ISBN: 978-981-18-1791-5
2021 the 11th International Workshop on Computer Science and Engineering (WCSE 2021)
doi: 10.18178/wcse.2021.06.047

Pilot Contamination Attack Detection Method Based on OFDM Pilot Signal

Jianqiu Xiong [1], Chen Xie [2] and Qiao Hu [1+]

[1] College of Computer Science and Electronic Engineering, Hunan University, China
[2] Yueyang Public Security Bureau, China

Abstract. Pilot contamination (PC) attack is a well-known active eavesdropping attack. Because of the public transmission protocol, frame structure and pilot sequence set, the eavesdropper (Eve) can send the same pilot as the legitimate user (Bob) to actively affect the channel estimation process at the base station(BS). This results in the contamination of the channel state information (CSI) measured at BS, the alteration of the legitimate beamformer design and information leakage. In this paper, we use the statistical characteristics of the channel to detect whether there is a pilot contamination attack. The secure region is deduced theoretically, which makes the detection of pilot contamination attack more accurate.

Keywords: Physical layer security, pilot contamination attack, OFDM.

1. Introduction

Due to the open nature of wireless medium, wireless communications are especially vulnerable to eavesdropping attacks, such as PC attack[1]-[3]. PC attack is the activity of eavesdropper launching attack in channel training phase. The purpose of eavesdropping is achieved by affecting the channel estimation at the BS. In time division duplex (TDD) transmission mode, the uplink and downlink are considered to be mutual. And Eve has a complete understanding of the pilot sequence set and transmission time of Bob. Therefore, Eve can send the same pilot signal, causing channel estimation deviation at the BS. This will not only reduce the quality of the received signal at Bob, but also lead to the leakage of information to Eve. When Eve's signal power is high, Bob's signal is masked. Since Eve can act like Bob, it's hard to detect. To prevent this situation, it is crucial to design appropriate detection methods.

The issue of PC attack is first noticed in [1], and the authors study how to improve the performance of eavesdropper. Another paper[4] discusses how to detect attacks through random training methods related to massive multiple-input multiple-output (MIMO) systems, but the authors do not theoretically derive the decision threshold. Then Wang et.al [5] improve previous method [4], uses random symbols to detect the existence of attacks, and gives the theoretical security region. However, the method in [5] has strict requirements for the modulation mode of the signal, and can not be applied to all modulation modes. A subspace-based method is considered in [6]. The idea is to compare the difference between the largest and second largest eigenvalues of the covariance matrix of the received random pilot to detect whether there is an attack. However, its performance depends on the knowledge of channel and noise covariance matrices.

In this paper, we propose an attack detection scheme based on OFDM pilot signal. The scheme uses the correlation coefficient of CSI at pilot and CSI at subcarrier to detect whether there is an attack. The statistical characteristics of received signals with and without attacks are studied in details in Section III. On this basis, we analyze the distribution of decision statistics. According to the distribution of decision statistics, the security region is constructed to obtain accurate detection performance.

Notation: Superscripts $(.)^*$ represent complex conjugate. $\Re e(.)$ and $\Im m(.)$ are the operations that extract the real and imaginary parts of a complex value. The notation $x \sim N_2(\mu, \Sigma)$ denotes a random variable x that is bivariate normal with mean μ and covariance Σ.[1]

[+] Corresponding author.
E-mail address: huqiao@hnu.edu.cn.

2. System Model

The OFDM system adopts similar OFDM signal technology in IEEE 802.11. For each OFDM symbol, a certain number of specified subcarriers are allocated to transmit pilot signals, while the other subcarriers carry data signals.

In order to launch PC attack in OFDM system, Eve tries to interfere with the communication between Bob and BS. Suppose a simple flat channel fading model and assume that the Eve has a complete understanding of the training sequence of Bob and the exact time of signal transmitting. At the same time, it is assumed that the channel is quasi-static, that is, the channel response remains constant during the transmission of each OFDM signal. Since the eavesdropper knows the public pilot signal and the exact time of transmission, the eavesdropper can also send the same pilot signal. Then we analyze these two cases in details below.

2.1. Without Attack

Here we analyze the case that there exists no PC attack. The signal received by BS in the *k*-th subchannel of OFDM is $r_k = h_B s_k + \eta_k$, where s_k is the signal transmitted by Bob, h_B is the complex Gaussian channel response of the legitimate user to the BS, with zero mean and unit variance, η_k is additive white Gaussian noise, which follows the distribution of zero mean and variance of N_0. From [7], we can get that when SNR equals to 20, the upper limit of symbol error rate is about 0.3%. Therefore, we can assume that the base station can correctly demodulate all signals. For non-pilot parts of the signals, the received signal is divided by the transmitting signal as $r_k / s_k = h_B + \eta_k / s_k$, while at the pilot subcarrier, it is $r_p / s_p = h_B + \eta_p / s_p$. We can recover the transmitting signal from the data acquired by the demodulation process. Let z_k be the relationship coefficient between the CSI at the pilot and the CSI at the non-pilot, then it can be expressed as

$$z_k = \left(\frac{r_p}{s_p}\right)^* * \frac{r_k}{s_k} = h_B^* h_B + h_B^* \frac{\eta_k s_k^*}{\| s_k \|^2} + h_B \frac{\eta_p^* s_p}{\| s_p \|^2} + \frac{\eta_p^* s_p \eta_k s_k^*}{\| s_p \|^2 \| s_k \|^2} \tag{1}$$

Therefore, z_k can be regarded as a function of the channel. It can be seen from the above formula that the first term of z_k is a real number about the channel response between the legitimate user and the BS, and the other three terms can be regarded as noise terms. Since the channel response follows the complex Gaussian distribution, then $h_B^* h_B$ is a random variable that follows the non-central chi-square distribution with 2 degrees of freedom, and the mean is 1 and the variance is 1.

2.2. With Attack

In the *k*-th channel of OFDM, Bob sends a random symbol s_k, Eve randomly selects a symbol sa_k from the modulation constellation set and sends it to BS. Therefore, the signal received by BS in the *k*-th channel is $r_k = \alpha h_E sa_k + h_B s_k + \eta_k$, h_E is the complex Gaussian channel response from eavesdropper to BS, which also has zero mean and unit variance, and α represents the power enhancement of the attack signal. Due to pilot contamination, the BS may not be able to demodulate the signal correctly. When there is only one eavesdropper, s_x is the signal generated from the result of the demodulation process at BS, z_k is expressed as

$$\begin{aligned} z_k = \left(\frac{r_p^{'}}{s_p}\right)^* * \frac{r_k}{s_x} &= \frac{\alpha^2 h_E^* h_E sa_k s_x^*}{\| s_x \|^2} + \frac{h_B^* h_B s_k s_x^*}{\| s_x \|^2} + \frac{\alpha h_E^* h_B s_k s_x^*}{\| s_x \|^2} + \frac{\alpha h_B^* h_E sa_k s_x^*}{\| s_x \|^2} \\ &+ \frac{\alpha h_E^* \eta_k s_x^*}{\| s_x \|^2} + \frac{h_B^* \eta_k s_x^*}{\| s_x \|^2} + \frac{\alpha h_E sa_k s_x^* \eta_p^* s_p}{\| s_x \|^2 \| s_p \|^2} + \frac{h_B s_k s_x^* \eta_p^* s_p}{\| s_x \|^2 \| s_p \|^2} + \frac{\eta_k s_x^* \eta_p^* s_p}{\| s_x \|^2 \| s_p \|^2} \end{aligned} \tag{2}$$

Obviously, when $sa_k = s_k$, s_x is equal to s_k. In this case, z_k is expressed as

$$\begin{aligned} z_k &= \alpha^2 h_E^* h_E + h_B^* h_B + \alpha h_E^* h_B + \alpha h_B^* h_E \\ &+ \frac{\alpha h_E^* \eta_k s_x^*}{\| s_x \|^2} + \frac{h_B^* \eta_k s_x^*}{\| s_x \|^2} + \frac{\alpha h_E \eta_p^* s_p}{\| s_p \|^2} + \frac{h_B \eta_p^* s_p}{\| s_p \|^2} + \frac{\eta_k s_x^* \eta_p^* s_p}{\| s_x \|^2 \| s_p \|^2} \end{aligned} \tag{3}$$

From the above formula, the first two terms of z_k are random variables that obey the non-central chi-square distribution with 2 degrees of freedom, and the means are α^2 and 1, and the variances are α^4 and 1, respectively.

3. Detection Scheme

The previous section gives the expression of the relationship coefficient z_k between the CSI at the pilot and the CSI at the non-pilot of a OFDM symbol with or without attack. Inspired by [5], we can use the statistical characteristics of the relationship coefficient z_k to detect whether there is a PC attack. The detection flow chart is shown in Figure 1. For multiple received OFDM symbols, since the channel response of each symbol is different, but they all follow the complex Gaussian distribution. We can select M subcarriers with the same demodulation result from multiple OFDM signals to calculate the correlation coefficient of CSI at the subcarrier signal and the corresponding pilot signal. Therefore, in the case of no attack, the decision statistic Z can be expressed as

$$\begin{aligned} Z &= \frac{1}{M}\sum_{i=1}^{M} z_i = \frac{1}{M}\sum_{i=1}^{M}\left(\frac{r_{i,p}}{s_p}\right)^* * \frac{r_i}{s_x} \\ &= \frac{1}{M}\sum_{i=1}^{M}\left(h_{i,B}^* h_{i,B} + h_{i,B}^* \frac{\eta_i s_i^*}{\| s_x \|^2} + h_{i,B}\frac{\eta_{i,p}^* s_p}{\| s_p \|^2} + \frac{\eta_{i,p}^* s_p \eta_i s_i^*}{\| s_p \|^2 \| s_x \|^2}\right) \end{aligned} \tag{4}$$

Where $h_{i,B}$ represents the channel response of the i-th OFDM signal, $r_i = h_{i,B}s_i + \eta_i$ represents the received signal in the i-th OFDM signal, and s_i represents the signal transmitted by Bob. $r_{i,p} = h_{i,B}s_p + \eta_{i,p}$ represents the signal received at the pilot subcarrier in the i-th OFDM signal. η_i and $\eta_{i,p}$ represent the noise at the subcarrier corresponding to s_x and s_p in the i-th OFDM signal, respectively. Assuming there is no attacker, the BS can demodulate the signal correctly, then $\forall i \in [1, M]$ $s_i = s_x$.

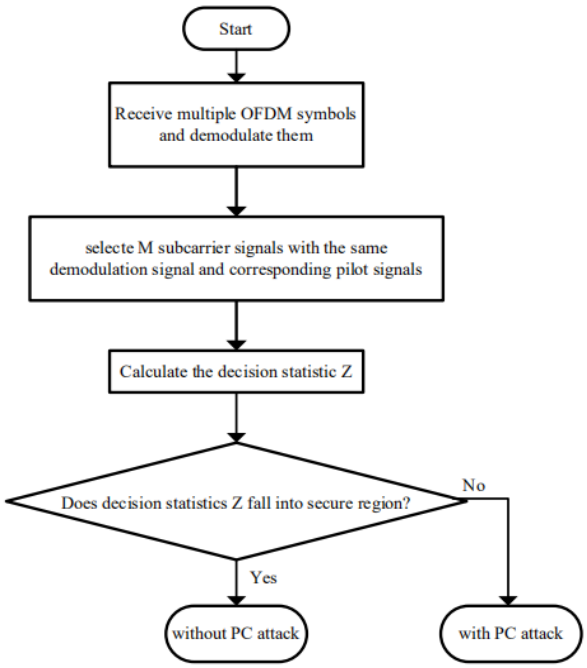


Fig. 1: Proposed PC attack detection flow chart

$\sum_{i=1}^{M} h_{i,B}^* h_{i,B}$ is a real number following the chi-square distribution of $2M$ degrees of freedom, i.e. χ_{2M}^2. Therefore, when M becomes larger, the first term of equation 4 follows the real valued normal distribution of $N(1, 1/M)$ according to the central limit theorem. For a given $h_{i,B}$, the last three terms are the sum of M complex-valued normal product Gaussian variables. According to Lyapunov central limit theorem, when $M \to \infty$, the sum is a complex Gaussian random variable with mean 0 and variance σ_M^2. Therefore, the overall distribution of Z can be expressed by the bivariate normal density function $Z \sim N_2(\mu, \Sigma)$, where

$$\mu = \begin{bmatrix} \Re\mathrm{e}(Z) \\ \Im m(Z) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \Sigma = \begin{pmatrix} \frac{1}{M} + \sigma_M^2 & 0 \\ 0 & \sigma_M^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{M} + \frac{N_0^2 + N_0 \| s_p \|^2 + N_0 \| s_x \|^2}{2M \| s_x \|^2 \| s_p \|^2} & 0 \\ 0 & \frac{N_0^2 + N_0 \| s_p \|^2 + N_0 \| s_x \|^2}{2M \| s_x \|^2 \| s_p \|^2} \end{pmatrix} \tag{5}$$

This distribution is used to determine the normal distribution of decision statistic Z which representing a secure transmission.

When there is an attack, we perform the same actions. And we assume that at the selected subcarrier, the data transmitted by the attacker is equal to the data transmitted by the legitimate user, i.e. $\forall i \in [1, M], s_i = sa_i = s_x$

$$
\begin{aligned}
Z &= \frac{1}{M}\sum_{i=1}^{M}\left(\frac{r_{i,p}^{'}}{s_p}\right)^{*} * \frac{r_i}{s_x} \\
&= \frac{1}{M}\sum_{i=1}^{M}\Big(\frac{\alpha^2 h_{i,E}^{*} h_{i,E} sa_i s_x^{*} + h_{i,B}^{*} h_{i,B} s_i s_x^{*} + \alpha h_{i,E}^{*} h_{i,B} s_i s_x^{*} + \alpha h_{i,B}^{*} h_{i,E} sa_i s_x^{*} + \alpha h_{i,E}^{*} \eta_i s_x^{*} + h_{i,B}^{*} \eta_i s_x^{*}}{\| s_x \|^2} \\
&\qquad + \frac{\alpha h_{i,E} sa_i s_x^{*} \eta_{i,p}^{*} s_p + h_{i,B} s_i s_x^{*} \eta_{i,p}^{*} s_p + \eta_i s_x^{*} \eta_{i,p}^{*} s_p}{\| s_x \|^2 \| s_p \|^2}\Big) \\
&= \frac{1}{M}\sum_{i=1}^{M}\Big(\alpha^2 h_{i,E}^{*} h_{i,E} + h_{i,B}^{*} h_{i,B} + \alpha h_{i,E}^{*} h_{i,B} + \alpha h_{i,B}^{*} h_{i,E} + \frac{\alpha h_{i,E}^{*} \eta_i s_x^{*} + h_{i,B}^{*} \eta_i s_x^{*}}{\| s_x \|^2} + \frac{\alpha h_{i,E} \eta_{i,p}^{*} s_p + h_{i,B} \eta_{i,p}^{*} s_p}{\| \| s_p \|^2} + \frac{\eta_i s_x^{*} \eta_{i,p}^{*} s_p}{\| s_x \|^2 \| s_p \|^2}\Big)
\end{aligned}
\tag{6}
$$

Where $r_i = \alpha h_{i,E} sa_i + h_{i,B} s_i + \eta_i$ is the subcarrier in the *i*-th OFDM signal, s_i and sa_i are the signals transmitted by Bob and Eve respectively. $r_{i,p}^{'} = \alpha h_{i,E} + h_{i,B} s_p + \eta_{i,p}$ is the signal received in the pilot subcarrier of the *i*-th OFDM signal. In this case, the overall distribution of $Z^{'}$ is expressed as $Z^{'} \sim N_2(\mu^{'}, \Sigma^{'})$, where

$$
\mu^{'} = \begin{bmatrix} \alpha^2 + 1 \\ 0 \end{bmatrix} \qquad \Sigma^{'} = \begin{pmatrix} \frac{\alpha^4 + \alpha^2 + 1}{M} + \frac{N_0^2 + N_0(\alpha^2 + 1)(\| s_p \|^2 + \| s_x \|^2)}{2M \| s_x \|^2 \| s_p \|^2} & 0 \\ 0 & \frac{\alpha^2}{M} + \frac{N_0^2 + N_0(\alpha^2 + 1)(\| s_p \|^2 + \| s_x \|^2)}{2M \| s_x \|^2 \| s_p \|^2} \end{pmatrix}
\tag{7}
$$

It is known from the above that the decision statistic Z follows the bivariate normal distribution. And we know the mean and variance of Z without attack. Therefore, the confidence interval of 99.5% can be calculated according to the cumulative probability density function of two-dimensional Gaussian distribution. The confidence interval of 99.5% is regarded as the detection area. If it is within this area, it is considered that there is no attack. If it is not in this area, it can be considered that there is an attack.

4. Simulations

In the case of different *M* and the signal power ratio of Eve and Bob is 1, figure 2 illustrates the detection probability of the proposed PC attack detection scheme relative to SNR, and compares it with the method in [5]. The method in [5], referred as “Random Symbol”. Both methods take the same assumption that the channel is quasi-static, and exclude the energy examination stage in [5]. It can be seen from the Figure 1 that the detection rate of PC attack increases with the increase of SNR. In the case of low SNR, the proposed method can also detect PC attack. Compared with “Random Symbol”, our method is significantly outperforms. In the case of lower SNR, our method has higher detection probability and requires less data. And with the increase of M, the robustness to noise is better.

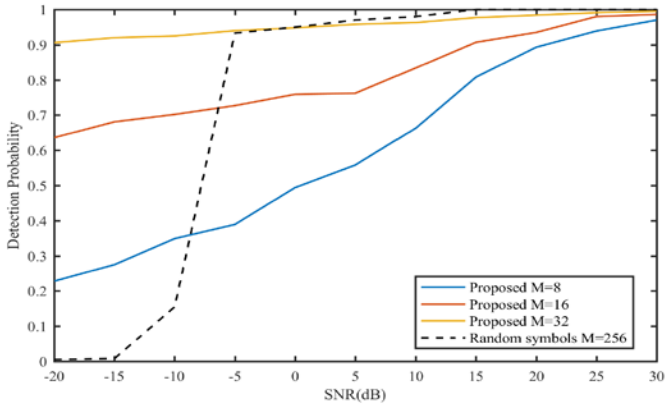
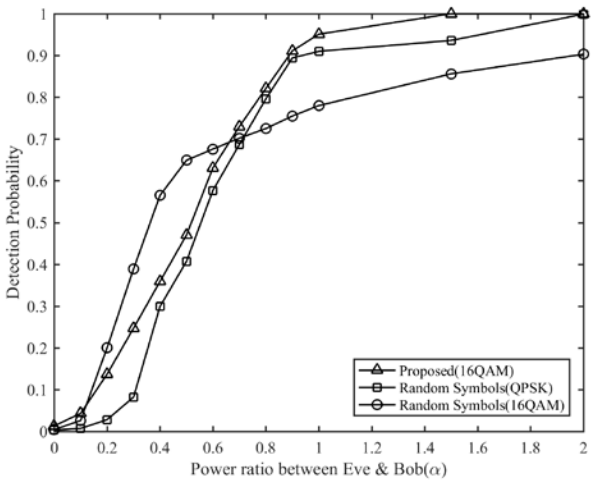
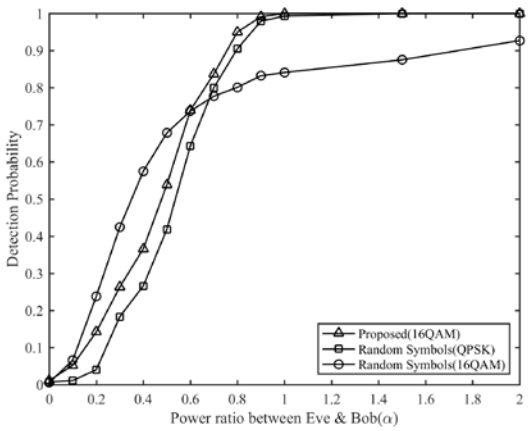


Fig. 2: Detection probability vs. SNR

Figure 3 shows the detection probability of our PC attack detection scheme relative to the power ratio between the eavesdropper signal and the legitimate user signal under the same SNR(SNR=10). It can be seen from the figure that the higher the power of the eavesdropper signal is, the higher the probability of detection will be. Under the same SNR, the performance of our method is similar to "Random Symbol" which signal modulation is QPSK. It is proved that our method is feasible. However, under the same modulation mode, the eavesdropper's signal power $\alpha \le 0.8$, the performance of the proposed method is similar to that of "Random Symbol". When $\alpha > 0.8$, the performance of proposed method is better than that of "Random Symbol".



(a) M=32



(b) M=64

Fig. 3: Detection probability vs. Different power ratios

5. Conclusion

To combat pilot contamination attack, this paper proposes a detection scheme based on OFDM pilot. The decision statistics are constructed by using the channel relationship coefficients between the pilot and non-pilot parts in OFDM signal. According to the bivariate normal distribution of decision statistics, the detection region is proposed. Simulation results show that the performance of the proposed method is better than that of the reference method at low SNR.

6. Reference

[1] Zhou, X., Maham, B., & Hjørungnes, A. (2012). Pilot Contamination for Active Eavesdropping. *IEEE Transactions on Wireless Communications, 11*, 903-907.

[2] Kapetanovic, D., Al-Nahari, A., Stojanovic, A., & Rusek, F. (2014). Detection of active eavesdroppers in massive

MIMO. *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, 585-589.

[3] Xiong, Q., Liang, Y., Li, K., & Gong, Y. (2015). An Energy-Ratio-Based Approach for Detecting Pilot Spoofing Attack in Multiple-Antenna Systems. *IEEE Transactions on Information Forensics and Security, 10*, 932-940.

[4] Kapetanovic, D., Zheng, G., Wong, K., & Ottersten, B. (2013). Detection of pilot contamination attack using random training and massive MIMO. *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 13-18.

[5] Wang, X., Liu, M., Wang, D., & Zhong, C. (2017). Pilot Contamination Attack Detection Using Random Symbols for Massive MIMO Systems. *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, 1-7.

[6] Kapetanovic, D., Zheng, G., & Rusek, F. (2015). Physical layer security for massive MIMO: An overview on passive eavesdropping and active attacks. *IEEE Communications Magazine, 53*, 21-27.

[7] Baxley, R. et al. "Complex Gaussian Ratio Distribution with Applications for Error Rate Calculation in Fading Channels with Imperfect CSI." *2010 IEEE Global Telecommunications Conference GLOBECOM 2010* (2010): 1-5.