Securing IoT monitoring device using PUF and physical layer authentication

Zheng, Yue; Dhabu, Sumedh Somnath; Chang, Chip Hong

2018

Zheng, Y., Dhabu, S. S., & Chang, C. H. (2018). Securing IoT monitoring device using PUF and physical layer authentication. 2018 IEEE International Symposium on Circuits and Systems (ISCAS). doi:10.1109/ISCAS.2018.8351844

https://hdl.handle.net/10356/80436

https://doi.org/10.1109/ISCAS.2018.8351844

Securing IoT Monitoring Device using PUF and Physical Layer Authentication

Yue Zheng, Sumedh Somnath Dhabu, and Chip-Hong Chang
School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
Email: yzheng015@e.ntu.edu.sg, sdhabu@ntu.edu.sg, echchang@ntu.edu.sg

***Abstract*—The IoT is rapidly becoming a reality. Forecasts predict more than 20 billion connected devices in 2020. These devices bring many benefits, but security them in IoT environment can be a quandary. With the advent of technology, it is very easy for an adversary to clone a device and replace it, or tamper the data. In the context of wireless communications in IoT, the definition of message authentication should be extended to include verification of the device along with the integrity of the message it produced. In this paper we propose a device- and data-dependent physical layer authentication scheme by using a device-specific, dynamically variable key to generate a data-dependent tag. This tag is embedded in the data transmission using an information hiding scheme to reliably extract it at the receiver, without compromising the performance of the underlying wireless communication system. Simulation results show that our scheme can achieve high authentication rate while rejecting the tampered transmissions in typical noisy communication channels.**

I. Introduction

The advent of wireless networking techniques has resulted in the well-known paradigm of Internet of Things (IoT). This connected environment is established by everyday devices equipped with appropriate sensors, computing power and transceivers for digital communication [1], [2]. Wireless communication is inherently vulnerable to various types of malicious attacks because of its broadcast nature, and the situation is exacerbated by multiple wireless entities in IoT. In application scenarios such as remote video-monitoring, e-health systems etc., a receiver in an IoT environment should be able to reliably verify the claimed identity of the transmitter and detect tampered received data in the transmission-reception process to avoid catastrophe. Consider the example of a remote video-monitoring system implemented on a large estate. Video cameras installed at various remote locations (or on unmanned aerial vehicles - UAVs) transmit data to the central unit situated in the home. In order to compromise the premises monitoring, an intruder (Eve) can pose as a legitimate video transmitter (Alice) to make the central unit receiver (Bob) accept her own video feed. To thwart such an impersonation attack, Bob should be able to authenticate the transmissions from Alice and reject the transmissions from Eve. To this end, Alice can create and transmit a message-dependent tag along with her transmissions. Bob endorses Alice as the sender of the message and the data is tamper-free only if a valid tag is detected in the received data.

Security primitives for such tagged authentication traditionally implemented at the upper layers of the protocol stack can be compromised [3], [4]. Simply appending the tag at the end of transmission also reduces the data throughput. As a complementary approach to higher layer security, authentication at physical layer provides greater trust assurance [5]–[12]. Similar to message authentication at the higher layers, active schemes [8]–[12] make use of a tag that Alice embeds in her transmission. It is generated using a fixed pre-shared key between Alice and Bob, which can be stored on the on-chip non-volatile memory of Alice's device. This method is not sufficiently secure and efficient as non-volatile memory can be cracked by advanced attack technology and the cost of integrating it onto the chip is higher than volatile memory. Besides, an attacker may retrieve such a key by physically accessing Alice's device.

Existing active methods [8]–[12] rely solely on the uncertainty of Eve's observations of tag and her limited computational ability. With sufficient observations, Eve could use machine learning attacks to break the key. A potential solution is to renew the Alice-Bob key at regular intervals, which incurs a heavy penalty of multiple back-and-forth communications between Alice and Bob over an insecure channel. A dynamically variable key can be created by appending the transmission number to pre-shared static key [8]–[12]. However, it is difficult to synchronize Alice and Bobs transmission numbers in a potentially adverse environment and retransmissions are often required. Alice and Bob may generate matching keys using the characteristics of the Alice-Bob channel, based on the assumption that the channel between a transmitter-receiver pair is unique, reciprocal, and multipath-rich [6]. Such channel may still be manipulated by the attacker to influence the key generation process. Moreover, such a scheme needs constant channel validation for successive transmissions. It may not be suitable when Alice-Bob channel varies abruptly, e.g., when Alice's device is not stationary (a camera mounted on a UAV) or when Eve-Bob channel is similar to Alice-Bob channel within permissible limits, e.g., when Eve replaces Alice's device with her own device or she places it within a close proximity. For transmission authentication, Bob can undertake a passive approach [5]–[7] and monitor Alice's RF transmission fingerprints, but Eve can also do so. With the contemporary cognitive radio and wireless transmitters, it may be possible for Eve to mimic Alice's transmission characteristics closely.

Advanced threat protection calls for dynamically variable key derivation function for Alice and Bob and more robust information hiding scheme to communicate the tag without unacceptably degrading the performance. The key derivation function should be independent of the operating environment so that the key and tag cannot be controlled and/or forged by Eve. The tag should be unique to Alice's device and data, such that Eve cannot reproduce the same tag and Bob is able to identify any possible tampering by Eve. With these considerations, we propose a device- and data-dependent physical layer authentication scheme for wireless remote monitoring. Our scheme uses physical unclonable function (PUF) [13] and perceptual image hash to create a device-specific key and a data-based tag respectively, and an information hiding scheme to transmit this tag with the data through a public channel. A PUF is an integrated circuit that transforms the physical disorder of random semiconductor fabrication process variations of its nanoscale devices into a unique and unpredictable digital bitstream (response) upon query (challenge). In our scheme, the

PUF forms the basis of unique, environment-independent, hard-to-break device key derivation function. Alice uses the PUF response of the image content capturing device to compute an authentication tag based on perceptual image hash. Perceptual hash is robust to addition of noise or content-preserving operations, but is otherwise sensitive to malicious tampering of image content. This device- and data-dependent authentication tag is embedded in the transmissions using the information hiding scheme from [14], which is based on the principle that small variations (by the information to be hidden into the transmitted message) in the frequency response of the pulse shaping filter do not affect the received information symbols, but can be reliably identified at the receiver's end by the intended recipient of the hidden message.

II. Proposed Physical Layer Authentication Method

For illustration, an end-to-end system model of the aforementioned wireless monitoring scenario is shown in Fig. 1. The following flow provides an overview of the proposed scheme.

1). The control room (Bob) creates a database of challenge-response pairs of the PUF on the monitoring device (Alice). This database is known only to Bob.
2). The monitoring device installed at the desired location (or mounted on the desired UAV) captures and transmits images/video to the control room using a suitable wireless communication system termed "reference communication system".
3). Using an on-chip true random number generator (TRNG), Alice generates a random challenge C_{tx} to query the PUF for a response R_{tx}.
4). Alice uses a perceptual image hash function to create an authentication tag H_I from R_{tx} of the image or video bitstream m_t to be transmitted to Bob.
5). Alice uses the proposed information hiding scheme [14] to secretly transmit t_t, which is a concatenation of C_{tx} and H_I, along with m_t.
6). From the received signal, Bob computes the received version of m_r and t_r in order to recover m_t and t_t, respectively. From t_r, Bob obtains an estimated $\hat{H}_I$ of H_I and C_{rx} of C_{tx} subjected to bit error rate introduced by channel noise. Using C_{rx}, Bob retrieves the response R_{rx} from the database.
7). Using m_r and R_{rx}, Bob generates his local copy of authentication tag $\tilde{H}_I$.
8). Bob compares $\tilde{H}_I$ against $\hat{H}_I$. The received message is accepted as originated from an authentic device and transmitted by Alice if $\tilde{H}_I \simeq \hat{H}_I$ within the tolerance of hamming distance. Otherwise it is rejected.

The steps involved are discussed in the following subsections.

A. Authentication Tag Generation

To securely authenticate the data transmitted by a remote monitoring device, its authentication tag should satisfy the following requirements. Firstly, the tag obtained from original image should be substantially different from the tag computed from the maliciously tampered image, but highly similar to the tag computed from the (noisy) image received through the communication channel. Secondly, the tag should also authenticate the device. A malicious attack performed by either tampering the image or changing the source device or a combination of these acts should cause discriminable change on the tag value while the inevitable noise contamination during the transmission process shall have insignificant effect. Last but not least, the bitstream of the tag should be sufficiently short for transmission efficiency. To meet these requirements, the following processes are involved in the proposed PUF based perceptual biohashing.

1) Adjoint Block-based DCT (Bb DCT) Concatenated Feature Extraction: Unlike biometric authentication problem, the images that are to be sent by the monitoring device could be highly diverse and dynamic. They cannot be pragmatically associated with pre-defined classes, which means that training phase is infeasible for such system. As a workaround of this problem, robust and concise features of each individual image are extracted. After applying pre-processing techniques (scaling, filtering, histogram equalization) to preserve the original salient features, Discrete Cosine Transform (DCT) is used to extract the local features from the DCT coefficients of the pre-processed image I computed in the frequency domain. The first DCT coefficient $C(0,0)$ is called the DC coefficient and the remaining coefficients are called the AC coefficients of I.

As tampering tends to focus on dominant local instead of global features, it can be better detected from the change in dominant block features by performing DCT at block level. Hence, the $N \times N$ image I is first divided into M non-overlapping blocks before the 2D-DCT transform is applied on each block to obtain its DCT coefficients. The DC coefficient is relatively large in magnitude compares to its AC coefficients. As perceptually sensitive content of an image resides predominantly in the DC coefficient and a few of its neighboring low frequency AC coefficients, only the first eight AC coefficients in a raster scan order from the top left corner of the DCT spectrum of each block are selected and quantized to extract the feature vector of the block, as shown in Fig. 2. The extracted feature vectors of all blocks are concatenated to form the feature vector F_I of the image I as follows:

$$F_I = [f_1, f_2, f_3, \cdots, f_M] \tag{1}$$

The 8-bit feature of the i_{th} block is given by

$$f_i = [C_i(1,2), C_i(2,1), \cdots, C_i(3,2), C_i(3,3)] \tag{2}$$

where $C_i(p,q)$ denotes the quantized coefficient of the i^{th} block at spatial location (p,q).

As the block feature f_i in (2) only has eight bits, it may be hard to differentiate between bit flipping of original hash image tag due to transmission noise from the modified hash value of maliciously tampered image. Including more AC coefficients does not help since they carry insignificant energy of the localized perceptual information. Enlarging the DCT block size will further reduce the granularity of energy concentration on local features. To mitigate this problem, after calculating the 8-bit DCT binary feature of each block, the four neighboring blocks are combined to form a larger "*cblock*", as shown in Fig. 2. The corresponding features are concatenated to form a 32-bit feature vector of this $cblock$ without compromising the resolution and energy of localized features. Therefore, the feature vector F_I of image I will have $M/4$ adjoined $cblock$ features $FB_{i,j}$ for $i,j \in 1,2,\cdots,M/2$, while each $FB_{i,j}$ is a concatenation of four smaller subblock features $fb_{m,n}$, where $(m,n) = (2i-1, 2j-1), (2i-1, 2j), (2i, 2j-1)$ and $(2i, 2j)$. $FB_{i,j}$ will be used as an input for the generation of perceptual hash value.

2) Perceptual Biohashing by Image Sensor PUF: A PUF-based biohashing scheme was first proposed in our previous work [15] to incorporate both image features and device fingerprint. The final biohash vector can have comparatively flexible bit length ($\leq$ image feature dimension) and is capable of

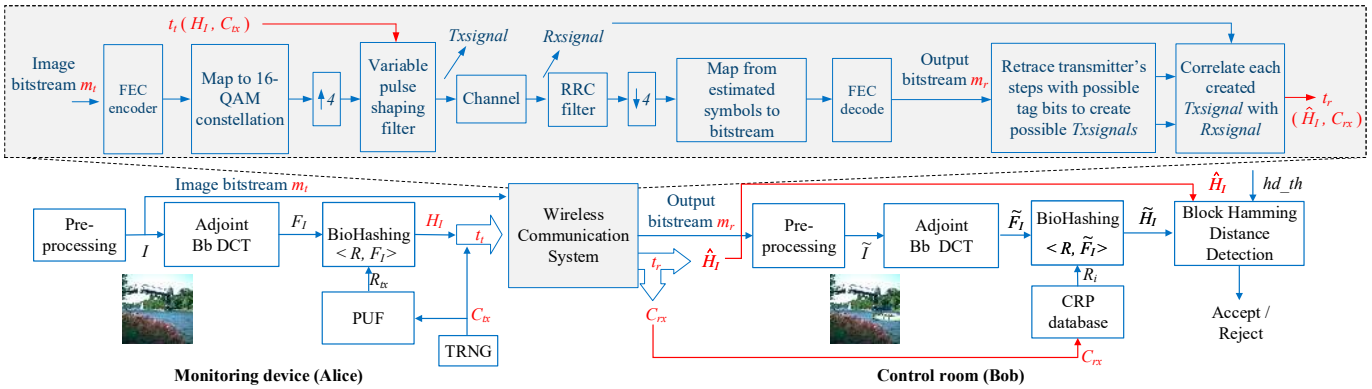

Fig. 1. End to end system diagram for tagged authentication based remote monitoring system.

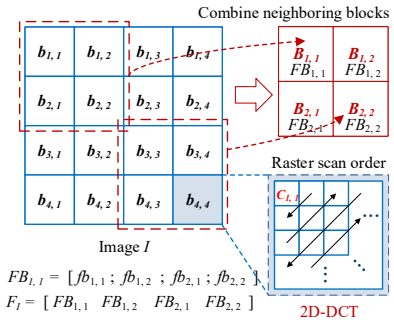

Fig. 2. An example of concatenated adjoint block-based DCT feature extraction for a 4×4 image subblock of I.

distinguishing different (image, device) combinations. In the proposed system, a CMOS image sensor PUF [16], where the CRPs are extracted from the fixed-pattern noise of its active pixel array, is used as a "device biometric" to birthmark the image source. A detailed description of the image sensor PUF is beyond the scope of this paper. Interested readers are referred to [16] for more details.

A perceptual biohash H_i of the i^{th} block can be obtained by the following steps. A device key R_i is first generated by stimulating the image sensor PUF with a randomly generated challenge C_i. The PUF response R_i is transformed into an orthonormal matrix $\{U_i \in \mathbb{R}^m | i = 1, 2, ..., n\}$ by Gram-Schmidt algorithm, where n is the length of the final hash vector ($n \leq m$).

Let FB_i be the extracted m-bit feature vector of the i^{th} $cblock$. A random projection of FB_i is made by its inner product with U_i, i.e.,

$$v_i = \langle FB_i \cdot U_i \rangle = (U_i)^T (FB_i) \tag{3}$$

The projection result $\{v_i\}_{k=1}^n \in \mathbb{R}$ is digitalized by comparing the k_{th} member $v_{i,k}$ of v_i against a pre-defined threshold τ to obtain an n-bit binary vector $\{H_i\}_{k=1}^n$, i.e.,

$$H_{i,k} = \begin{cases} 0 & \text{if } v_{i,k} \leq \tau \\ 1 & \text{if } v_{i,k} > \tau \end{cases} \tag{4}$$

where $H_{i,k} \in 0, 1 \forall k = 0, 1, \cdots, n-1\}$ is the k^{th} binary bit of H_i.

The hash value of I with M blocks of concatenated adjoint DCT features is given by:

$$H_I = [H_1, H_2, ..., H_M] \tag{5}$$

The hash vectors computed from the transmitted image and received image are denoted as H_I and $\tilde{H}_I$, respectively. Their

Fig. 3. Two groups of examples showing the original image, tampered image and the detected region, respectively.

hamming distance (HD) is calculated at $cblock$ level. If the HD HD_i between the hash vectors for the i^{th} $cblock$ of the two images is detected to exceed a pre-defined threshold t_{tamp}, the i_{th} $cblock$ of the received image is considered to be tampered. Otherwise the received image is accepted as authentic.

B. Wireless Communication System

We note that the proposed authentication scheme will work if and only if $\tilde{H}_I \simeq \hat{H}_I$. In order to satisfy this condition, Bob should receive the authentication tag without any error, i.e., $C_{rx} = C_{tx}$ and $\hat{H}_I = H_I$. Other desired properties are the tag transmission should be stealthy, transparent to a receiver which is not aware of the presence of the tag, and should not degrade the performance of the underlying communication system. A recently proposed information hiding scheme from [14] satisfies all these requirements, and can therefore be used to realize this physical layer authentication with PUF-based perceptual image hash. The effectiveness of this scheme and its advantages over [8]–[12] have been shown in [14] and the same system model (with addition of forward error correction encoder and decoder) is used here. The information hiding scheme will not be reiterated here due to the page limit .

III. Results and Discussions

We constructed a database with 50 color images, and corresponding manually tampered images using PhotoShop. In each tampered image, the area of the tampered part is around 5% of that of the original image. Fig. 3 shows two examples of the original and tampered images from this database. The images are resized to 256×256 pixels. The sizes of DCT block and $cblock$ are 16×16 and 32×32, respectively. The CMOS image sensor based PUF is simulated in Cadence environment to produce the CRP database for the authorized remote monitoring camera device. More details on the uniqueness and randomness of this PUF can be found in [16]. For illustration purpose, the lengths of the challenge and response used are 12 and 512 bits, respectively. The latter is rearranged into a 32×16-bit array for hashing operation. All the image processing, transmission-reception and authentication operations are implemented in Matlab.

Perceptual robustness describes the ability of the image hashing method to be resilient against non-malicious or content-preserving modifications. In our case, it refers to resiliency against the addition of noise due to wireless communication channel, i.e., the hash values generated from the original images and received (noisy) images should be within the threshold of hamming distance, when generated using the same key in each pair comparison. At the same time, the image hashing method should also provide good tampering detection accuracy, i.e., the hash values generated using the same key for original image and tampered image should be different enough. We simulated the wireless communication system shown in Fig. 1, for channel with additive white Gaussian noise (AWGN) with SNR varying from 8 dB to 15 dB. Table. I second row shows the average detection accuracy by comparing the hash values generated from the original images and received (noisy) images. The results obtained by comparing the hash values for original images and hash values for received (noisy) tampered images are shown in Table. I third row. We note that all the results in Table I are obtained for same (empirically determined) threshold $t_{tamp} = 0.3$, which is used to classify the received image as either original or tampered. Our image hashing method has good perceptual robustness as well as good tampering detection ability. Fig. 3 also shows the detected regions of the previous two examples.

TABLE I
ACCURACY OF PERCEPTUAL ROBUSTNESS (PR) AND TAMPERING DECTION (TD) PERFORMANCE.

SNR(dB)	8	9	10	11	12	13	14	15
PR^1	0.27	0.6	0.87	0.94	1	1	1	1
TD^2	0.98	0.96	1	0.96	0.94	0.92	0.92	0.92

PR^1: Detection Rate of Noisy but Authentic Images.
TD^2: Detection Rate of Noisy Tampered Images.

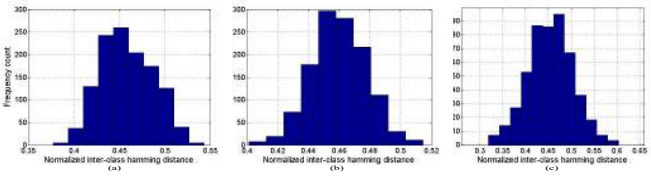

Fig. 4. Hamming distance (HD) between (a)(different image, different device key) (mean (HD)= 0.458); (a)(different image, same device key) (mean (HD)= 0.460); (b)(same image, different device key) (mean (HD)= 0.456)

The discriminability of the proposed system is also evaluated by verifying the distinguishability of hash vectors generated for the (different image, different device key) cases. Two additional cases need to be evaluated. Firstly, as it is impractical for each image to have a unique device key for transmission, the system must be capable of differentiating the hash vectors among (different image, same device key) cases. Besides, the control unit must be able to distinguish the hash vectors of the same images sent by different devices. The discriminability of all the three cases are evaluated by measuring their corresponding inter-class hamming distances. By setting the t_{tamp} to 0.3 as before, the simulation results in Fig. 4 show 100% discriminability for all the three cases as the normalized inter-class hamming distances are all above 0.3.

Our simulation results on the bit error rate (BER) performance of the recovered data and secret bitstreams from the transmission and reception of 50 images based on the system shown in Fig. 1 are in agreement with the results obtained in [14]. The results showed that modifying the pulse shaping filter coefficients of transmitter by the authentication tag (secret bitstream) does not have any significant effect on the BERs of image data bitstream. The received images are practically identical without and with pulse shaping filter modification by authentication tag. The authentication tag was retrieved successfully without any error, i.e., $t_r = t_t$, for every simulation of each SNR. Appropriate parameters were selected to take advantage of embedding small number of tag bits into the transmissions of large number of data bits.

For security analysis, we assume that Eve is a powerful adversary who knows all about the system except the CRP database. Eve succeeds in impersonating Alice if Bob accepts her message as authentic. We note that Eve cannot simply replay the previous transmissions from Alice. This is because the authentication tag for every image is generated using the PUF response to a random challenge, hence the tag is time- and message-dependent. For Eve to impersonate Alice, the authentication tag embedded in Eve's transmission should look authentic to Bob, i.e., with Eve's C_{rx}, the response R_i obtained from Bob's CRP database should generate an acceptable hash value. As Eve also receives m_r and t_r, she can succeed in this task provided that she can $1)$. obtain the correct responses (R_i) to generate sufficiently accurate hash values $(\hat{H}_I)$ for multiple images, $2)$. prepare a database of the observed challenges $(C_r x)$ and response calculated in point $1)$. As perceptual image hash is an intractable one-way function without the key [17], Eve must first conduct an exhaustive search for right R_i to generate the observed hash $(\hat{H}_I)$ value for each image. Then she has to model the PUF based on the database she created from her observations. Both these tasks are infeasible with sufficiently long challenges, response and hash values. Alternatively, Eve may attack by embedding a random authentication tag in her transmission using her own transmitter, but the probability that it will result in successfully authenticating Eve as Alice is $\frac{\sum_{\lceil(1-t_{tamp})\times 16\rceil}^{16}[16C_i]^{64}}{2^{1024}}$. For a threshold of 0.3, this probability is less than 10^{-74}.

IV. CONCLUSION

A novel physical layer authentication method for remote monitoring system in IoT is proposed in this paper. PUF-based perceptual image hash not only enables the intended receipient of a digital media to detect small malicious content tampering from the received copy through a noisy public channel, but also allows the control center to authenticate the source device used to generate the digital content. The scheme is evaluated to achieve good perceptual robustness, high tampering detection rate and excellent discriminability for different content and device scenarios.

REFERENCES

[1] S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K. S. Kwak, "The internet of things for health care: A comprehensive survey," *IEEE Access*, vol. 3, pp. 678–708, 2015.

[2] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, Feb 2014.

[3] N. Borisov, I. Goldberg, and D. Wagner, "Intercepting mobile communications: the insecurity of 802.11," in *Proc. 7th annual international conference on Mobile computing and networking*, 2001.

[4] A. Mishra, M. Shin, and W. A. Arbaugh, "Your 802.11 network has no clothes," *IEEE Communications Magazine*, pp. 44–51, 2002.

[5] L. Xiao, L. Greenstein, N. Mandayam, and W. Trappe, "Fingerprints in the Ether: Using the Physical Layer for Wireless Authentication," in *2007 IEEE Int. Conf. Communications*, June 2007, pp. 4646–4651.

[6] W. Trappe, "The challenges facing physical layer security," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 16–20, June 2015.
[7] W. Wang, Z. Sun, S. Piao, B. Zhu, and K. Ren, "Wireless physical-layer identification: Modeling and validation," *IEEE Trans. Information Forensics and Security*, vol. 11, no. 9, pp. 2091–2106, Sept 2016.
[8] P. L. Yu, J. S. Baras, and B. M. Sadler, "Physical-layer authentication," *IEEE Trans. Information Forensics and Security*, vol. 3, no. 1, pp. 38–51, March 2008.
[9] G. Verma, P. Yu, and B. M. Sadler, "Physical Layer Authentication via Fingerprint Embedding Using Software-Defined Radios," *IEEE Access*, vol. 3, pp. 81–88, 2015.
[10] N. S. Goergen, W. S. Lin, K. J. R. Liu, and T. C. Clancy, "Extrinsic Channel-Like Fingerprinting Overlays Using Subspace Embedding," *IEEE Trans. Information Forensics and Security*, vol. 6, no. 4, pp. 1355–1369, Dec. 2011.
[11] N. Goergen, W. S. Lin, K. J. R. Liu, and T. C. Clancy, "Extrinsic Channel-Like Fingerprint Embedding for Authenticating MIMO Systems," *IEEE Trans. Wireless Communications*, vol. 10, no. 12, pp. 4270–4281, Dec. 2011.
[12] V. Kumar, J. M. Park, T. C. Clancy, and K. Bian, "PHY-layer authentication using hierarchical modulation and duobinary signaling," in *Proc. 2014 Int. Conf. Computing, Networking and Communications (ICNC)*, Feb. 2014, pp. 782–786.
[13] C. H. Chang, Y. Zheng, and L. Zhang, "A Retrospective and a Look Forward: Fifteen Years of Physical Unclonable Function Advancement," *IEEE Circuits and Systems Magazine*, vol. 17, no. 3, pp. 32–62, Aug. 2017.
[14] S. Dhabu and C. H. Chang, "A novel scheme for information hiding at physical layer of wireless communication system," in *Proc. 14th Int. SoC Design Conf.*, South Korea, Nov. 2017, to be published.
[15] Y. Zheng, Y. Cao, and C. H. Chang, "Facial Biohashing based User-Device Physical Unclonable Function for Bring Your Own Device Security," in *Proc. 2018 Int. Conf. Consumer Electronics (ICCE)*, Las Vegas, US, Jan. 2018, to be published.
[16] Y. Cao, L. Zhang, S. S. Zalivaka, C. H. Chang, and S. Chen, "CMOS Image Sensor based Physical Unclonable Function for Coherent Sensor-Level Authentication," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 62, no. 11, pp. 2629–2640, Nov. 2015.
[17] X. Lv and Z. J. Wang, "Perceptual Image Hashing Based on Shape Contexts and Local Feature Points," *IEEE Trans. Information Forensics and Security*, vol. 7, no. 3, pp. 1081–1093, Jun. 2012.