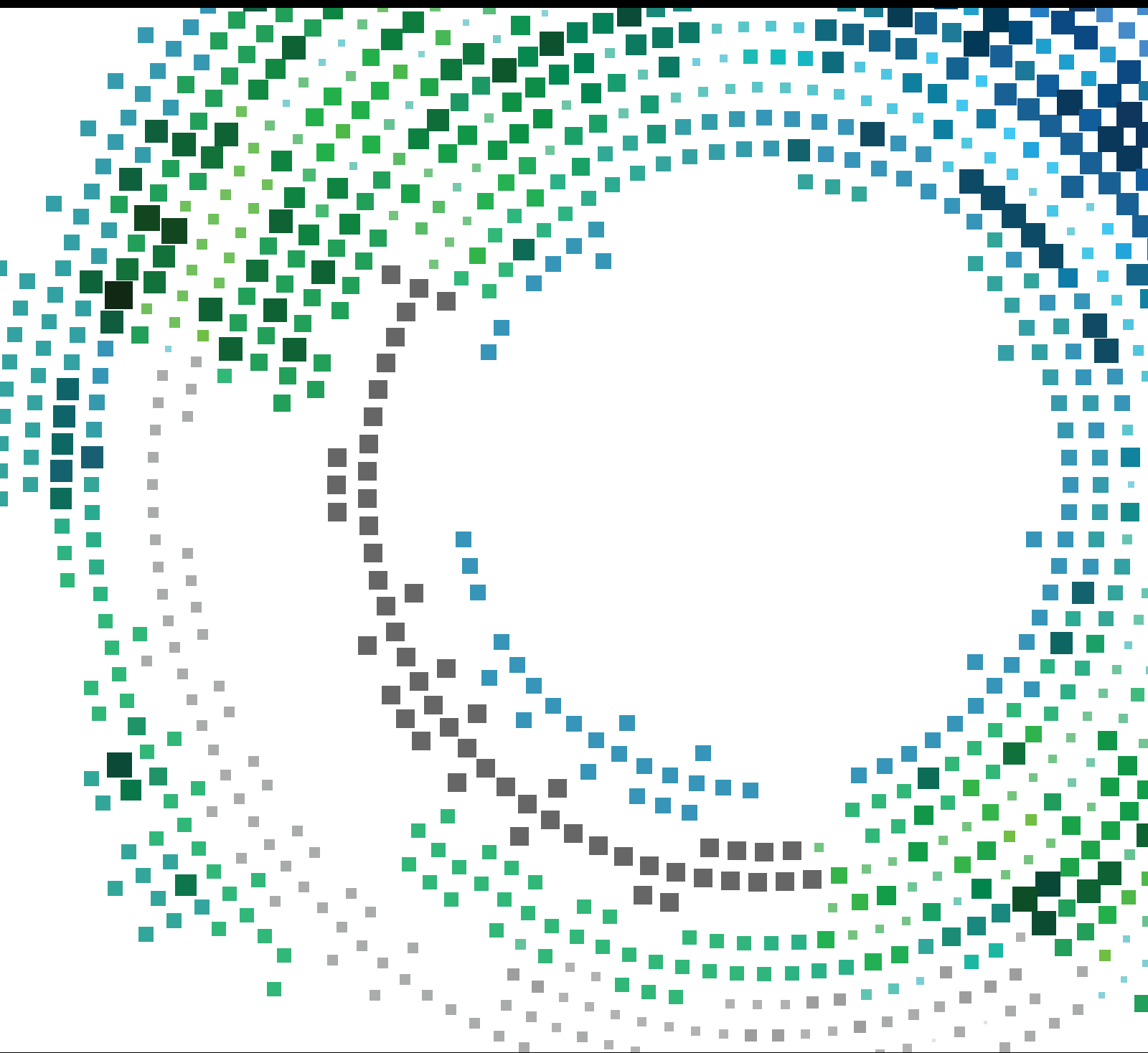
Mobile Information Systems

Mobile Cyber-Physical System

Lead Guest Editor: Xiping Hu
Guest Editors: Jun Cheng, Xitong Li, Wei Tan, Qiang Liu, and Zhengguo Sheng



Mobile Cyber-Physical System

Mobile Information Systems

Mobile Cyber-Physical System

Lead Guest Editor: Xiping Hu
Guest Editors: Jun Cheng, Xitong Li, Wei Tan, and Zhengguo Sheng

Editorial Board

Contents

Hindawi
Mobile Information Systems
Volume 2017, Article ID 4970290, 2 pages
https://doi.org/10.1155/2017/4970290

Editorial
Mobile Cyber-Physical System

Xiping Hu,[1] Jun Cheng,[1] Xitong Li,[2] Wei Tan,[3] Qiang Liu,[4] and Zhengguo Sheng[5]

[1]*Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China*
[2]*HEC Paris, Paris, France*
[3]*IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA*
[4]*National University of Defense Technology, Changsha, China*
[5]*University of Sussex, Brighton, UK*

Correspondence should be addressed to Xiping Hu; xipingh@bravolol.com

Received 15 May 2017; Accepted 16 May 2017; Published 3 August 2017

In recent years, the capabilities of contemporary mobile devices have been improving a lot. These capabilities, such as significant computational resources, multiple communication radios, various sensor modules, and high level programming languages, enable mobile devices to form mobile cyber-physical system (CPS) in our daily lives. Mobile CPS integrates distributed sensing with computing and ubiquitous connectivity of Internet. Also, mobile CPS coordinates computational, virtual, and physical resources and facilitates the interaction of digital world with physical world, potentially driving the pervasive effect in the citizens' everyday life anytime and anywhere. Thus, mobile CPS could provide a convenient and economical platform that facilitates sophisticated and ubiquitous intelligent applications between humans and the surrounding physical world. Mobile CPS could find applicability in multiple areas and disciplines, including (but not limited to) (1) mobile intelligent robots and robotics systems, making use of multiple smart sensors, mobile devices, intelligent service, cloud robots, and so on, to improve efficiency and scalability in processing of complex tasks that are impossible under local resource constraints in different application areas; (2) intelligent transportation systems, for example, leveraging sensing, computing, and communication capabilities with control vehicles in the physical world to deal with the challenges of safe (e.g., decreasing time latency of reaction in traffic accidents), efficient, and green transportations; and (3) smart living technologies, for example, smart city, environmental monitoring, healthcare systems, and smart grid, which improve intelligence, convenience, operational safety, and green energy of human society. The purpose of this special issue aims to involve multidiscipline research contributions of unpublished research on the recent methodologies, innovations, and experimental validations regarding mobile CPS, so as to facilitate the real-world deployment of such systems in the future. In this special issue on mobile CPS, we have invited a few papers that achieve such goal.

The paper "Sliding Window Based Feature Extraction and Traffic Clustering for Green Mobile Cyber-Physical Systems" focuses on a green mobile cyber-physical system to ensure network coverage and to reduce the total energy consumption, which proposes a feature extraction method using sliding window to extract the distribution feature of mobile user equipment (UE). Furthermore, traffic clustering analysis and corresponding optimized control strategy are presented for the rapid control of base stations. Experimental results demonstrate the superior performance of the proposed method in terms of UE coverage compared with the grid method.

The paper "An Anonymous Access Authentication Scheme Based on Proxy Ring Signature for CPS-WMNs" proposes a novel anonymous access authentication scheme based on proxy ring signature to address the anonymous access authentication issue for CPS-Wireless Mesh Network (CPS-WMN). In the scheme, a hierarchical authentication architecture is presented firstly. Then, intergroup and intragroup anonymous mutual authentication are carried out through proxy ring signature mechanism and certificateless signature mechanism, respectively. A formal security proof with SVO logic and simulation results demonstrate the security and the performance of the proposed scheme.

The paper "Improved Object Proposals with Geometrical Features for Autonomous Driving" aims at generating high quality object proposals for object detection in autonomous driving, which proposes several geometrical features suited for autonomous driving and integrates them into current general proposal generation methods. Experiments over KITTI benchmark demonstrate the significant performance improvement of the existing methods by using the proposed geometrical features.

The paper "A Formal Approach to Verify Parameterized Protocols in Mobile Cyber-Physical Systems" targets the challenging task of protocol verification in mobile CPS and proposes a formal approach to verify the safety properties of parameterized protocols. In this approach, the protocol is modeled as a Petri net using counterabstraction; then a verification algorithm is presented to verify the Petri net model. Experimental results show higher capability but lower memory consumption compared with existing approaches.

The paper "Design and Voluntary Motion Intention Estimation of a Novel Wearable Full-Body Flexible Exoskeleton Robot" proposes to use steel wire as the flexible transmission medium and design a group of special wire locking structures. Moreover, passive points for partial joints of the exoskeleton are designed and a novel gait phase recognition method for full-body exoskeletons is proposed. Experimental results demonstrate the performance of the proposed method in terms of high correct rates of motion pattern classification and phase recognition.

The paper "Efficient Physical Layer Secret Key Generation and Authentication Schemes Based on Wireless Channel-Phase" focuses on the inefficiency of current physical-layer secret key generation schemes and the unpractical assumption of the pre-known knowledge of the shared key in existing authentication schemes. Then, a novel physical-layer secret key generation scheme and an authentication scheme for Orthogonal Frequency-Division Multiplexing (OFDM) systems are proposed, which both exploit the randomness and reciprocity of the channel-phase response. Simulation results demonstrate the superior performance of the proposed schemes under the NIST randomness test and various attacks.

Xiping Hu
Jun Cheng
Xitong Li
Wei Tan
Qiang Liu
Zhengguo Sheng

Hindawi
Mobile Information Systems
Volume 2017, Article ID 7393526, 13 pages
https://doi.org/10.1155/2017/7393526

Research Article

Efficient Physical-Layer Secret Key Generation and Authentication Schemes Based on Wireless Channel-Phase

Longwang Cheng,[1] Li Zhou,[1,2] Boon-Chong Seet,[3] Wei Li,[1] Dongtang Ma,[1] and Jibo Wei[1]

[1]*School of Electronic Science and Engineering, National University of Defense Technology, Changsha, China*
[2]*Science and Technology on Information Transmission and Dissemination in Communication Networks Laboratory, Shijiazhuang, China*
[3]*Department of Electrical and Electronic Engineering, Auckland University of Technology, Auckland, New Zealand*

Correspondence should be addressed to Li Zhou; zhouli2035@nudt.edu.cn

Received 3 February 2017; Accepted 30 April 2017; Published 10 July 2017

Academic Editor: Wei Tan

Exploiting the inherent physical properties of wireless channels to complement or enhance the traditional security mechanisms has attracted prominent attention recently. However, the existing secret key generation schemes suffer from miscellaneous extracting procedure. Many PHY-layer authentication schemes assume that the knowledge of the shared key is preknown. In this paper, we propose PHY-layer secret key generation and authentication schemes for orthogonal frequency-division multiplexing (OFDM) systems. In the secret key generation scheme, to simplify the extracting procedure, only one legitimate party is chosen to probe the channel and quantize the measurements to obtain the preliminary key. The preliminary key is masked by the channel-phase after the mapping and before equalization and distributed to the other party. The final shared key is used for the PHY-layer authentication scheme in which random signals and the shared key masked by the channel-phase are exchanged at the PHY-layer. Then, a binary hypothesis test is formulated for authentication. Simulation results show that the proposed secret key generation scheme outperforms the existing schemes. For the PHY-layer authentication scheme, it is immune to various passive and active attacks and a high successful authentication rate is acquired even at low signal-to-noise ratio region.

1. Introduction

With the continuous development of the wireless communications, people pay more and more attention to the security issue. The security mechanisms of traditional communications networks mainly rely on symmetric or asymmetrical encryption algorithms to achieve confidentiality and authentication. However, due to the lack of key management infrastructures and limited resources of the devices, the conventional security mechanisms may be inapplicable in wireless communications. In addition, the broadcasting nature of the wireless channels causes the wireless communication channel easily to be eavesdropped on or intercepted by an adversary [1, 2]. Therefore, the interest in exploiting the characteristics of the wireless channels at the PHY-layer to enhance and complement the conventional security mechanisms is growing, such as the secret key extraction from the characteristics of wireless channels [3–19] and PHY-layer authentication [20–27].

From an information-theoretic perspective, the authors of [3, 4] demonstrated that it is possible to extract secret key bits from the correlated random sources. Fortunately, with the properties of randomness, location-specific, and reciprocity, the wireless channels can be seen as natural correlated random sources. In [5], Hassan et al. firstly introduced the idea of generating secret keys from the characteristics of the wireless channels. Since then, many investigators pay attention to extract secret keys from received signal strength (RSS) [6–9], since the RSS is easy to acquire from the off-the-shelf wireless cards. However, these methods suffer from scalability and low secret key generation rate (KGR) which is defined as the average amount of secret key bits produced in one measurement/second [10–12]. To resolve these issues, researchers exploit the channel state information (CSI) to

extract secret keys [13–15]. Besides, to increase the KGR, the orthogonal frequency-division multiplexing (OFDM) [18, 19] and multi-input-multi-output (MIMO) techniques in the PHY-layer are adopted.

Various efforts also have been made towards PHY-layer authentication, which can be recognized as a complement or an enhancement to the higher layer authentication mechanisms. In general, according to whether a shared secret key between the legitimate parties is utilized to authenticate each other or not, the existing PHY-layer authentication schemes can be divided into key based or keyless [20]. In some practical cases, it might be difficult to implement the keyless authentication schemes [21–23]. This is because the features of either the transmitting device or the specific channel between the legitimate users, which are exploited to authenticate the transmission, are required to be identified. Instead, the authentication schemes based on the shared key between two legitimate users [25–27] are closer to the conventional challenge-response authentication protocols. In [25], the specific spatial and temporal multipath fading channel between the transmitter and the receiver was exploited for an authentication algorithm. The authors of [26] proposed a PHY-layer challenge-response authentication mechanism (PHY-CRAM) where the randomness and reciprocity of the wireless signal amplitude are exploited for authentication. In [27], a PHY-layer phase challenge-response authentication scheme (PHY-PCRAS) was proposed. It exploited the randomness and reciprocity of the channel-phase response to protect shared key from possible eavesdropping and achieve authentication.

However, the existing secret key generation schemes suffer from miscellaneous extracting procedure which may lead to a high secret key bits mismatched rate (BMR, defined as the ratio of the number of bits unmatched between Alice and Bob's preliminary keys to the number of the preliminary key [7]) and an inefficiency in secret key generation. The key based authentication schemes assume that the knowledge of the shared key is preknown between the authenticated parties, but how to implement the secret key distribution is not given. In this paper, we propose PHY-layer secret key generation and authentication schemes for OFDM systems. Both schemes exploit the randomness and reciprocity of the channel-phase response that is very sensitive to the distance between the legitimate parties. In the secret key generation scheme, to simplify the extracting procedure, only one legitimate party is chosen to probe the channel and quantize the measurements to obtain the preliminary key. After mapping and before equalizing, the preliminary key is masked in the channel-phase and then distributed to the other legitimate party. The final shared key is used for the PHY-layer authentication scheme in which random signals and the generated shared key masked by the channel-phase through mapping and before equalizing are exchanged at the physical layer. Then, a binary hypothesis test is formulated for the authentication procedure.

The security strength of our proposed schemes relies heavily on the randomness of the fading channel and the relative geographic location of the attacker and legitimate users, because the channel-phase response is sensitive to the distance between the legitimate parties. That is, even when the attacker's computational power is increased, the security of our schemes is guaranteed.

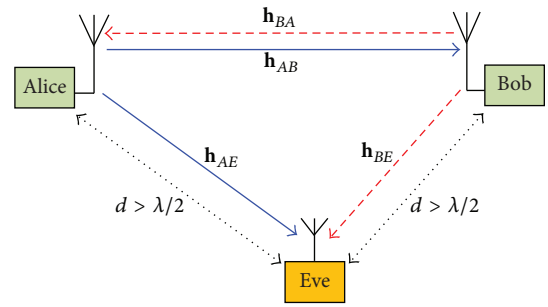


Figure 1: The system model.

The major contributions of this paper are summarized as follows:

(i) To simplify the secret key extraction procedure, we propose a secret key generation scheme based on the channel-phase response in which only one node is chosen to generate the preliminary key and further the preliminary key is masked by the channel-phase after mapping and before equalizing and distributed to the other node.

(ii) Extensive simulations are conducted to compare the secret key generation performance of the prior works with the proposed scheme and the security of the keys is evaluated under passive attack.

(iii) With the aid of the secret keys extracted in the proposed secret key generation scheme, we propose a PHY-layer authentication scheme based on the channel-phase response in which the shared key masked by the channel-phase is exchanged at the PHY-layer.

(iv) The judgement of the authentication is transformed into a binary hypothesis test and the security strength is analyzed under various types of attacks.

The remainder of this paper is organized as follows. In Section 2, we introduce the system model for the two proposed schemes. The procedure and performance of the proposed secret key generation scheme are analyzed in detail in Section 3. In Section 4, the proposed PHY-layer authentication is presented and the security and performance of the scheme are evaluated. Finally, concluding remarks are made in Section 5.

2. System Model

2.1. System Model Description. As shown in Figure 1, an OFDM network with three nodes, in which Alice and Bob are legitimate nodes and Eve is an adversary, is considered. Each node is equipped with a single antenna. All nodes work in half-duplex mode and a time-division duplex (TDD) system is employed. The forward and reverse propagation channels are identical during coherence time by reciprocity.

The distance from Eve to both Alice and Bob is more than $\lambda/2$, where λ is the wavelength of the radio waves; thus the wiretap channels and the legitimate channel are uncorrelated [28].

2.2. The Assumptions. It is assumed that the subcarriers are well separated for ensuring independent fading, which ensures the randomness of the extracted secret keys and the independence of the subchannels.

For the secret key generation case, Alice and Bob want to build a shared secret key. Eve, a passive adversary who tries to obtain the key by eavesdropping, can monitor all the communications during the secret key extraction and neither modify any messages exchanged between Alice and Bob nor jam the legitimate channel. For the PHY-layer authentication case, Alice and Bob try to authenticate each other. Eve is an active attacker who not only can listen to the communications for authentication but also perform various active attacks.

The procedures and parameters of the secret key generation scheme and the PHY-layer authentication scheme adopted by Alice and Bob are assumed to be open to Eve.

2.3. Channel Reciprocity. The principle of the short-term reciprocity of the radio channel is the basis of the two proposed schemes. As discussed in [29], this is guaranteed because, in the real environments compared to channel coherence time T_C, the processing time of channel probing or authentication can be much smaller. For example, we consider the 2.4 GHz radio frequency carrier. For a mobile scenario, the channel variation is mainly due to Doppler effects and when the relative speed between the transmitter and receiver is $v = 60$ km/h, the Doppler frequency is $f_d = vf/c = 16.67 \times 2.4 \times 10^9/(3 \times 10^8) = 133.3$ Hz. Empirically, the channel coherence time T_C which is related to the maximum Doppler frequency shift can be calculated as $T_C = 9/16\pi f_d = 9/(16\pi \times 133.3) = 1.3$ ms. In our proposed schemes, the processing time, which demands for channel coherence, includes double propagation time T_p and transmitting time T_t and one operation delay T_d. For 5 MHz sampling rate, it takes about $T_t = 16$ us to transmit an OFDM symbol with 64 subcarriers and 16 cyclic prefix samples. When the distance is 3000 m, the propagation time $T_P = 10$ us. In general, the transmitting time is in the same order of the operation delay. Then the total processing time $2T_t + 2T_P + T_d$ is much smaller than the coherence time T_c in our two proposed schemes.

3. Secret Key Generation and Performance Analysis

3.1. The Proposed Secret Key Generation Scheme. Compared to single carrier systems, OFDM systems can provide extra randomness in view of the use of multiple subchannels. In this paper, we propose a secret key generation scheme based on the channel-phase response of OFDM systems in frequency domain.

In general, a secret key generation scheme consists of the following four steps:

(1) Channel probing: Alice and Bob alternately and periodically send the probe signals to each other to obtain the characteristics of channel between them.

(2) Measurement quantization: Alice and Bob separately quantize the collected channel characteristics into bit vector to obtain a preliminary secret key bits.

(3) Information reconciliation: due to the nature of half-duplex and noise, a small number of Alice and Bob's preliminary keys may be mismatched. They exchange messages to agree on a synchronized key.

(4) Privacy amplification: since the messages exchanged in the information reconciliation phase are open to Eve, they may be exploited by Eve to infer the generated keys. To address this issue, Alice and Bob apply privacy amplification method to eliminate Eve's partial information about the key and obtain a shared key.

We can find that, for half-duplex mode, the legitimate nodes have to transmit probe signals alternately to characterize the channel, which means they cannot probe the channel simultaneously. After collecting sufficient measurements, they quantize the measurements into preliminary keys separately. These phases may bring estimation error and quantization error, which may lead to many mismatched bits between the preliminary keys generated by the legitimate parties. Thus, the cost to reconcile the mismatched bits is high.

In this paper, to address this problem, we propose a scheme in which only one of the legitimate nodes is chosen to perform the channel probing and measurement quantization. This simplifies the procedure of secret key generation and eliminates the estimation error and quantization error.

Under the principle of reciprocity, we set $\mathbf{h}_{AB} = \mathbf{h}_{BA} = \mathbf{h}$. To maintain the reciprocity requirement, for each of Alice to Bob's channel probes, the corresponding Bob to Alice's channel probe event must be conducted within the coherence time of the channel. The process of the proposed secret key generation scheme is depicted in Figure 2 and the detailed steps are as follows. (Note that throughout this paper, the signals and equations are in frequency domain.)

Step 1 (channel probing). The experiments in [7] revealed that, in certain environments, due to lack of variations in the wireless channel, the extracted bits have very low entropy making these bits unsuitable for a secret key, which can cause predictable key generation by an adversary in these static environments. To prevent this, during channel probing, we utilize random signals to probe the channel. Suppose that Bob is chosen to probe the channel and quantize the channel measurements. Thus, Alice transmits the random probe signal $\mathbf{s}_a = [s_{a,1}, s_{a,2}, \ldots, s_{a,N}]$ to Bob, where $s_{a,i} = \exp(j\theta_{a,i})$, $\theta_{a,i} \sim U[0, 2\pi]$ for $i = 1, 2, \ldots, N$, and N is the number of the subcarriers of the OFDM system. The random signal $\mathbf{s}_a$ is unknown to Bob and Eve.

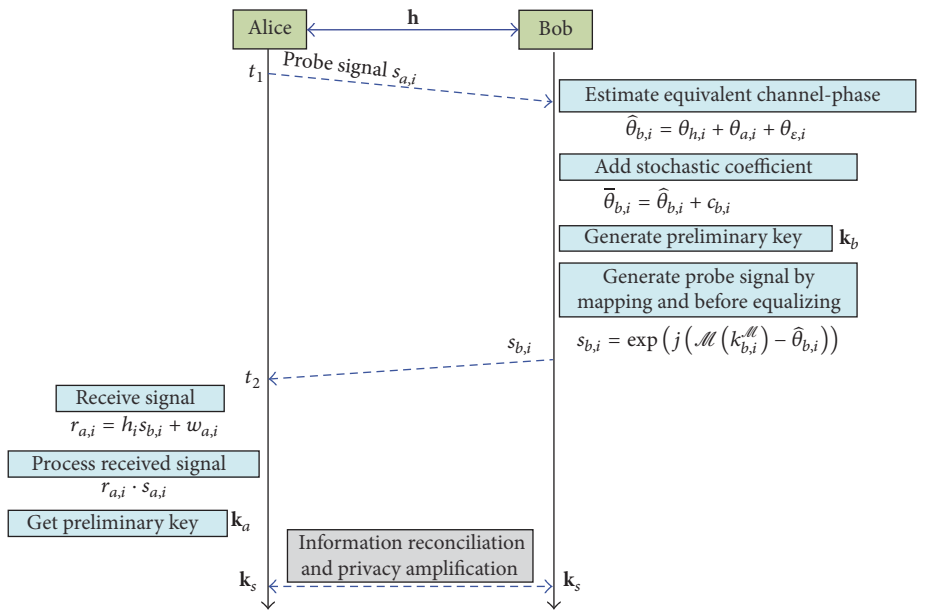


Figure 2: The process of the secret key generation.

Without loss of generality, we only take the ith ($1 \le i \le N$) subcarrier, for example. Thus, the received signal at Bob can be expressed as

$$\begin{aligned} r_{b,i} &= h_i s_{a,i} + w_{b,i} = |h_i| \exp\left(j\theta_{h,i}\right) s_{a,i} + w_{b,i} \\ &= |h_i| \exp\left(j\theta_{h,i} + j\theta_{a,i}\right) + w_{b,i}, \end{aligned} \tag{1}$$

where h_i denotes the ith subchannel response of the legitimate channel in frequency domain and $\theta_{h,i}$ is the underlying subchannel-phase response. The subchannels are independent and identically distributed (i.i.d.) and $h_i \sim \mathcal{CN}(0, \sigma_h^2)$. $w_{b,i}$ is the i.i.d. complex Gaussian noise with zero mean and variance σ_n^2.

Based on the received signal, Bob gets the subchannel-phase response estimation as

$$\widehat{\theta}_{b,i} = \tan^{-1}\left(\frac{\text{imag}\left(r_{b,i}\right)}{\text{real}\left(r_{b,i}\right)}\right) = \theta_{h,i} + \theta_{a,i} + \varepsilon_{b,i}, \tag{2}$$

where $\varepsilon_{b,i}$ is the phase estimation error. Note that the phase of the random probe signal $\theta_{a,i}$ is contained in the subchannel-phase response estimation, so we treat $\widehat{\theta}_{b,i}$ as equivalent subchannel-phase response estimation.

If Eve is in close proximity to Bob (here the "close" means that the distance between Eve and Bob is much smaller than $\lambda/2$, which may lead to highly correlated $\mathbf{h}_{AB}$ and $\mathbf{h}_{AE}$), she may infer the preliminary key easily based on her observations. To reduce the risk, Bob adds a stochastic coefficient to $\widehat{\theta}_{b,i}$ as

$$\overline{\theta}_{b,i} = \widehat{\theta}_{b,i} + c_{b,i}, \tag{3}$$

where $c_{b,i}$ is uniformly distributed over $[0, 2\pi]$ and $\mathbf{c}_b = [c_{b,1}, c_{b,2}, \ldots, c_{b,N}]$. Bob obtains the vector records as $\overline{\boldsymbol{\theta}}_b = [\overline{\theta}_{b,1}, \overline{\theta}_{b,2}, \ldots, \overline{\theta}_{b,N}]$.

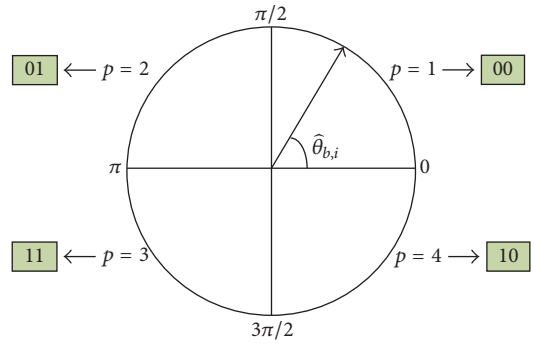


Figure 3: A quantization example with $M = 2$.

Step 2 (measurement quantization and preliminary key distribution). Bob quantizes the vector records $\overline{\theta}_b$ into bit vector to obtain the preliminary key. Firstly, Bob divides the interval $[0, 2\pi)$ into 2^M subintervals, where 2^M is the number of quantization levels, which is bounded by the mutual information between Alice and Bob [30]. Thus each $\overline{\theta}_{b,i}$ can be quantized to M binary bits. The pth ($1 \le p \le 2^M$) subinterval is $[2\pi(p-1)/2^M, 2\pi p/2^M)$. Secondly, gray code is used to assign a binary code word with M bits to each subinterval. A quantization example with $M = 2$ is illustrated in Figure 3.

The ith record can be quantized as

$$\begin{aligned} &Q\left(\overline{\theta}_{b,i}\right) = p_i, \\ &\text{if } \text{mod}\left(\overline{\theta}_{b,i}, 2\pi\right) \in \left[\frac{2\pi\left(p_i - 1\right)}{2^M}, \frac{2\pi p_i}{2^M}\right). \end{aligned} \tag{4}$$

After quantization, Bob obtains the preliminary key as $\mathbf{k}_b = [k_{b,1}, k_{b,2}, \ldots, k_{b,NM}]$.

Then, Bob sends "probe signal" to Alice. Different from the constant probe signal and $\mathbf{s}_a$, this probe signal, in fact, is the preliminary key after mapping and before equalizing. Thus, the probe signal can be expressed as

$$s_{b,i} = \exp\left(j\mathscr{M}\left(k_{b,i}^{\mathscr{M}}\right) - j\widehat{\theta}_{b,i}\right), \tag{5}$$

where $\mathscr{M}(\cdot)$ denotes the mapping operation on the preliminary key and $k_{b,i}^{\mathscr{M}} = [k_{b,(i-1)l+1}, \ldots, k_{b,(i-1)l+l}]$ is the ith input secret key sequence with length l ($Nl \leq NM$). When $l = 2$, a mapping function can be designed as

$$\mathscr{M}\left(k_{b,i}^{\mathscr{M}}\right) = \begin{cases} 0, & k_{b,i}^{\mathscr{M}} = \begin{bmatrix} 0 & 0 \end{bmatrix} \\ \dfrac{\pi}{2}, & k_{b,i}^{\mathscr{M}} = \begin{bmatrix} 0 & 1 \end{bmatrix} \\ \pi, & k_{b,i}^{\mathscr{M}} = \begin{bmatrix} 1 & 1 \end{bmatrix} \\ \dfrac{3\pi}{2}, & k_{b,i}^{\mathscr{M}} = \begin{bmatrix} 1 & 0 \end{bmatrix}. \end{cases} \tag{6}$$

The mapping function is known to Alice. The subtraction term in (5) denotes the preequalization process using the equivalent subchannel-phase response estimation. In fact, the preequalization process can be seen as an encryption operation; thus the preliminary key is masked by the subchannel-phase response estimation.

Alice's received signal can be expressed as

$$\begin{aligned} r_{a,i} &= h_i s_{b,i} + w_{a,i} \\ &= |h_i| \exp\left(j\left(\theta_{h,i} + \mathscr{M}\left(k_{b,i}^{\mathscr{M}}\right) - \widehat{\theta}_{b,i}\right)\right) + w_{a,i}, \end{aligned} \tag{7}$$

where $w_{a,i} \sim \mathscr{CN}(0, \sigma_n^2)$ is the i.i.d. complex Gaussian noise. Based on the reciprocity between the forward and reverse links and substituting $\widehat{\theta}_{b,i}$ with (2), (7) can be simplified as

$$r_{a,i} = |h_i| \exp\left(j\left(\mathscr{M}\left(k_{b,i}^{\mathscr{M}}\right) - \theta_{a,i} - \varepsilon_{b,i}\right)\right) + w_{a,i}. \tag{8}$$

As observed in (8), during the receiving, Alice completes the subchannel-phase equalization and eliminates the encryption based on the channel reciprocity. Alice further multiplies (8) by the random probe signal $s_{a,i}$ and gets

$$s_{a,i} r_{a,i} = |h_i| \exp\left(j\left(\mathscr{M}\left(k_{b,i}^{\mathscr{M}}\right) - \varepsilon_{b,i}\right)\right) + s_{a,i} w_{a,i}. \tag{9}$$

Then, Alice performs unmapping on the phase of $s_{a,i} r_{a,i}$ to acquire the preliminary key transmitted by Bob.

In conclusion, in this step, Bob firstly obtains the preliminary key by quantizing the vector records and randomly chooses Nl key bits from the preliminary key as the input key sequences of the mapping function. Secondly, these key sequences are mapped, preequalized, and transmitted to Alice. Lastly, Alice acquires these key sequences based on the channel reciprocity. So the preliminary key is distributed from Bob to Alice.

Step 3 (information reconciliation and privacy amplification). Note that, in our scheme, we assume that the length of the preliminary key is NM. In practical systems, this length may be much longer, which in turn may require more rounds of channel probing and secret key distribution. Alice and Bob need to update the random probe signal vector $\mathbf{s}_a$ and the stochastic coefficient vector $\mathbf{c}_b$, respectively, after each round.

Due to the noise, a small number of mismatched bits may exist in the preliminary keys of Alice and Bob. Then, the mismatched bits are reconciled by using BCH codes to get synchronized keys. The privacy of the synchronized keys is subsequently enhanced by using a hash function to obtain a secure and common key.

During the secret key generation process, Alice and Bob should do Steps 1 and 2 fast enough to ensure that $t_2 - t_1$ is not more than the coherence time. We can observe that due to the random probe signal, the randomness of the channel is ensured even if the environments are static. So it also can address the highly correlated and unsecure key bits problem in stationary environments [7].

3.2. Performance Analysis. In this subsection, we will analyze the proposed secret key generation scheme and evaluate its performance in terms of the secret key capacity, bits mismatched, and key generation rates.

3.2.1. Security Analysis. Eve is a passive attacker and only can listen to the communications during secret key generation. For ease of analysis, we neglect the effect of noise in Step 1 so that Eve's received signal from Alice is

$$r_{ea,i} = h_{AE,i} s_{a,i} = |h_{AE,i}| \exp\left(j\theta_{hAE,i} + j\theta_{a,i}\right), \tag{10}$$

where $h_{AE,i} = |h_{AE,i}| e^{j\theta_{hAE,i}}$ is the ith subchannel from Alice to Eve. In Step 2, Eve's received signal from Bob can be expressed as

$$\begin{aligned} r_{eb,i} &= h_{BE,i} s_{b,i} = |h_{BE,i}| \\ &\cdot \exp\left(j\left(\theta_{hBE,i} + \mathscr{M}\left(k_{b,i}^{\mathscr{M}}\right) - \widehat{\theta}_{b,i}\right)\right) = |h_{BE,i}| \\ &\cdot \exp\left(j\left(\theta_{hBE,i} + \mathscr{M}\left(k_{b,i}^{\mathscr{M}}\right) - \theta_{h,i} - \theta_{a,i} - \varepsilon_{b,i}\right)\right), \end{aligned} \tag{11}$$

where $h_{BE,i} = |h_{BE,i}| e^{j\theta_{hBE,i}}$ is the ith subchannel from Bob to Eve. We can find that the factors which influence Eve to derive the key bits are the phases of h_i, $h_{AE,i}$, $h_{BE,i}$, and $s_{a,i}$, that is, $\theta_{h,i}$, $\theta_{hAE,i}$, $\theta_{hBE,i}$, and $\theta_{a,i}$. Besides, the stochastic coefficient $c_{b,i}$ also impairs Eve's inference to some extent.

To reduce the factors, Eve can multiply (10) by (11) and obtains

$$\begin{aligned} r_{eb,i} r_{ea,i} &= |h_{BE,i} h_{AE,i}| \\ &\cdot \exp\left(j\left(\theta_{hBE,i} + \theta_{hAE,i} + \mathscr{M}\left(k_{b,i}^{\mathscr{M}}\right) - \theta_{h,i} - \varepsilon_{b,i}\right)\right). \end{aligned} \tag{12}$$

Then, the factors are reduced to $\theta_{h,i}$, $\theta_{hAE,i}$, and $\theta_{hBE,i}$. Note that since the phase of the signals transmitted by Alice and Bob in Steps 1 and 2 is random, it is hard for Eve to estimate the phases of $\theta_{hAE,i}$ and $\theta_{hBE,i}$. Thus, it is difficult for Eve to derive the generated keys and we will analyze various cases in the following.

Firstly, both Alice and Bob are far away from Eve, so that the wiretap channels (i.e., h_{AE} and h_{BE}) and the legitimate

channel (i.e., h) are uncorrelated, along with the random signal $\mathbf{s}_a$ and stochastic coefficient $\mathbf{c}_b$; for Eve, it is almost impossible to obtain the generated secret keys from her measurements.

Then, an aggressive case, where Eve is close to Bob, is considered. In this case, $h_{AE,i} \approx h_{AB,i} = h_i$. Then (10) can be approximately simplified as

$$r_{ea,i} = |h_i| \exp\left(j\theta_{h,i} + j\theta_{a,i}\right). \quad (13)$$

Then the phase of $r_{ea,i}$ is approximately equal to Bob's equivalent subchannel-phase response estimation $\widehat{\theta}_{b,i}$. So for Eve it is possible to infer the preliminary key by the same quantization approach. However, due to the random coefficient $\mathbf{c}_b$, which is unknown to Eve, the probability of obtaining the key based on $r_{ea,i}$ is low. In (11) and (12), since $h_{BE,i}$ is uncorrelated with $h_{BA,i}$, for Eve it is improbable to derive the distributed preliminary key.

Lastly, we consider that Eve is close to Alice. Under this circumstance, $h_{BE,i} \approx h_{BA,i} = h_i$, so (11) becomes

$$r_{eb,i} = |h_{BE,i}| \exp\left(j\left(\mathscr{M}\left(k_{b,i}^{\mathscr{M}}\right) - \theta_{a,i} - \varepsilon_{b,i}\right)\right). \quad (14)$$

Since $\theta_{a,i}$ is random and unknown to Eve, she cannot infer the key based on $r_{eb,i}$. In this situation, $h_{AE,i}$ is uncorrelated with $h_{AB,i}$, so it is obvious that Eve cannot obtain the key based on (10) and (12).

In conclusion, when Eve is a passive attacker, the secret key cannot be derived only by her observations and Alice and Bob can still establish a secure key. In fact, the channel-phase response is sensitive to the distance between Alice and Bob, so for Eve it is more difficult to infer the secret key bits from her measurements of the channel-phase. In addition, in [31], the authors pointed out that applying error-correcting codes on the preliminary key with reasonable rate can ensure the correct preliminary key for Alice in theory, while ensuring useless information for Eve. It means that we can design an error-correcting code with proper rate to further ensure the security of the proposed scheme.

3.2.2. The Secret Key Generation Performance. Firstly, the secret key capacity which is defined as the maximum available key generation rate [3] is considered. Since the subchannels are independent, for ease of analysis, we only take one of the subchannels, for example. Our proposed scheme can be approximatively modeled as Bob generating a random source $X = h$ and Alice observing the random source as $Y = X + W_a = h + W_a$, where $h \sim \mathscr{CN}(0, \sigma_h^2)$ is one of the subchannel responses and $W_a \sim \mathscr{CN}(0, \sigma_w^2)$ is the observed noise. Alice and Bob extract secret keys from the phases of X and Y, that is, θ_X and θ_Y, respectively. Assuming that Eve's observations are uncorrelated with Alice's, the secret key capacity can be expressed as [3]

$$C_P = I\left(\theta_X; \theta_Y\right), \quad (15)$$

where symbol $I(\cdot;\cdot)$ denotes the mutual information between two random variables. Since the joint probability density function of θ_X and θ_Y is difficult to calculate, we adopt the information theoretical estimators (ITE) toolbox to estimate the secret key capacity [32]. The input of the ITE is $(\boldsymbol{\theta}_X; \boldsymbol{\theta}_Y)$, where $\boldsymbol{\theta}_X = [\theta_{X,1}, \theta_{X,2}, \ldots, \theta_{X,k}]$, $\boldsymbol{\theta}_Y = [\theta_{Y,1}, \theta_{Y,2}, \ldots, \theta_{Y,k}]$, and k is the length of the input. In the simulations, the signal-to-noise ratio (SNR) is defined as σ_h^2/σ_w^2.

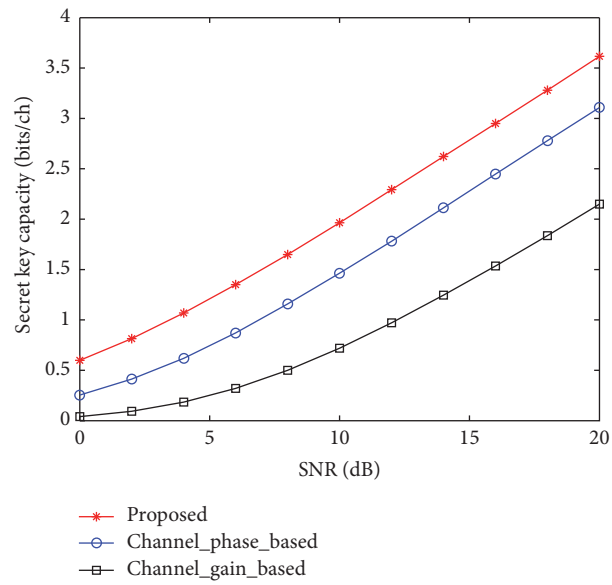


Figure 4: The secret key capacity of different schemes.

The secret key capacity of the proposed secret key generation scheme is compared with the existing channel-phase based secret key generation scheme [15] and channel-gain based secret key generation scheme [14] in Figure 4. During channel probing, the channel condition of these three schemes is identical. We can clearly observe that, in contrast to the channel-phase based and channel-gain based schemes, our proposed scheme achieves a greater secret key capacity. For example, when SNR is 10 dB, the secret key capacity of the proposed scheme is 34% and 170% greater than the channel-phase based scheme and channel-gain based scheme, respectively. Note that the secret key capacity of the discussed OFDM systems is N times of those shown in Figure 4.

Secondly, we analyze the secret key bits mismatched and key generation rates. The BMR and KGR of the proposed scheme are evaluated through Monte-Carlo simulations under multipath channels and further are compared with the channel-phase based and channel-gain based schemes. Since it does not need to estimate the CSI during channel probing, in the simulations, the probe signal contains two OFDM symbols, that is, one pilot symbol for synchronization and one symbol for signal phase estimation. The carrier frequency of the OFDM system is 2.4 GHz and the number of the subcarriers is $N = 64$. We consider the multipath Rayleigh fading channel with 2 us constant delay time and the maximum Doppler frequency is 0 Hz (suppose that Alice and Bob remain static in the simulations). The sample interval is 0.25 us.

These three schemes adopt the same quantization method, that is, equal interval quantization method, in

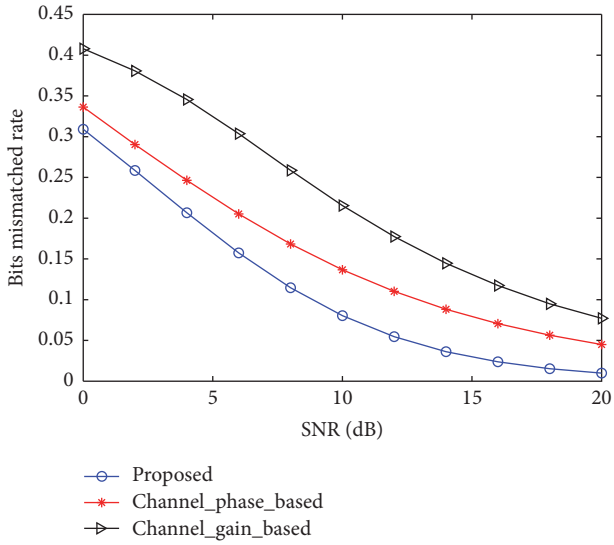


Figure 5: The BMR performance of different schemes.

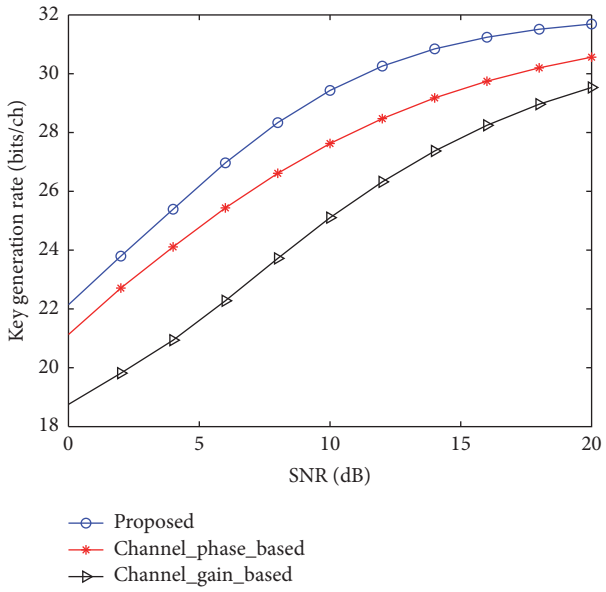


Figure 6: The KGR performance of different schemes.

which the characteristics space is divided on average into $2^M = 4$ subspaces. The same information is reconciled and privacy amplification approaches are employed. Note that, theoretically, the bit length of the resulting quantization should be bounded by the mutual information between Alice and Bob [33]. In other words, the quantization level 2^M should not be higher than the secret key capacity, that is, $2^M \leq 2^{C_P}$. However, for ease of analysis, the quantization levels for the three schemes are all set as $2^M = 4$.

Figures 5 and 6 show the BMR and KGR performance of these schemes, respectively. From these two figures, we can observe that the BMRs of these schemes decrease, while the KGRs increase as SNR increases. The primary reason is that the accuracy of the channel estimates obtained in the channel probing phase increases as SNR increases. The BMR and KGR performances of the proposed scheme exhibit apparent superiority to the other schemes. For example, when SNR = 10 dB, the BMR and KGR of the proposed scheme decreases by 41% and increases 6.6%, respectively, compared to the channel-phase based scheme.

Table 1: The evaluation of randomness test.

Test	P value
Frequency	0.65
Block Frequency	0.55
Cumulative sum (Rev)	0.30
Cumulative sum (Fwd)	0.64
Approximate entropy	0.69
Runs	0.48
Longest run	0.54
Serial	0.72 0.55

Lastly, the randomness of the secret key is analyzed. A cryptographic key should be substantially random; otherwise, an adversary can crack the key with low cost. A widely used randomness test suite NIST [34] is employed to verify the randomness of our generated secret key bits. The NIST test suite is a statistical package consisting of 16 tests which were developed to test the randomness of binary sequences. To pass the test, all the P values of the 16 tests should be at least greater than 0.01. We randomly select 10-bit sequences generated from our simulations at SNR = 10 dB. Due to the limitation of bit length, we run eight typical tests. The results in Table 1 shows that our generated bit sequences pass the NIST test and their average entropy is close to that of a truly random sequence.

4. PHY-Layer Authentication Scheme and Performance Analysis

4.1. PHY-Authentication Scheme. In the proposed secret key generation scheme, the legitimate parties establish a shared key which can be used for the encryption and authentication. The existing key based authentication schemes assume that the knowledge of the shared key is preknown between the authenticated parties, but how to achieve the secret key distribution is not given. Consider that the shared key between the authenticated parties is generated by our proposed secret key generation scheme.

After establishing a shared key and a period of time without communication, if Alice and Bob want to establish a communication, they need to authenticate each other. In this section, we propose a challenge-response PHY-layer authentication scheme for OFDM systems, which exploits the short-term reciprocity and randomness of the channel-phase response in TDD mode. Generally, Alice and Bob need a two-way authentication process to achieve the mutual authentication. However, the one-way authentication process is enough to describe the process, since both directions of the two-way authentication employ the same regulation. We assume that Alice wants to communicate with Bob, who

in turn needs to verify the identity of "Alice" based on the proposed challenge-response PHY-layer authentication scheme (here Bob is assumed to be legitimate, while, for the illegitimate Bob, it will be analyzed later). The process of the proposed PHY-layer authentication scheme is shown in Figure 7 and the detailed stages are as follows.

Stage 1. Alice transmits an authentication request signal to Bob. The authentication request signal contains the frame type, time stamp information, media access control address, and so forth.

Stage 2. After receiving the authentication request, Bob contemplates that "Alice" wants to communicate with himself. Then Bob generates a response signal vector $\mathbf{s}_b = [s_{b,1}, s_{b,2}, \ldots, s_{b,N}]$ and sends to "Alice," where $s_{b,i} = \exp(j\theta_{b,i})$ and $\theta_{b,i} \sim U[0, 2\pi]$ for $i = 1, 2, \ldots, N$. N is the number of the subcarriers. The random response signal vector $\mathbf{s}_b$ is unknown to Alice and Eve.

Stage 3. The received signal of the ith subcarrier in frequency domain at Alice is

$$\begin{aligned} r_{a,i} &= h_{BA,i} s_{b,i} + w_{b,i} \\ &= \left|h_{BA,i}\right| \exp\left(j\theta_{hBA,i} + j\theta_{b,i}\right) + w_{a,i}, \end{aligned} \tag{16}$$

where $h_{BA,i}$ denotes the ith subchannel response from Bob to Alice and $\theta_{hBA,i}$ is the underlying subchannel-phase response. The subchannels are i.i.d. and $h_{BA,i} \sim \mathcal{CN}(0, \sigma_h^2)$. $w_{a,i}$ is the i.i.d. complex Gaussian noise with zero mean and variance σ_n^2. Alice is not concerned with what Bob transmits but only estimates the phase of the received signal. The estimation can be expressed as

$$\widehat{\theta}_{a,i} = \tan^{-1}\left(\frac{\operatorname{imag}\left(r_{a,i}\right)}{\operatorname{real}\left(r_{a,i}\right)}\right) = \theta_{hBA,i} + \theta_{b,i} + \varepsilon_{a,i}, \tag{17}$$

where $\varepsilon_{a,i}$ is the phase estimation error.

Then, to generate a tagged signal vector $\mathbf{s}_a$ for authentication, Alice processes her shared secret key by using the mapping function and preequalizes the mapped key by subtracting the phase estimation vector $\widehat{\boldsymbol{\theta}}_a$ (here the length of the shared key is assumed to be long enough). The tagged signal at the ith subcarrier is

$$s_{a,i} = \exp\left(j\mathscr{M}\left(k_{a,i}^{\mathscr{M}}\right) - j\widehat{\theta}_{a,i}\right). \tag{18}$$

Alice sends the tagged signal to Bob. As processed in (18), the secret key for authentication is masked by phase estimation and it is difficult for a passive attacker to crack the authenticated secret key.

Stage 4. Bob's received signal is

$$\begin{aligned} r_{b,i} &= h_{AB,i} s_{a,i} + w_{b,i} \\ &= \left|h_{AB,i}\right| \exp\left(j\left(\theta_{hAB,i} + \mathscr{M}\left(k_{a,i}^{\mathscr{M}}\right) - \widehat{\theta}_{a,i}\right)\right) + w_{b,i}, \end{aligned} \tag{19}$$

where $h_{AB,i}$ denotes the ith subchannel response from Alice to Bob, and $\theta_{hAB,i}$ is the underlying subchannel-phase response. $w_{b,i} \sim \mathcal{CN}(0, \sigma_n^2)$ is the i.i.d. complex Gaussian noise. Alice and Bob perform these steps fast enough to ensure the time interval from Stages 2–4 is smaller than the coherence time; thus $h_{AB,i} = h_{BA,i} = h_i$. Then (19) can be simplified as

$$r_{b,i} = \left|h_i\right| \exp\left(j\left(\mathscr{M}\left(k_{a,i}^{\mathscr{M}}\right) - \theta_{b,i} - \varepsilon_{a,i}\right)\right) + w_{b,i}. \tag{20}$$

We can find that the channel-phase equalization has been completed during the receiving. Bob multiplies $s_{b,i}$ by his response signal $r_{b,i}$ and gets $\mathbf{y} = \mathbf{r}_b \odot \mathbf{s}_b$, where $\odot$ denotes element-wise multiplication.

The ith element of $\mathbf{y}$ can be expressed as

$$y_i = r_{b,i} s_{b,i} = \left|h_i\right| \exp\left(j\left(\mathscr{M}\left(k_{a,i}^{\mathscr{M}}\right) - \varepsilon_{a,i}\right)\right) + w_{b,i} s_{b,i}. \tag{21}$$

Then Bob obtains the signal y_i which only contains the mapped secret key from Alice and estimation error. Based on y_i, combining his shared key, Bob needs to judge whether the other party is Alice or not. There are two solutions for this judgement. A straightforward solution is to check the difference between the obtained authenticated key $\mathbf{k}_a^{\mathscr{M}}$ from "Alice" and his own secret key $\mathbf{k}_b^{\mathscr{M}}$, where $\mathbf{k}_a^{\mathscr{M}} = \mathscr{M}^{-1}(\angle \mathbf{y})$. The difference is defined as $D = \operatorname{sum}(\mathbf{k}_a^{\mathscr{M}} \oplus \mathbf{k}_b^{\mathscr{M}})$, where $\oplus$ is the XOR operation. If $D < D_0$, Bob determines that the other party is Alice; otherwise it is not, where D_0 is a constant real number. However, it is hard to determine D_0 in practical systems. Thus, we provide another solution in which the authentication judgement is formulated as a binary hypothesis test.

From (21), we can find that the phase of y_i is mainly affected by the mapped authenticated key. In order to eliminate the influence of the authenticated key, similar to [27], we generate a variable with the expression as

$$C = \left|e^{-j\mathscr{M}\left(\mathbf{k}_b^{\mathscr{M}}\right)} \mathbf{y}^T\right|, \tag{22}$$

where $()^T$ denotes transpose operation. Thus, the binary hypothesis test can be expressed as

$$\begin{aligned} \mathscr{H}0\text{: } \mathbf{k}_x^{\mathscr{M}} &= \mathbf{k}_e^{\mathscr{M}}, \\ \mathscr{H}1\text{: } \mathbf{k}_x^{\mathscr{M}} &= \mathbf{k}_b^{\mathscr{M}}, \end{aligned} \tag{23}$$

where $\mathbf{k}_x^{\mathscr{M}}$ denotes the authenticated key possessed by "Alice." For $\mathscr{H}0$ and $\mathscr{H}1$, the corresponded $C_{\mathscr{H}0}$ and $C_{\mathscr{H}1}$ are

$$\begin{aligned} C_{\mathscr{H}0} &= \left|\sum_{i=1}^{N}\left(\left|h_i\right| e^{j\left(\mathscr{M}\left(k_{e,i}^{\mathscr{M}}\right) - \mathscr{M}\left(k_{b,i}^{\mathscr{M}}\right) - \varepsilon_{a,i}\right)} + e^{-j\mathscr{M}\left(k_{b,i}^{\mathscr{M}}\right)} w_{b,i} s_{b,i}\right)\right|, \\ C_{\mathscr{H}1} &= \left|\sum_{i=1}^{N}\left(\left|h_i\right| e^{j\left(-\varepsilon_{a,i}\right)} + e^{-j\mathscr{M}\left(k_{b,i}^{\mathscr{M}}\right)} w_{b,i} s_{b,i}\right)\right|. \end{aligned} \tag{24}$$

Based on (24), Bob makes a final decision by comparing with a threshold T.

$$C \underset{\mathscr{H}0}{\overset{\mathscr{H}1}{\gtrless}} T. \tag{25}$$

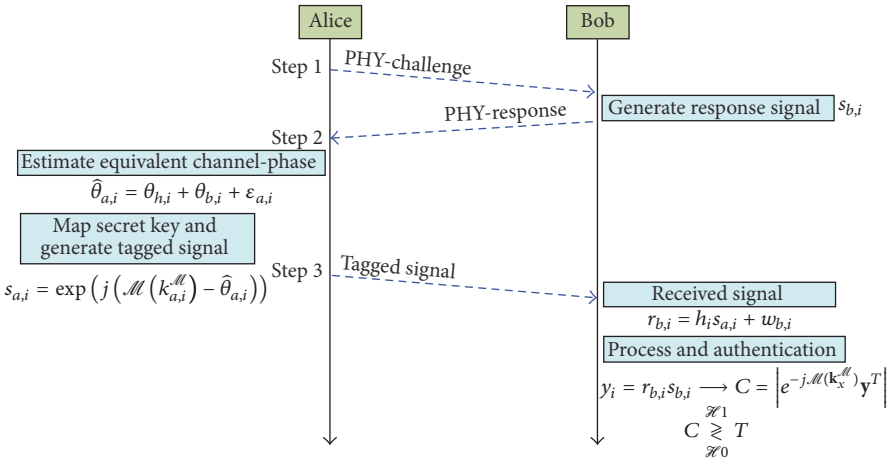


FIGURE 7: The process of the PHY-layer authentication.

If $\mathcal{H}1$ is true, then Bob judges that the other party is Alice; otherwise, it is not.

Now, it is essential to find the threshold T. We can see that, in both hypotheses, $e^{-j\cdot\mathcal{M}(\mathbf{k}_b^{\mathcal{M}})}\mathbf{y}^T$ is the sum of N dependent normally distributed random variables. The resulting sum is still normally distributed [27]; thus its amplitude C obeys Rice distribution. The probability density function of C is

$$f_{C_{\mathcal{H}i}}(x) = \frac{x}{\sigma_{\mathcal{H}i}^2}\exp\left(-\frac{x^2+\overline{C}_{\mathcal{H}i}^2}{2\sigma_{\mathcal{H}i}^2}\right) I_0\left(\frac{x\overline{C}_{\mathcal{H}i}}{\sigma_{\mathcal{H}i}^2}\right), \quad (26)$$

where $x \geq 0$, $i = 0, 1$. $\overline{C}_{\mathcal{H}i}$ and $\sigma_{\mathcal{H}i}^2$ denote the mean and variance of $C_{\mathcal{H}i}$, respectively. $I_0(\cdot)$ is the zero-order modified Bessel function of the first kind. Based on $f_{C_{\mathcal{H}i}}(x)$, we can calculate the false acceptance rate (the rate that the attacker passes the authentication) as

$$\begin{aligned} P_f &= \int_T^{+\infty} f_{C_{\mathcal{H}0}}(x)\,dx \\ &= \int_T^{+\infty} \frac{x}{\sigma_{\mathcal{H}0}^2}\exp\left(-\frac{x^2+\overline{C}_{\mathcal{H}0}^2}{2\sigma_{\mathcal{H}0}^2}\right) I_0\left(\frac{x\overline{C}_{\mathcal{H}0}}{\sigma_{\mathcal{H}0}^2}\right) dx \\ &= \int_{T/\sigma_{\mathcal{H}0}}^{+\infty} \frac{x}{\sigma_{\mathcal{H}0}}\exp\left(-\frac{\left(x/\sigma_{\mathcal{H}0}\right)^2+\left(\overline{C}_{\mathcal{H}0}/\sigma_{\mathcal{H}0}\right)^2}{2}\right) \\ &\quad \cdot I_0\left(\frac{x}{\sigma_{\mathcal{H}0}}\frac{\overline{C}_{\mathcal{H}0}}{\sigma_{\mathcal{H}0}}\right) d\left(\frac{x}{\sigma_{\mathcal{H}0}}\right) = Q\left(\frac{\overline{C}_{\mathcal{H}0}}{\sigma_{\mathcal{H}0}}, \frac{T}{\sigma_{\mathcal{H}0}}\right), \end{aligned} \quad (27)$$

where $Q(a,b) = \int_b^{+\infty} x\exp(-(x^2+a^2)/2) I_0(ax)dx$. $Q(\cdot,\cdot)$ is Marcum Q function. Thus, for a given false acceptance rate P_f, the threshold value T can be calculated by (27). Furthermore, we can get the successful authenticate rate (the rate that the legitimate user passes the authentication) as

$$P_d = \int_T^{+\infty} f_{C_{\mathcal{H}1}}(x)\,dx = Q\left(\frac{\overline{C}_{\mathcal{H}1}}{\sigma_{\mathcal{H}1}}, \frac{T}{\sigma_{\mathcal{H}1}}\right). \quad (28)$$

4.2. Security Analysis. To evaluate the security of the proposed scheme, in this section, we analyze various types of attackers.

Eve, as the adversary, knows Alice and Bob's PHY-layer authentication scheme. When Eve is a passive attacker, she only can listen to all the communications inside the network and attempts to learn the shared key from the information that she eavesdropped. In Section 3.2, we have analyzed that it is almost impossible for Eve to crack and infer the shared key during the secret key generation. Thus, during the authentication, it is also difficult for Eve to derive the shared key and pass the authentication as a passive attacker, and the analysis process is similar. Therefore, we mainly consider the case that Eve is an active attacker. When Eve is an active attacker, she can perform three types of attacks, namely, impersonation attacks, jamming attacks, and replay attacks.

4.2.1. Impersonation Attacks. Eve can impersonate Alice or Bob under impersonation attacks. If Eve initiates Stage 1 (sends authentication request to Bob), she can hardly succeed. The reason is that Bob's response contains no information about the shared key in Stage 2. If Eve impersonates Alice in Stage 3 and sends a tagged signal to Bob, she will not be authenticated by Bob as she has no information about the authenticated key. Compared to the other two stages, Stage 2 is more vulnerable, since, during Stage 1, Alice does not know the legitimacy of its counterpart. In this case, Eve impersonates Bob and may steal the authenticated key from the tagged signal of Alice. To solve this problem, the authors in [26] proposed a mutual authentication approach by sharing two distinguished keys, K_A and K_B, between Alice and Bob. However, the keys of Alice and Bob generated by the secret key generation scheme are identical in our scheme, which means that the mutual authentication approach cannot be applied directly. To solve this problem in our scheme, after Alice has been authenticated by Bob, they drop the authenticated key. When Alice authenticates Bob, they choose new authenticated key from the remaining shared key. If Bob cannot provide a valid tagged signal, Alice would consider

that this "Bob" is impersonated. Thus, Eve cannot actively steal the shared key and pass the authentication under impersonation attacks.

4.2.2. Jamming Attacks. As discussed in [35], Eve can attempt to disrupt the authentication procedure by jamming attacks. When she performs jamming in Stage 2, it may make Bob unable to authenticate Alice, which means the denial of service. However, the frequent jamming in Stage 2 is apt to be detectable. When Eve jams Stages 1 and 3, the jamming signal may be viewed as interference, and, if the jamming signal is not AWGN-like, it can be suppressed through conventional interference rejection techniques.

4.2.3. Replay Attacks. In Stages 2 and 3, Eve's received signals in the noiseless setting are, respectively, given by

$$r_{eb,i} = h_{BE,i}s_{b,i} = \left|h_{BE,i}\right| \exp\left(j\theta_{hBE,i} + j\theta_{b,i}\right), \tag{29}$$

$$\begin{aligned} r_{ea,i} &= h_{AE,i}s_{a,i} \\ &= \left|h_{AE,i}\right| \exp\left(j\left(\theta_{hAE,i} + \mathscr{M}\left(k_{a,i}^{\mathscr{M}}\right) - \widehat{\theta}_{a,i}\right)\right), \end{aligned} \tag{30}$$

where $h_{BE,i} = |h_{BE,i}|e^{j\theta_{hBE,i}}$ and $h_{AE,i} = |h_{AE,i}|e^{j\theta_{hAE,i}}$ are the ith subchannel from Bob to Eve and Alice to Eve, respectively. In a signal replay attack, Eve can store the waveforms as shown in (29) and (30) and simply replay the waveforms (since the signal in Step 1 contains no information about the shared key, we ignore Eve's replay of this signal). If the waveform in (29) is replayed, Alice will send the tagged signal since Alice does not know the legitimacy of its counterpart. Then Eve can get the authenticated key from the tagged signal. This case is similar to the impersonation attack in which Bob is impersonated by Eve, so we can employ the mutual authentication approach to address this problem. If the waveform in (30) is replayed, since the channel-phase response between two legitimate users is unique and cannot be revealed to Eve, the signal received by Bob will be

$$\begin{aligned} r_{b,i}^{e} &= h_{EB,i}r_{ea,i} = \left|h_{EB,i}h_{AE,i}\right| \\ &\cdot \exp\left(j\left(\theta_{hEB,i} + \theta_{hAE,i} + \mathscr{M}\left(k_{a,i}^{\mathscr{M}}\right) - \widehat{\theta}_{a,i}\right)\right). \end{aligned} \tag{31}$$

It is obvious that Bob will not accept it.

From the analysis above, we can find that, under various active attacks, it is nearly impossible for Eve to obtain the authenticated secret key or be authenticated by Alice or Bob.

4.3. Performance Evaluation. In this subsection, we present extensive simulations to demonstrate the effectiveness of the proposed scheme.

In the simulations, assuming that the receiver achieves ideal synchronization, so that the response message and tagged message contain only one OFDM symbol, respectively. The carrier frequency of the OFDM system is 2.4 GHz. The propagation environment is simulated by Rayleigh fading with 2 us constant delay time and the maximum Doppler frequency is 10 Hz. The sample interval is 0.25 us. The length of the mapping function input bits is set to be 2, so $2N$ key bits are needed for the simulations.

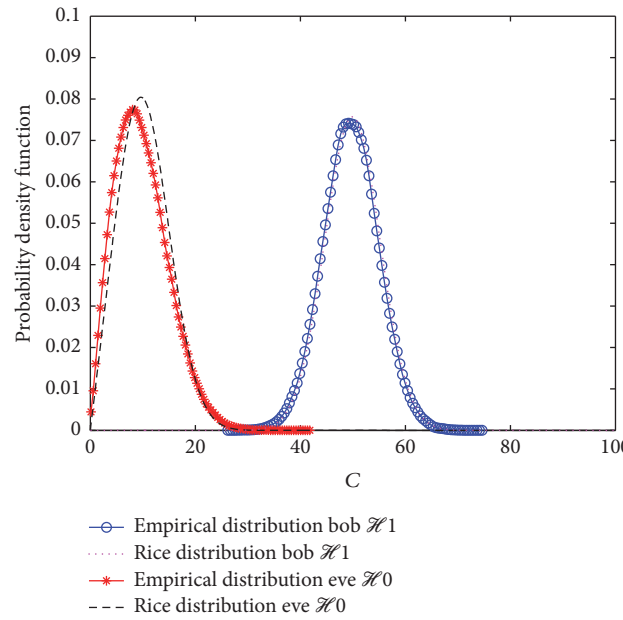


Figure 8: PDFs of $C_{\mathscr{H}0}$ and $C_{\mathscr{H}1}$ at SNR = 5 dB for $N = 64$.

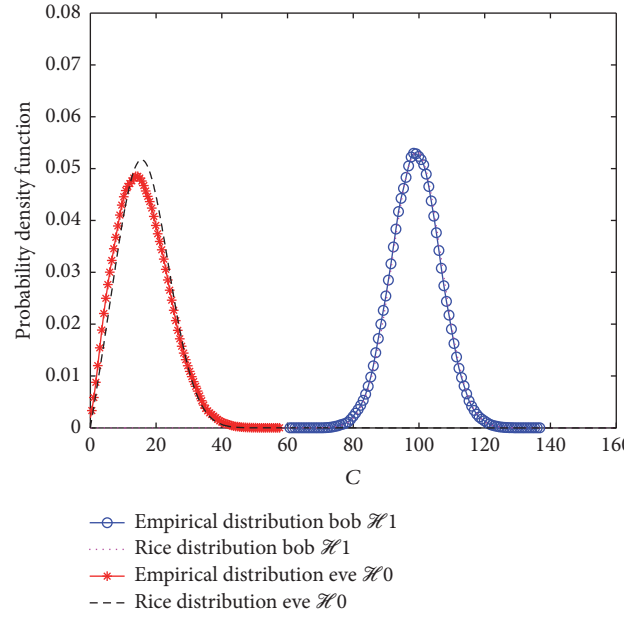


Figure 9: PDFs of $C_{\mathscr{H}0}$ and $C_{\mathscr{H}1}$ at SNR = 5 dB for $N = 128$.

Extensive Monte-Carlo simulations are conducted to investigate the PDFs of C under two hypothesis $\mathscr{H}0$ and $\mathscr{H}1$, which can be utilized to evaluate false acceptance rate and successful authentication rate. Furthermore, the appropriate choice of the threshold T also can be determined by these PDFs.

Figures 8 and 9 show the empirical PDFs of $C_{\mathscr{H}0}$ and $C_{\mathscr{H}1}$ at SNR = 5 dB for $N = 64$ and $N = 128$, respectively. As claimed in Section 4.1, $C_{\mathscr{H}0}$ and $C_{\mathscr{H}1}$ obey Rice distribution. Hence, Rice distributions according to (26) are also given in both figures, where the mean and variance are directly

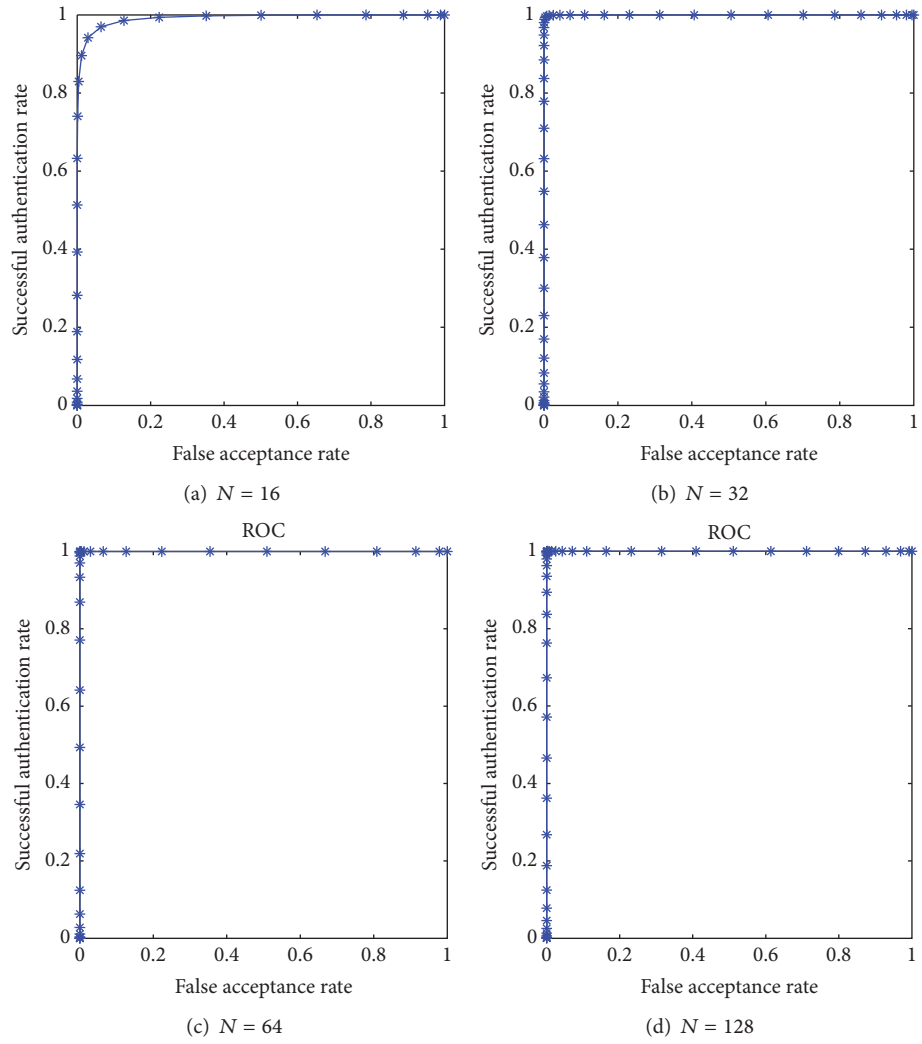


FIGURE 10: Successful authentication rate versus false acceptance rate for different N when SNR = 5 dB.

estimated through Monte-Carlo simulations [36]. From these two figures, we can find that the empirical distributions are coincided well with the theoretical Rice distributions. We also note that the PDFs of $C_{\mathscr{H}0}$ and $C_{\mathscr{H}1}$ are distinguished clearly in Figure 8 and in Figure 9 and the PDF of $C_{\mathscr{H}1}$ is far apart from that of $C_{\mathscr{H}0}$ even at SNR = 5 dB. Thus, it is easy to calculate threshold T if the successful authentication rate and false acceptance rate are given.

The receiver operating characteristic (ROC) describes the correlation between the false acceptance rate and the successful authentication rate. Figure 10 plots the ROC performance for different N when SNR = 5 dB. From these four subfigures, we can find that the ROC performance becomes better as N increases. Furthermore, when $N = 32$, the ROC are nearly ideal even at SNR = 5 dB.

4.4. Comparison with PHY-CRAM and PHY-PCRAS. The PHY-layer authentication schemes PHY-CRAM [26] and PHY-PCRAS [27] were shown to be simple and feasible. In the following, we will compare our proposed scheme with these two schemes.

As illustrated in Figure 11, for ROC performance, our scheme is better than PHY-CRAM and very similar to PHY-PCRAS. The reason is that our proposed scheme and PHY-PCRAS employ the channel-phase response, while amplitude modulation is employed in PHY-CRAM, which in performance is usually worse than phase modulation. Since the amplitude of all the subcarriers are not the same, the received performance may be impacted due to different SNR at each subchannel. Furthermore, in PHY-CRAM, high-peak fluctuations may occur, and in practice it is required to suppress the high peak with additional complexity. However, since OFDM technique is employed, compared to PHY-CRAM, our proposed scheme and PHY-PCRAS are more sensitive to the frequency offset.

As discussed in [26, 27], for impersonation attacks, our proposed scheme and PHY-PCRAS are more secure than PHY-CRAM. This can be explained by the better ROC performance and the fact that the channel-phase response

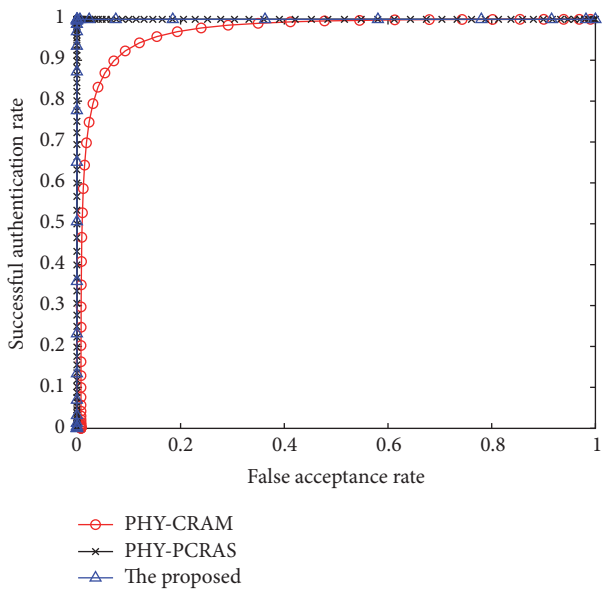


Figure 11: The ROC performance comparisons of the proposed authentication scheme, PHY-CRAM, and PHY-PCRAS at SNR = 5 dB with $N = 64$.

is more sensitive than channel-amplitude response to the distance between Alice and Bob.

The shared key of our proposed scheme is obtained from the proposed secret key generation scheme, while the other two schemes suppose that the shared key is preknown.

5. Conclusion

In this paper, to simplify the secret key extracting procedure based on the characteristics of the wireless channels and reduce the secret key bits mismatched rate, we propose a secret key generation scheme based on the channel-phase response. In the scheme, only one node is chosen to probe the channel and perform the quantization phase. Then the preliminary key is distributed after mapping and before equalizing. Further a PHY-layer authentication scheme is proposed utilizing the extracted secret key. This scheme exploits the short-term reciprocity of the channel-phase response and the sensitivity to the distance between the legitimate parties. The simulation results reveal that the proposed secret key generation scheme achieves a better performance compared to the existing scheme in terms of KGR, BMR, and secret key capacity. Besides, the extracted key passes the NIST randomness test. For the PHY-layer authentication scheme, the numerical results show that it performs better than existing work even at SNR = 5 dB when the shared key is obtained from the proposed secret key generation scheme, ensuring the randomness and security.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grants nos. 61601482, 61601516, 61601480, and 61502518) and sponsored by the Foundation of Science and Technology on Information Transmission and Dissemination in Comm. Networks Lab, National Key Laboratory of Anti-Jamming Communication Technology.

References

[1] W. Li, M. Ghogho, B. Chen, and C. Xiong, "Secure communication via sending artificial noise by the receiver: outage secrecy capacity/region analysis," *IEEE Communications Letters*, vol. 16, no. 10, pp. 1628–1631, 2012.

[2] W. Li, Y. Tang, M. Ghogho, J. Wei, and C. Xiong, "Secure communications via sending artificial noise by both transmitter and receiver: optimum power allocation to minimise the insecure region," *IET Communications*, vol. 8, no. 16, pp. 2858–2862, 2014.

[3] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE. Transactions on Information Theory*, vol. 39, no. 3, pp. 733–742, 1993.

[4] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography. I. Secret sharing," *IEEE. Transactions on Information Theory*, vol. 39, no. 4, pp. 1121–1132, 1993.

[5] A. A. Hassan, W. E. Stark, J. E. Hershey, and S. Chennakeshu, "Cryptographic key agreement for mobile radio," *Digital Signal Processing*, vol. 6, no. 4, pp. 207–212, 1996.

[6] S. Mathur, W. Trappe, N. Mandayam, C. Ye, and A. Reznik, "Radiotelepathy: extracting a secret key from an unauthenticated wireless channel," in *Proceedings of the 14th Annual International Conference on Mobile Computing and Networking (MobiCom '08)*, pp. 128–139, San Francisco, Calif, USA, September 2008.

[7] S. Jana, S. N. Premnath, M. Clark, S. K. Kasera, N. Patwari, and S. V. Krishnamurthy, "On the effectiveness of secret key extraction from wireless signal strength in real environments," in *Proceedings of the 15th Annual ACM International Conference on Mobile Computing and Networking (MobiCom '09)*, pp. 321–332, Beijing, China, September 2009.

[8] X. Hu, X. Li, E. C.-H. Ngai, V. C. M. Leung, and P. Kruchten, "Multidimensional context-aware social network architecture for mobile crowdsensing," *IEEE Communications Magazine*, vol. 52, no. 6, pp. 78–87, 2014.

[9] X. Hu, J. Zhao, B.-C. Seet, V. C. M. Leung, T. H. S. Chu, and H. Chan, "S-aframe: agent-based multilayer framework with context-aware semantic service for vehicular social networks," *IEEE Transactions on Emerging Topics in Computing*, vol. 3, no. 1, pp. 44–63, 2015.

[10] L. Cheng, W. Li, D. Ma, J. Wei, and X. Liu, "Moving window scheme for extracting secret keys in stationary environments," *IET Communications*, vol. 10, no. 16, pp. 2206–2214, 2016.

[11] L. Cheng, W. Li, L. Zhou, C. Zhu, J. Wei, and Y. Guo, "Increasing secret key capacity of OFDM systems: a geometric program approach," *Concurrency and Computation: Practice and Experience*, 2016.

[12] L. Zhou, Z. Sheng, L. Wei et al., "Green cell planning and deployment for small cell networks in smart cities," *Ad Hoc Networks*, vol. 43, pp. 30–42, 2016.

[13] K. Zeng, "Physical layer key generation in wireless networks: challenges and opportunities," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 33–39, 2015.

[14] E. Zhang, P. Yuan, and J. Du, "Verifiable rational secret sharing scheme in mobile networks," *Mobile Information Systems*, vol. 2015, Article ID 462345, 7 pages, 2015.

[15] Q. Wang, H. Su, K. Ren, and K. Kim, "Fast and scalable secret key generation exploiting channel phase randomness in wireless networks," in *Proceedings of the IEEE INFOCOM*, pp. 1422–1430, Shanghai, China, April 2011.

[16] X. Hu, T. H. S. Chu, V. C. M. Leung, E. C.-H. Ngai, P. Kruchten, and H. C. B. Chan, "A survey on mobile social networks: applications, platforms, system architectures, and future research directions," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1557–1581, 2014.

[17] L. Zhou, X. Hu, E. C.-H. Ngai et al., "A dynamic graph-based scheduling and interference coordination approach in heterogeneous cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3735–3748, 2016.

[18] H. Liu, Y. Wang, J. Yang, and Y. Chen, "Fast and practical secret key extraction by exploiting channel response," in *Proceedings of the 32nd IEEE Conference on Computer Communications (INFOCOM '13)*, pp. 3048–3056, April 2013.

[19] X. Wu, Y. Peng, C. Hu, H. Zhao, and L. Shu, "A secret key generation method based on CSI in OFDM-FDD system," in *Proceedings of the IEEE Globecom Workshops (GC Wkshps '13)*, pp. 1297–1302, December 2013.

[20] X. Wu, Z. Yang, C. Ling, and X. G. Xia, "Artificial-noise-aided physical layer phase challenge-response authentication for practical OFDM transmission," *IEEE Transactions on Wireless Communications*, vol. 15, no. 10, pp. 6611–6625, 2016.

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

[22] X. Hu, T. H. S. Chu, H. C. B. Chan, and V. C. M. Leung, "Vita: a crowdsensing-oriented mobile cyber-physical system," *IEEE Transactions on Emerging Topics in Computing*, vol. 1, no. 1, pp. 148–165, 2013.

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

[24] L. Zhang, Q. Wu, A. Solanas, and J. Domingo-Ferrer, "A scalable robust authentication protocol for secure vehicular communications," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1606–1617, 2010.

[25] L. Xiao, L. J. Greenstein, N. B. Mandayam, and W. Trappe, "Using the physical layer for wireless authentication in time-variant channels," *IEEE Transactions on Wireless Communications*, vol. 7, no. 7, pp. 2571–2579, 2008.

[26] D. Shan, K. Zeng, W. Xiang, P. Richardson, and Y. Dong, "PHY-CRAM: physical layer challenge-response authentication mechanism for wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 1817–1827, 2013.

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

[28] W. C. Jakes Jr., *Microwave Mobile Communications*, Wiley-IEEE Press, Piscataway, NJ, USA, 1994.

[29] X. Du, D. Shan, K. Zeng, and L. Huie, "Physical layer challenge-response authentication in wireless networks with relay," in *Proceedings of the 33rd IEEE Conference on Computer Communications (INFOCOM '14)*, pp. 1276–1284, IEEE, Ontario, Canada, May 2014.

[30] R. Wilson, D. Tse, and R. A. Scholtz, "Channel identification: secret sharing using reciprocity in ultrawideband channels," in *Proceedings of the IEEE International Conference on Ultra-Wideband (ICUWB '07)*, pp. 270–275, September 2007.

[31] Q. Dai, H. Song, L. Jin, and K. Huang, "Physical layer authentication and secret key distribution mechanism based on euivalent channel," *Science China*, vol. 44, no. 12, pp. 1580–1592, 2014.

[32] Z. Szabó, "Information theoretical estimators toolbox," *Journal of Machine Learning Research*, vol. 15, pp. 283–287, 2014.

[33] K. Zeng, D. Wu, A. Chan, and P. Mohapatra, "Exploiting multiple-antenna diversity for shared secret key generation in wireless networks," in *Proceedings of the IEEE (INFOCOM '10)*, pp. 1–9, San Diego, Calif, USA, March 2010.

[34] A. Rukhin, J. Sota, J. Nechvatal et al., "A statistical test suite for random and pseudorandom number generators for cryptographic applications," Special Publication NIST 800-22, National Institute of Standards and Technology, 2010.

[35] X. Wu, Z. Yang, C. Ling, and X. G. Xia, "A Physical-Layer Authentication Assisted Scheme for Enhancing 3GPP Authentication," 2015, http://arxiv.org/abs/1502.07565v1.

[36] T. R. Benedict and T. T. Soong, "The joint estimation of signal and noise from the sum envelope," *IEEE Transactions on Information Theory*, vol. 13, no. 3, pp. 447–454, 1967.

Hindawi
Mobile Information Systems
Volume 2017, Article ID 8682168, 11 pages
https://doi.org/10.1155/2017/8682168

Research Article

Design and Voluntary Motion Intention Estimation of a Novel Wearable Full-Body Flexible Exoskeleton Robot

Chunjie Chen,[1,2,3] Xinyu Wu,[1,2,4] Du-xin Liu,[1,2,3] Wei Feng,[1,2] and Can Wang[1,2]

[1] *Guangdong Provincial Key Laboratory of Robotics and Intelligent System, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China*
[2] *Chinese Academy of Sciences (CAS) Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen 518055, China*
[3] *Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Beijing, China*
[4] *Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong*

Correspondence should be addressed to Xinyu Wu; xy.wu@siat.ac.cn and Can Wang; can.wang@siat.ac.cn

Received 28 January 2017; Accepted 30 March 2017; Published 27 June 2017

Academic Editor: Qiang Liu

The wearable full-body exoskeleton robot developed in this study is one application of mobile cyberphysical system (CPS), which is a complex mobile system integrating mechanics, electronics, computer science, and artificial intelligence. Steel wire was used as the flexible transmission medium and a group of special wire-locking structures was designed. Additionally, we designed passive joints for partial joints of the exoskeleton. Finally, we proposed a novel gait phase recognition method for full-body exoskeletons using only joint angular sensors, plantar pressure sensors, and inclination sensors. The method consists of four procedures. Firstly, we classified the three types of main motion patterns: normal walking on the ground, stair-climbing and stair-descending, and sit-to-stand movement. Secondly, we segregated the experimental data into one gait cycle. Thirdly, we divided one gait cycle into eight gait phases. Finally, we built a gait phase recognition model based on k-Nearest Neighbor perception and trained it with the phase-labeled gait data. The experimental result shows that the model has a 98.52% average correct rate of classification of the main motion patterns on the testing set and a 95.32% average correct rate of phase recognition on the testing set. So the exoskeleton robot can achieve human motion intention in real time and coordinate its movement with the wearer.

1. Introduction

The number of China's elderly and disabled people had reached 260 million at the end of 2014, China's rapid aging has caused widespread concern, and it is difficult for China to afford the issue of a rapidly aging population [1]. Additionally, the number of hemiplegic and paraplegic patients and people with walking difficulties increased year by year. Most of these people can only use a wheelchair to achieve self-care. Walking and standing are their greatest desire, the physical function of the elderly declines, and their daily activities are limited with the increase of age. It brings great pressure and burden to the families and society to take care of the elderly, which has become a major social problem. Thus, it is necessary to develop a power exoskeleton device to assist the elderly in their needs. The exoskeleton robot is a kind of wearable human-machine integration device, which combines the science of robotics and rehabilitation engineering. The application of an exoskeleton robot can help the patients who have lost their walking ability or have walking difficulties stand and walk normally again.

The concept of mobile cyberphysical system (CPS) has emerged as a promising tool where the operations of the physical and engineered systems are monitored, controlled, coordinated, and integrated by means of a computing and communication core [2]. The exoskeleton robot must determine the wearer's moving intention accurately and quickly due to its close link with the wearer. Recent advances in microelectromechanical systems (MEMS) have led to the rapid development of microsensors [3, 4]. Correlations among the data representing an event are usually reflected in multiple measurements at different locations [5, 6]. The

acquisition of electroencephalogram, electromyography, and other biological signals have become a hot research topic in the field of human-machine interaction owing to their quick response [6, 7]. However, the measurement of biological signals involves poor robustness because of their low frequency, weak amplitude, and low signal-to-noise ratio [8]. On the other hand, the traditional force feedback and position tracking control strategies are also widely used for exoskeleton control, which is based on physical signals; a wearer with walking difficulties has difficulty in achieving free walking by simply relying on physical signals because of physical signals' obvious signal delay and too many sensors. As research based on biological signals, force, position, and other physical signals has not yet achieved breakthroughs in the motion intention estimation for exoskeleton robots, researchers have been searching for new methods to assess the wearer's moving intentions accurately and quickly.

The Cybernetics Laboratory of Tsukuba University developed the HAL series of wearable power-assist robot systems to help the elderly or the disabled to achieve normal walking in 2002. This exoskeleton robot can help disabled people stand up or reduce the labor intensity of workers. The control strategy of the HAL series includes two kinds of control modes: one is based on EMG and the other one involves gait prelearning and gait pattern generation. Researchers who developed HAL series established a joint torque control method based on EMG signals to achieve control and maintain coordination with the wearer's lower limb movement [9].

In the field of exoskeleton robots research, one walking gait cycle is usually divided into multiple phases. In [10], the walking gait cycle was divided into the stance phase and the swing phase. Murray et al. [11] proposed an assistive approach without dictating the spatiotemporal nature of joint movement for the lower limb exoskeleton; a finite-state machine consisting of six gait phase states was used to govern the exoskeleton controller. Kazerooni et al. [12] adopted a hybrid control strategy for different gait phases to control the Berkeley Lower Extremity Exoskeleton. Liu et al. [13] proposed an approach of gait phase recognition for a lower limb exoskeleton with only joint angular sensors. Oh et al. [14] considered that a single control method is not the best option for all motion phases during the gait cycle. It is difficult to use a fixed model to describe the process of walking and obtain a fixed output setting due to individual differences in human walking and the changes in road conditions. Normal walking is the result of a coordinated body movement that allows the body to move in the most efficient way [15]. In this study, we developed a set of novel wearable full-body flexible exoskeleton robots and presented a method of voluntary motion intention estimation, which can help wearable exoskeleton robot walk comfortably.

2. Materials and Methods

2.1. Design of a Novel Wearable Full-Body Flexible Exoskeleton Robot. The human body has three basic planes: the sagittal plane, the coronal plane, and the horizontal plane, and these three surfaces are perpendicular to each other at the center of the body [16]. The exoskeleton robot is positioned on the human body and the wearer coordinates his movement with the robot; the design of the robot's dimensions must be in accordance with the specifications of the human body. The length of the connecting rod of a full-body exoskeleton robot should be adjustable to a certain range for better compatibility. The size design of the exoskeleton in this paper refers to the GB10000-88 (Chinese adult body size) standard. Although the size of various parts of the body and their proportions are basically fixed, there are differences in proportions of the body between men and women; the lower limb size of men is generally larger than that of women [17]. The height range of a subject was limited from 1550 to 1850 mm, and then the regulating length range of the connecting rod can be calculated as Table 1.

Table 1: Adjustment range of the connecting rod of the exoskeleton robot.

Adjustable Component	Range
Lower leg	340 mm–410 mm
Thigh	315 mm–385 mm
Upper arm	318 mm–379 mm
Forearm	218 mm–261 mm

The exoskeleton robot must be a mechanical shadow of the wearer and it must be able to mimic each of his actions in real time, and even a millisecond's hesitation can create a burden that makes the wearer feel as impeded as walking in water. Therefore, its sensors must be able to quickly read every minor action and its microprocessor must be sufficiently powerful to convert these data into instructions to the robot in real time. The speed of traditional hydraulic transmission is too slow which increases the weight of robot; additionally, if a motor drive was mounted directly on the joint that will require a larger body structure. Therefore, in this study, we positioned all the drive units in the backpack to obtain a more delicate exoskeleton robot system.

Figure 1 shows the overall structure of the exoskeleton robot system. The system uses a distributed control architecture, which is composed of the main control computer, data acquisition board, motor drive board and various sensors, and so forth.

2.1.1. Steel Wire Transmission. The steel wire and tube are forced along the tangent direction of the wire line when the motor drives the wire. The tube is rigid and incompressible in the longitudinal direction although the wire and tube appear to be soft, the interior of the tube is made of steel, and the wall is rather thick. The steel wire and the tube can be bent only to a limited degree according to different types, while the deformation of the steel wire is negligible compared to the moving distance of the motor output. As a result, the motor torque can be transferred to the terminal almost without loss [18].

The steel wire in one loop is divided into three parts: the upper end of the wire between the upper part of tube and the lower tube, the tube, and the part between the tube and the lower end of the lower part of the wire, their length is L_1,

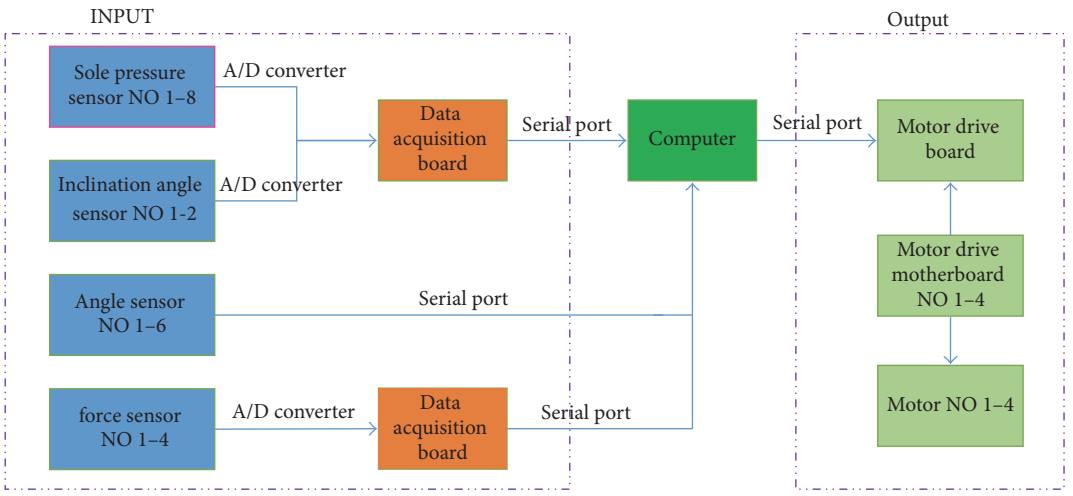


Figure 1: Exoskeleton robot system structure.

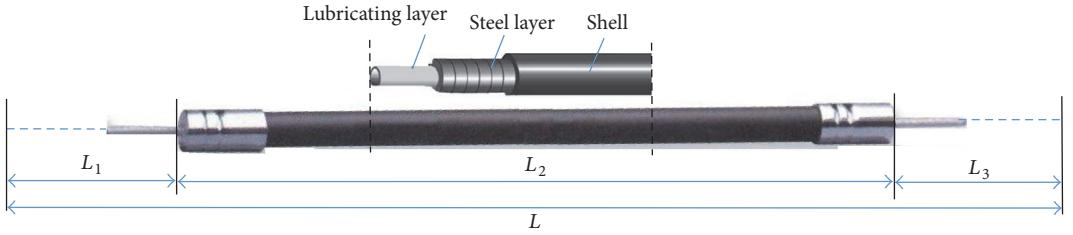


Figure 2: Steel wire in one loop.

L_2, and L_3, respectively, and the total length of steel wire is L which is shown in Figure 2.

$$L = L_1 + L_2 + L_3. \tag{1}$$

L_1 will increase when the motor rotates, L_2 does not change because it is actually the length of the tube, and L does not change because it is the length of the steel wire; consequently, L_3 has to decrease. Thus, a wire drive circuit is formed.

The advantage of this transmission approach described in this paper is combined with hydraulic and pneumatic transmission, gears drive, and any other traditional transmission, which can realize flexible transmission with a rather simple structure as shown in Figure 3. Additionally, it can avoid the noise caused by hydraulic pumps or air pumps. Therefore, it constitutes a major innovation for the structural design of the exoskeleton robot, which can easily transfer the torque to any joint of the robot without interfering with the other joints.

Most steel wire transmission systems achieve the torque transmission through a pulley or pulley group. Furthermore, most of steel wire transmission systems involve only one-way transmission, which require a spring or other devices to assist in the return time of a loop, and they also need large pretightening force of the steel wire to prevent slipping. In this study, we designed a group of special wire-locking pulley structures as shown in Figure 4 that resulted in a wire without slippage; thus, transmission could be fully implemented as needed.

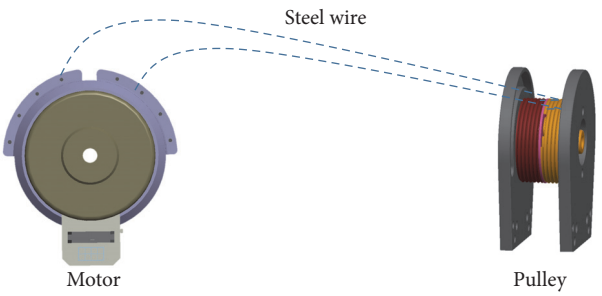


Figure 3: Steel wire transmissions.

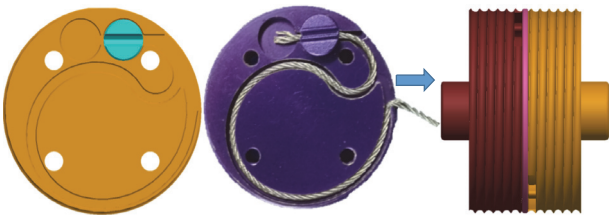


Figure 4: Wire-locking pulley.

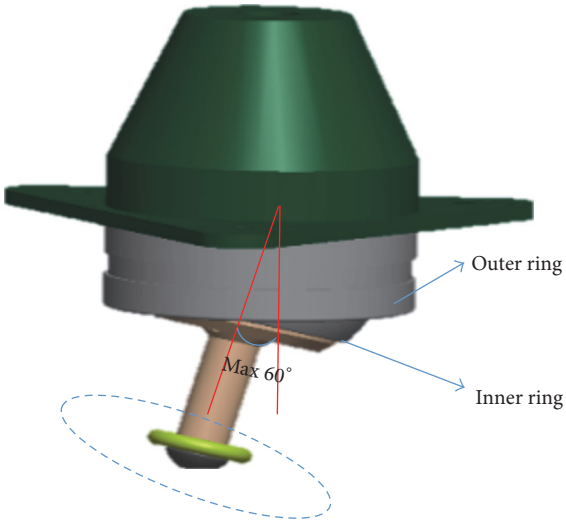


Figure 5: Passive spherical joint.

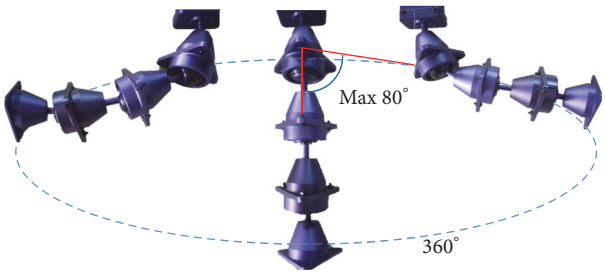


Figure 6: Passive spherical joint group.

2.1.2. Passive Spherical Joint. This joint is composed of a special sliding bearing structure containing an inner ring with an outer spherical surface and an outer ring with an inner sphere. The joint which is shown in Figure 5 has two bonded spherical surfaces that can withstand large loads and it is usually used for the low speed of the swing movement. It can also be tilted in a certain angle because of the sliding surface of the spherical shape, which can operate normally even when the support shaft and the shaft shell hole are not concentric. Several joints can be connected to a group as needed, which is shown in Figure 6.

This type of joint is very suitable for the flexible robot developed in this study, which can be worn without being bound by the mechanical structure of exoskeleton robot, and it greatly improves comfort to the wearer.

2.1.3. System Overview. The wearable exoskeleton robot developed in this study is a complex intelligent system integrating mechanics, electronics, computer science, and artificial intelligence, which is necessary to capture the movement state and motion intention of the wearer through signals from the sensor system. The sensor system mainly includes angle sensors, inclination sensors, plantar pressure sensors, and pressure sensors between the human limbs and the exoskeleton robot as shown in Figure 7. The gait acquisition system was established in advance, and the sensors were installed at each joint of the gait acquisition system, the gait acquisition system could adapt to wearers of different heights, and the following are the main tasks of the four kinds of sensors:

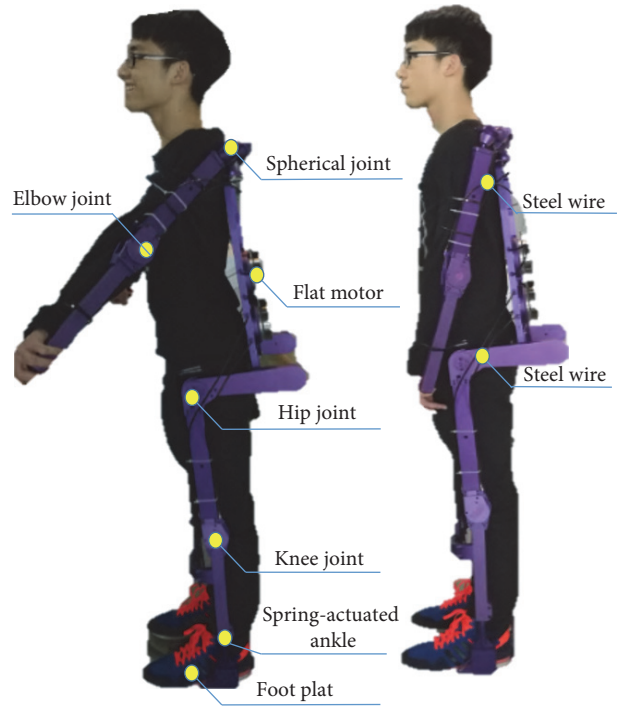


Figure 7: Full-body flexible exoskeleton robot.

(1) The inclination sensors help detect body posture and prevent falling.

(2) The plantar pressure sensors are used to determine the motion state.

(3) The joint angle sensors are used to determine the joint angle of the robot.

(4) The pressure sensor is used to collect the real-time interaction data between the human limb and the exoskeleton robot.

Energy consumption of the knee joint is the most of all the joints according to the data curve of the joint movement angle and output torque [19]. The exoskeleton robot is difficult to be designed with the same freedom as with human shoulder using only a motor drive because of complex human body structure, whereas the elbow joint movement is simple. The exoskeleton robot developed in this study was designed to only add active driving in the knee and elbow joints, and we used flexible passive degrees of freedom for all the other joints.

2.2. Sensor Data Acquisition System and Experiment. The lower limb of human is generally believed to have seven degrees of freedom [20]: hip adduction, abduction, external

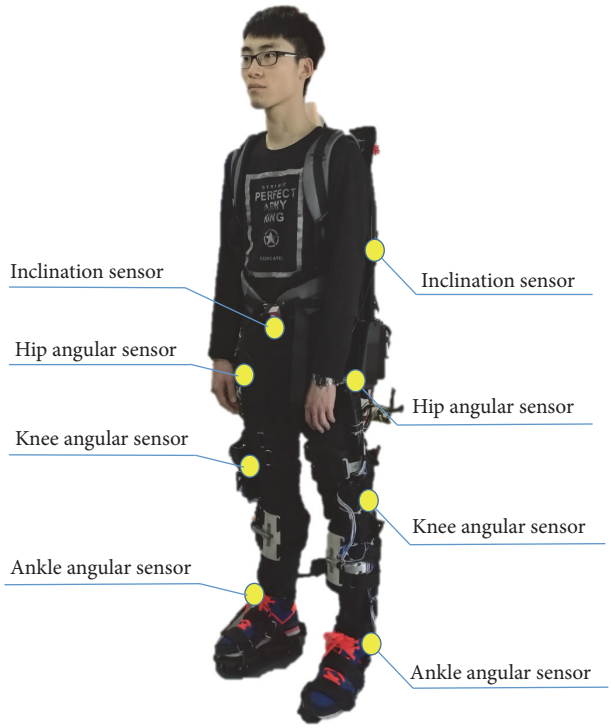


Figure 8: Sensor data acquisition system.

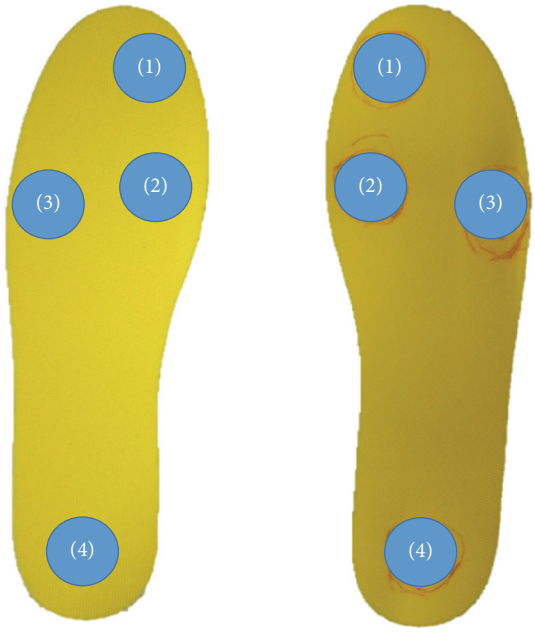


Figure 9: Plantar pressure sensor distribution.

rotation, internal rotation, ankle valgus and virus, external rotation, and internal rotation motion. The angles of the hip, knee, and ankle changed regularly during the walking. We only collected data from the angle of the hip, knee, and ankle in the sagittal plane as shown in Figure 8 to facilitate the analysis of the gait. The distribution of plantar pressure sensors is shown in Figure 9. In walking, the upper part of the body exhibits a slight swing with the swing of the feet [21]. We installed sensors to acquire upper body swinging data to more accurately reproduce the process of walking. We collected and analyzed the changes of the hip, knee, ankle angle, and body posture that occur in the process of walking, stair-climbing, and sit-to-stand (STS) movement to understand the real human gait; Table 2 shows the number of three types of sensors.

Table 2: Exoskeleton sensor data.

Type	Number
Plantar pressure	8
Joint angle	6
Inclination angle	4

Subjects. All the recruited subjects before the experiment were informed about the experimental nature, procedure, effect, and the potential risk. They were very willing to participate in experiment and sign the informed consent forms. Meanwhile, this experiment has been approved by Shenzhen Institutes of Advanced Technology Ethics Committee.

We selected 10 young men with normal lower limb function to perform walking on the ground, treadmill walking, stair-climbing and stair-descending, and STS tests using our sensor data acquisition system. The mean age of the group was 25.3 years, with an average height of 173.3 cm and an average weight of 67.7 kg. Before the experiment, the participants' thigh and leg length and waist thickness were measured, and the various links of the collector were adjusted to the appropriate size to conduct the experiment.

Gait analysis is a method to study walking patterns that aims to reveal the key links and factors affecting gait by means of mechanics and kinematics. This technique is also helpful to guide the robot in the later control stage. Compared with other biometrics, such as fingerprint, iris, and face recognition, gait represents a type of external dynamic performance that is closely related to spatial and temporal information. Gait recognition must begin with the lower limb of the human body and the gait signal is divided periodically. It must accurately and quickly determine the wearer's movement intentions and make decisions because the robot has a close relationship with the wearer. People always maintain the center of gravity while walking [22]. In the initial stage of the gait, the wearer is not yet skilled at controlling the robot. The motion planning algorithm requires more intervention and learning to adapt to the gait. The algorithm will gradually stabilize and finally achieves a regular gait when a stable complete gait is achieved.

2.2.1. Walking at Different Speeds on the Ground and a Treadmill. The walking gait of a human has the characteristics of periodicity, left-right symmetry, and coordination, and the walking motion is mainly performed in the sagittal plane [23]. The time of the process that the same heel touches the ground again is called a complete gait cycle during walking. Both sides have their own gait cycle as the left and right legs

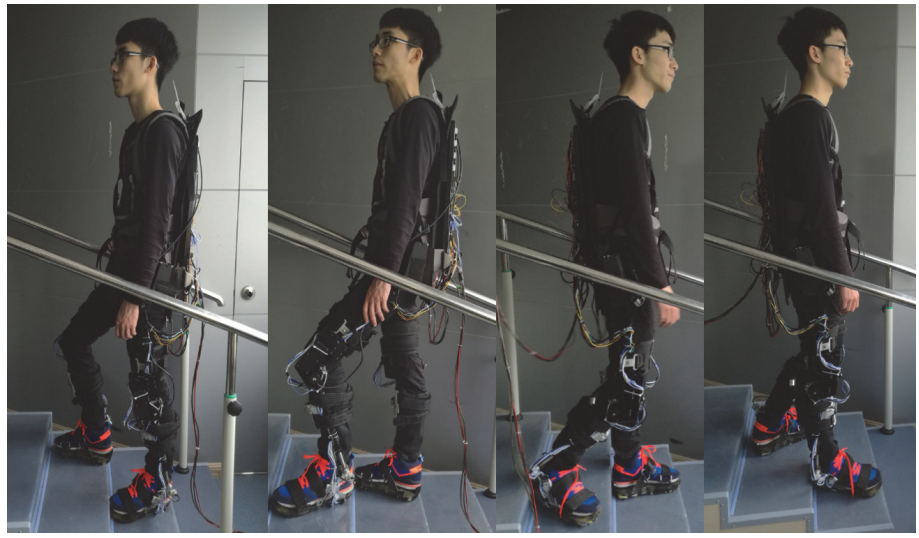


Figure 10: Stair-climbing and stair-descending.

Figure 11: STS.

alternate. The right lower limb is considered as an example because of the symmetry of human gait; the same person's left and right leg gait are usually only a phase difference.

2.2.2. Stair-Climbing and Stair-Descending. The cyclic pattern of the lower limbs during the task of stair-climbing and stair-descending as shown in Figure 10 is very similar to the pattern observed during walking on the ground [24].

For both tasks, periods of support (stance) and non-support (swing) can be defined. Joseph and Watson have cited similar figures (i.e., 60% stance, 40% swing) for a stair-descending task. In contrast, for the lower limbs of subjects in stairs-climbing task it was observed that 66% time of the gait cycle is stance phase and the remaining 33% time is the swing phase [25].

2.2.3. STS. STS movement is also divided into phases as shown in Figure 11 for better understanding, the most commonly used classification for which contains two phases [26], one is the preparation phase, which includes the flexion process of the upper part of the body and the process of the body begins the seat-off motion, the other one is the rising phase, which contains the process from the seat-off action (maximum anterior flexion of the trunk) to the standing posture, and finally people maintain their body in a quasi-stationary position.

2.3. Voluntary Motion Intention Estimation. The phases of the gait data are firstly classified offline before the motion intention estimation.

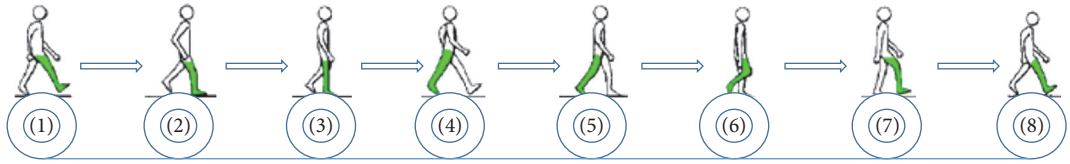


Figure 12: Human walking gait cycle.

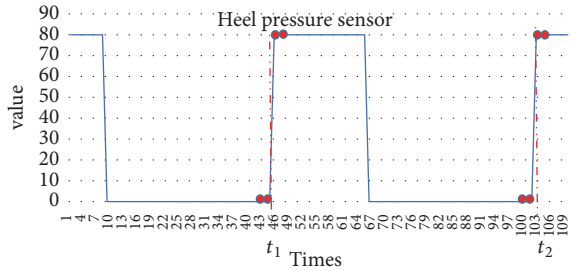


Figure 13: Gait cycle cutting.

2.3.1. Gait Cycle Division. The lower limbs repetitively deviate from the standard standing position in walking [27]. Gait of walking is a cyclic phenomenon that can be divided into phases from the initial contact with the floor to the final contact with the floor in the swing phase; it begins when the tibia is vertical to the floor and ends when the foot strikes the floor again.

In this study, one gait cycle was divided into eight phases as shown in Figure 12; the proportion of each phase in one gait cycle is as follows:

(1) Initial contact phase (0%).
(2) Loading response phase (0–10%).
(3) Middle stance phase (10–30%).
(4) Terminal stance phase (30–50%).
(5) Preswing phase (50–60%).
(6) Initial swing phase (60–73%).
(7) Middle swing phase (73–87%).
(8) Terminal swing phase (87–100%).

One gait cycle is defined by the time that the right heel hits the ground two times. The moment of mutation is used as the dividing point according to the data of the plantar pressure sensors. The gait sequence can be automatically divided into several gait cycles by using this method. We defined the moment when two consecutive point values of the heel pressure sensors are greater than the setting threshold as the dividing point which is shown in Figure 13.

$$T_{\text{cycle}} = t_{i+1} - t_i, \quad i = 1, 2, \ldots, N,$$
$$\Delta t_i = \left(\Delta t_1, \Delta t_2, \Delta t_3, \ldots, \Delta t_n\right). \tag{2}$$

We can divide the gait cycle as required by determining two such successive points. From the 10 recruited subjects that performed the walking test with the sensor data acquisition system, 4500 gait cycles were obtained. Each captured gait cycle dataset can be expressed in the form of a matrix.

$$\begin{bmatrix} s1 & \cdots & s18 \\ \vdots & \ddots & \vdots \\ s1j & \cdots & s18j \end{bmatrix}, \quad j = 1, 2, \ldots, N, \tag{3}$$

where j is the number of data points of each captured gait cycle, which depends on the walking speed and sampling frequency, and $s1$ to $s18$ are all the data of the sensors installed in the sensor data acquisition system.

2.3.2. Gait Phase Labels on Gait Data. We verified the classification method after dividing the gait cycle. Each of the obtained 4500 gait cycles was divided into eight phases by the proportion of time in one gait cycle.

One gait cycle was taken for analysis; a gait cycle determined with the above method begins approximately at the moment that one foot leaves the ground and ends with the ipsilateral foot leaving the ground. The gait phase classification resulted in the values as shown in Figure 14. The cyclic pattern of the lower limbs during the task of stair-climbing and stair-descending is very similar to the pattern of walking on the ground [28]. Therefore, we used the same proportion of time in a cycle as normal walking in this study. The STS movement was also divided into two phases for better understanding [28]. The first phase is the flexion phase, which occurred during the first 35% of the STS cycle. The second phase is the extension phase; a gait cycle is also generally divided into two main phases: the stance phase (60%) and the swing phase (40%) [29]. The opposite process applies and the phases are easy to divide during the process of standing up and sitting down.

2.3.3. Gait Phase Recognition. In the actual movement process, we can use the following methods to obtain the real-time behavior of the robot once the instantaneous sensor data information is obtained. In this study, all the actions were divided into the three categories of normal walking on the ground, stair-climbing and stair-descending, and STS. Firstly, we used the following method to distinguish these three categories. In each category, the phase in the gait cycle needs to be found out in real time; walking on the ground was taken as an example, and then Figure 15 shows how we further identified the specific gait phases, the eight different color blocks represent the eight phases in a gait cycle, the

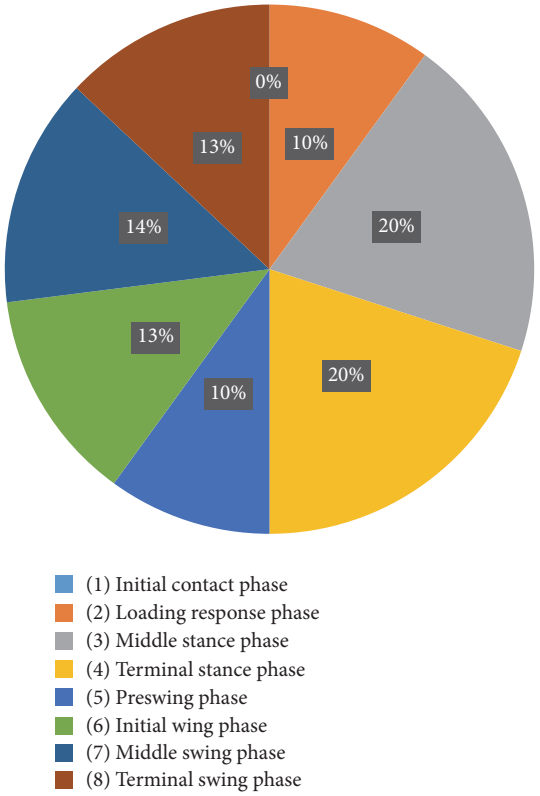


Figure 14: Gait phases in one cycle.

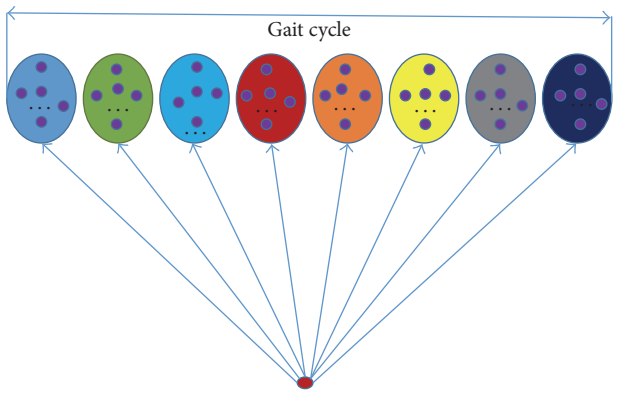


Figure 15: Gait phase recognition.

blue dot inside the blocks represents the specific data points in each phase, the nearest blue point to the target point was found by calculating, and the same method is also applied to other categories.

Distinguishing of Walking on the Ground, Stair-Climbing, and Stair-Descending. A novel method was developed to distinguish these two categories as shown in Figure 16 based on the gait cycle classification data described above and then a k-Nearest Neighbor classifier was created for data training.

$$\begin{aligned} \theta i &= \boldsymbol{\theta}_{\mathrm{LK}i} - \boldsymbol{\theta}_{\mathrm{RK}i} \quad i = 1, 2, \ldots, N, \\ \Delta\boldsymbol{\theta}\mathbf{i} &= (\Delta\boldsymbol{\theta}1, \Delta\boldsymbol{\theta}2, \Delta\boldsymbol{\theta}3, \ldots, \Delta\boldsymbol{\theta}n), \end{aligned} \tag{4}$$

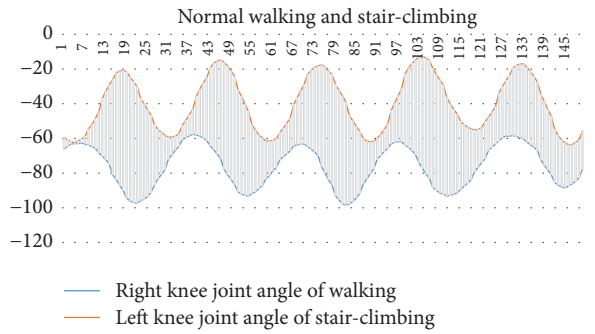


Figure 16: Distinguishing of normal walking and stair-climbing.

where $\boldsymbol{\theta}_{\mathrm{LK}}$ denotes the left knee joint angle for stair-climbing and $\boldsymbol{\theta}_{\mathrm{RK}}$ indicates the right knee joint angle for normal walking.

STS. Standing from a seated position is an activity that we perform many times every day. STS is a rather action different from normal walking and stair-climbing and stair-descending; the angular values that we recorded reflected the relationships of the lower limb joints. This difference was confirmed by a reversal of the rapid increase of angle in knee extension. The angle of the hip and ankle joints also increased. The left and right leg have the same phase during the STS action; this is the greatest difference between STS and other actions, in which the two legs have the same phases. Thus, we can easily distinguish STS categories.

$$\begin{aligned} \boldsymbol{\theta}_{\mathrm{LA}} &= \boldsymbol{\theta}_{\mathrm{RA}}, \\ \boldsymbol{\theta}_{\mathrm{LK}} &= \boldsymbol{\theta}_{\mathrm{RK}}, \\ \boldsymbol{\theta}_{\mathrm{LH}} &= \boldsymbol{\theta}_{\mathrm{RH}}, \end{aligned} \tag{5}$$

where $\boldsymbol{\theta}_{\mathrm{LA}}$ denotes the left ankle joint angle, $\boldsymbol{\theta}_{\mathrm{RA}}$ is the right ankle joint angle, $\boldsymbol{\theta}_{\mathrm{LK}}$ is the left knee joint angle, $\boldsymbol{\theta}_{\mathrm{RA}}$ indicates the right knee joint angle, $\boldsymbol{\theta}_{\mathrm{LH}}$ is the left hip joint angle, and $\boldsymbol{\theta}_{\mathrm{RH}}$ denotes the right hip joint angle.

3. Results and Discussion

We observed that angle changes of the hip joint, knee joint, and ankle joint had obvious regularity in the process of walking through the analysis of the gait data. Furthermore, we verified that the walking gait described above was periodic and left-right symmetrical. In addition, we determined that there is a slight swing around the waist; the body naturally inclines to the side of the supporting leg to focus on the foot when people walk, which is to maintain the balance of the body and avoid falling.

We firstly divided them into normal walking, stair-climbing and stair-descending, and STS after obtaining the instantaneous 18-dimensional sensor data and then inserted them into the *K*NN model, which was built in a Python environment. We finally got a recognition rate of 98.21%, which is nearly with no intersection of other phases, except the very standing stationary points that have the same phase.

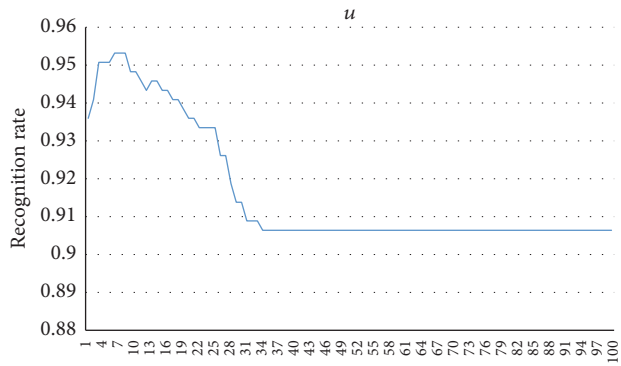


Figure 17: The best coefficient u.

Secondly, normal walking is taken as an example; we used the six angle sensors data to calculate the nearest point to the instantaneous point and then obtained the corresponding gait phase and the most probable point.

$$\text{dist\_angle} = \sqrt[2]{\sum_{1}^{6} \left(\text{test}_i - \text{target}_i\right)^2}, \quad i = 1, 2, 3, 4, 5, 6, \tag{6}$$

where test_i denotes the data from the six ankle joint angles after training, target_i represents the instantaneous ankle joint angle data, and dist_angle indicates the distance between the target point and instantaneous point in the gait cycle.

$$\text{dist\_pressure} = \sqrt[2]{\sum_{1}^{8} \left(\text{test}_j - \text{target}_j\right)^2}, \quad j = 1, 2, 3, 4, 5, 6, 7, 8, \tag{7}$$

where test_j represents the data from the eight plantar pressure sensors after training and target_j indicates the instantaneous plantar pressure sensor data.

In an actual gait, the plantar pressure sensors and the angle sensors data are not uniform, and the weight of the impact on the calculation results is unknown. Therefore, we must define a new coefficient u:

$$\text{dist} = \text{dist\_angle} + \text{dist\_pressure} * u, \tag{8}$$

where dist denotes the distance between the target point and instantaneous point and u is the weight coefficient.

We inserted the target data into the training model and determined the best coefficient u (from 1 to 100).

We calculated that the best recognition rate was 95.32% when $u = 7, 8, 9$ and that the recognition rate no longer increases when $u \geq 35$; they are shown in Figure 17, so we did not choose the number larger than 100 to try. Finally, we selected $u = 8$ and input four normal walking gait cycles into the model to identify the corresponding phase and pose, which provided the following result.

From Figure 18, we can easily see that the length of each gait cycle may be different according to the walking speed. We can acquire approximately 55 to 58 data points in a gait cycle while walking at a normal pace. The abscissa shows each point in the gait cycle; the ordinate is the every phase of the gait cycle. The blue curve shows the actual gait phases data, and the red one is predicted value through the gait phase recognition method.

In Figure 18(a) the middle stance phase exhibits recognition errors, while all the other phases are predicted accurately. In Figure 18(b), only the loading response phase and the preswing phase have small recognition errors and Figure 18(c) shows four gait phase recognition errors. However, each gait phase had only a few gait data points due to the sampling rate. Therefore, erroneously identified data points affected severely the correct recognition rate. In Figure 18(d), only one phase has a few recognition errors. From Figure 18, the correct rate of phase (CRP) of the four cycles was above 90% and the CRP of the entire gait set was 95.32%. Further analysis determined that, in each of the phases that exhibited identification errors, only one gait data point was identified incorrectly. Therefore, the gait phase recognition model described in this paper was able to predict the gait phase labels accurately.

4. Conclusions

In this study, we designed a novel wearable full-body flexible exoskeleton robot as one application of mobile cyberphysical system (CPS). We used a signal acquisition system to collect data for normal walking on the ground, treadmill walking in a setting speed, stair-climbing and stair-descending, and STS. We employed a novel simple method to distinguish all kinds of action mentioned above according to the posture characteristics of angle and pressure. To date, many approaches have been developed to identify the gait phase. Here, we propose a novel gait phase recognition method using lower limb joint angle, plantar pressure, and inclination sensors. According to the characteristics of the gait data, we defined eight gait phases to extract the gait phase features. The deviation distances were calculated and classified by fixed proportion of time in one gait cycle. Then, the gait phase labels of the gait set were obtained. Through offline gait data classification, one gait cycle was divided into eight phases. We built a gait phase recognition model using the gait phase-labeled data. By training the model with a 14-dimensional input vector, we can recognize the gait phase in real time through the lower limb joint angle and plantar pressure sensor data. For the tested set, the model had a CRP of 95.32%; the experimental results demonstrate the effectiveness of the gait phase recognition technique. This novel method can accurately and quickly assess the wearer's moving intentions with a rather simple sensor system. In future work, we intend to improve the existing control strategy of the full-body exoskeleton with the novel gait phase recognition method and develop a gait evaluation method for the exoskeleton robot. The same method can be applied to the upper-limb movement of the robot. Thus, the flexible full-body exoskeleton robot system will be realized in the near future. In the future, we can also use a mobile phone to receive the real-time sensor data of exoskeleton robot and choose the best control mode and walking route for the wearers.

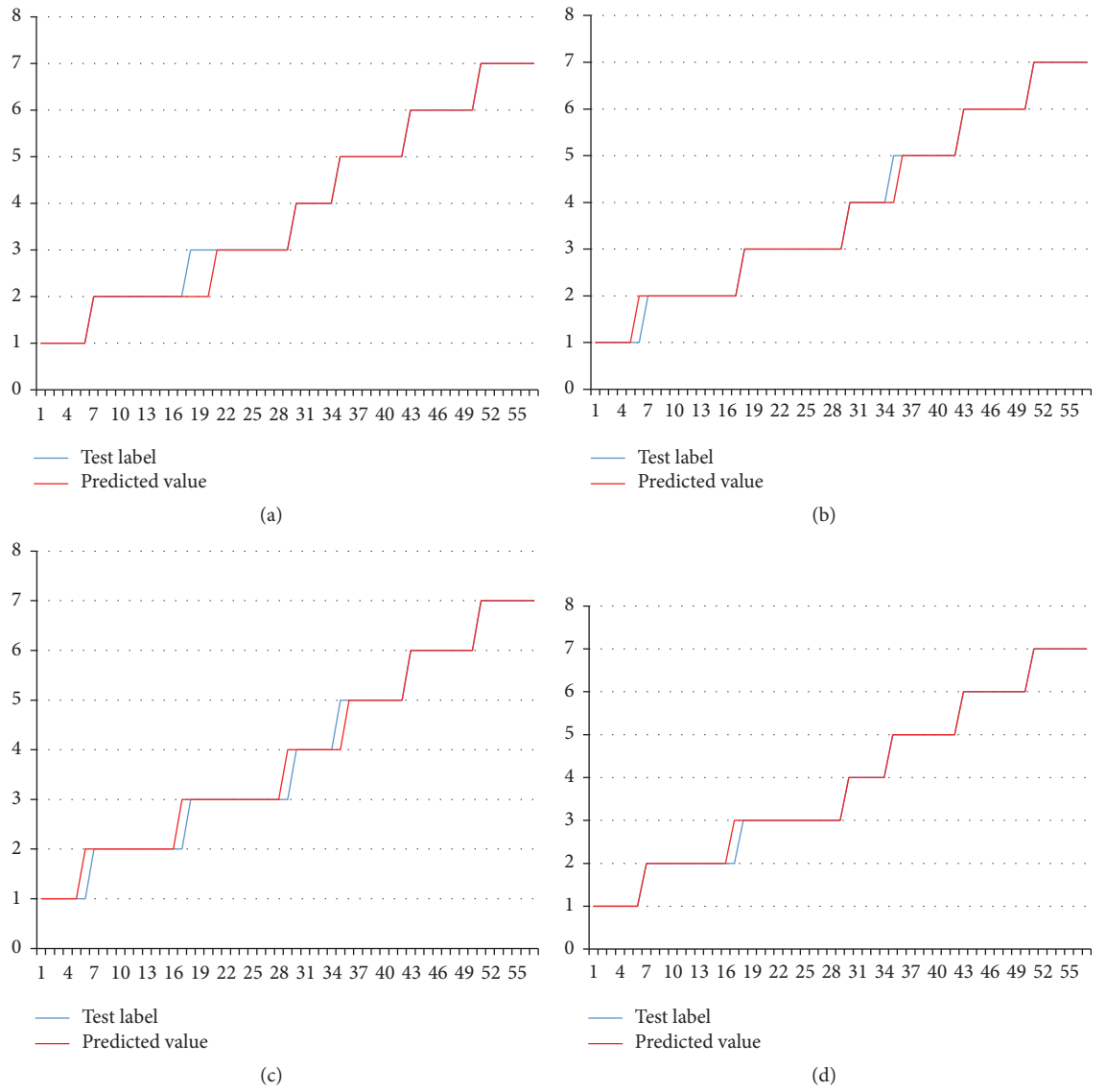


FIGURE 18: The gait phase recognition results of four cycles acquired from four different subjects. *K*NN, gait phase recognition model.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The work described in this paper is partially supported by the National Basic Research Program of China, 973 Program (2015CB351706), NSFC-Shenzhen Joint Fund for Robotics Research Center (U1613219), Guangdong Natural Science Foundation (2016A030313177), Shenzhen Fundamental Research and Discipline Layout project (JCYJ20150925163244742), and Shenzhen Fundamental Research (JCYJ20140901003938996). Finally, the authors thank the members of the SIAT exoskeleton team for supporting the research.

References

[1] Q. Jiang, S. Yang, and J. J. Sánchez-Barricarte, "Can China afford rapid aging?" *SpringerPlus*, vol. 5, no. 1, article 1107, 2016.

[2] N. Jabeur, N. Sahli, and S. Zeadally, "Enabling cyber physical systems with wireless sensor networking technologies, multiagent system paradigm, and natural ecosystems," *Mobile Information Systems*, vol. 2015, Article ID 908315, 2015.

[3] R. Ishak, S. Salleh, S. Olariu, and M. Abdul Aziz, "SPLAI: computational finite element model for sensor networks," *Mobile Information Systems*, vol. 2, no. 1, pp. 77–92, 2006.

[4] S. Kumar and S.-J. Park, "Probability model for data redundancy detection in sensor networks," *Mobile Information Systems*, vol. 5, no. 2, pp. 195–204, 2009.

[5] X. Hu, T. H. S. Chu, H. C. B. Chan, and V. C. M. Leung, "Vita: a crowdsensing-oriented mobile cyber-physical system," *IEEE Transactions on Emerging Topics in Computing*, vol. 1, no. 1, pp. 148–165, 2013.

[6] X. Hu, J. Zhao, B.-C. Seet, V. C. M. Leung, T. H. S. Chu, and H. Chan, "S-aframe: agent-based multilayer framework with context-aware semantic service for vehicular social networks," *IEEE Transactions on Emerging Topics in Computing*, vol. 3, no. 1, pp. 44–63, 2015.

[7] T. Lenzi, S. M. M. De Rossi, N. Vitiello, and M. C. Carrozza, "Intention-based EMG control for powered exoskeletons," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 8, pp. 2180–2190, 2012.

[8] Y. Sankai, "HAL: Hybrid assistive limb based on cybernics," *Springer Tracts in Advanced Robotics*, vol. 66, pp. 25–34, 2010.

[9] K. Suzuki, G. Mito, H. Kawamoto, Y. Hasegawa, and Y. Sankai, "Intention-based walking support for paraplegia patients with robot suit HAL," *Advanced Robotics*, vol. 21, no. 12, pp. 1441–1469, 2007.

[10] A. Pantall and D. Ewins, "Muscle activity during stance phase of walking: comparison of males with transfemoral amputation with osseointegrated fixations to nondisabled male volunteers," *Journal of Rehabilitation Research and Development*, vol. 50, no. 4, pp. 499–514, 2013.

[11] S. Murray, K. H. Ha, C. Hartigan, and M. Goldfarb, "An assistive control approach for a lower-limb exoskeleton to facilitate recovery of walking following stroke," *IEEE Transactions on Neural Systems & Rehabilitation Engineering A Publication of the IEEE Engineering in Medicine & Biology Society*, vol. 23, no. 3, pp. 441–449, 2015.

[12] H. Kazerooni, R. Steger, and L. Huang, "Hybrid control of the berkeley lower extremity exoskeleton (BLEEX)," *International Journal of Robotics Research*, vol. 25, no. 5-6, pp. 561–573, 2006.

[13] D. Liu, X. Wu, W. Du, C. Wang, and T. Xu, "Gait phase recognition for lower-limb exoskeleton with only joint angular sensors," *Sensors*, vol. 16, no. 10, p. 1579, 2016.

[14] S. Oh, E. Baek, S.-K. Song, S. Mohammed, D. Jeon, and K. Kong, "A generalized control framework of assistive controllers and its application to lower limb exoskeletons," *Robotics and Autonomous Systems*, vol. 73, pp. 68–77, 2015.

[15] S. M. H. J. Jaegers, L. D. W. Vos, P. Rispens, and A. L. Hof, "The relationship between comfortable and most metabolically efficient walking speed in persons with unilateral above-knee amputation," *Archives of Physical Medicine and Rehabilitation*, vol. 74, no. 5, p. 521, 1993.

[16] D. Winter, "Human balance and posture control during standing and walking," *Gait and Posture*, vol. 3, no. 4, pp. 193–214, 1995.

[17] GB10000, "Human dimensions of Chinese adults," in *General Administration of Quality Supervision, Inspection and Quarantine of the Peoples Republic of China*, 1988.

[18] G. Q. Duan, "Manufacturing technique of aluminium-clad steel wire and its application in overhead transmission conductor," *Electric Wire & Cable*, 2005.

[19] F. Yamasaki, K. Hosoda, and M. Asada, "An energy consumption based control for humanoid walking," in *proceeding of the IEEE/RSJ International Conference on Intelligent Robots and Systems IEEE Xplore*, vol. 3, pp. 2473–2477, 2002.

[20] N. Sánchez, A. M. Acosta, A. H. A. Stienen, and J. P. A. Dewald, "A multiple degree of freedom lower extremity isometric device to simultaneously quantify hip, knee, and ankle torques," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 5, pp. 765–775, 2015.

[21] C. Li, S. T. Hsieh, and D. I. Goldman, "Multi-functional foot use during running in the zebra-tailed lizard (*Callisaurus draconoides*)," *The Journal of Experimental Biology*, vol. 215, no. 18, pp. 3293–3308, 2012.

[22] H. Kagaya, S. Ito, T. Iwami, G. Obinata, and Y. Shimada, "A computer simulation of human walking in persons with joint contractures," *Tohoku Journal of Experimental Medicine*, vol. 200, no. 1, pp. 31–37, 2003.

[23] K. C. Foucher, B. R. Schlink, N. Shakoor, and M. A. Wimmer, "Sagittal plane hip motion reversals during walking are associated with disease severity and poorer function in subjects with hip osteoarthritis," *Journal of Biomechanics*, vol. 45, no. 8, pp. 1360–1365, 2012.

[24] J. E. Zachazewski, P. O. Riley, and D. E. Krebs, "Biomechanical analysis of body mass transfer during stair ascent and descent of healthy subjects," *Journal of Rehabilitation Research and Development*, vol. 30, no. 4, pp. 412–422, 1993.

[25] L. A. Livingston, J. M. Stevenson, and S. J. Olney, "Stairclimbing kinematics on stairs of differing dimensions," *Archives of Physical Medicine and Rehabilitation*, vol. 72, no. 6, p. 398, 1991.

[26] S. L. Pavão, A. N. Santos, A. B. Oliveira, and N. A. C. F. Rocha, "Postural control during sit-to-stand movement and its relationship with upright position in children with hemiplegic spastic cerebral palsy and in typically developing children," *Brazilian Journal of Physical Therapy*, vol. 19, no. 1, pp. 18–25, 2015.

[27] G. Trevelyan, "Observational gait analysis," 238. rev.sfr.net.

[28] S. Nuzik, R. Lamb, A. VanSant, and S. Hirt, "Sit-to-stand movement pattern. A kinematic study," *Physical Therapy*, vol. 66, no. 11, pp. 1708–1713, 1986.

[29] M. J. Stephens and J. F. Yang, "Loading during the stance phase of walking in humans increases the extensor EMG amplitude but does not change the duration of the step cycle," *Experimental Brain Research*, vol. 124, no. 3, pp. 363–370, 1999.

Hindawi
Mobile Information Systems
Volume 2017, Article ID 4078521, 11 pages
https://doi.org/10.1155/2017/4078521

Research Article

An Anonymous Access Authentication Scheme Based on Proxy Ring Signature for CPS-WMNs

Tianhan Gao,[1] Quanqi Wang,[1] Xiaojie Wang,[2] and Xiaoxue Gong[3]

[1]*Software College, Northeastern University, Shenyang 110819, China*
[2]*School of Software, Dalian University of Technology, Dalian 116024, China*
[3]*School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China*

Correspondence should be addressed to Xiaojie Wang; wangxj1988@mail.dlut.edu.cn
and Xiaoxue Gong; gongxiaoxue@stumail.neu.edu.cn

Received 27 January 2017; Accepted 12 April 2017; Published 4 June 2017

Academic Editor: Jun Cheng

Access security and privacy have become a bottleneck for the popularization of future Cyber-Physical System (CPS) networks. Furthermore, users' need for privacy-preserved access during movement procedure is more urgent. To address the anonymous access authentication issue for CPS Wireless Mesh Network (CPS-WMN), a novel anonymous access authentication scheme based on proxy ring signature is proposed. A hierarchical authentication architecture is presented first. The scheme is then achieved from the aspect of intergroup and intragroup anonymous mutual authentication through proxy ring signature mechanism and certificateless signature mechanism, respectively. We present a formal security proof of the proposed protocol with SVO logic. The simulation and performance analysis demonstrate that the proposed scheme owns higher efficiency and adaptability than the typical one.

1. Introduction

With the prosperous development of mobile communication and versatile mobile devices [1, 2] and the diversification of the network environment [3–5], the requirement of accessing ubiquitous network becomes more and more imperative for Cyber-Physical Systems (CPS) [6]. Owing to the advantages of low cost, expansible, self-healing, fine mobility support, and high efficiency, Wireless Mesh Network (WMN) is regarded as a critical accessing technology of the next generation CPS network [7, 8]. As for the open nature of transmission medium free users' movement, as well as the multihop transmission method, WMN suffers from security issues in both wired and wireless environment. Efficient and secure access authentication technology forms the baseline of CPS-WMN's security. Moreover, user's privacy should also be preserved during the access authentication process. Thus, the security and privacy in CPS-WMNs become the research focus recently [9].

In the past few years, a lot of researches have been carried out for WMN's access authentication. The authors in [10] present an efficient identity-based authentication scheme for WMN using tickets, which avoids multihop wireless communications in order to minimize the authentication delay, while in a complex network environment, with the increasing number of MRs, handover authentication efficiency decreases. The authors of [11] propose an authentication scheme for WMN based on EAP-TLS, although the scheme offers mutual authentication and robustness against malicious attacks. But the asymmetric cryptography mechanisms result in high computation cost. The author [12] improves the access control function of IEEE 802.1X by the port operation so that user may acquire message through the dynamic channel under current or previous access point. However, the requirement of keeping the channel alive during the authentication procedure limits the adaptability of the scheme. Some distributed authentication schemes to reduce the authentication delay have been discussed in [13], while the scheme performs poorly when handling multiple mobile users. A symmetric key generation scheme based on hierarchical multivariable function for WMN is presented in [14], which achieves efficient mutual authentication and key generation for entities, whereas the scheme is not suitable for the scenario when the network users grow rapidly. The identity information of

mobile users is divided into critical information and noncritical information that the critical information is only visible to the mobile user and his/her group manager in [15]. With the help of improved short ring signature mechanism and special binding policy, the scheme is able to provide anonymity during authentication. However, the key escrow problem is inevitable since the private key is generated by the group manager. In general, the literature WMN access authentication schemes suffer from security, privacy, efficiency, and adaptability issues. The needs of an efficient and anonymous authentication scheme for CPS-WMNs are impending.

In terms of the security issues shown above, an anonymous authentication scheme based on proxy ring signature is proposed in this paper. The scheme utilizes a high-efficient proxy ring signature mechanism to achieve proxy-authorization and anonymous authentication which are able to preserve mobile users' privacy. In addition, certificateless signature mechanism is incorporated into our intragroup authentication to obtain high handover efficiency. The formal security proof based on SVO logic and other security analyses show that the proposed scheme possesses such advantages as reliability, anonymity, unforgeability, and reliability. Through the simulation and performance analysis, we demonstrate the efficiency and adaptability of our scheme.

The rest of this paper is organized as follows. Section 2 briefly describes the related preliminaries. Section 3 elaborates the proposed anonymous mutual authentication scheme. Sections 4 and 5 present the security and performance analysis of the scheme, respectively. Finally, we make a conclusion of the scheme and discuss the future research work in Section 6.

2. Preliminaries

2.1. Bilinear Pairing. Let G be an additive group and let G_T be a multiplicative group of the same prime order q and I_{GT} is the generator of G_T. Assume that the discrete logarithm problem is hard on both G and G_T [16]. A mapping e: $G \times G \to G_T$ which satisfies the following properties is called bilinear pairing:

(1) Bilinearity: for all $P, Q \in G$ and , $b \in Z_q^*$, $e(aP, bQ) = e(bP, aQ) = e(P, Q)^{ab}$.

(2) Nondegeneracy: there exists $P, Q \in G$, so that $e(P, Q) \neq I_{GT}$.

(3) Computability: for all $P, Q \in G$, there is an efficient algorithm to compute $e(P, Q) \in G_T$.

2.2. BB1 Encryption. BB1 [17], nonadaptive selective-ID encryption, was presented by Boneh and Franklin in 2003. The BB1 works as follows.

(1) BB1-Setup. Given a security parameter $k \in Z_q^*$, the algorithm works as the following steps.

Step 1. Run G on input k to generate a prime q, two cycle groups $(G_1, +)$, $(G_2, \times)$ of order q, and an admissible bilinear pairing $e : G_1 \times G_1 \to G_2$. Choose a random generator $\in G_1$.

Step 2. Pick a random $s \in Z_q^*$ and set $P_{\text{pub}} = sP$.

Step 3. Choose a cryptographic hash function $H_1 : \{0,1\}^* \to G_1$. Choose a cryptographic hash function $H_2 : G_2 \to \{0,1\}^n$ for some n. The message space is $M = \{0,1\}^n$. The ciphertext space is $C = G_1 \times \{0,1\}^n$. The system parameters are $\{H_1, H_2, G_1, G_2, q, e, P, P_{\text{pub}}\}$. The master key is $s \in Z_q^*$.

(2) BB1-Extract. For a given string ID $\in \{0,1\}^*$, compute $Q_{\text{ID}} = H_1(\text{ID}) \in G_1$ and set the private key d_{ID} to be $d_{\text{ID}} = sQ_{\text{ID}}$.

(3) BB1-Encrypt. To encrypt $m \in M$ under the public key ID, compute $Q_{\text{ID}} = H_1(\text{ID}) \in G_1$, choose a random $r \in Z_q^*$, and set the ciphertext to be $C = \langle rP, m \oplus H_2(g_{\text{ID}}^r)\rangle$, where $g_{\text{ID}} = e(Q_{\text{ID}}, P_{\text{pub}})$.

(4) BB1-Decrypt. Let $c = \langle U, V\rangle \in C$ be a ciphertext encrypted using the public key ID. To decrypt c using the private key $d_{\text{ID}} \in G_1$, compute $m = V \oplus H_2(e(d_{\text{ID}}, U))$.

2.3. Certificateless Signature. Certificateless signature (CLS) [18] allows that users' private key is comprised by the key issued by system and the secret generated by user. In addition, users' public key is conducted by their own secret which avoids key escrow problem. The CLS scheme is mainly used in the Intra-WMN authentication in this paper. The algorithms of CLS [18] are shown as follows.

(1) CLS-Setup. Given security parameter l, prime $q \geq 2^l$, $(G_1, +)$, and $(G_2, \times)$ are cycle groups of order q. Three hash functions are as follows: $H_1 : \{0,1\}^* \to G_1$, $H_2 : \{0,1\}^* \times G_1 \to G_1$, and $H_3 : \{0,1\}^* \times G_1 \times G_2 \to Z_q^*$. Private key generator (PKG) chooses $s \in Z_q^*$ as private key and generates system public key $P_0 = sP$, where P is the generator of G_1. Let $g = e(P, P)$; system public parameters Param = $(G_1, G_2, e, q, P, P_0, g, H_1, H_2, H_3)$.

(2) CLS-Extract-sk. User A sends identity $\text{ID}_\text{A} \in \{0,1\}^*$ to PKG. After authenticating ID_A, PKG generates partial private key of A : $D_\text{A} = sH_1(\text{ID}_\text{A})$.

(3) CLS-Gen-sk. A chooses $x_\text{A} \in Z_q^*$ as secret. A's private key is (x_A, D_A).

(4) CLS-Gen-pk. A computes $P_\text{A} = x_\text{A}P$ as A's public key.

(5) CLS-Sign. A signs message $m \in \{0,1\}^*$, and outputs $\sigma = \{U, \delta\}$ through following steps:

(a) Choose $r \in Z_q^*$ and calculate $U = g^r$.

(b) $V = H_2(\text{ID}_\text{A} \parallel P_\text{A})$.

(c) $h = H_3(m \parallel U \parallel \text{ID}_\text{A} \parallel P_\text{A})$.

(d) $\delta = rP - h(D_\text{A} + x_\text{A}V)$.

(6) CLS-Verify. Verifier B uses $P_0, P_\text{A}, \text{ID}_\text{A}$ to verify the signature $\sigma = \{U, \delta\}$.

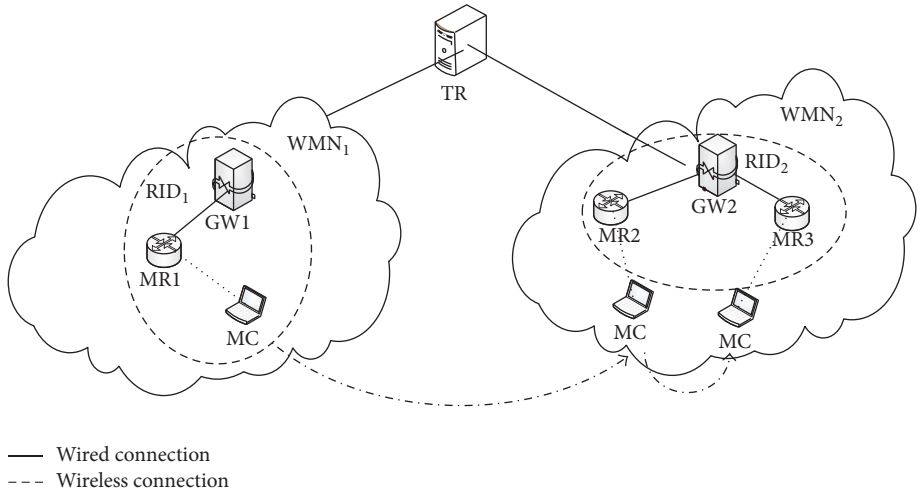


Figure 1: Hierarchical mobile network architecture for CPS-WMNs.

(a) Compute $V = H_2(\mathrm{ID_A} \parallel P_A)$ and $h = H_3(m \parallel U \parallel \mathrm{ID_A} \parallel P_A)$.

(b) Check if the equation $U = e(\delta, P)(e(H_1(\mathrm{ID_A}), P_0)e(P_A V))^h$ is hold. If yes, σ is valid; otherwise, σ is invalid.

2.4. Proxy Ring Signature. Proxy ring signature (PRS) [19] allows an original signer delegate authorization to a group of signers in which every member in the group can represent the original signer to sign the message and is able to keep anonymous. In this paper, we incorporate proxy ring signature into the access authentication process of WMN, which not only achieves mutual authentication between mobile user and accessed network but also solves the problem of privacy preserving for mobile user. The algorithms of PRS are as follows.

(1) PRS-Setup. Given secure parameter K as system input and the output is (G_1, G_2, q, e, P). G_1 is a cyclic additive group generated by the generator P, whose order is prime q, and G_2 is a cyclic multiplicative group of the same prime order of q. $e : G_1 \times G_1 \to G_2$ is a bilinear pairing map. In addition, there are two hash functions: $H_0 : \{0,1\}^* \times G_1 \to Z_q^*$ and $H_1 : \{0,1\}^* \to G_1$.

(2) PRS-Generation. Original signer O chooses $x_0 \in Z_q^*$ as the private key and calculates the public key $Y_o = x_o P$. u_i belonging to proxy signer group U randomly chooses private key $x_i \in Z_q^*$ and calculates the public key $Y_i = x_i P$.

(3) PRS-Delegation. O generates a warrant m_ω which includes the descriptions of the relationship between O and proxy signer. O chooses a random number $r \in Z_q^*$, calculates $R = rp$, $s = r + x_o H_o(m_\omega, R) \bmod q$, and then sends (m_ω, R, s) to the group $U = \{u_1, \ldots, u_n\}$ of proxy signers.

(4) PRS-Verify-Auth. After receiving (m_ω, R, s), each proxy signer μ_i checks $sP \stackrel{?}{=} R + Y_o H_o(m_\omega, R) \bmod q$. If the verification fails, the authorization will be rejected. Otherwise, u_i calculates his own proxy signing key $\mathrm{psk}_i = s + x_i H_o(m_\omega, R) \bmod q$.

(5) PRS-Sign. The proxy signer u_s signs message $m \in \{0,1\}^*$ as follows:

(a) For all $i \in \{1, 2, \ldots, n\}$ and $i \neq s$, choose a random number $r_i \in Z_q^*$ and calculate $\delta_i = r_i P$.

(b) Calculate $\delta_s = (1/\mathrm{psk}_s)(H_1(m \parallel m_\omega) - \sum_{i \neq s} r_i (R + H_o(m_\omega, R)(Y_o + Y_i))$.

(c) Send $\delta = (\delta_1, \delta_2, \ldots, \delta_n, m, m_\omega, R)$ to the verifier.

(6) PRS-Verify-Sign. After receiving δ from the proxy signer, the verifier checks if the following equation holds with the public key $Y_0, Y_1, \ldots, Y_n$:

$$\prod_{i=1}^{n} e\left(R + H_0\left(m_\omega, R\right)\left(Y_o + Y_i\right), \delta_i\right) \stackrel{?}{=} e\left(P, H_1\left(m \parallel m_\omega\right)\right). \quad (1)$$

If yes, δ is valid. Otherwise, δ is invalid.

3. Anonymous Mutual Authentication Scheme

3.1. Hierarchical Mobile Network Architecture. As shown in Figure 1, a hierarchical mobile network architecture is designed for CPS-WMNs. In the first level, Trusted Root (TR), as original signer who can delegate signing right to proxy signers, is creditable to all of the network entities. In the second level, there are many WMNs that each one can be regarded as a group of proxy signers including Gateway (GW), Mesh Routers (MRs), and mobile Mesh Clients (MCs). MC is able to handover across different WMNs or between different MRs in the same WMN. To achieve mutual authentication between MC and visiting network based on PRS, we build the group of proxy rings for network entities in terms of the hierarchical mobile network architecture shown above. Assuming that a group of the proxy ring (abbreviated as a

Table 1: Symbols and descriptions.

Symbols	Descriptions
RID_i	Ring i's identification
PK_X/SK_X	The public/private key of entity X
Param	System parameters
W_{ri}	The warrant for the members in ring i
$Auth_X$	The authorization of proxy signature for X
$K_{X\text{-}Y}$	The session key between X and Y
Enc_$K\{M\}$	Using symmetric key K to encrypt message M
ENCR_ALG_$PK_X\{M\}$	Using algorithm ALG and Xs public key PK_X to encrypt message M
SIGN_ALG_$SK_X\{M\}$	Using algorithm ALG to sign message M with X's signing key SK_X
TS_i	The current timestamp
$M_1 \parallel M_2$	Concatenation of messages M_1 and M_2

ring) is composed of GW, MRs (connected with the GW), and MCs (connected with the MRs). We denote ring ID as RID_i in Figure 1 (RID_1 means ring 1 and RID_2 means ring 2). GW takes the role of a manager of the ring and is responsible for managing and maintaining the members in the ring.

The symbols used sections are shown in Table 1.

3.2. Trust Model. As shown in Figure 2, the trust model is presented according to the mobile network architecture. TR is trusted by all the entities. GW in different CPS-WMNs does not trust each other. Moreover, different MR belonging to the same GW does not own trust relationship, the same as the MRs in different CPS-WMNs. In addition, we assume that GW is trusted by the MR which is connected to itself. MC only trusts the MR in its home CPS-WMN. The main objective of our proposed scheme is to set up the trust relationship between MC and the accessed MR during MC's roaming.

3.3. System Initialization. As the trusted root, TR generates Param and broadcasts it to all entities. Param = $\{G_1, G_2, e : G_1 \times G_1 \to G_2, P \in G_1, PK_{TR} = sP, H_1 : \{0,1\}^* \to G_1, H_2 : \{0,1\}^* \times G_1 \to Z_q^*, g = (P, P)\}$, where q is the order of G_1 and G_2 and s is the master key of TR. All entities' public key in the ring should be delivered to TR. In addition, GW generates the ring's public and private keys through random choosing of $SK\_R_i \in Z_q^*$ as the private key; the corresponding public key is $PK\_R_i = SK\_R_i \cdot P$. PK_R_i is shared by all the members in the ring, while SK_R_i is only allocated to the legitimate members who are authenticated by TR in system initialization phrase.

3.4. Inter-WMN Authentication Protocol. When MC wants to leave the WMN it belonged to and accesses another WMN, the MC needs to achieve mutual Inter-WMN authentication with the visiting WMN. As shown in Figure 1, when the MC in WMN_1 wants to access WMN_2 and connect to MR_2, MC triggers mutual authentication with MR_2. The mutual authentication details are shown in Figure 3.

(1) $MR_2 \to MC : PK_{GW_2}, PK_{MR_2}$. MR_2 broadcasts PK_{MR_2} and PK_{GW_2} to MC. MC executes PRS-Verify-auth to verify PK_{GW_2}.

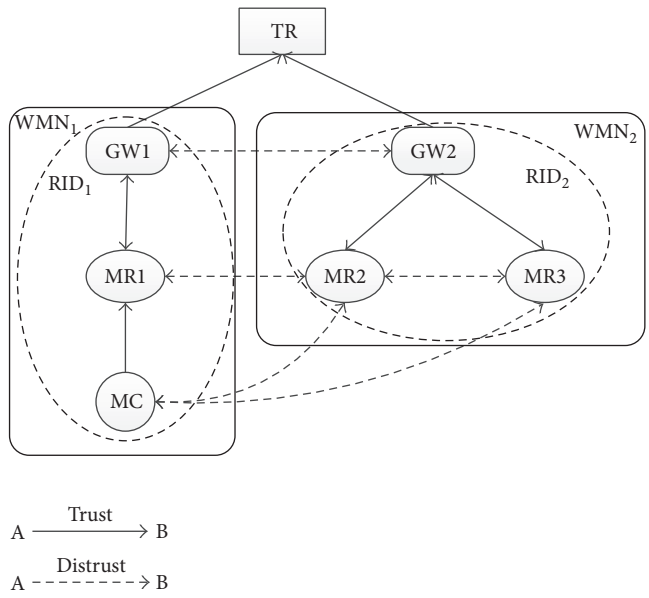


Figure 2: Trust model.

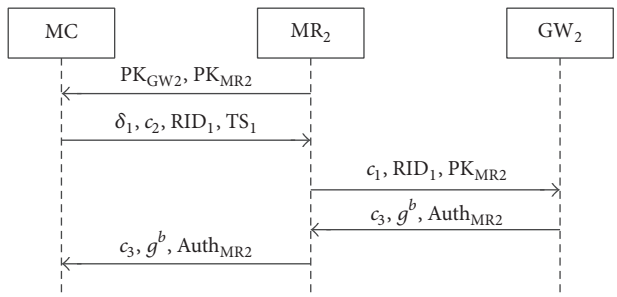


Figure 3: The workflow of Inter-WMN authentication protocol.

(2) $MC \to MR_2 : \delta_1, c_2, TS_1, RID_1$. MC calculates $c_1 = ENCR\_BB1\_PK_{GW_2}\{g^a\}$, $c_2 = ENCR\_BB1\_PK_{MR_2}\{c_1\}$, and $\delta_1 = SIGN\_PRS\_SK_{MC}\{TS_1\}$, where $g^a \in G_1$ is the parameter for session key negotiation and TS_1 is the current timestamp. MC sends δ_1, c_2, TS_1, and RID_1 to MR_2.

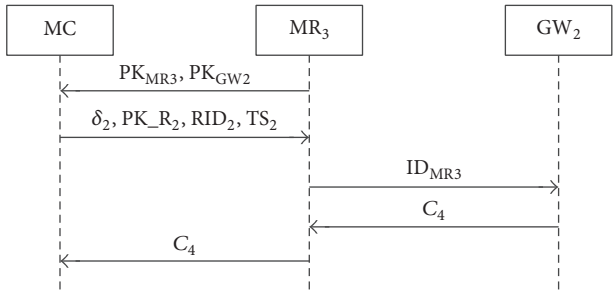


Figure 4: The workflow of Intra-WMN authentication protocol.

(3) $MR_2 \rightarrow GW_2 : c_1, PK_{MR_2}, RID_1$. After receiving (2) from MC, MR_2 checks the freshness of TS_1. If fresh, MR_2 decrypts c_2 to obtain c_1. MR_2 then sends c_1, RID_1, and PK_{MR_2} to GW_2. Meanwhile, MR_2 utilizes PRS-Verify-sign to verify δ_1 through requesting the ring members public key from TR in terms of RID_1. If δ_1 is valid, MC is regarded as a legal user.

(4) $GW_2 \rightarrow MR_2 : c_3, Auth_{MR_2}, g^b$. After receiving (3) from MR_2, GW_2 decrypts c_1 to obtain g^a, choose $g^b \in G_1$ and generate the warrant $W_{r_2} = sk_{GW_2} H_1(ID_{r2})$, where sk_{GW_2} is GW_2's private key. GW_2 computes $AUTH_{MR_2} = sk_{GW_2} H_1(PK_{MR_2})$, session key $K_{GW_2\text{-}MC} = g^{ab}$, and $c_3 = Enc\_K_{GW_2\text{-}MC}\{W_{r_2}\}$. GW_2 then sends c_3, $Auth_{MR_2}$, and g^b to MR_2. Finally, GW_2 stores $K_{GW_2\text{-}MC}$.

(5) $MR_2 \rightarrow MC : c_3, Auth_{MR_2}, g^b$. MR_2 relays c_3, $Auth_{MR_2}$, and g^b to MC when getting (4) from GW_2.

After receiving (5) from MR_2, MC calculates session key $K_{MC\text{-}GW_2} = g^{ab}$ to decrypt c_3 for obtaining W_{r_2}. MC checks $e(P, Auth_{MR_2}) \stackrel{?}{=} e(PK_{GW_2}, H_1(PK_{MR_2}))$, if equal, MC makes sure to access a legal WMN. MC then calculates $U = g^r$ and $V = H_2(RID_2 \parallel PK\_R_2)$, where $r \in Z_q^*$. MC stores U, V, and $K_{MC\text{-}GW_2}$.

3.5. Intra-WMN Authentication Protocol. After finishing Inter-WMN authentication, MC will obtain W_{r_2} issued by GW. When MC moves from one MR to another in the same WMN, we use CLS [14] to achieve efficient Intra-WMN authentication. As shown in Figure 1, assuming that MC and MR_2 finished Inter-WMN authentication, when MC wants to move from MR_2 to MR_3, the Intra-WMN authentication protocol is triggered as shown in Figure 4.

(1) $MR_3 \rightarrow MC : PK_{MR_3}, PK_{GW_3}$. MR_3 broadcasts PK_{MR_3} and PK_{GW_2} to MC.

(2) $MC \rightarrow MR_3 : \delta_2, PK\_R_2, TS_2, RID_2$. MC calculates $h = H_2(TS_2 \parallel U \parallel RID_2 \parallel PK\_R_2)$ and $\delta_2 = rP - h(W_{r_2} + SK\_R_2 V)$, where U and V are generated and stored in the process of Inter-WMN authentication. TS_2 is the current timestamp. MC sends TS_2, δ_2, PK_R_2, and RID_2 to MR_3.

(3) $MR_3 \rightarrow GW_2 : ID_{MR_3}$. After receiving (2) from MC, MR_3 checks the freshness of TS_2. If fresh, MR_3 adopts CLS-Verify to verify δ_2. If δ_2 is valid, MC is regarded as a legal user. MR_3 then sends ID_{MR_3} to GW_2.

(4) $GW_2 \rightarrow MR_3 : C_4$. After receiving ID_{MR_3}, GW_2 uses the previously saved key $K_{GW_2\text{-}MC}$ to encrypt ID_{MR_3} to produce the ciphertext C_4. Then, GW_2 sends C_4 to MR_3.

(5) $MR_3 \rightarrow MC : C_4$. MR_3 relays C_4 to MC after getting (4) from GW_2.

MC uses the previously saved key $K_{GW_2\text{-}MC}$ to decrypt C_4. If the decryption is successful, MC makes sure to access a legal MR.

4. Security Analysis of the Proposed Scheme

In order to prove the security of our scheme, we first take a fundamental security analysis. Then we choose SVO logic [20] to analyze the proposed protocols. SVO logic was presented by Syverson and van Oorshot in 1994 based on BAN logic, GNY logic, AT logic, and VO logic [21]. SVO holds the features of complete semantics, expansibility, and practicality.

4.1. Fundamental Security Analysis. According to the mobile network architecture shown in Figure 1, we will first present fundamental security analysis of the proposed scheme in the following aspects: anonymity, unforgeability, and reliability.

Anonymity. During Inter-WMN authentication, the accessed network checks the legality of MC through verifying the signature $\sigma_1 = SIGN\_PRS\_SK_{MC}\{TS_1\}$ offered by MC. The accessed network is able to know the ring where MC comes from but cannot tell the real identity of MC since it is hidden in the ring. So the anonymity of MC is guaranteed. In addition, when handover occurred, accessed network verifies the certificateless signature $\sigma_i = rP - h(W_{r_i} + SK\_R_i V)$ to authenticate MC. In this paper, the proposed scheme adopts enhanced certificateless signature mechanism: $V = H_2(RID_i \parallel PK\_R_i)$, $h = H_2(TS_2 \parallel U \parallel RID_i \parallel PK\_R_i)$, and $\sigma_i = rP - h(W_{r_i} + SK\_R_i V)$. Thus, with the help of the ring, the identity of MC is also kept private to achieve anonymity.

Unforgeability. Firstly, only TR can calculate the authority for the proxy group. If the adversary does not know TR's private key, he fails to compute the legal authority. Secondly, the only legal proxy signer can generate legal proxy ring signature. If the adversary cannot obtain the authority, he cannot generate the legal signature. Thus, the proxy ring signature is unforgeable. Finally, only trusted GW can issue W_{r_i} to foreign MC, if the adversary does not know GW's private key for certificateless signature, the legal cannot be computed. Moreover, if the adversary cannot obtain the other part of the private key SK_R_i, the legal certificateless signature also cannot be computed. Consequently, certificateless signature is unforgeable based on the security of related entity's private key.

Reliability. In Inter-WMN authentication, if adversary does not know the BB1 secret key of GW_2, then $c_1 = \text{ENCR\_BB1\_PK}_{GW_2}\{g^a\}$ cannot be decrypted. The adversary thus cannot negotiate the correct key with MC. So GW_2 is legal. Likewise, if adversary does not know MR_{2BB1_SK}, he fails to decrypt $c_2 = \text{ENCR\_BB1\_PK}_{MR_2}\{c_1\}$ to obtain c_1, thus MR_2 is legal. Furthermore, the legal proxy ring signature cannot be generated since adversary does not know MC_{PRS_SK}, so the Inter-WMN authentication protocol is reliable. In addition, during Intra-WMN authentication, adversary fails to generate legal signature σ_2, if he cannot obtain W_{r_2}, then MC is thus legal.

4.2. Security Proof of the Proposed Protocols under SVO. SVO logic is not only semantic sound, but also convenient. In terms of our scheme, SVO owns advantages over other logic analysis methods in the following aspects: (1) The axioms in SVO can be adjusted or expanded easily to meet the security proof needs rather than BAN or other logical approaches. (2) SVO is detailed and legible which helps to accurately express the actual meaning of the protocol and thus avoid the misunderstandings. (3) SVO is rigorous and reliable, and the semantics is clear. We first give the grammatical components of SVO logic as follows.

P believes X: indicating that P believes that proposition is right.

P received X: indicating that P received the message including X.

P says X: indicating that P sends a message including X.

P controls X: indicating that P is a trusted authority on X.

P sees X: indicating that P possesses message X.

fresh(X): indicating that X is random number generated in running scheme.

$P \overset{K}{\longleftrightarrow} Q$: indicating that K is a key shared exclusively by P and Q.

$\{X\}_K$: indicating that the ciphertext is output by encrypting X through key.

$[X]_K$: indicating that the message is generated by signing X through key.

$\text{PK}_\sigma(A, K)$: indicating that K is the public signature verification key associated with principal A.

$\text{PK}_\delta(A, K)$: indicating that K is the key agreement key associated with principal A.

$\text{PK}_\psi(A, K)$: indicating that K is the public encryption key associated with principal A.

$\text{SV}(X, K, Y)$: indicating that given signed message X, applying K to it as a signature verification key verifies Y as the message signed with the corresponding private key.

SVO logic includes two initial rules and twenty axioms, part of which are regular axioms and others are axiom templates that include formula variables. We only present part of the axioms used in the following security proof. All the axioms can be found in [20].

Two inference rules are as follows:

(1) Modus ponens MP: φ and $\varphi \supset \psi$ infer ψ

(2) Necessitation Nec: $\vdash \varphi$ infer $\vdash P$ believes φ

φ and ψ are metalinguistic symbols used to refer to arbitrary formula. $\vdash$ is a metalinguistic symbol. $\vdash \varphi$ means that φ is a theorem.

There are twenty SVO axioms. We list only several axioms associated with this article. For any principal P, Q and formula φ, ψ:

(A1) P believes φ and P believes $(\varphi \supset \psi) \supset P$ believes ψ

(A2) $\text{PK}_\sigma(Q, K)$ and R received $X \wedge \text{SV}(X, K, Y) \supset Q$ says Y

(A3) P received $(X_1, \ldots, X_n) \supset p$ received X_i

(A4) P received $\{X\}_K$ and P sees $K^{-1} \supset P$ received X

In SVO, some generic goals should be satisfied. This does not mean a definitive list of the goals that our protocol should meet. In our paper, we should achieve the mutual authentication between MC and MR. For this purpose, we just need that MR and MC could make sure of the legality for each other. So on the basis of the generic goals, we make the appropriate modifications. The goals of Inter-WMN authentication protocol could be described as follows.

(G1′) MR believes $[\text{TS}_1]^{-1}_{K_s}$

(G2′) MC believes $[H(\text{PK}_{\text{MR}})]^{-1}_{K_{\text{GW}}}$

SVO Logic Initial Assumptions

(P1) MC believes $\text{PK}_\sigma(S, K_s)$

(P2) MC believes $\text{SV}([\text{TS}_1]^{-1}_{K_s}, K_s, \text{TS}_1)$

(P3) MC believes fresh(TS_1)

(P4) MR believes (MC says $(\text{TS}_1, \text{RID}_i) \supset \text{TS}_1$)

(P5) MR believes MC says $\text{TS}_1 \supset$ MC says TS_1

(P6) MR believes MR received $\{\{*_1\}_{K_{\text{GW}}}\}_{K_{\text{MR}}}, *_2, \text{RID}_i, [*_3]_{*_4}$

(P7) MR believes MR received $\{\{*_1\}_{K_{\text{GW}}}\}_{K_{\text{MR}}}, *_2, \text{RID}_i, [*_3]_{*_4} \supset$ MR received $\{\{*_1\}_{K_{\text{GW}}}\}_{K_{\text{MR}}}, \text{TS}_1, \text{RID}_i, [\text{TS}_1]^{-1}_{K_s}$

(P8) GW believes $\text{PK}_\sigma(\text{GW}, K_{\text{GW}})$

(P9) GW believes $\text{SV}([H(\text{PK}_{\text{MR}})]^{-1}_{K_{\text{GW}}}, K_{\text{GW}}, H(\text{PK}_{\text{MR}}))$

(P10) MC believes (GW says $(H(\text{PK}_{\text{MR}})) \supset H(\text{PK}_{\text{PK}_{\text{MR}}})$)

(P11) MC believes GW says $H(\text{PK}_{\text{MR}}) \supset$ GW says $H(\text{PK}_{\text{MR}})$

(P12) MC believes MC received $\{[*_5]^{-1}_{K_{\text{GW}}}\}_{K_{\text{GW-MC}}}, *_6, [*_7]_{*_8}$

(P13) MC believes MC received $\{[*_5]^{-1}_{K_{\text{GW}}}\}_{K_{\text{GW-MC}}}, *_6, [*_7]_{*_8} \supset$ MC received $\{[*_5]^{-1}_{K_{\text{GW}}}\}_{K_{\text{GW-MC}}}, g^b, [H(\text{PK}_{\text{MR}})]^{-1}_{K_{\text{GW}}}$

Where K_s is the public key for proxy signer and TS_1 is the current timestamp, $*_i$ means the part that subject cannot understand. The proof is as follows.

From (P6), (A1), (A3), and Nec, we have

$$\text{MR believes MR received } [*_3]_{*_4}. \quad (2)$$

From (2), (P7), (A1), (A3), and Nec, we have

$$\text{MR believes MR received } [TS_1]^{-1}_{K_s}. \quad (3)$$

From (P5), (P1), (P2), (3), (A1), (A2), and Nec, we have

$$\text{MR believes MC says } [TS_1]^{-1}_{K_s}. \quad (4)$$

From (4), (P4), and (A1), we have the following.

MR believes $[TS_1]^{-1}_{K_s}$; *(GI′) is then proved*. In the same way as above, we can get

$$\text{GW believes } [TS_1]^{-1}_{K_s}. \quad (5)$$

From (P12), (A1), (A3), and Nec, we have

$$\text{MC believes MC received } [*_7]_{*_8}. \quad (6)$$

From (6), (P13), (A1), (A3), and Nec, we have

$$\text{MC believes MC received } [H(PK_{MR})]^{-1}_{K_{GW}}. \quad (7)$$

From (P11), (P8), (P9), (7), (A1), (A2), and Nec we have

$$\text{MC believes GW says } [H(PK_{MR})]^{-1}_{K_{GW}}. \quad (8)$$

From (8), (P10), and (A1), we have the following.

MC believes $[H(PK_{MR})]^{-1}_{K_{GW}}$; *(G2′) is thus proved*.

Similar to Inter-WMN authentication protocol, the goal of Intra-WMN authentication is also mutual authentication between MR and MC. The difference is that the MR is in MC's accessed WMN. The security proof of Intra-WMN authentication protocol is described as follows:

(G3′) MR believes $[TS_2]^{-1}_{K_{MC}}$

(G4′) MC believes $[H(PK_{MR})]^{-1}_{K_{GW}}$

SVO Logic Initial Assumptions

(P14) MC believes $PK_\sigma(MC, K_{MC})$

(P15) MC believes $SV([TS_2]^{-1}_{K_s}, K_s, TS_2)$

(P16) MC believes $\text{fresh}(TS_2)$

(P17) MR believes (MC says $(TS_2, RID_i) \supset TS_2$)

(P18) MR believes MC says $TS_2 \supset$ MC says TS_2

(P19) MR believes MR received $PK_\sigma(W_{r_i}, K_{W_{r_i}}), *_1, RID_i, [*_2]^{-1}_{K_{MC}}$

(P20) MR believes MR received MR received $PK_\sigma(W_{r_i}, K_{W_{r_i}}), *_1, RID_i, [*_2]^{-1}_{K_{MC}} \supset$ MR received $PK_\sigma(W_{r_i}, K_{W_{r_i}}), TS_2, RID_i, [TS_2]^{-1}_{K_{MC}}$

(P21) GW believes $PK_\sigma(GW, K_{GW})$

(P22) GW believes $SV([H(PK_{MR})]^{-1}_{K_{GW}}, K_{GW}, H(PK_{MR}))$

(P23) MC believes (GW says $(H(PK_{MR})) \supset H(PK_{PK_{MR}})$)

(P24) MC believes GW says $H(PK_{MR}) \supset$ GW says $H(PK_{MR})$

(P25) MC believes MC received $\{[*_3]_{*_4}\}$

(P26) MC believes MC received $\{[*_3]_{*_4} \supset$ MC received $[H(PK_{MR})]^{-1}_{K_{GW}}\}$

Where TS_2 is the current timestamp, $*_i$ means the part that subject cannot understand. The proving process is shown as follows.

From (P19), (A1), (A3), and Nec, we have

$$\text{MR believes MR received } [*_2]^{-1}_{K_{MC}}. \quad (9)$$

From (9), (P20), (A1), (A3), and Nec, we have

$$\text{MR believes MR received } [TS_2]^{-1}_{K_{MC}}. \quad (10)$$

From (P18), (P14), (P15), (10), (A1), (A2), and Nec, we have

$$\text{MR believes MC says } [TS_2]^{-1}_{K_{MC}}. \quad (11)$$

From (11), (P17), and (A1), we have the following.

GR believes $[TS_2]^{-1}_{K_{MC}}$; *(G3′) is proved*.

From (P25), (A1), (A3), and Nec, we have

$$\text{MC believes MC received } [*_3]_{*_4}. \quad (12)$$

From (12), (P26), (A1), (A3), and Nec, we have

$$\text{MC believes MC received } [H(PK_{MR})]^{-1}_{K_{GW}}. \quad (13)$$

From (P24), (P21), (P22), (13), (A1), (A2), and Nec, we have

$$\text{MC believes GW says } [H(PK_{MR})]^{-1}_{K_{GW}}. \quad (14)$$

From (14), (P23), and (A1), we have the following.

MC believes $[H(PK_{MR})]^{-1}_{K_{GW}}$; *(G4′) is thus proved*.

5. Simulation and Performance Analysis

CPS-WMN has limited resource in the computation ability of nodes and operating bandwidth, so the performance of authentication scheme plays an important role in the practicability of CPS-WMNs. The simulation and performance analysis focus on the efficiency of system initialization and the handover process. In addition, in order to demonstrate the high efficiency of our scheme, we give a comparison analysis between our scheme and PEACE [15].

5.1. Simulation Environment. We do simulations for PRS and PEACE using OMNET++ (4.4) simulation platform to get average results based on 20-time experiments. In the process of bilinear group instantiation, we use Tate pairing in the MNT curve [22].

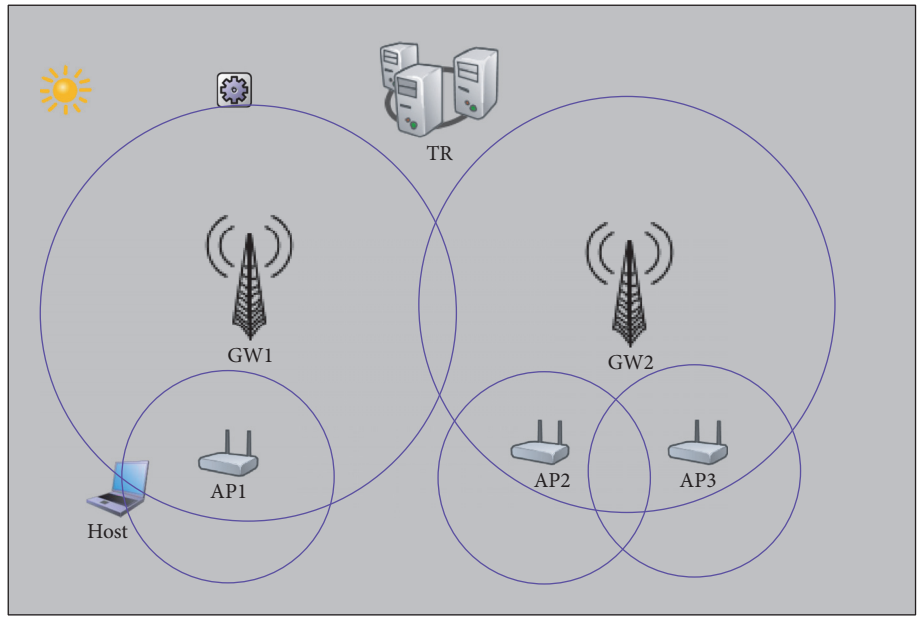


Figure 5: Simulation network topology.

As shown in Figure 5, the initial topological structure of simulation environment is composed of one TR, two GWs, three APs, and one host. These nodes are arranged in a 420 m 300 m simulation space according to the hierarchical network architecture. The TR generates initial parameters for the system. The wireless covering radius is 100 m. AP represents MR, whose covering radius is 45 m. TR, GW, and AP are fixed nodes. Host represents MC, which will take a movement from coordinate (10,250) to coordinate (400,250) by speed 1 m/s. During this process, host firstly accesses the coverage of AP1 and triggers the Inter-WMN authentication. Then, host leaves AP1 to AP2 and the Inter-WMN authentication takes place again. When host moves on from AP2 to AP3, the Intra-WMN authentication protocol should be executed. The details of the parameters and values are shown in Table 2.

(1) The internal structure of the network node shown in Figure 6.

(2) Wlan and eth module: implementation of ethernet and 802.11 capabilities.

(3) NetworkLayer: to achieve network-level functions and as the interface of upper and lower layer.

(4) TCPapp: template for TCP applications.

(5) RoutingTable: the table of routing status.

(6) InterfaceTable: the table of network interfaces.

(7) NotificationBoard: notification about "events" such as wireless handovers.

5.2. Performance Analysis of System Initialization. The delay of system initialization is the period from the simulation start to the first movement of the host. The relationship between the number of nodes and system initialization delay is shown in Figure 7, where the number of nodes could be adjusted as needed. The system initialization includes authorization from original signer to proxy signers, the public key registration for ring members, and the generation of public and private keys for the ring members. Figure 6 shows that the delay of system initialization would increase with the increasing network scale.

Table 2: Parameters and values.

Parameters	Values
Scenario area	420 m × 300 m
Number of nodes	7
Routing protocol	AODV
MAC protocol	IEEE 802.11
Channel	Wireless channel
Simulation time	480 S
Packet length	512 bytes
Node energy	1 J

5.3. Performance Analysis of Authentication Protocols. In this section, we focus on the delay of Inter-WMN authentication and Intra-WMN authentication. The delay of Inter-WMN authentication means the period from AP receiving an access requirement of a new host to the end of Inter-WMN authentication. The delay of Intra-WMN authentication is the period from AP receiving a handover requirement of a host to the end of Intra-WMN authentication.

Figure 8 shows the relationship between the number of ring members and the delay of access authentication scheme. From the result we can see that the efficiency of Intra-WMN authentication is higher than that of Inter-WMN authentication with the increasing number of ring members. During Inter-WMN authentication, the main cost is from

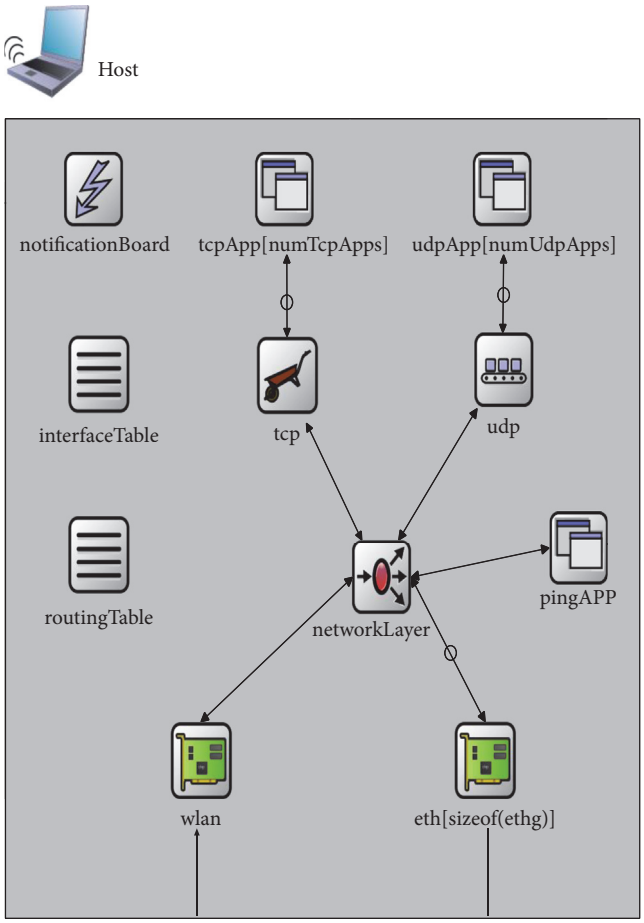


Figure 6: Node internal structure.

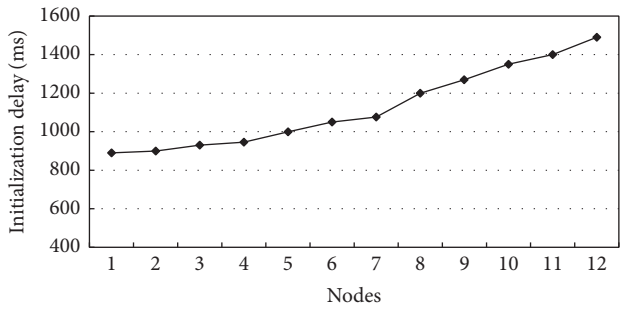


Figure 7: Relationship between system initialization delay and the number of nodes.

verifying the proxy ring signature. For the use of high-efficient ring setup policy, the verifier could acquire all ring members' public keys from TR at once, which help to reduce the delay of communication. In addition, in the process of Intra-WMN authentication, the utilization of certificateless signature makes the scheme independent of the number of ring members that would not lead to obvious delay.

5.4. The Efficiency Analysis of Intra-WMN Authentication Protocol. As shown in Figure 9, we make the comparison analysis of Intra-WMN authentication delay between PRS and PEACE [15]. The delay of PRS is obviously lower than PEACE since PEACE adopts multiple bilinear pairing operations and exponential operations which lead to high computation cost. In the Intra-WMN authentication, we use more efficient certificateless signature which only includes two scalar multiplications in group and one hash operation. Moreover, we just need one bilinear pairing operation, two exponential operations, and one hash operation during the verification process. Thus, the computation cost is obviously reduced in PRS.

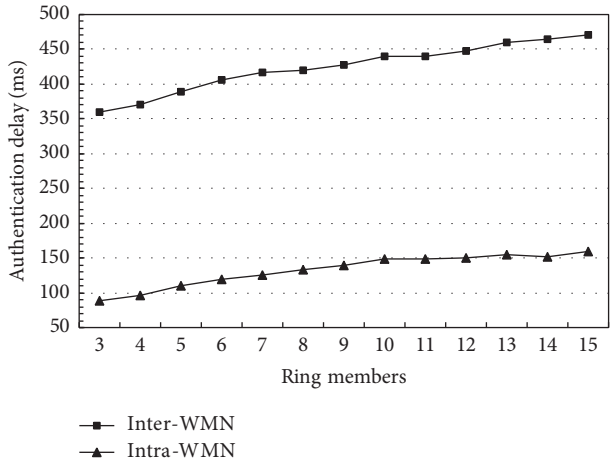


Figure 8: Relationship between authentication delay and the number of ring members.

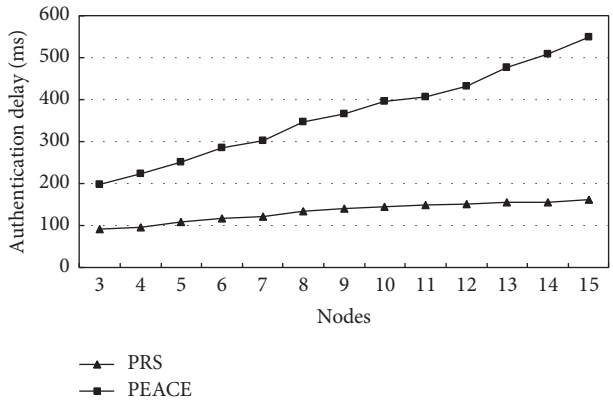


Figure 9: Comparison of the Intra-WMN authentication delay between PRS and PEACE.

In short, the main cost of PRS is from the process of system initialization, while the access authentication delay is obviously dropped down. In addition, the delay of access authentication will not elevate much with the increasing number of nodes in the ring. Although the delay of system initialization increases with the increasing number of ring members, the result of simulation shows that the delay would be controlled in a reasonable range. Comparing to the typical scheme (PEACE), our proposed scheme performs more efficiently, especially during the Intra-WMN authentication.

We further compared the computational overhead of PRS scheme and PEACE scheme during the signing and

Table 3: Comparison of the computational overhead between PRS and PEACE.

Scheme	Signing algorithm	Verifying algorithm
PEACE	2BP + 8SM	(3 + 2|URL|) BP + 6SM
Our scheme	1SM + 2E	3BP + E

verifying phases. In Table 3, BP represents a bilinear mapping operation, SM represents scalar multiplication in $G1$, E represents exponentiation in $G1$, and |URL| represents the time of searching revocation list. From the result we can see that PRS performs more efficiently than PEACE in terms of computational overhead.

6. Conclusions

Anonymous access authentication is an essential approach to address the security issue of CPS-WMNs. In this paper, we propose a novel anonymous access authentication scheme based on proxy ring signature for CPS-WMNs. The scheme is elaborated with the hierarchical mobile network architecture and the corresponding mutual authentication protocols, which achieve high-efficient mutual authentication and satisfy the privacy requirements. The fundamental security and the security proof of the authentication protocols under SVO logic demonstrate the robustness of our scheme. Moreover, the simulation and performance analysis show that the proposed scheme owns higher efficiency and adaptability than the typical.

In our future research, some novel and robust encryption and signature mechanisms will be introduced to make our scheme more resilient. Moreover, how to secure the routing procedure of WMNs under the proposed hierarchical architecture forms another future task.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by National Natural Science Foundation of China under Grant no. 61402095 and Grant no. 61300196 and China Fundamental Research Funds for the Central Universities under Grant no. N120404010 and Grant no. N130817002. This work was also supported in part by Soonchunhyang University Research Fund and the Scientific and Technological Research Program of Chongqing Municipal Education Commission (KJ1500440).

References

[1] Z. Ning, F. Xia, X. Kong, and Z. Chen, "Social-oriented resource management in cloud-based mobile networks," *IEEE Cloud Computing*, vol. 3, no. 4, pp. 24–31, 2016.

[2] Z. Ning, F. Xia, X. Hu, Z. Chen, and M. S. Obaidat, "Social-oriented adaptive transmission in opportunistic internet of smartphones," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 810–820, 2017.

[3] X. Hu, X. Li, E. C.-H. Ngai, V. C. M. Leung, and P. Kruchten, "Multi-dimensional context-aware social network architecture for mobile crowdsensing," *IEEE Communications Magazine*, vol. 52, no. 6, pp. 78–87, 2014.

[4] X. Hu, J. Zhao, B.-C. Seet, V. C. M. Leung, T. H. S. Chu, and H. Chan, "S-aframe: agent-based multilayer framework with context-aware semantic service for vehicular social networks," *IEEE Transactions on Emerging Topics in Computing*, vol. 3, no. 1, pp. 44–63, 2015.

[5] Z. Ning, L. Liu, F. Xia, B. Jedari, I. Lee, and W. Zhang, "CAIS: a copy adjustable incentive scheme in community-based socially aware networking," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3406–3419, 2017.

[6] X. Hu, T. H. S. Chu, V. C. M. Leung, E. C.-H. Ngai, P. Kruchten, and H. C. B. Chan, "A survey on mobile social networks: applications, platforms, system architectures, and future research directions," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1557–1581, 2015.

[7] Q. Liu, X. Hu, E. C. Ngai et al., "A security patch addressing bandwidth request vulnerabilities in the IEEE 802.16 standard," *IEEE Network*, vol. 30, no. 5, pp. 26–34, 2016.

[8] Z. Ning, Q. Song, L. Guo, Z. Chen, and A. Jamalipour, "Integration of scheduling and network coding in multi-rate wireless mesh networks: optimization models and algorithms," *Ad Hoc Networks*, vol. 36, pp. 386–397, 2016.

[9] X. Hu, T. H. S. Chu, H. C. B. Chan, and V. C. M. Leung, "Vita: a crowdsensing-oriented mobile cyber-physical system," *IEEE Transactions on Emerging Topics in Computing*, vol. 1, no. 1, pp. 148–165, 2013.

[10] C. Li, U. Nguyen, and H. Nguyen, "Efficient authentication for fast handover in wireless mesh networks," *Computers and Security*, vol. 47, pp. 124–142, 2013.

[11] Y. Yu, Z. Ning, and L. Guo, "A secure routing scheme based on social network analysis in wireless mesh networks," *Science China Information Sciences*, vol. 59, no. 12, article 122310, 2016.

[12] D. S. Wong, "Security analysis of two anonymous authentication protocols for distributed wireless networks," in *Proceedings of the 3rd IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom '05)*, pp. 532–536, Kauai Island, Hawaii, USA, 2005.

[13] M. Ayyash, H. Elgala, A. Khreishah et al., "Coexistence of WiFi and LiFi toward 5G: concepts, opportunities, and challenges," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 64–71, 2016.

[14] Y. Li, F. Yao, T. Lan, and G. Venkataramani, "SARRE: semantics-aware rule recommendation and enforcement for event paths on android," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 12, pp. 2748–2762, 2016.

[15] K. Ren, S. Yu, W. Lou, and Y. Zhang, "PEACE: a novel privacy-enhanced yet accountable security framework for metropolitan wireless mesh networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 2, pp. 203–215, 2010.

[16] V. S. Miller, "Use of elliptic curves in cryptography," in *Proceedings of the Advances in Cryptology (CRYPTO '85)*, Lecture Notes in Computer Science, pp. 417–426.

[17] D. Boneh and M. Franklin, "Identity-based encryption from the Weil pairing," *SIAM Journal on Computing*, vol. 32, no. 3, pp. 586–615, 2003.

[18] V. Kumar and R. Kumar, "Prevention of blackhole attack using certificateless signature (CLS) scheme in MANET," in *Security*

Solutions for Hyperconnectivity and the Internet of Things, vol. 130, IGI Global, 2016.

[19] Y. Yu, C. Xu, and X. Huang, "An efficient anonymous proxy signature scheme with provable security," *Computer Standards and Interfaces*, vol. 31, no. 2, pp. 348–353, 2009.

[20] P. F. Syverson and P. C. van Oorschot, "On unifying some cryptographic protocol logics," in *Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy*, pp. 14–28, May 1994.

[21] W. Zhao, A. Zhang, J. Li et al., "Analysis and design of an authentication protocol for space information network," in *Proceedings of the IEEE Military Communications Conference (MILCOM '16)*, pp. 43–48, Baltimore, MD, USA, 2016.

[22] S. D. Galbraith, K. Harrison, and D. Soldera, "Implementing the tate pairing," in *Proceedings of the International Algorithmic Number Theory Symposium*, pp. 324–337, Sydney, Australia, 2002.

Hindawi
Mobile Information Systems
Volume 2017, Article ID 2409830, 10 pages
https://doi.org/10.1155/2017/2409830

Research Article

Sliding Window Based Feature Extraction and Traffic Clustering for Green Mobile Cyberphysical Systems

Jiao Zhang,[1,2] Li Zhou,[1,3] Angran Xiao,[4] Sai Zeng,[5] Haitao Zhao,[1] and Jibo Wei[1]

[1] *College of Electronic Science and Engineering, National University of Defense Technology, Changsha, China*
[2] *Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China*
[3] *Science and Technology on Information Transmission and Dissemination in Communication Networks Laboratory, Shijiazhuang, China*
[4] *Department of Mechanical Engineering Technology, New York City College of Technology, City University of New York, Brooklyn, NY 11201, USA*
[5] *IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, USA*

Correspondence should be addressed to Li Zhou; zhouli2035@nudt.edu.cn

Received 16 February 2017; Accepted 4 May 2017; Published 30 May 2017

Academic Editor: Jun Cheng

Both the densification of small base stations and the diversity of user activities bring huge challenges for today's heterogeneous networks, either heavy burdens on base stations or serious energy waste. In order to ensure coverage of the network while reducing the total energy consumption, we adopt a green mobile cyberphysical system (MCPS) to handle this problem. In this paper, we propose a feature extraction method using sliding window to extract the distribution feature of mobile user equipment (UE), and a case study is presented to demonstrate that the method is efficacious in reserving the clustering distribution feature. Furthermore, we present traffic clustering analysis to categorize collected traffic distribution samples into a limited set of traffic patterns, where the patterns and corresponding optimized control strategies are used to similar traffic distributions for the rapid control of base station state. Experimental results show that the sliding window is more superior in enabling higher UE coverage over the grid method. Besides, the optimized control strategy obtained from the traffic pattern is capable of achieving a high coverage that can well serve over 98% of all mobile UE for similar traffic distributions.

1. Introduction

The rising demands for network resources and quality of service are forcing operators of wireless cellular networks to continuously add capacities to their networks. One of the means to do so is densification: deploying heterogeneous networks with a multitude of smaller base stations, such as picobase stations and femtobase stations [1–4]. Such a heterogeneous network will be significantly more complex than today's system and hence will require more effective state control strategies to achieve energy efficient coverage. The control strategy decides the active or sleep state of each base station in the network. If all base stations were active when the traffic demand is low, the unnecessary energy consumption can be significant [5, 6]. On the other hand, excessively frequent switching on/off the base stations is not practical, considering the control precision, protocol limitation, and lifespan of the stations. Therefore, the control strategy of a heterogeneous network must be optimized to reduce energy consumption, while maintaining its QoS including coverage, response time, and spectral efficiency.

To determine an optimized state control strategy for a network, it is necessary to capture its traffic distribution, that is, the geographical distribution of mobile UE in the network [7]. Uniform Poisson modeling method was presented based on IMT-advanced evaluation guidelines in [8]. Although very popular, this method is less effective when representing the heterogeneous user activities and geographical characteristics of the network using only Poisson point processes. References [9, 10] modeled traffic distribution using a two-dimensional sub-Poisson point process in order to capture the heterogeneity; that is, a perfect lattice and a

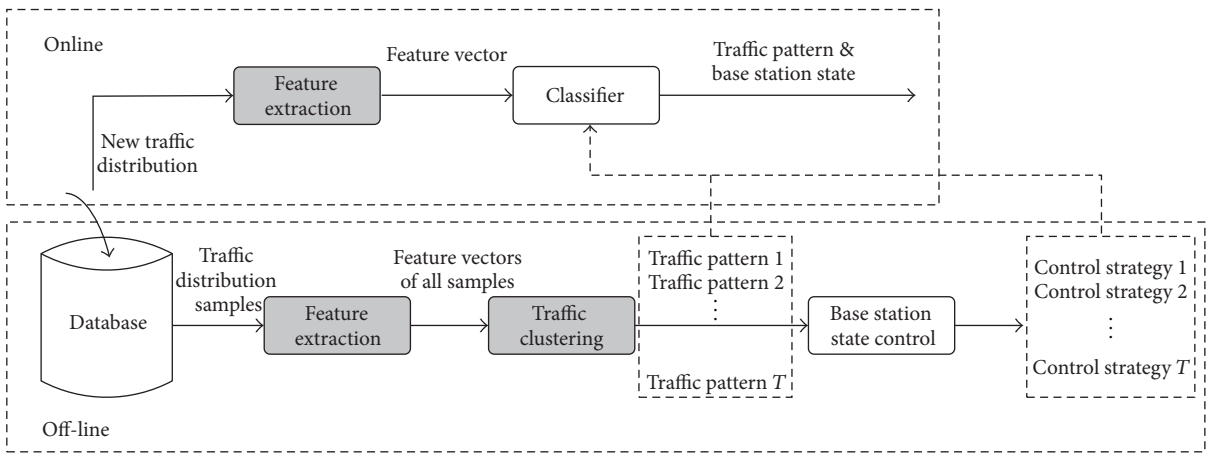


Figure 1: A green mobile cyberphysical system framework.

random perturbation were applied to generate sub-Poisson UE distribution. A thorough review of point processes was presented in [11], in which clustering properties of point processes in space are compared. In [12], Mirahsan et al. introduced a heterogeneous traffic modeling method that allows statistical adjustment. This method is continuously scalable from uniform to heterogeneous point process.

Optimization algorithms were developed to determine control strategies while satisfying a set of conflicting requirements such as coverage and energy efficiency. In [13], a nonlinear integer programming method was used to determine optimized control strategy. In [14], Lorincz et al. used an efficient linear integer programming method. In [15], Jung et al. presented a two-step optimization algorithm that identified the minimal set of base stations needed to maintain coverage and analyzed the bandwidth and power allocation of each base station to minimize energy consumption. Minimax algorithm was adopted in the development of a distributed base stations switch-off method in [16], while the optimization problem was solved using time-consuming exhaustive search method. In [17], Al-Kanj et al. proposed a green radio network planning approach which jointly optimized the number of active base stations and the base station on/off switching strategies. It is worth noting that most of the optimization approaches considered only the traffic distribution at a certain time. Monitoring the dynamic traffic distributions and controlling base stations in a real-time manner are not yet achieved. Mobile cyberphysical systems (MCPSs) were presented for this purpose.

A MCPS is a combination of computation and communication systems. It is capable of sensing, processing, and responding to the dynamic changes of traffic distribution of a network [18–21]. MCPS inherits many essential and important characteristics of traditional cyberphysical system (CPS) [22], such as intelligent network, interaction between human and MCPS, and computation with physical process [23]. It is also capable of integrating wireless sensor network (WSN) and cloud environment. For the tasks requiring more resources than what is available locally, MCPS gives customers rapid access to other WSN and clouds [24]. In a MCPS, mobile UE, such as smart phones and tablets, acts like cyber terminals with storages and processing capabilities. Sensors in the system like GPS and Bluetooth collect physical data such as the geographical positions of UE in the system and their receiving power from the surrounding base stations [25, 26]. The data is transmitted to the computational backbone of MCPS, the clouds. The clouds analyze the data to determine optimal control strategies for base stations, resource allocation, and network scheduling.

In [27], we presented a green MCPS. To maximize the energy efficiency of the network, a heuristic method was developed to determine the control strategy according to the dynamic traffic distributions. The two major components of the green MCPS are introduced in this paper: (1) Modeling: we use a sliding window approach to extracting the overlapped features from traffic distributions; (2) Optimization: a traffic clustering algorithm is presented to classify all the samples of traffic distributions into a set of traffic patterns. The patterns and corresponding optimized control strategies will be used to handle new traffic distributions collected in real time.

The remainder of the paper is organized as follows. The framework of the green MCPS is explained in Section 2. The feature extraction and traffic clustering algorithms are introduced in detail in Section 3. A set of experiments are presented to validate the algorithms in Section 4. Section 5 concludes the paper.

2. Green MCPS Framework

The framework of the green MCPS is shown in Figure 1. It consists of two processes, online and off-line.

The off-line process is a learning process. Its input is a set of traffic distributions sampled at different times when the network is deployed. The outputs of this process are a set of traffic patterns and the corresponding optimized control strategies of the base stations. When the off-line process starts, a set of traffic distribution samples have been collected at various times and saved in the database of the system. Each sample contains the geographical locations of all mobile UE

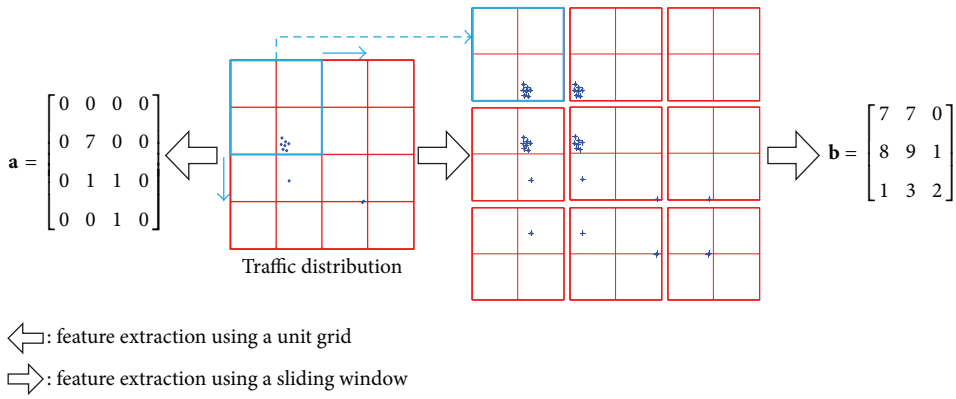


Figure 2: Feature extraction example.

in the 2-dimensional target region at certain time. It must be noted that we must collect enough samples in order to capture the characteristics of the network's traffic distribution during its daily operation. These samples are then transferred to the feature extraction module, in which if the target region is partitioned into equal-sized areas, a feature vector with a dimension equal to the number of areas is used to represent the UE densities in each area. The detailed feature extraction algorithm will be explained in the next section. After that, the feature vectors are transferred to a clustering module in order to group all traffic distribution samples into a certain number of traffic patterns using the clustering algorithm, as introduced in the following section. Each pattern is a traffic distribution that represents a group of samples with similar characteristics of geographical distribution. Each pattern has an optimal control strategy for the target region. The optimized strategy is decided using a heuristic method as introduced in previous papers [27]. The traffic patterns and the corresponding control strategies are stored to the database.

The online process starts when a new traffic distribution is identified. In this case, the system starts the feature extraction module to extract and represent its distribution feature into a feature vector. Then the vector is imported into a classifier, which acts as a decision-making tool that is capable of matching the sample with the traffic patterns according to the Euclidean distance between feature vectors [28]. If the Euclidean distance of the new feature vector is with a defined range from that of a pattern, the sample is matched to the pattern and the corresponding control strategy is used to control the base stations of the new sample. Therefore, the control strategies of the new distribution can be rapidly obtained. If there is no match, the new sample will be stored into the database waiting for the next matching.

3. Feature Extraction and Traffic Clustering Algorithms

3.1. Feature Extraction Using Sliding Window. If we partition a 2-dimensional target region into equal-sized grids, that is, unit grid, and count the amount of mobile UE within each grid, we can represent the characteristics of the traffic distribution into a feature vector. However, elements of the feature vector are always independent of each other. Since the detail of the feature vector is limited by the grid size, the unit grid method is not effective in capturing the clustering characteristics of the traffic distribution.

The high density distribution of mobile UE forms cluster, a common and critical distribution form to affect the coverage and QoS of heterogeneous networks [7]. In order to effectively capture the clustering characteristics and detect changes in traffic distribution, we propose a new method to represent traffic distribution. We use a sliding window with a sliding step smaller than the length of the window instead of using even grids. Using a traffic distribution as shown in Figure 2 as an example, the feature matrix captured using 4×4 red grids is **a**. A square sliding window is defined as being as large as 2×2 grids, shown as the blue cell in Figure 2, and a sliding step size is the same as the grid size. The sliding window slides from left to right until it reaches the right end of the region. Then the window slides down one step and scans from left to right. The process stops when the whole target region is scanned, and the resulting matrix is **b**. It can be seen that a cluster is formed in the target region. This clustering feature is represented by element 7 of matrix **a**, while four elements (7, 7, 8, and 9) of matrix b reserve the paradigm of this cluster similarly. If there are clusters of UE distributions, the sliding window method creates the feature matrix with more elements to depict the clustering distribution, which has the advantage of reserving the integrity of clustering distribution in case of the loss of a cluster in the process of feature extraction. In the sliding window method, both the window size and sliding step size are important design parameters.

Given a sample of traffic distribution in a region with area L^2, it contains N_u mobile UE and the coordinate of mobile UE k is (x_k, y_k), $k \leq N_u$. In order to extract the clustering characteristics, we partition the target region into $N_g \times N_g$ equal-sized grids, and the size of grid is $e = L/N_g$. The size of the sliding window e_w and sliding step size s_w are given as the integer multiple of e for the convenience of data processing, denoted as $e_w = pe$ ($p \in \mathbf{z}^+$, $\mathbf{z}^+$ is a set of positive integers) and $s_w = qe$ ($e_w = pe$, $p \in \mathbf{z}^+$). Besides, the sliding step size should be smaller than the window size; both of them should

Input: the number of mobile UEs N, the coordinates of mobile UEs and the area L^2 of the region.
Output: Traffic distribution feature matrix **a**, feature matrix **b**
(1) Determine the value of N_g, p, q and M.
(2) **for** $k = 1 : N_u$ **do**
(3) calculate the row i and the columns j of mobile UE k in $N_g \times N_g$ grids by $i = \lceil x_k/e + \delta \rceil$ and $j = \lceil y_k/e + \delta \rceil$.
(4) count the number of mobile UEs $a_{i,j}$ in the grid (i, j), and $a_{i,j} = a_{i,j} + 1$.
(5) **end for**
(6) **for** $i = 1 : M$ **do**
(7) **for** $j = 1 : M$ **do**
(8) count the number of mobile UEs $b_{i,j}$ in the window (i, j), and $b_{i,j} = \sum_{k=0}^{p-1} \sum_{l=0}^{p-1} a_{i+k,j+l}$.
(9) **end for**
(10) **end for**

Algorithm 1: The feature extraction using sliding window.

be smaller than the target region. That is, $0 < q < p < N_g$. If a window needs to slide M times to cover the length of the region L, we have

$$L = N_g e = Mqe + pe, \quad \left(0 < q < p < N_g\right). \tag{1}$$

In general, q is a small integer. Given q, M and p are

$$\begin{aligned} M &= \left\lceil \frac{N_g}{q} \right\rceil - n_0 \\ p &= N_g - Mq, \end{aligned} \tag{2}$$

where a small integer $n_0 \in \mathbf{z}^+$ is added to ensure that $q < p$, and it is usually assigned manually. In fact, the optimal value of p and q can only be determined after numerical experiments. With the above variables, the feature extraction method is presented as shown in Algorithm 1.

In Algorithm 1, δ is defined to keep both i and j from becoming 0, and its value is infinitesimal. Matrix $\mathbf{a} \in \mathbf{R}^{N_g \times N_g}$ is the traffic distribution feature matrix obtained using $N_g \times N_g$ grids. Matrix **b** is obtained using the sliding window method.

Subsequently, we convert matrix **b** to feature vector **s** with M^2 elements by array rearrangement; this function is implemented using the *reshape* function of MATLAB. In the off-line process, after conducting feature extraction for N traffic distribution samples, we obtained feature vector set $\mathbf{s} = \{s_1, s_2, \ldots, s_N\}$, which will be transferred to the following clustering module. In order to process data conveniently, we normalize these vectors.

3.2. Traffic Clustering. Traffic clustering module categorizes all traffic distributions samples into a limited set of traffic patterns. Clustering is an unsupervised classification approach, which is capable of analyzing the internal characteristic and mutual relationship of objects without label [29, 30]. The K-means clustering is a well-known algorithm for classification based on the distance measurement and the squared error [31, 32]. Nevertheless, it has significant shortcomings, including predetermined number of clusters, unguaranteed global optimum, and being sensitive to noises. K-medoids algorithm [33] uses the medians of a cluster as its centroid to reduce the influence of noises. Kernel K-means algorithm is presented in [34], which transforms original features set of objects into a higher-dimensional space to make objects more separable. Spectral clustering method performs dimensionality reduction using Laplacian eigenmap [35]. Other advanced clustering algorithms include iterative self-organizing data analysis technique (ISODATA), Gaussian mixture (GM), and density based spatial clustering of applications with noise (DBSCAN). Considering the computation complexity and multidimensional features of the traffic distributions, we decide to combine the K-means and the spectral clustering algorithm presented in [36] to analyze the internal similarity of traffic distribution vectors. The process of traffic clustering primarily comprises two parts, the determination of optimal number of traffic patterns and the classification of all traffic distribution vectors, as presented in Algorithm 2.

In Algorithm 2, the optimal number of traffic patterns K $(1 < K < N)$ is decided using the average silhouette method. The silhouette value measures how well a sample lies within its cluster comparing to other clusters [37]. In order to obtain the silhouette value of each sample, we classify all samples into k clusters at first, where k is a variable quantity. For each sample i, let $a(i)$ be the average dissimilarity of i to all other samples within the same cluster, and the average dissimilarity of sample i can be defined as the average of the distance from i to all other samples within the same cluster. $b(i)$ denotes the lowest average dissimilarity of i to any other sample in other clusters. Then, the silhouette value can be obtained by combining $a(i)$ and $b(i)$.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}. \tag{3}$$

From the above definition, it can be seen that $-1 < s(i) < 1$ for each sample i.

$s(i)$ is close to 1 for $a(i) \ll b(i)$, which indicates that sample i is well clustered because $a(i)$ represents how dissimilar i is to its own cluster. When $s(i)$ is around zero for $a(i) = b(i)$, sample i is on the border of two clusters. The worst situation takes place when $s(i)$ is close to -1 when $a(i) \gg b(i)$,

Input: Feature vectors of all traffic distribution samples $\mathbf{s} = \{s_1, s_2, \ldots, s_N\}$.
Output: Traffic patterns $T_1, T_2, \ldots, T_K$.
(1) Normalize Feature vectors to $\mathbf{s}' = \{s'_1, s'_2, \ldots, s'_N\}$.
(2) Determine the number of traffic patterns K by the average silhouette method.
(3) Construct an affinity matrix $\mathbf{A}$ with Gaussian kernel function, in which $A_{ij} = \exp(-d^2(s'_i, s'_j))/2\delta^2$ holds for $i \neq j$ and $A_{ii} = 0$.
(4) Define the diagonal degree matrix $\mathbf{D}$ ($D_{ii} = \sum_j A_{ij}$). Normalize the affinity $\mathbf{A}$ to $\mathbf{L}$, and $\mathbf{L} = \mathbf{A}^{-1/2}\mathbf{D}\mathbf{A}^{-1/2}$.
(5) Compute the first K eigenvectors of $\mathbf{L}$. Construct a matrix $\mathbf{X} = \{x_1, x_2, \ldots, x_K\} \in \mathbf{R}^{N\times K}$.
(6) Construct a matrix $\mathbf{Y}$ from $\mathbf{X}$ by normalizing the rows of $\mathbf{X}$ to norm 1, and $Y_{ij} = X_{ij}/(\sum_j X_{ij}^2)^{1/2}$.
(7) Treating each row of $\mathbf{Y}$ as a point, cluster them into K traffic patterns by K-means.
(8) Assign the original feature vector s_i of traffic distribution sample to traffic pattern j according to the assigned label j of the row i of the matrix $\mathbf{Y}$.
(9) Compute the features of traffic patterns, and $u_j = 1/|T_j| \sum_{s_i \in T_j} s_i$ $(j = 1, 2, \ldots, K)$.

Algorithm 2: The traffic clustering algorithm.

meaning that the sample is misclassified. Furthermore, the average silhouette s_{ave} is the average of silhouette $s(i)$ of all samples for a traffic pattern; it shows how tightly all samples are grouped in the cluster and, hence, evaluates clustering validity. If there are too many or too less clusters, some narrow silhouette may occur; thus it is used to determine the optimal number of clusters. It is believed that an optimal number of traffic patterns K is the one at which the average silhouette is maximized over a range of values for k [38].

In order to obtain the optimal K traffic patterns, the spectral clustering method [36] is applied. First, an affinity matrix $\mathbf{A}$ is formed, where $d(s'_i, s'_j)$ is the Euclidean distance between s'_i and s'_j, and δ^2 is the scaling parameter that determines the speed of the affinity matrix falling off. Then, the selection of K eigenvectors of $\mathbf{L}$ leads to the feature dimensionality reduction. Finally, K-means algorithm is adopted to assort all traffic distribution samples into various traffic patterns.

After traffic clustering, the optimal control strategy for each pattern is decided using the method presented in our previous work [27]. Briefly, considering the constraint conditions of the UE association, the received signal to interference and noise ratio (SINR) of UE, and capacity of base station, the optimal control strategy maximizes the energy efficiency of a traffic pattern. Assume that N_u UE is scattered in the region under traffic pattern t and N_b base stations are deployed as well. The UE associate is denoted as $s_{i,k}^t$, in which $s_{i,k}^t$ represents that UE k is ($s_{i,k}^t = 1$) or is not ($s_{i,k}^t = 0$) associated with base station i under traffic pattern t. The SINR of UE k from base station i is denoted as $\gamma_{i,k}^t$. $M_i^{\max}$ represents the maximum amount of servable UE for base stations i. For each traffic pattern, the control strategy is represented using vector $\mathbf{a}^t = [a_i^t]_{1\times N_b}$, where a_i^t denotes that the state of base station i is active ($a_i^t = 1$) or sleep ($a_i^t = 0$). Hence, the optimal control model can be formulated as follows:

$$\max \eta_{EE} = \frac{\sum_{i=1}^{N_b} a_i^t \sum_{k=1}^{N_u} s_{i,k}^t R_{i,k}^t}{\sum_{i=1}^{N_b} \left[a_i^t \left(k_i P_i^{\max} \sum_{k=1}^{N_u} s_{i,k}^t / M_i^{\max} + P_i^A\right) + \left(1 - a_i^t\right) P_i^S\right]}, \quad (4)$$

where $R_{i,k}^t$ is the transmission rate calculated using $R_{i,k}^t = \log_2(1 + \gamma_{i,k}^t)$. k_i denotes the power amplifier inefficiency factor. $P_i^{\max}$ represents the total transmit power. P_i^A is the circuit power consumption of active base station i and P_i^S is the circuit power consumption when it is in sleep mode. A heuristic algorithm is adopted to solve the problem, and the states of all base stations, the number of active base stations, and the association of UE with base stations can be obtained eventually [27].

4. Simulation Results and Discussions

The presented method is demonstrated using a 1600 m × 1600 m target region. In order to verify the efficacy of the presented method in different traffic distributions, we create 1000 different traffic distributions including three distribution models:

(i) A *rand* function in MATLAB is used to generate a complete random distribution (Poisson).

(ii) A perfect lattice and a random perturbation [10] are used to generate a uniform distribution (sub-Possion).

(iii) Thomas process [11] is used to generate a clustered distribution (sub-Possion).

Each of these traffic distribution samples consists of a random amount of UE between 250 and 650. In order to differentiate the samples, the difference of the UE numbers between any two samples is at least 50. 113 base stations are in cochannel deployment in a two-tier heterogeneous cellular network, including one macro base station and 112 small base stations. In addition, other major network parameters, such as the bandwidth and transmission power of base stations, bit error rate, and outage probability, are set referring to [27].

According to the optimal control strategy, the amount of mobile UE served by the active base stations in each sample is counted. Hence, we select the rate of UE coverage as the

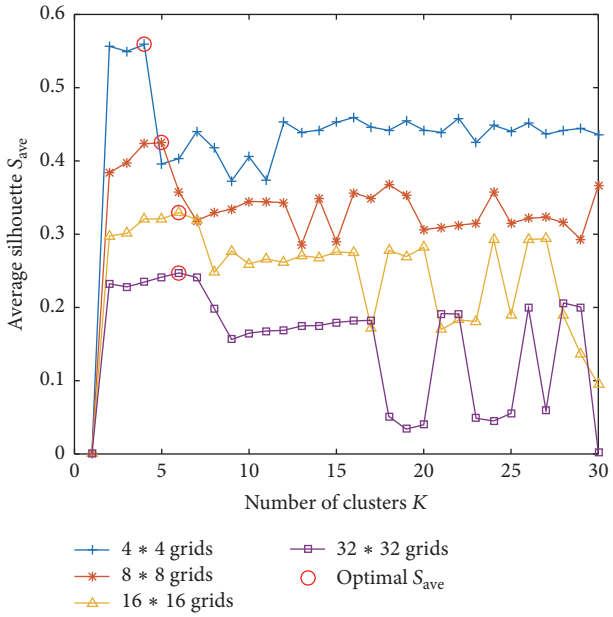


Figure 3: Average silhouettes value for different number of traffic patterns.

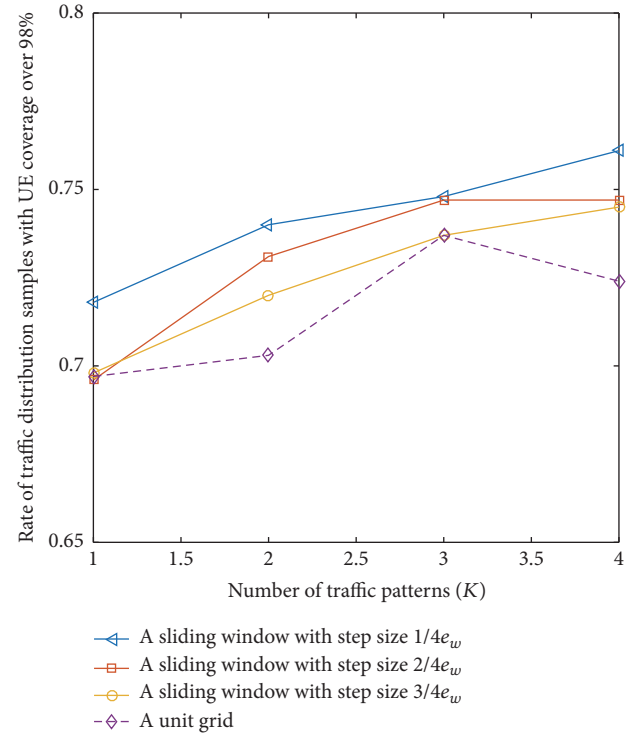


Figure 4: Feature extraction using sliding window with different step sizes and a unit grid.

performance metric of green MCPS.

$$\eta = \frac{n_s}{N_u}, \tag{5}$$

where n_s is the amount of mobile UE served by the active base stations and N_u is the total mobile UE in the target region. According to [27], the UE coverage for a traffic pattern should be more than 98% in order to satisfy the required QoS. In this paper, we believe 98% or higher coverage is a "good coverage."

We start by estimating an optimal number of traffic patterns. The sliding window size e_w is predefined as $2e$ (e is the size of a grid) and the sliding step size is e. The average silhouettes s_{ave} of 1000 traffic distribution samples for different number of traffic patterns and different partition grids are illustrated in Figure 3.

It is observed that the average silhouettes value achieves their peak over a range of possible value for k from 1 to 30. Thus, the optimal number of traffic patterns can be determined as the corresponding value of k at the peak. That is, the optimal number of traffic patterns is 4 when the partitioned grids are 4×4, 5 when the grids are 8×8, and 6 when the grids are 16×16 and 32×32. It is noted that the number of traffic patterns at the optimal average silhouettes has slightly been raised but not sharply changed with the increasing of the number of traffic patterns. Hence, an appropriate number of traffic patterns clustered can be determined to be 6. The performance analysis is carried out for different number of traffic patterns next.

4.1. Feature Extraction Using Sliding Window with Different Sliding Step Sizes. In this experiment, we count the rate of traffic distribution samples with UE coverage over 98% based on the feature vectors extracted by sliding window with different step sizes and a unit grid. The number of grids partitioned is set to be 16×16, and the size of the sliding window e_w is four times the size of a grid ($e_w = 4e$). Thus, the value of q can be 1, 2, or 3.

We can observe from Figure 4 that the smaller the sliding step size is, the more the traffic distribution samples achieve coverage over 98%. The percentage of samples with good coverage (at least 98% UE coverage) at step size $1/4e_w$ is always higher than others. That is, sliding window with smaller sliding step size, we can not only extract more features but also reserve the clustering feature more faithfully. This leads to a good coverage. In addition, it shows that, by using the proposed sliding window method, we can always acquire better coverage than that of using the unit grid, shown with the dotted line on the figure.

4.2. UE Coverage under Different Sliding Window Sizes and Traffic Patterns. In order to discuss the effects of sliding window size and the number of traffic patterns on the coverage, we use 32×32 grids to partition the region, and grid edge is e. The sliding step size s_w is always half of the window size e_w. Three different intervals of UE coverage rate are counted in Figure 5, namely, $[0, 0.9]$, $[0.9, 0.98]$, and $[0.98, 1]$, respectively.

It can be seen that the smaller the size of the sliding window is, the higher the percentage of samples achieves good coverage. This is because that more heterogeneous distribution features are preserved by using a smaller sliding window, which contributes to better coverage. However, smaller window size means larger feature vector for a sample, and this unavoidably reduces the operational speed. It is

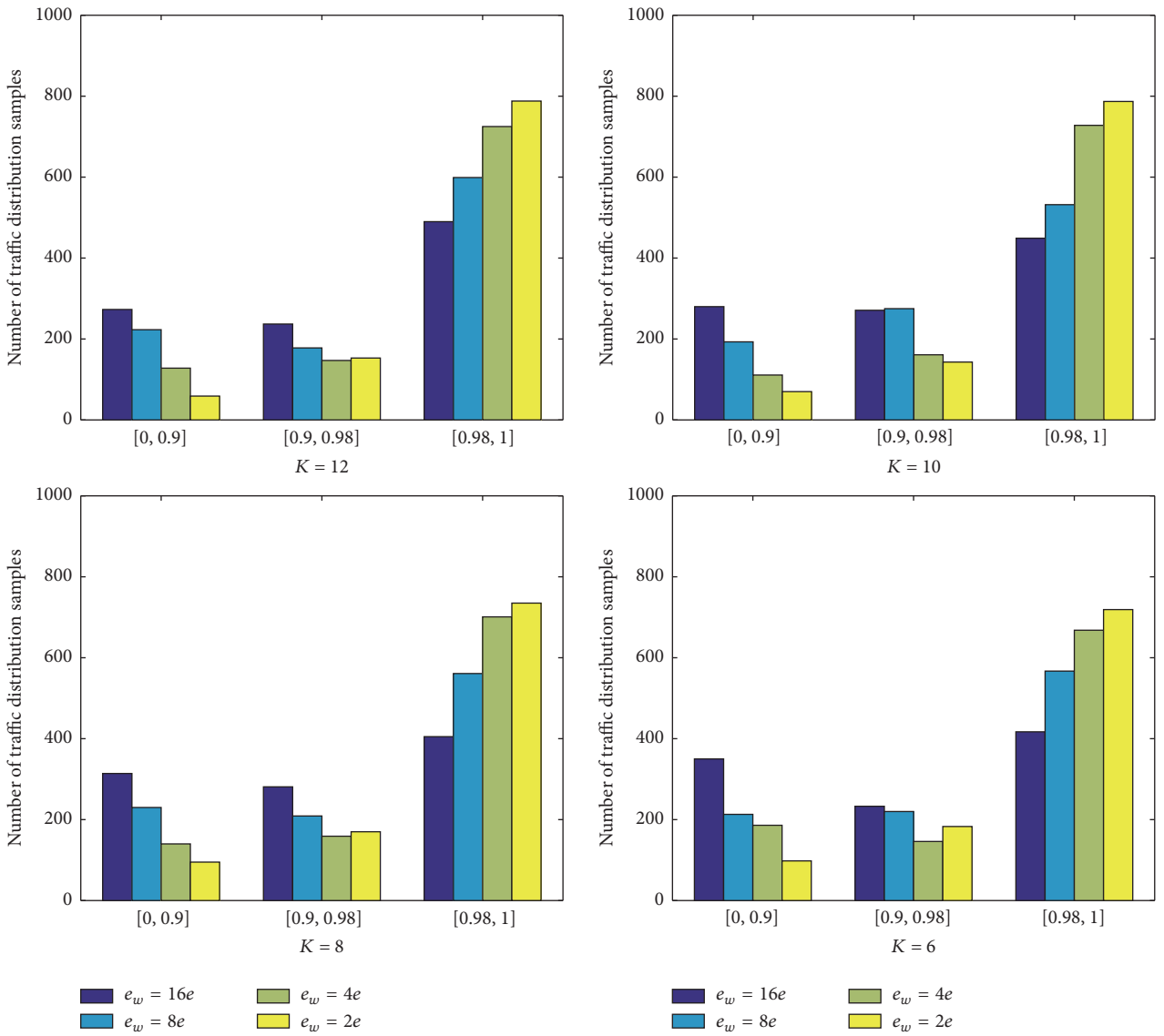


Figure 5: Coverage for different number of patterns.

necessary to compromise between the coverage and operational speed when we decide the window size in practice. We can choose different window size according to different requirements for the rates of coverage. At the same time, we can observe that, by increasing the number of traffic patterns, the number of samples with good coverage rises slightly but is not significant. Therefore, we conclude that the number of traffic patterns is less important than other factors such as sliding step size and window size. In practice, we can decide the pattern number by the average silhouette method.

4.3. Number of Active Base Stations for Different Traffic Clusters. After adopting the optimal control strategy of the traffic pattern to its samples, we can observe the number of active base stations in the traffic distribution samples. Here, in the process of feature extraction, 32×32 grids are used, the size of the sliding window e_w is $2e$, and the sliding step size is e.

As shown in Figure 6, the difference between the maximal and minimal number of active base stations for all traffic distribution samples within the same traffic pattern is shown. That is, most of the number of active base stations for a traffic pattern is slightly more than that of the maximal number of active base stations for samples in this pattern. This means that applying the control strategy of a traffic pattern to the samples belonging to this pattern will ensure good coverage, and this can achieve significant energy saving as well.

4.4. SINR Distribution under a Traffic Pattern. Finally, a SINR distribution example under a traffic pattern is depicted in Figure 7, in order to observe the coverage of active base stations.

In this traffic pattern, UE is represented as dots, and base stations are uniformly distributed within the target region. As shown in Figure 7, the triangles represent small base

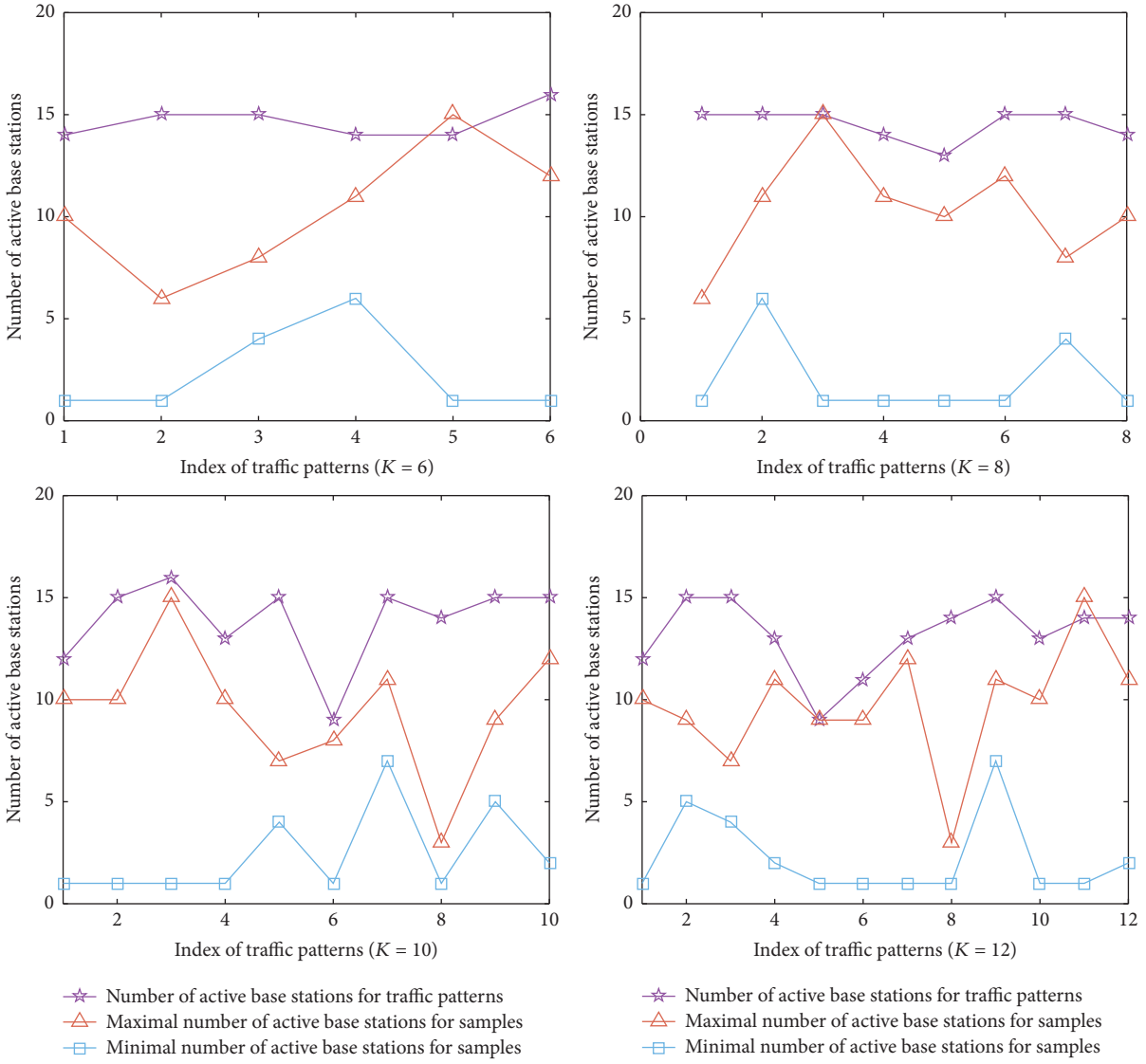
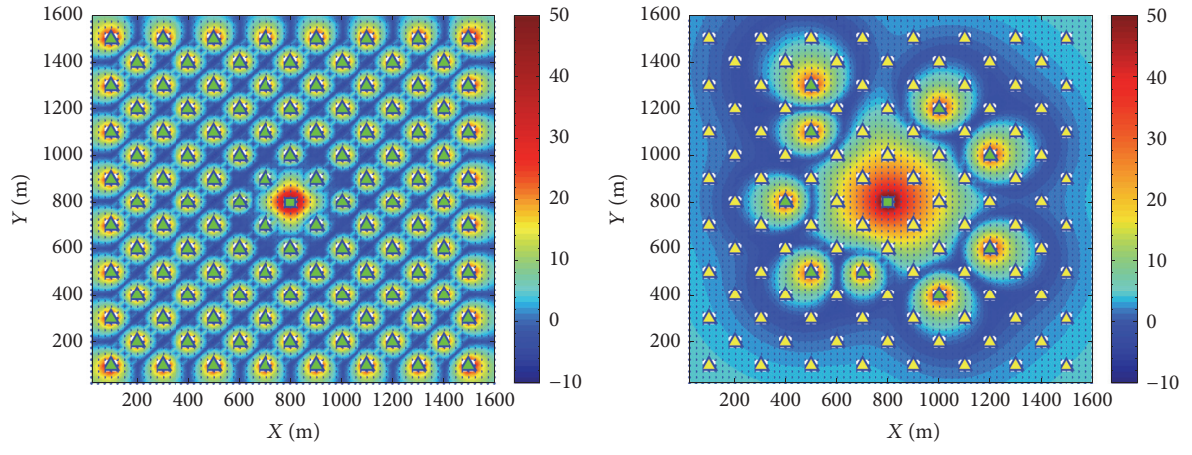


Figure 6: Number of active base stations for different traffic clusters.



(a) Initial state

(b) Final state

Figure 7: SINR distributions.

stations, and the squares represent macro base station. The state of base stations is represented by colors, where yellow means sleep and green means active. The values of SINR are in the range of −10 dB to 50 dB, which are represented by different gradient colors. Figure 7(a) shows the initial SINR distribution of UE when all base stations are active. We can observe that the coverage of active base stations is limited due to the ubiquitous interference between base stations. After adopting the optimal control strategy, only ten base stations are activated, and the coverage with higher SINR of UE around the active base stations increases in Figure 7(b). This indicates that the coverage of active base station is enlarged under the condition of over 98% UE coverage.

5. Conclusion

In this paper, we present a feature extraction method using sliding window for traffic distributions in a green mobile cyberphysical system. The method has the advantages of reserving more clustering distribution features over using the grid method. In order to implement rapid base station state control and extend the lifespan of a heterogeneous network, we apply clustering analysis for all traffic distributions to obtain a limited set of traffic patterns. Numerical results demonstrate that the proposed method helps obtain better UE coverage comparing with using the grid method. It is worth noting that both smaller sliding step size and smaller sliding window size can lead to good UE coverage but slow the operational speed of the network.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported in part by the National Natural Science Foundation of China (Grant no. 61601482) and Guangdong Technology Project (2016B010125003 and 2016B010108010) and sponsored by the Foundation of Science and Technology on Information Transmission and Dissemination in Comm. Networks Lab, National Key Laboratory of Anti-Jamming Communication Technology, and State Joint Engineering Laboratory for Robotics and Intelligent Manufacturing funded by National Development and Reform Commission (no. 2015581).

References

[1] J. F. Monserrat, I. Alepuz, J. Cabrejas et al., "Towards user-centric operation in 5G networks," *Eurasip Journal on Wireless Communications and Networking*, vol. 2016, no. 1, article no. 6, pp. 1–7, 2016.

[2] J. F. Monserrat, G. Mange, V. Braun, H. Tullberg, G. Zimmermann, and Ö. Bulakci, "METIS research advances towards the 5G mobile and wireless system definition," *Eurasip Journal on Wireless Communications and Networking*, vol. 2015, no. 1, pp. 1–16, 2015.

[3] L. Zhou, Z. Sheng, L. Wei et al., "Green cell planning and deployment for small cell networks in smart cities," *Ad Hoc Networks*, vol. 43, pp. 30–42, 2016.

[4] L. Zhou, C. Zhu, R. Ruby et al., "QoS-aware energy-efficient resource allocation in OFDM-based heterogenous cellular networks," *International Journal of Communication Systems*, vol. 30, no. 2, p. e2931, 2017.

[5] M. Hoshino, Y. Yuda, T. Takata, and A. Nishio, "Performance evaluation of jt-comp under non full buffer traffic condition on heterogeneous network with dense small cells," vol. 112, pp. 29–34, 2012.

[6] L. Zhou, X. Hu, E. C.-H. Ngai et al., "A dynamic graph-based scheduling and interference coordination approach in heterogeneous cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3735–3748, 2016.

[7] M. Mirahsan, Z. Wang, R. Schoenen, H. Yanikomeroglu, and M. St-Hilaire, "Unified and non-parameterized statistical modeling of temporal and spatial traffic heterogeneity in wireless cellular networks," in *Proceedings of 2014 IEEE International Conference on Communications Workshops, ICC 2014*, pp. 55–60, aus, June 2014.

[8] ITU-R, "Guidelines for evaluation of radio interface technologies for imt-advanced," M.2135-1, 2009.

[9] J. Rataj, I. Saxl, and K. Pelikán, "Convergence of randomly oscillating point patterns to the Poisson point process," *Applications of Mathematics*, vol. 38, no. 3, pp. 221–235, 1993.

[10] V. Lucarini, "From symmetry breaking to Poisson point process in 2D Voronoi tessellations: the generic nature of hexagons," *Journal of Statistical Physics*, vol. 130, no. 6, pp. 1047–1062, 2008.

[11] B. Błaszczyszyn and D. Yogeshwaran, "Clustering comparison of point processes, with applications to random geometric models," *Lecture Notes in Mathematics*, vol. 2120, pp. 31–71, 2015.

[12] M. Mirahsan, R. Schoenen, and H. Yanikomeroglu, "HetHetNets: Heterogeneous Traffic Distribution in Heterogeneous Wireless Cellular Networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 10, pp. 2252–2265, 2015.

[13] K. Son, S. Nagaraj, M. Sarkar, and S. Dey, "QoS-aware dynamic cell reconfiguration for energy conservation in cellular networks," in *Proceedings of 2013 IEEE Wireless Communications and Networking Conference, WCNC 2013*, pp. 2022–2027, chn, April 2013.

[14] J. Lorincz, A. Capone, and D. Begušić, "Optimized network management for energy savings of wireless access networks," *Computer Networks*, vol. 55, no. 3, pp. 514–540, 2011.

[15] H.-S. Jung, H.-T. Roh, and J.-W. Lee, "Energy and traffic aware dynamic topology management for wireless cellular networks," in *Proceedings of 2012 IEEE International Conference on Communication Systems, ICCS 2012*, pp. 205–209, sgp, November 2012.

[16] X. Su, E. Sun, M. Li, F. R. Yu, and Y. Zhang, "An Energy-Efficient User Location-Aware Switch-Off Method for LTE-A Cellular Networks," *Wireless Personal Communications*, vol. 84, no. 3, pp. 1817–1833, 2015.

[17] L. Al-Kanj, W. El-Beaino, A. M. El-Hajj, and Z. Dawy, "Optimized joint cell planning and BS on/off switching for LTE networks," *Wireless Communications and Mobile Computing*, vol. 16, no. 12, pp. 1537–1555, 2015.

[18] X. Hu, T. H. S. Chu, H. C. B. Chan, and V. C. M. Leung, "Vita: a crowdsensing-oriented mobile cyber-physical system," *IEEE Transactions on Emerging Topics in Computing*, vol. 1, no. 1, pp. 148–165, 2013.

[19] X. Hu, T. H. S. Chu, V. C. M. Leung, and C. H. Ngai, "A survey on mobile social networks: Applications, platforms, system architectures, and future research directions," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1557–1581, 2014.

[20] C.-L. Fok, M. Hanna, S. Gee et al., "A platform for evaluating autonomous intersection management policies," in *Proceedings of IEEE/ACM Third International Conference on Cyber-Physical Systems*, pp. 87–96, Beijing, China, April 2012.

[21] S. A. Haque, S. M. Aziz, and M. Rahman, "Review of cyber-physical system in healthcare," *International Journal of Distributed Sensor Networks*, vol. 2014, Article ID 217415, 2014.

[22] J. Sztipanovits, "Composition of cyber-physical systems," in *Proceedings of 14th Annual IEEE International Conference and Workshops on the Engineering of Computer-Based Systems, ECBS 2007*, pp. 3-4, usa, March 2007.

[23] E. A. Lee, "Cyber physical systems: design challenges," in *Proceedings of the 11th IEEE International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing (ISORC '08)*, pp. 363–369, May 2008.

[24] B. Perumal, P. Rajasekaran, and M. H. M. Ramalingam, "Wsn integrated cloud for automated telemedicine (atm) based e-healthcare applications," in *Proceedings of International Proceedings of Chemical Biological & Environmenta*, 2012.

[25] X. Hu, J. Zhao, B.-C. Seet, V. C. M. Leung, T. H. S. Chu, and H. Chan, "S-aframe: agent-based multilayer framework with context-aware semantic service for vehicular social networks," *IEEE Transactions on Emerging Topics in Computing*, vol. 3, no. 1, pp. 44–63, 2015.

[26] X. Hu, X. Li, E. C.-H. Ngai, V. C. M. Leung, and P. Kruchten, "Multidimensional context-aware social network architecture for mobile crowdsensing," *IEEE Communications Magazine*, vol. 52, no. 6, pp. 78–87, 2014.

[27] L. Zhou, J. Zhang, B. Seet et al., "Software Defined Small Cell Networking under Dynamic Traffic Patterns," in *Proceedings of IEEE Cyber Science and Technology Congress*, Auckland, New Zealand, August 2016.

[28] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999.

[29] B. S. Everitt, G. Dunn, B. S. Everitt, and G. Dunn, "Cluster analysis," 1em plus 0.5em minus 0.4em Wiley, 2011.

[30] A. K. Jain and R. C. Dubes, "Algorithms for clustering data," *Technometrics*, vol. 32, no. 2, pp. 227–229, 1988.

[31] E. W. Forgy, "Cluster analysis of multivariate data efficiency vs. interpretability of classification," *Biometrics*, vol. 21, no. 3, pp. 41–52, 1965.

[32] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297, University of California Press, Berkeley, California, 1967.

[33] V. Estivill-Castro and J. Yang, "A fast and robust general purpose clustering algorithm," in *Proceedings of the Pacific Rim International Conference on Artificial Intelligence*, pp. 208–218, Springer-Verlag, London, UK, 1999.

[34] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[35] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.

[36] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering Analysis and an algorithm," in *Proceedings of the Advances in Neural Information Processing Systems*, vol. 14, pp. 849–856, San Francisco, CA, USA, 2002.

[37] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, no. 20, pp. 53–65, 1987.

[38] L. Kaufman and P. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley & Sons, New York, NY, USA, 1990.

Hindawi
Mobile Information Systems
Volume 2017, Article ID 5731678, 10 pages
https://doi.org/10.1155/2017/5731678

Research Article

A Formal Approach to Verify Parameterized Protocols in Mobile Cyber-Physical Systems

Long Zhang,[1] Wenyan Hu,[2] Wanxia Qu,[1] Yang Guo,[1] and Sikun Li[1]

[1]*College of Computer, National University of Defense Technology, Changsha, China*
[2]*Carnegie Mellon University, Pittsburgh, PA, USA*

Correspondence should be addressed to Yang Guo; guoyang@nudt.edu.cn

Received 16 February 2017; Accepted 12 April 2017; Published 10 May 2017

Academic Editor: Jun Cheng

Mobile cyber-physical systems (CPSs) are very hard to verify, because of asynchronous communication and the arbitrary number of components. Verification via model checking typically becomes impracticable due to the state space explosion caused by the system parameters and concurrency. In this paper, we propose a formal approach to verify the safety properties of parameterized protocols in mobile CPS. By using counter abstraction, the protocol is modeled as a Petri net. Then, a novel algorithm, which uses IC3 (the state-of-the-art model checking algorithm) as the back-end engine, is presented to verify the Petri net model. The experimental results show that our new approach can greatly scale the verification capabilities compared favorably against several recently published approaches. In addition to solving the instances fast, our method is significant for its lower memory consumption.

1. Introduction

A cyber-physical system (CPS) [1] is an integration of computation and physical components. The improvement of contemporary mobile devices, such as smartphones and wearable electronics, enables the formation of mobile CPSs [2, 3]. A mobile CPS could be considered as a subcategory of CPSs with inherent mobile features [4, 5]. Different from traditional CPSs, mobile CPSs could be built on mobile devices that travel with their owners. For example, mobile social networking [6, 7] helps people communicate with each other on their daily commute to and from work, traveling along the same rotes at about the same time. Due to distributed interactions of cyber word, physical word, and human behaviors, the mobile CPS becomes more complex. Asynchronous communication and the arbitrary number of components make the mobile CPS appear similar to a parameterized system. It is difficult to guarantee the parameterized system's correctness for any natural number [8, 9].

Due to the tight market windows and safety-critical nature of their applications, it has become an urgent need to design error-free mobile CPSs and thus a significant amount of time is spent on ensuring the correctness of mobile CPS designs. The verification of the CPS designs becomes an important issue [10]. Formal methods, replacing the traditional testing methods for large mobile CPSs, have been successfully used for verifying software, hardware, and physical systems in the past decades [11]. Model checking [12, 13] is an automatic formal approach to verify if the specification satisfies the properties and has been used in finite and infinite state system verification successfully.

Abstraction [14, 15] is a good way to reduce the state space. By abstraction, each agent of the mobile CPS can be modeled as a finite state automaton in which local transitions model one of the following: an internal action, a broadcast, or a reception of a message. A mobile CPS is defined as the composition of a finite but arbitrary number of copies of the automaton running in parallel. A mobile CPS which combined with an arbitrary number of components is a parameterized system, which is a wide class infinite system, including cache coherence protocols and mutual exclusion protocols.

Parameterized systems arise naturally in the modeling of mutual exclusion algorithms, distributed protocols, or cache coherence protocols. Parameterized verification [8] is aimed at verifying families of transition systems for all values of the parameter. Counter abstraction [14] is natural to model

parameterized systems into Petri nets and their extensions. Petri net is a powerful mathematical tool and has been used widely for modeling and verifying CPSs [10, 16, 17].

In this paper, we propose a formal approach to verify the safety properties of parameterized protocols in mobile CPSs. By using counter abstraction, the protocols of mobile CPSs are described as Petri nets, and then the state-of-the-art model checker is used to check the safety properties.

The significant contributions of this paper are as follows:

(i) We propose a new method based on the SAT-based model checking algorithm to verify the parameterized protocols of mobile CPSs. By using counter abstraction, we describe the parameterized protocol as a Petri net and then translate it into a finite state machine (FSM), so that IC3 [18, 19], the state-of-the-art finite state model checking algorithm, can be used as the back-end engine.

(ii) A smart encoding technique is introduced to make the verification efficient. A bounded Petri net is transformed into a FSM and described as a general format for most model checkers.

(iii) To improve the scalability of parameterized protocols verification, an incremental algorithm is proposed to make IC3 perform more efficiently.

The rest of this paper is organized as follows. In Section 2, we review the related work. Section 3 presents necessary preliminaries used in this paper. In Section 4, we propose our new method based on the model checking algorithm and give more details of the implementation and optimization. Section 5 shows the experimental evaluation on parameterized protocols. Section 6 concludes this paper and discusses future works.

2. Related Works

As a successful application in traditional hardware and software verification, model checking has been frequently used in CPS verification, especially for safety-critical CPSs. Akella and McMillin [20] encoded the physical system into an event-based discretized system and modeled the associated CPS by Security Process Algebra. The model checker, CoPS, was used to check the confidentiality properties. A statistical model checker has been recently utilized to analyze some aspects of CPSs [21]. However, this method also suffers from the classical model checking problems, such as the state space explosion and the lack of ability to reason about mathematical relations. Bae et al. [22] combined model checking and Multirate PALS (physically asynchronous, logically synchronous) methodology for the first time to verify an airplane turning control system. More cases should be studied for the verification of distributed cyber-physical systems using Multirate PALS.

Petri nets are well-known tools for modeling and verification of distributed systems and CPSs. Xu and Deng [16] proposed a Petri nets-based method for architectural modeling of mobile agent systems. Chen et al. [17] investigated the use of Petri nets for modeling coordinated cyber-physical attacks on the smart grid. A novel hierarchical method was proposed to construct large Petri nets from a number of smaller Petri nets that can be created separately by different domain experts. Vita [23, 24], which is a novel mobile cyber-physical system for crowdsensing applications, introduced Petri nets to design a high level service state synchronization mechanism to address the possible unavailable situations of mobile devices in mobile CPSs. In order to define the functionality of traveler information systems (TIS) and integrate new functions and technologies based on cloud computing and mobile communications, Nemtanu et al. [25] presented a Petri nets-based model of this system. Zhang et al. [26] proposed a mechanism to model fault tolerated mobile agents by using colored Petri nets.

There is a large amount of related works on automating the parameterized verification problem [27–29]. The theorem prover PVS, for example, has been successfully applied to verify Small Aircraft Transportation System (SATS) [30]. By using the Model Checker Modulo Theories, Johnson and Mitra presented a model checking method for SATS [31]. Guo et al. [32, 33] proposed a new method to reduce the state space of parameterized systems by two-dimensional abstraction (TDA). Asynchronous composition was the key part of TDA but suffered from higher memory consumption.

3. Preliminaries

Petri nets have been popular models for various types of asynchronous or concurrent processes. A Petri net is a directed graph consisting of places (drawn as circles), transitions (typically boxes), and directed arcs. Input places point to a transition, and a transition points to output places. A number of tokens move around the net from place to place, and the distribution of tokens among the places (called the *marking*) represents the dynamic state of the entire modeled system. The formal definition of Petri nets is as follows.

Definition 1 (Petri net). A Petri net (PN) is a triple $\mathcal{PN} = (P, T, F)$, where P is a finite set of places, T is a finite set of transitions disjoint from P, and $F : (P \times T) \cup (T \times P) \to \mathbb{N}$ is flow relations for the set of arcs.

The configuration of $\mathcal{PN}$ is *markings*, which can be seen as the multisets of places. The semantics of $\mathcal{PN}$ is given by *markings*. A *marking* is a function $m : P \to \mathbb{N}$, which describes the number of tokens $m(p)$ in place $p \in P$. In the sequel, if places are ordered by $P = \{p_1, p_2, \ldots, p_n\}$, we often identify m and the vector $\langle m(p_1), m(p_2), \ldots, m(p_n)\rangle$.

Example 2. As shown in Figure 1, a simple example contains all components of a Petri net. There are two places and three transitions. Arcs have capacity 1 by default; if other than 1, the capacity is marked on the arc. Places have infinite capacity, and transitions have no capacity and cannot store tokens at all. The current marking is $\langle 1, 0\rangle$. If t_1 is fired, the next marking will be $\langle 2, 0\rangle$. If t_2 is fired, the next marking will be $\langle 0, 1\rangle$. The transition t_3 cannot be fired now.

In particular, it has been shown that certain communication procedures, which are common when programming

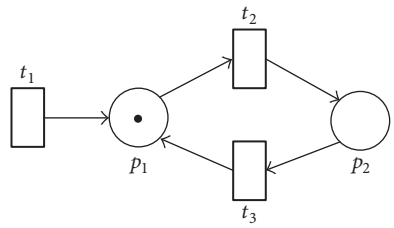


Figure 1: A Petri net with two places and three transitions.

distributed systems, can easily be modeled by plain Petri nets, but other communication procedures such as broadcast are not easily captured by plain Petri nets. For that reason, we address two extensions of the model in this paper, following the definition in [34].

(i) Petri nets with transfer arcs, in which a transition can also consume all tokens present in one place and move them to another

(ii) Petri nets with reset arcs, in which a transition can delete all tokens present in one place

This significance of these extensions in terms of modeling power has been demonstrated, for instance, in the modeling and verification of parameterized protocols in [35].

A common way to model a parameterized system by a Petri net is to apply the idea of the counter abstraction [8]. This principle consists of mapping each single process to a token and representing each state of each type of process by a place. In this case, the presence of a token in a given place $p \in P$ indicates that, in the current global state of the system, there is a process that corresponds to p in its local states. The transition of the Petri net then consumes and produces tokens to move the associated process from one state to another. This formalization has the drawback of abstracting away the actual identities of the processes. Still, some interesting properties, for instance, safety properties, can be verified at this level of abstraction.

In this paper, we focus on the safety property, which is equal to the coverability of well-structured transition systems (WSTSs) when expressing the safety property as the upward-closed set. Petri nets are WSTSs (with respect to $\leq$) [36]. We present the related notions of WSTSs and then define the PN safety problem.

Definition 3 (well-quasi-ordering). A well-quasi-ordering (wqo) is a reflexive and transitive binary relation $\preceq$ over set X, and for every infinite sequence $x_0, x_1, x_2, \ldots$ of elements from X, there exists $i < j$ such that $x_i \preceq x_j$. For $Y \subseteq X$, the upward-closure of Y is the set $\uparrow Y = \{x \mid \exists y \in Y,\ y \preceq x\}$. A set U is said to be $\preceq$-upward-closed (or simply upward-closed if $\preceq$ is clear from the context) if $U = \uparrow U$.

In Figure 1, the safety property is defined whether the tokens number of place p_2 is greater than or equal to 2; then we can use the upward-closed set $p_2 \geq 2$ to express the property.

Definition 4 (well-structured transition systems). A well-structured transition system (WSTS) is a transition system equipped with a wqo on its states that satisfies the monotonicity property. A WSTS is a triple $(S, \rightarrow, \preceq)$ such that

(1) S is the (possibly infinite) state space,

(2) $\rightarrow \subseteq S \times S$ is transition relation,

(3) $\preceq$ is a wqo over S,

(4) for all $x, x', y \in S$, if $x \rightarrow x'$ and $x \preceq y$, there exists y' such that $y \rightarrow y'$ and $x' \preceq y'$.

The covering relation $\leq$ between Petri net markings is a wqo. A $\mathcal{PN} = (P, T, F)$ and the initial marking m_0 give rise to a WSTS $(S, I, \rightarrow, \leq)$, where S is the set of markings and I corresponding with m_0 is the initial states. The transition relation is defined as follows: there is an edge $m \rightarrow m'$ if and only if there is some transition $t \in T$ such that when transition t was fired, the marking m yields a new marking m'. The coverability problem for PN is defined as the coverability problem on this WSTS.

Definition 5 (PN safety problem). Given a Petri net $\mathcal{PN} = (P, T, F)$ and the initial marking m_0, we get a WSTS $(S, I, \rightarrow, \leq)$. Then given an $\leq$-upward-closed set $U \subseteq S$, does there exist a sequence $x_0 \rightarrow x_1 \rightarrow \cdots \rightarrow x_l$ such that $x_0 \in I$ and $x_l \in U$? We write safe$(\mathcal{PN}, I, U)$ if the answer is "no."

Example 6. As shown in Figure 1, if the initial marking is $m_0 = \langle 1, 0\rangle$, and the safety property is described by the upward-closed set $p_2 \geq 2$, there exists a sequence $\langle 1, 0\rangle \rightarrow \langle 2, 0\rangle \rightarrow \langle 1, 1\rangle \rightarrow \langle 0, 2\rangle$, which means the Petri net is not *safe*.

4. Incremental Bounded Model Checking Algorithm

This section describes how to bound a Petri net to an equivalent FSM and then verifies the safety properties by using SAT-based model checkers. An incremental method is proposed, which benefits more from the modern SAT solver by assumption.

4.1. Cut Off the Petri Net to FSM. In general, Petri nets are infinite state systems, as the number of tokens can be assigned with arbitrary $n \in \mathbb{N}$. In order to use the finite state model checking algorithms, we cut off the Petri net to FSM by a given boundary $B \in \mathbb{N}$. As the counter abstraction is used to model the parameterized systems, the boundary B can be defined as the current process number.

Definition 7 (finite state machine). A finite state machine is a quadruple $\mathcal{M} = \langle S, R, S_0, \text{In}\rangle$, where S is a finite set of states, $R \subseteq S \times S$ is the transition relations, $S_0 \subseteq S$ is the initial states, and In is a finite set of inputs.

To cut off the Petri net to FSM, we represent Petri nets as follows. Let $P = (p_1, p_2, \ldots, p_n)$ be the set of places. A marking m is represented as the tuple of natural numbers

$\langle m(p_1), m(p_2), \ldots, m(p_n)\rangle$. A transition t is represented as a pair $(\mathbf{g}, \mathbf{e}) \in \mathbb{N}^n \times \mathbb{N}^n$, where $\mathbf{g}$ is the guards and $\mathbf{e}$ is the effects. Formally, $\mathbf{g} = (F(p_1, t), F(p_2, t), \ldots, F(p_n, t))$, which represents the enabling condition, and $\mathbf{e} = (F(t, p_1), F(t, p_2), \ldots, F(t, p_n))$, which represents the yield marking m'.

The Petri net is a concurrent model, in which just one transition can be fired at one time. In the equivalent FSM, we introduce extra inputs to simulate and select one transition to be fired randomly.

Definition 8 (B-bounded Petri net, B-bounded upward-closed set). Given a Petri net $\mathscr{PN} = (P, T, F)$ and a boundary $B \in \mathbb{N}$, a B-bounded Petri net PN_B is a $\mathrm{FSM}\langle S, R, S_0, \mathrm{In}\rangle$, where S is the subset for all markings, so that, for a marking m, $m(p_i) \leq B$ for all $1 \leq i \leq n$; R is the subset of T, so that, for $t = (\mathbf{g}, \mathbf{e}) \in T$, the guard $\mathbf{g}$ can be fired under the boundary B; S_0 is the bounded initial marking m_0; In is the set of extra inputs for selecting which rule to be fired. Given an upward-closed set U, U_B is the B-bounded upward-closed set, which cuts off the infinite U to finite U_B by the boundary B.

B is the total token's boundary, so the summary of each place's token number should be less than or equal to B. Here an adder is used to count the total tokens, which will be discussed in Section 4.3.

The bad states are presented with an upward-closed set U in Petri nets. We bound U with the boundary B, by cutting off the upward-closed set with B. For instance, for an upward-closed set $p_2 \geq 2$, and the boundary 10, the bounded upward-closed set is $2 \leq p_2 \leq 10$.

Here, we introduce a function $cut\text{-}off(^*, B)$ to map the Petri net PN or upward-closed set U to B-bounded Petri net or B-bounded upward-closed set. That is to say, $\mathrm{PN}_B = cut\text{-}off(\mathrm{PN}, B)$, and $U_B = cut\text{-}off(U, B)$.

4.2. SAT-Based Model Checking Algorithms for Petri Nets. The Boolean Satisfiability (SAT) problem is a well-known NP-complete constraint satisfaction problem. With the introduction of bounded model checking (BMC) [37], it becomes clear that SAT solvers can be used for model checking [12]. There has been significant progress on SAT-based model checking techniques in the past two decades, including BMC, interpolation [38], and IC3.

IC3, known also as property directed reachability (PDR), is a recently proposed SAT-based model checking technique for the analysis of sequential circuits. IC3 maintains a list of trace: $[R_0, R_1, \ldots, R_N]$. The first element R_0 is special; it is simply identified with the initial states. For $k > 0$, R_k is a set of clauses that represents an overapproximation of the states reachable from the initial states in k steps or less. Together with the trace, the IC3 algorithm consists of a set of proof-obligations, which consists of a frame number k and a cube s. By manipulating the trace and the set of proof-obligations, IC3 gets new facts and adds them into the trace until it either (1) produced an inductive invariant proving the property or (2) added a proof-obligation at frame 0 with a cube that intersects the initial states, which is a counterexample. Without unrolling the model, IC3 performs better than most SAT-based model checking algorithms, especially in memory consumption.

Given a Petri net model PN, the safety property U, and the boundary $B \in \mathbb{N}$, U is expressed as an upward-closed set, and B is the current process number of the parameterized system to be verified. By Definition 8, we create the bounded Petri net model and bounded upward-closed set by the function $cut\text{-}off(*, B)$. PN_B is a FSM and can be translated into propositional logic directly. U_B represents the safety property for PN_B. An SAT-based model checker is used to check the property. There are multiple choices of model checkers. Here the state-of-the-art model checker IC3 was used as the back-end engine, because of its lower memory consumption. If the model checker returns UNSAT, it means PN is safe for the current boundary B; otherwise, a counterexample will be found.

We note this method as B-bounded model checking algorithm. Because of the use of finite state model checking algorithms, the B-bounded model checking algorithm will terminate when we find a counterexample or prove the safety.

For parameterized verification problem, we want to verify the system on an arbitrary parameter. If the maximum process number is M, we need M times single runs for $1 \leq B \leq M$. For each single run, the FSM is encoded into a new propositional formula to use an SAT solver, separately.

Fortunately, the modern SAT solver supports the incremental mechanism, so that the new SAT problem can be solved based on the previous solve result. The back-end SAT engine used in this paper is MINISAT [39], which supports the incremental mechanism by assumption. There is a vector to store the assumption variables which will be assigned to *True*. Hence, we introduce some extra variables, named active literals, to control the boundary of the bounded Petri net model.

Definition 9 (Inc-bounded Petri net, Inc-bounded upward–closed set). Given a Petri net $\mathscr{PN} = (P, T, F)$, a boundary $n \in \mathbb{N}$, and a maximum boundary $m \in \mathbb{N}$, an Inc-bounded Petri net $\mathrm{PN}_{n\to m}$ is a m-bounded Petri net PN_m equipped with the active literals vector $\mathbf{v}$ for controlling the boundary incrementally. An Inc-bounded upward-closed set $U_{n\to m}$ is a m-bounded upward-closet set U_m equipped with $\mathbf{v}$.

If $\mathbf{v} = \{a_{n+1}, a_{n+2}, \ldots, a_m\}$, the current boundary is n. If a_{n+1} is popped out, the boundary increases to $n + 1$. If $\mathbf{v} = \emptyset$, the boundary reaches the maximum boundary m. A new function *inc-cut-off(*,n,m)* is introduced to map the Petri net PN or upward-closed set U to Inc-bounded Petri net or Inc-bounded upward-closed set. We write that $\mathrm{PN}_{n\to m} = inc\text{-}cut\text{-}off(\mathrm{PN}, n, m)$, and $U_{n\to m} = inc\text{-}cut\text{-}off(U, n, m)$.

Algorithm 1 shows our new algorithm to verify the bounded Petri net from boundary n to m incrementally. The inputs are a Petri net model PN, safety property U, a base boundary n, and a maximum boundary m. U is expressed as an upward-closed set. The algorithm returns *unsafe* if it finds a counterexample; otherwise, it returns *safe* and proves the system is safe from parameters n to m.

Lines 1–5. Generate the active literals and push them into the assumption vector $\mathbf{v}$. Then, set the current boundary as n.

Input:
PN: a Petri net model to describe the parameterized protocol
U: an upward-closed set to describe the safety property
n: base boundary
m: maximum boundary
Output:
safe or *unsafe*
(1) initial assumption vector $\mathbf{v} \coloneqq \varnothing$ // push all active literals into assumption vector
(2) **for** $i \coloneqq m;\ i > n;\ i \coloneqq i - 1$ **do**
(3) $\mathbf{v}$.push(a_i)
(4) **end for**
(5) $i \coloneqq n$ // the initial boundary is n
(6) $\mathrm{PN}_{n\to m} \coloneqq$ *inc-cut-off*(PN, n, m) // create the incremental bounded Petri net
(7) $U_{n\to m} \coloneqq$ *inc-cut-off*(U, n, m)
(8) **while** $\mathbf{v} \neq \varnothing$ **do**
(9) // use an SAT-based model checker to verify the property at the boundary i
(10) **if** $\mathrm{PN}_{n\to m} \models U_{n\to m}$ **then**
(11) print CEX
(12) RETURN *unsafe*
(13) **else**
(14) print "PN is safe for current boundary i"
(15) $\mathbf{v}$.pop() // the boundary is increased by 1
(16) $i \coloneqq i + 1$
(17) **end if**
(18) **end while**
(19) RETURN *safe*

Algorithm 1: Incremental bounded model checking algorithm.

Lines 6-7. Create the incremental bounded Petri net model by using the assumption vector $\mathbf{v}$.

Lines 8–19. The while-loop is the main routine to verify the model incrementally. If the condition is satisfied at line 10, the model checker finds a counterexample and returns *unsafe*. If the model bounded by i is safe, then the algorithm increases the boundary at lines 15 and 16 and calls the SAT-based model checker again to solve the new model with higher boundary. When vector $\mathbf{v}$ is empty, the algorithm proves that the input PN is safe from boundaries n to m.

Algorithm 1 reuses the context from the previous solving results. The main routine in Algorithm 1 is based on an SAT-based model checker. Hence, the algorithm terminates when it finds a counterexample at line 12 or proves safety at line 19.

4.3. Implementation and Optimization. In this section, we introduce key points which make a great contribution in improving the performance.

The Petri Net Format. The input Petri net is encoded in the MIST format (https://github.com/pierreganty/mist/). Each place is mapped to a variable, and each transition corresponds to a rule. For each rule, there are guards and effects, as described in Definition 8. The guards are the conditions under which this rule can be fired, and the effects describe how tokens transfer from places. The target is the safety property to be verified, which is expressed as an upward-closed set.

The FSM Format. AIGER (http://fmv.jku.at/aiger/) is a format, library, and set of utilities for And-Inverter Graphs (AIGs). The hardware model checking competition (HWMCC) uses AIGER as input format, and most modern model checkers support AIGER as the input model. AIGER is a good way to describe FSM and can be translated into a propositional logic for an SAT solver. The bounded Petri net is encoded as an AIGER model, where each place corresponds to state variables in Boolean value. Extra input variables are introduced to select which rule to be fired and then update the state variables to set up the transition relations equally.

Encoding: Binary versus Unary. Encoding is important for SAT solvers. In this paper, both binary and unary encodings

Table 1: Encoding: binary versus unary.

n	Binary	One-hot
0	000	00000001
1	001	00000010
2	010	00000100
3	011	00001000
4	100	00010000
5	101	00100000
6	110	01000000
7	111	10000000

were used to encode the places in the Petri net model. As shown in Table 1, binary and unary encodings are used to encode the natural numbers. One-hot encoding is one possible unary encoding, where just one bit is "1" and the others are all "0." The binary encoding needs $\lceil \log_2[N] \rceil$ bits, and the unary encoding needs N bits to encode the natural number $0 \sim N$.

Full Adder. A full adder is designed to count the total token numbers to bound the Petri net. As binary and unary encodings are used to represent the token numbers for each place, we present how to design an n-bit binary and unary full adder, respectively.

A n-bit binary full adder is just combing n single 1-bit binary adders together. The 1-bit binary adder can be described using the following logics. The logics AND, OR, and XOR are represented as $\wedge$, $\vee$, and $\oplus$, respectively.

$$\begin{aligned} S &= A \oplus B \oplus C_{\text{in}} \\ C_{\text{out}} &= (A \wedge B) \vee \left(C_{\text{in}} \wedge (A \oplus B)\right) \end{aligned} \tag{1}$$

There are many different ways to design unary adder, but a simple n-bit unary adder is presented as follow:

$$S_i = \bigwedge_{j=0}^{i} \left(A[j] \wedge B[i-j]\right), \quad 0 \le i < n. \tag{2}$$

Structural Information. When using unary encoding to represent the token's number of places, it is important to give this structural information to the SAT solver. Hence, we add some extra logics as a constraint to the AIGER circuits. Figure 2 shows the logics to check if the n-bit vector $\mathbf{x}$ is encoded in one-hot. The total number of gates is $3 * n + 2$.

5. Experimental Evaluation

We have implemented the incremental bounded model checking algorithm in a tool named PNPV. PNPV is implemented with C++ and uses MINISAT as the back-end SAT solver. All input instances are encoded in the MIST format.

To measure PNPV's performance, we compare with TDA [32], which was used to verify parameterized cache coherence protocols. We also compare with the MIST toolkit (https://github.com/pierreganty/mist) with the classical backward [40] and EEC [41] algorithm for WSTS coverability on Petri nets instances.

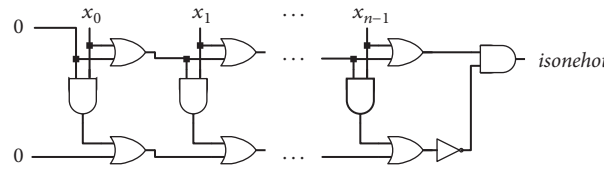


Figure 2: One-hot encoding checking logics.

All experiments are performed on a machine, with Intel 2.60 GHz CPU and 16 GB main memory, running CentOS 6.5 in 64-bit. The running time is limited to 600 seconds and memory to 2 GB.

5.1. Benchmarks. Memory coherence is very important for both the multicore processors and mobile CPSs. We verify several classical parameterized cache coherence protocols as described in [35]. We collected 12 Petri nets from the MIST repository, where six bounded Petri nets are all safe and six plain Petri nets are all unsafe. Some of those Petri nets are used to model the communication protocols in mobile CPSs.

5.2. Evaluation. As shown in Table 2, the unary encoding performs about 4~19 times better than binary. By using unary encoding, we can solve 160 processes for Illinois and Firefly protocols, but just 22 and 18 when using binary. For Berkeley, PNPV solves 146 and 18 processes by unary and binary, respectively. German protocol is an industry-like cache coherence protocol, for which 97 and 5 processes are solved by using unary and binary encoding, respectively. The CSM-broad and Dragon protocols both have about 90 processes solved by using the unary encoding but just 9 and 24 processes by using the binary encoding, respectively. The unary encoding is more competitive than binary on all protocols. We identify two reasons for unary's better performance.

The first reason is the difference in expressing the transition in FSM. Though there are more variables used to encode the places, the transition relations can be generated by shifting the variables. The logics for transitions are simpler than binary, and the SAT problem is easy to solve.

The second reason is the use of the IC3 algorithm. IC3 is an incremental inductive algorithm, which builds the overapproximation from the last SAT results incrementally. The unary encoding is more efficient for learning clauses.

Figure 3 shows a comparison of the B-bounded model checking algorithm with the incremental bounded model checking algorithm on five parameterized protocols. The incremental bounded model checking algorithm is competitive on all instances, especially for Firefly, Illinois, and Berkeley. The main reason is that the B-bounded algorithm cannot

Table 2: The maximum process number solved by different encodings. As presented in Section 4.3, one-hot encoding is used for unary. Six parameterized protocols are used to test the performance, and the maximum process number is represented in the table. All results are collected by running Algorithm 1.

Protocols	Unary	Binary
Illinois	160	22
Firefly	160	18
Berkeley	146	18
German	97	5
CSMbroad	93	9
Dragon	90	24

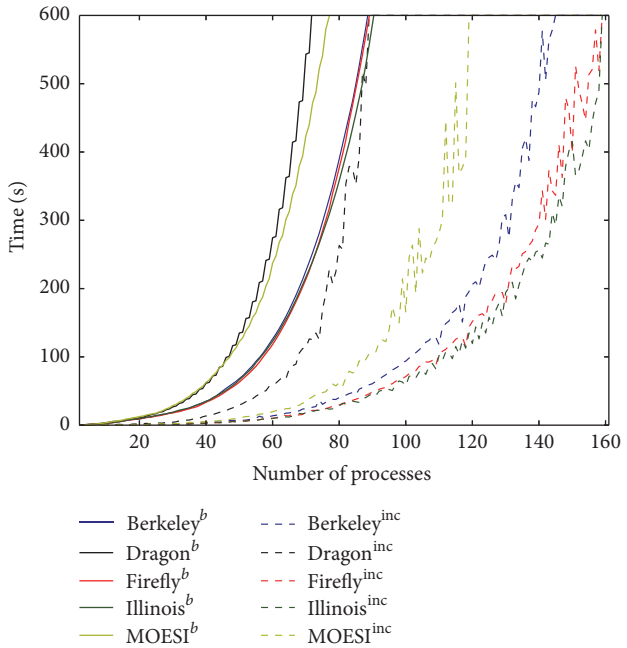


Figure 3: The performance comparison of B-bounded with incremental bounded model checking algorithm. The superscripts b and inc represent B-bounded model checking algorithm and incremental bounded model checking algorithm, respectively.

reuse the previous result to prove the current verification problem. Algorithm 1 gives us a good solution with little expenses when adding active literals for the generated FSM.

TDA [32, 33] uses X-abstract and Y-abstract to reduce the state space of parameterized systems and speed up the verification performance. Asynchronous composition is the key part of TDA, but it suffers from higher memory consumption. As shown in Table 3, PNPV performs better than TDA on all parameterized protocols. For Berkeley and Dragon protocols, TDA solves 14 processes, but PNPV solves 146 and 90 processes, respectively. PNPV handles 160 processes for Firefly and Illinois, but just 22 processes are solved by TDA. Table 3 shows that TDA suffers from high memory consumption, as most instances hit the memory limit before hitting the time limit. PNPV is also competitive with respect to speed when comparing with TDA, especially for large process numbers. As the results are all zero when the process number $n < 10$, we present the data from $n = 10$.

MIST is a tool to check safety properties against Petri net-like models. To compare with PNPV, we select the classical backward and EEC algorithms to run 12 Petri net benchmarks. Six out of 12 instances are bounded Petri nets. The total token numbers of the bounded Petri nets are limited. All of the six bounded Petri nets are safe. To test PNPV's ability of bug finding, six unsafe instances were collected from the MIST toolkit to evaluate the performance.

As shown in Table 4, PNPV performs better than both backward and EEC algorithms for all unsafe instances. Both in time and memory usage, PNPV is competitive. For six bounded Petri nets, PNPV wins on four out of six instances.

6. Conclusion and Future Works

We introduced an incremental bounded model checking algorithm to verify the safety properties of parameterized protocols in mobile CPS. By using counter abstraction, the protocol is modeled as a Petri net. Then the state-of-the-art SAT-based model checking algorithm is used to verify the safety properties. The algorithm can be used to verify parameterized systems, including cache coherence protocols, mutual exclusion communication protocols, and common concurrency primitives in mobile CPSs. The results show that our new approach can greatly scale the verification capabilities compared favorably against several recently published approaches. Due to using IC3 as the back-end model checking algorithm, our method is significant for its lower memory consumption.

There are two directions to extend the current work in the future. The first one is to study the property to be verified. Liveness would be an interesting direction, as the liveness property can be converted into safety. Security problems and run time verification would also be a good direction in the future. The second one is to model more complex systems. The ideas we have presented are naturally applicable to other concurrency systems modeled by Petri net or its extension. It is natural to shift SAT solver to SMT solver, and it would be a good way to improve the scalability.

Table 3: Comparison of running time and memory consumption between PNPV and TDA. *n* is the number of processes, which is the parameter of the parameterized protocols. *Max* stands for the maximum number of solved processes. Memory consumption is in MB and running time in seconds.

	Berkeley				Dragon				Firefly				Illinois			
n	TDA		PNPV		TDA		PNPV		TDA		PNPV		TDA		PNPV	
	Time	Mem.	Time	Mem.	Time	Mem.	Time	Mem.	Time	Mem.	Time	Mem.	Time	Mem.	Time	Mem.
10	0.97	16.00	0.62	0.00	1.03	15.50	0.60	0.00	0.59	0.00	0.37	0.00	0.04	0.00	0.49	0.00
11	1.69	46.20	0.66	0.00	0.75	46.20	0.72	0.00	0.61	0.00	0.41	0.00	0.06	0.00	0.54	0.00
12	3.03	130.8	0.71	0.00	3.11	135.1	0.90	6.10	0.64	0.00	0.47	0.00	0.09	0.00	0.60	0.00
13	6.03	416.0	0.78	0.00	8.07	431.4	1.08	6.80	0.69	0.00	0.52	0.00	0.14	0.00	0.67	0.00
14	19.55	1168	0.85	0.00	24.47	1210	0.37	7.10	0.77	0.00	0.60	0.00	0.22	0.00	0.75	0.00
15	Memout		0.94	0.00	Memout		0.64	7.60	0.95	7.90	0.67	0.00	0.42	7.40	0.84	0.00
16	Memout		1.06	0.00	Memout		1.07	8.10	1.24	19.10	0.79	0.00	0.74	18.6	0.96	0.00
17	Memout		0.20	5.00	Memout		0.48	8.70	1.14	40.60	0.89	0.00	1.59	40.60	1.10	5.20
18	Memout		0.37	5.30	Memout		1.13	9.80	1.56	76.20	1.04	5.20	2.25	78.20	0.25	5.20
19	Memout		0.56	5.60	Memout		0.72	10.20	3.68	161.1	0.18	5.40	3.38	161.0	0.42	5.40
20	Memout		0.79	5.90	Memout		1.67	10.90	6.77	311.8	0.39	5.80	6.81	315.4	0.64	5.80
21	Memout		1.04	6.10	Memout		1.52	11.50	13.75	638.2	0.57	5.90	14.33	644.9	0.89	6.00
22	Memout		0.35	6.50	Memout		1.49	12.60	29.61	1249	0.84	6.20	27.28	1224	1.17	6.20
23	Memout		0.70	6.80	Memout		1.98	13.20	Memout		1.09	6.60	Memout		0.49	6.80
Max	14		146		14		90		22		160		22		160	

Table 4: Comparison of running time and memory consumption for different algorithms on Petri net benchmarks. Memory consumption is in MB, and running time is in seconds.

Problem	PNPV		Backward		EEC	
Instance	Time	Mem.	Time	Mem.	Time	Mem.
			Safe instances			
kanban	7.39	6.60	427.62	54.90	**0.83**	**0.00**
lamport	**0.42**	0.00	0.63	0.00	0.84	0.00
newdekker	**0.55**	3.30	0.71	0.00	0.85	0.00
newrtp	**0.58**	0.00	0.72	0.00	0.86	0.00
peterson	**0.68**	0.00	0.75	0.00	0.86	0.00
read-write	2.31	4.50	**0.81**	0.00	0.91	0.00
			Unsafe instances			
kanban	**1.05**	**5.20**	459.92	54.90	Timeout	
leabasicapproach	**0.08**	0.00	0.93	0.00	0.08	0.00
pingpong 2	**0.10**	0.00	0.93	0.00	0.13	0.00
pingpong_wrong	**0.13**	0.00	0.94	0.00	0.20	0.00
pncsacover	**6.22**	**4.90**	Timeout		23.89	8.70
pncsasemiliv	**1.06**	**4.40**	1.40	6.00	23.42	8.70

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to acknowledge that this work was supported by the National Natural Science Foundation of China (Grant no. 61133007). The authors thank Carl Kwan for helpful and detailed comments and suggestions.

References

[1] R. Rajkumar, I. Lee, L. Sha, and J. Stankovic, "Cyber-physical systems: the next computing revolution," in *Proceedings of the 47th Design Automation Conference (DAC '10)*, pp. 731–736, New York, NY, USA, June 2010.

[2] M. Conti, S. K. Das, C. Bisdikian et al., "Looking ahead in pervasive computing: challenges and opportunities in the era of cyberphysical convergence," *Pervasive and Mobile Computing*, vol. 8, no. 1, pp. 2–21, 2012.

[3] X. Hu, K. Bai, J. Cheng et al., "MeDJ: multidimensional emotion-aware music delivery for adolescent," in *Proceedings of the World Wide Web*, pp. 793–794, Proceedings of the World Wide Web, 2017.

[4] J. White, S. Clarke, C. Groba, B. Dougherty, C. Thompson, and D. C. Schmidt, "R&D challenges and solutions for mobile cyber-physical applications and supporting internet services," *Journal of Internet Services and Applications*, vol. 1, no. 1, pp. 45–56, 2010.

[5] L. Zhou, X. Hu, E. C.-H. Ngai et al., "A dynamic graph-based scheduling and interference coordination approach in heterogeneous cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3735–3748, 2016.

[6] X. Hu, T. H. S. Chu, V. C. M. Leung, E. C.-H. Ngai, P. Kruchten, and H. C. B. Chan, "A Survey on mobile social networks: applications, platforms, system architectures, and future research directions," *IEEE Communications Surveys and Tutorials*, vol. 17, no. 3, pp. 1557–1581, 2015.

[7] X. Hu, J. Zhao, B.-C. Seet, V. C. M. Leung, T. H. S. Chu, and H. Chan, "S-aframe: agent-based multilayer framework with context-aware semantic service for vehicular social networks," *IEEE Transactions on Emerging Topics in Computing*, vol. 3, no. 1, pp. 44–63, 2015.

[8] S. M. German and A. P. Sistla, "Reasoning about systems with many processes," *Journal of the Association for Computing Machinery*, vol. 39, no. 3, pp. 675–735, 1992.

[9] R. Bloem, S. Jacobs, A. Khalimov et al., "Decidability in parameterized verification," *Synthesis Lectures on Distributed Computing Theory*, vol. 6, no. 1, pp. 1–170, 2015.

[10] W. Hunt, "Modeling and verification of cyber-physical systems," in *Proceedings of the National Workshop on High-Confidence Automotive Cyber-Physical Systems*, 2008.

[11] M. U. Sanwal and O. Hasan, "Formal verification of cyber-physical systems: coping with continuous elements," in *Proceedings of the International Conference on Computational Science and Its Applications*, pp. 358–371, Springer, 2013.

[12] E. M. Clarke, O. Grumberg, and D. Peled, *Model Checking*, MIT Press, 1999.

[13] K. L. McMillan, "Symbolic model checking," in *Symbolic Model Checking*, pp. 25–60, Springer, 1993.

[14] A. Pnueli, J. Xu, and L. Zuck, "Liveness with $(0, 1, \infty)$-counter abstraction," in *Proceedings of the International Conference on Computer Aided Verification*, pp. 107–122, Springer, Berlin, Germany, 2002.

[15] E. Clarke, M. Talupur, and H. Veith, "Environment abstraction for parameterized verification," in *Proceedings of the International Workshop on Verification, Model Checking, and Abstract Interpretation*, pp. 126–141, Springer, 2006.

[16] D. Xu and Y. Deng, "Modeling mobile agent systems with high level Petri nets," in *Proceedings of the 2000 IEEE International Conference on Systems, Man and Cybernetics*, vol. 5, pp. 3177–3182, October 2000.

[17] T. M. Chen, J. C. Sanchez-Aarnoutse, and J. Buford, "Petri net modeling of cyber-physical attacks on smart grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 741–749, 2011.

[18] A. R. Bradley, "SAT-based model checking without unrolling," in *Proceedings of the International Workshop on Verification, Model Checking, and Abstract Interpretation*, pp. 70–87, Springer.

[19] N. Een, A. Mishchenko, and R. Brayton, "Efficient implementation of property directed reachability," in *Proceedings of the Formal Methods in Computer-Aided Design FMCAD '11*, pp. 125–134, November 2011.

[20] R. Akella and B. M. McMillin, "Model-checking BNDC properties in Cyber-physical systems," in *Proceedings of the 33rd Annual IEEE International Computer Software and Applications Conference COMPSAC '09*, pp. 660–663, July 2009.

[21] L. Bu, Q. Wang, X. Chen et al., "Toward online hybrid systems model checking of cyber-physical systems' time-bounded short-run behavior," *ACM SIGBED Review*, vol. 8, no. 2, pp. 7–10, 2011.

[22] K. Bae, J. Krisiloff, J. Meseguer, and P. C. Ölveczky, "Designing and verifying distributed cyber-physical systems using Multirate PALS: an airplane turning control system case study," *Science of Computer Programming*, vol. 103, pp. 13–50, 2015.

[23] X. Hu, T. H. S. Chu, H. C. B. Chan, and V. C. M. Leung, "Vita: a crowdsensing-oriented mobile cyber-physical system," *IEEE Transactions on Emerging Topics in Computing*, vol. 1, no. 1, pp. 148–165, 2013.

[24] X. Hu, X. Li, E. C.-H. Ngai, V. C. M. Leung, and P. Kruchten, "Multidimensional context-aware social network architecture for mobile crowdsensing," *IEEE Communications Magazine*, vol. 52, no. 6, pp. 78–87, 2014.

[25] F. C. Nemtanu, I. M. Moise, M. G. Beldescu, and V. Iordache, "Model of cloudified traveller information system based on petri nets," in *Proceedings of the 36th International Spring Seminar on Electronics Technology Automotive Electronics*, 2013.

[26] S.-Z. Zhang, Z.-H. Ding, and J.-L. Hu, "Modeling fault tolerated mobile agents by colored petri nets," in *International Conference on Intelligent Computing*, pp. 607–617, Springer, 2015.

[27] M. C. Browne, E. M. Clarke, and O. Grumberg, "Reasoning about networks with many identical finite state processes," *Information and Computation*, vol. 81, no. 1, pp. 13–31, 1989.

[28] K. L. McMillan, "Parameterized verification of the flash cache coherence protocol by compositional model checking," in *Proceedings of the Advanced Research Working Conference on Correct Hardware Design and Verification Methods*, pp. 179–195, Springer, Berlin Heidelberg, Germany.

[29] J. Esparza, P. Ganty, and R. Majumdar, "Parameterized verification of asynchronous shared-memory systems," *Journal of the ACM*, vol. 63, no. 1, article 10, 2016.

[30] C. Munoz, V. Carreño, and G. Dowek, "Formal analysis of the operational concept for the small aircraft transportation system," in *Rigorous Development of Complex Fault-Tolerant Systems*, pp. 306–325, Springer Berlin Heidelberg, 2006.

[31] T. T. Johnson and S. Mitra, "Parametrized verification of distributed cyber-physical systems: an aircraft landing protocol case study," in *Proceedings of the IEEE/ACM 3rd International Conference on Cyber-Physical Systems ICCPS '12*, pp. 161–170, April 2012.

[32] Y. Guo, W. Qu, L. Zhang, and W. Xu, "State space reduction in modeling checking parameterized cache coherence protocol by two-dimensional abstraction," *Journal of Supercomputing*, vol. 62, no. 2, pp. 828–854, 2012.

[33] L. Zhang, W. Qu, Y. Guo, and S. Li, "Automatic abstraction for verification of parameterized systems," *Journal of Computer-Aided Design and Computer Graphics*, vol. 26, no. 6, pp. 991–998, 2014.

[34] G. Geeraerts, "On the expressive power of petri nets with transfer arcs vs. petri nets with reset arcs," Tech. Rep. 572, Université Libre de Bruxelles, 2007.

[35] G. Delzanno, "Automatic verification of parameterized cache coherence protocols," in *Proceedings of the International Conference on Computer Aided Verification*, pp. 53–68, 2000.

[36] A. Finkel and P. Schnoebelen, "Well-structured transition systems everywhere!," *Theoretical Computer Science*, vol. 256, no. 1-2, pp. 63–92, 2001.

[37] A. Biere, A. Cimatti, E. Clarke, and Y. Zhu, "Symbolic Model Checking without BDDs," in *Proceedings of the International conference on tools and algorithms for the construction and analysis of systems*, vol. 1579, pp. 193–207, Springer.

[38] K. L. McMillan, "Interpolation and SAT-based model checking," in *Proceedings of the International Conference on Computer Aided Verification*, vol. 2725, pp. 1–13, Springer.

[39] N. Sorensson and N. Een, "Minisat v1.13-a sat solver with conflict-clause minimization," *SAT*, vol. 2005, no. 53, 2005.

[40] P. A. Abdulla, K. Cerans, B. Jonsson, and Y.-K. Tsay, "General decidability theorems for infinite-state systems," in *Proceedings of the 11th Annual IEEE Symposium on Logic in Computer Science LICS'96*, pp. 313–321, 1996.

[41] G. Geeraerts, J.-F. Raskin, and L. Van Begin, "Expand, enlarge and check: new algorithms for the coverability problem of WSTS," *Journal of Computer and System Sciences*, vol. 72, no. 1, pp. 180–203, 2006.

Hindawi
Mobile Information Systems
Volume 2017, Article ID 3175186, 11 pages
https://doi.org/10.1155/2017/3175186

Research Article

Improved Object Proposals with Geometrical Features for Autonomous Driving

Yiliu Feng, Wanzeng Cai, Xiaolong Liu, Huini Fu, Yafei Liu, and Hengzhu Liu

College of Computer, National University of Defense Technology, Changsha, China

Correspondence should be addressed to Yiliu Feng; fengyiliu11@nudt.edu.cn

Received 13 February 2017; Accepted 22 March 2017; Published 26 April 2017

Academic Editor: Zhengguo Sheng

This paper aims at generating high-quality object proposals for object detection in autonomous driving. Most existing proposal generation methods are designed for the general object detection, which may not perform well in a particular scene. We propose several geometrical features suited for autonomous driving and integrate them into state-of-the-art general proposal generation methods. In particular, we formulate the integration as a feature fusion problem by fusing the geometrical features with existing proposal generation methods in a Bayesian framework. Experiments on the challenging KITTI benchmark demonstrate that our approach improves the existing methods significantly. Combined with a convolutional neural net detector, our approach achieves state-of-the-art performance on all three KITTI object classes.

1. Introduction

Object detection has been developed in many years and there are a variety of robust approaches [1–5]. In the early years, most of them follow the sliding-window paradigm. But enormous numbers of windows would waste a large amount of efforts on no-object areas. In order to overcome this problem, an effective framework is proposed: object proposals generation followed by a classifier. Most of the methods are designed to generate object proposals for general object detection, such as Edgeboxes [6] and Selective Search [7]. They both work well on the PASCAL VOC dataset [8].

However these methods would suffer a great performance degradation, when they are applied to autonomous driving scene, such as the challenging KITTI benchmark [9], which contains many small objects, occlusion, high saturated areas, and even shadows.

In this paper, we propose an effective approach to improve the results of object proposals in autonomous driving scene. Our work is motivated by the following observations. First, there are three primary objects, in autonomous driving scene, Car, Cyclist, and Pedestrian. These three objects usually lie on the ground with different height. So the proposals should lie on the ground. Second, the real-world size of objects in one category would vary far less than their image-world size, but the real-world size of different categories are also different. It is helpful to use the object size prior of object as an indicator to generate proposals. The details are discussed in Section 3.

This paper has two fundamental contributions.

(1) We propose two new geometric features, AR and SD2, to represent the object size prior. We exploit D2R as an indicator to constraint the proposals lying on the ground. These features are demonstrated to be effective for generating fewer proposals with higher recall.

(2) We deeply analyze the four geometric features, AR, SD2, DMD, and D2R, and propose a method to combine these features with existing methods efficiently. The final results on the KITTI object detection benchmark achieve the state-of-the-art performance in stereo-based methods.

Since it is inevitable to use the depth information to compute the geometric features, we assume a stereo image pair as an input and obtain depth information via the state-of-the-art approach by Yamaguchi et al. [10].

2. Related Work

The main idea of object proposal method is to generate relatively fewer number of bounding boxes that contain the

objects in an image that we are interested in with high recall. Existing proposal generation methods are often based on low-level image features, which can be divided into two categories generally: grouping methods and window scoring methods.

2.1. Grouping Methods. Grouping proposal methods aim to generate multiple segments that are likely to correspond to objects. To cover different objects with various size, most methods attempt to merge the output of a hierarchical image segmentation algorithm. The decision to merge segments is designed manually typically based on superpixel shape, appearance features, and boundary estimates.

Selective Search [7] is one of the most well-known grouping methods which greedily merges superpixels to generate proposals. The method has no learned parameters and has been broadly used as the proposal method of choice by many state-of-the-art object detectors, such as the R-CNN detector.

In order to detect objects with different size, MCG [11] propose an algorithm for fast computing multiscale hierarchical segmentation. They merge the segments based on edge strength and ranking the results using appropriate features.

Since SS and MCG both need an initial image segmentation which impacts the object proposal results, CPMC [12] does not have initial segmentations and uses graph cut directly on pixels. Then it ranks the resulting segments based on a large pool of features.

2.2. Window Scoring Methods. Window scoring methods are to score each candidate window to indicate how likely an object of interest is contained in it. Compared to grouping approaches these methods usually directly return bounding boxes with fast speed. However, they tend to generate proposals with low localization accuracy unless the window sampling is performed very densely.

Objectness [13, 14] is one of the earliest window scoring proposal methods. A model is trained to distinguish objects from the background and an initial set of proposals is generated from salient locations in an image. Then each proposal is scored by a Bayesian framework combining several image features including color, edge density, saliency, and superpixels straddling.

BING [15] is an extremely fast object proposal method (300 fps/s on CPU). Gradient features are used to train a simple linear classifier to detect object proposals in a sliding-window framework which can yield 96.2% recall with 1000 proposals at the IOU threshold of 0.5. Meanwhile BING needs to resize the candidate window to $8 * 8$ which leads to low localization accuracy when mapping the $8 * 8$ window back to the original image. The recall drops rapidly when the IOU threshold gets larger.

Edgeboxes [6] is a very fast and efficient region proposal method, which can generate millions of candidate boxes in a fraction of one second and achieve nearly 96% recall at overlap threshold of 0.5 by using 1000 proposals on the PASCAL VOC dataset. The main contribution of the method is that the number of contours wholly falling into a bounding box is indicative of the possibility of a box covering an object. All of the bounding boxes are generated by sliding-window algorithm and then scored by measuring the number of edge groups that exist in the box minus some of them that overlap the box's boundary.

However most previous methods are designed for general objects; they do not perform well in a particular scene such as the KITTI [9] benchmark. 3DOP [16] is an excellent proposal generation method which exploits object size priors, ground plane, and several depth informed features such as free space, point densities inside the box, visibility, and distance to the road to place proposals in the form of 3D bounding boxes. After generating a large number of proposals, the method scores every proposal by minimizing an energy function. The energy function encodes object size priors, ground plane, and a variety of depth informed features. Their final results achieve a 25% higher recall with 2,000 proposals than the state-of-the-art RGB-D method MCG-D [17] on the KITTI benchmark.

Most grouping and scoring methods mentioned above either purely use RGB appearance features or only use depth informed geometric features which ignore their complement of those two features. Although some methods, such as MCG-D, use RGB and depth features simultaneously, it is not suitable for autonomous driving because of the complex outdoor environment. In this paper, we propose a method to exploit both the appearance features and the geometric features. Our work formulates the problem by fusing those two complementary features in a Bayesian framework for obtaining high-quality object proposals in autonomous driving.

3. Methodology

As mentioned in previous sections, geometric features are important for improving the quality of object proposals. We introduce four geometric features: aspect Ratio, diagonal multiplication distance, area multiplication of the square of the object depth, and distance to the road.

3.1. Geometric Features

3.1.1. Aspect Ratio (AR). Objects in different classes usually have magnificent difference on appearance while those in the same class vary far less. Since an object is tightly bounded by a square box whose aspect ratio of the same class should vary in a specific range, based on this intuition, we use AR as a feature to assess the possibility of an image window covering a specific class. The aspect ratio of a square box is calculated as follows:

$$\mathrm{AR} = \frac{w_b}{h_b}, \tag{1}$$

where w_b is the width of a given bounding box while h_b is the height.

3.1.2. Area Multiplication of the Square of the Object Depth (SD2). Objects' sizes in the image can be measured by the bounding boxes covering them and they vary significantly across the dataset. Meanwhile, the real-world size of objects in the same class varies far less as mentioned in [18]. According to the optical imaging principles, the real-world size A_T and the image size A_I of the object have a specific relationship.

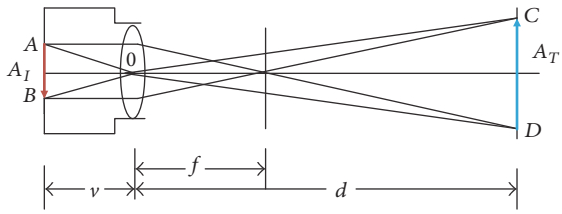


Figure 1: The imaging principle of the camera.

As shown in Figure 1, by using the homothetic triangle theory, the relationship between A_T and A_I can be described as follows:

$$A_T = \frac{d^2}{v^2} A_I, \tag{2}$$

where d is the real-world distance of the object and v is the camera focal length which is usually considered to be fixed.

Depth information has been utilized for object detection in recent years; it can be computed from disparity map or directly obtained by depth sensors, such as Kinect. In this paper, we use a stereo image pair as an input, compute the disparity map via the state-of-the-art approach by Yamaguchi et al. [10], and then calculate the depth by binocular vision theory:

$$\text{depth} = \frac{f * l}{\text{disparity}}, \tag{3}$$

where f is the focal length of the two lenses, l is the distance between two optical centers, and disparity is the horizontal disparity of two stereo-corresponding points. After calculating the depth of all pixels for each image, the average depth of a $3 * 3$ area around the center is used to approximate the depth of an object enclosed by the box:

$$d_{\text{box}} = \frac{1}{9} \sum_{x_i=x_c-1}^{x_c+1} \sum_{y_i=y_c-1}^{y_c+1} \text{depth}(x_i, y_i), \tag{4}$$

where $\text{depth}(x_i, y_i)$ is the depth of point (x_i, y_i) in image and $x_c = x_l + w_b/2$ and $y_c = y_l + h_b/2$ is the center point of the box.

As mentioned above, the relationship between the image size and the depth information of the object can be utilized as a proxy for the real-world object size approximately. The camera focal length can be ignored as it is considered to be a constant. Inspired by the observation of the relationship between real-world size and image size, we use the product of area of the bounding box and the square distance to the camera as an approximate representation of the object size in real-world. The SD2 can be written as

$$\text{SD2} = w_b * h_b * {d_{\text{box}}}^2, \tag{5}$$

where $w_b * h_b$ is the area of the bounding box and can be used as a representation of an object image size approximately.

3.1.3. Diagonal Multiplication Distance (DMD). DMD is the feature that could approximately represent the real-world object size [18].

$$\text{DMD} = \sqrt{{w_b}^2 + {h_b}^2} * d_{\text{box}}, \tag{6}$$

where $\sqrt{{w_b}^2 + {h_b}^2}$ is the diagonal of a bounding box and d_{box} is the depth of the box.

The distributions of DMD and SD2 on Car, Cyclist, and Pedestrian are shown in the second and the third row in Figure 2. It is obvious that DMD and SD2 vary a few in the same class and vary in different ranges which prove the analysis we discussed before.

3.1.4. Distance to the Road (D2R). Since all the annotated objects in the KITTI benchmark are on the ground, the ground plane can be used as an important indicator to predict the possibility that a proposal contains an object. It is more likely to cover an object when the proposal is close to ground plane and is less likely when the proposal is far away from the ground plane. We use the same method in [16] to compute the distance of every pixel to the ground. Then, as in (5), the average of a $3 * 3$ area around the center is used to measure the distance to the road of an object enclosed by the box:

$$\text{D2R} = \frac{1}{9} \sum_{x_i=x_c-1}^{x_c+1} \sum_{y_i=y_c-1}^{y_c+1} \text{Dist}(x_i, y_i). \tag{7}$$

The distribution of D2R on Car, Cyclist, and Pedestrian is shown in the last row in Figure 2.

3.2. Bayesian Framework. As the four proposal features are relatively complementary, using some of them at the same time may appear promising. AR gives only the proportion of object projection size in the image. DMD or SD2 is the replacement for the real-world object size, but either of them depends on precise depth calculated from disparity map. D2R denotes the distance to the road, which can roughly distinguish positive examples from negative examples.

To combine these features (AR, SC, DMD, SD2, and D2R), we train a Bayesian classifier to distinguish between positives and negatives. SC is the initial result of the existing method. For each training image, we sample all the proposals that have an IOU $\geq$ 0.6 with any ground truth as positive W_{obj} and IOU $<$ 0.35 as negative W_{bg}. As there are too many negative proposals we just select 600 randomly for each training image. In this paper, we choose a Naive Bayes approach [13, 14]. In the Naive Bayes model, the features are independent, so training consists of estimating the priors $p(\text{obj})$, $p(\text{bg})$ by relative frequency and the individual feature likelihoods $p(\text{feature} \mid c)$, feature $\in C$ and $c$obj, and bg from the training set we chosen before.

After training, when given a proposal we calculate its posterior probability using the following equation:

$$p\left(\text{obj} \mid C^1\right) = \frac{p\left(C^1 \mid \text{obj}\right) p\left(\text{obj}\right)}{p\left(C^1\right)}, \tag{8}$$

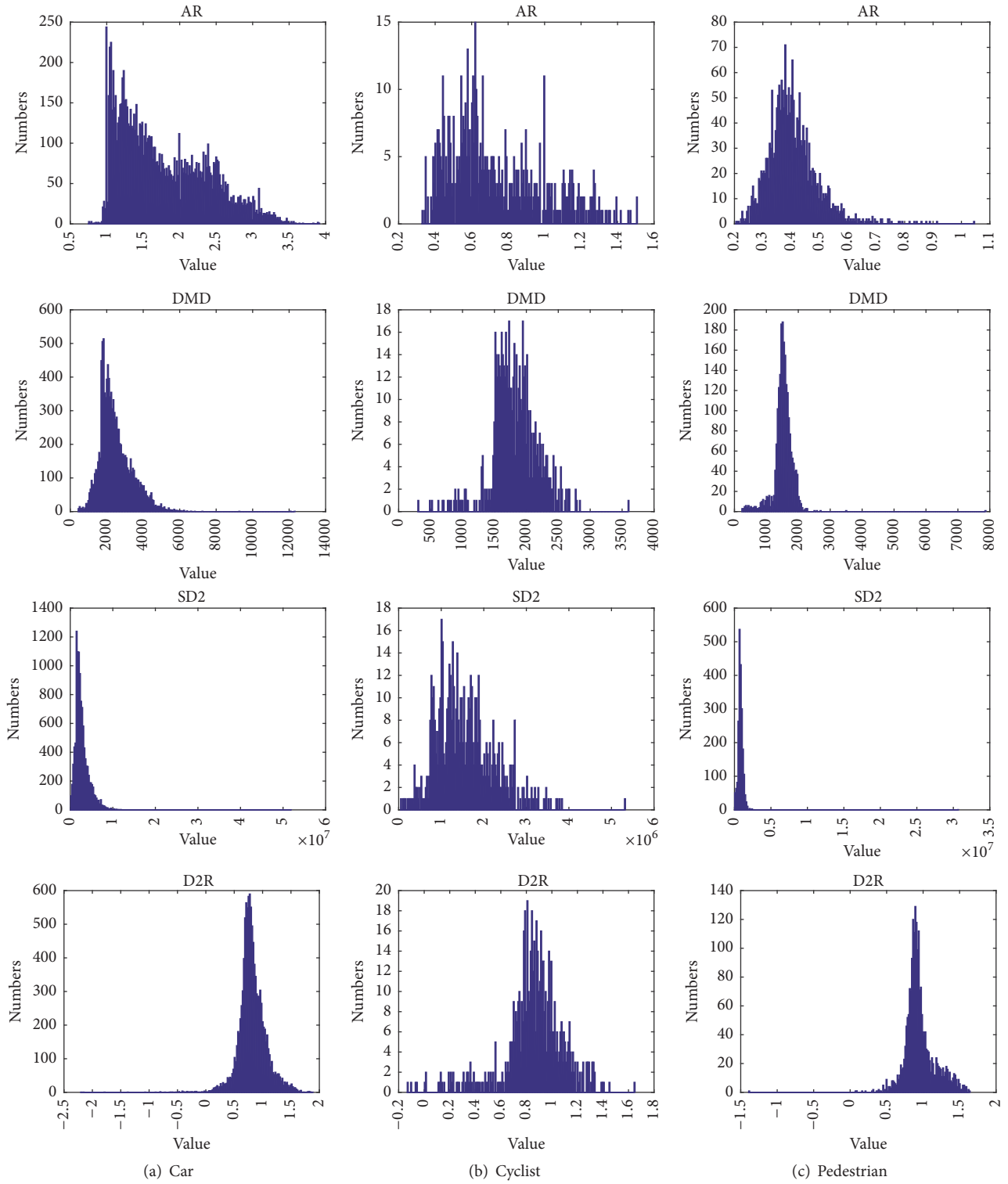


FIGURE 2: Statistic of four object features. For each object class, Car, Cyclist, and Pedestrian, from top to down the features are AR, DMD, SD2, and D2R. We could normalize them to zero mean and unit variance (mean subtraction and division by the standard deviation).

where $C^1 \subseteq C$. This posterior probability constitutes the final proposal score, which is used as the indication of the possibility of a proposal that tends to cover an object.

3.3. Implementation Details. After a large number of positive and negative proposals are sampled, the distribution of their image features (AR, SC, DMD, SD2, and D2R) is demonstrated via the histogram (we sample all the proposals that have an IOU ≥ 0.6 with any ground truth as positive and 600 negative proposals that have IOU < 0.35 for each image). The values of the feature$_i$ V(feature$_i$ | c) are divided into K bins in a range $[V_{\min}, V_{\max}]$. Therefore, the priors p(feature$_i$ | c) are set by relative frequency:

$$p\left(\text{feature}_i \mid c\right) = \frac{N_{\text{Bin}}j}{N}, \quad 1 \leq j \leq K, \tag{9}$$

where $N_{\text{Bin}}j$ is the number of V(feature$_i$ | c) falling into the Bin^j and N is the total number of V(feature$_i$ | c). When any proposal is given, the bin which the value V of the (feature$_i$ | c) falls into is first determined. Then, the individual feature likelihood p(feature$_i$ | c) is roughly equivalent to (9) for each proposal. And the final posterior probability can be calculated according to (8). Noted that (8) allows us to combine any subset C of features, for example, pairs of features $C = \{\text{AR}, \text{SD2}\}$, triplets $C = \{\text{AR}, \text{SD2}, \text{D2R}\}$, or all features $C = \{\text{AR}, \text{SD2}, \text{SC}, \text{DMD}, \text{D2R}\}$. Function (8) can combine any subset rapidly without recomputing the likelihoods.

4. Experiments and Analysis

In this section, we evaluate our method on the challenging KITTI benchmark [9] for all three object classes, which contains 7481 right, 7481 left training images, and 7518 test images. Since the test images do not have any annotations, we split the KITTI training set into train (3,712 images) and validation (3769 images) sets as described in [16]. Bayes model is trained on the train set. All the experiments results are reported on the validation set in three regimes: easy, moderate, and hard, which are defined according to the occlusion and truncation levels of objects.

Following [6], we evaluate the quality of object proposals by using the recall metric. Recall is calculated as the fraction of ground truth objects covered above an IOU threshold. We use curve of the recall versus the number of proposals to depict accuracy at different proposal budgets and recall versus IOU curve to show the variety of recall over different localization precision. In addition, in order to measure the overall accuracy of proposals, we use Area Under the Curve (AUC), which is the area under "recall versus the number of proposals" curve. AUC is a canonical metric which has been shown in [6].

The results of analyzing features and features integration are tested on the hard validation set for all three objects, while the comparison results to the state of the art are on all three object classes and three regimes which use the same metrics depicted in previous section.

4.1. Various Features Integration. We first verify the effectiveness of all the geometric features independently. As our goal is to analyze the performance of each of the features and their combinations which is independent of the baseline method, we only evaluate our method based on Edgeboxes. The results of the baseline method are named SC. As shown in Figure 3, we analyze the baseline and the four proposed geometric features independently to observe the performance of these features. The first row of Figure 3 is the recall versus IOU curve on 500 proposals while the second row is curve of the recall versus the number of proposals on different IOU threshold. For Car, the IOU threshold is 0.7, and it is 0.5 for Cyclist and Pedestrian. We find that all the four proposed features work better than the baseline in which we only use a single feature to generate the proposals. Based on experiments on the three objects we find that D2R is the most useful feature while our proposed feature SD2 is second. DMD has a similar performance to SD2, because they both catch the constancy of object size in real-world. AR is also a useful feature.

Then we combine those geometric features and SC together in a Bayesian framework using different combination to find the best way for fusion of these features. In order to use Bayesian function, the prior probabilities p(obj), p(bg), and p(feature | c) should be first computed. The p(obj) and p(bg) are constant value which are computed in the training stage. The probability p(feature | c) is calculated by using histogram as described in (9). Before we construct the histogram, we normalize them to zero mean and unit variance (mean subtraction and division by the standard deviation). The mean and standard deviation values of each feature are computed on the entire training set.

We combine the five features in a Bayesian framework with all possible combinations. The combinations include 10 ways for any pairs of features, 10 for any triplets, 5 for any four, and 1 for all features together. We have evaluated all the combinations. Since plotting all the combinations is difficult to observe, we only choose 2 top results from pairs of features combinations and triplets of features combinations, 1 from four features combinations, and 1 for all five features. The results are shown in Figure 4. It can be seen that the best performance is obtained by the combination of {AR, D2R, SD2, DMD} and {AR, D2R, SD2}. The results also hint that D2R is the most effective feature, followed by SD2 and DMD, which is consistent with previous observations in Figure 3. We also can find that SD2 and DMD are highly dependent on each other. So we just use SD2 because SD2 is lightly better than DMD. Usually more features make better results. However, it is noteworthy that when combining SC with all other four features the SC does not improve the performance but depresses it. A possible reason is that boxes with larger SC do not mean having higher possibility of containing Car, Cyclist, or Pedestrian. Finally, as shown in Table 1 we summarize the statistics accuracy measures including Area Under the Curve (AUC), the top recall the method can reach (recall), and the number of proposals to achieve recall = 0.75 (M).

4.2. Comparison to the State of the Art. Based on the analysis on the features in previous section, we choose D2R, SD2, and AR as our final choice. As our method can be integrated into any object proposal generation method, we verify its

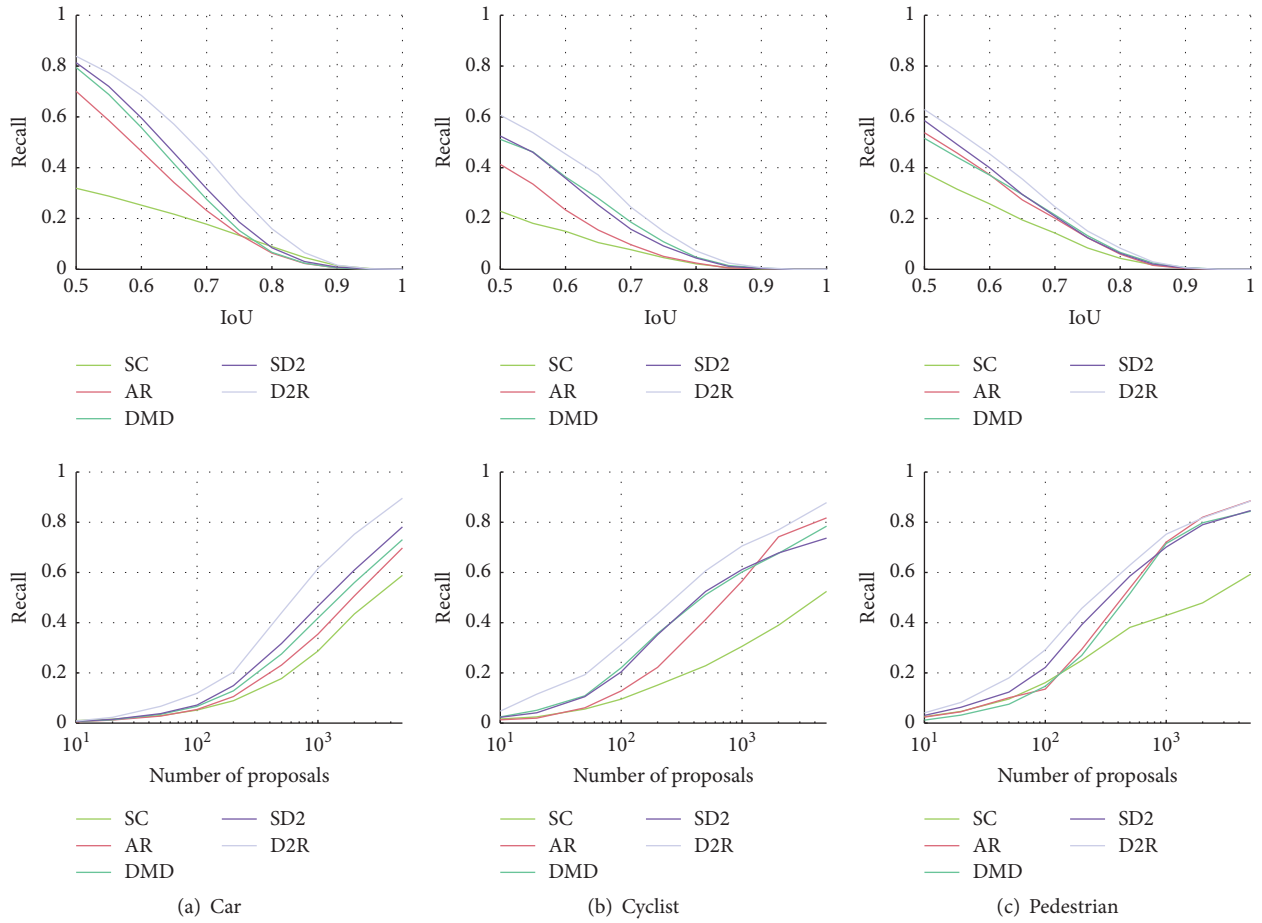


Figure 3: Single feature results: the first row is the recall versus IOU curve on 500 proposals while the second row is curve of the recall versus the number of proposals on different IOU threshold. For Car the IOU threshold is 0.7, and it is 0.5 for Cyclist and Pedestrian. We analyze the original results and the four proposed features independently to observe the usefulness of these features. We find that all the four proposed features work better than the original result when we just use a single feature to generate the proposals. With experiments on the three objects we find that D2R is the most useful feature while our proposed feature SD2 ranks second. DMD have similar performance with SD2, because they both catch the constancy of object size in real-world. AR is also a useful feature.

Table 1: Results on the hard validation sets for all three object classes. AUC is the abbreviation for Area Under the Curve, recall is the maxima recall the method can achieve, and M is the number of proposals when the recall reaches 75%. Inf means the maxima recall cannot reach 75%.

Features		Cars			Cyclist			Pedestrian		
		AUC	Recall (%)	M	AUC	Recall (%)	M	AUC	Recall (%)	M
Single features	AR	0.13	59	Inf	0.14	52	Inf	0.2	89	Inf
	SC	0.15	70	Inf	0.24	82	2209	0.29	89	1226
	DMD	0.17	73	Inf	0.27	78	3735	0.28	84	1337
	SD2	0.19	78	4426	0.27	74	Inf	0.31	85	1463
	D2R	**0.25**	**90**	**1977**	**0.33**	**88**	**1616**	**0.34**	**88**	**986**
Single features	D2R + DMD	0.39	92	682	0.43	89	832	0.46	90	307
	D2R + SD2	0.41	92	509	0.42	88	927	0.46	89	286
	D2R + DMD + AR	0.45	92	413	0.47	89	564	0.53	90	609
	D2R + SD2 + AR	**0.46**	**92**	**352**	**0.47**	**89**	**536**	0.54	89	667
	D2R + DMD + SD2 + AR	0.46	92	392	0.47	89	625	**0.54**	**91**	**129**
	All	0.45	92	423	0.44	89	568	0.52	91	234

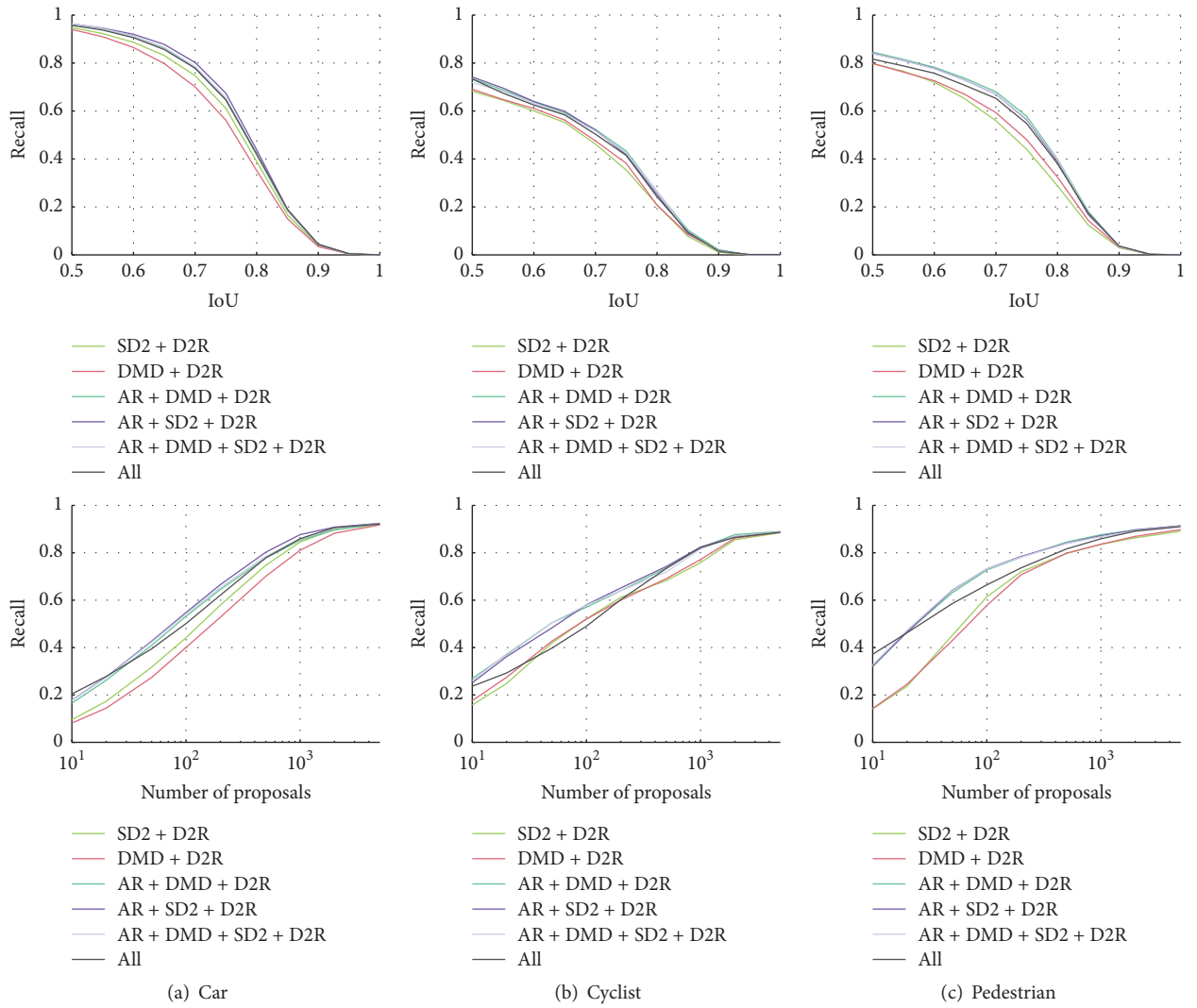


Figure 4: Features combination results: The first row is the recall versus IOU curve on 500 proposals while the second row is curve of the recall versus the number of proposals on different IOU threshold. For Car the IOU threshold is 0.7, and it is 0.5 for Cyclist and Pedestrian.

effectiveness on two representativeness methods: EB (Edge-boxes) and SS (Selective Search). Correspondingly, we name their improved versions Our-EB145 and Our-SS145, where 1 represent AR, 2 represent SC, 3 represent DMD, 4 represent SD2, and 5 represent D2R. Our-EB145 means the results obtained by fusing those three geometric features, AR, SD2, and D2R, with EB in a Bayesian framework. In the paper, we just use Our-EB instead of Our-EB145, the same to Our-SS. We also compare our results to 3DOP because it is the state-of-the-art method that exploits geometric features to generate object proposals.

Figure 5 shows recall versus IOU on 500 proposals and we can see that Our-EB and Our-SS obtain significant improvement compared to the original EB and SS. For Car, our method is better than 3DOP when the IOU is below 0.7, while, with the IOU getting larger, 3DOP obtain better results. This phenomenon also appears in Cyclist. A possible reason is that the original results are good enough when the IOU is high. However, for Pedestrian Our-EB always shows better performance than 3DOP.

Figure 6 shows recall versus the number of candidates. For Car, we can achieve nearly 90% recall when the number of candidates is 1000 for moderate and hard regimes while for easy regimes we only need 200 candidates to get the same results. However, the baseline cannot achieve 90% recall no matter how many candidates are used. For Cyclist and Pedestrian our results show similar improvements over the baselines. Compared to 3DOP our method obtains different degrees of improvements. For example, by using 100 proposals for Pedestrian our method achieves 89%, 80%, and 70% recall for easy, moderate, hard regimes while the 3DOP is around 70%, 60%, and 52%. However when the number of proposals gets larger our method achieves similar result with 3DOP.

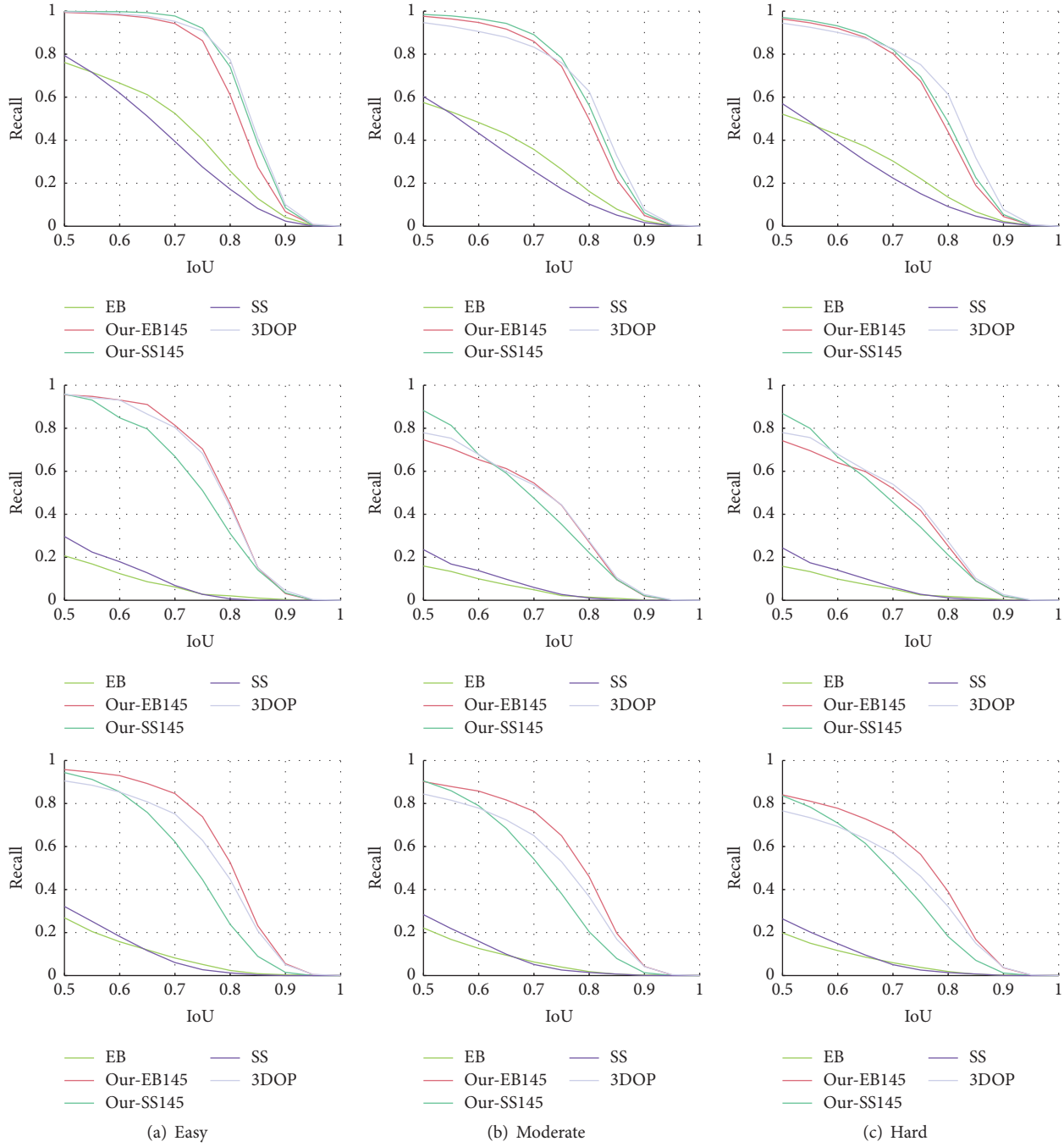


Figure 5: Recall versus IOU for 500 proposals in three regimes. From top to down: Car, Cyclist, and Pedestrian.

4.3. Running Time. Given the depth map, our features can be computed efficiently. Combined with the existing method, our approach can obtain significant improvement with only 0.2 s additional runtime on a single core by MATLAB. Table 2 shows the running time of different proposal methods.

4.4. Object Detection. To evaluate the object detection performance based on our proposal generation method, we apply the state-of-the-art fast R-CNN object detector on the bounding box proposals generated by our method, as 3DOP in [16]. We report results on the validation set of the KITTI benchmark and compare our methods (Our-EB and Our-SS) with those whose bounding box proposals are generated by Selective Search and Edgeboxes. Experiments show that the detection performance can be improved around 18% and 15%, respectively. We also compare our results with that of 3DOP. The results are presented in Table 3. Our approach can achieve comparable or better performance across all three categories.

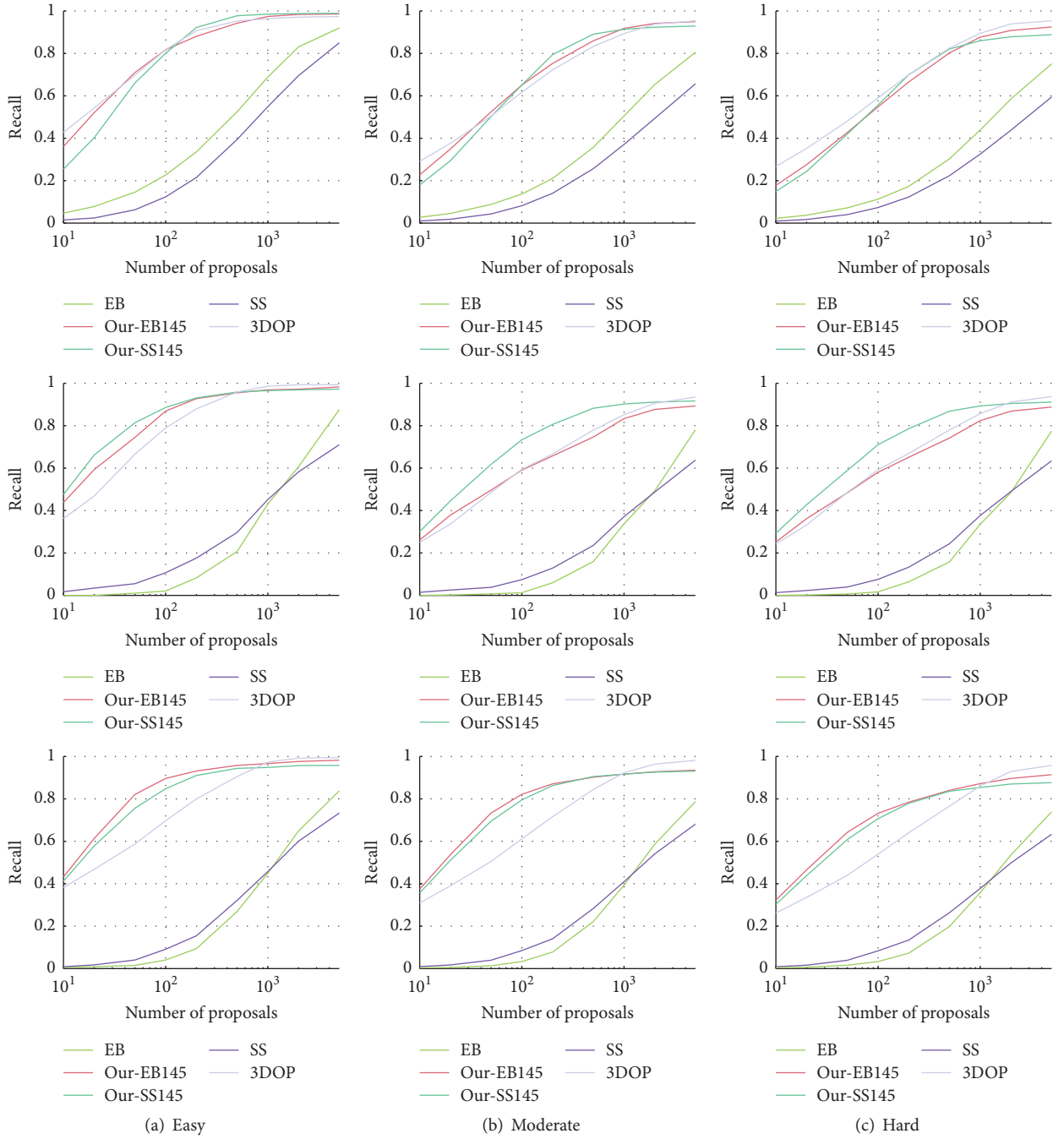


Figure 6: Recall versus number of proposals: the overlap threshold for Car is 0.7, and it is 0.5 for Pedestrian and Cyclist. From top to down: Car, Cyclist, and Pedestrian.

Table 2: Running time of different proposal methods.

Method	Selective Search	Edgeboxes	3DOP	Our-SS	Our-EB
Time (second)	15	1.5	1.2	15.2	1.7

4.5. Visual Results. The visual results of our object detection framework are shown in Figure 7. It would be best to enlarge and view it in color. The odd rows are the ground truth bounding box while the even rows are detection bounding box. Different colors indicate different difficulties. Green means not occluded, yellow means partly occluded, and red means fully occluded. Our approach produces precise detection result even for distant and occluded objects. But it more failed if the object is too distant and fully occluded, since we can not obtain enough depth or appearance information

Figure 7: The visual results of our object detection framework. The odd rows are the ground truth bounding box while the even rows are detection bounding box. Different colors indicate different difficulties. Green means not occluded, yellow means partly occluded, and red means fully occluded. The first four rows are the results of Cars, while the second four rows are the results of Pedestrian and Cyclist.

Table 3: Average Precision (AP) (in %) on the validation set of the KITTI object detection benchmark with 1000 proposals, while, for EB and SS, the number of proposals is 2000.

Metric	Method	Cars			Cyclist			Pedestrian		
		Easy	Moderate	Hard	Easy	Moderate	Hard	Easy	Moderate	Hard
AP	SS [16]	75.91	60.00	50.98	56.23	39.16	38.83	54.06	47.55	40.56
	EB [16]	86.81	70.47	61.16	55.01	37.87	35.80	57.79	49.99	42.19
	3DOP	94.47	87.09	**78.72**	84.65	57.38	55.63	72.47	65	57.24
	Our-SS	**95.36**	**87.84**	78.57	**84.71**	**57.74**	**55.8**	74.23	66.54	57.9
	Our-EB	88.92	87.40	78.43	83.38	57.72	55.69	**74.39**	**66.73**	**58.17**

for object detection. And when a person rides a Cyclist, the ground truth just has an annotation of Cyclist while our method gives two detection results, Cyclist and Pedestrian, as shown in the sixth row. People sitting in a chair are detected; however they are not marked as ground truth in the KITTI datasets.

5. Conclusion

In this paper, we propose several geometric features which are suitable for object proposals in the autonomous driving scene and integrate them with existing object proposal generation methods in a Bayesian framework. We deeply analyze

the effectiveness of each geometric feature and different combinations of features. Experiments on the challenging KITTI benchmark demonstrate that, by integrating these geometric features into existing object proposal methods, we achieve significant improvement on all three object classes. Subsequently we improve the object detection performance. Our future work will focus on integrating geometric features into a totally CNN framework for boosting their performance in the autonomous driving scene.

Conflicts of Interest

The authors declare that they have no competing interests.

References

[1] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 1, pp. 886–893, IEEE, June 2005.

[2] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, June 2008.

[3] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2010.

[4] S. Ravishankar, A. Jain, and A. Mittal, "Multi-stage contour based detection of deformable objects," in *Proceedings of the European Conference on Computer Vision (ECCV '08)*, pp. 483–496, Springer, 2008.

[5] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[6] C. L. Zitnick and P. Dollár, "Edge boxes: locating object proposals from edges," in *Proceedings of the European Conference on Computer Vision*, pp. 391–405, Springer, 2014.

[7] K. E. A. Van De Sande, J. R. R. Uijlings, T. Gevers, and A. W. M. Smeulders, "Segmentation as selective search for object recognition," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 1879–1886, November 2011.

[8] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: a retrospective," *International Journal of Computer Vision*, vol. 111, no. 1, pp. 98–136, 2015.

[9] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the KITTI vision benchmark suite," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 3354–3361, June 2012.

[10] K. Yamaguchi, D. McAllester, and R. Urtasun, "Efficient joint segmentation, occlusion labeling, stereo and flow estimation," in *Computer Vision—ECCV 2014*, Springer, 2014.

[11] P. Arbeláez, J. Pont-Tuset, J. Barron, F. Marques, and J. Malik, "Multiscale combinatorial grouping," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 328–335, IEEE, June 2014.

[12] J. Carreira and C. Sminchisescu, "Constrained parametric min-cuts for automatic object segmentation," *IEEE Transactions on Software Engineering*, vol. 23, no. 3, pp. 3241–3248, 2010.

[13] B. Alexe, T. Deselaers, and V. Ferrari, "What is an object?" in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 73–80, June 2010.

[14] B. Alexe, T. Deselaers, and V. Ferrari, "Measuring the objectness of image windows," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2189–2202, 2012.

[15] M.-M. Cheng, Z. Zhang, W.-Y. Lin, and P. Torr, "BING: Binarized normed gradients for objectness estimation at 300fps," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 3286–3293, June 2014.

[16] X. Chen, K. Kundu, Y. Zhu, H. Ma, S. Fidler, and R. Urtasun, "3d Object proposals using stereo imagery for accurate object class detection," https://arxiv.org/abs/1608.07711.

[17] S. Gupta, R. Girshick, P. Arbeláez, and J. Malik, "Learning rich features from RGB-D images for object detection and segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV '14)*, pp. 345–360, Springer, 2014.

[18] A. Janoch, S. Karayev, Y. Jia et al., "A category-level 3-D object dataset: putting the kinect to work," in *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV '11)*, pp. 1168–1174, November 2011.