On Radio Frequency Fingerprint Identification for DSSS Systems in Low SNR Scenarios

Yuexiu Xing, Aiqun Hu, Junqing Zhang, Linning Peng, and Guyue Li

***Abstract*—Radio frequency fingerprint (RFF) is an intrinsic hardware characteristic and has been employed for device identification. Its application in low signal-to-noise-ratio (SNR) has never been explored because its identification performance is greatly affected by the received signal quality. This paper proposes a novel RFF identification scheme for spread spectrum systems in low SNR scenarios. In the scheme, a signal preprocessing method, information data estimation based stacking (IDES) algorithm, is proposed, which leverages the repeated spreading sequences and stacks them together to eliminate the noise and interference effect. Simulation results demonstrate that the proposed scheme can achieve 98% identification rate when the received signal SNR is - 15 dB and the length of spreading sequence is 1023.**

***Index Terms*—Radio frequency fingerprint, spread spectrum, low SNR, device identification**

I. Introduction

RADIO frequency fingerprint (RFF) is an intrinsic characteristic of wireless device itself, which is caused by hardware imperfection resulted from the manufacturing process [1]. The RFF features, e.g., clock skew [2], cannot be tampered and has become an emerging device identification technique. Numerous RFF identification prototypes have been implemented with WiFi, ZigBee, Blutooth [1].

The performance of the RFF-based identification is greatly affected by the SNR of the received signals [3]. The work in [4] achieved an identification rate of about 100% for ZigBee device when the SNR is 30 dB. However, the identification in low SNR is very challenging because of the inaccurate estimation of the RFF features. For example, work [5] only achieves no more than 5% identification rate when the SNR is 0 dB. The performance can be improved by using a smarter classifier. Work [6] employes non-parametric random forest and multi-class adaboost ensemble classifier and identifies 40% ZigBee devices correctly when SNR = 0 dB. Multiple RFF features can also be leveraged to improve the performance [3].

The SNR of received signals can be very low in some scenarios, e.g., in satellite communications with extreme long communication distance. Direct sequence spread spectrum (DSSS) is able to work in very low SNR situations, e.g., as low as - 15 dB. While the DSSS receiver can identify different transmitters by detecting their particular spreading sequences in normal situations, the rogue devices can initiate attacks by spoofing and jamming [7]. RFF-based device identification scheme for DSSS systems is thus strongly required and its application in low SNR scenarios becomes essential.

This paper proposes a novel and robust RFF identification scheme for DSSS systems which can achieve a good identification performance even in low SNR scenarios. This scheme is consisted of three parts, including RFF model for DSSS systems, information data estimation based stacking (IDES) algorithm, RFF extraction and identification. The main contribution of this paper is that for the first time it proposes a signal preprocessing method, namely IDES algorithm, to improve the SNR of received signal without destroying RFF features. It utilizes the repeatability of spreading sequences to eliminate the noise and interference. Therefore, this scheme can be extended to other systems which have the same sequences in each frame signal, e.g., WiFi systems with preambles. The simulation results show that our scheme can achieve 98% identification rate when the signal SNR is as low as -15 dB, with the scheme configured with the length of spreading sequence as 1023 and number of stacked periods as 900.

II. RFF Model of DSSS Systems

We investigated an DSSS system modulated with offset quadrature phase shift keying (OQPSK) as a case study which is common in satellite and mobile communications because of its strong anti-interference capability.

DSSS employs a spreading sequence, $C(t)$, to modulate the information data, $D(t)$. The spreading sequence has N_c chips per period and each chip has a duration of $2T_c$. The sequence usually uses pseudorandom code and has a good autocorrelation property, i.e., $R_C(\tau) = 0$ when $\tau = 0$ and $R_C(\tau) = -\frac{1}{N_c}$ when $\tau \neq 0$. The duration of the information data, T_d, is equal to $2N_cT_c$.

The complex form of the baseband OQPSK symbols in $2T_c$ duration can be described as

$$X(t) = \begin{cases} A^I(t)\sin\left(\frac{\pi t}{2T_c}\right) + jA^Q(t)\cos\left(\frac{\pi t}{2T_c}\right), & 0 \le t \le T_c \\ A^I(t)\sin\left(\frac{\pi t}{2T_c}\right) - jA^Q(t+T_c)\cos\left(\frac{\pi t}{2T_c}\right), & T_c < t \le 2T_c \end{cases} \quad (1)$$

where $A^I(t) = C^I(t)D^I(t)$, $A^Q(t) = C^Q(t)D^Q(t)$, the superscript I and Q represent the real and imaginary part of the signals, respectively, and j is the imaginary multiplexing.

Manuscript received June 14, 2018; revised August 24, 2018; accepted September 15, 2018. Date of publication xx, 2018; date of current version xx 2018. This work was supported in part by National Natural Science Foundation of China 61571110, 61601114, and 61801115. The associate editor coordinating the review of this letter and approving it for publication was xxx. (*Corresponding author: A. Hu.*)

Y. Xing, A. Hu, L. Peng, and G. Li are with the Institute of Information Science and Engineering, Southeast University, No. 2 Sipailou, Nanjing, China. (e-mail: {yxxing, aqhu, pengln, guyuelee}@seu.edu.cn.)

J. Zhang is with the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, L69 3GJ, United Kingdom. (email: junqing.zhang@liverpool.ac.uk)

Digital Object Identifier xxx

Algorithm 1 IDES Algorithm

Input:
The received signal after carrier wipe-off, $Y(t)$;
The spreading sequence, $C(t)$;
The number of spreading sequences used in stacking, M;
Output:
The stacked signal in expected SNR, $Y_b(t)$;
1: Estimate code phase offset τ_d and frequency offset Δf of $Y(t)$ with the help of $C(t)$;
2: Estimate the information data $D(t)$ of $Y(t)$ with help of $C(t)$, τ_d , and Δf;
3: Stack M spreading sequences with the same polarity;
4: **return** The signal after stacking, $Y_b(t)$;

We chose three features as the transmitter's RFF, namely I/Q phase mismatch, I/Q DC offset, and I/Q gain imbalance. These features are generally considered constant [8] and stable [9] over the signal bandwidth. The baseband of the transmitted signal with RFF can be given as

$$T(t) = \left(\alpha X^I(t) + \lambda_I\right) + j\left(\beta X^Q(t|\phi) + \lambda_Q\right), \tag{2}$$

where α and β are the coefficients of I/Q gain, λ_I and λ_Q are the coefficients of I/Q DC offset, $X^Q(t|\phi) = A^Q(t)\cos\left(\frac{\pi t}{2T_c}+\phi\right)$ and ϕ is the normalized coefficient of I/Q phase mismatch.

It is common to use the same receiver to identify different devices, which means the receiver's RFF has the same influence to different devices and is not considered for simplicity. Therefore, the received signal after carrier wipe-off is

$$\begin{aligned} Y(t) =& \Big(\left(\alpha X^I(t-\tau_d) + \lambda_I\right) + j\left(\beta X^Q(t-\tau_d|\phi) + \lambda_Q\right)\Big) \\ & \exp(-j2\pi\Delta f t) + n(t), \end{aligned} \tag{3}$$

where τ_d and Δf are the code phase offset and frequency offset between the received signal and the transmitted signal, respectively, and $n(t)$ is the additive white Gaussian noise with mean of 0 and variance of $N_0/2$. Assuming the signal power is 1, then SNR of the received signal is $\gamma_s = \frac{1}{N_0/2}$.

III. IDES Algorithm

The received DSSS signal is buried in noise when the channel condition is bad. We proposed the IDES algorithm to improve the received signal SNR γ_s without destroying RFF features, which is explained in Algorithm 1.

A. Code Phase Offset and Frequency Offset Estimation

This section shows the estimation of τ_d and Δf of the in-phase of DSSS signal as an example. The correlation of one period $Y(t)$ and the spreading sequence $C^I(t)$ can be given as

$$\rho = \int_t^{t+2N_cT_c} Y(\tau)\cos(2\pi\hat{\Delta f}(\tau))C^I(\tau - \hat{\tau}_d - t)d\tau \tag{4}$$

where $\hat{\tau}_d = \tau_0, \tau_1, \cdots, \tau_{N_c}$, $\hat{\Delta f} = f_{min}, f_{min} + f_{step}, \cdots, f_{max}$. The two-dimensional and serial search method are used to realize the synchronization of code phase offset and frequency offset. When ρ is larger than a threshold, $\hat{\tau}_d$ and $\hat{\Delta f}$ are taken as the estimated values of τ_d and Δf, respectively. After the compensation of $\hat{\tau}_d$ and $\hat{\Delta f}$, the in-phase of DSSS signal can be written as

$$\hat{Y}^I(t) = \alpha X^I(t) + \lambda_I = \alpha C^I(t)D^I(t)\sin\left(\frac{\pi t}{2T_c}\right) + \lambda_I. \tag{5}$$

B. Information Data Estimation

In order to get the polarity of spreading sequences, we first correlate $\hat{Y}^I(t)$ with $C^I(t)$ to estimate the information data, which can be given as

$$\hat{\rho} = \int_t^{t+2N_cT_c} \left(\alpha C^I(t)D^I(t)\sin\left(\frac{\pi t}{2T_c}\right) + \lambda_I\right)C^I(\tau - t)d\tau. \tag{6}$$

When $t = iT_d, i = 0, 1, 2, \cdots$, $C^I(\tau)$ aligns with $C^I(\tau - t)$ and $\hat{\rho}$ will reach peaks and troughs, given as

$$\hat{\rho}' = \alpha D^I(t)\frac{4N_cT_c}{\pi}. \tag{7}$$

We can get the values of $D^I(t)$. Same approach can be applied to get the values of $D^Q(t)$.

C. Stacking

Finally, we choose M spreading sequences with the same polarity from the received signal, and stack them together. The signal after stacking is given in (8) and $n_M(t)$ is the noise after stacking. The equivalent SNR of the stacked signal is

$$\gamma_e = \frac{M^2}{MN_0/2} = M\frac{1}{N_0/2}, \tag{9}$$

which increases M times compared to the received signal SNR γ_s. This scheme utilizes the repeatability of spread sequences which is irrelevant of modulations.

D. Discussion

The estimation of τ_d and Δf affect the synchronization, which is a general step in RFF extraction. Information data estimation is a new and essential step to get the signal polarity because if the stacked signals do not have the same polarity, some RFF features would be damaged, such as I/Q DC offset. Stacking of spreading sequences is a new signal preprocessing operation to boost the signal quality without destroying RFFs.

Those steps mainly involve correlation and addition operations. Correlation is a cost effective operation in the hardware as it can be implemented by multipliers and addition is a simple operation. Therefore, the algorithm complexity is low. In addition, this algorithm exploits the existing repeated sequences, which does not require to generate spread spectrum signals.

IV. RFF Extraction And Identification

There are many RFF extraction methods. This paper adopted the differential constellation trace figure (DCTF) based-method as a case study, because DCTF has been demonstrated good identification performance [3]. A brief introduction is given below and a detailed description can be found in [10]. A blind differential process is performed with a differential interval of one chip duration of $C(t)$, which is given as

$$S(t) = Y_b(t)Y_b^*(t + 2T_c), \tag{10}$$

$$Y_b(t) = \begin{cases} M\left[\left(\alpha A^I(t)\sin\left(\frac{\pi t}{2T_c}\right)+\lambda_I\right)+j\left(\beta A^Q(t)\cos\left(\frac{\pi t}{2T_c}+\phi\right)+\lambda_Q\right)\right]+n_M(t), & 0\le t\le T_c \\ M\left[\left(\alpha A^I(t)\sin\left(\frac{\pi t}{2T_c}\right)+\lambda_I\right)-j\left(\beta A^Q(t+T_c)\cos\left(\frac{\pi t}{2T_c}+\phi\right)+\lambda_Q\right)\right]+n_M(t), & T_c\le t\le 2T_c \end{cases} \tag{8}$$

where $(\cdot)^*$ is the conjugate operation.

When the bits of $[A^I(t), A^I(t+2T_c), A^Q(t), A^Q(t+2T_c)]$ belongs to the following four groups:

$G_{11}=[1,1,1,1],[1,1,-1,-1]; G_{12}=[-1,-1,1,1],[-1,-1,-1,-1];$
$G_{21}=[1,-1,1,-1],[1,-1,-1,1]; G_{22}=[-1,1,1,-1],[-1,1,-1,1],$

the expectations of each groups are

$$G_{11}: E(S_{11}(t))=M^2(\lambda_I^2+\lambda_Q^2+\frac{\alpha^2+\beta^2}{2}+\frac{8T_c}{\pi}\alpha\lambda_I)+N_0M,$$
$$G_{12}: E(S_{12}(t))=M^2(\lambda_I^2+\lambda_Q^2+\frac{\alpha^2+\beta^2}{2}-\frac{8T_c}{\pi}\alpha\lambda_I)+N_0M,$$
$$G_{21}: E(S_{21}(t))=M^2(\lambda_I^2+\lambda_Q^2-\frac{\alpha^2+\beta^2}{2}-\frac{8T_c}{\pi}\alpha\lambda_Q j)+N_0M,$$
$$G_{22}: E(S_{22}(t))=M^2(\lambda_I^2+\lambda_Q^2-\frac{\alpha^2+\beta^2}{2}+\frac{8T_c}{\pi}\alpha\lambda_Q j)+N_0M.$$

There are four gathering centers in DCTF for groups G_{11}, G_{12}, G_{21} and G_{22}. I/Q gain imbalance, α and β, and I/Q DC offset, λ_I and λ_Q, will change the coordinates of those gathering centers. However, the I/Q phase mismatch ϕ influences the deviations but not the expectations at I channel.

The DCTF can be obtained by plotting $S(t)$ and an example is given in Fig. 1. Image processing algorithms can be used to analyze the features of the DCTF and the k-means clustering algorithm is used here. The RFF feature vector is given as

$$\vec{F}=\left[E\left(S_{11}(t)\right),E\left(S_{12}(t)\right),E\left(S_{21}(t)\right),E\left(S_{22}(t)\right)\right]. \tag{11}$$

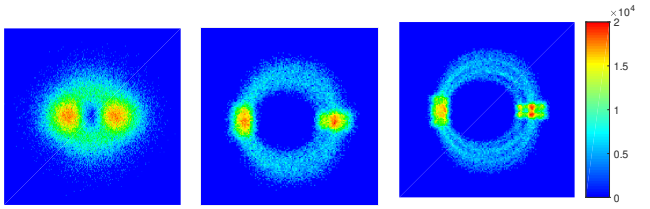
RFF identification usually involves a training stage and an identification stage. During the training stage, the RFF feature library is established by obtaining reference RFF features, $\vec{F}_{\text{ref}}$, for verified devices. The channel SNR of obtaining the library is denoted as γ_l. The library can be obtained with high SNR, e.g., trained by the manufacturer. Low SNR in this paper means that the SNR of received signals in identification stage is low but the γ_l can be high. During the identification stage, the system will return the corresponding features, $\vec{F}$. A minimum distance classifier is designed as

$$\arg\min_{d_i,d_j}|\vec{F}_{\text{ref}}(d_i)-\vec{F}(d_j)|, \tag{12}$$

when a device d_j needs to be identified and d_i is the device index if it is correctly identified. When $d_i\neq d_j$, an identification error happens.

V. Simulation Results and Discussion

The scheme was evaluated by Matlab simulation and characterized by the identification rate σ in different SNR scenarios. Ten devices were simulated by adding a set of different RFF features. In this paper, I/Q gain imbalance, I/Q DC offset and I/Q phase mismatch randomly locates in $0.85\sim1$, $0.01\sim0.28$, and $0.01\sim0.09$, respectively. In addition, frequency offset ranges from 2 kHz to 2.9 kHz, with a jitter of 40 Hz.



(a) $M=1$ (b) $M=20$ (c) $M=60$

Fig. 1. DCTF with different stacked periods. $\gamma_s=5$ dB.

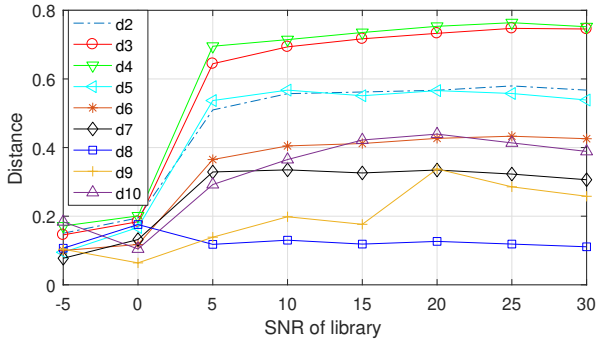


Fig. 2. Distances between device 1 and others devices

Fig. 1 illustrates the DCTFs with different stacked periods when $\gamma_s=5$ dB. Fig. 1(a) is the DCTF of the original signal whose four clustering centers are blurred into two decentralized signal point areas. However, as shown in Fig. 1(b)(c), the clustering centers are becoming clearer with the increase of stacked periods. In other words, the clustering centers can be extracted more accurately thanks to the IDES algorithm.

Fig. 2 describes the average distances between device d_1 and other devices under different library SNR γ_l. In order to find the reasonable γ_l, Fig. 2 investigated γ_l ranging from -5 dB to 30 dB. When the γ_l is lower than 5 dB, the distances are very small. The DCTF is seriously blurred in low SNR scenarios and the receiver can hardly extract accurate clustering centers. Fig. 2 also shows that distances are growing when γ_l ranges from 5 dB to 20 dB, and then tend to be stable. Libraries with high SNR are used in the rest of the paper.

Fig. 3 shows the equivalent SNR γ_e, and the identification rate σ with different stacked periods and received signals' SNR γ_s. We plotted γ_e, both from the simulation and the analytic expression (9), for the scenario when $\gamma_s=10$ dB. These results match very well. When the γ_e of stacked signal is higher than the library SNR γ_l, 20 dB in this example case, the σ will keep steady first but then decrease. It is more robust when the received signal SNR and library SNR is

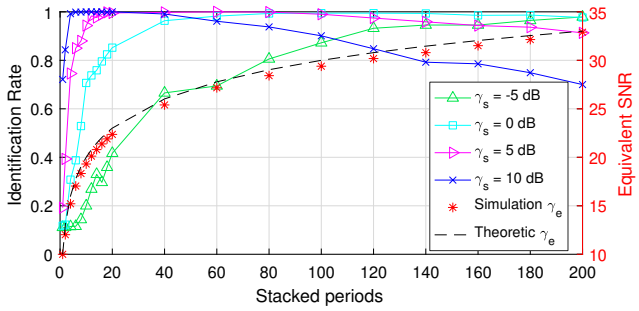


Fig. 3. Identification rate σ and equivalent SNR. γ_I = 20 dB.

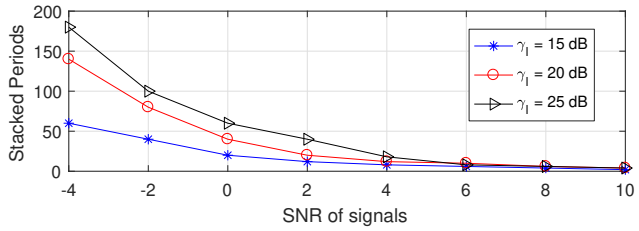


Fig. 4. The minimum M to achieve a high identification rate σ as 95%.

similar. This phenomenon also has been observed in [11]. On the other hand, when $\gamma_e < \gamma_I$, the σ rises with the increase of stacked periods because the IDES algorithm improves the SNR successfully without the destruction of RFF features. In addition, the σ increases with the γ_s when the stacked periods is fixed, because a higher SNR is beneficial to obtain more accurate RFF features. When M=1, the results in Fig. 3 represent the performance of DCTF based method in [10], i.e., without IDES algorithm. It achieved an identification rate of 11% and 72% when γ_s is -5 dB and 10 dB, respectively. On the other hand, the identification rate can reach over 98% when IDES was adopted. This demonstrates that the IDES algorithm can significantly improve the identification performance, especially in low SNR scenarios.

Fig. 4 shows the minimum number of stacked periods M required to achieve a high identification rate σ, e.g., 95%, with different libraries. Less stacked periods are required when the γ_I is closer to the γ_s. For example, when γ_s is -4 dB, 60 stacked periods are required when γ_I is 15 dB, while the required stacked periods increase to more than 140 when γ_I is 20 dB or 25 dB. More signal stacking will result in longer process time and higher computational cost. The selection of libraries thus should be under very careful consideration with a tradeoff among the processing time, computing costs and the stability and accuracy of library.

Finally, Table. I presents the identification rate σ when the received signals' SNR γ_s is very low, i.e., -15 dB. The γ_I is chosen as 15 dB, based on the balance of the time and computing costs and the stability and accuracy of library. The length of the spreading sequence used is 1023, which has about 30 dB spread spectrum gain. The DSSS system is able to operate when the SNR is as low as - 15 dB in this case [12]. The obtained result shows that the scheme can achieve 98% identification rate when M = 900. Our scheme is thus totally applicable to very low SNR.

TABLE I
IDENTIFICATION RATE σ. γ_s = −15 DB AND γ_I = 15 DB.

M	1	100	200	300	400	500	600	700	800	900
σ(%)	0	0	5.4	32.3	50.6	61.2	78.1	88.7	97.3	98.5

VI. Conclusion

This paper proposed a novel and robust RFF identification scheme for DSSS, which is able to operate even in very low SNR. The proposed IDES algorithm in the scheme leveraged the repeated spreading sequences and stacked the signals together to eliminate the noise and interference effect. Extensive simulation was carried out to investigate the identification performance. The scheme was demonstrated to be effective in very low SNR scenarios, which achieved 98% identification rate when the length of spreading sequence is 1023 and the stacking period is 900. Our future work will focus on recognition of unauthorized device, e.g., smart attacker as investigated in [13].

References

[1] Q. Xu, R. Zheng, W. Saad, and Z. Han, "Device fingerprinting in wireless networks: Challenges and opportunities," *IEEE Commun. Surveys Tuts.*, vol. 18, pp. 94–104, Sep. 2016.

[2] M. Cristea and B. Groza, "Fingerprinting smartphones remotely via ICMP timestamps," *IEEE Commun. Lett.*, vol. 17, pp. 1081–1083, Apr. 2013.

[3] L. Peng, A. Hu, J. Zhang, Y. Jiang, and Y. Yan, "Design of a hybrid RF fingerprint extraction and device classification scheme," *IEEE Internet Things J.*, p. in press.

[4] C. G. Wheeler and D. R. Reising, "Assessment of the impact of CFO on RF-DNA fingerprint classification performance," in *Proc. ICNC*, Santa Clara, CA, USA, Jan. 2017, pp. 110–114.

[5] P. Hao, X. Wang, and A. Behnad, "Relay authentication by exploiting I/Q imbalance in amplify-and-forward system," in *Proc. IEEE GLOBECOM*, Austin, TX, USA, Dec. 2014, pp. 613–618.

[6] H. J. Patel, M. A. Temple, and R. O. Baldwin, "Improving ZigBee device network authentication using ensemble decision tree classifiers with radio frequency distinct native attribute fingerprinting," *IEEE Trans. Rel.*, vol. 64, pp. 221–233, Mar. 2015.

[7] A. Broumandan, A. Jafarnia-Jahromi, V. Dehghanian, J. Nielsen, and G. Lachapelle, "GNSS spoofing detection in handheld receivers based on signal spatial correlation," in *Proc. IEEE/ION PLANS*, Myrtle Beach, SC, USA, Apr. 2012, pp. 479–487.

[8] G. Xing, M. Shen, and H. Liu, "Frequency offset and I/Q imbalance compensation for direct-conversion receivers," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 673–680, Mar. 2005.

[9] N. T. Nguyen, G. Zheng, Z. Han, and R. Zheng, "Device fingerprinting to enhance wireless security using nonparametric bayesian method," in *Proc. IEEE INFOCOM*, Shanghai, China, Apr. 2011, pp. 1404–1412.

[10] L. Peng, A. Hu, Y. Jiang, Y. Yan, and C. Zhu, "A differential constellation trace figure based device identification method for ZigBee nodes," in *Proc. WCSP*, Yangzhou, China, Oct. 2016, pp. 1–6.

[11] P. Robyns, E. Marin, W. Lamotte, P. Quax, D. Singelée, and B. Preneel, "Physical-layer fingerprinting of LoRa devices using supervised and zero-shot learning," in *Proc. WiSec*, New York, NY, USA, Jul. 2017, pp. 58–63.

[12] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread spectrum communications handbook*. USA: McGraw-Hill, Inc., 1994.

[13] A. C. Polak and D. L. Goeckel, "Identification of wireless devices of users who actively fake their RF fingerprints with artificial data distortion," *IEEE Trans. Wireless Commun.*, vol. 14, pp. 5889–5899, Nov. 2015.