAG-based scheme can be reduced. Therefore, the subsequent iteration point is updated with a constant step size γ as follows

$$\boldsymbol{\beta}\leftarrow\boldsymbol{\beta}-\gamma g. \tag{27}$$

Thus, $\hat{\boldsymbol{\beta}}$ can be estimated via the iterations to calculate $\hat{y}_i$ by (4) in the PHY-layer authentication.

VII. Performance evaluation and discussion

We analyze the performance of the proposed authentication schemes such as the convergence of the dFW-based authentication and compare it with the IAG-based scheme. In order to analyze the convexity of the objective function $f(\boldsymbol{\beta})$ in (6), we denote its Hessian matrix as $\mathbf{M}=[M_{jn}]_{1\leq j,n\leq MN}$, where $M_{jn}=\partial^2 f(\boldsymbol{\beta})/\partial\beta_j\partial\beta_n$.

Algorithm 4 IAG-based PHY-layer Authentication

```
Initialize: $\boldsymbol{\beta} = \mathbf{0}$, $\gamma$
for $k = 1, 2, ..., T$, $\tau_k^0 = 0$.
Channel estimation to form $\mathbf{H}_i$
Calculate $\hat{\beta}_0$ via (11)
For $n$ =0, 1, 2,... do
    Determine $\tau_k^n$ via (25)
    Compute $g$ via (26)
    Update $\boldsymbol{\beta}$ via (27)
    $n \leftarrow n+1$
End for
Obtain $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$
Calculate $\hat{y}_i$ via (4)
If $\hat{y}_i = 1$ then
    Accept message $i$
Else
    Send spoofing alarm for message $i$
End if
```

Theorem 2. *The function $f(\boldsymbol{\beta})$ is convex and the dFW-based authentication in Algorithm 2 and 3 converges if updating $\boldsymbol{\beta}$ with (20).*

Proof. By (6), we have

$$M_{jn} = \begin{cases} \sum\limits_{k=i-T-1}^{i-1} \frac{e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k} H_{kj} H_{kn}}{\left(1+e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k}\right)^2}, & \text{if } j \neq n \\ \sum\limits_{k=i-T-1}^{i-1} \frac{e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k} H_{kj}^2}{\left(1+e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k}\right)^2}, & \text{otherwise.} \end{cases} \tag{28}$$

Let $\mathbf{x} = [x_j]_{1 \leq j \leq MN} \in R^{MN}$ and by (28), we have

$$\mathbf{x}^T \mathbf{M} \mathbf{x} = \sum_{k=i-T-1}^{i-1} \frac{e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k} \left(\mathbf{x}^T \mathbf{H}_k\right)^2}{\left(1+e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k}\right)^2} > 0, \tag{29}$$

indicating that $\mathbf{M}$ is positive semi-definite and thus $f(\boldsymbol{\beta})$ is convex. By Theorems 1 and 2 in [20], $f(\boldsymbol{\beta})$ can be minimized by the dFW-based algorithm given by (20). □

We now prove the convergence of the IAG-based authentication by analyzing the strong convexity of the objective function $f(\boldsymbol{\beta})$ given by (6) and the gradients of the component functions $f_k(\boldsymbol{\beta})$ given by (23).

Theorem 3. *The IAG-based authentication in Algorithm 4 is linearly convergent.*

Proof. By Theorem 2, we see that $f(\boldsymbol{\beta})$ is convex. According to (6), we define $g(\boldsymbol{\beta})$ with $\mu > 0$ as

$$\begin{aligned} g(\boldsymbol{\beta}) &= f(\boldsymbol{\beta}) - \frac{\mu}{2} ||\boldsymbol{\beta}||^2 \\ &= \sum_{k=i-1-T}^{i-1} \left(\ln\left(1+e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k}\right) - y_k\left(\beta_0 + \boldsymbol{\beta}\mathbf{H}_k\right)\right) \\ &\quad - \frac{\mu}{2} ||\boldsymbol{\beta}||^2. \end{aligned} \tag{30}$$

The Hessian matrix of $g(\boldsymbol{\beta})$ is denoted by $\mathbf{G} = [G_{jn}]_{1 \leq j,n \leq MN}$ and given by

$$\begin{aligned} G_{jn} &= \frac{\partial^2 g(\boldsymbol{\beta})}{\partial \beta_j \beta_n} \\ &= \begin{cases} \sum\limits_{k=i-T-1}^{i-1} \frac{e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k} H_{kj} H_{kn}}{\left(1+e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k}\right)^2}, & \text{if } j \neq n \\ -\mu + \sum\limits_{k=i-T-1}^{i-1} \frac{e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k} H_{kj}^2}{\left(1+e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k}\right)^2}, & \text{otherwise.} \end{cases} \end{aligned} \tag{31}$$

Let $\mathbf{x} = [x_j]_{1 \leq j \leq MN}$. By (31), we have

$$\mathbf{x}^T \mathbf{G} \mathbf{x} = \sum_{k=i-T-1}^{i-1} \frac{e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k} \left(\sum\limits_{j=1}^{MN} x_j H_{kj}\right)^2}{\left(1+e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k}\right)^2} - \mu \sum_{j=1}^{MN} x_j^2. \tag{32}$$

According to Appendix A, we see that $\mathbf{G}$ is positive semi-definite and thus $f(\boldsymbol{\beta})$ is strongly convex.

We denote the Lipschitz constant for each component function f_k by $L_k \geq 0$. The standard Euclidean norm is given by

$$||\nabla f_k(\boldsymbol{\beta}_1) - \nabla f_k(\boldsymbol{\beta}_2)|| = \frac{e^{\beta_0} \left|e^{\boldsymbol{\beta}_1\mathbf{H}_k} - e^{\boldsymbol{\beta}_2\mathbf{H}_k}\right| ||\mathbf{H}_k||}{\prod\limits_{j=1}^{2} |1 + e^{\beta_0+\boldsymbol{\beta}_j\mathbf{H}_k}|}. \tag{33}$$

We denote the function $D_k(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ as follows

$$\begin{aligned} D_k(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) &= \frac{||\nabla f_k(\boldsymbol{\beta}_1) - \nabla f_k(\boldsymbol{\beta}_2)||}{||\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2||} \\ &= \frac{e^{\beta_0} \left|e^{\boldsymbol{\beta}_1\mathbf{H}_k} - e^{\boldsymbol{\beta}_2\mathbf{H}_k}\right| ||\mathbf{H}_k||}{\prod\limits_{j=1}^{2} |1 + e^{\beta_0+\boldsymbol{\beta}_j\mathbf{H}_k}| \, ||\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2||}. \end{aligned} \tag{34}$$

By Appendix B, we have $\exists\ L_k \geq 0,\ L_k \geq D_k(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ and thus f_k has Lipschitz continuous gradients. The IAG-based PHY-authentication uses (25)-(27) to update $\boldsymbol{\beta}$. If γ satisfies

$$0 < \gamma < \frac{8\mu}{25\,(T-1) \sum\limits_{k=i-T-1}^{T-1} L_k \left(\mu + \sum\limits_{k=i-T-1}^{T-1} L_k\right)}, \tag{35}$$

and $f(\boldsymbol{\beta})$ is strongly convex and has Lipschitz continuous gradients, the IAG-based algorithm linearly converges according to Theorem 3 in [22]. □

Remark: The PHY-layer authentication schemes based on FW, dFW, and IAG as shown in Algorithms 1-3 are all based on the logistic regression model. In the FW-based and IAG-based scheme, each landmark sends all the local channel vector $\mathbf{H}_m^i$ to the security agent before iteration, and thus produces the same communication overheads. At each iteration in the dFW-based authentication, landmarks first evaluate their local channel information and only send the resulting vector computing $\{\nabla f(\boldsymbol{\beta})_{\psi_m}, \psi_m, S_m\}$ to Bob, and then Bob broadcasts $\{\psi_\rho,\ \nabla f(\boldsymbol{\beta})_{\psi_\rho}\}$ to landmarks. In the multiple MIMO landmark system with a large number of training data, dFW-based authentication can further reduce the communication overhead with the similar detection accuracy to FW-based scheme. Both

FW-based and dFW-based authentication calculated gradients of all the training data at each iteration, whereas only the gradient of single message in the training data set needs to be calculated. Therefore, the IAG-based authentication produce a reduced computation overhead.

Lemma 4. *The dFW-based PHY-authentication achieves the same spoofing detection accuracy after convergence with the FW-based authentication.*

Proof. The FW algorithm and FW algorithm update $\boldsymbol{\beta}$ by (14) and (20), respectively. By (16) and (18), we have

$$|\nabla f(\boldsymbol{\beta})_j| \leq \left|\nabla f(\boldsymbol{\beta})_{\psi_\rho}\right|, \forall j \in [1, MN]. \tag{36}$$

Then by (13), we have $\psi_\rho = \psi$. Therefore, (20) is equivalent to (14).

By (15), (17) and (19), we have

$$\begin{aligned}\chi_{dFW} &= \sum_{m=1}^{M} \sum_{j \in \mathcal{A}_m} \beta_j \nabla f(\boldsymbol{\beta})_j + C|\nabla f(\boldsymbol{\beta})_{\psi_\rho}| \\ &= \sum_{j=1}^{MN} \beta_j \nabla f(\boldsymbol{\beta})_j + C\left|\nabla f(\boldsymbol{\beta})_{\psi_\rho}\right| = \chi_{FW}. \end{aligned} \tag{37}$$

Therefore, the dFW algorithm has the same accuracy with the FW algorithm after convergence. □

The communication overhead of the PHY-authentication assisted by multiple landmarks is denoted by Γ and the computation overhead is denoted by Ψ. Let T be the number of training data and K be the number of the non-zero entry in $\boldsymbol{\beta}$. The authentication overhead can be given by the follows.

Lemma 5. *The communication overhead of the proposed PHY-authentication with M landmarks each with N antennas and T training data is given by*

$$\Gamma = \begin{cases} 2MNT, \text{ IAG or FW} & \text{(38a)} \\ 5Mn + K(M-1)T + MT, \text{ dFW.} & \text{(38b)} \end{cases}$$

Proof. According to Algorithm 1, M landmarks equipped with N antennas send T channel vectors to Bob with communication overhead MNT. Then, M landmarks send $\mathbf{H}_i^m$ to Bob and form channel vector $\mathbf{H}_i$ with communication overhead MN. Thus the total overhead to authenticate T messages is $2MNT$. We can prove the case of IAG similarly.

In each of the n iterations of the dFW-based authentication, each landmark sends information set $\{\nabla f(\boldsymbol{\beta})_{\psi_m}, \psi_m, S_m\}$ to Bob, and Bob broadcasts $\{\psi_\rho, \nabla f(\beta)_{\psi_\rho}\}$ to each landmark, with the communication overhead $5M$. Landmark ρ sends the ψ_ρ-th column of the $T \times MN$ matrix $\mathbf{H}$ to each landmark, and thus totally sends K columns of $\mathbf{H}$ to the other landmarks with communication overhead $K(M-1)T$. Bob then broadcasts $\hat{y}_i$ to each landmark with communication overhead M. Therefore, the total communication overhead is given by (38b). □

Lemma 6. *The total computation overhead of the proposed PHY-authentication with M landmarks each with N antennas and T training data is given by*

$$\Psi = \begin{cases} (T+n)(3MN+4) + MN(n+1)(T+1), \\ \text{IAG} & \text{(39a)} \\ n\big((T+1)(3MN+4) + MNT\big), \\ \text{FW} & \text{(39b)} \\ Mn\big(2T(MN+N+2) + 4MN+3\big), \\ \text{dFW.} & \text{(39c)} \end{cases}$$

Proof. By (12), the IAG-based authentication takes $3MN+4$ runs to compute $\nabla f_k(\boldsymbol{\beta})$. By (26) and (27), the computation overhead at the first time is given by $T(3MN+4)+MN(T-1)+2MN$. In each subsequent iteration, the computation overhead to update $\boldsymbol{\beta}$ is $(3MN+4)+MN(T-1)+2MN$. Therefore, the total computation overhead is given by (39a).

By (12), the FW-based authentication takes $T(3MN+4)+MN(T-1)$ runs to compute $\nabla f(\boldsymbol{\beta})$ and takes $MN-1$ runs to find the largest entry. The overhead to update $\boldsymbol{\beta}$ is $(MN+3)+(3MN+1)$. Thus, the total computation overhead is given by (39b). Similarly, we can prove (39c) for the dFW-based scheme.

□

VIII. Simulation and Experimental Results

A. Simulation Results

Simulations have been performed to evaluate the spoofing detection performance of the proposed PHY-layer authentication systems with M landmarks each equipped with N antennas. The false alarm rate denoted by P_f is defined as the probability of rejecting a legal message, i.e., $P_f = \Pr(\hat{y}_i = 0 | y_i = 1)$, and the miss detection rate denoted by P_m is the probability to accept a spoofing massage, i.e., $P_m = \Pr(\hat{y}_i = 1 | y_i = 0)$. In the simulation, the radio propagation model in (1) was used to generate the channel states and y_k was chosen randomly from $\{0, 1\}$ following the binomial distribution with probability 0.5. We set $A = 20\log_{10}(4\pi d_0/\lambda)$, $\lambda = 0.06$m, $d_0 = 100$m, $\tau = 4, \sigma = 0.2$dB, $M = 6$, $N = 6$, $T = 100$, $C = 10$, and $\epsilon = 0.1$, if not specified otherwise. The locations of the 6 landmarks, Alice, Bob, and Eve are shown in Fig. 2.

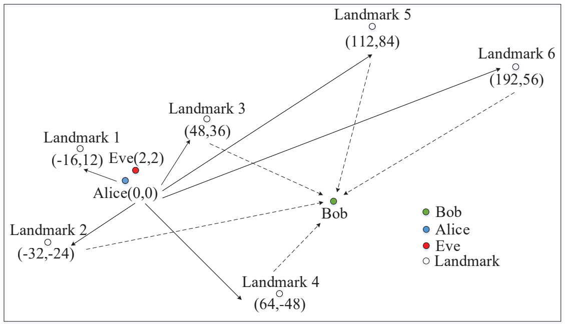
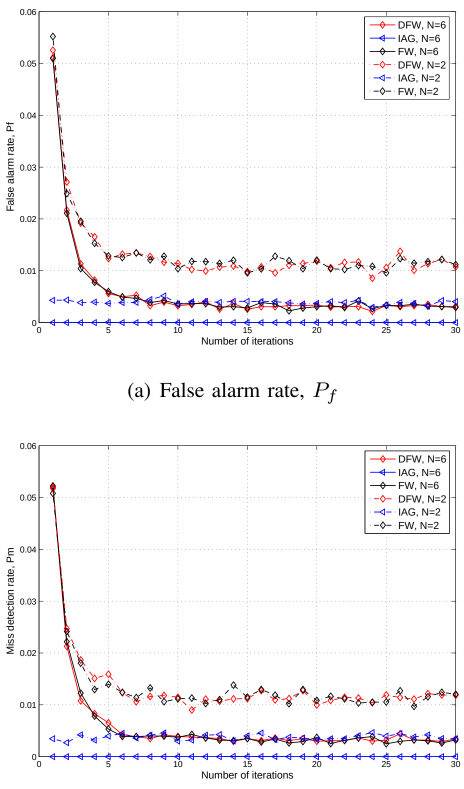


Fig. 2. Network topology of the simulation in meters, in which Alice and Eve are equipped with a single antenna and each of the 6 landmarks has N antennas.

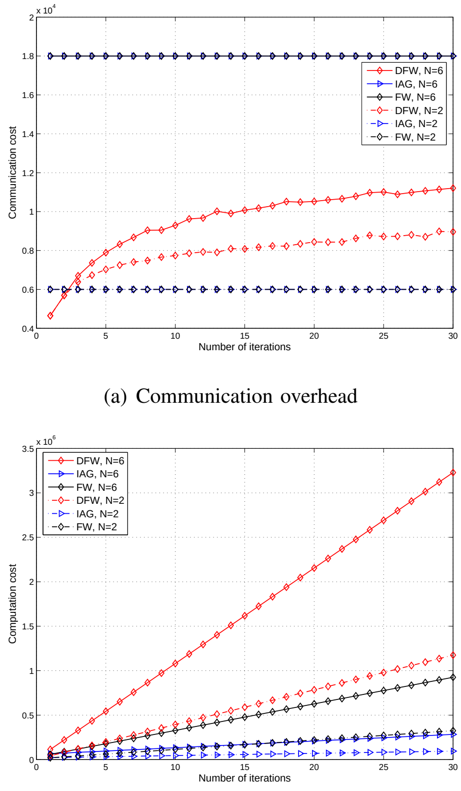
As shown in Fig. 3, both the dFW-based and IAG-based authentication schemes achieve accurate spoofing detection, e.g., the average detection error rate is less than 5% at iteration



(a) False alarm rate, P_f

(b) Miss detection rate, P_m

Fig. 3. Spoofing detection accuracy of the PHY-layer authentications with $M = 6$, $T = 100$, and $C = 10$.



(a) Communication overhead

(b) Computation overhead

Fig. 4. Authentication overhead with $M = 6$, $T = 100$, and $C = 10$.

15. The IAG-based authentication is more accurate than both FW and dFW, e.g., its miss detection rate and the false alarm rate decreases by 67.3% and 68.4% compared with the dFW-based scheme at iteration 20. In Fig. 4(a), the dFW-based authentication significantly reduces the communication overhead especially in the MIMO landmark system, e.g., its overhead is reduced by 44.0% compared with FW-based authentication at iteration 15, if locating 6 landmarks equipped with 2 antennas. In addition, Fig. 4(b) shows that the IAG-based authentication reduces the computation overhead, e.g., the overhead is reduced by 68.3% or 90.9%, respectively, at iteration 25 compared with FW or dFW algorithm, if using 6 landmarks with 6 antennas.

As shown in Fig. 5, the detection accuracy of the authentication decreases with both the number of landmarks and antennas, e.g., P_f and P_m of the dFW-based scheme reduce by 90.2% and 89.2%, respectively, as M changes from 2 to 5, with $N = 6$. In another instance, the error rates of the IAG-based scheme decrease by 88.1% and 89.1%, as N changes from 1 to 6, if $M = 6$. Moreover, the IAG-based scheme decreases the false alarm rate and the miss detection rate especially in the MIMO landmark system, e.g., P_f and P_m reduce by 98.3% and 99.1% compared with FW-based authentication, if $M = 6$ and $N = 6$.

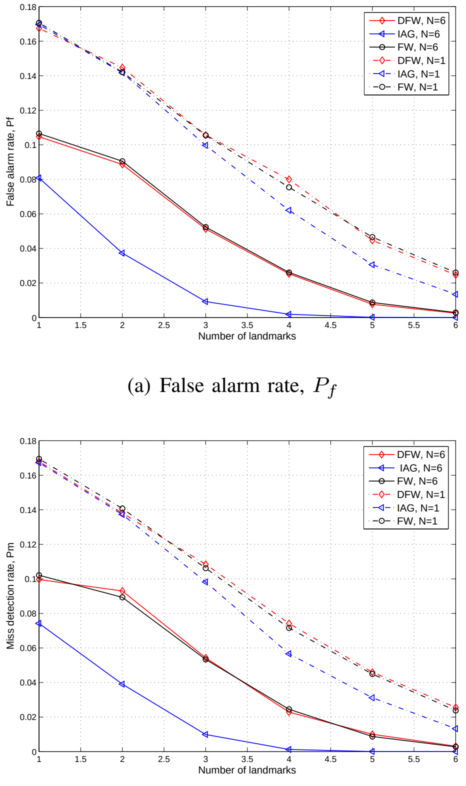
As shown in Fig. 6(a), the dFW-based authentication reduces the communication overhead compared with the FW-based scheme. For example, compared with the FW-based scheme, the communication overhead of dFW-based authentication decreases by 37.4%, if using 6 landmarks with 6 antennas. In Fig. 6(b), the IAG-based authentication produces lower computation overhead with the same communication cost as that of FW algorithm, e.g., the overhead is reduced by 71.3% compared with the FW-based scheme in this case.

As shown in Fig. 7, both the dFW-based PHY-layer authentication and the IAG-based scheme can achieve accurate spoofing detection with a small number of the training data. For instance, the false alarm rate and the miss detection rate of the dFW-based scheme in the spoofing detection is as low as 1.3% and 1.6% with 5 landmarks each with 4 antennas, respectively, if the number of the training data is no less than 50. The IAG-based scheme can also achieve an accurate detection with the false alarm rate and the miss detection rate lower than 0.2% in that case.

B. Experimental Results

Experiments have been performed to evaluate the spoofing detection performance of the proposed schemes. As shown in Fig. 8(a), 10 radio nodes each equipped with laptops and USRPs were placed in a $12 \times 9.5 \times 3$ m^3 office room to act as Alice, Eve, and landmarks. Both Alice and Eve were equipped with one antenna, and each landmark had 2 antennas. The minimum distance between Alice and Eve is 2.3 m.

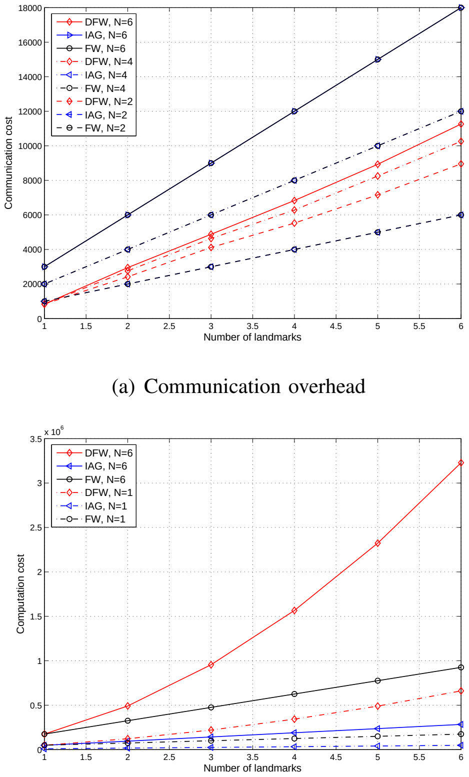
As shown in Fig. 8(b), both the dFW-based and IAG-based authentication schemes are accurate even if Alice is close to



(a) False alarm rate, P_f

(b) Miss detection rate, P_m

Fig. 5. Spoofing detection performance of the PHY-layer authentication system with N antennas at each landmark with $C = 10$, $T = 100$, and $\epsilon = 0.1$.



(a) Communication overhead

(b) Computation overhead

Fig. 6. Overhead of the PHY-layer authentication system with N antennas at each landmark with $C = 10$, $T = 100$, and $\epsilon = 0.1$.

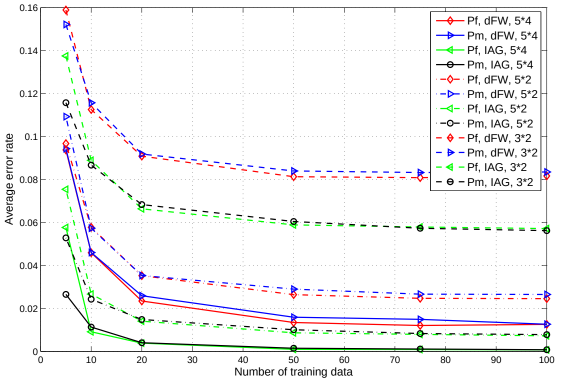


Fig. 7. Spoofing detection accuracy of the PHY-layer authentication for the $M \times N$ system with M landmarks and N antennas to collect T RSSI data with $C = 10$ and $\epsilon = 0.1$.

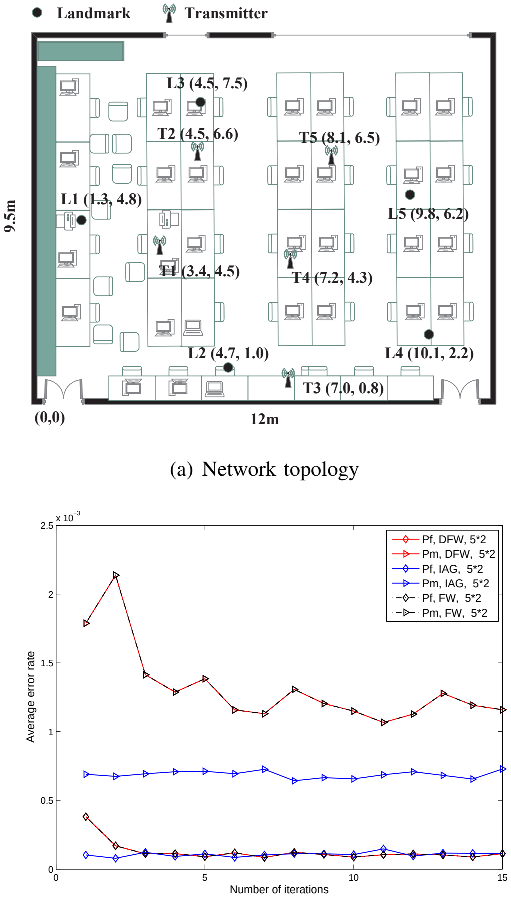
Eve. For instance, the miss detection rate and the false alarm rate of the dFW-based scheme with 10 iterations are 1.1×10^{-3} and 8.8×10^{-5}, respectively. The IAG-based authentication is the most accurate, e.g., the miss detection rate is 31.2% less than the dFW-based scheme at iteration 15, and the false alarm rate is also less than 1.1×10^{-4}.

IX. Conclusion

In this paper, we proposed a PHY-layer authentication system that applies multiple MIMO landmarks to measure the signal RSSI and uses the logistic regression to avoid being restricted to a known radio channel model. In this system, the dFW algorithm is used to estimate the coefficients of the logistic regression model to reduce the communication overhead between the landmarks and the security agent and the ℓ_1-norm type constraint is applied to reduce the computational complexity and avoid over-fitting in the learning process. An IAG-based authentication is proposed to reduce the computation overhead. Simulation and experimental results show that the proposed authentication system can improve the spoofing detection accuracy, e.g., both the false alarm rate and miss detection rate are less than 1% in the MIMO landmark system with FW. Moreover, the communication cost is further reduced by the dFW algorithm, e.g., the communication overhead of the dFW-based authentication with 2 landmarks each with 6 antennas is only 49.1% of the FW-based authentication. The computation overhead is reduced by the IAG algorithm and the authentication performance is significantly improved, e.g., the computation overhead of the IAG-based authentication is reduced by 72.6% and the detection error rate is 57.8% less than the FW-based authentication.

References

[1] Y. Chen, W. Trappe, and R. P. Martin, “Detecting and localizing wireless spoofing attacks,” in *IEEE Commun. Society Conf. Sensor, Mesh and Ad Hoc Commun. and Networks*, pp. 193–202, San Diego, CA, Jun. 2007.



(a) Network topology

(b) Spoofing detection accuracy

Fig. 8. Experimental results in a 12 × 9.5 × 3 m^3 office room, consisting of 5 transmitters each equipped with a single antenna to act as Alice or Eve, and 5 landmarks each with 2 antennas, with $T = 100$, $C = 10$ and $\epsilon = 0.1$.

[2] R. Liu and W. Trappe, *Securing Wireless Communications at the Physical Layer*. New York: Springer-Verlag, 2010.

[3] D. Shan, K. Zeng, W. Xiang, P. Richardson, and Y. Dong, "PHY-CRAM: Physical layer challenge-response authentication mechanism for wireless networks," *IEEE J. Sel. Areas in Commun.*, vol. 31, no. 9, pp. 1817–1827, Aug. 2013.

[4] H. Wen, S. Li, X. Zhu, and L. Zhou, "A framework of the PHY-layer approach to defense against security threats in cognitive radio networks," *IEEE Network*, vol. 27, no. 3, pp. 34–39, May, 2013.

[5] H. Alipour, Y. B. Al-Nashif, P. Satam, and S. Hariri, "Wireless anomaly detection based on IEEE 802.11 behavior analysis," *IEEE Trans. Information Forensics and Security*, vol. 10, no. 10, pp. 2158 – 2170, Oct. 2015.

[6] A. Ferrante, N. Laurenti, C. Masiero, and M. Pavon, "On the error region for channel estimation-based physical layer authentication over rayleigh fading," *IEEE Trans. Information Forensics and Security*, vol. 10, no. 5, pp. 941–952, May 2015.

[7] V. Bhargava and M. L. Sichitiu, "Physical security perimeters for wireless local area networks," *Int'l J. Network Security*, vol. 3, no. 1, pp. 73–84, Jul. 2006.

[8] Y. Chen, J. Yang, W. Trappe, and R. P. Martin, "Detecting and localizing identity-based attacks in wireless and sensor networks," *IEEE Trans. Vehicular Technology*, vol. 59, no. 5, pp. 2418–2434, Jun. 2010.

[9] C. Pei, N. Zhang, X. S. Shen, and J. W. Mark, "Channel-based physical layer authentication," in *Proc. IEEE Global Commun. Conf (GLOBECOM)*, pp. 4114–4119, Austin, TX, Dec. 2014.

[10] J. Yang, Y. Chen, W. Trappe, and J. Cheng, "Detection and localization of multiple spoofing attackers in wireless networks," *IEEE Trans. Parallel and Distributed systems*, vol. 24, no. 1, pp. 44–58, Jan. 2013.

[11] X. Wu and Z. Yang, "Physical-layer authentication for multi-carrier transmission," *IEEE Commun. Letters*, vol. 19, no. 1, pp. 74–77, Jan. 2015.

[12] J. Liu and X. Wang, "Physical layer authentication enhancement using two-dimensional channel quantization," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4171 – 4182, Jun. 2016.

[13] Y. Zou, X. Wang, and W. Shen, "Optimal relay selection for physical-layer security in cooperative wireless networks," *IEEE J. Sel. Areas in Commun.*, vol. 31, no. 10, pp. 2099–2111, Oct. 2013.

[14] J. K. Tugnait, "Wireless user authentication via comparison of power spectral densities," *IEEE J. Sel. Areas in Commun.*, vol. 31, no. 9, pp. 1791–1802, Sep. 2013.

[15] H. Liu, Y. Wang, J. Liu, J. Yang, and Y. Chen, "Practical user authentication leveraging channel state information (CSI)," in *Proc. ACM Symp. Inform., Computer and Commun. Security*, pp. 389–400, Kyoto, Japan, Jun. 2014.

[16] T. Y. Liu, P. H. Lin, S. C. Lin, and Y. W. P. Hong, "To avoid or not to avoid CSI leakage in physical layer secret communication systems," *IEEE Commun. Magazine*, vol. 53, no. 12, pp. 19–25, Dec. 2015.

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

[18] X. Wan, L. Xiao, Q. Li, and Z. Han, "PHY-layer authentication with multiple landmarks with reduced communication overhead," in *Proc. IEEE Int'l Conf. Commun. (ICC)*, Paris, France, May 2017.

[19] M. Jaggi, "Revisiting Frank-Wolfe: Projection-free sparse convex optimization," in *Proc. Int'l Conf. Machine Learning (ICML)*, pp. 427–435, Atlanta, GA, Jun. 2013.

[20] A. Bellet, Y. Liang, A. B. Garakani, M.-F. Balcan, and F. Sha, "A distributed Frank-Wolfe algorithm for communication-efficient sparse learning," in *Proc. SIAM Int'l Conf. Data Mining*, pp. 478–486, Vancouver, Canada, Apr. 2015.

[21] D. Blatt, A. O. Hero, and H. Gauchman, "A convergent incremental gradient method with a constant step size," *SIAM J. Optim.*, vol. 18, no. 1, pp. 29–51, Feb. 2007.

[22] M. Gurbuzbalaban, A. Ozdaglar, and P. Parrilo, "On the convergence rate of incremental aggregated gradient algorithms," *arXiv preprint arXiv:1506.02081*, Jun. 2015.

[23] H. Wen, P.-H. Ho, C. Qi, and G. Gong, "Physical layer assisted authentication for distributed ad hoc wireless sensor networks," *IET information security*, vol. 4, no. 4, pp. 390–396, Dec. 2010.

[24] X. Du, D. Shan, K. Zeng, and L. Huie, "Physical layer challenge-response authentication in wireless networks with relay," in *Proc. IEEE INFOCOM*, pp. 1276–1284, Toronto, Canada, May 2014.

[25] W. Hou, X. Wang, J. Y. Chouinard, and A. Refaey, "Physical layer authentication for mobile systems with time-varying carrier frequency offsets," *IEEE Trans. Commun.*, vol. 62, no. 5, pp. 1658–1667, May 2014.

[26] P. L. Yu, G. Verma, and B. M. Sadler, "Wireless physical layer authentication via fingerprint embedding," *IEEE Commun. Magazine*, vol. 53, no. 6, pp. 48–53, Jun. 2015.

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

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

[29] G. Caparra, M. Centenaro, N. Laurenti, S. Tomasin, and L. Vangelista, "Energy-based anchor node selection for IoT physical layer authentication," in *IEEE Int'l Conf. Commun. (ICC)*, pp. 1–6, Kuala Lumpur, Malaysia, May 2016.

[30] M. Dong, K. Ota, H. Li, S. Du, H. Zhu, and S. Guo, "Rendezvous: towards fast event detecting in wireless sensor and actor networks," *Computing*, vol. 96, no. 10, pp. 995–1010, Oct. 2014.

[31] M. Ozay, I. Esnaola, F. T. Yarman Vural, S. R. Kulkarni, and H. V. Poor, "Machine learning methods for attack detection in the smart grid," *IEEE Trans. Neural Networks and Learning Systems*, vol. 27, no. 8, pp. 1773–1786, Aug. 2015.

[32] M. Esmalifalak, L. Liu, N. Nguyen, and R. Zheng, "Detecting stealthy false data injection using machine learning in smart grid," *IEEE Systems J*, vol. 283, no. 1, pp. 808 – 813, Aug. 2014.

[33] L. Xiao, Y. Li, G. Han, G. Liu, and W. Zhuang, "PHY-layer spoofing detection with reinforcement learning in wireless networks," *IEEE Trans. Vehicular Technology*, vol. 65, no. 12, pp. 10037–10047, Dec. 2016.

[34] L. Xiao, T. Chen, G. Han, W. Zhuang, L. Sun, “Game theoretic study on channel-based authentication in MIMO systems,” *IEEE Trans. Vehicular Technology*, pp. 1–11, Jan. 2017, in press.

[35] L. Xiao, L. J. Greenstein, N. B. Mandayam, and W. Trappe, “Using the physical layer for wireless authentication in time-variant channels,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2571–2579, Jul. 2008.

[36] A. Goldsmith, *Wireless communication*. New York: Cambridge university press, 2005.

[37] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer-Verlag, 2009.

[38] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

Appendix A

Let $\mathbf{x} \in R^{MN}$ be a non-zero MN-dimension column vector. If $\forall x_j x_n > 0$ with $1 \leq j \neq n \leq MN$ and $\mu \leq \sum_{k=i-T-1}^{i-1} \frac{e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k} \min\limits_{1\leq j\leq MN}\{H_{kj}^2\}}{\left(1+e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k}\right)^2}$ holds, we have

$$\begin{aligned}\mu &\leq \sum_{k=i-T-1}^{i-1} \frac{e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k} \min\limits_{1\leq j\leq MN}\{H_{kj}^2\}\,||\mathbf{x}||^2}{\left(1+e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k}\right)^2||\mathbf{x}||^2} \\ &\leq \sum_{k=i-T-1}^{i-1} \frac{e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k} \sum\limits_{j=1}^{MN} x_j^2 H_{kj}^2}{\left(1+e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k}\right)^2||\mathbf{x}||^2} \\ &\leq \sum_{k=i-T-1}^{i-1} \frac{e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k} \left(\mathbf{x}^T\mathbf{H}_k\right)^2}{\left(1+e^{\beta_0+\boldsymbol{\beta}\mathbf{H}_k}\right)^2||\mathbf{x}||^2}. \end{aligned} \tag{40}$$

Thus, we have $\exists\ \mu > 0,\ \mathbf{x}^T\mathbf{G}\mathbf{x} \geq 0$, indicating that $\mathbf{G}$ is positive semi-definite.

Appendix B

As $MN = 1$, the 2nd-order Taylor expansion of $e^{\beta H_k}$ at $\beta H_k = 0$ is given by

$$e^{\beta H_k} = 1 + \beta H_k + \frac{\beta^2 H_k^2}{2} + O(\beta^2 H_k^2). \tag{41}$$

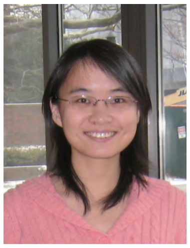
By (34) and (41), we have

$$\begin{aligned} D_k(\beta_1,\beta_2) &= \frac{e^{\beta_0}||H_k||^2||\beta_1-\beta_2||\left|2+\sum\limits_{j=1}^{2}\beta_j H_k\right|}{2\prod\limits_{j=1}^{2}|1+e^{\beta_0+\beta_j H_k}|\,||\beta_1-\beta_2||} \\ &\leq \frac{||H_k||^2\left|2+\sum\limits_{j=1}^{2}\beta_j H_k\right|}{2\left|e^{\beta_0+\sum\limits_{j=1}^{2}\beta_j H_k}\right|}. \end{aligned} \tag{42}$$

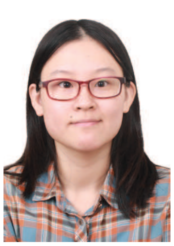
If $\sum_{j=1}^{2}\beta_j H_k > 0$ and $L_k \geq \frac{3||H_k||^2}{2e^{\beta_0}}$, we have by (42)

$$\begin{aligned} L_k &\geq \frac{3||H_k||^2}{2e^{\beta_0}} \\ &\geq \frac{||H_k||^2}{\left|e^{\beta_0+\sum\limits_{j=1}^{2}\beta_j H_k}\right|} + \frac{||H_k||^2\left|\sum\limits_{j=1}^{2}\beta_j H_k\right|}{2\left|e^{\beta_0+\sum\limits_{j=1}^{2}\beta_j H_k}\right|} \geq D_k(\beta_1,\beta_2). \end{aligned} \tag{43}$$

Similarly, $\exists L_k \geq 0$, we can prove $L_k \geq D_k(\beta_1,\beta_2)$.



Liang Xiao (M’09, SM’13) is currently a Professor in the Department of Communication Engineering, Xiamen University, Fujian, China. She has served as an associate editor of IEEE Trans. Information Forensics and Security and guest editor of IEEE Journal of Selected Topics in Signal Processing. She is the recipient of the best paper award for 2016 INFOCOM Big Security WS and 2017 ICC. She received the B.S. degree in communication engineering from Nanjing University of Posts and Telecommunications, China, in 2000, the M.S. degree in electrical engineering from Tsinghua University, China, in 2003, and the Ph.D. degree in electrical engineering from Rutgers University, NJ, in 2009. She was a visiting professor with Princeton University, Virginia Tech, and University of Maryland, College Park.



Xiaoyue Wan (S’16) received the B.S. degree in communication engineering from Xiamen University, Xiamen, China, in 2016. She is currently pursuing the M.S. degree with the Department of Communication Engineering, Xiamen University, Xiamen, China.

Zhu Han (S’01 – M’04-SM’09-F’14) received the B.S. degree in electronic engineering from Tsinghua University, in 1997, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, in 1999 and 2003, respectively.

From 2000 to 2002, he was an R&D Engineer of JDSU, Germantown, Maryland. From 2003 to 2006, he was a Research Associate at the University of Maryland. From 2006 to 2008, he was an assistant professor at Boise State University, Idaho. Currently, he is a Professor in the Electrical and Computer Engineering Department as well as in the Computer Science Department at the University of Houston, Texas. His research interests include wireless resource allocation and management, wireless communications and networking, game theory, big data analysis, security, and smart grid. Dr. Han received an NSF Career Award in 2010, the Fred W. Ellersick Prize of the IEEE Communication Society in 2011, the EURASIP Best Paper Award for the Journal on Advances in Signal Processing in 2015, IEEE Leonard G. Abraham Prize in the field of Communications Systems (best paper award in IEEE JSAC) in 2016, and several best paper awards in IEEE conferences. Currently, Dr. Han is an IEEE Communications Society Distinguished Lecturer. Dr. Han is 1% highly cited researcher 2017 according to Web of Science.