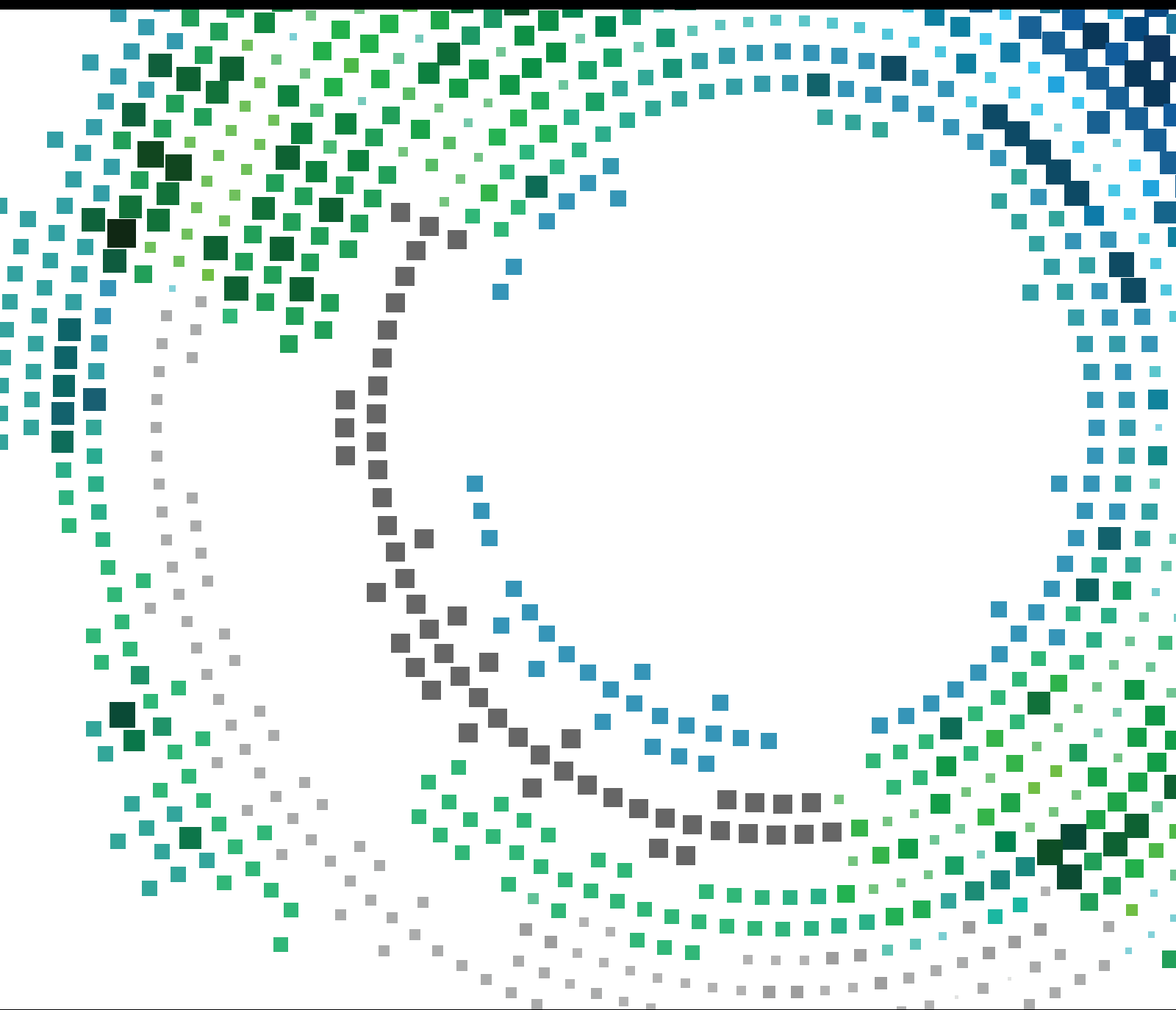
Mobile Information Systems

Enabling Technologies towards 5G Mobile Networks

Lead Guest Editor: Jing Zhao
Guest Editors: Yejun He, Jian Qiao, Ben Niu, and Qiben Yan



Enabling Technologies towards Mobile Networks

Mobile Information Systems

Enabling Technologies towards 5G Mobile Networks

Lead Guest Editor: Jing Zhao
Guest Editors: Yejun He, Jian Qiao, Ben Niu, and Qiben Yan

Editorial Board

Contents

Enabling Technologies towards 5G Mobile Networks
Jing Zhao, Yejun He, Jian Qiao, Ben Niu, and Qiben Yan
Volume 2017, Article ID 7401863, 2 pages

Training Based Channel Estimation for Multitaper GFDM System
Shravan Kumar Bandari, Venkata Mani Vakamulla, and A. Drosopoulos
Volume 2017, Article ID 4747256, 8 pages

5G Development in China: From Policy Strategy to User-Oriented Architecture
Qian Liu, Xiaochuan Shi, Xu Wang, and Jia Li
Volume 2017, Article ID 2358618, 11 pages

Effective Feature Selection for 5G IM Applications Traffic Classification
Muhammad Shafiq, Xiangzhan Yu, Asif Ali Laghari, and Dawei Wang
Volume 2017, Article ID 6805056, 12 pages

An Automata Based Intrusion Detection Method for Internet of Things
Yulong Fu, Zheng Yan, Jin Cao, Ousmane Koné, and Xuefei Cao
Volume 2017, Article ID 1750637, 13 pages

ULMAP: Ultralightweight NFC Mutual Authentication Protocol with Pseudonyms in the Tag for IoT in 5G
Kai Fan, Panfei Song, and Yintang Yang
Volume 2017, Article ID 2349149, 7 pages

Privacy-Preserving Billing Scheme against Free-Riders for Wireless Charging Electric Vehicles
Xingwen Zhao, Jiaping Lin, and Hui Li
Volume 2017, Article ID 1325698, 9 pages

A Window-Based, Server-Assisted P2P Network for VoD Services with QoE Guarantees
Noé Torres-Cruz, Mario E. Rivero-Angeles, Gerardo Rubino, Ricardo Menchaca-Mendez, and Rolando Menchaca-Mendez
Volume 2017, Article ID 2084684, 18 pages

A Novel Secure Transmission Scheme in MIMO Two-Way Relay Channels with Physical Layer Approach
Qiao Liu, Guang Gong, Yong Wang, and Hui Li
Volume 2017, Article ID 7843843, 12 pages

Efficient and Privacy-Aware Power Injection over AMI and Smart Grid Slice in Future 5G Networks
Yinghui Zhang, Jiangfan Zhao, and Dong Zheng
Volume 2017, Article ID 3680671, 11 pages

Hindawi
Mobile Information Systems
Volume 2017, Article ID 7401863, 2 pages
https://doi.org/10.1155/2017/7401863

Editorial
Enabling Technologies towards 5G Mobile Networks

Jing Zhao,[1] Yejun He,[2] Jian Qiao,[3] Ben Niu,[4] and Qiben Yan[5]

[1]*Google Inc., Mountain View, CA, USA*
[2]*Shenzhen University, Shenzhen, China*
[3]*Oracle Inc., Mississauga, ON, Canada*
[4]*Chinese Academy of Sciences, Beijing, China*
[5]*University of Nebraska, Lincoln, NE, USA*

Correspondence should be addressed to Jing Zhao; juz139@cse.psu.edu

Received 9 November 2017; Accepted 9 November 2017; Published 26 November 2017

Future fifth-generation (5G) mobile networks denote the next-generation mobile networks beyond the current 4G mobile networks. The 5G networks are provisioned by the Next Generation Mobile Networks Alliance to provide much higher capacity and support various types of emerging applications with stringent quality of service (QoS) requirements. The objective of this special issue is to solicit the state-of-the-art research contributions that present key and emerging results on 5G-enabling technologies to optimize spectrum efficiency and provide heightened security and privacy.

This special issue offers a selected and articulated overview of the examined topics. It contains nine papers, and the details were listed as follows:

Shafiq et al. proposed a feature selection algorithm named "WMI_ACC," which filters most of the features with WMI metric. It further used a wrapper method to select features for ML classifiers with accuracy (ACC) metric. Their proposed algorithm can achieve 99% flow accuracy results, which is very promising.

Mani et al. investigated channel estimation methods for multitaper GFDM (MGFDM) systems with and without discrete Fourier transform (DFT). The performance of the MMSE estimator is proved to provide a better estimate of the channel response in both cases of with and without DFT.

Zhang et al. proposed an efficient and privacy-aware power injection (EPPI) scheme suitable for advanced metering infrastructure and 5G smart grid network slice. Throughout the EPPI system, both the gateway and the utility company cannot know individual bids, and hence, user privacy is preserved.

Instead of focusing on the security architecture in the upper layer, Liu et al. investigated the secure transmission for a basic channel model in the heterogeneous network, that is, two-way relay channels. Two different approaches have been introduced and proved to be secure under three different adversarial models.

Zhao et al. described a billing scheme against free riders with several cryptographic tools, namely, the encryption scheme, signature scheme, and hash function. The proposed scheme is privacy preserving, so the charging will not disclose the locations and routine routes of each vehicle.

Liu et al. analyzed 5G development and its significant shift from a strategy policy to a user-oriented development showing the coplay of technology and society in China. They proposed a hybrid routing protocol TOHRP and a distributed channel assignment algorithm LBCA in the multichannel environment.

Cruz et al. described a peer-to-peer (P2P) network that is designed to support video on demand (VoD) services. This network is based on a video-file-sharing mechanism that classifies peers according to the window (segment of the file) that they are downloading.

Fan et al. proposed an ultralightweight mutual authentication protocol, named "ULMAP," which uses Bit and XOR operations to complete the mutual authentication and prevent the denial-of-service (DOS) attack. They also used the subkey and subindex number into the key update process to achieve the forward security. Experiment results are also involved to show its lightweight, economical, practical, and resistance to the synchronization attack.

Fu et al. analyzed the intrusion detection requirements of IoT networks and proposed a uniform intrusion detection method for the vast heterogeneous IoT networks based on an automata model. The proposed method can detect and report the possible IoT attacks with three types: Jam-attack, False-attack, and Reply-attack automatically.

The guest editors hope the information provided in this special issue is useful and offers stimulation to the new development of 5G. Finally, we would like to thank the authors for an excellent contribution of their research works as well as we also very warmly acknowledge the reviewers for an excellent contribution of their valuable review results.

Jing Zhao
Yejun He
Jian Qiao
Ben Niu
Qiben Yan

Hindawi
Mobile Information Systems
Volume 2017, Article ID 4747256, 8 pages
https://doi.org/10.1155/2017/4747256

Research Article

Training Based Channel Estimation for Multitaper GFDM System

Shravan Kumar Bandari,[1] Venkata Mani Vakamulla,[1] and A. Drosopoulos[2]

[1]*Electronics & Communication Engineering, National Institute of Technology, Warangal 506004, India*
[2]*Electrical Engineering, TEI of Western Greece, 26334 Patras, Greece*

Correspondence should be addressed to Shravan Kumar Bandari; shravnbandari@gmail.com

Received 27 January 2017; Revised 27 July 2017; Accepted 20 August 2017; Published 1 October 2017

Academic Editor: Jian Qiao

Recent activities in the cellular network world clearly show the need to design new physical layer waveforms in order to meet future wireless requirements. Generalized Frequency Division Multiplexing (GFDM) is one of the leading candidates for 5G and one of its key features is the usage of circular pulse shaping of subcarriers to remove prototype filter transients. Due to the nonorthogonal nature of the conventional GFDM system, inherent interference will affect adversely channel estimation. With Discrete Prolate Spheroidal Sequences (DPSSs) or multitapers as prototype filters an improved orthogonal GFDM system can be developed. In this work, we investigate channel estimation methods for multitaper GFDM (MGFDM) systems with and without Discrete Fourier Transform (DFT). The simulation results are presented using Least Squares (LS) and Minimum Mean Square Error (MMSE) channel estimation (CE) methods. DFT based CE methods provide better estimates of the channel but with an additional computational cost.

1. Introduction

The basic idea behind the multicarrier technique was to divide the total available bandwidth into a number of subbands, allowing the handling of channel effects in an efficient manner. Each generation of mobile communication systems offered many services after which these systems are classified. There are enormous capability advancements in the mobile communications from 1G to the current 4G. The upcoming future generation, 5G, is expected to have an even better coverage area capability, high data connectivity of 1 Gbps, less energy consumption, high security, and better spectral utilization. Moreover, emerging applications such as self-driving cars, real time gaming, and industrial automation control demand a low latency period of the order of less than 1 ms. Even though there is no standard yet for 5G, the industrial and academic research communities are exploring a number of possible implementation options.

A widely used air interface multicarrier (MC) technique in many present wireless standards is Orthogonal Frequency Division Multiplexing (OFDM). Some of the standards include Wireless Local Area Networks (WLAN 802.11), 4G-LTE (Long Term Evolution) Advanced, Digital Video Broadcasting for Terrestrial Television (DVBT), and Digital Audio Broadcasting (DAB) [1]. OFDM gains its popularity due to its efficient implementation with Fast Fourier Transform (FFT)/Inverse Fast Fourier Transform (IFFT) signal processing blocks and its robustness to intersymbol interference (ISI) with simple low complexity equalization. Despite its advantages, in the area of future vision applications, OFDM suffers from a number of drawbacks, such as high Peak to Average Power Ratio (PAPR), addition of cyclic prefix (CP) per symbol that reduces bandwidth efficiency, high out-of-band (OOB) emission due to the use of rectangular filters, and high synchronisation issues due to orthogonality mismatch.

A great deal of effort has been expended in the search for an alternate multicarrier waveform to serve future generation, 5G. Some of the possible contenders with a variety of properties are Filter Bank Multicarrier (FBMC) [2], Universal Filtered Multicarrier (UFMC) [3], and GFDM (Generalized Frequency Division Multiplexing) [4]. More insights into each waveform and their individual properties, such as

spectral efficiency and bit error rates (BER), can be found in the literature and references therein.

GFDM [4] is a flexible multicarrier technique in which the data is divided into a number of subcarriers and subsymbols, using circular pulse shaping for each subcarrier. GFDM covers OFDM as a special case, retaining the advantages of OFDM and dealing with the limitations. For example, GFDM uses one cyclic prefix (CP) per group of subsymbols and prototype filtering with low OOB emission to improve bandwidth utilization. However, due to the nonorthogonal nature of conventional GFDM, channel estimation is not straightforward.

Generally, the channel impulse response can be estimated using training sequences or pilot symbols. Reference [5] introduces two scattered pilot based channel estimation methods for GFDM, Pilot Interference Cancellation (IC) and Transmitter IC. The article gives an insight into interference cancellation at either the transmitter or the receiver of the communication system. Blind estimation of channel block duration, symbol duration, and number of subcarriers are discussed in [6]. Filters with low OOB are desirable for GFDM systems which is quite challenging. In [7], the authors introduce multitapers as prototype filters in order to improve the orthogonality of a conventional GFDM system, naming the result as multitaper GFDM (MGFDM). Another advantage of using multitapers is the low OOB emission which in turn will increase the spectral efficiency. Reference [8] presented the basic framework of how GFDM is used for physical layer services of 5G networks. In this context as multitaper GFDM is the modified version of the conventional GFDM waveform, where we use tapers to pulse-shape each subcarrier instead of circular pulse shaping as in GFDM, MGFDM is related to 5G networks and can be regarded as a potential candidate for 5G networks.

OFDM channel estimation is addressed by [9] based on time domain channel statistics. Channel estimation for various pilot patterns is discussed in [10] and references therein. A great amount of research covering channel estimation for OFDM has been done over the years and much of it can be modified and applied to GFDM systems.

However the subcarrier orthogonality which exists in OFDM is no longer valid for GFDM, as it uses pulse shaping for the subcarriers. There are very few papers available in the literature for GFDM signal channel estimation [5] and to the best of the authors knowledge there is no literature covering channel estimation using Least Squares (LS) and Minimum Mean Square Error (MMSE) and in combination with DFT to multitaper GFDM. These methods are well discussed in the past for various multicarrier signals. However, as multitaper GFDM is the most advanced modulation scheme using a sophisticated block based structure, implementing the above estimation algorithms under severe channel conditions is a difficult task. It should also be noted that any real-life implementation of a wireless system will include a method of channel equalization to improve performance and therefore requires a way to do a running channel estimation to handle effectively changing channel conditions.

Motivated by the above facts, therefore, in this paper, pilot symbol based channel estimation (CE) is carried out for the MGFDM system model. This approach is applicable in general, to any multicarrier technique. Pilot symbols are multiplexed along with the information data symbols. These pilot symbols are known to both transmitter and receiver and a variety of interpolation techniques can be employed to estimate the channel response. As is widely done for CE when training symbols are available, the techniques used are Least Squares (LS) and Minimum Mean Square Error (MMSE).

The article is organized as follows: Section 2 gives a brief introduction about the MGFDM system model. Channel estimators under consideration are discussed in Section 3. Results are discussed in Section 4, followed by conclusions in Section 5.

2. System Model

GFDM is a two-dimensional multicarrier technique in which the data samples are divided among the time and frequency domains [4]. The time and frequency domain plots of root raised cosine (RRC), first DPSS, and PHYDYAS prototype filters are shown in Figure 1. Note, from the plot, that multitapers have deeper sidelobes compared to other pulse shaping filters, which leads to an overall improved system performance [7]. In this section we will give a brief introduction on the MGFDM transceiver system model. Figure 2 shows the system model for training based channel estimation for the MGFDM system.

2.1. MGFDM Signal Model. The binary data generated from the source are grouped together to form P-QAM modulated data, where $P = 2^{\mu}$ and μ is the modulation index. The training symbols or pilot information bits are inserted at an equal spacing between data symbols. The time domain signal after passing through the MGFDM modulator is given by [7]

$$x(n) = \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} d_k(m)\, g_{k,m}(n) \quad n = 0, 1, \ldots, KM-1, \quad (1)$$

where $k = 0, 1, \ldots, K-1$ denotes the frequency index, $m = 0, 1, \ldots, M-1$ denotes the subsymbol index, and n is the sample index. $g_{k,m}(n) = g_m(n)e^{j2\pi kn/K}$ with $g_m(\cdot)$ corresponding to the mth taper of length KM is chosen from the Discrete Prolate Spheroidal Sequences (DPSSs) [11]. The exponential term is the frequency shifted version of the prototype filter $g_m(\cdot)$. $d_k(m)$ is the data symbol on the kth subcarrier and mth subsymbol. It is important to note that MGFDM also covers OFDM with $M = 1$ and $g[n] = 1/\sqrt{K}$ and Single Carrier Frequency Domain Equalization (SC-FDE) with $K = 1$ and $g[n] = \delta[n]$, as special cases.

In matrix notation equation (1) can be written as [4]

$$\mathbf{x} = \mathbf{A}\mathbf{d}, \quad (2)$$

where $\mathbf{A}$ is the transmitter modulation matrix as described in [12]. Matrix $\mathbf{A}$ incorporates all the signal processing steps, while $\mathbf{d}$ is a vector containing the data symbols $d_k(m)$.

Guard interval bits are added at the front side of the symbol in order to prevent intersymbol interference that may possibly affect the MGFDM system. The signal is then

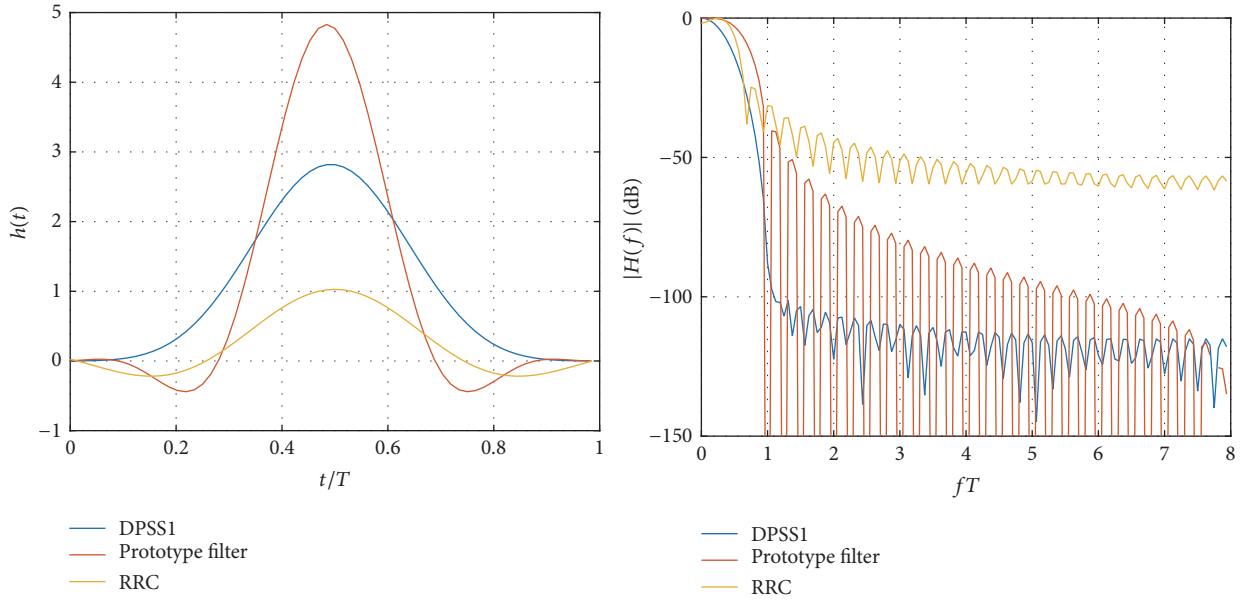


FIGURE 1: Time and frequency domain plots of RRC, DPSS1, and the PHYDYAS prototype filter.

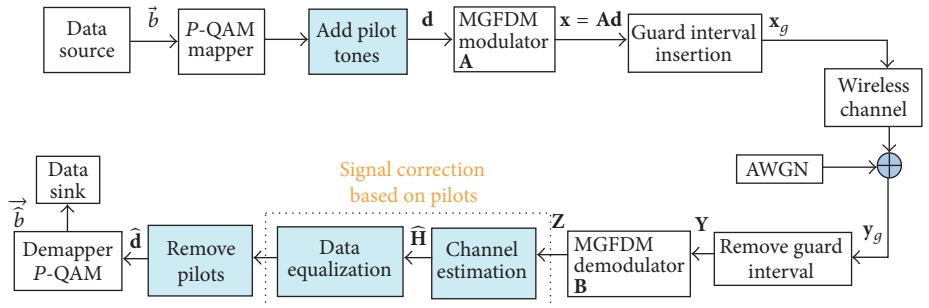


FIGURE 2: Baseband model for pilot based MGFDM system.

transmitted through the frequency selective multipath fading channel $h(n)$. The signal received can be represented as [5]

$$y_g[n] = x_g[n] * h[n] + w[n], \tag{3}$$

where $*$ denotes linear convolution, $h[n]$ represents the channel impulse response, and $w[n]$ is Additive White Gaussian Noise (AWGN).

The guard interval bits are removed from the corrupted signal with AWGN, that is, $y_g[n]$, and are sent to the MGFDM demodulator block. There are several receivers found in the literature to demodulate the data [4]. In this article we have used the zero forcing receiver given by

$$\mathbf{B} = \left(\mathbf{A}^{\mathbf{H}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathbf{H}} \tag{4}$$

$\mathbf{B}$ is known as the MGFDM demodulation matrix, which incorporates the reverse signal processing steps involved at the transmitter side (MGFDM modulator). The demodulated output is written as

$$\mathbf{Z} = \mathbf{BY}, \tag{5}$$

where $\mathbf{Y}$ is the received data after removal of the guard interval. It should be noted that the length of the guard interval is required to be greater than the length of the channel impulse response in order to avoid intersymbol interference. The demodulated output is passed through the signal correction block which is done based on the pilot symbols that are added at the transmitter side. The steps involved in this block are elaborated in Figure 3. In this block, the channel estimation of the pilot tones is first carried out, followed by channel interpolation (linear/spline). Finally, the demodulated data are equalized with the estimated channel coefficients.

3. Channel Estimator

In this section we will briefly discuss the most popularly used channel estimation techniques, Least Squares (LS) and Minimum Mean Square Error (MMSE), applicable to any multicarrier technique. This work was inspired by [13] which was originally proposed for OFDM systems. As MGFDM is the generalization of FDM systems, we carried out the same approach to estimate the channel response.

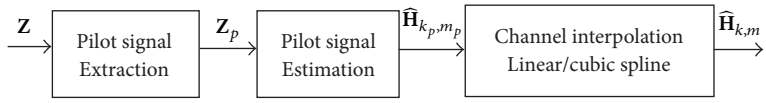


Figure 3: Steps involved in channel estimation.

3.1. LS. Let $\mathbf{H}_p$ be the channel response of the pilot symbols and $\mathbf{Z}_p$ be the received pilot signal vector. Based on the LS approach [13], the channel estimates of the pilot signals are given by

$$\begin{aligned}\widehat{\mathbf{H}}_{p,\mathrm{LS}} &= \mathbf{X}_p^{-1}\mathbf{Z}_p \\ &= \left[\frac{Z_p(0)}{X_p(0)}, \frac{Z_p(1)}{X_p(1)}, \ldots, \frac{Z_p\left(N_p-1\right)}{X_p\left(N_p-1\right)}\right],\end{aligned} \tag{6}$$

where N_p is the number of the pilot symbols used; $\mathbf{X}_p$ and $\mathbf{Z}_p$ are the pilot symbols data at the transmitted side and the pilot information obtained at the receiver side after demodulation, respectively. Let us state that the FFT of (2) is given by

$$X[k,m] = \mathrm{FFT}(\mathbf{x}). \tag{7}$$

Then the MSE of the kth subcarrier and mth subsymbol is given by

$$\mathrm{MSE}_{\mathrm{LS}}[k,m] = \frac{\beta}{\mathrm{SNR}}, \tag{8}$$

where the factor β is given by $\beta = E[|X[k,m]|^2]E[|X[k,m]|^{-2}]$ and is a constant depending on the constellation. For example, β for 16-QAM is 17/9 and $\beta = 1$ for BPSK [14]. SNR $= E[|X[k,m]|^2]/\sigma_w^2$ is the average SNR.

3.2. MMSE. Let us consider the LS solution obtained in the above approach. The MMSE estimate of the channel is given by [10]

$$\widehat{\mathbf{H}}_{\mathrm{mmse}} = \mathbf{R}_{\mathbf{H}_p\mathbf{H}_{p,\mathrm{LS}}}\left(\mathbf{R}_{\mathbf{H}_p\mathbf{H}_p} + \sigma_w^2\left(\mathbf{X}_p\mathbf{X}_p^{\mathbf{H}}\right)^{-1}\right)^{-1}\widehat{\mathbf{H}}_{\mathrm{LS}}, \tag{9}$$

where $\mathbf{R}_{\mathbf{H}_p\mathbf{H}_{p,\mathrm{LS}}} = E[\mathbf{H}_p\mathbf{H}_{p,\mathrm{LS}}^{\mathbf{H}}]$ is the cross correlation matrix between $\mathbf{H}_p$ and $\mathbf{H}_{p,\mathrm{LS}}$. $(\cdot)^{\mathbf{H}}$ is the Hermitian operator of a matrix, $\widehat{\mathbf{H}}_{\mathrm{LS}}$ is the least square channel estimation, and σ_w^2 is the variance of the noise.

3.3. Channel Interpolation. To estimate the complete response of the channel, we make use of the interpolation concept. In this, the samples of the channel response are interpolated according to the estimated channel response of the pilot tones. This kind of work was considered for OFDM systems [13] and is shown to be efficient. In order to increase the overall system performance of such an advanced multicarrier technique (MGFDM), the complete channel state/coefficients should be estimated more accurately. Hence, information about the channel at both pilot and nonpilot locations is essentially what is required to enhance performance. The interpolation method is an efficient way to estimate the nonpilot (data location) channel characteristics by making use of the channel coefficients at the pilots. In this article we examined two interpolation techniques, namely, piecewise linear interpolation and cubic spline interpolation. For more details on these techniques, interested readers can refer to [13].

3.4. DFT Based Channel Estimation. In order to improve the performance of LS and MMSE, a DFT based channel estimation is considered. The performance of such DFT based channel estimators is analyzed for OFDM systems and [14, 15] provide more insights into the method. The scope of the paper is to investigate such estimators for MGFDM system using Mean Square Error (MSE) as a measuring parameter. Even though all existing methods of channel estimation techniques are directly applicable to GFDM, if the pilot symbols require a simpler and different orthogonal demodulation approach, this should be taken into account at the receiver; otherwise it might be questionable to detect pilot symbols due to the nonorthogonality nature of GFDM and pulse shaping property in MGFDM. Our implementation uses the same demodulation approach to both pilot and data symbols and we do not observe the above problem. A simple block diagram of DFT based CE is shown in Figure 4. In this, the effect of noise outside the channel delay (L) is eliminated. Let $\widehat{\mathbf{H}}_{k,m}$ denote the estimation of the channel response at the kth subcarrier and mth subsymbol, calculated by using either the LS or the MMSE method. This is a two-step process. First, we take the IDFT of the $\widehat{\mathbf{H}}_{k,m}$ and ignore the coefficients outside the maximum channel delay. Next, the obtained time domain channel estimation coefficients are converted back to frequency domain by taking an N point DFT. This can be mathematically represented as follows:

$$\begin{aligned}\mathrm{IDFT}\left[\widehat{\mathbf{H}}_{k,m}[\mathbf{n}]\right] &= \widehat{h}_{k,m}[n] = \widehat{h}_{\mathrm{DFT},k,m}[n], \\ \widehat{\mathbf{H}}_{\mathrm{DFT},k,m}[\mathbf{n}] &= \mathrm{DFT}\left[\widehat{h}_{\mathrm{DFT},k,m}[n]\right],\end{aligned} \tag{10}$$

where $n = 0, 1, \ldots, KM-1$.

The individual MSE of the DFT based channel estimation is given by

$$\mathrm{MSE}_{\mathrm{DFT}}[k,m] = \frac{L}{N}\frac{\beta}{\mathrm{SNR}}, \tag{11}$$

where N is a constant and L depends on the channel environment. According to [16], $L/N = T_G/T_s$ in IEEE 802.11 and IEEE 802.16 standards is selected from $\{1/32, 1/16, 1/8, 1/4\}$.

The simulated MSE is given by the average of the error matrix:

$$\mathrm{MSE} = \frac{1}{KM}E\left\{\left(\widehat{\mathbf{H}}-\mathbf{H}\right)\left(\widehat{\mathbf{H}}-\mathbf{H}\right)^{T}\right\}, \tag{12}$$

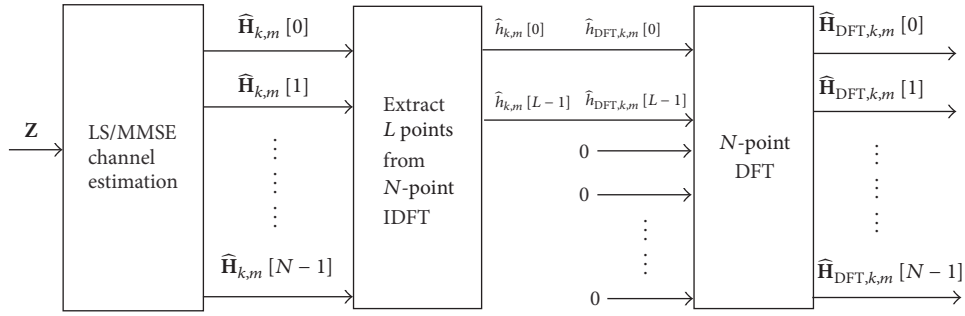


Figure 4: Block diagram of DFT based channel estimation.

Table 1: Simulation parameters.

Description	Value
Number of subcarriers K, subsymbols M	32, 4
Total number of samples, $N = K * M$	128
Length of CP, N_{cp}	8
Modulation order, μ	4
Length of pilot symbols, N_p	16
Length of the channel, N_{ch}	2

where $E(\cdot)$ is the expectation operator, $(\cdot)^T$ is the transpose of the matrix, and $\widehat{\mathbf{H}}$ is the channel estimate.

4. Results and Discussions

In this section, we will discuss the performance of the channel estimation methods that are investigated in this article for the MGFDM system. We have considered a 2-tap random channel model. In Matlab we use the interp1() command for interpolation, as this command is based on piecewise linear and cubic interpolation. The simulation parameters are considered are as shown in Table 1. We have used a total of N samples in which N_p number of pilots are inserted in between the data points with a pilot spacing of 8.

Figure 5 shows the channel estimates that are obtained by using various types of channel estimation methods and are compared with the true channel. We have assumed that the guard interval is greater than the channel delay L. From the figure it is clear that the MMSE estimation shows an improvement compared to that of the LS approach. In the LS estimate we have used linear interpolation and cubic spline interpolation methods to estimate the complete channel response.

A DFT based channel estimation method using LS and MMSE is shown in Figure 6. A good match between the true channel and the estimated channel response is observed. Comparing Figures 5 and 6, an improvement is observed in estimating the channel using DFT based channel estimation.

Figure 7 illustrates the MGFDM system performance in terms of MSE for different SNR values. The MSE of the proposed three techniques with and without the use of DFT based channel estimation is provided. To strengthen the simulated results, theoretical plots for the expressions

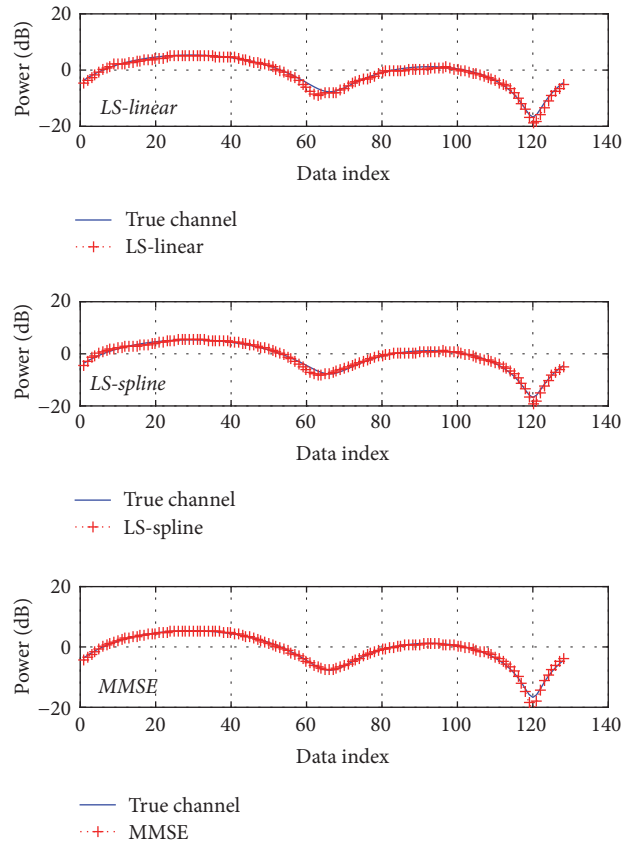


Figure 5: Channel estimation without DFT.

provided in the above sections under LS and DFT based channel estimation methods are also given. It should be noted that both simulated and theoretical values are in good agreement and match almost at every SNR. MMSE results in a better performance when compared to the LS approach, without DFT estimation. To enhance the performance, DFT based CE is proposed in this article and from the figure we can infer that there is a reduction in MSE significantly. The results suggest the benefit of using DFT based CE methods but this comes at a cost of additional computational complexity.

The received signal constellation of a GFDM system with 16-QAM before and after channel compensation is shown in Figure 8. We can infer from the figure that the data

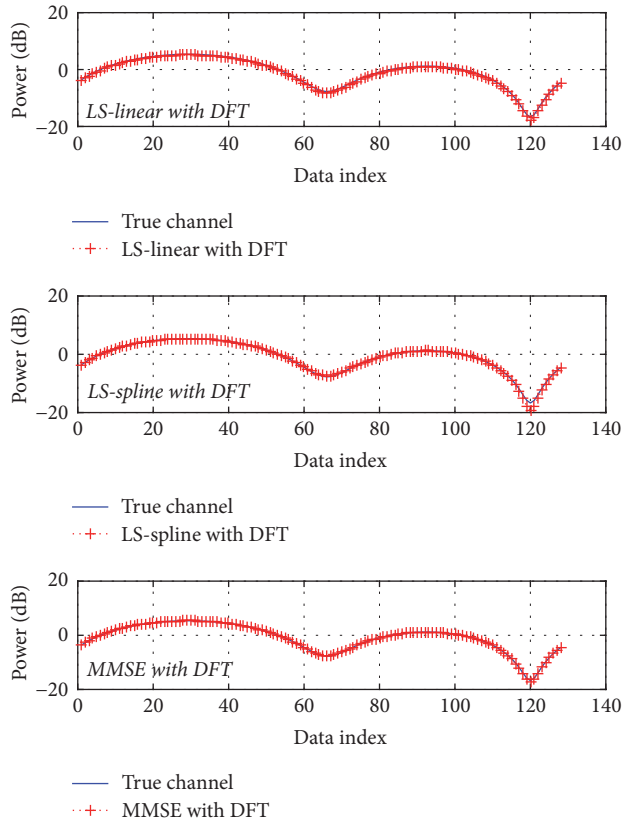


Figure 6: DFT based channel estimation.

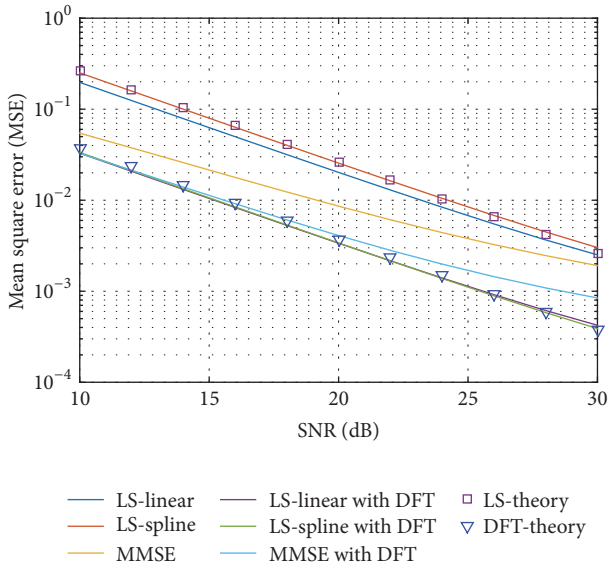


Figure 7: Performance evaluation of MGFDM with different channel estimation techniques.

symbols are properly mapped after channel compensation using MMSE channel estimation with DFT technique. From these results it is obvious that DFT based channel estimation methods reduce the MSE at a faster rate.

Lastly, the power spectral densities (PSD) for the various multicarrier techniques are displayed in Figure 9. The simulations are carried out in Matlab and the parameters that are used are given in Table 1 with a sampling frequency of 20 MHz using the *pwelch* spectral estimation method. From the figure we can observe that the PSD of GFDM has a better spectral efficiency property compared to an OFDM system. Interestingly, the spectra of GFDM using multitapers exhibit the same spectral properties as that of a conventional GFDM system. Thus with the same spectral properties, deeper sidelobes of the pulse shaping filter can be achieved, as discussed earlier.

5. Conclusion

In this paper, we have presented channel estimation techniques, namely, LS (with linear and spline interpolation) and MMSE with and without the DFT method.

The channel estimation at the pilot tones and the interpolation of the channel at remaining points are presented. From the simulation results, among the investigated methods, when compared to the LS estimator, the performance of the MMSE estimator is observed to provide a better estimate of the channel response in both cases with and without DFT.

There is an additional computational cost that has to be paid in using DFT based channel estimation. Moreover as we increase the number of data sample points, the complexity of MMSE increases exponentially. Also, the overall efficiency

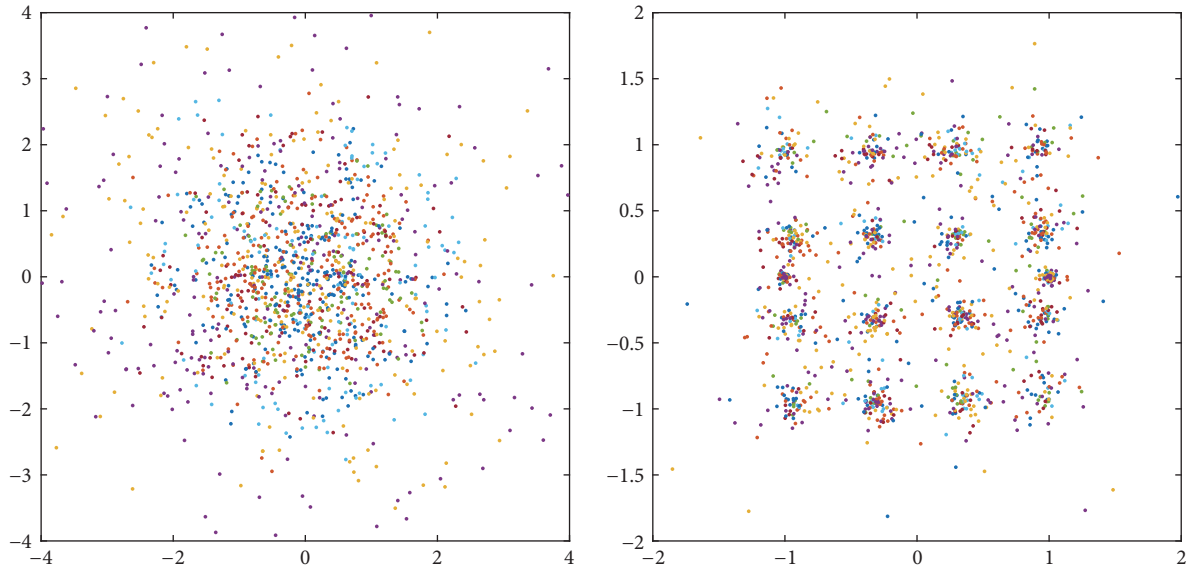


Figure 8: Received signal constellation before and after channel compensation.

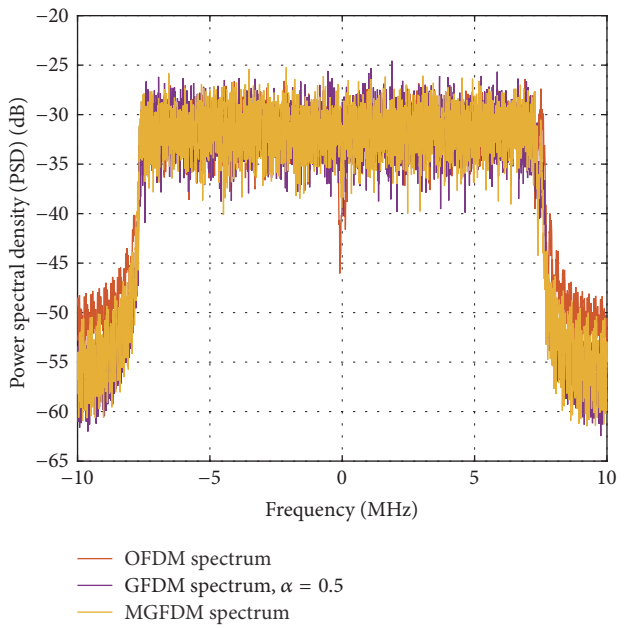


Figure 9: Spectrum comparison of OFDM, conventional GFDM with root raised cosine filter of roll-off factor 0.5, and MGFDM systems.

of the system may be reduced due to the addition of the overhead pilot symbols to that of data symbols. Therefore a trade-off exists between better channel estimation system performance and additional complexity of the overall system. This paper addresses the preliminary investigation of channel estimation methods on the novel multicarrier MGFDM system technique. It will be of great interest to implement some of the advanced channel estimation techniques in usage today and propose new such techniques for MGFDM, which is where our future work focuses.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] A. R. Bahai, B. R. Saltzberg, and M. Ergen, *Multi-Carrier Digital Communications: Theory and Applications of OFDM*, Springer Science & Business Media, Berlin, Germany, 2004.

[2] B. Farhang-Boroujeny, "OFDM versus filter bank multicarrier," *IEEE Signal Processing Magazine*, vol. 28, no. 3, pp. 92–112, 2011.

[3] V. Vakilian, T. Wild, F. Schaich, S. Ten Brink, and J.-F. Frigon, "Universal-filtered multi-carrier technique for wireless systems beyond LTE," in *Proceedings of the 2013 IEEE Globecom Workshops, GC Wkshps 2013*, pp. 223–228, Atlanta, Ga, USA, December 2013.

[4] N. Michailow, M. Matthe, I. S. Gaspar et al., "Generalized frequency division multiplexing for 5th generation cellular networks," *IEEE Transactions on Communications*, vol. 62, no. 9, pp. 3045–3061, 2014.

[5] U. Vilaipornsawai and M. Jia, "Scattered-pilot channel estimation for GFDM," in *Proceedings of the 2014 IEEE Wireless Communications and Networking Conference, WCNC 2014*, pp. 1053–1058, Istanbul, Turkey, April 2014.

[6] L. Chang, G. Y. Li, J. Li, and R. Li, "Blind parameter estimation of GFDM signals over frequency-selective fading channels," *IEEE Transactions on Communications*, vol. 64, no. 3, pp. 1120–1131, 2016.

[7] S. K. Bandari, V. V. Mani, and A. Drosopoulos, "Multi-Taper implementation of GFDM," in *Proceedings of the 2016 IEEE Wireless Communications and Networking Conference, WCNC 2016*, qat, April 2016.

[8] M. Danneberg, N. Michailow, I. Gaspar et al., "Implementation of a 2 by 2 MIMO-GFDM transceiver for robust 5G networks," in *Proceedings of the 2015 International Symposium on Wireless Communication Systems (ISWCS)*, pp. 236–240, Brussels, Belgium, August 2015.

[9] J.-J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, "On channel estimation in OFDM systems," in *Proceedings of the 1995 IEEE 45th Vehicular Technology Conference. Part 2 (of 2)*, pp. 815–819, July 1995.

[10] F. Tufvesson and T. Maseng, "Pilot assisted channel estimation for OFDM in mobile cellular systems," in *Proceedings of the 1997 47th IEEE Vehicular Technology Conference. Part 1 (of 3)*, pp. 1639–1643, May 1997.

[11] D. Slepian, "Prolate spheroidal wave functions, fourier analysis, and uncertainty—V: the discrete case," *Bell System Technical Journal*, vol. 57, no. 5, pp. 1371–1430, 1978.

[12] A. Farhang, N. Marchetti, and L. E. Doyle, "Low-complexity modem design for GFDM," *IEEE Transactions on Signal Processing*, vol. 64, no. 6, pp. 1507–1518, 2016.

[13] M.-H. Hsieh and C.-H. Wei, "Channel estimation for OFDM systems based on comb-type pilot arrangement in frequency selective fading channels," *IEEE Transactions on Consumer Electronics*, vol. 44, no. 1, pp. 217–225, 1998.

[14] O. Edfors, M. Sandell, J.-J. Van De Beek, S. K. Wilson, and P. O. Börjesson, "Analysis of DFT-based channel estimators for OFDM," *Wireless Personal Communications*, vol. 12, no. 1, pp. 55–70, 2000.

[15] M. J. F.-G. García, J. M. Páez-Borrallo, and S. Zazo, "DFT-based channel estimation in 2D-pilot-symbol-aided OFDM wireless systems," *IEEE Vehicular Technology Conference*, vol. 2, no. 53, pp. 810–814, 2001.

[16] Y. Shen and E. Martinez, "Channel estimation in OFDM systems," 2006.

Hindawi
Mobile Information Systems
Volume 2017, Article ID 2358618, 11 pages
https://doi.org/10.1155/2017/2358618

Research Article

5G Development in China: From Policy Strategy to User-Oriented Architecture

Qian Liu,[1] Xiaochuan Shi,[2] Xu Wang,[2] and Jia Li[1]

[1]*Jinan University, Guangzhou 510632, China*
[2]*International School of Software, Wuhan University, Wuhan 430079, China*

Correspondence should be addressed to Xiaochuan Shi; shixiaochuan@whu.edu.cn

Received 28 October 2016; Revised 7 January 2017; Accepted 24 January 2017; Published 15 August 2017

Academic Editor: Jing Zhao

5G encompasses the development of various key wireless communication technology standards. The development entails both technological advancement and social interaction. This paper analyzes 5G development and its significant shift from a strategy policy to a user-oriented development showing the coplay of technology and society in China. Based on this theory, relevant proposals are recommended for future technical development and industrial regulation. Also, to increase the throughput and spectrum efficiency, channel assignment and load balance are considered. A hybrid routing protocol TOHRP (tree-based and on-demand hybrid routing protocol) and a distributed channel assignment algorithm LBCA (Load Based Channel Assignment) in the multichannel environment are proposed and they solve the traditional problem of (1) waste of spectrum and (2) self-interference. The computer-simulated experiment shows that the algorithm improves the performance.

1. Introduction

5G encompasses the development of various key wireless communication technology standards. The development entails both technological advancement and social interaction. Numerous scholars have worked in this field, for developing improved technical standards through an understanding of the interaction between technology and society. Actor network theory is one of the important theories in this regard.

Scientists and scholars are working on new approaches for energy-efficient and cost-effective 5G technologies, for example, using reduced radio spectrum. They also seek to commercialize 5G technologies in 5 years. In China, from 1G to 5G, the development of each generation is highly competitive. 2G mainly uses Time Division Multiple Access (TDMA), providing digital voice and low-speed data services. 3G provides high speed voice and data services using for instance Code Division Multiple Access (CDMA). 4G has a peak data rate from 100 Mbps to 1 Gbps, using Orthogonal Frequency Division Multiple Access (OFDMA) and supports various mobile broadband data services. Technical research on 5G focuses on system architecture design, millimeter wave, new air interface, light MAC (Media Access Control), RRM (Radio Resource Management), multicell joint processing, large-scale massive multiple-input multiple-output (MIMO), antennas and propagation, intelligent devices, Machine-to-Machine (M2M) communication, and so forth.

To provide theoretical support for 5G development based on the actor network theory, this paper offers an extensive review of current literature, discusses the relevant topics, and provides an overview of China's Information and Communications Technology (ICT) development. It highlights the importance of the close interaction between society and technology. Based on the analysis for preliminary 5G standardization both in China and abroad, participating actors in China and their relationship changes have shown a shift from a policy strategy to user orientation. The relevant key issue is analyzed, such as spectrum efficiency. Also, to increase the throughput and spectrum efficiency, channel assignment and load balance are considered. A hybrid routing protocol TOHRP (tree-based and on-demand hybrid routing protocol) and a distributed channel assignment algorithm LBCA (Load Based Channel Assignment) in the multichannel environment are proposed and they solve the traditional problem of (1) waste of spectrum and (2) self-interference.

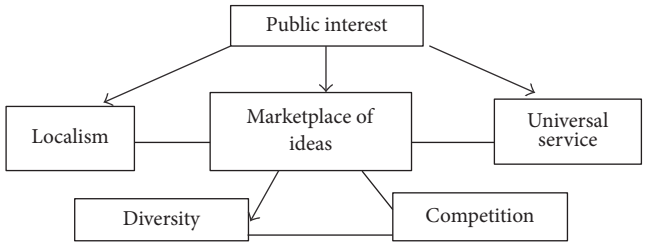


Figure 1: Foundation principles of communications policy [1].

2. Related Work

2.1. The Foundation Principles of Telecommunications Policy. Government departments and international organizations have played policy and standard makers' role in the recent decades. Scholars have tried to define the influence policymakers could have on communication policy. Napoli (2001) stated 6 principles of communications policy in his book, *Foundations of Communications Policy Principles and Process in the Regulation of Electronic Media*, which is also applicable to telecommunications policy. It contains public interest, marketplace of ideas, localism, universal service, diversity, and competition. In his book, the 6 parts are the central motivators or justification for policy decisions. Each of them has endured for decades as an important concept in communications policy making but has remained in contested territory. They are prominent, enduring, and also controversial [1].

Napoli (2001) indicated that "public interest" was the most important concept, and three policy principles would extend from it: "marketplace of ideas", "localism," and "universal service." "Diversity" and "competition" were similar but they differed from each other. Universal service is what the Chinese government had been keen on for decades. An alternative for public interest is the idea of collective users in the 5G era as shown in Figure 1. These principles will be employed in this paper for analyzing the development of ICT and of 5G.

2.2. 5G Academic Research in a Preliminary State. 5G related researches have covered varied topics. The very first book on 5G was published in June 2016. This book *5G Mobile and Wireless Communications Technology* covered different topics varying from special cases to spectrum, describing 5G system architecture [2]. Some most cited papers have shown a variety of perspectives: general concepts, basic infrastructure, and disruptive technology directions [3]. For a more efficient use of the spectrum in the millimeter wave bands, 5G cellular communications were studied with outdoor and indoor penetration tests and a beam-forming prototype was conducted [4]. Emphasizing millimeter wave mobile communications, scholars from USA collected data from experiments around the University of Texas at Austin (38 GHz) and New York University (28 GHz) and proved that indoor networks would be isolated from outdoor networks and a large number of devices should be installed. They found that consistent coverage could be reached when the base stations had a cell-radius of 200 m and dense urban environment could cause path losses. Reflection coefficients and penetration losses for outdoor materials were significantly higher and larger [5].

On the Web of Science platform, there were a total number of 2,577 5G related academic papers covering computer science, engineering, and telecommunications: engineering electrical and electronic field (2,019), telecommunications (1,683), computer science information systems (520), computer science hardware architecture (375), and computer science theory methods (172), respectively. By using the CiteSpace software the 5G key citation situation is studied. 5G academic papers exhibit scattered citation relations and some papers are cited within a small group. Certain authors have more reputation, such as Andrews JG and Rappaport TS. However, for 3GPP, Bhushan N and Demestichas P are different, although they are cited frequently, but they are normally cited along together. This implies a preliminary state and a diversity of research directions.

2.3. Research around the World. The distribution of relevant research institutions around the world includes some of the most prestigious ones, for example, NASA, Machine-to-Machine Intelligence (M2Mi) Corporation, and South Korean IbjngT R&D program focusing on 5G technology researches back in 2008.

Research centers of universities also participated. NYU WIRELESS center from New York University and UK's University of Surrey had pioneering work in 5G technology.

These academic research centers have also collaborated with key industry partners, including Telefonica, Vodafone, regional SMEs, EM3, Huawei, Samsung, Fujitsu Laboratories Europe, Rohde & Schwarz, and Aircom International [6].

"Mobile and wireless communications Enablers for the Twenty-twenty Information Society" (METIS) project started for 5G standardization and definition in 2012, trying to build consensus worldwide. Since then, many projects have emerged out with different focuses: the iJOIN EU project launched for "small cell" technology, ITU-R Working Party 5D (WP 5D) started for a better understanding of future technical issues, EU research project focusing on a CROWD ubiquitous, ultrahigh bandwidth "5G" infrastructure, the TIGRE5-CM (integrated technologies for management and operation of 5G networks) project for future mobile network architecture design, and METIS-II project on 5G radio access network.

Key stakeholders in the industry have made heavy research investments into 5G technology R&D, including the following: Samsung Electronics, Huawei, NTT DoCoMo, Alcatel Lucent, Ericsson, Fujitsu, NEC, Nokia, Verizon, Orange (French Operator), and Google. Moreover, NTT (Nippon Telegraph and Telephone) is operating the world's first 5G networks. NTT DoCoMo and Ericsson pioneered on the 5G outdoor trials, with a cumulative 20 Gbps, and Samsung and Verizon joined late in February 2016.

3. History of China's Telecommunications from 3G to 5G

3.1. Histories of Telecommunication Development in China. By looking back at the history, we can visualize the trend and tendency, so as to generate new knowledge for future development. 5G technology trend can be studied through the following six phases:

3.1.1. 1949–1993: A Tough Beginning—Planned Economy and Regulation from the Central Government. China was founded in the year 1949, October 1. And the Ministry of Posts and Telecommunications (MTP) was established right after that on November 1 in the same year. The newly founded government did not plan for the telecommunications development. Thus, the infrastructure building of telecommunications was given the back seat. With limited poor financial resources, the development of telecommunications was slow. China's national economy got back on track at the time of the first Five-Year Plan, from 1953 to 1957. With 3 years of development after the second Five-Year Plan, covering 1958 to 1962, the most important long-distance telephone building in Beijing exhausted the fund and had to be put off.

In the third Five-Year Plan of 1966–1970 and the fourth Five-Year Plan of 1971–1975, relatively large investments were given to the telecommunications and post industry. Unfortunately, the cultural revolution took place.

In the Fifth Five-Year Plan of 1976–1980 and with the reform of the economics, dramatic changes took place in the economic structure. The telecommunications industry experienced rapid growth. In Paul's (1997) book on telecommunications and development in China, he emphasized that this period "had a tremendous impact on future 5-year plans and on the development of the telecommunications in China" [7].

3.1.2. 1993-1994: Telecommunications Liberalization—The Rise of Operators. The development of private networks and the introduction of China Unicom in 1994 and of Ji Tong in 1993 marked the milestone for the liberalization of China's telecommunications industry. Foreign investments increased the telecommunications competition.

China Unicom was established in July, 1994. It was a joint venture with stakeholders from MEI (Ministry of Electronic Industry), the MOR (Ministry of Railway), the MEP (Ministry of Electrical Power), and 13 other corporations. The origination of China Unicom was a milestone in the Chinese government's telecommunications policy, because it stopped the monopoly model that Xu and Douglas (2002) claimed [8].

Ji Tong was established as a corporation in June 1993, with one aim of seeking joint ventures with overseas companies. Stakeholders include China International Trust and Investment Corporation and 30 other state-owned enterprises and research institutes.

About the emergence of Lian Tong and Ji Tong, Mueller and Tan (1997) indicated that it mixed ministerial politics and business.

> In some ways they are perfect embodiment of China's reform process, and their policy rationale incoherently mixes socialist industrial policy with market competition, they are supposed to be a national, state-centered initiative, but any real energy they have comes from local activities that hardly seem related to the purpose of the central government. Where this process will be leading, is almost impossible to predict. (Mueller and Tan, 1997, p. 63-64)

3.1.3. 1995–1998: The Internal Regroup of MPT—The Transfer of Power from the Central Government. China started its journey of separating the government from enterprise, in the year of 1995. MPT formally separated government functions from enterprise management. It was a huge turning point that the central government officially transferred some of its management power to the public. Thus, the administration power turned out to be a supervision power. In 1998, the post and telecom entities got separated. Since then, China Telecom has focused on telecommunications.

3.1.4. 1998–2002: Entering WTO and Restructuring of China Telecom—Open Market Influence. The entry into WTO in December 1, 2001, has provided significant opportunities for China's telecommunications development. Competition in an open market largely contributes to the improvement of the national services.

In 2000, after the separation of paging, satellite, and mobile services, the services in China Telecom were divided into four parts, thus ending the initial monopoly. The four parts were run by different companies: China Unicom was responsible for paging services, China Telecom was responsible for fixed lines, China Mobile and China Satellite were responsible for mobile and satellite services, respectively. In 2001, China Telecom was further divided into the south part and north part, to optimize the competition inside the industry. In May 2002, China Telecom was formed. With the development, upstream and downstream manufacturers got strengthened.

3.1.5. 2003–2013: 3G to 4G—New Reform Was under Deliberation in MII and Vendors and Suppliers Become More Influential. China's State-owned Assets Supervision and Administration Commission (SASAC) was established in March 2003. It accelerated the change of management and development of the state-owned China Telecom. Since then, ownership, manufacturing rights and management were separated from each other. The development of China's telecommunications was thus better regulated.

3.1.6. 2013–Now: 5G Technology Standardization, Preliminary Stage—User Oriented. From 1G to 5G, the telecommunication development history has shown key components of the 5G research group. The IMT-2020 (5G) established in 2013, Ministry of Industry and Information Technology, National Development and Reform Commission and Ministry of Science and Technology have exhibited great influence on

the development of 5G technical standards in the global context. 5G as a transformative technology calls for many universities, operators, research institutes and manufacturers to participate. Telecommunication is driven by not only government agencies, the operators or manufacturers, but also by users. For example, customers request for very high speed and uninterrupted usage scenarios. Drastic challenges and huge changes will occur to reshape the stakeholders' weight in the future.

3.2. Existing Players: Actor Analysis. After analyzing the 6 phases in the history of telecommunication in China, we could conclude that there has been an obvious shift from planned economy to free market, from policy strategy to customer-oriented strategy. In the actor network theory, these important stakeholders are constantly distributing interests as well as power, in order to form a stable and sustainable environment for 5G growth.

Actor network theory belongs to the category of scientific philosophy [9]. The main scholars of this theory are Latour and Laur and Caron [10]. The actor network theory provides a perspective of the social construction for the interpretation of technical standardization [10]. It is a process-oriented, description-relational theory (Law, 1992) [11] that could be used to describe the process of social construction of science and technology, with different actors being influencing factors. In order to classify these stakeholders, the important concept in the actor network is quoted as "punctualization," which Callon (1991) [12] puts forward, arguing about "the process of punctualization thus converting an entire network into a single point or node in another network" and that "everything is an actor as well as a network—it simply depends on the perspective."

Wireless industry, as an important part of industrial ecology, was divided into market systems, regulatory systems, and innovation systems in academic literature [13]. This paper is based on the classification, forming three systems: the regulatory system, the innovation system, and the market system.

3.2.1. Innovation System. Each system is equivalent to an actor node, which is equivalent to a black box, which encapsulates more components, connected to the network. The subcomponents of the innovation system include regulation of subordinate innovation institutions, manufacturers' research and development departments, and other independent innovation institutions [14]. As a more specific actor, the subdivision component is actively involved in the construction of the actor network, interacting with other actors in the network and embodying their own independent functional attributes.

Through several in-depth interviews, it is revealed that innovation systems, especially private enterprises, call for the continuity in innovation. In an interview, a former technology leader of Huawei Technologies and Microtech described the innovative work of chip and terminal manufacturers in the industry, describing innovation work focusing on integrating the existing technologies, rather than on revolutionary innovation. The original words are quoted as follows:

> For the chip industry, the domestic chip cannot be said to possess independent property rights. They are based on existing technology. We are only processing integration. For the chip, we use the CPU from the ARM (UK); EDN tools from the United States. Chip design language is also from the United States. From front to back they are all westernized. What we do is only to integrate and the core work technology is still dominated by foreign countries. Our advantage lies in its integration capabilities. National chip manufacturers need to have revolutionary innovation to seek stronger role in the 5G development.

Thus, the demand for innovative systems is not just regulatory system funds for stimulation, because the innovation itself is driven by interests and it is impossible to separate innovation from the market system. Also, the innovative continuity is highly demanded for the long-term interests of the innovation system.

3.2.2. Regulatory System. Regulatory systems in China include MII (Ministry of Information Industry), Ministry of Science and Technology, Science and technology commission of the Beijing municipal government, SARFT (State Administration of Radio, Film, and Television), and SASAC (State-owned Assets Supervision and Administration Commission of the state council). In many countries, when facing the basic interests of infrastructure of telecommunication, the government intervenes in the free market. This is the only reason why the Chinese government has a strong system to supervise and manipulate the industry in every aspect, including intellectual property rights protection, innovation encouragement, and asset value management. However, in the 5G initiating phase, a shift of power away from the government was observed.

3.2.3. Market System. In the market system, the components are the service providers, manufacturers, content providers, users, and so forth. It is still early to get solid information, as the 5G technologies are still immature.

Service providers always play important roles in telecommunications development, and this issue becomes even more critical in China's 5G development where China Telecom, China Mobile, China Unicom, and China Net Com are major players. China Unicom already provides unlimited data packages for users to form user habits for future 5G technology in selected developed cities. Customers could enjoy unlimited data package for 139 RMB per month. Content providers also emerge as a significant player in the fields, as 5G enables high speed good quality data everywhere. IoT (Internet of Things), VR (Virtual Reality), and other technologies with high data requirements also boomed recently with content creation.

Manufacturers play an important role in the development of 5G, developing from 3G and 4G, and Da Tang successfully

established the TD-SCDMA standard. To support new 5G air interface and spectrum together with LTE and WiFi, a large numbers of devices are needed to ensure high-rate coverage and a seamless user experience. Manufacturers are striving hard to improve themselves to grasp this 5G opportunity. Based on the insight and understanding of the future development of mobile broadband (MBB), Huawei commercialized its 4.5G Giga-radio solution in June 2016, which continued to innovate with distributed base stations, SingleRAN base stations, and Blade Site base stations and with mobile broadband solution, aiming to help the development of 5G.

In the actor network theory, the meaning of the individual is huge. As Latour described, it is "more complex than the whole," when an individual is no longer a self-sufficient atomic structure but has a series of differences and complexities, such as the properties of a variety of information [15]. Also, in the 5G key network technologies, users play a vital role. The customer-centric network requires customer-centric access, simplifying the multiconnection management mechanism and the service provisions are to be based on user preference as well. In the distribution of interests among different actors (innovation system, regulatory system, and market system), it is important to maintain the balance. In the user-centric orientation, there is a demand for more scenarios for dealing with the technological development challenges. Data traffic explosion is the main driver behind 5G and could be handled through three key technologies: (1) increased area spectral efficiency with extreme densification and offloading, (2) improved bandwidth with mmWave spectrum, and (3) increased spectral efficiency with MIMO [16].

4. Background and Methods

4.1. Background Description. From 1G to 5G, the development of global technical standards has been based on both technology and social interaction. The move from 1G to 4G has paradigm shifts breaking backward compatibility. 5G as a transformative technology requires very high carrier frequencies with numerous base stations and antennas. It will also drastically affect our private lives with the deployment of 5G-based social applications such as Internet of Things (IoT). The 4G technology widely used at present is referred to as the 4th-generation mobile communication system, which mainly includes TD-LTE (Long-Term Evolution, LTE) and FDD-LTE modes. TD and FDD modulation differ by air interface, which could be considered as one type of technology. Regarding the performance, 4G technology, which integrates 3G and WLAN, supports speed of downloads or data transference up to 100 Mbps. Furthermore, it is expandable and easy to be deployed. It however has a few shortcomings. One major disadvantage is that it has too many bands of frequency ranging from 700 MHz to 3.6 GHz, which leads to the complications for terminal design.

The fifth-generation (5G) mobile communication system is the next generation of technical standard. 5G fixed many of the technical weaknesses in 4G technologies, which significantly improved the quality of service, time delay, I/O speed, energy efficiency, and system performance. 5G communication technologies is 10 to 100 times faster than 4G, which reaches 10 Gbps at maximum data transference rate compared to 100 Mbps for 4G technology. And the terminal-to-terminal time delay was shortened by 5 to 10 times compared to 4G. As far as the network capacity is concerned, the mobile data per unit area for 5G is 1000 times more than 4G. And the frequency efficiency of 5G is 5 to 10 times higher than 4G.

The development of 5G has two schemes. One is to improve the technology step by step, which means developing the technology based on currently used 4G LTE technologies to improve the network capacity and performance. This scheme employed unitary TDD and FDD technologies, enhanced relay, 3D-MIMO, enhanced CoMP, LTE-Hi with small sized base, and so forth. Another scheme is to design completely new network structures and wireless technologies to construct a whole new mobile communication network, which requires the following key technologies.

(1) The Employment of High Frequency Bandwidth. To mobile communication, most of the function frequency is below 3 GHz, which provides limited spectrum for more users. The bandwidth spectrum over 3 GHz however has not been fully used. The employment of communication frequency higher than 3 GHz will effectively relieve the shortage of available spectrum resources. High frequency spectrum (60 GHz, e.g.) has the feature of high anti-interference, sufficient bandwidth with reusable bandwidth, smaller size equipment, and antenna with high gain rate. On the other hand, it also has some disadvantages such as shorter broadcasting range and stronger diffraction to the signal, which is also easily affected by the weather or larger obstacles. Therefore it is necessary to consider the work condition for high frequency bandwidth, which is often used cooperatively with other communication techniques.

(2) Innovative Multiantenna Technologies. With the rapid development of wireless communications, the demand for data traffic is growing, and the available spectrum resources are limited. Therefore, it is very important to improve the efficiency of spectrum utilization. Multiantenna technology is an effective way to improve network reliability and spectrum efficiency. At present, it is being applied to all aspects of wireless communication, such as 3G, LTE, and LTE-A. The increased number of antennas provides guarantee for the reliability of data transmission as well as spectral efficiency.

(3) Cofrequency Cotime Full Duplex (CCFD). Traditional wireless communication technology has limitations; it cannot realize bidirectional communication of the same frequency at the same time. This results in great waste of resources and CCFD technology can realize simultaneous use of the same upload and download frequency resources for two-way communication. It theoretically utilizes the system two times. But CCFD also encounters a technical problem of serious self-interference; hence the primary problem is to eliminate interference. In addition, there is also the problem of cofrequency interference with adjacent cells and full duplex. The CCFD application in multiantenna environment will be more difficult and it requires further research.

Table 1: 5G scenarios.

Scenarios	Main applications	Objectives	User requirements	MIMO	UDN	Novel multiple access	All-spectrum access
Wide-area coverage	Everywhere	Consistency	100 Mbps user experienced data rate	V			V
Hot spot	Bus station, football fields, theater,	Tens of Gbps peak data rate and tens of Tbps/km^2 traffic volume density	1 Gbps user experienced data rate	V	V	V	V
Massive-connections	IoT	Low-power massive-connections Low-latency high-reliability, low-power	Reduce device cost Air interface latency: 1 ms End-to-end latency: ms level Reliability: nearly 100%				V

(4) Dense Network. In the future 5G communications, the wireless communication network is evolving towards network diversification, broadband, and integrated, intelligent direction of evolution. With the popularity of a variety of intelligent terminals, data traffic will experience exponential growth. The future of data services will be mainly distributed in the indoor and hot spots, which makes the ultradense network the main method to achieve 1000 times of the mean demand for data traffic for 5G in the future. Ultradense networks will be able to improve network coverage, significantly increase system capacity, and shunt traffic, with more flexible network deployment and more efficient frequency reuse. In the future, for high-band large bandwidth, one will adopt a more intensive network solution for the demand of high frequency spectrum and large bandwidth, which might deploy up to 100 small districts/sector.

(5) Device-to-Device (D2D) Techniques. Traditional cellular communication system is based on the base station as the center to achieve cellular coverage, and while the base stations and relay stations cannot move, the flexibility of the network structure has some restrictions. With the increasing number of wireless multimedia services, the traditional base station, as the center of the business, has been unable to meet the business requirements of massive users in different environments.

D2D technology can achieve direct communication between communication terminals, expanding network connectivity and access methods without the help of the base station.

Due to short-range direct communication, D2D technology can realize higher quality data communication, lower latency and lower power consumption, improve coverage and efficient use of spectrum resources through widely distributed terminals, and support more flexible network architecture and connection methods to enhance link flexibility and network reliability. Currently, D2D technology uses broadcast, multicast, and unicast technological solutions. More technologies will be enhanced in the future, including D2D-based relay technology, multiantenna technology, and joint coding technology.

4.2. Scenarios. 5G can offer high speed, quality, and seamless service, meeting the needs for different applications as well as scenarios. Typical applications include connected gadgets, robotics, social web of things, autonomous vehicles, micropayments, retail logistics, shopping assistance, pollution surveillance, smart grid, water/waste management, remote monitoring, assisted living, integrated environments, optimized operations, and automation [17]. Reduced device costs, better performances, improved coverage, higher speeds, higher capacity, and improved battery life are required for different scenarios; three typical scenarios could be derived from the main applications mentioned above, according to drastically different objectives, performance requirements, user requirements, service requirements, and key challenges. They are as follows: wide-area coverage scenario, hot-spot scenario, and massive-connections scenario as shown in Table 1.

4.2.1. Wide-Area Coverage Scenario. Wide-area coverage scenario is the basic seamless scenario providing service everywhere at anytime, with data rate exceeding 100 Mbps. Some extreme situations could be moving vehicles, such as highway situations which require seamless services on moving vehicles with a speed of 500 km/h. In order to achieve spectral efficiency with a large number of users, massive MIMO technology is needed and it requires more antennas and low-cost implementation. Novel multiple access technology could be applied to improve spectral efficiency and access capability: sparse code multiple access (SCMA), multiuser shared access (MUSA), pattern division multiple access (PDMA), and nonorthogonal multiple access (NOMA) are the possible future schemes.

4.2.2. Hot-Spot High-Capacity Scenario. Hot-spot scenario provides high-capacity as well as quality ultrahigh data

rates, where ultrahigh volume traffic is being handled. Via deploying dense base stations, ultradense networking will increase capacity with novel multiple access technologies. All-spectrum access could also be applied to a variety of spectrum resources, with enhanced data rates and system capacity.

4.2.3. Massive-Connections IoT Scenario. Massive-connections scenarios are mainly IoT, sensing/collecting data, and vertical industries. IMT-2020 concluded it with low-latency high-reliability and low-power massive-connections. Their scenarios require low-power and low costs for a large number of devices applied. Low-latency with high-reliability is required for IoT and vertical industries.

4.3. Methodology. In this paper, a mixed method containing science and social science is used. The method of field study is used and is supported by in-depth interview and document collection and historical data comparison, so as to reveal the trend in the 5G development: an obvious shift from policy strategy to user-oriented architecture. A hybrid routing protocol TOHRP and a distributed channel assignment algorithm LBCA in the multichannel environment are proposed and they solve the traditional problem of (1) waste of spectrum and (2) self-interference.

Field study is a very suitable research method. The purpose of this paper is to analyze the formation and diffusion of 5G technical standards in the 5G's preliminary stage. Semistructured in-depth interviews were used with document collection and comparison of historical data in this study for combining historical data, like the historical changes in regulatory agencies, the evolution of the telecommunications broadcasting system, and so on.

Also, to increase the throughput and spectrum efficiency, channel assignment and load balance are considered. A hybrid routing protocol TOHRP and a distributed channel assignment algorithm LBCA in the multichannel environment are proposed. They solve the traditional problem of (1) waste of spectrum and (2) self-interference. The tree base protocol and the traditional AODV (Ad hoc On-demand Distance Vector) routing protocol are integrated together to reduce delay in the routing protocol, which is proposed by HWMP (Hybrid Wireless Mesh Protocol). A new routing metric is used in this routing protocol and the computer-simulated experiment shows that the algorithm improves the performance.

The relevant computer-simulated experiment follows the selection, preparation, environment, steps, and assessment procedures.

4.4. Experiment

4.4.1. Selection of the Experiment. As the amount of data processing and transmission is enormous in 5G network and the performance is closely related to routing protocol, the key solution for a network with optimized throughput and spectrum efficiency is to choose the best path for large amounts of data to transmit. To improve efficiency, Multiradio Multichannel system is selected, and each node has both the receiving function and the sending function to simulate the 5G scenarios. To solve the problem of self-interference, different information channels are assigned and load balancing issue is also considered.

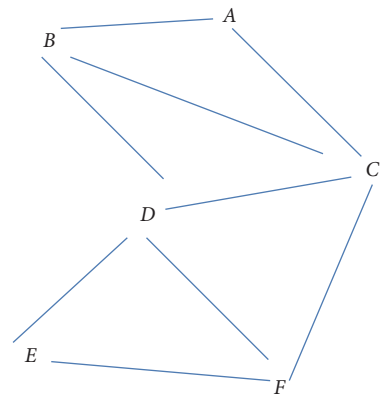


Figure 2: A 5G network transmission model.

A 5G network transmission can be simulated as a model of weighted graph $G = \langle V, E \rangle$, in which weight of edge $e \in E$ is to be determined with each line (shown in Figure 2). There are several traditional algorithms to calculate the optimal path [18–21].

4.4.2. Preparation of the Experiment

ETT (Expected Transmission Time) Algorithm. On link AB, set a time window with length of s at node B to record the number of times (c) that B successfully receives detected packets from node A in the past s seconds. c/s is set to be p_f, the success rate of forward packet transmission on link AB. Similarly, node B regularly sends detective packets to node A and p_r the success rate of backward packet transmission on link AB can be estimated. The ETX (Expected Transmission Count) for a single-hop link is

$$\begin{aligned} \mathrm{ETX} &= \sum_{r=1}^{\infty} r \times p^{r-1} \times (1-p) = (1-p) \sum_{r=1}^{\infty} \frac{dp^r}{dp} \\ &= (1-p)\, \frac{d \sum_{r=1}^{\infty} p^r}{dp} = \frac{1}{1-p} = \frac{1}{1-p_f-p_r}. \end{aligned} \tag{1}$$

Here r is defined as the resending times after the failure of packet transmission. Then the ETT (Expected Transmission Time) for a single-hop link is

$$\mathrm{ETT} = \mathrm{ETX} \times \frac{S}{B}. \tag{2}$$

Here S is the size of the detective packet and B the bandwidth of the link. ETT is the Expected Transmission Time Algorithm.

ETTI (Expected Transmission Time with Interference) for a single-hop link is

$$\mathrm{ETTI} = \mathrm{ETTI}_{\mathrm{Inter}} + \mathrm{ETTI}_{\mathrm{Intra}}. \tag{3}$$

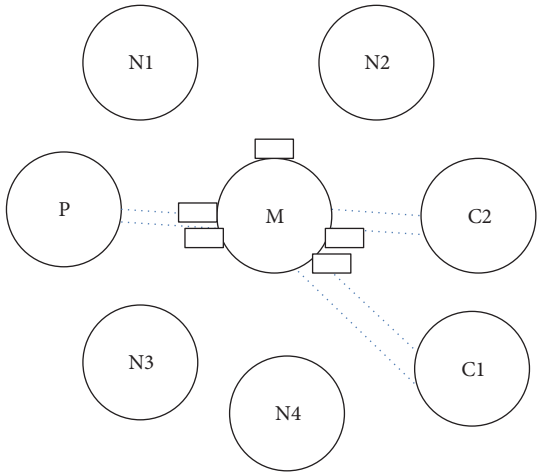


Figure 3: Bind neighbor node to the T transceiver.

Define I_A (and I_B) as the collection of nodes that interferes with the communication of node A (and node B) within A's (and B's) coverage of interference. Thus,

$$\mathrm{ETTI}_{\mathrm{Inter}} = \mathrm{ETT} \times \left| I_A \cup I_B \right|. \tag{4}$$

Equation (4) shows that the interference towards a link is caused by outside data flows, and the interference within the link is determined as (5) shows:

$$\mathrm{ETTI}_{\mathrm{Intra}} = \sum_{i=1}^{k} \mathrm{ETTI}_i \times e,$$
$$e = \begin{cases} 1 & \mathrm{channel}_i = \mathrm{channel}_{AB},\ i \cap \left(I_A \cup I_B\right) \neq \varphi \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

where k means that AB is the kth hop on the route.

The calculation method of $\mathrm{AETTI}_{\mathrm{Intra}}$ is the same as that for $\mathrm{ETTI}_{\mathrm{Intra}}$. AETTI (All Expected Transmission Time with Interference) takes into account not only traditional factors such as delay but also channel interference and node load.

$$\mathrm{AETTI} = \sum_{i=1}^{n} \mathrm{ETTI}_i. \tag{6}$$

Through the routing protocol, nodes on the network can communicate with other nodes and establish a path to connect to the Internet. Neighbor nodes transferred on the same channel will interfere with each other. This is a key problem of 5G network, and the channel allocation needs to be addressed. In this paper, a distributed algorithm called Channel Load Based Channel Assignment (LBCA) is proposed to eliminate interference and improve channel utilization.

According to the characteristics of the tree topology, each node can be classified as transceiver to a parent node (marked as P in Figure 3) or a child node (marked as C in Figure 3). Each node needs a common transceiver (marked as M in Figure 3) for the transmission of various control information.

As shown in Algorithm 1, each node has public transceiver and two other parts: an upstream transceiver for communicating with the parent node, and a downstream transceiver for communicating with the child nodes. When the channel of the downstream transceiver of each node changes, it only affects the upstream transceiver of its child node and will not cause the channel dependency problem. And the LBCA algorithm code is as Algorithm 1 shows.

For Algorithm 1,

(1) each node is responsible for allocating the downstream transceiver channel of its own, and the upstream transceivers channel is the same as the downstream transceiver channel of the corresponding parent nodes;

(2) the closer it is to the gateway, the heavier its load is. So each node must also be assigned a priority according to the hops to gateway node. A node with a lower value has a higher priority;

(3) each node periodically sends a CHANUSAGE packet to $(k + 1)$ hop neighbor nodes (k is the ratio of the node interference radius to the communication radius), or when the channel used by a node changes, to broadcast its own channel usage;

(4) each node will recalculate the channel load of the neighbor node according to the channel load received in the CHANUSAGE packet. The calculation method is as follows:

$$\mathrm{LOAD}_{\mathrm{new}} = \frac{\mathrm{LOAD}_{\mathrm{old}} + \mathrm{LOAD}_{\mathrm{current}}}{2}; \tag{7}$$

(5) the node periodically determines the vector of an optimal channel according to the channel usage of the neighbor node in its neighbors table and the priority of the neighbor node:

$$V_{\mathrm{optimal}} = \left\{\mathrm{channel}_1, \mathrm{channel}_2, \mathrm{channel}_3, \ldots, \mathrm{channel}_n\right\}. \tag{8}$$

4.4.3. Specification of Environment. To verify Revised Channel Selection Algorithm, simulation has been conducted with NS2. Comparison on average delay and throughput of mesh network estimated with other frequently used algorithms are carried out. Parameters of the simulated environment are shown in Table 2.

4.4.4. Steps and Subjective Assessment of the Data. Revised Channel Selection Algorithm is tested in this experiment and the result is compared with frequently used metrics hop count and ETT.

In the simulated experiment, every node is set as fixed and changes of throughput and end-to-end delay are compared at different rates of packet transmission under different routing metrics.

```
(1) Set m = any channel in Voptimal
(2)  Set k = the count of the channels
(3)  Set Timer Tsend to T1
(4)  If Tsend is expired
(5)       broadcast CHANUSAGE message to its (k + 1) hop neighbors
(6)  End If
(7)  Set Timer Tchange to T2
(8)  If Tchange is expired
(9)       Look up the neighbor table
(10)       use heap sort to find the most idle channel vector Voptimal
(11)      Set n = the element count of the Voptimal
(12)      If n > k
(13)           For Each channel m in Voptimal
(14)                If m is not used
(15)                     change the channel of one DOWN-RADIO to m
(16)                     End If
(17)           End For Each
(18)       Else
(19)           For Each channel m in the first n channels of Voptimal
(20)                If m is not used
(21)                     change the channel of one DOWN-RADIO to m
(22)                End If
(23)           End For Each
(24)      End If
(25)      If any radio has changed
(26)           broadcast CHANCHANGE message to its (k + 1) hop neighbors
(27)      End If
(28) End If
(29) If receive a CHANUSAGE message
(30)      update its neighbor table
(31) End If
(32) If receive a CHANCHANGE message
(33)      If the sender is parent and its DOWN-RADIOs changed
(34)          change UP-RADIOs to the channels of parent's DOWN-RADIOs
(35)          broadcast CHANCHANGE message to its (k + 1) hop neighbors
(36)      End If
(37)      update its neighbor table
(38) End If
```

Algorithm 1: LBCA algorithm code.

5. Key User-Oriented Issues

To meet users' extreme requirements and to suit a variety of 5G scenarios with efficiency and performance, the key issue to be addressed is spectrum efficiency and throughput.

Increased spectrum efficiency and throughput could be achieved through Cofrequency Cotime Full Duplex (CCFD), all-spectrum access technology, Device-to-Device (D2D) technology, and other technologies. Spectrum efficiency is supposed to be at least three times that for 4.5G. Innovation is needed for achieving that goal. Cooperation among global organizations, universities, and research institutes should be enhanced. Enterprises, especially small and medium-sized ones, should be promoted and mobilized, so as to form a better industry alliance as a healthy actor network, shaping the consensus.

This paper simulates a 5G network, and, through the experiment, a hybrid routing protocol TOHRP and a distributed channel assignment algorithm LBCA in the multichannel environment are proposed and they solve the traditional problem of (1) waste of spectrum and (2) self-interference.

Simulated experiment results shown in Figures 4 and 5 illustrate the comparison of throughput. Hop count, ETT, and AETTI listed together, we could easily observe a higher rate of packet transmission, throughput of AETTI, followed by ETT. Figure 5 demonstrates the comparison of the end-to-end delay. Hop count has the longer delay.

Table 2: Specification of the environment.

Parameters	Set value
Campaign topology	1000 m × 1000 m
Number of nodes	60
Data rate	100 Mbps
Number of channels	3
Number of radios	1
Range of data transmission	250 m
Range of interference	500 m
Length of packet	1024 bytes
Type of flow	UDP
Ways of data transmission	CBR

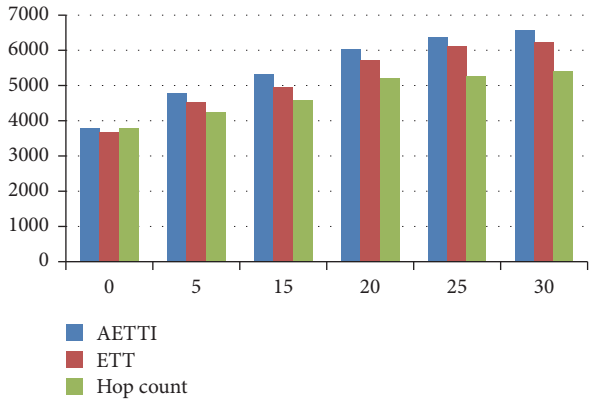


Figure 4: Comparison of throughput.

The analysis above indicates that AETTI has taken channel information interference and node load into consideration. AETTI has better performance than hop count and ETT in the multichannel situation.

6. Conclusion and Suggestions

In China, the current status of the 5G test is limited to technology assessment and related technologies, including Cofrequency Cotime Full Duplex (CCFD), Device-to-Device (D2D) technology, all-spectrum access, massive MIMO, and ultradense network. Innovation systems including corporations and government supported research centers, as well as colleges, are working on the related research for power in the standardization process. Only by satisfying interests to each actor with actual involvement could the 5G development grow healthily. As different scenarios require different technology approaches and in order to meet user's needs, the following issues need to be addressed.

6.1. Smooth Evolution from 4G to 5G Requires More Private Enterprises Enrollment in Spectrum Efficiency Research as Innovation System. Support to the development of small and medium-sized private enterprises, for sustainable and revolutionary innovation, especially in spectrum efficiency

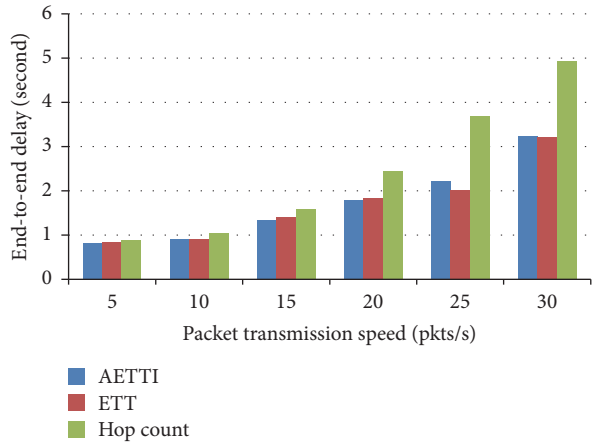


Figure 5: Comparison of end-to-end delay.

researches is needed. A smooth transition from 4G to 5G should be put to schedule. The market demand for data services growth will not be slowed down because 5G is not yet a mature technology. Thus operators must continue to improve the existing 4G networks and to take full account of the smooth transition for the future.

Spectrum efficiency should be improved at least three times that of 4.5G now, and related technologies will be applied directly so as to serve a more smooth transition from 4G to 5G.

6.2. Standardization in a Global Context Needs Cooperation. Standardization work should pay attention to both international and local aspects at the same time. Under the framework of the Union and 3GPP, it has to emphasize both eMBB and IoT. Only by generating compatible, reliable, and efficient technologies, standards could be recommended by 3GPP and ITU. Premature overseas agendas will outweigh the benefits.

The standardization of 5G requires cooperation and sharing of results. Experimental results sharing program is indispensable and cooperation platform should be enhanced.

6.3. User-Oriented Architecture Requires High Throughput and Spectrum Efficiency. As the main observation of a shift from policy strategy to being customer oriented is revealed, service providers, content providers, and manufacturers could benefit by scientifically collecting, studying, and considering user-needs at the initiating stage for a more satisfying end user experience.

Improving throughput and spectrum efficiency especially for 5G data is an important issue to be addressed in the vast growing network in 5G scenarios.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work has been partially supported by the 2016 Humanities and Social Sciences Project of the Ministry of

Education (16YJC630037), partially funded by a grant from Hubei Provincial Natural Science Foundation of China (2013CFB294), and partially funded by Project in the 2017 Guangzhou Philosophy and Sociology Science Program during the Thirteen Five-Year Plan Period (no. 2017GZQN04). Many thanks go to Mr. John Yip for his valuable advice and revision suggestions.

References

[1] P. M. Napoli, *Foundations of Communications Policy: Principles and Process in the Regulation of Electronic Media*, Hampton Press, 2001.

[2] M. Dohler and T. Nakamura, *5G Mobile and Wireless Communications Technology*, Cambridge University Press, 2016, Edited by A. Osseiran, J. F. Monserrat, O. Queseth and P. Marsch.

[3] F. Boccardi, R. Heath Jr., A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, 2014.

[4] W. Roh, J.-Y. Seol, J. Park et al., "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 106–113, 2014.

[5] T. S. Rappaport, S. Sun, R. Mayzus et al., "Millimeter wave mobile communications for 5G cellular: it will work!," *IEEE Access*, vol. 1, pp. 335–349, 2013.

[6] http://www.surrey.ac.uk/.

[7] S. N. L. Paul, *Telecommunications and Development in China*, Hampton Press, Cresskill, NJ, USA, 1997.

[8] Y. Xu and P. Douglas, *Chinese Telecommunications Policy*, Artech House, Norwood, Mass, USA, 2002.

[9] Y. Yoo, K. Lyytinen, and H. Yang, "The role of standards in innovation and diffusion of broadband mobile services: the case of South Korea," *The Journal of Strategic Information Systems*, vol. 14, no. 3, pp. 323–353, 2005.

[10] S. Guoxun, "Sociology and social construction theory," *Foreign Social Sciences*, vol. 1, pp. 4–13, 2002.

[11] J. Law, "Notes on the theory of the actor-network: ordering, strategy, and heterogeneity," *Systems Practice*, vol. 5, no. 4, pp. 379–393, 1992.

[12] M. Callon, "Techno-economic networks and irreversibility," in *A Sociology of Monsters: Essays on Power, Technology and Domination*, J. Law, Ed., p. 153, Routledge, London, UK, 1991.

[13] K. Lyytinen and J. L. King, "Around the cradle of the wireless revolution: the emergence and evolution of cellular telephony," *Telecommunications Policy*, vol. 26, no. 3-4, pp. 97–100, 2002.

[14] http://www.imt-2020.org.cn/.

[15] B. Latour, "Networks, societies, spheres: reflections of an actor-network theorist," *International Journal of Communication*, vol. 5, no. 1, pp. 796–810, 2011.

[16] J. G. Andrews, S. Buzzi, W. Choi et al., "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[17] J. Holler, V. Tsiatsis, C. Mulligan, S. Avesand, S. Karnouskos, and D. Boyle, *From Machine-to-Machine to the Internet of Things: Introduction to a New Age of Intelligence*, Academic Press, 2014.

[18] J. C. Park and S. K. Kasera, "Expected data rate: an accurate high-throughput path metric for multi-hop wireless routing," in *Proceedings of the 2nd Annual IEEE Communications Society Conference on Sensor and AdHoc Communications and Networks (SECON '05)*, pp. 218–228, Santa Clara, Calif, USA, September 2005.

[19] P. KyaSanur and N. H. Vaidya, "Routing and interface assignment in multi-channel multi-interface wireless networks," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '05)*, New Orleans, La, USA, March 2005.

[20] P. Kyasanur and N. H. Vaidya, "Routing and link-layer protocols for multi-channel multi-interface ad hoc wireless networks," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 10, no. 1, pp. 31–43, 2006.

[21] M. Malekesmaeili, M. Shiva, and M. Soltan, "Topology optimization for backbone wireless mesh networks," in *Proceedings of the 5th Annual Conference on Communication Networks and Services Research (CNSR '07)*, pp. 221–227, Fredericton, Canada, May 2007.

Hindawi
Mobile Information Systems
Volume 2017, Article ID 6805056, 12 pages
https://doi.org/10.1155/2017/6805056

Research Article

Effective Feature Selection for 5G IM Applications Traffic Classification

Muhammad Shafiq,[1] Xiangzhan Yu,[1] Asif Ali Laghari,[1] and Dawei Wang[2]

[1]*School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China*
[2]*National Computer Network Emergency Response Technical Team, Coordination Center of China, Beijing, China*

Correspondence should be addressed to Xiangzhan Yu; yuxiangzhan@hit.edu.cn

Received 11 January 2017; Accepted 4 April 2017; Published 22 May 2017

Academic Editor: Ben Niu

Recently, machine learning (ML) algorithms have widely been applied in Internet traffic classification. However, due to the inappropriate features selection, ML-based classifiers are prone to misclassify Internet flows as that traffic occupies majority of traffic flows. To address this problem, a novel feature selection metric named weighted mutual information (WMI) is proposed. We develop a hybrid feature selection algorithm named WMI_ACC, which filters most of the features with WMI metric. It further uses a wrapper method to select features for ML classifiers with accuracy (ACC) metric. We evaluate our approach using five ML classifiers on the two different network environment traces captured. Furthermore, we also apply Wilcoxon pairwise statistical test on the results of our proposed algorithm to find out the robust features from the selected set of features. Experimental results show that our algorithm gives promising results in terms of classification accuracy, recall, and precision. Our proposed algorithm can achieve 99% flow accuracy results, which is very promising.

1. Introduction

Accurate network traffic classification is extremely important for the network management including IP network management, deploying QoS-aware mechanisms, monitoring security, bandwidth management, and intrusion detection. For instance, it is useful for the Internet Service Providers (ISPs), network operators, and network administrators to understand the traffic composition and prioritize some sensitive bandwidth traffic such as video conferencing and voice over IP (VoIP). Moreover, from the perspective of network security, network traffic classification technique can help us in blocking unwanted or attack traffic.

In the last few years, several traffic classification models [1, 2] have been proposed in this regard. Traditionally, port-based technique was proposed, which is based on well-known port numbers for traffic classification. This technique is easy to be deployed and implemented. However, many Internet applications use dynamic port number for their communication instead of well-known port number, which makes it difficult for the network operator to identify network traffic composition by port numbers. Moore and Papagiannaki [3] showed that port-based traffic classification technique does not give more than 50–70 percent accuracy.

To address the above problems, payload-based technique was proposed [4, 5] which inspects the packet payload signatures. Though this traffic classification technique is easy and accurate, this classification technique is ineffective for encrypted applications as numerous applications such as Skype use encrypted methods to protect their data from being detected. Furthermore, this technique is against the privacy laws to inspect the packet payload. In recent years, machine learning (ML) algorithms have been presented to classify Internet traffic flows, whose features are extracted from the flow statistics. Moreover, this ML technique is user privacy friendly and does not inspect the packet payload. However, the accurate feature selection problems challenge this technique. Features selection refers to the selection or filtration of accurate features in more than available features. For instance, Moore et al. in 2005 [6] presented the most wide features of extraction and selection method and they selected 248 statistical features, based on a whole traffic flow.

Using these selected statistical features, classifiers can achieve high performances results in Internet traffic classification. Nevertheless, in real words, it is not applicable to use these features in traffic classification. Recently, Zhang et al. [7] presented feature selection algorithm based on imbalance traffic classification. However, it is important to study more deeply accurate features selection techniques for Internet traffic classification.

In the previous work [8], we classify instant messages (IM) WeChat application message services (including text and picture messages) using four machine learning (ML) classifiers. We select 50 statistical flow-based features and got very effective accuracy results. Similarly, in [9, 10], we classify IM traffic accurately and find out effective features for IM traffic classification using machine learning algorithms. However, more than 50 features increase computational complexity and decrease accuracy results. That is why it is important to study more deeply accurate features selection technique for IM applications traffic classification.

In this paper, we proposed a feature selection algorithm to improve the performance of ML-based traffic classification technique and select effective features set for IM applications traffic classification. Our main contributions are given as follows:

(i) In order to deal with accurate features selection problem in Internet traffic identification, a hybrid feature selection algorithm named WMI_ACC is proposed. It includes two metrics for accurate features selection: weighted mutual information (WMI) metric and accuracy (ACC) metric of the selected classifiers. Firstly, WMI_ACC algorithm assigns WMI values to the features. After assigning the WMI values, WMI_ACC algorithm selects the best features with high ACC values for the specific classifier. Our study is the first study where the WMI metric is put forward in Internet traffic classification. Furthermore, this is the first time we apply WMI combined with ACC metrics for feature selection in traffic identification.

(ii) We present the features that are selected by our proposed algorithm and report their actual values with metric values. Experimental results show that ten common flow-based features selected from different network environment traces have discriminative power for classifying instant messaging (IM) applications TCP and UDP flows. These features are (1) max_fpktl, (2) mean_fpktl, (3) max_fpktl, (4) std_fpktl, (5) min_bpktl, (6) mean_bpktl, (7) max_bpktl, (8) std_bpktl, (9) max_fiat, and (10) total_fpacket.

(iii) Our third contribution is to select the robust features from the selected features of our proposed WMI_ACC algorithm. We use Wilcoxon statistical test to find out the robust feature from the selected features.

The rest of the paper is organized as follows: Section 2 gives an overview of the related works. Section 3 elaborates our proposed WMI_ACC algorithm. Section 4 demonstrates the evaluation methodology with details, experimental work, and utilized datasets. Analysis and discussions are given in Section 5. Finally, conclusions and future works are shown in Section 6.

2. Related Work

Recently, machine learning (ML) algorithm has widely been applied in Internet traffic classification in [11–19]. Some of them are applied for traffic flow classification and some of them are applied for bandwidth management. However, most of the methods are applied for improving the performance of classification by using ML algorithms methods. These proposed methods are able to get 80% accuracy results. For feature extraction and selection, mostly using feature selection method as presented by Moore et al. in 2005 [6], based on whole traffic flow, they selected 248 statistical features, such as RTT and minimum, maximum, and average values of packet size. Using these selected statistical features, the applied classifiers can get very effective performance results in Internet traffic classification. But, in real circumstances, it is not good for traffic classification.

Thus, we must select accurate features for Internet traffic classification so that we can manage subsequent management and security policies. In 2012, in [7], Zhang et al. proposed two different algorithms for feature selection for optimization of traffic classification. They evaluated their results based on true positive rate (TPR) and false positive rate (FPR) and proved that their algorithm can achieve greater than 90% flow accuracy. Similarly, Peng et al. in 2016 [20] evaluated the effectiveness of statistical features. But their research study was limited to only early stage Internet traffic classification. Bernaille et al. [21] studied the problem of effective features in network traffic classification. In their study, they used K-Means and GMM and HMM model for Internet traffic classification. They selected packet size as a feature and extracted more features for early stage Internet traffic classification. Lim et al. in [22] used the packet size, connection level, and statistical features for Internet traffic classification. Van Der Putten and Van Someren in [23] revealed that features selection is very important for performance optimization compared to the choice of classification classifiers.

Zheng et al. in [24] showed that how many types of selected features metrics affects the identification performances. For accurate features selection purpose, Chen and Wasikowski in [25] proposed new feature selection metric using area under the ROC curve to evaluate features for Internet traffic classification. Kamal et al. [26] proposed three different filtering techniques, Balanced Minority Repeat (BMR), Differential Minority Repeat (DMR), and Higher Weight (HW), for the identification of effective features.

From the last few years, daily use of Internet applications increases day by day due to free of cost availability of instant messaging (IM) applications. It is also important to accurately classify IM applications. For instance, WeChat is an IM and free calling application developed by Tencent Holdings in China. After launching the WeChat application, its online users reached 300 million [27] and, in November 2015, its

active users reached 650 million all over the world. Apart from China, its online users reached 100 million [28] in the rest of the world. So, day to day increasing number of active users and traffic of this application can affect performance of the network. It is also important to classify WeChat messages and audio and video call traffic accurately to manage the quality of services (QoS) as Huang et al. [29] proposed measurement ChatDissect tool to measure WeChat application traffic and distinguish 150K users and 16 GB traffic of WeChat from real-world network traces. In 2013, Church and Rodrigo de Oliveira [30] studied the performance of mobile instant messaging sending service with traditional short messages. In 2014, O'Hara et al. [31] studied instant messaging application WhatsApp in smartphone and conducted some interviews and a survey to study the user activity using WhatsApp application. In 2014, Fiadino et al. [32] also studied WhatsApp application's flow stream and collected data in European Network, which consisted of millions of data flow streams. They also studied audio and video flow data stream. In 2014, Liu and Guo [33] studied video messaging services in WeChat and WhatsApp applications. They captured the traffic using mobile devices for their study. Furthermore, in our previous work in [8], we classify IM applications; however, we only classify WeChat text messages service flow traffic. Nevertheless, it is more important to select accurate features for IM applications traffic classification.

3. Proposed Method

In this section, we explain our proposed WMI_ACC algorithm in detail. First of all, we examine the problem of features selection. Then we introduce WMI based feature metric and then design the WMI_ACC algorithm that uses the WMI combined with ACC metric to select effective features for Internet traffic classification.

3.1. Feature Selection Metrics

3.1.1. Mutual Information Based Metric. In information theory, mutual information is extensively used for features selection [9, 34], image processing [35], speech recognition [36], and so forth. It measures the mutual dependency between two random variables X and Y, which describes the amount of information held by random variable. The mutual information between two random variables is described as

$$\begin{aligned} I(X;Y) &= H(X) - H(X \mid Y) = H(Y) - H(Y \mid X) \\ &= H(X) + H(Y) - H(X,Y) \\ &= H(X,Y) - H(XY) - H(YX). \end{aligned} \tag{1}$$

In (1), the marginal entropies of X and Y are $H(X)$ and $H(Y)$, while conditional entropies are $H(X \mid Y)$ and $H(Y \mid X)$ and joint entropies of X and Y are $H(X,Y)$. Moreover, the relationship between $H(X)$, $H(Y)$, $H(X \mid Y)$, $H(Y \mid X)$, $H(X,Y)$, and $I(X;Y)$ is shown in Figure 1. According to Shannon definition of entropy theory, we have

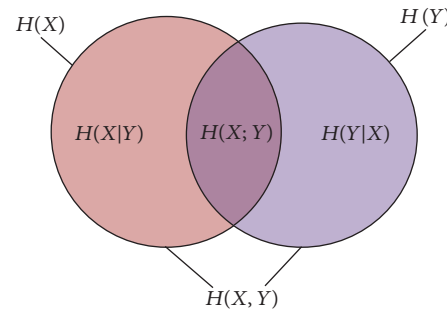


Figure 1: The relationship between mutual information and entropies.

$$\begin{aligned} H(X) &= -\sum_{x \in X} p(x) \log(p(x)), \\ H(Y) &= -\sum_{x \in Y} p(y) \log(p(y)), \\ H(X,Y) &= -\sum_{x \in Y} \sum_{y \in Y} p(x,y) \log(p(x,y)), \end{aligned} \tag{2}$$

where $p(\cdot)$ indicates the probability distribution of a random variable. As in [11], use the three equations in (1) to achieve the computational formula for mutual information. We also use the same method that the authors therein have used for mutual information.

$$I(X;Y) = -\sum_{x \in Y} \sum_{y \in Y} p(x,y) \log\left(\frac{p(x,y)}{p(x)\,p(y)}\right). \tag{3}$$

In case of continuous random variables, the summation will be replaced by a definite double integral.

$$I(X;Y) = \int_Y \int_X p(x,y) \log\left(\frac{p(x,y)}{p(x)\,p(y)}\right) dx\,dy. \tag{4}$$

3.1.2. Weighted Mutual Information (WMI) Metric. To address accurate features selection problem, we proposed weighted mutual information based on weighted entropy. If the total number of features is N, the weight value is calculated as follows:

$$W_i = 1 - \frac{ni}{N}, \tag{5}$$

where ni is the number of features assigned to features set; then the weighted mutual information (WMI) between two random variables can be defined as

$$\begin{aligned} I_w(X;Y) &= H_w(X) - H_w(XY) = H_w(X) - H_w(XY) \\ &= H_w(X) + H_w(XY) - H_w(X,Y) \\ &= H_w(X,Y) - H_w(XY) - H_w(XY). \end{aligned} \tag{6}$$

In (5), the weighted marginal entropies of X and Y are $H_w(x)$ and so on. Again $p(\cdot)$ is the probability distribution of random

Input: $D(F_1, F_2, F_3, \ldots, F_N, F)$ // training data set,
Output: *feature* [] //selected feature set
(1) **begin**
(2) **for** i = 1 to **M**
(3) calculate weight value $w[i]$ for each features;
(4) **end for**
(5) **for** i = 1 to N;
(6) calculate WMI(F_i);
(7) **if** (WMI(F) > δ);
(8) insert F_i into descending order;
(9) **end if**
(10) **end for**
(11) F_p = getfirstfeatures(list);
(12) **end until** (Fp == NULL);
(13) X is a data set of samples
Values of features;
(14) last_accuracy ← classify X;
(15) Insert the feature into S_{wrappers};
(16) Feature = get next features;
(17) **For** feature is not NULL
(18) insert the feature into S_{wrappers};
(19) X is a data set of samples values for S_{wrappers};
(20) Accuracy ← classify X with a specific classifiers;
(21) **If** (ACC <= *last_ACC*)
(22) Remove features from S_{wrappers};
(23) **else**
(24) feature = getNextfeature(list, feature);
(25) **end if**
(26) **end for**
return S_{wrappers};

Algorithm 1: Feature selection algorithm based on weighted mutual information combined with ACC (WMI_ACC).

variable. So, by using the three equations in (6), the weighted mutual information can be obtained as follows:

$$H_w(X) = -\sum_{x=X} w_i p(x) \log\left(p(x)\right),$$

$$H_w(Y) = -\sum_{x=Y} w_i p(Y) \log\left(p(Y)\right),$$

$$H_w(X;Y) = -\sum_{x\in Y}\sum_{y\in Y} w_i p(x) \log\left(w_i p(x)\right), \tag{7}$$

$$I_w(X;Y) = \int_Y \int_X w_i p(x,y) \log\left(\frac{w_i p(x,y)}{w_i p(x)\, w_i p(y)}\right) dx\, dy.$$

For mutual information computational analysis, there is a bundle of software applications publically available, but we select H. Peng's mutual information Matlab toolbox [37] for our study.

3.1.3. ACC Metric. After using WMI metric, it is essential to select the effective features for specific ML classifier to obtain effective performance results. For this purpose, a wrapper method based on accuracy (ACC) metric is applied. On the other hand, to achieve high performance accuracy with regard to classification of applications, the AUC metric is not suitable to rank the features. The highest ACC implies that ML classifier can obtain effective performance. Thus, we rank the features by using ACC metric and select those features with highest ACC values. We used C4.5, Random Forest, and Random Tree machine learning classifiers for ACC metric.

3.1.4. WMI_ACC Algorithm. In this section, we propose effective features selection algorithm, named as WMI_ACC. WMI_ACC is a hybrid features selection algorithm based on WMI combined with ACC metric. Firstly, it filters most of the features with WMI metric and then selects the effective features with ACC metric for a specific algorithm. Algorithm 1 shows the detailed pseudocodes for our implemented features selection algorithm.

In Algorithm 1, there are two steps, given dataset D with M classes and N features. In the first step (lines (1)–(10)), WMI_ACC algorithm filters most of the features with WMI value. The weight values for each of the features (line (3)) are calculated according to (4) (illustrated in Section 3.1.2). A good feature has greater mutual information values related to other features. WMI_ACC firstly calculates the value of WMI between each of the features (line (6)). However, if the value of WMI is greater than the predetermined threshold value

(line (7)), it inserts features in the list in descending order. The greater threshold value speeds up the feature selection process but decreases the classification accuracy [20]. Thereafter, in line (11), the algorithm will get the list of WMI features set.

In the second step (lines (13)–(26)), WMI_ACC algorithm selects effective features with ACC metric for a particular ML classifier. It gets the features from the desired list one by one and finds the feature that produces high ACC (accuracy) value. Exactly, from lines (13)–(16), firstly it achieves the values of ACC based on S_{wrapper} which consists of first feature list and then it takes the next feature from the list and then inserts it into S_{wrapper}. If the ACC value of new inserted feature is low, WMI_ACC algorithms remove the features from the list in line (21). Lastly, S_{wrapper} includes the effective features set.

3.1.5. Statistical Test. In more depth, to select the robust features from the selected features list of our proposed WMI_ACC algorithm and to find the significant difference among the results of the applied method, statistical tests are conducted. In this study, we executed Wilcoxon pairwise statistical test on the results of methods [38, 39]. The detailed introduction to the Wilcoxon pairwise statistical test is given as follows:

(i) *Wilcoxon test:* we also used Wilcoxon signed-rank statistical test in this research. Wilcoxon test is also a nonparametric method used for pairwise comparison between two methods [40] and is also used in many research areas [9]. If d_i is the variance between two methods' performance scores on ith out of n problem and the score is in different ranges, then it can be normalized on interval 0 and 1 in [41]. Afterwards, the variations are ranked by their absolute values and in ties practitioner will be conducted on one method as in [42]. In this case, the positive values indicate that the method performed well and vice versa.

$$R^{+} = \sum_{di>0} \text{rank}\left(d_j\right) + \frac{1}{2}\sum_{di=0} \text{rank}\left(d_i\right),$$
$$R^{-} = \sum_{di<0} \text{rank}\left(d_i\right) + \frac{1}{2}\sum_{di=0} \text{rank}\left(d_j\right). \quad (8)$$

R^{+} is used for the sum of positive values and R^{-} will be used for the sum of negative variation values. It means that if the difference between these R^{-} and R^{+} is very high, then the hypothesis will be disallowed, that is, rejected. This statistical test is also used like Friedman test to determine whether the hypothesis will be rejected or not on the specific significant values α.

4. Evaluation Methodology

This section includes traces traffic, evaluation criteria, and analysis of experimental results.

Table 1: Characteristics of HIT Trace 1 dataset.

Application	Duration time	# instances	Date
WTCP	1 hour	20512	28 April 2016
WUDP	1 hour	16400	28 April 2016
P2P	1 hour	1501	27 December 2015
IM	1 hour	7911	27 December 2016
IMAP	1 hour	15832	27 December 2015
FTP	1 hour	25251	27 December 2015

4.1. Datasets. In this paper, we select two sets of network traces for our experimental study. One dataset is our set of traces collected in our lab, while the other set is an open network trace dataset. The selected two traces are different network environment datasets. We applied our proposed feature selection algorithm on both datasets, respectively, not on only one dataset for better understanding of the composition of Internet traffic. We used two different network environment datasets, because these datasets are different from each other; for example, in our trace dataset, we capture mostly WeChat instant messaging application's traffic, while in NIMS dataset GTALK IM application's traffic is traced.

4.1.1. HIT Trace I Dataset. In this paper, we used the same dataset that we have used in our previous paper in [9]. However, for developing HIT Trace 1 dataset, we capture WeChat instant messaging (IM) application traffic. WeChat IM application includes multiple functions, but we only trace text messages, pictures messages, and audio and video calls traffic; also, in more depth, we trace IM, IMAP, and FTP applications' traffic for our research study. In this research study, we are interested in finding out the effective features for IM application traffic classification. Thus, we trace only WeChat IM application text messages, pictures messages, and audio and video calls traffic, respectively, with a Wireshark tool [43]. We captured the traffic with duration of one hour for our research study at the laboratory at School of Computer Science & Information Technology, Harbin Institute of Technology, Harbin, China, on 27 December 2015 and 28 April 2016. It should be noted that we only trace the traffic that has none zero payload packets and we are only interested in TCP and UDP traffic of WeChat IM application and P2P, IM, IMAP, and FTP traffic. In this dataset, WTCP traffic and WUDP traffic mean TCP traffic and UDP traffic of WeChat application. The detailed characteristics of HIT Trace 1 dataset are shown in Table 1.

4.1.2. NIMS Dataset. NIMS dataset includes packets collected at the research tested network. The dataset consists of SSH servers outside connection and application behaviors traffic such as DNS, HTTP, SFTP, and P2P traffic. However, we are also interested in instant messaging applications traffic classification. In this case, we also added NIMS GTALK trace traffic, which includes TCP GTALK traffic and UDP GTALK traffic. Moreover, in NIMS dataset, we select only DNS, HTTP, SFTP, GTALK TCP, and GTALK UPD traffic for

Table 2: Characteristics of NIMS dataset.

Application	# instances	Location	Date
GTALKTCP	482	Dalhousie University Network	2010
GTALKUDP	9176	Dalhousie University Network	2010
DNS	12734	Dalhousie University Network	2010
FTP	1728	Dalhousie University Network	2010
HTTP	3840	Dalhousie University Network	2010
SFTP	2269	Dalhousie University Network	2010

Table 3: Accuracy result of HIT Trace 1 dataset.

Applications	Bayes Net	Naïve Bayes	C4.5	R/Forest	Random Tree
WTCP	98.40	92.68	99.56	99.82	99.70
WUDP	100	97.75	99.56	99.87	99.93
P2P	99.95	97.78	99.87	99.93	99.95
IM	94.09	92.08	94.39	91.59	90.94
IMAP	93.31	91.81	94.17	91.42	91.05
FTP	99.98	97.53	99.85	99.98	99.95

our research work study. The detailed characteristics of NIMS data are shown in Table 2.

4.2. Performance Measures. For the evaluation performance of five machine learning (ML) classifiers/algorithms, classification accuracy, recall, and precision values are employed. All the measuring metrics are described as follows:

(i) Classification accuracy: it is the number of correctly classified traffic flows divided by total classified flows

(ii) Recall: it is the percentage of specific traffic flows Class Z correctly classified as belonging to that Class Z

(iii) Precision: it is the percentage of the traffic flows which exactly have Class Z between all those that were classified as Class Z

These performance evaluation metrics are important for flow-based traffic classification in network traffic identification. However, flow accuracy is used to measure the overall performance of an ML classifier.

4.3. Experimental Results and Analysis. In this section, our objective is to evaluate the performance of our proposed algorithm, comparing the results of HIT Trace 1 dataset with NIMS dataset's results. Our experiments include three phases. First of all, on HIT dataset and NIMS dataset, we validate that our proposed feature selection algorithm is effective for feature selection with respect to accuracy results. Then, we validate that our proposed algorithm is effective with precision and recall results. Table 3 depicts the classification accuracy results of HIT trace dataset using Bayes Net, Navies Bayes, C4.5 decision tree, Random Forest, and Random Tree machine learning algorithms. Figure 2 shows the detailed accuracy result chart of HIT Trace 1 dataset. However, we use Weka application for our experiments using training and testing method to classify IM applications traffic accurately.

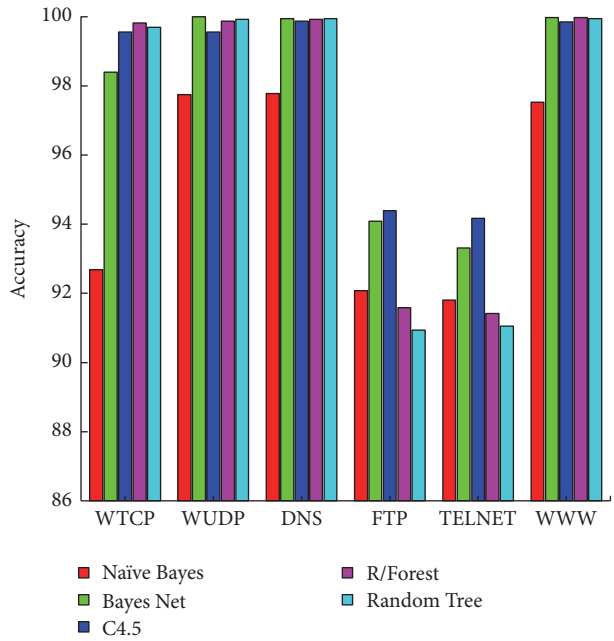


Figure 2: Accuracy result of HIT Trace 1 dataset.

It is clear from Table 3 and Figure 2 that the applied machine learning classifiers give maximum classification

Table 4: NIMS dataset's accuracy results.

Applications	Bayes Net	Naïve Bayes	C4.5	R/Forest	Random Tree
GTALK TCP	100	98.60	100	100	99.99
GTALK UDP	100	86.55	100	100	100
DNS	99.78	86.43	100	100	100
FTP	100	99.34	99.97	100	100
HTTP	100	98.58	99.97	100	99.99
SFTP	100	99.97	100	100	100

Table 5: Recall results of HIT dataset.

Applications	Bayes Net	Naïve Bayes	C4.5	R/Forest	Random Tree
WTCP	97.81	96.09	99.61	99.74	99.61
WUDP	100	92.25	100	100	100
P2P	96.10	89.61	90.91	94.81	96.10
IM	36.82	14.77	38.86	31.14	45.91
IMAP	80.91	49.90	80.32	50.89	31.61
FTP	100	89.69	92.79	98.97	99.95

accuracy results and values, but C4.5 decision tree machine learning classifier provides overall maximum classification accuracy for HIT Trace 1 dataset, 97.90%, which is effective classification accuracy value as compared to other machine learning classifiers' accuracy results. On the other side, all applications are classified with very effective values; however, FTP and Telnet applications are slightly classified when compared to other classified applications for HIT Trace 1 dataset. Telnet and FTP applications give maximum 94% accuracy results but WTCP, WUDP, and WWW applications are classified with very effective accuracy results as compared to other applications. While classifying Telnet and FTP applications, the maximum accurate results are given by Bayes Net and C4.5 machine learning classifiers when compared to other machine learning classifiers.

From these experimental results, it is evident that Random Forest and C4.5 machine learning classifiers give better performance in terms of classification accuracy as compared to other machine learning classifiers for the NIMS dataset. However, Random Forest machine learning classifier gives very effective results in terms of classification accuracy. The details are shown in Table 4 and Figure 3. In NIMS dataset, SFTP application is classified 100% as compared to other traffic applications and the applied machine learning classifiers give very accurate identification results for SFTP. The applied machine learning classifiers give very accurate performance results for NIMS dataset, but Naïve Bayes ML classifiers give slightly low accuracy results for all traffic applications except SFTP application. However, all the traffic applications are classified using five machine learning classifiers accurately.

Similarly, Bayes Net machine learning classifier gives better recall values for HIT Trace 1 dataset as shown in Table 5 and Figure 4. All the traffic applications are classified accurately with respect to recall metrics, but IM and IMAP traffic applications give very poor results with respect to recall metrics, particularly the IM application that gives very low performance results of recall metric for HIT Trace 1 dataset

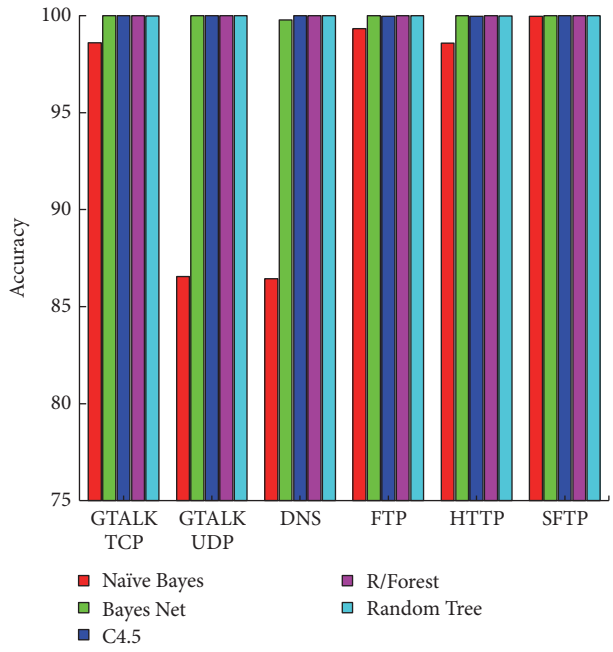


Figure 3: Accuracy result of NIMS dataset.

traffic classification, while ML classifiers C4.5, Bayes Net, and Random Tree give effective recall results. From the table, WTCP application is classified very effectively using the ML classifiers with respect to recall metrics but Random Forest ML classifier gives maximum recall results as compared to other machine learning classifiers and then Random Tree and C4.5 ML classifiers give maximum same performance recall results using WTCP application. Similarly, using WUDP traffic application, only Naïve Bayes ML classifier gives low results and all other ML classifiers give 100 recall results for WUDP traffic application. For P2P, Random Forest and Bayes Net give good recall results, while, for IM, Random Forest

Table 6: Precision result of HIT dataset.

Applications	Bayes Net	Naïve Bayes	C4.5	R/Forest	Random Tree
WTCP	99.63	92.73	99.69	99.97	99.92
WUDP	100	99.31	99.81	99.97	99.74
P2P	100	36.51	98.59	100	100
IM	60.67	37.14	63.81	35.69	36.27
IMAP	54.41	50	58.89	44.83	39.75
FTP	97.98	39.55	97.83	100	100

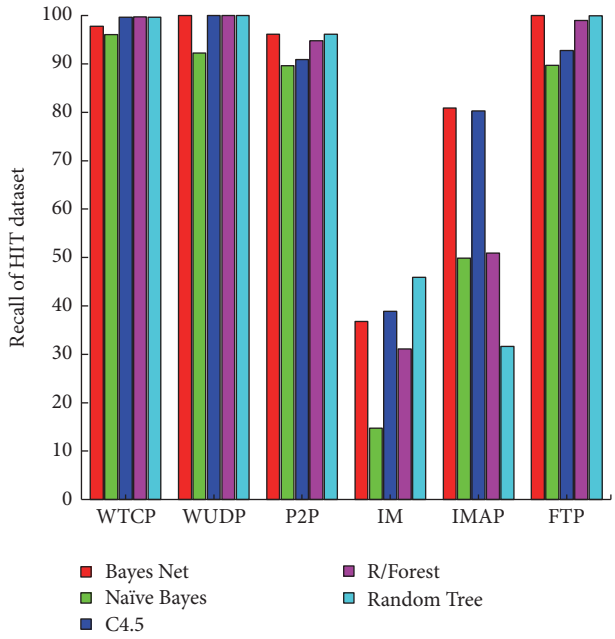


Figure 4: Recall results for HIT Trace 1 dataset.

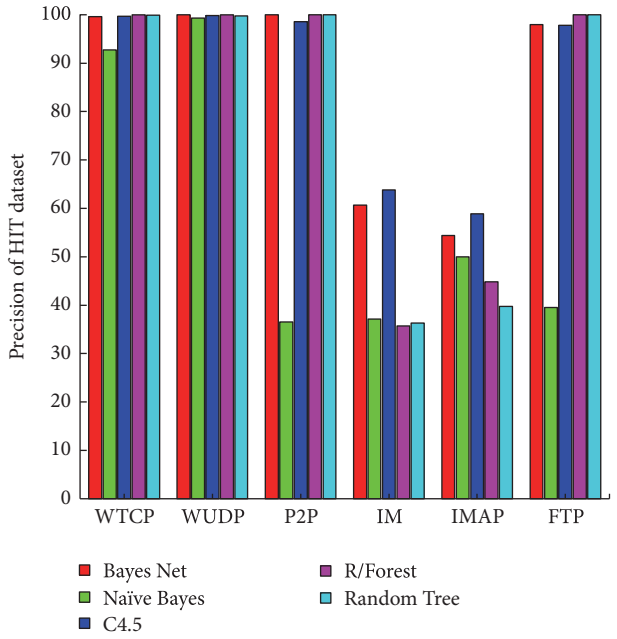


Figure 5: Precision results for HIT Trace 1 dataset.

gives maximum recall results. Similarly, for FTP, Bayes Net and Random Tree give maximum recall results.

However, for precision results, as shown in Table 6 and Figure 5, Random Forest machine learning classifier gives effective precision results for HIT Trace 1 dataset. It is clear from the experimental results using HIT Trace 1 dataset that the entire selected machine learning classifiers get high performance results values in terms of classification accuracy, recall, and precision.

Though all the traffic applications are classified very efficiently using machine learning classifiers, IM and IMAP traffic applications give very low precision performance results for HIT Trace 1 dataset as compared to other traffic applications. Using machine learning algorithms, WTCP application is classified efficiently and the applied ML classifiers give good results with respect to precision metric, which are more than 99% results. Similarly, WUDP traffic application is also classified very accurately and the applied classifiers got more than 99% precision results, but Bayes Net classifier gives 100% precision results, which are promising precision results. For P2P application, Random Forest, Random Tree, and Bayes Net ML classifiers give 100% precision results, while IM application traffic is classified very effectively and the applied classifiers do not get more than 63% precision results. Similarly, IMAP application is also poorly classified with respect to precision metric. However, FTP traffics are classified with respect to precision values and mostly classifiers got 100% precision results for HIT Trace 1 dataset.

From the experimental result of NIMS dataset, it is evident again that C4.5 and Random Forest ML classifiers give very effective precision results for NIMS dataset. However, all the applied ML classifiers give very attractive results but C4.5 and Random Forest ML classifiers' results are very promising results in terms of precision values. Using ML classifiers, all the applied machine learning classifiers give very effective precision results for GTALK TCP application, but Random Forest, C4.5, and Bayes Net ML classifiers give 100% results. For GTALK UDP application, all the applied classifiers got good precision results, but only Naïve Bayes got low precision result. Similarly, for DNS, FTP, HTTP, and SFTP, all the applied classifiers got promising precision results but only Naïve Bayes and Random Tree got slightly low precision results. However, all the precision results are good using NIMS dataset.

Recalling results for NIMS dataset, C4.5, Random Forest, and Random Tree ML classifiers give accurate results in terms

Table 7: Precision result of NIMS dataset.

Applications	Bayes Net	Naïve Bayes	C4.5	R/Forest	Random Tree
GTALK TCP	100	58.11	100	100	99.42
GTALK UDP	99.26	71.85	100	100	100
DNS	100	91.77	100	100	100
FTP	100	100	100	100	100
HTTP	100	99.58	100	100	99.92
SFTP	100	100	100	100	100

Table 8: Recall results of NIMS dataset.

Applications	Bayes Net	Naïve Bayes	C4.5	R/Forest	Random Tree
GTALK TCP	100	100	100	100	100
GTALK UDP	100	91.63	100	100	100
DNS	99.47	75.05	100	100	100
FTP	100	89.88	99.48	100	100
HTTP	100	90.60	99.97	100	100
SFTP	100	99.63	100	100	100

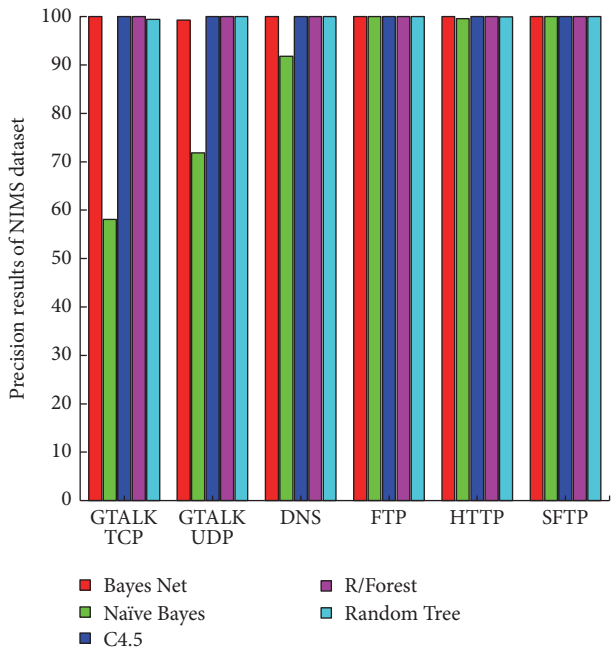


Figure 6: Precision results for NIMS dataset.

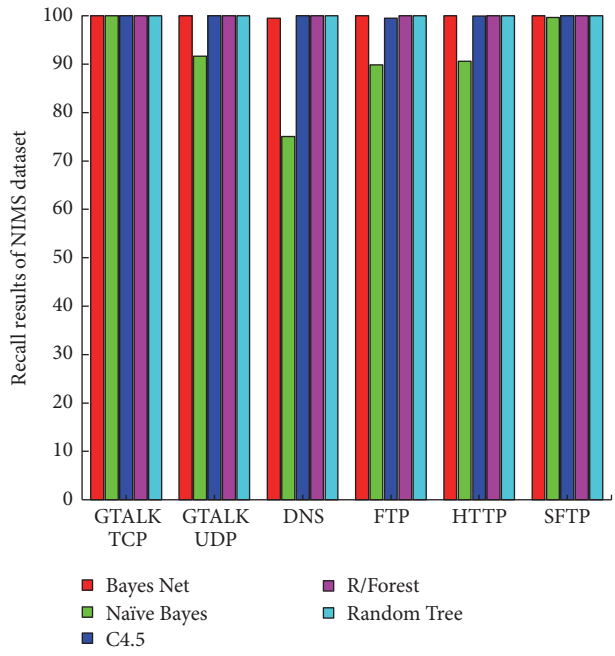


Figure 7: Recall results for NIMS dataset.

of recall for the NIMS dataset. The detailed results are shown in Tables 7 and 8 and Figures 6 and 7. Similarly, all the utilized traffic applications of NIMS dataset are classified accurately with respect to recall metric but SFTP and GTALK TCP got 100% recall results using the applied ML classifiers. From the classifiers' point of view, only Naïve Bayes ML classifier got low recall results as compared to other ML classifiers. Moreover, using ML classifiers for GTALK TCP application, the traffic is classified vey accurately as all the applied ML classifiers got 100% recall results, while for GTALK UDP only Naïve Bayes ML classifier gives slightly low recall results. For DNS application, only Naïve Bayes got poor recall value and the remaining applications got 100% recall values. Similarly, FTP, HTTP, and SFTP applications are classified accurately about 100% but only Naïve Bayes gives slightly low recall results.

4.3.1. Wilcoxon Pairwise Statistical Test Result. Table 9 shows the Wilcoxon pairwise test results for the robust features selection from the selected features of WMI_ACC proposed algorithm. From Table 9, p value of features is greater than 0.05 for the accuracy results. Thus, we conclude that there is no significant difference between the results of 9 features and other features for the selected features. We conclude that

Table 9: Wilcoxon pairwise test results.

9 Ftrs versus min_fpktl	Accuracy		
	R^+	R^-	p value
Mean_fpktl	0.00	3.00	0.180
Max_fpktl	0.00	1.00	0.317
Std_fpktl	3.00	3.00	1.000
Min_bpktl	0.00	1.00	0.317
Mean_bpktl	3.00	3.00	0.317
Max_bpktl	0.00	1.00	1.000
Std_bpktl	3.00	3.00	1.000
Max_fiat	5.00	1.00	0.276
Total_fpackets	0.00	1.00	0.317

Table 10: Selected feature of our proposed algorithm.

S. number	Feature name
1	Mean_fpktl
2	Max_fpktl
3	Std_fpktl
4	Min_bpktl
5	Mean_bpktl
6	Max_bpktl
7	Std_bpktl
8	Max_fiat
9	Total_fpackets
10	Max_fpktl

all the selected features of our proposed algorithm are very effective.

4.3.2. Selected Features of Our Proposed Algorithm. The features that are selected by our proposed algorithm are given in Table 10.

4.3.3. Average Accuracy Comparison of Datasets. In this section, we compare average accuracy results of both (HIT and NIMS) datasets to find out the effective machine learning classifier out of five applied machine learning classifiers. In Figure 8, it is clear that NIMS dataset application traffic flows are classified very effectively and all the applied ML classifiers got very promising accuracy results as compared to HIT Trace 1 dataset. However, all the applied ML classifiers got very effective average accuracy results, but Bays Net ML classifier gives maximum average performance results for HIT Trace 1 dataset as compared to other applied classifiers. Similarly, using NIMS dataset, the applied classifiers performances are very effective, but Random Forest and Random Tree ML classifiers' performance results are very effective as shown in Table 11. However, using Wilcoxon statistical test, we conclude that Random Forest ML classifier is a very effective classifier for IM applications traffic classification.

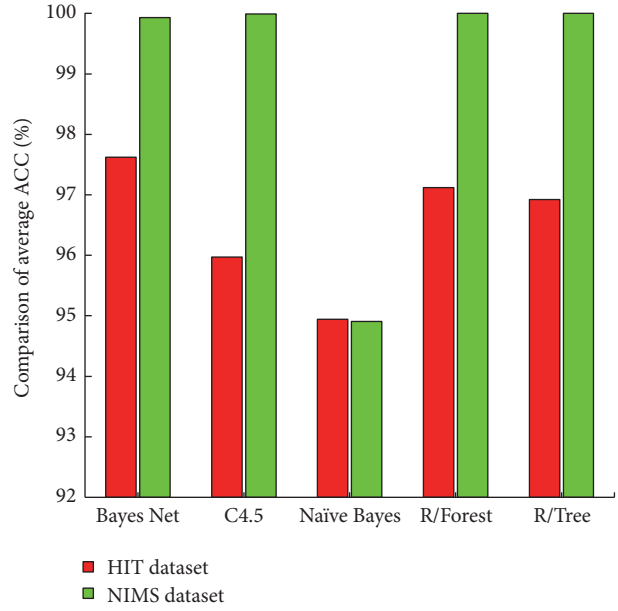


Figure 8: Datasets' average accuracy comparison.

5. Analysis and Discussion

Though the results of the five applied machine learning classifiers are different with respect to accuracy, recall, and precision using HIT Trace 1 dataset and NIMS dataset, some information can be obtained from experimental study for IM traffic classification:

(i) From this study, it is clear that our proposed algorithm selects effective features set for IM traffic classification using two different network environment datasets in terms of classification accuracy, recall, and precision metrics.

(ii) From the experimental results, all the applied machine learning classifiers give very effective performance results for all application classifications, but only FTP and Telnet applications are classified a little bit low in both utilized datasets as compared to other applications.

(iii) In this research study, our proposed algorithm gives effective features sets and it is evident that all the features carry enough identification information for IM traffic classification.

(iv) Through accuracy results, the classification performance can be easily evaluated for the instant messaging (IM) traffic classification. But, in some cases, some classifiers get high identification performance results and in some cases they do not get very effective results. It is due to imbalance traffic composition found in the datasets.

(v) We discuss that all the applied ML classifiers give very effective performance results. However, C4.5 decision tree and Random Forest ML classifiers give very accurate performance results as compared to other machine leaning classifiers.

Table 11: Datasets' average accuracy comparison.

Algorithms	Average accuracy of HIT dataset	Average accuracy of NIMS dataset
Bayes Net	97.62%	99.93%
C4.5	95.97%	99.99%
Naïve Bayes	94.94%	94.91%
Random Forest	97.12%	100%
Random Tree	96.92%	100%

6. Conclusion

This paper proposed feature selection algorithm named WMI_ACC used to select effective features for IM traffic classification. The performance of our proposed algorithm WMI_ACC is very promising for 5G IM traffic classifications. The experimental results showed that our approach is able to improve the classification accuracy, recall, and precision mostly in 5G high dimension traffic. Furthermore, ten flow-based features selected by our approach are very important for 5G IM traffic classification. They are (1) max_fpktl, (2) mean_fpktl, (3) max_fpktl, (4) std_fpktl, (5) min_bpktl, (6) mean_bpktl, (7) max_bpktl, (8) std_bpktl, (9) max_fiat, and (10) total_fpacket. Using Wilcoxon pairwise statistical test, it is evident from the experimental study that these features carry enough classification information. Moreover, all the applied ML classifiers get very effective performance results, but we found that C4.5 and Random Forest ML classifiers with WMI_ACC selected features have very effective performance as compared to other applied machine learning classifiers. In our experiments, some ML classifiers get very efficient performance results in terms of classification accuracy, recall, and precision and some ML classifiers get little bit low classification results. It is due to imbalance of dataset. However, there is still a gap for further research in the 5G instant messaging (IM) traffic classification. A new approach should be designed to select robust feature for IM applications traffic classification and this is our future research work.

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. 61571144.

References

[1] T. T. T. Nguyen and G. Armitage, "A survey of techniques for internet traffic classification using machine learning," *IEEE Communications Surveys and Tutorials*, vol. 10, no. 4, pp. 56–76, 2008.

[2] P. A. W. E. Ł. Foremski, "On different ways to classify Internet traffic: a short review of selected publications," *Theoretical and Applied Informatics*, vol. 25, 2013.

[3] A. W. Moore and K. Papagiannaki, "Toward the accurate identification of network applications," in *Proceedings of the International Workshop on Passive and Active Network Measurement*, Springer, Berlin, Heidelberg, Germany, 2005.

[4] S. Sen, O. Spatscheck, and D. Wang, "Accurate, scalable in-network identification of P2P traffic using application signatures," in *Proceedings of the 13th International World Wide Web Conference (WWW '04)*, pp. 512–521, ACM, May 2004.

[5] P. Haffner, S. Sen, O. Spatscheck, and D. Wang, "ACAS: automated construction of application signatures," in *Proceedings of the ACM SIGCOMM 1st Workshop on Mining Network Data (MineNet '05)*, pp. 197–202, ACM, August 2005.

[6] A. Moore, D. Zuev, and C. Michael, *Discriminators for Use in Flow-Based Classification*, Queen Mary and Westfield College, Department of Computer Science, 2005.

[7] H. Zhang, G. Lu, M. T. Qassrawi, Y. Zhang, and X. Yu, "Feature selection for optimizing traffic classification," *Computer Communications*, vol. 35, no. 12, pp. 1457–1471, 2012.

[8] M. Shafiq, X. Yu, and A. A. Laghari, "WeChat text messages service flow traffic classification using machine learning technique," in *Proceedings of the 6th International Conference on IT Convergence and Security (ICITCS '16)*, pp. 1–5, IEEE, Prague, Czech Republic, September 2016.

[9] M. Shafiq and X. Yu, "Effective packet number for 5G IM wechat application at early stage traffic classification," *Mobile Information Systems*, vol. 2017, Article ID 3146868, 22 pages, 2017.

[10] M. Shafiq, X. Yu, A. A. Laghari, L. Yao, N. K. Karn, and F. Abdesssamia, "WeChat text and picture messages service flow traffic classification using machine learning technique," in *Proceedings of the IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), IEEE 18th International Conference on (IEEE '16)*, 2016.

[11] L. Peng, B. Yang, and Y. Chen, "Effective packet number for early stage internet traffic identification," *Neurocomputing*, vol. 156, pp. 252–267, 2015.

[12] L. Peng, H. Zhang, B. Yang, Y. Chen, M. T. Qassrawi, and G. Lu, "Traffic identification using flexible neural trees," in *Proceedings of the IEEE 18th International Workshop on Quality of Service (IWQoS '10)*, June 2010.

[13] G. Lu, H. Zhang, X. Sha, C. Chen, and L. Peng, "TCFOM: A robust traffic classification framework based on OC-SVM combined with MC-SVM," in *Proceedings of the 2010 International Conference on Communications and Intelligence Information Security (ICCIIS '10)*, pp. 180–186, chn, October 2010.

[14] A. W. Moore and D. Zuev, "Internet traffic classification using bayesian analysis techniques," *ACM SIGMETRICS Performance Evaluation Review*, vol. 33, no. 1, p. 50, 2005.

[15] T. Auld, A. W. Moore, and S. F. Gull, "Bayesian neural networks for internet traffic classification," *IEEE Transactions on Neural Networks*, vol. 18, no. 1, pp. 223–239, 2007.

[16] D. A. Cieslak, N. V. Chawla, and A. Striegel, "Combating imbalance in network intrusion datasets," in *Proceedings of the IEEE International Conference on Granular Computing (GrC '06)*, pp. 732–737, May 2006.

[17] D. Nechay, Y. Pointurier, and M. Coates, "Controlling false alarm/discovery rates in online internet traffic flow classification," in *Proceedings of the 28th Conference on Computer Communications, IEEE (INFOCOM '09)*, pp. 684–692, April 2009.

[18] W. Li, M. Canini, A. W. Moore, and R. Bolla, "Efficient application identification and the temporal and spatial stability of classification schema," *Computer Networks*, vol. 53, no. 6, pp. 790–809, 2009.

[19] D. G. Gomes, N. Agoulmine, Y. Bennani, and J. N. de Souza, "Predictive connectionist approach for VoD bandwidth management," *Computer Communications*, vol. 30, no. 10, pp. 2236–2247, 2007.

[20] L. Peng, B. Yang, Y. Chen, and Z. Chen, "Effectiveness of statistical features for early stage internet traffic identification," *International Journal of Parallel Programming*, vol. 44, no. 1, pp. 181–197, 2016.

[21] L. Bernaille, R. Teixeira, I. Akodkenou, A. Soule, and K. Salamatian, "Traffic classification on the fly," *ACM SIGCOMM Computer Communication Review*, vol. 36, no. 2, pp. 23–26, 2006.

[22] Y.-S. Lim, H.-C. Kim, J. Jeong, C.-K. Kim, T. Kwon, and Y. Choi, "Internet traffic classification demystified: on the sources of the discriminative power," in *Proceedings of the 6th International Conference on Emerging Networking Experiments and Technologies (Co-NEXT' 10)*, December 2010.

[23] P. Van Der Putten and M. Van Someren, "A bias-variance analysis of a real world learning problem: the CoIL challenge 2000," *Machine Learning*, vol. 57, no. 1-2, pp. 177–195, 2004.

[24] Z. Zheng, X. Wu, and R. Srihari, "Feature selection for text categorization on imbalanced data," *ACM SIGKDD Explorations Newsletter*, vol. 6, no. 1, pp. 80–89, 2004.

[25] X.-W. Chen and M. Wasikowski, "FAST: a roc-based feature selection metric for small samples and imbalanced data classification problems," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '08)*, pp. 124–132, August 2008.

[26] A. H. M. Kamal, X. Zhu, A. Pandya, and S. Hsu, "Feature selection with biased sample distributions," in *Proceedings of the International Conference on IEEE Information Reuse and Integration (IRI '09)*, 2009.

[27] WeChat attracts 300m users in less than 2 years, http://www.chinadaily.com.cn/cndy/2013-01/17/content_16128915.htm.

[28] By the numbers: 50+ Amazing WeChat Statistics, http://expandedramblings.com/index.php/wechat-statistics/.

[29] Q. Huang, P. P. C. Lee, C. He, J. Qian, and C. He, "Fine-grained dissection of WeChat in cellular networks," in *Proceedings of the 23rd IEEE International Symposium on Quality of Service (IWQoS '15)*, pp. 309–318, Portland, Ore, USA, June 2015.

[30] K. Church and R. de Oliveira, "What's up with whatsapp?: comparing mobile instant messaging behaviors with traditional SMS," in *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '13)*, pp. 352–361, ACM, Munich, Germany, August 2013.

[31] K. P. O'Hara, M. Massimi, R. Harper, S. Rubens, and J. Morris, "Everyday dwelling with whatsApp," in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1131 pages, ACM, Baltimore, Md, USA, February 2014.

[32] P. Fiadino, M. Schiavone, and P. Casas, "Vivisecting WhatsApp through large-scale measurements in mobile networks," in *Proceedings of the ACM conference on (SIGCOMM '14)*, pp. 133-134, usa, August 2014.

[33] Y. Liu and L. Guo, "An empirical study of video messaging services on smartphones," in *Proceedings of the 24th ACM Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV '14)*, pp. 79–84, ACM, March 2014.

[34] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[35] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 187–198, 1997.

[36] L. R. Bahl, P. Brown, P. de Souza, and R. Mercer, "Maximum mutual information estimation of hidden Markov model parameters for speech recognition," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '86)*, Tokyo, Japan, April 1986.

[37] H. Peng, Mutual information Matlab toolbox, http://www.mathworks.com/matlabcentral/fileexchange/14888-mutual-information-computation.

[38] S. García, A. Fernandez, J. Luengo, and F. Herrera, "Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: experimental analysis of power," *Information Sciences*, vol. 180, no. 10, pp. 2044–2064, 2010.

[39] F. Gringoli, L. Salgarelli, M. Dusi, N. Cascarano, F. Risso, and K. C. Claffy, "GT: picking up the truth from the ground for internet traffic," *ACM SIGCOMM Computer Communication Review*, vol. 39, no. 5, pp. 12–18, 2009.

[40] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, CRC Press, 2003.

[41] D. Quade, "Using weighted rankings in the analysis of complete blocks with additive block effects," *Journal of the American Statistical Association*, vol. 74, no. 367, pp. 680–683, 1979.

[42] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference*, Springer, Berlin, Heidelberg, Germany, 2011.

[43] Trace traffic WireShark, http://www.wireshark.org/ (last Accessed Sept 2015).

Hindawi
Mobile Information Systems
Volume 2017, Article ID 1750637, 13 pages
https://doi.org/10.1155/2017/1750637

Research Article

An Automata Based Intrusion Detection Method for Internet of Things

Yulong Fu,[1] Zheng Yan,[2,3] Jin Cao,[1] Ousmane Koné,[4] and Xuefei Cao[1]

[1]*School of Cyber Engineering, Xidian University, Xian, China*
[2]*Aalto University, Espoo, Finland*
[3]*The State Key Lab of ISN, Xidian University, Xian, China*
[4]*University of Pau and Academy of Bordeaux, Mont-de-Marsan, France*

Correspondence should be addressed to Yulong Fu; ylfu@xidian.edu.cn

Received 25 January 2017; Revised 12 March 2017; Accepted 28 March 2017; Published 2 May 2017

Academic Editor: Jing Zhao

Internet of Things (IoT) transforms network communication to Machine-to-Machine (M2M) basis and provides open access and new services to citizens and companies. It extends the border of Internet and will be developed as one part of the future 5G networks. However, as the resources of IoT's front devices are constrained, many security mechanisms are hard to be implemented to protect the IoT networks. Intrusion detection system (IDS) is an efficient technique that can be used to detect the attackers when cryptography is broken, and it can be used to enforce the security of IoT networks. In this article, we analyzed the intrusion detection requirements of IoT networks and then proposed a uniform intrusion detection method for the vast heterogeneous IoT networks based on an automata model. The proposed method can detect and report the possible IoT attacks with three types: jam-attack, false-attack, and reply-attack automatically. We also design an experiment to verify the proposed IDS method and examine the attack of RADIUS application.

1. Introduction

Due to the rapidly advancing technologies of network communication, the Internet is going to connect everything from everywhere. New concept of Internet of Things (IoT) appears and is associated with the future Internet of 5G. IoT connects a large number of heterogeneous devices, such as "instance cameras," "wireless sensor network" (WSN), "smart meters," and "vehicles," while providing open access to a variety of data generated by such devices to provide new services to citizens and companies [1]. However, as the resources of IoT's front devices are constrained, many security mechanisms are hard to be implemented to protect the IoT networks. Some lightweight encryption methods are considered as the core technology to build the security mechanism of IoT [2], but considering the increments of the hacker's computation ability (the usage of Cloud Computing, Distributed Computing, Quantum computation, etc.), those lightweight cryptography methods are going to be crushed in the foreseeable future. Other kinds of security enforcement methods, such as intrusion detection system should be considered to protect the IoT networks [3].

Intrusion detection system (IDS) is an efficient technique to detect attackers when cryptography is broken [4]. It can detect malicious activities or policy violations by monitoring the network traffics or system actives [5]. IDS is normally a stand-by device or third-part software which will not inquire many changes to the current system. It is suitable for the resource constrained or inherited systems to protect their network security.

Many recent works have noticed the security problem of IoT system, and a number of intrusion detection methods are proposed and developed, such as [4, 6–10]. However, most of the proposed methods are still limited to data mining and can only give an intrusion view of WSN, MANET, Zigbee, or other subnets of IoT, and a uniform intrusion detection method for the whole IoT networks is rarely discussed. Meanwhile, as the network packets digging and statistic

feature training usually require many computation resources, such methods are hard to be implemented in some cases of IoT environments.

In this article, we present an automata based intrusion detection method for the networks of Internet of Things. Our method uses an extension of Labelled Transition Systems to propose a uniform description of IoT systems and can detect the intrusions of IoT networks. The used automata model can describe the combination of heterogeneous networks with terms and graphs, and the proposed IDS structure and algorithm can detect the intrusions by comparing the abstracted actions flows, which can solve the aforementioned problems.

Paper Contribution. By using automata theory, many complicated problems can be described and solved. In this article, we use an extension of Input Output Labelled Transition System to solve the uniform description problem of the heterogeneous IoT networks and propose a corresponding intrusion detection mechanism for IoT network. To achieve this purpose, a set of procedures including collected data grouping, packet data translation, anomaly data detection, and intrusion classification are designed and proposed. Comparing with the existing methods, the benefits of our work can be listed as below:

(1) To our knowledge, this is the first time of using automata theory to model and detect the intrusions of IoT networks. By using the proposed automata methods, we can map the IoT system to an abstract space, where a uniform security evaluation structure can be built.

(2) We defined and proposed a set of intrusion detection mechanisms by using the proposed automata method.

(3) We developed a GUI tools to automatically analyze and graphically present the abstract action flows and to detect the possible intrusions.

(4) We also analyzed and classified the detected intrusions, and three kinds of attacks, including replay-attack, jam-attack, and fake-attack, can be distinguished in our method.

The following sections are organized as below: In Section 2, the background, problem description, and related works of developing the IDS system over IoT are discussed. In Section 3, the entire approach of the automata based intrusion detection method will be described. In Section 4, to illustrate the use of the proposed IDS methods, we present an example of using the proposed method to analyze a simplified IoT system, and the results demonstrate the correctness of our method. And finally, in Section 5, we conclude this work and discuss some possible future works.

2. Background, Problems, and Related Works

2.1. Internet of Things and Its Security

2.1.1. Internet of Things. IoT is the network of things, with clear element identification, embedded with software intelligence, sensors, and ubiquitous connectivity to the Internet [11]. IoT enables things or objects to exchange information with the manufacturer, operator, and other connected devices utilizing the telecommunications infrastructure of the Internet. It allows physical objects to be sensed (by providing the specific information such as the RFID tags and QR code) and controlled remotely across the Internet. IoT will create opportunities for more direct integration between the physical world and computer-based systems, resulting in improved efficiency, accuracy, and economic benefit, for example, monitoring and controlling things by experts such as telemedicine and searching for things (keys, passports) directly that search engines do not provide today.

Normally, three basic elements should be included by an IoT system: the unique identity per thing (e.g., IP address), the ability to communicate between things (e.g., wireless communications), and the ability to sense specific information about the things (sensors) [11]. Therefore, for an IP based system, the IoT gateway is a good solution to form the IoT networks. The IEEE 802.15 Task Group 4 has defined the personal area network (PAN) coordinator to take in charge of the network domain. The PAN allocates local addresses and acts as a gateway to other domains or networks [12]. IEEE 802.15.4 also defined two types of IoT devices: the full-function device (FFD), which implements all of the functions of the communication stack and allows it to communicate with any other device in the network; and the reduced-function devices (RFDs) which are meant to be extremely simple devices with very modest resource and communication capabilities. Hence, RFDs can only communicate with FFDs and can never act as PAN coordinators.

2.1.2. IoT Security Attacks. Considering the specific features of IoT networks, we found that the following three kinds of attack scenarios likely happen in the real world and are important to be studied.

(i) Attack Scenario 1. For a given IoT network, such as the one presented in Figure 1, an authorized user, User1, may want to control the specific device in the IoT. The user needs to use the IoT networks to find the right device and to communicate with the device. For some security reason, the IoT device has to verify the authentication of User1. During this process, a cryptography method is normally needed to verify the authentication and to protect against the malicious attacks. However, a malicious user, User2, may be able to listen the communication between User1 and the corresponding IoT device. User2 may fake himself as User1 and create a replay-attack to the IoT system. To solve such problem, the RFD may ask FFD or PAN to help him to verify the authentication of the user and record the passed IDs of the user. A group authentication protocol and cryptography functions can help RFD to protect itself from such kind of attack. However,

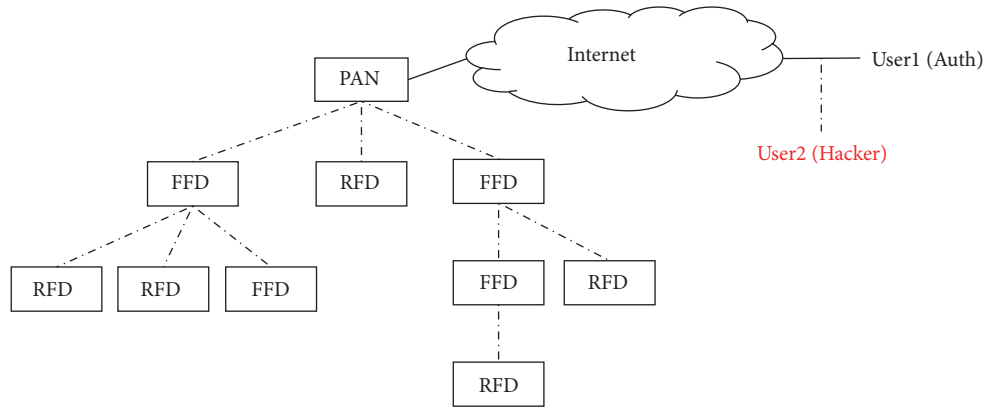


Figure 1: Attack scenario 1.

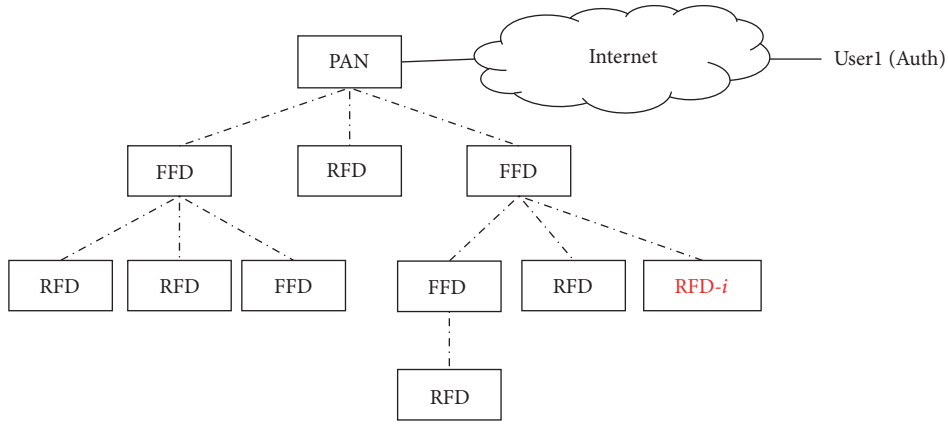


Figure 2: Attack scenario 2.

the FFD is also a resource constrained device, and the communication delay and calculation consuming will be too much for him to hold.

(ii) Attack Scenario 2. As most of the IoT networks are not closed, a malicious device may be able to present its willingness to join the IoT networks. For example, in Figure 2, a powerful device RFD-*i* (such devices can listen the communication channel of IoT devices), which is controlled by an attacker, may want to join the IoT network. Such powerful device can detect the communication information on the IoT networks and can execute many kinds of attacks such as DoS/DDoS to the corresponding FFD or PAN. Simply using the cryptography methods on IoT device will be hard to defense this kind of attacks.

(iii) Attack Scenario 3. Because the structure of IoT networks is dynamic, some authorized IoT device may be captured by the attacker. The attacker then can modify some functions or inject some virus and trojans to such device. Then the attacker can put such compromised devices to rejoint the IoT networks (see Figure 3). Because the device will be still recognized by the IoT system, it will pass the security verification of IoT network. This kind of attack is also difficult to be protected through the cryptography methods.

As we can see, by simply using the cryptography methods, some kinds of attack are hard to be detected in IoT networks. Although the usage of some complex security protocols may be able to achieve the security goals of IoT, they are hard to be implemented on the resource constrained IoT devices. Other ways of defensing the security of the system, such as the usage of intrusion detection system, should be considered for IoT network security.

2.2. Intrusion Detection System. The concept of intrusion detection was first proposed by Anderson in the year of 1980 [13] and is introduced to network system by Heberlein in 1990 [14]. After 2 decades of developing, the researches on IDS are becoming mature and have helped the industries to protect their system security for many years. An IDS may be either host or network-based [15]. A host based IDS analyzes events mainly related to OS information, while a network-based IDS analyzes network related events, such as traffic volume, IP addresses, and service ports. Meanwhile, according to the way of detecting the intrusion, two main categories of IDS are usually discussed: misuse IDS and anomaly IDS. The former uses the traces or templates of the known attacks, while the latter builds profiles of nonanomalous behaviors of computer system's active subjects. For example, IDIOT [16]

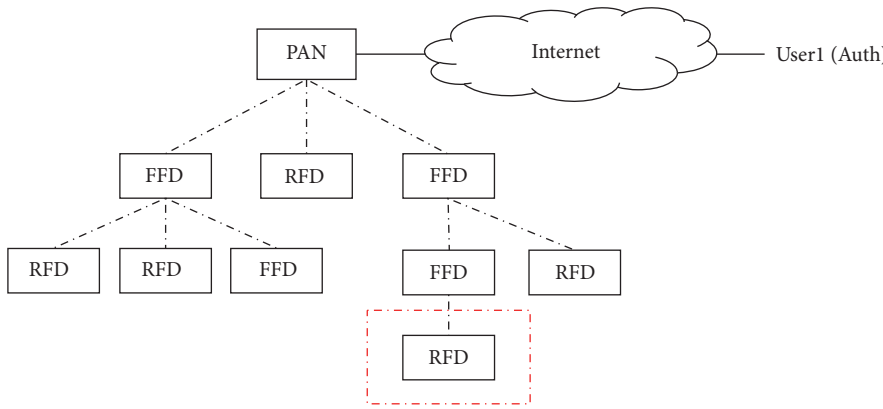


Figure 3: Attack scenario 3.

and STAT [17] use patterns of well-known attacks or weak spots in the system to match and identify known intrusions. The main advantage of misuse IDS is that it can accurately and efficiently detect instances of known attacks. The principal disadvantage is that it lacks the ability to detect the truly innovative attacks. On the other hand, anomaly IDS [18] does not require prior knowledge of intrusion and can thus detect new intrusions. But it may not be able to describe what the attack is and may have a high false positive rate.

An IDS normally contained four major components: Event Monitor, Event Database, Event Analyzer, and Response Unit [19]. The Event Monitor is responsible for detecting the system or environment actives and converts them as some specific formats and store them in the Event Database. The Event Analyzer retrieves the modeled actives from the Event Database and analyzes them in order to detect the intrusions. Once the unusual actives are detected, the Response Unit produces reports to a management station to warn a risk. IDS focuses on detecting and preventing the intrusive activities, which were not detected by conventional system security mechanisms. For some inherited systems, because of some historical or economic reasons, some powerful security mechanisms are hard to be deployed. However, the IDS can be used to solve this problem, because it needs nothing to change the target system.

2.3. Existing Intrusion Detection Works on IoT Networks. In recent years, along with the development of Internet of Things, Intelligent Hardware, and Virtual Reality, the intrusion detection method under IoT has become a trend in the development of information technology. However, the researches on such problem are still in its infancy. As IoT can be thought of as a vast heterogeneous network, most of the existing works began to study the components of IoT to find a suitable intrusion detection method. In [1], based on the use of Game Theory, Sedjelmaci et al. proposed a hybrid intrusion detection method, which mixed the usage of signature and anomaly ways for IoT intrusion detection. By creating the game model of intruder and normal user, the Nash Equilibrium Value was calculated and was used to decide when to use the intrusion detection method of anomaly. In [20], J. Chen and C. Chen proposed a real-time pattern matching system for IoT devices by using the Complex Event Processing (CEP). The advantage of this method is that it uses the features of the events flows to judge the intrusions, which can reduce the false alarm rate comparing with the traditional intrusion detection methods. Although this method will increase the consumption of system computing resources, it can obviously reduce the feedback delay of the IDS system. In [7], Nadeem and Howarth summarized the intrusion detection methods for MANET, which is one kind of network structure of the IoT. By analyzing and comparing the attack methods and detection algorithms of MANET, this paper analyzes the existing CRADS, GIDP, and other intrusion detection frameworks for MANET.

Although these existing methods can solve the intrusion detection problems of IoT from different levels, a uniform intrusion detection method is still needed to give an entire intrusion view of the IoT networks. As what have been pointed by Gendreau and Moorman in their survey of [10], the research of intrusion detection system for IoT system should focus on solving the problems of "lacking complete interoperability between different IoT parts."

3. An Automata Based Intrusion Detection Approach for IoT Security

In order to give a complete intrusion view for the different cases of IoT networks, a uniform intrusion detection method is required. In this article, by using the proposed automata model, we can project the different cases of IoT to an abstract algebra space, where a uniform security evaluation structure can be built. Meanwhile, in the real word of IoT system, by adopting a data collector and analyzing the transmitting packets, the real-time actions flows of the IoT networks can be achieved and translated into the formal format of automata. Then by comparing the real-time action flows with the anomaly or standard libraries, we can detect the intrusions of IoT quickly and solve the aforementioned problems.

3.1. The Automata Model. A finite automata (or finite state machine) [21] can present the network system with a finite

number of states and transitions, where the states represent the current status of the device and the transitions represent the active actions between different states. The current state changes only if it receives the corresponding actions. An Input/Output Labelled Transition System (IOLTS) [22] is a special case of automata, which emphasizes the input and output interactions of the system. An IOLTS system can be presented as a 4-tuple algebra set $\langle S, L, T, s_0\rangle$, where S represents a countable, nonempty set of states; L represents a countable set of labels; T represents the set of transition relations, $T \subseteq S \times (L \cup \{\tau\}) \times S$ (here, τ represents an internal action of the system that will not be achieved from outside); and s_0 is the initial state. Notice that L contains two subsets: input label L_I and output label L_O ($L_I \cap L_O = \emptyset$, $L_I \cup L_O = L$). If $s \in S$, then we denote $\mathrm{In}(s)$ and $\mathrm{Out}(s)$ to represent the set of input and output labels of state s. A transition is denoted as $s_i \xrightarrow{!l} s_j$, where $s_i, s_j \in S$ and $l \in L$. The symbol ! or ? representing l is an output label or input label, respectively. IOLTS can be used to describe an interactive system and can present the system with a graphic view. However, as the IoT networks contain multiple components, an extension of IOLTS, the Glued-IOLTS [23], is needed to present the networked system.

In a Glued-IOLTS, in order to describe the communication medium between different components, a normal state $s \in S$ of IOTS(L) is defined as the following two levels:

(i) higher_level state s_i_u, which connects to the environment or other states of the same component;

(ii) lower_level state s_i_l, which connects to the states of other components.

And then, the communication medium can be defined by such transition, which begins from the lower_level state of one component and ends with the lower_level of initial state of another component. If we use S_i and L_i to denote the states and labels in IOTS(L_i) and S_j and L_j to denote the state and labels in IOTS(L_j), then if $\exists !l \in L_i$, $\exists s_i \in S_i$, $!l \in \mathrm{Out}(s_i)$, and $\exists s_j \in S_j$, $?l \in L_j$, $?l \in \mathrm{In}(s_j)$. The transition of the common medium between IOTS(L_i) and IOTS(L_j) is presented as $s_i\_l \xrightarrow{!l} s_0\_l$. We use S_{medium} and T_{medium} to denote the states and transitions in the medium, and we give the definition of Glued-IOLTS as below.

Definition 1 (Glued-IOLTS). A Glued-IOLTS represents a set of IOLTS $\langle S_i, L_i, T_i, s_{i0}\rangle$ ($i = 1, \ldots, n$) and a medium M, which is still a 4-tuple system $\langle S_{\mathrm{glu}}, L_{\mathrm{glu}}, T_{\mathrm{glu}}, s_{\mathrm{glu}_0}\rangle$, where

(i) $S_{\mathrm{glu}} = \langle S_1 \cup S_2 \cup \cdots \cup S_n \cup S_M\rangle$,

(ii) $L_{\mathrm{glu}} = \langle L_1 \cup L_2 \cup \cdots \cup L_n\rangle$,

(iii) $s_{\mathrm{glu}_0} = \langle s_1\_0, s_2\_0, \ldots, s_n\_0\rangle$ is the initial state,

(iv) $T_{\mathrm{glu}} \subset S_{\mathrm{glu}} \times L_{\mathrm{glu}} \times S_{\mathrm{glu}}$,

$$T_{\mathrm{glu}} = \Big\{\big(s_1, s_2, \ldots, s_i, \ldots, s_m\big) \xrightarrow{\alpha} \big(s_1, s_2, \ldots, s_i', \ldots, s_m\big) \mid \big(s_i, \alpha, s_i'\big) \in T_i \cup T_M\Big\}, \quad (1)$$

$$T_M = \Big\{\big(s_{il}, \mu, s_{jl}\big) \mid i \neq j, \mu \in \mathrm{Out}\big(s_{il}\big) \cap \mathrm{In}\big(s_{jl}\big)\Big\}.$$

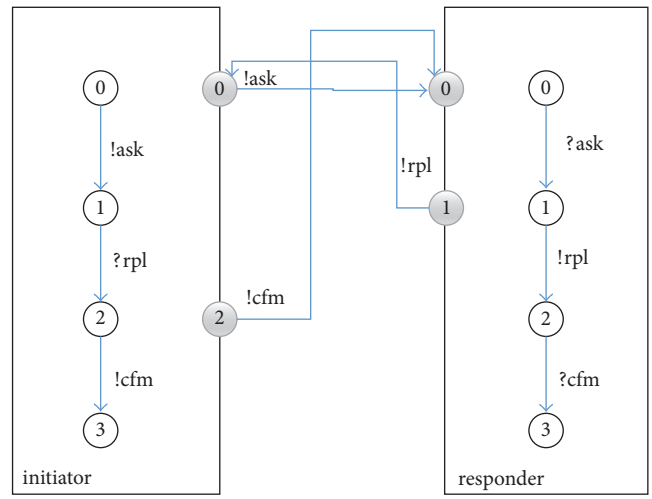


Figure 4: Glued-IOLTS of NSPK.

Example 2. The Needham-Shroeder Public Key (NSPK) protocol [24] is an asymmetric cryptography based authentication protocol, which defines the handshakes between two participations: the initiator i and the responder r. The brief protocol narrations can be presented with the three-message exchanging as below:

Msg 1 (Ask). $i \rightarrow r$: $\{n_i, i\}_{\mathrm{pk}_r}$

Msg 2 (Rpl). $r \rightarrow i$: $\{n_i, n_r\}_{\mathrm{pk}_i}$

Msg 3 (Cfm). $i \rightarrow r$: $\{n_r\}_{\mathrm{pk}_r}$

A networked security system implementing the NSPK protocol can be described and modeled with the Glued-IOLTS, and the result is presented in Figure 4.

3.2. Intrusion Detection Approaches of IoT Networks. Although the proposed automata model can be used to describe the communications of an IoT system and can make the comparison of different subnets of IoT become possible, to adopt this model into an intrusion detection system, a set of cooperated devices and some existing approaches are also needed. Just like the general IDS system, the proposed automata based IDS of IoT networks also consist of four major components: Event Monitor, Event Database, Event Analyzer, and Response Unit. A general view of the proposed IDS can be presented in Figure 5. In this article, although the four components are developed in our system, our description will mainly focus on the Event Analyzer and Response Unit.

3.2.1. Event Monitor. For the purpose of collecting the data traffics through the IoT network, a network collector (the component labelled with C in Figure 5) should be implemented on the PAN coordinator or other IoT gateways to monitor the network traffic. Such collector will be embedded software or hardware to obtain the received and sent packets through the network device. The collector needs to record the transmitting data into digital files and send the files to the IDS Event Analyzer.

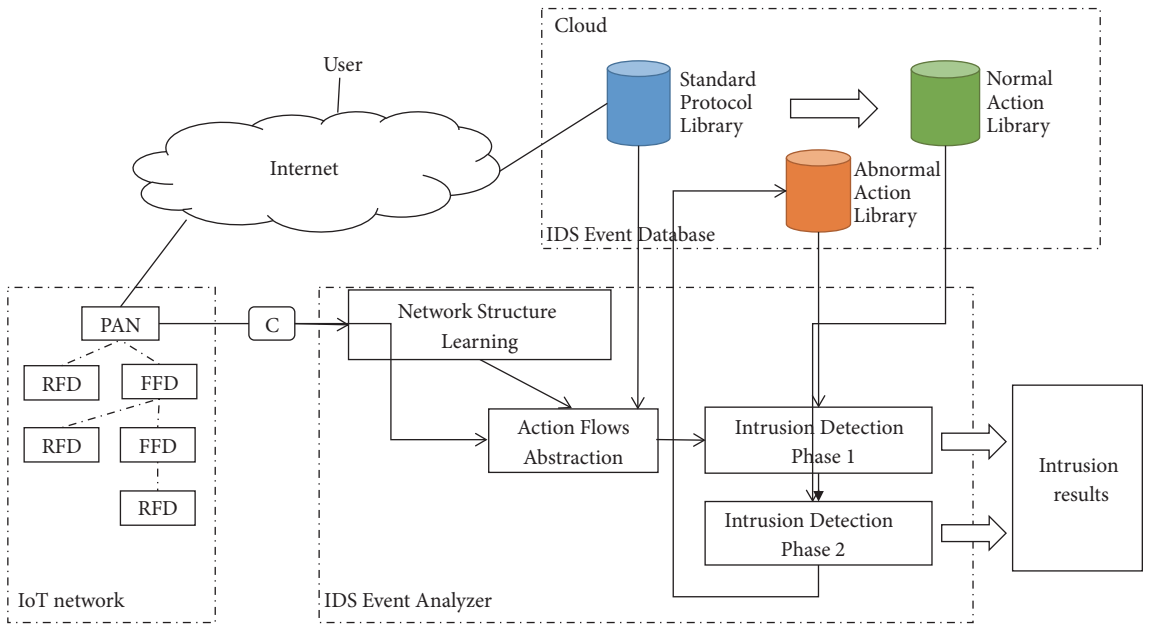


Figure 5: IDS structure.

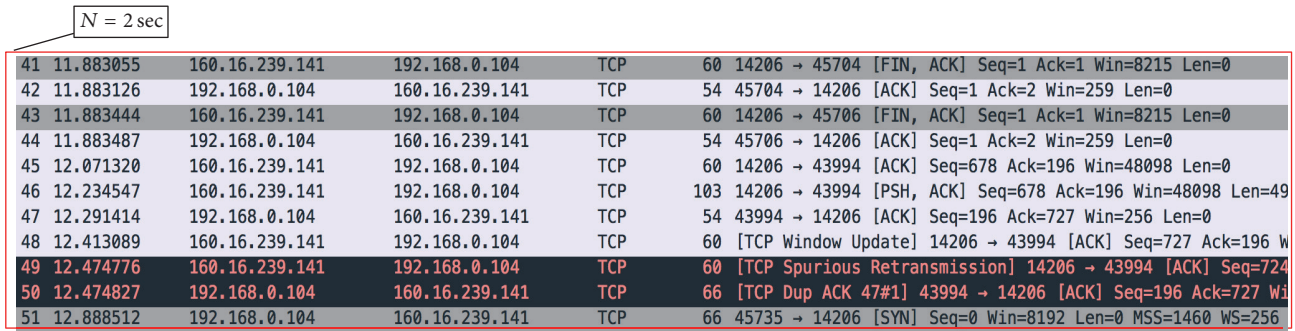
3.2.2. Event Database. In our method, the network event is described as the abstract action flows, and such network actions are described with transitions of the proposed Glued-IOLTS model. Three databases should be implemented in our IDS: Standard Protocol Library, Abnormal Action Library, and Normal Action Libraries are required. The Standard Protocol Libraries store the description of the standard protocols through Glued-IOLTS. The Normal Action Libraries store the possible action flows which are created from the Standard Protocol Libraries. The Abnormal Action Libraries store the recognized anomaly actions flows for the system. These three databases should be stored on the cloud and can be visited directly by the Event Analyzer.

3.2.3. Event Analyzer. The IDS Event Analyzer is an important part of our IDS system. It contains three basic models: Network Structure Learning Model, Action Flows Abstraction Model, and Intrusion Detection Model.

(i) Network Structure Learning Model. In our method, the collected packet data should be sent to this model first to make the IDS system get a general view of the network topologies. As the IoT devices can be distinguished with the unique ID, by analyzing the collected information of the data packets, such as the source IP, destination IP, port number, timestamp, and protocol type, we can distinguish the IoT devices from the others. For example, because the IoT devices are usually connected to the same IoT gateway, the first three fields of the IPv4 address of such devices will be the same. In this case, by counting the frequency of each IPv4 field, we can achieve the IP segment of the IoT devices. These unique IDs of the IoT devices will be recorded and sent to the Action Flows Abstraction Model.

(ii) Action Flows Abstraction. The collected real-time packets from IoT also need to be sent to the Action Flows Abstraction Model. Through this model, the packets will be allocated according to the device belonging, session ID, timestamps, and protocol types which are recognized through the aids of Network Structure Learning Model and the Standard Protocol Library. Through the information detected, the network traffics can be classified into message sequences. However, if the IoT serves multiple customers, different sessions may happen in parallel, which may make the messages become hard to be distinguished. In this article, we assume that the network connections from different services happen sequently; then by using one selected window size N, by comparing the other detected information, such as IP address, protocol type, and info (see Figure 6), we can allocate the packets to be the message sequence. The selected window size N relates to the efficiency of the Event Analyzer. The greater the value of N is selected, the more accurate the sequence detection is. But at the same time, it also means more memory and computing times consuming. We suggest N should be considered bigger than the amount of messages which happened during one session of the protocol specification and less than the whole detected messages space of the Event Monitor.

After we can allocate the packets to be message, we need to translate these messages to abstract action flows. To do this, the help from the Standard Protocol Library is needed. From the results of the message allocation, together with the protocol type information of each packet, we can know the main protocol type of such selected message. Then after we get the protocol type of the selected message, we can search for the basic formal action primitives from the Standard Protocol Library. And by comparing with the Info information of each packet, we can represent the packets



41	11.883055	160.16.239.141	192.168.0.104	TCP	60	14206 → 45704 [FIN, ACK] Seq=1 Ack=1 Win=8215 Len=0
42	11.883126	192.168.0.104	160.16.239.141	TCP	54	45704 → 14206 [ACK] Seq=1 Ack=2 Win=259 Len=0
43	11.883444	160.16.239.141	192.168.0.104	TCP	60	14206 → 45706 [FIN, ACK] Seq=1 Ack=1 Win=8215 Len=0
44	11.883487	192.168.0.104	160.16.239.141	TCP	54	45706 → 14206 [ACK] Seq=1 Ack=2 Win=259 Len=0
45	12.071320	160.16.239.141	192.168.0.104	TCP	60	14206 → 43994 [ACK] Seq=678 Ack=196 Win=48098 Len=0
46	12.234547	160.16.239.141	192.168.0.104	TCP	103	14206 → 43994 [PSH, ACK] Seq=678 Ack=196 Win=48098 Len=49
47	12.291414	192.168.0.104	160.16.239.141	TCP	54	43994 → 14206 [ACK] Seq=196 Ack=727 Win=256 Len=0
48	12.413089	160.16.239.141	192.168.0.104	TCP	60	[TCP Window Update] 14206 → 43994 [ACK] Seq=727 Ack=196 W
49	12.474776	160.16.239.141	192.168.0.104	TCP	60	[TCP Spurious Retransmission] 14206 → 43994 [ACK] Seq=724
50	12.474827	192.168.0.104	160.16.239.141	TCP	66	[TCP Dup ACK 47#1] 43994 → 14206 [ACK] Seq=196 Ack=727 Wi
51	12.888512	192.168.0.104	160.16.239.141	TCP	66	45735 → 14206 [SYN] Seq=0 Win=8192 Len=0 MSS=1460 WS=256
52	13.123550	160.16.239.141	192.168.0.104	TCP	66	14206 → 45735 [SYN, ACK] Seq=0 Ack=1 Win=63443 Len=0 MSS=
53	13.123624	192.168.0.104	160.16.239.141	TCP	54	45735 → 14206 [ACK] Seq=1 Ack=1 Win=66560 Len=0
54	13.123744	192.168.0.104	160.16.239.141	TCP	1089	45735 → 14206 [PSH, ACK] Seq=1 Ack=1 Win=66560 Len=1035
55	13.354638	160.16.239.141	192.168.0.104	TCP	60	14206 → 45735 [ACK] Seq=1 Ack=1036 Win=1048576 Len=0
56	13.430259	192.168.0.104	14.18.245.211	TCP	66	45736 → 80 [SYN] Seq=0 Win=65535 Len=0 MSS=1460 WS=256 SA
57	13.471336	14.18.245.211	192.168.0.104	TCP	66	80 → 45736 [SYN, ACK] Seq=0 Ack=1 Win=14400 Len=0 MSS=144
58	13.471442	192.168.0.104	14.18.245.211	TCP	54	45736 → 80 [ACK] Seq=1 Ack=1 Win=262144 Len=0
59	13.471571	192.168.0.104	14.18.245.211	HTTP	1022	POST /cgi-bin/qqshow_user_props_info HTTP/1.1 (applicati
60	13.523540	14.18.245.211	192.168.0.104	TCP	60	80 → 45736 [ACK] Seq=1 Ack=969 Win=16384 Len=0

Figure 6: Example of selecting $N = 2$ sec.

to be the automata primitives. Then the abstracted action sequences can be achieved. For example, the selected message in Figure 7 can be translated as [?FIN, !ACK, ?ACK + FIN, !ACK, ?ACK, ?PSH, !ACK, ?UPDATE, !SYN] through the processes presented in Figure 7.

(iii) Intrusion Detection. The result of the Action Flows Abstraction Model will be the list of automata transition sequence of the target system. Such transition sequences are then taken as the input to the intrusion verification part. In our method, we have two phases of intrusion verification.

Intrusion Detection Phase 1. The results of Action Flows Abstraction Model are used to be checked with an Abnormal Action Library, which is stored in the Event Databases. This library is a predefined database that is stored on the cloud next to the IoT system (Fog Computing [11]). If the transition sequence matches with the one stored in the Abnormal Action Library, we remark such message as an intrusion and output it as the result of the intrusion detection system. If the input sequence does not match any stored sequences in the Abnormal Action Library, the action flows go to the second phase of the intrusion detection.

Intrusion Detection Phase 2. In the second phase of intrusion, an anomaly detection method will be used to check the intrusion. In this phase, a Normal Action Library will be used to check whether the input transition sequence is a normal one. The Normal Action Library is generated from the Standard Protocol Library, by using the techniques of Fuzzing [25] and Robustness Testing [26]. If the comparing results show that the input sequence is abnormal, we take such message as a suspected one and ask for a manual verification from the experts to avoid the false positive. If the suspected transition sequence is confirmed as intrusion by the experts, we then record such message into the Abnormal Action Library and use it for the next time of intrusion detection. The method of verifying transition sequences in the Normal Action Library is to find the *walk* in the Glued-IOLTS graph of the library. During the verification process, we may need to adapt some past transitions into the detected sequence to complete the walk in Glued-IOLTS; for the detailed algorithm, please check [27]. After doing this, if the transition sequence can find the corresponding walk, it means the detected messages traffics are normal messages. Otherwise, message traffic contains some possible attacks to the system.

3.2.4. Response Unit. The Response Unit produces reports to a management station to warn an intrusion risk to the IoT networks. In the report, the following three types of attacks are going to be classified, which correspond to the attack scenarios presented in Section 2.

(i) Replay-attack: this attack corresponds to the aforementioned attack scenario 1. In this kind of attack scenario, the attacker can listen the communication between an authenticated user and the IoT device; then the attacker uses the transition which happened to attack the system. This kind of attacks can be distinguished by our IDS because the corresponded transition sequence can not be found in the normal library. The *walk* will stop at an inopportune transition, and also this transition can be found in the past transitions.

(ii) Jam-attack: this attack corresponds to the aforementioned attack scenario 2. In this kind of attack, the powerful attacker can detect the communication information on the IoT networks and can execute attacks such as DoS/DDoS to the corresponding FFD or PAN to block the communication channel. In this case, on our IDS system, after translating the collected messages into automata transition sequences, the

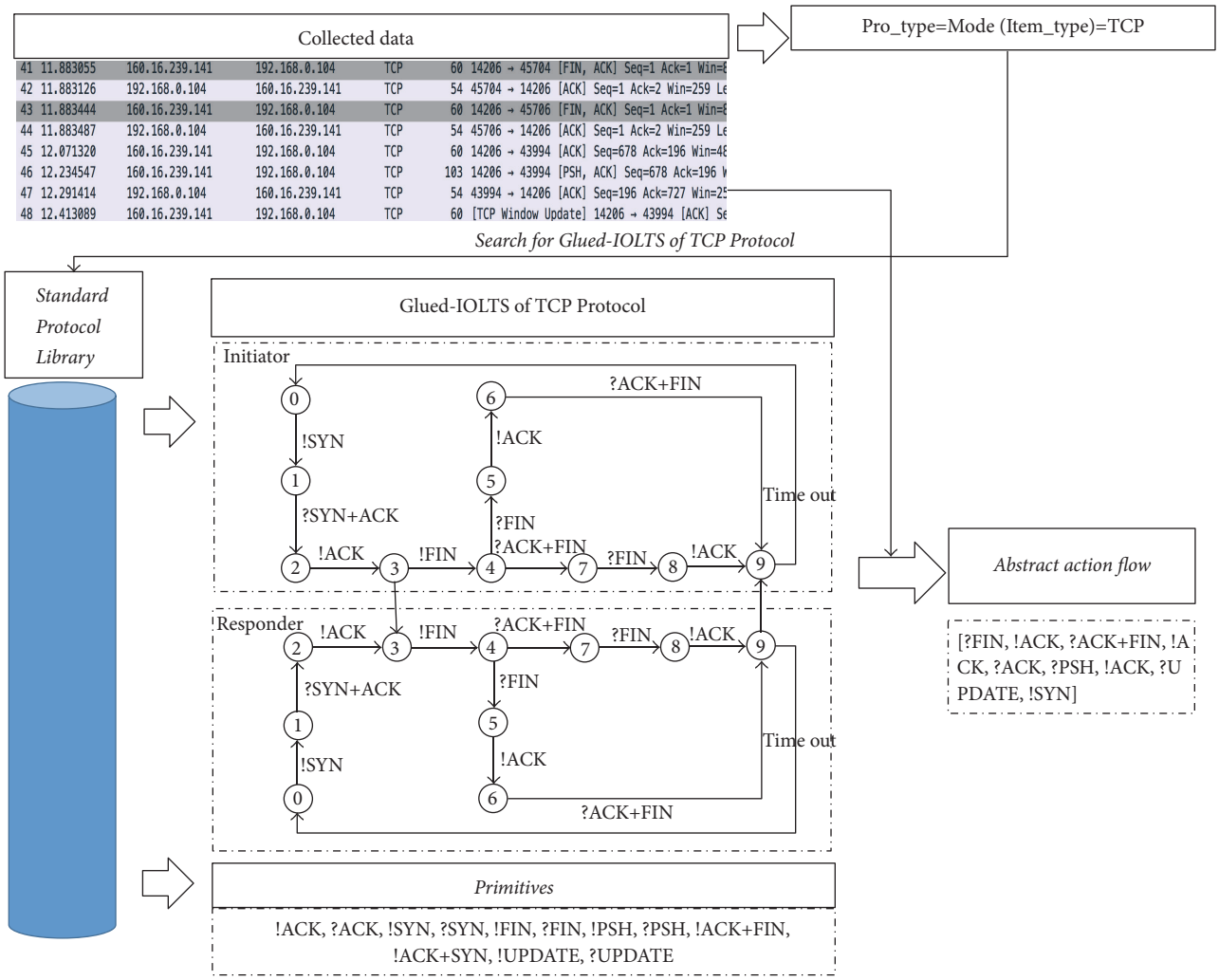


Figure 7: Example of translating abstract action flow.

corresponding *walk* can be found in the Glued-IOLTS graph, but the end state of this walk will not be the end state of the transition machine. It is a partial sequence of Glued-IOLTS.

(iii) Fake-attack: this attack corresponds to the aforementioned attack scenario 3. In this kind of attack, the compromised IoT devices may modify the transmitting message and inject some malicious codes to the message and send it to the receiver. This kind of attack may contain many strategies of modification, but here, we only consider the modifications which causes the changes on the automata primitives (the model transition label will change). If a sequence contains the fake-attack, the verification cannot find the corresponding *walk* in the Glued-IOLTS. But the fake actions may happen at the transition which makes the walk stopped or may happen before.

In order to detect those attacks automatically, we propose an algorithm in Algorithm 1. The inputs to the algorithm are one of the modeled label sequences (l_{ids}) which is detected by the IDS monitors and the glued transition system (T_{sys}). First of all, the algorithm searches for the transitions in T_{sys}, which have the same label as the first label of l_{ids} and record the results in a transition list of t_temp. Then for each transition t_i in t_temp, the algorithm compares the label of the next transition of t_i and the next label of l_{ids}. Remove t_i from t_temp. If the transition with the same label can be found, record it in t_temp. Backup this t_temp as t_temp_bac. Repeat the process until the end of l_{ids} or the t_temp is empty. During the loop, the algorithm records the past labels of l_{ids} in l_{pass}. The algorithm will stop if it checks all of the items in l_{ids} or T_{sys}. When it stops, if it found all labels of l_{ids} in T_{sys}, we go to check the final state of the walk in T_{sys}. If the finial state is an "end" state, l_{ids} is secure. Otherwise, l_{ids} contains jam-attack. If the algorithm stops when comparing l_n of l_{ids} with result of the t_temp being empty, then for each transition t_j in t_temp_bac, compare the label of the next transition of t_j and the passed label l_i in l_{pass}. If l_i is the same as the label of the next transition of t_j, record the next transition of t_j in t_temp, backup t_temp to t_temp_bac, record l_i in l_{pass}. Then, compare l_n with the next transitions of t_temp.

```
Input:
Label Array l_ids; //one transition sequence detected by IDS.
Transition Array T_sys; //the transition system of the protocol.
Output:
secure, fake-attack, jam-attack, replay-attack
Begin
Transition Array t_temp;
Transition Array t_next;
Label Array l_pass;
String result;
int flag=0; Search l_ids[0] in T_sys and record the results in t_temp;
For each transition t_i in t_temp{
   record the next transition of t_i in t_next;
   record l_ids[0] in l_pass;
  }For (int i = 1; i < l_ids.length; i++){
   flag++;
   If (t_tempisnotempty){
      record the next transition of t_i in t_next;
      t_temp_bac=t_temp;
      remove t_i from t_temp;
      Search l_ids[i] in t_next and record the results in t_temp;
      record l_ids[i] in l_pass;     }else{
      For each l_k in l_pass{
         Search l_k in t_next and record the results in t_temp;
         If (t_tempisnotempty){
            continue;
         }
      }
      If (l_ids[i] in l_pass){
         result="replay-attack";
         return result;
      }
      else{
         result="fake-attack";
         return result;
       }
   }   }
If(flag==l_ids.length){
   If(t_i.nexState().getStatus.equals("end")){
      result="secure";
      return result;
   }else{
      result="jam-attack";
      return result;
      result="secure";
   }
}
End
```

Algorithm 1: Algorithm for intrusion detection.

If l_n can be found in the next transition, record l_n in l_{pass} and move to the next label of l_{ids}. Otherwise, reconsider the passed labels until the end of l_{pass}. If after considering the labels of l_{pass}, l_n still cannot be found in the transition sequence, then l_{ids} must contain some modifications. The algorithm returns "fake-attack." Meanwhile, if l_{pass} contains l_n, then l_{ids} contains a replay, and the algorithm returns "replay-attack."

4. An Experiment over a Tested IoT System

In order to verify the proposed intrusion detection method, we design a IoT experiment environment like Figure 8. In the tested environment, we use two *Raspberry Pi 3* as the reduced-function device, an *Android Phone* (HUAWEI Mate 9) as a full-function device, and a wireless router

type:RADIUS
source:c0 a8 01 84
dest:c0 a8 01 0a
time:16:16:09
data:01 00 00 14 74 68 69 73 20 69 73 20 63 6c 69 65 6e 74 20 31
category:send
type:RADIUS
source:c0 a8 01 0a
dest:c0 a8 01 84
time:16:16:12
data:0b 00 00 3c 4e 61 73 74 6f 63 6c 69 65 6e 74 63 68 61 6c 6c 12 1e
69 6e 70 75 74 20 75 73 65 72 6e 61 6d 65 20 61 6e 64 20 70 61 73 73
77 61 72 64 73 18 0a 33 32 37 36 39 34 33 30
category:receive
type:RADIUS
source:c0 a8 01 84
dest:c0 a8 01 0a
time:16:17:12
data:01 00 00 3a 74 68 69 73 20 69 73 20 63 6c 69 65 6e 74 20 31 01 08 79
75 6c 6f 6e 67 02 12 0d be 70 8d 93 d4 13 ce 31 96 e4 3f 78 2a 0a ee 04
06 c0 a8 01 84 05 06 00 00 12 0c
category:send
...

Box 1: An example of IDS1 records traffics.

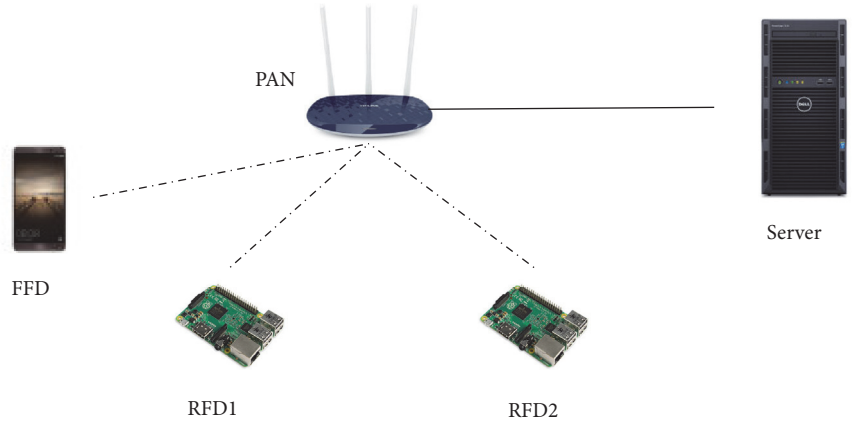


Figure 8: Experiment IoT networks.

(OpenWrt router) to be the IoT gateway (PAN coordinator). The router is connected with a server, and on the server, we use MySQL to build three database tables: Standard_Protocol, Abnormal_table, and Normal_table, which are corresponding to the three databases in our IDS methods. We use port mirroring on the router (a plug-in is needed to be installed on the OpenWrT router) and mirror the packets of WAN to the connected server. We install Wireshark [28] on the server side to collect and analyze the forwarded transmitting packets from IoT gateway. In our experiment, the RADIUS applications are taken as the services executed on the tested IoT networks [29]. The RADIUS protocol is an application layer protocol, which transmits data through UDP traffics. It uses the port number 1812 or 1645 to communicate. So when the monitor (Wireshark) obtains the IP traffics, by checking the port number of the UDP messages, the RADIUS messages can be distinguished.

For the simplicity of the experiment, we make the FFDs and RFDs only execute the RADIUS applications: we install the FreeRADIUS [30] on the server and the RADIUS client (NTRadPing [31]) on the client side (RFD1, RFD2 and FFD) to construct an experiment environment. We take the FFD device as an attacker and send the RADIUS requests as we need. Because the IoT gateway mirrored all of the WAN ports packets to the server, the Wireshark can record the sent/received data of each of the IoT devices, analyze them, and restore them. For better understanding, we select several packets and write them as the format of Box 1.

The IDS Event Analyzer in this experiment is an application we developed with Java. It can concatenate

Wc1	Wc2	Wc3	Wc4	Wc5	Lc1	Lc2	Lc3	R1	S1
xxxx !Ac_req_w1									
								?Ac_req_w1 !Ac_req_n_w1	
									?Ac_req_w1 !Ac_req_n_w1
								?Ac_accept_n_w1 !Ac_accept_w1	
?Ac_accept_w1 xxxx									
	xxxx !Ac_req_w2								
		?Ac_req_w2 !Ac_req_w2							
								?Ac_req_w2 !Ac_req_n_w2	
									?Ac_req_n_w2 !Ac_accept_n_w2
								?Ac_accept_n_w2 !Ac_accept_w2	
	?Ac_accept_w2 xxxx								
					xxxx !Ac_req_n				
								?Ac_req_l1 !Ac_req_n_l1	
				.....					

Figure 9: Message concatenation.

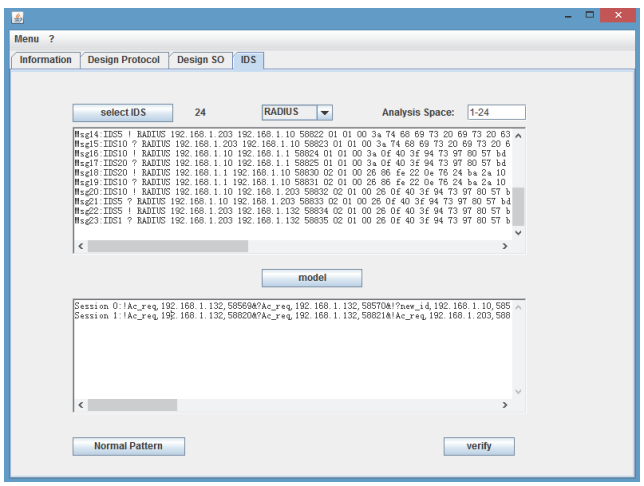


Figure 10: GUI of IDS.

the IDS detected messages as sequences, model those message sequences, and implement our algorithm to detect the possible intrusion (see Figure 10). As the network traffics happen sequently, the detected traffic data from different IoT devices may happen as Figure 9, where Wc1, Wc2, and Wc3 represented the RFD1, RFD2, and FFD of Figure 9, respectively. R1 represents the router, and S1 represents the server. For example, we choose a window size of 1 sec and found three modeled message sequences: {xxxx, !Ac_req_w1, ?Ac_req_w1, !Ac_req_w1_n, ?Ac_req_n_w1, !Ac_accept_n_w1, ?Ac_accept_n_w1, !Ac_accept_w1, ?Ac_accept_w1, xxxx}, {xxxx, !Ac_req_w2, ?Ac_req_w2, !Ac_req_w2, ?Ac_req_w2, !Ac_req_n_w2, ?Ac_accept_n_w2, !Ac_accept_w2, ?Ac_accept_w2, xxxx}, and {xxxx, !Ac_req_l1}. In this case, the first transition sequence is a normal connection sent from the client Wc1 to the server. The second sequence is a connection from Wc2 to Wc3 (this is maybe because the Wc3 declares himself as a NAS server); then Wc3 forwards the request of Wc2 to the real server. This sequence contains a replay-attack. And the third sequence is not a complete sequence. If the IDS only verifies the signature of the message, it will not find the problem of the second transition sequence. In our IDS approach, we only need to search this transition trace in the corresponding reachable graph, which is a nonanomalous profile of the target system.

The proposed Java tools will visit the Standard_Protocol table (the Standard Protocol Library) on MySQL database, and the nonanomalous profile of RADIUS protocol can be presented as the Glued-IOLTS of Figure 11. In this selected experiment, the verified traffics contain two RADIUS sessions and after the "message concatenation and classification," two different message sequences are obtained (they are listed in the bottom-left of Figure 11). Then through

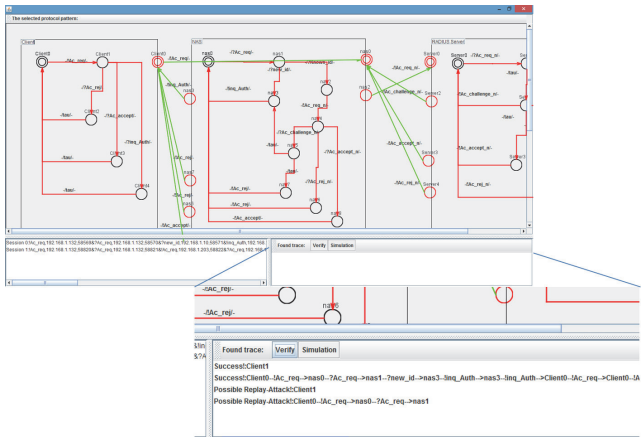


Figure 11: IDS verification panel.

the algorithm proposed, the program can verify the detected traffics automatically. The verification results of each detected sequence are presented in the bottom-right of Figure 11 (which identified that the first sequence is normal and the second sequence contains "replay-attack," and an alarm will be triggered when verifying the second message traffics).

5. Advances of the Proposed Method

The proposed intrusion detection method uses automata transitions to describe the network traffic flows and can map the different subnets of IoT to the same algebra space. In this case, different types of IoT, such as WSN, MANET, and Zigbee, can be described and compared with the same IDS method. Meanwhile, the way of using transition and graphic also makes the Standard Library, Anomaly Action Library, and Normal Action Library become easy to be implemented. However, because, in the process of finding abnormal action flows, the algorithm we used is a state based algorithm, which may cause the "state space explosion" problem, the complicity of the analyzed system should not be too much high. In fact, as the IoT devices are resources contained, the complexity of the IoT system is normally simple, and our IDS methods will be fine for the IoT intrusion detection.

6. Conclusion

Internet of Things is an important part of the future 5G, and the security of IoT will relate to many important scenarios of the future 5G and has become the core requirement of the network development. However, as the resources of IoT devices are constrained, many security mechanisms are hard to be implemented to protect the security of IoT networks. In this article, based on the automata theory, we proposed a uniform intrusion detection method for the vast heterogeneous IoT networks. Our method uses an extension of Labelled Transition Systems to propose a uniform description of IoT systems and can detect the intrusions by comparing the abstracted actions flows. We designed the intrusion detection approach, built the Event Databases, and implemented the Event Analyzer to achieve the IDS approaches. The result of the proposed IDS detects three types of IoT attacks: jam-attack, false-attack, and reply-attack. We also design an experiment environment to verify the proposed IDS method and examine the attack of RADIUS application in this article.

For the future work, we plan to continue enrich date types in our Standard Protocol Library and to improve the fuzzy method to make the creating of Normal Action Library become more efficient and accurate. Another line of our future research is to develop the suitable method to describe and evaluate the contents of the translating packets.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is sponsored by the National Key R&D Program of China (Grant 2016YFB0800700), the NSFC (Grants 61602359 and 61402354), the China Postdoctoral Science Foundation Funded Project (no. 2015M582618), the 111 project (Grant B16037), and the Fundamental Research Funds for the Central Universities (JB150115 and JB161508).

References

[1] H. Sedjelmaci, S. M. Senouci, and M. Al-Bahri, "A lightweight anomaly detection technique for low-resource IoT devices: a game-theoretic methodology," in *Proceedings of the IEEE International Conference on Communications (ICC '16)*, pp. 1–6, IEEE, Kuala Lumpur, Malaysia, May 2016.

[2] N. Boggs, W. Wang, S. Mathur, B. Coskun, and C. Pincock, "Discovery of emergent malicious campaigns in cellular networks," in *Proceedings of the 29th Annual Computer Security Applications Conference (ACSAC '13)*, pp. 29–38, New Orleans, La, USA, December 2013.

[3] C. X. Wang, X. Gao, X. You et al., "Cellular architecture and key technologies for 5g wireless communication networks," *IEEE Communications Magazine*, vol. 5, no. 2, pp. 122–130, 2014.

[4] B. Arrington, L. Barnett, R. Rufus, and A. Esterline, "Behavioral modeling intrusion detection system (BMIDS) using internet of things (IoT) behavior-based anomaly detection via immunity-inspired algorithms," in *Proceedings of the 25th International Conference on Computer Communication and Networks (ICCCN '16)*, pp. 1–6, Waikoloa, Hawaii, USA, August 2016.

[5] A. R. Baker and J. Esler, *Snort Intrusion Detection and Prevention Toolkit*, Andrew Williams, Norwich, NY, USA, 1st edition, 2007.

[6] C. Liu, J. Yang, Y. Zhang, R. Chen, and J. Zeng, "Research on immunitybased intrusion detection technology for the internet of things," in *Proceedings of the 7th International Conference on Natural Computation (ICNC '11)*, Shanghai, China, 2011.

[7] A. Nadeem and M. P. Howarth, "A survey of manet intrusion detection & prevention approaches for network layer attacks," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 4, pp. 2027–2045, 2013.

[8] Z. Yan, R. Kantola, G. Shi, and P. Zhang, "Unwanted content control via trust management in pervasive social networking," in *Proceedings of the 12th IEEE International Conference on*

Trust, Security and Privacy in Computing and Communications (TrustCom '13), pp. 202–209, Melbourne, Australia, July 2013.

[9] C. Modi, D. Patel, B. Borisaniya, H. Patel, A. Patel, and M. Rajarajan, "A survey of intrusion detection techniques in cloud," *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 42–57, 2013.

[10] A. A. Gendreau and M. Moorman, "Survey of intrusion detection systems towards an end to end secure internet of things," in *Proceedings of the IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud '16)*, pp. 84–90, Vienna, Austria, August 2016.

[11] A. Rayes and S. Samer, *Internet of Things—From Hype to Reality*, Springer International Publishing, Cham, Switzerland, 2017.

[12] Z. Hanzálek and P. Jurčík, "Energy efficient scheduling for cluster-tree wireless sensor networks with time-bounded data flows: application to IEEE 802.15.4/ZigBee," *IEEE Transactions on Industrial Informatics*, vol. 6, no. 3, pp. 438–450, 2010.

[13] J. P. Anderson, "Computer security threat monitoring and surveillance," Tech. Rep., 1980.

[14] L. T. Heberlein, "A network security monitor," in *Proceedings of the IEEE Computer Society Symposium, Research in Security and Privacy*, pp. 296–303, Oakland, Calif, USA, 1990.

[15] P. García-Teodoro, J. Díaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, "Anomaly-based network intrusion detection: techniques, systems and challenges," *Computers and Security*, vol. 28, no. 1-2, pp. 18–28, 2009.

[16] S. Kumar and E. H. Spafford, "A software architecture to support misuse intrusion detection," in *Proceedings of the 18th National Information Security Conference*, pp. 194–204, Baltimore, Md, USA, October 1995.

[17] K. Ilgun, R. A. Kemmerer, and P. A. Porras, "State transition analysis: a rule-based intrusion detection approach," *IEEE Transactions on Software Engineering*, vol. 21, no. 3, pp. 181–199, 1995.

[18] T. Lunt, A. Tamaru, F. Gilham et al., "A real-time intrusion detection expert system (ides)-final technical report," Technical Report, Computer Science Laboratory, SRI International, Menlo Park, Calif, USA, 1992.

[19] S. Staniford-Chen, B. Tung, P. Porras et al., "The common intrusion detection framework-data formats," Internet draft draft-staniford-cidf-dataformats-00.txt, 1998.

[20] J. Chen and C. Chen, "Design of complex event-processing IDS in internet of things," in *Proceedings of the 6th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA '14)*, pp. 226–229, January 2014.

[21] D. Lee and M. Yannakakis, "Principles and methods of testing finite state machines—a survey," *Proceedings of the IEEE*, vol. 84, no. 8, pp. 1090–1123, 1996.

[22] J. Tretmans, "Conformance testing with labelled transition systems: implementation relations and test generation," *Computer Networks*, vol. 29, no. 1, pp. 49–79, 1996.

[23] Y. Fu and O. Koné, "Security and robustness by protocol testing," *IEEE Systems Journal*, vol. 8, no. 3, pp. 699–707, 2014.

[24] G. Lowe, "Breaking and fixing the Needham-Schroeder Public-Key Protocol using FDR," in *Tools and Algorithms for the Construction and Analysis of Systems*, vol. 1055 of *Lecture Notes in Computer Science*, pp. 147–166, Springer, Berlin, Germany, 1996.

[25] P. Tsankov, M. T. Dashti, and D. Basin, "SECFUZZ: fuzz-testing security protocols," in *Proceedings of the 7th International Workshop on Automation of Software Test (AST '12)*, pp. 1–7, Zurich, Switzarland, June 2012.

[26] B. Lei, X. Li, Z. Liu, C. Morisset, and V. Stolz, "Robustness testing for software components," *Science of Computer Programming*, vol. 75, no. 10, pp. 879–897, 2010.

[27] Y. Fu and O. Koné, "Validation of security protocol implementations from security objectives," *Computers and Security*, vol. 36, pp. 27–39, 2013.

[28] Wireshark, "Wireshark network protocol analyzer.," 2017, http://www.wireshark.org/.

[29] C. Rigney, S. Willens, and A. Rubens, "Remote authentication dial in user service (radius)," Tech. Rep. RFC2865, The Internet Society, Reston, Va, USA, 2000.

[30] FreeRADIUS, "Freeradius-the world's most popular radius server," 2017, http://freeradius.org/.

[31] mastersoft, "Ntradping-radius test utility," 2017, http://www.mastersoft-group.com/.

Hindawi
Mobile Information Systems
Volume 2017, Article ID 2349149, 7 pages
https://doi.org/10.1155/2017/2349149

Research Article

ULMAP: Ultralightweight NFC Mutual Authentication Protocol with Pseudonyms in the Tag for IoT in 5G

Kai Fan,[1] Panfei Song,[1] and Yintang Yang[2]

[1]*State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an, China*
[2]*Key Laboratory of Ministry of Education for Wide Band-Gap Semiconductor Materials and Devices, Xidian University, Xi'an, China*

Correspondence should be addressed to Kai Fan; kfan@mail.xidian.edu.cn

Received 24 January 2017; Revised 3 March 2017; Accepted 20 March 2017; Published 27 April 2017

Academic Editor: Jing Zhao

As one of the core techniques in 5G, the Internet of Things (IoT) is increasingly attracting people's attention. Meanwhile, as an important part of IoT, the Near Field Communication (NFC) is widely used on mobile devices and makes it possible to take advantage of NFC system to complete mobile payment and merchandise information reading. But with the development of NFC, its problems are increasingly exposed, especially the security and privacy of authentication. Many NFC authentication protocols have been proposed for that, some of them only improve the function and performance without considering the security and privacy, and most of the protocols are heavyweight. In order to overcome these problems, this paper proposes an ultralightweight mutual authentication protocol, named ULMAP. ULMAP only uses Bit and XOR operations to complete the mutual authentication and prevent the denial of service (DoS) attack. In addition, it uses subkey and subindex number into its key update process to achieve the forward security. The most important thing is that the computation and storage overhead of ULMAP are few. Compared with some traditional schemes, our scheme is lightweight, economical, practical, and easy to protect against synchronization attack.

1. Introduction

IoT [1] is a large network that consists of various information sensing devices and the Internet. As a new technology, the NFC [2, 3] is one of the core technologies of IoT and is listed as one of the most promising technologies.

NFC is a short-range, high-frequency, noncontact automatic identification wireless communication technology using the 13.56 MHz frequency band at a distance of less than 10 cm. It is a development and breakthrough of the RFID [4–6] technology. NFC now has been widely used in electronic ticket, product security, and other fields. But the security issues, especially the authentication problem between the reader and the tag, have become an important factor restricting its development. The problem of authentication is to confirm the validity of the tag and the reader. Since NFC communication is completely exposed to the wireless environment, it faces a lot of malicious attacks such as clone attack [7, 8], man-in-the-middle attack, and packet losses attack. Once the authentication protocol is under the above attack, the authentication will be failed. Meanwhile, because the NFC system is limited by many factors, such as computing power, storage space, and power supply, it is a challenging task to design a secure and efficient NFC authentication protocol.

So far, although a lot of security authentication schemes for NFC are presented, researchers at home and abroad do not put forward a universal applicability scheme. For example, Yun-Seok et al. [3] proposed a scheme that uses the asymmetric encryption and hash function to try to eliminate the security and privacy thread. Although the solution can solve the problem of mutual authentication and prevent replay attack and the man-in-the-middle attack, it lacks some necessary security attributes, such as the message authentication. In 2013, Eun et al. [9] presented a new conditional privacy preserving security protocol to protect the user's privacy. In 2015, Kannadhasan et al. [10] proposed the similar approach as presented in CPPNFC. In the same year, He et al. [11] proposed a pseudonym-based NFC protocol, but it cannot solve the forward security. In order to better promote

the NFC technology, a scheme is needed to be proposed to solve the security and privacy thread.

Therefore, in this paper, we propose an ultralightweight mutual authentication protocol (ULMAP). Compared with the old NFC scheme, this protocol not only solves the security and privacy problem but also reduces the computation and storage cost.

Our Contributions. In this paper, we propose an ultralightweight mutual authentication protocol (ULMAP) for NFC using less memory storage and computational power for low-cost NFC tags. Our scheme has the following features:

(1) Ultralightweight: the scheme is designed only with simple shift and XOR operations, not hash or other encryption operations.

(2) Secure and efficient: the scheme we proposed could meet requirements of forward security, mutual authentication, synchronization, and non-denial of service by subkey and pseudonym.

Paper Organization. The remainder of the paper is organized as follows: In Section 2, we will present the detailed protocol of our new NFC mutual authentication protocol (ULMAP). In Section 3, the security proof with BAN logic of the proposed protocol will be provided. Section 4 provides the security and performance analysis of our protocol. Finally, our conclusion is shown in Section 5.

2. NFC Authentication Protocol for Mobile Device

In this section, we will propose ULMAP and basic ideas are as follows: the scheme only with a simple shift and XOR operations, greatly reducing the cost of operations. And it uses the concept of pseudonym, thus improving the system of security. And the scheme uses the concept of subkeys, preventing the man-in-the-middle attack as compared to the related existing authentication protocols.

2.1. Initialization. The explanations of symbols are shown in Abbreviation.

MixBits(X, Y) [12] is defined as follows:

$$
\begin{aligned}
&Z = \text{MixBits}(X, Y)\\
&\vdots\\
&Z = X;\\
&\text{for } (i = 0; i < 32; i++)\\
&\{\\
&\qquad Z = (Z \gg 1) + Z + Z + Y;\\
&\}
\end{aligned}
$$

In this scheme, the message $(\text{IDS}, \text{ID}, K_1, K_2)$ is stored in each tag. Meanwhile, $(\text{ID}, (\text{IDS}_{\text{old}}, K^1_{\text{old}}, K^2_{\text{old}}), (\text{IDS}_{\text{new}}, K^1_{\text{new}}, K^2_{\text{new}}))$ is stored in the server corresponding to each tag.

2.2. The Authentication Process. The authentication process of ULMAP is shown in Figure 1. The protocol involves three entities: tag, reader, and database. The channel between the reader and the database is assumed to be secure, but that between the reader and the tag faces all the possible potential attacks [13–15]. Each tag has a unique static identification (ID) and preshares a pseudonym (IDS) and two keys K_1, K_2 with the database.

Each database actually has two entries of $(\text{ID}, (\text{IDS}_{\text{old}}, K^1_{\text{old}}, K^2_{\text{old}}), (\text{IDS}_{\text{new}}, K^1_{\text{new}}, K^2_{\text{new}}))$: one is for the old values and the other is for the potential next values. The reader first sends "Query" and T_R message to the tag. The tag will respond with its IDS after it verifies that the timestamp T_R is larger than T_t. Then, the reader will use the tag's response IDS to find a matched entry in the database and goes to the mutual authentication stage if a matched entry is found no matter what $\text{IDS} = \text{IDS}_{\text{old}}$ or $\text{IDS} = \text{IDS}_{\text{new}}$. In the mutual authentication phase, the reader and the tag authenticate each other, and they, respectively, update their local pseudonym and the keys after successful authentication, which are shown in Figure 1.

There are four stages in the scheme that we proposed, such as initialization, tag identification, mutual authentication, and index-pseudonym and key updating. Then, we will in detail introduce the four stages as follows.

Initialization. The database selects a pseudorandom generator PRNG [16] $g : \{0,1\}^k \to \{0,1\}^{2k}$ to generate pseudorandom number. The database generates the key $K = K_1 \mid K_2$, which is initialized to $K_1 = \text{Rot}(\text{Rot}(n_{i,2}+\text{ID}+n_{i,3}, n_{i,2})+\text{ID}, n_{i,1})+n_{i,3}$ and $K_2 = \text{Rot}(\text{Rot}(n_{i,1} + \text{ID} + n_{i,3}, n_{i,1}) + \text{ID}, n_{i,2}) \oplus n_{i,3}$, and places it in a valid tag and the legitimate reader. $n_{i,j}$ is the jth random number of the ith tag in the initialization phase, $j = 1, 2, 3$. The database, reader, and tag will store the IDS and $K = K_1 \mid K_2$ corresponding to the tag.

Tag Identification. The reader generates the random timestamp T_R and the random number n_2 and sends authentication queries n_2, Query, and T_R to the tag. Then, the tag judges whether $T_R > T_t$; if T_R is not larger than T_t, the authentication is failed. Otherwise, the mutual authentication phase will begin.

Mutual Authentication. After identification phase, the tag will generate a random number n_2, calculate A, B, and C as shown in Figure 1, and send IDS, A, B, and C to the reader. Using the IDS, the reader tries to find an identical entry in the database. If this search succeeds, the reader can get the nonce from submessages A and B. Then, the reader will compute n'_3 and $\overline{K^*_1}/\overline{K^*_2}$ and build a local version of submessage C' as shown in Figure 1. It will be compared with the received value. If it is verified, the tag is authenticated. Finally, the reader sends message $D = (\overline{K} \oplus \text{ID}) \oplus ((K_2 + K_1) \cup \overline{K^*_2})$ to the tag. When the message D is received by the tag, it will be compared with a computed local version $D' = (\overline{K_1} \oplus \text{ID}) \oplus ((K_2 + K_1) \cup \overline{K_2})$. If comparison is successful, the reader is authenticated. Otherwise, the authentication protocol is failed.

Index-Pseudonym and Key Updating. After successfully completing the mutual authentication phase between the tag and

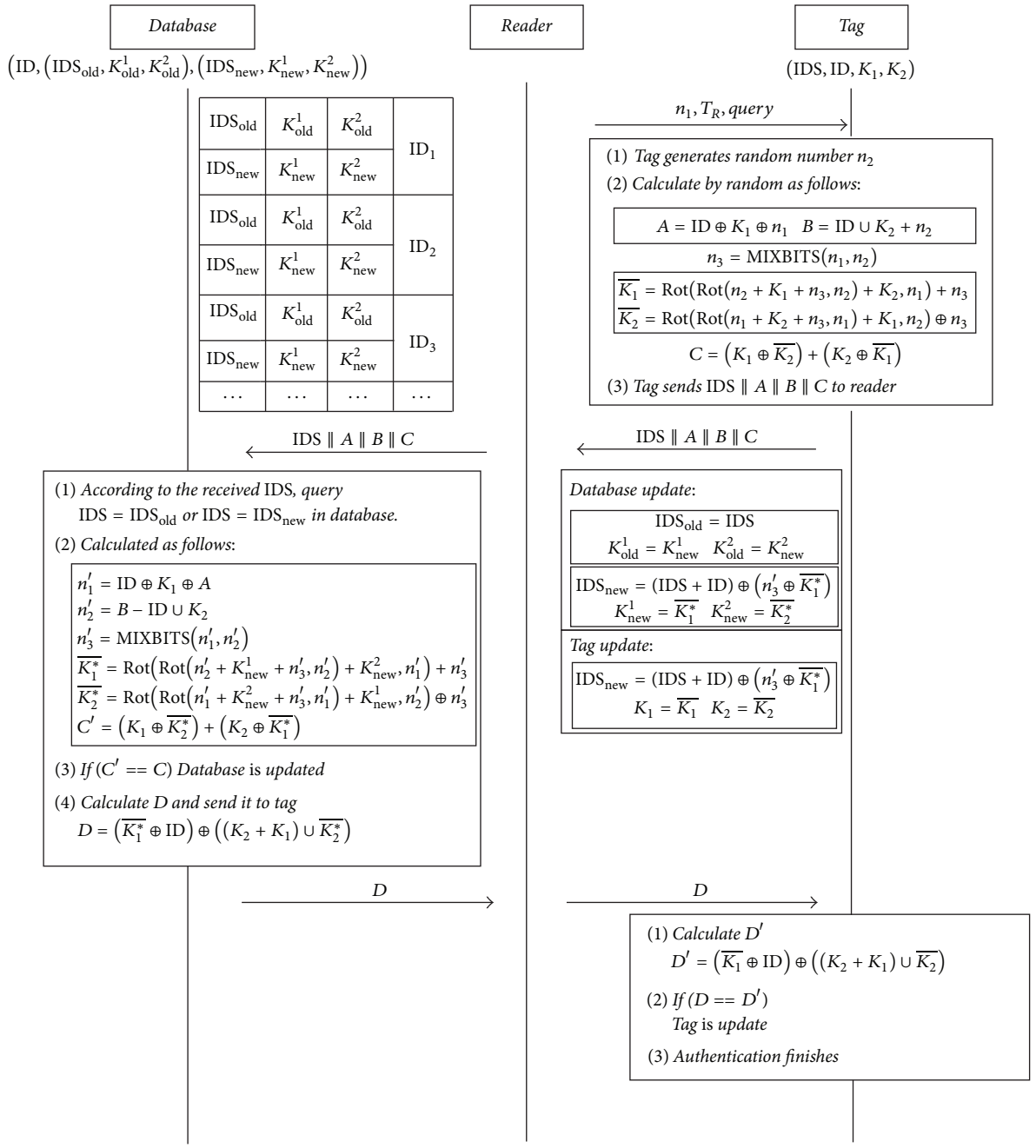


Figure 1: Authentication process of ULMAP.

the reader, they locally update IDS and key as indicated in Figure 1.

3. Security Proof with BAN Logic

The security assurance of the proposed protocol is the secure mutual authentication, which means the following security aims should be achieved.

Security Aim 1. The database needs to make sure the received message IDS ‖ A ‖ B ‖ C is exactly the one sent by the tag. This means that we need to achieve Database|≡ Tag|~ (IDS, A, B, C) and Database|≡ Tag|≡ (IDS, A, B, C).

Security Aim 2. The tag needs to make sure the received message D is exactly the one sent by the database, which means the following formulas need to be achieved: Tag|≡ Database|~ D and Tag|≡ Database|≡ D.

3.1. Security Assumption. According to the given protocol and the assumption that the server and the reader are connected securely, the following conditions can be achieved:

AS1: Database$|\equiv$ Database $\overset{n_{i,j}}{\rightleftarrows}$ Tag$_i$.

AS2: Tag$_i|\equiv$ Database $\overset{n_{i,j}}{\rightleftarrows}$ Tag$_i$.

AS3: Reader $\Longrightarrow (n_1)$.

AS4: Reader$|\equiv \#(n_1)$.

AS5: Database$|\equiv \#(n_1)$.

AS6: Tag$_i \Longrightarrow (n_2)$.

AS7: Tag$_i|\equiv \#(n_2)$.

3.2. Security Analysis. According to the proposed protocol (ULMAP) $K_1 = \mathrm{Rot}(\mathrm{Rot}(n_{i,2} + \mathrm{ID} + n_{i,3}, n_{i,2}) + \mathrm{ID}, n_{i,1}) + n_{i,3}$ and $K_2 = \mathrm{Rot}(\mathrm{Rot}(n_{i,1} + \mathrm{ID} + n_{i,3}, n_{i,1}) + \mathrm{ID}, n_{i,2}) \oplus n_{i,3}$, together with the assumptions AS1 and AS2, we can deduce Tag$_i|\equiv$ Database $\overset{K_{i,j}}{\rightleftarrows}$ Tag$_i$ and Database$|\equiv$ Database $\overset{K_{i,j}}{\rightleftarrows}$ Tag$_i$, because, in this scheme, the database will receive the message (IDS, A, B, C) forwarded from the reader, where $C = (K_1 \oplus \overline{K_2}) + (K_2 \oplus \overline{K_1})$. As we have achieved $K_{i,j}$ as secret between the database and the tag, we can take $K_{i,j}$ as the secret key to protect messages. So we can simply write the received message of database as (IDS, A, B, $C)_{K_{i,j}}$, and we have Database $\triangleleft$ (IDS, $A, B, C)_{K_{i,j}}$. For the reason of "message-meaning rule" of BAN $(P|\equiv Q \overset{Y}{\parallel} P, P \triangleleft \langle X\rangle_Y)/(P|\equiv (Q|\sim X))$, we can deduce Database$|\equiv$ Tag$_i|\sim$ (IDS, A, B, C).

From the assumption AS5 : Server$|\equiv \#(n_1)$ and the BAN rule of $(P|\equiv \#(X))/(P|\equiv \#(X, Y))$, we know Database$|\equiv \#$(IDS, A, B, C). Because we have achieved Database$|\equiv$ Tag$_i|\sim \#$(IDS, A, B, C), together with the "nonce-verification" rule $(P|\equiv (\#(X)), P|\equiv (Q|\sim X))/(P|\equiv (Q|\equiv X))$, we will achieve Database$|\equiv$ Tag$_i|\equiv$ (IDS, A, B, C), and the first security aim of the given protocol is achieved.

For the same reason, we can also deduce Tag$_i|\equiv$ Database$|\sim D$ and Tag$_i|\equiv$ Database$|\equiv D$, and the second security aim is also achieved, and the security of mutual authentication of the proposed protocol has been proved.

4. Evaluation

In this section, we will analyze the proposed protocol (ULMAP) from the security and performance point of view.

4.1. Security Analysis. It is obvious, from the protocol specification, that not only can the tag and the reader successfully authenticate each other, but also ULMAP is able to resist the common NFC attacks effectively. In particular, it makes the scheme have the anti-DoS attack capability through using the timestamp. We now analyze our proposed scheme from the point of view of security as follows.

4.1.1. Mutual Authentication. The tag and the reader can authenticate each other by messages C and D, because only the genuine tag has the subkeys K_1 and K_2 which generate the consistent message C with random numbers n_1, n_2. Similarly, only the genuine reader keeps the ID that is used to generate the response message D. In this way, the reader and the tag can achieve mutual authentication.

4.1.2. Tag Anonymity. The tag uses the pseudonym in the whole authentication process. The pseudonym of each tag will be updated after every successful authentication by the random numbers n_1, n_2. So the pseudonym from the same tag looks different at each session authentication and the attackers cannot get the real identity of the tag. Moreover, even if the attackers intercept authentication pseudonym IDS, they cannot analyze the practical information from it.

4.1.3. Resistance to Tracking. The data stored in the database and the tag will be updated after the successful authentication process. So the message and the response message are different at each session authentication, making it almost impossible for the attackers to track the tag. In addition, the tag uses the pseudonym which improves the difficulty of tracking.

4.1.4. Data Confidentiality [17]. The calculation of each value of A, B, C, and D involves at least two secret values, including the subkey and random number. So, it is very hard to get the tag ID except for the tag itself that has K_1, K_2 and n_1, n_2.

4.1.5. Forward Security. After each successful session, the key and IDS value will be updated in the tag and the database. So even if the attacker achieves some session information, he cannot use it to trace back to previous communications. In addition, ULMAP makes the subkey and random number involved in the entire update process, which makes the entire update process have stronger stochastic properties. So ULMAP is forward security.

4.1.6. Nonreplaying. Because the value of IDS will be updated after the successful authentication process, the response message IDS $\parallel A \parallel B \parallel C$ from the same tag is different in each session authentication process. Moreover, the timestamp T_R is constantly changing over time. Therefore, the attacker cannot priorly disguise information to achieve legality certification.

4.1.7. Non-Denial of Service (Non-DoS) [18]. When the reader starts a new session, the tag will judge whether $T_R > T_t$. If not, the authentication is failed. Otherwise, the authentication process will continue. Compared with all most schemes responding to the query, ULMAP can reduce the number of denial of service attacks to some extent and prevent unauthorized readers from continuing to send queries which consume lots of resources of the tag. Therefore, this scheme can resist denial of service attacks in some cases.

The comparison between LMAP [19], SASI [20], and ULMAP in security is shown in Table 1. "$\surd$" means satisfaction, "$\times$" means to dissatisfy, and "#" means satisfaction to a certain extend.

It is very obvious, in Table 1, that neither of SASI and LMAP can resist desynchronization and DoS attacks. However, in addition to the forward security, data confidentiality,

Table 1: The security and functionality comparison.

Scheme	Mutual authentication and forward security	Confidentiality and anonymity	Resistance to tracking	Nonreplaying	Resistance to desynchronization attack	Non-DoS
LMAP	×	×	×	×	×	×
SASI	√	√	√	√	×	×
ULMAP	√	√	√	√	√	#

Table 2: The storage overhead comparison.

Scheme	Database	Reader	Tag
LMAP	$6ml$	0	$6L$
SASI	$4ml$	0	$7L$
ULMAP	$7ml$	0	$5L$

Table 3: The cost of communication comparison.

Scheme	The number of interactions	Total cost of communication
LMAP	4	$5L$
SASI	4	$5L$
ULMAP	3	$7L$

Table 4: Computation cost comparison.

Scheme	LMAP	SASI	ULMAP
Cost	$\oplus, +, \wedge, \vee$	$\oplus, +, \wedge, \vee$, Rot	$\oplus, +, \wedge, \vee$ Rot^2, MixBits

nonreplaying, and so forth, the proposed protocol ULMAP can prevent synchronicity attacks effectively and prevent DoS attacks to some extent. In summary, ULMAP improves the security.

4.1.8. Synchronization. In a normal session, if the tracker heads off the last message that the database sends to the tag, the database cannot be successfully verified. Once this case happens, the tag cannot be updated, but the database has been updated successfully. So the tag and the database will lose the synchronization. However, in the ULMAP protocol, the IDS, K_1, K_2, used in the last session is stored in (ID, ($\mathrm{IDS}_{\mathrm{old}}$, K^1_{old}, K^2_{old}), ($\mathrm{IDS}_{\mathrm{new}}$, K^1_{new}, K^2_{new})) in the database, so that this tag is still able to finish the authentication and get the synchronization again successfully.

4.2. Performance and Complexity Analysis. We will compare ULMAP with SASI and LMAP in performance and complexity. In order to compare easily, assume there are m tags in the system and the length of data is L.

4.2.1. The Cost of Storage. To achieve the authentication, in SASI protocol, the tag stores the message (ID, ($\mathrm{IDS}_{\mathrm{new}}$, $K_{1\mathrm{new}}$, $K_{2\mathrm{new}}$)($\mathrm{IDS}_{\mathrm{old}}$, $K_{1\mathrm{old}}$, $K_{2\mathrm{old}}$)) and (ID, IDS, K_1, K_2) is stored in the database, so the cost of storage in the tag and database is $7L$ and $4mL$, respectively. As it is shown in Table 2, in LMAP, the tag storage space needs $6L$ and the corresponding database storage space requires $6mL$. But in our protocol, the cost of storage space in the tag is $5L$ and the cost of storage space in the database is $7mL$.

Usually, the database has more resources than the tag, so the resource of tag is more valuable. Comparing with other protocols, the ULMAP needs smaller storage space in the tag that will greatly reduce the cost of the tag and increase a little cost of storage space in the database. Therefore, the proposed protocol can greatly reduce input cost. The specific storage overhead is shown in Table 2.

4.2.2. The Cost of Communication. The cost of communication consists of the number of interactions and the length of the communication data. From Table 3, we can know that the interaction times of both SASI and LMAP are 4. Although the transmitted data is increased a little, our protocol is just transmitted three times between the reader and the tag, which are four times in other protocols. Therefore, ULMAP has a relatively low communication overhead.

Comparing with other protocols, the ULMAP uses the timestamp for the first time. This will make the ULMAP resist the attack of DoS to a certain extent. Moreover, the subkey and random numbers are used widely in the database and the tag in the authentication update phase. This can make the whole protocol have stronger random feature which will greatly improve the ability of resisting desynchronization and the forward security of ULMAP.

4.2.3. The Cost of Computation Time. In order to better compare the computation performance of different protocols in Table 4, + represents AND operation, $\oplus$ represents the XOR operation, Rot is the displacement Rot(x, y) operation, Rot^2 is two displacement Rot(x, y) operations, and T represents the pseudorandom number or timestamp.

From Table 4, it is shown that the tag in ULMAP needs one random number generation. In addition, ULMAP also needs more computation operation (like Rot, MixBits) in the tag compared with SASI and Gossamer. Although this will increase the cost of computation, the computations also become more secure and effective with it.

By comparing our protocol with other schemes, it shows that our proposed protocol not only can provide mutual authentication function but also has the advantage of higher level of security and performance.

5. Conclusions

This paper proposes a new NFC mutual authentication protocol, named ULMAP. ULMAP can achieve not only mutual authentication but also complete anonymity. Moreover, the proposed scheme possesses higher security and performance. Because the database stores the new and old session private key and IDS, when the new session private key of the tag fails to update, the corresponding old private key and IDS can also be used. So the proposed protocol can effectively resist the desynchronization attack.

Abbreviations

IDS:	The pseudonym of tag identity
IDS_{old}:	The index number used last time
IDS_{new}:	The index number successfully used this time
ID:	The unique static identification of tag
K:	The shared key of the tag and database, which is divided into two parts
T_R:	The random timestamp generated by the reader
T_t:	The last time timestamp
K_{old}:	The key of the tag successfully used in the last round session
K_{new}:	The key of the tag used in this session
n_1, n_2:	The random number generated by the tag and the reader
Rot(x, y):	The operation of rotation $x \ll W(y)$, where $W(y)$ denotes Hamming weight of y.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work has been financially supported by the National Natural Science Foundation of China (nos. 61303216, 61272457, U1401251, and 61373172), the National High Technology Research and Development Program of China (863 Program) (no. 2012AA013102), the Open Research Project of the State Key Laboratory of Industrial Control Technology, Zhejiang University, China (no. ICT170312), and National 111 Program of China (nos. B16037 and B08038).

References

[1] H. Ning and B. Wang, *RFID Major Projects and the State Internet of Things*, Mechanical Industry Press, Beijing, China, 2008.

[2] V. Odelu, A. K. Das, and A. Goswami, "SEAP: secure and efficient authentication protocol for NFC applications using pseudonyms," *IEEE Transactions on Consumer Electronics*, vol. 62, no. 1, pp. 30–38, 2016.

[3] L. Yun-Seok, K. Eun, and J. Min-Soo, "A NFC based authentication method for defense of the man in the middle attack," in *Proceedings of the 3rd International Conference on Computer Science and Information Technology*, Bali, Indonesia, January 2013.

[4] K. Fan, J. Li, H. Li, X. Liang, X. Shen, and Y. Yang, "RSEL: revocable secure efficient lightweight RFID authentication scheme," *Concurrency Computation Practice and Experience*, vol. 26, no. 5, pp. 1084–1096, 2014.

[5] E. Haselsteiner, "Security in near field communication (NFC)," in *Proceedings of the Workshop on RFID Security*, Malaga, Hungary, 2006.

[6] K. Fan, Y. Gong, C. Liang, H. Li, and Y. Yang, "Lightweight and ultralightweight RFID mutual authentication protocol with cache in the reader for IoT in 5G," *Security and Communication Networks*, vol. 9, pp. 3095–3104, 2016.

[7] M. Dong, K. Ota, L. T. Yang, A. Liu, and M. Guo, "LSCD: a low-storage clone detection protocol for cyber-physical systems," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 35, no. 5, pp. 712–723, 2016.

[8] L. Zhang, L. Wei, D. Huang, K. Zhang, M. Dong, and K. Ota, "MEDAPs: secure multi-entities delegated authentication protocols for mobile cloud computing," *Security and Communication Networks*, vol. 9, no. 16, pp. 3777–3789, 2016.

[9] J. C. Paillès, C. Gaber, V. Alimi, and M. Pasquet, "Payment and privacy: a key for the development of NFC mobile," in *Proceedings of the International Symposium on Collaborative Technologies and Systems (CTS '10)*, pp. 378–385, May 2010.

[10] M. Hassinen, K. Hyppönen, and E. Trichina, "Utilizing national public-key infrastructure in mobile payment systems," *Electronic Commerce Research and Applications*, vol. 7, no. 2, pp. 214–231, 2008.

[11] Z. Kabir, *User centric design of an NFC mobile wallet framework [M.S. thesis]*, The Royal Institute of Technology (KTH), Stockholm, Sweden, 2011.

[12] E. G. Ahmed, E. Shaaban, and M. Hashem, "Lightweight mutual authentication protocol for low cost RFID tags," *International journal of Network Security & Its Applications*, vol. 2, no. 2, pp. 27–37, 2010.

[13] A. Juels, "Strengthening EPC tags against cloning," in *Proceedings of the ACM Workshop on Wireless Security (WiSe '05)*, pp. 67–75, Cologne, Germany, September 2005.

[14] C. Mulliner, "Vulnerability analysis and attacks on NFC-enabled mobile phones," in *Proceedings of the 4th International Conference on Availability, Reliability and Security*, pp. 695–700, IEEE, Fukuoka, Japan, March 2009.

[15] L. Francis, G. P. Hancke, K. Mayes et al., "Practical NFC peer-to-peer relay attack using mobile phones," in *Proceedings of the 6th International Workshop on Radio Frequency Identification: Security and Privacy Issues (RFID-SEC '10)*, pp. 35–49, Istanbul, Turkey, 2010.

[16] P. Peris-Lopez, J. C. Hernandez-Castro, J. M. Estevez-Tapiador, and A. Ribagorda, "LAMED—a PRNG for EPC class-1 generation-2 RFID specification," *Computer Standards and Interfaces*, vol. 31, no. 1, pp. 88–97, 2009.

[17] L. Gu, L. Wang, K. Ota, M. Dong, Z. Cao, and Y. Yang, "New public key cryptosystems based on non-Abelian factorization problems," *Security and Communication Networks*, vol. 6, no. 7, pp. 912–922, 2013.

[18] F. Fahrianto, M. F. Lubis, and A. Fiade, "Denial-of-service attack possibilities on NFC technology," in *Proceedings of the 4th International Conference on Cyber and IT Service Management*, pp. 1–5, IEEE, April 2016.

[19] P. Peris-Lopez, J. C. Hernandez-Castro, J. M. E. Tapiador et al., "LMAP: a real lightweight mutual authentication protocol for low-cost RFID tags," in *Proceedings of the Workshop on RFID Security*, Graz, Austria, July 2006.

[20] H.-Y. Chien, "SASI: a new ultralightweight RFID authentication protocol providing strong authentication and strong integrity," *IEEE Transactions on Dependable and Secure Computing*, vol. 4, no. 4, pp. 337–340, 2007.

Hindawi
Mobile Information Systems
Volume 2017, Article ID 1325698, 9 pages
https://doi.org/10.1155/2017/1325698

Research Article

Privacy-Preserving Billing Scheme against Free-Riders for Wireless Charging Electric Vehicles

Xingwen Zhao,[1,2] Jiaping Lin,[1,2] and Hui Li[1,2]

[1]*State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China*
[2]*School of Cyber Engineering, Xidian University, Xi'an 710071, China*

Correspondence should be addressed to Xingwen Zhao; sevenzhao@hotmail.com

Received 25 October 2016; Revised 24 March 2017; Accepted 30 March 2017; Published 10 April 2017

Academic Editor: Jing Zhao

Recently, scientists in South Korea developed on-line electric vehicle (OLEV), which is a kind of electric vehicle that can be charged wirelessly while it is moving on the road. The battery in the vehicle can absorb electric energy from the power transmitters buried under the road without any contact with them. Several billing schemes have been presented to offer privacy-preserving billing for OLEV owners. However, they did not consider the existence of free-riders. When some vehicles are being charged after showing the tokens, vehicles that are running ahead or behind can switch on their systems and drive closely for a free charging. We describe a billing scheme against free-riders by using several cryptographic tools. Each vehicle should authenticate with a compensation-prepaid token before it can drive on the wireless-charging-enabled road. The service provider can obtain compensation if it can prove that certain vehicle is a free-rider. Our scheme is privacy-preserving so the charging will not disclose the locations and routine routes of each vehicle. In fact, our scheme is a fast authentication scheme that anonymously authenticates each user on accessing a sequence of services. Thus, it can be applied to sequential data delivering services in future 5G systems.

1. Introduction

As more and more people concern about air pollution and the exhaustion of fossil energy, the increasing use of combustion engines will receive more criticisms than before. In order to alleviate these problems, electric vehicles (EVs) were introduced as a good replacement for combustion engines. The engines in EVs can use electrical power much more efficiently than the combustion engines [1]. However, the battery price is high and its size is limited. Moreover, plug-in EVs have to stop periodically for a period of time to recharge the battery.

In order to handle the above problems of PEVs, wireless charging vehicle called the on-line electric vehicle (OLEV) was introduced and tested in South Korea [2]. The technology of charging the vehicle while it is moving along the road will greatly reduce the number of times that a driver needs to stop for recharging. Such a convenient method makes battery-powered vehicle more favorable. By this way, the industry can decrease the size of the battery and then the price of EVs [3]. In the OLEV system, power transmitters (PTs) are installed underneath the road and electric vehicles can be charged wirelessly when users drive them along the road. With enough segments of PTs, it is not necessary to stop the vehicle to recharge. When the OLEV is put to use, it needs to be charged frequently when it is on the road. Therefore, there should be a suitable billing scheme to control the authenticating and charging interactions between the vehicles and the PTs under the road. And the scheme should protect the location privacy of the vehicle users. If not, an adversary can collect the user location information along the road during the authenticating and charging processes. Collections of charging locations can be used to deduce a driver's residential address, working office, and places of interest, which can be misused for crimes such as robberies or automobile thefts.

However, it is not desirable to provide unconditional location privacy to vehicles' owners because the vehicles need to be traceable in some situations, for example, when the vehicle is stolen or the vehicle is occupied by some criminals. In that case, the police would like to trace the vehicle. Moreover, there should be a trusted party who can trace a user if he/she has violated traffic regulations. In these conditions,

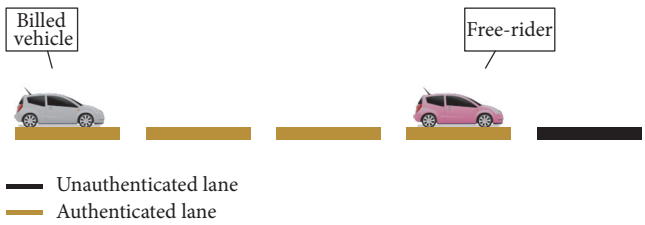


Figure 1: Free-rider in front of a billed vehicle on short lanes.

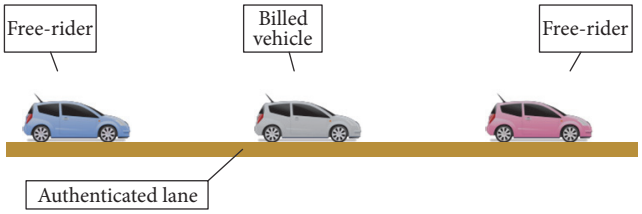


Figure 2: Free-riders near a billed vehicle on a long lane.

providing unconditional privacy is not suitable. Therefore, it is necessary to design a location privacy-preserving billing scheme for OLEV system in which the vehicles can only be traceable by a trusted party. The billing scheme should be efficient and can be carried out via vehicular ad hoc network (VANET) over 4G/5G when the vehicles are moving.

Currently, there exist several anonymous payment methods [6, 8, 9] that are suitable for wireless charging on the move. However, their scheme did not consider the cases of free-riders, and free-riders do exist since the charging is wireless and the charging segment is long enough for several closely moving vehicles (50 m in [10] and longer than one mile in [11]).

There will be several cases of free-riders. If the charging lane is short, each vehicle should authenticate itself with several charging lanes in front of it, in order to keep a moving speed. Then the vehicle driving ahead can slow down to get free charging from behind vehicle that is billed for wireless charging, as shown in Figure 1. If the charging lane is long, the vehicle driving ahead can slow down and the vehicle driving behind can speed up to get free charging, as shown in Figure 2. If the power supply is not constant for each charging lane, several vehicles can move together in a clustered group to get charging with only one vehicle paying for it. Or some vehicles can pay less money to get more supply by moving near other vehicles. These circumstances should have proper solutions before OLEV is put into wide application.

1.1. Our Contributions. In this paper, we present an efficient privacy-preserving billing scheme against free-riders for wireless charging electric vehicles by using several cryptographic tools including encryption scheme, signature scheme, and hash function. The proposed scheme achieves the following features:

(i) The scheme can fight against free-riders. Free-riders can be detected by checking their power levels and their authentication state. Proof of free-riding can be shown to the bank, so that the service provider can receive compensation from the free-riders. The compensation is much more than the fee of a full charging so the punishment can help to restrain free-riders.

(ii) The scheme is privacy-preserving because the billed user receives an anonymous token from bank. When the billed user charges his/her vehicle from any service provider, he/she shows only the anonymous token. His/her identity is not revealed so the location privacy of the user is enhanced.

(iii) The scheme can prevent double spending by employing online double spending checking. Banks only check the double spending of certificates but not the whole transactions, so that the efficiency of transactions is not affected. Banks can also cooperate to setup several distributed servers to alleviate the burdens.

(iv) The scheme is presented as a framework which can adopt the latest efficient public key encryption scheme, digital signature scheme, and hash function, so it can be implemented with the recent advances of modern technologies.

(v) The scheme is a fast authenticating framework which enables the vehicle to access one segment after another sequentially and securely. It uses hash chain receiving-and-acknowledging method to achieve fast authentication, which can be applied to sequential data delivering services in future 5G systems such as high-definition video streaming service.

1.2. Organization. The remainder of this paper is organized as follows. In Section 2 we describe the related works of billing schemes for wireless charging electric vehicles. In Section 3 we describe the system model and security requirements. In Section 4, we describe the proposed privacy-preserving billing scheme. The security of the proposed scheme is analyzed and features are compared in Section 5. Section 6 concludes the paper.

2. Related Works

Though we can obtain many benefits from OLEV, a privacy-preserving billing system is needed before OLEV is widely adopted. Recently, there are several contributions [6–9] that design privacy-preserving authentication and payment methods for OLEV.

Hussain et al. [6, 7] introduce a secure and privacy-aware fair billing framework for OLEV on the move through the charging plates installed under the road. They first propose two extreme lightweight mutual authentication mechanisms, namely, a direct mutual authentication method and a pure hash chain based authentication method. These methods can be used for different vehicular speeds on the road. Then they propose two power transfer and billing schemes separately based on the above two methods.

Zhao et al. [8] propose a secure and privacy-preserving billing scheme for OLEV. Users can buy electric energy

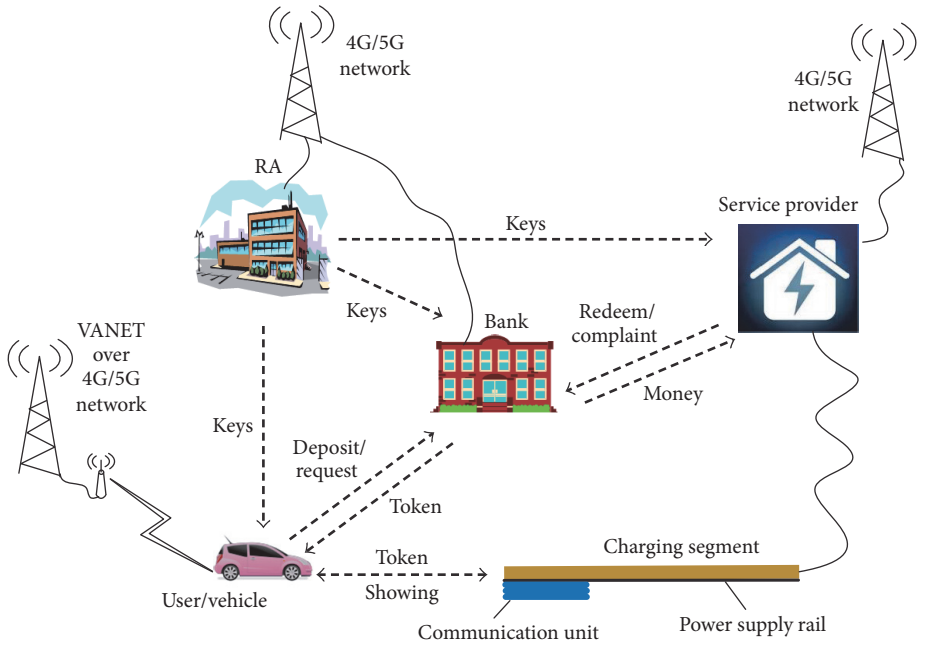


Figure 3: Network model used in the proposed scheme.

from power provider and charge their EVs anonymously and unlinkably. They assume that each PT transmits a fixed amount of energy to the EV and the energy supply company bills the EV the same amount of money for the electric energy from every PT. EVs can buy the energy according to the levels of their batteries.

Rezaeifar et al. [9] propose an efficient payment method based on "tokens" for wireless charging on the move, which minimizes the communications between service providers and users during the charging process. The proposed scheme prevents a user and the service provider from cheating each other, and it is robust to support different values for the price. However, their scheme cannot prevent double spending. A malicious user (e.g., a criminal) can obtain a token with small sum of deposit and distribute it to his/her colleagues. They can charge their vehicles from different service providers or the same service provider without a central certificates management server.

The above schemes cannot fight against free-riders, which may cause insufficient charging to billed user or electric energy lost to service providers.

3. System Model and Security Requirements

3.1. System Participants and Network Model. In summary, our proposed method consists of the following four main parties. First, there is a trusted registration authority (RA) that will generate system parameters for the system and public/private keys and signing/verifying keys for other parties. The second party is the bank, which is responsible for issuing authentication tokens to the users. There can be many banks and all the banks are also trustful that they will not disclose users' privacy and can help to track illegal users if needed. The third is the user who interacts with a bank to obtain his/her tokens and then connects to the power charging service provider to receive an electric charge for his/her EV by using the tokens. The fourth is the power charging service provider who owns the billing server and electrical power delivery service. The electrical power delivery service is to provide vehicles with an electric charge through a charging plate under the road.

The EVs can connect to the bank to receive tokens using vehicle ad hoc networks (VANETs) via a road side unit (RSU) over 4G/5G network when they move on the road, before reaching the charging lane. A certain length of power transmitter is installed under the road to form a charging segment. The charging segment also includes a hardware section for communication and computation purposes. There are many charging segments along the road so each vehicle can obtain enough charging. The communication channel between an EV and a charging segment is based on the 5.9 GHz Dedicated Short Range Communication (DSRC) standards. The charging segment is connected to both the electrical power delivery service and the billing server, and the service provider can communicate with the bank server. We assume that these communications can be done through a secure channel using 4G/5G network. Figure 3 shows the network model of our proposed method.

3.2. Security Requirements. As mentioned in [12], the general requirements for Internet payment systems include security, reliability, and scalability to support various users and service providers without losing efficiency, anonymity, and flexibility. However, we will only discuss some of them in brevity though we have achieved all of them. The focused requirements are listed as follows.

(1) Security against Double Spending. It is one of the main requirements for the payment system. Since the payment

system for wireless charging of EVs is on the networks which are open to the public, the payment method should be secured to avoid such attacks that may occur in an open environment where all interactions can be sniffed by others. And each user can also attempt to spend his/her token several times in order to get multiple charging by paying the money once. If a malicious user can double spend his/her token many times in different service providers, some service providers may not get back the money that they deserve to have.

(2) Fast Authentication. One of the most important features of wireless charging on the move is the fast operation, as vehicles are running with high speeds. Therefore, time-consuming payment methods such as Bitcoin and methods using iterative zero-knowledge proofs are not practical for wireless charging on the move. Therefore, to make our proposed payment system more efficient, it is important to minimize the verification time and the number of exchanged messages during the transactions.

(3) Location Privacy. Since an EV needs to charge frequently throughout the day, the location privacy of the EV can be abused to profile the owners of the EV. Therefore, providing anonymity and preserving location privacy against the service providers and the eavesdroppers are important requirements that we consider in our scheme. We assume that none of service providers and eavesdroppers use camera to record the physical identities of vehicles, since no one can provide anonymity in that case.

(4) Security against Free-Riders. The scheme should be secure against free-riders. As shown in previous section, the free-riders do exist if the OLEV systems are widely deployed. Billing scheme that is resistant against free-riders should be deployed together to avoid electricity loss and uncertain burden imbalance.

3.3. Assumption. The proposed scheme is based on the following assumptions.

Each bank is a trusted entity and only the bank can link the real identity of the user to the token number. Each bank has public and private keys to communicate with other entities, and all the entities can verify the bank's signature with its public key. Moreover, we assume that each bank has secure connections with all service providers.

Electric vehicle users should make an account for this purpose at certain bank. They can receive a defined number of prepaid tokens if they have enough money in the deposit. The bank can also allow users to book a number of tokens according to their credit scores (e.g., with their credit cards), so it does not matter whether there is money in the deposit. How much can be paid by credit cards and how the credit system is operated are out of scope of this paper.

Vehicles are equipped with sensors which can show the users how much charging they need. Each vehicle also uses an On-Board Unit (OBU) to communicate with the charging plate wirelessly. Electric vehicle has a switch for wireless charging. An honest user can switch off the charging circuit if he/she just wants to pass the route and take no charging. Then the OBU on the vehicle will authenticate itself with the charging plate setting the number of charging segments to zero.

The power charging service provider checks battery power level values of a vehicle periodically to decide whether it is actually in the state of charging. Since some user may report the values dishonestly in order to avoid being detected when his/her vehicle is getting a free-riding, we assume that each OBU is equipped with tamper-proof meter to carry out battery power level checking. The tamper-proof meter reports the value in a certified form [13]. Similar assumptions can be found in many smart metering schemes such as [14–16]. They assume that smart meters are fully trusted or the readings from meters are certified.

3.4. Thread Model. In our threat model, we assume that both the service provider and the EV (representing a user) can be malicious. We consider several kinds of attacks for malicious behavior, namely, free-rider, location privacy infringement, and double spending. Free-riding behaviors can be malicious in terms of bypassing the billing process such as refusing to give an authenticated hash value after receiving a charge, or driving in front of or following behind a charging vehicle closely to get a free charging. Besides, the service provider can abuse the users' privacy by tracking the EV and giving location information to a third party, such as advertising agencies, and so forth. Furthermore, the user can double spend his/her tokens or the adversaries can sniff the communication between charging plates and the EVs to collect information for double spending and user identification.

4. The Proposed Billing Schemes against Free-Riders

In this section, we describe the privacy-preserving billing schemes for wireless charging electric vehicles against free-riders.

4.1. Our Idea. In order to get rid of free-riders, we suggest that two methods be used together. One is prepaid anonymous token. Each vehicle should always obtain a token from bank by depositing money more than a maximum compensation fee of free-riders before using the route for OLEV. Or the token is paid by credit cards. The token can be a charging token which means the vehicle wants to choose the route to get to the destination at the same time charging for a number of segments. The token can be an entrance token which means the vehicle wants to take the route without charging for some reasons (e.g., the way is a shortcut or other ways are jammed). If one vehicle enters a route that is for OLEV, it should authenticate itself with one token. If this vehicle is detected for free-riders, the token is sent to bank together with proof of free-riding. The service provider will get money for compensation. Another required method is detecting the power level of each vehicle now and then. The detection is interaction taken between the charging plates (or the service provider) and tamper-proof meter inside OBU, so the vehicle cannot cheat on the power level of its battery. If the power

level of some vehicle is increased while the vehicle is not in an authenticated state of charging or has finished charging, this vehicle is blamed for free-riders.

4.2. Cryptographic Tools. We need several cryptographic tools to construct the proposed billing scheme, including a secure signature scheme supporting batch verification, a secure encryption scheme, and a secure hashing function. Some notations used in this paper are listed in Notations.

(i) Encryption Scheme. It can be any efficient secure public key encryption scheme which can be denoted as the tuple (ES-SETUP(1^λ), $\mathrm{ENC}_{\mathrm{PK}}(\cdot)$, $\mathrm{DEC}_{\mathrm{SK}}(\cdot)$). ES-SETUP($1^\lambda$) is used to generated parameters for the scheme given a secure parameter λ and then public/private key pair (PK, SK) for each participant. $\mathrm{ENC}_{\mathrm{PK}}(\cdot)$ means encrypt something with key PK. $\mathrm{DEC}_{\mathrm{SK}}(\cdot)$ means decrypt something with key SK. For instance, encryption schemes such as Elgamal [17] and elliptic curve cryptography [18] can be used.

(ii) Signature Scheme. It can be any efficient secure digital signature scheme, and it will be better if batch verification is supported. The scheme can be denoted as the tuple (DS-SETUP(1^λ), $\mathrm{SIG}_{\mathrm{CK}}(\cdot)$,$\mathrm{VER}_{\mathrm{VK}}(\cdot)$). DS-SETUP($1^\lambda$) is used to generate parameters for the scheme given a secure parameter λ and then signing/verifying key pair (CK, VK) for each participant. $\mathrm{SIG}_{\mathrm{CK}}(\cdot)$ means signing something with key CK. $\mathrm{VER}_{\mathrm{VK}}(\cdot)$ means verify some signature with key VK. Schemes with batch verification ability [19, 20] will be better because the service provider can verify multiple signatures in one round and detect bogus signatures quickly.

(iii) Hash Function. We require the hash function (denoted as $H(\cdot)$) to be an efficient collusion secure one-way hash function with input and output values in the same domain. In other words, the output value can be fed as input. We need this feature to generate hash chains like $h_1 = H(h_0)$, $h_2 = H(h_1)$, $h_3 = H(h_2)$, and so on, if given an initial value h_0.

4.3. The Schemes

4.3.1. Setup. In this phase, the trusted registration authority (RA) selects a large number λ according to the security requirement of the application and generates the system parameters *Params* by running ES-SETUP(1^λ) and DS-SETUP(1^λ). We assume that all entities are properly authenticated with RA and receive the required system parameters. RA generates their permanent private and public keys and transmits to each entity securely. A bank with identity B_i ($i = 1, 2, \ldots$) will receive a pair of public/private keys ($\mathrm{PK}_{B_i}, \mathrm{SK}_{B_i}$) and a pair of signing/verifying key ($\mathrm{CK}_{B_i}, \mathrm{VK}_{B_i}$). A service provider with identity S_i ($i = 1, 2, \ldots$) will receive a pair of public/private keys ($\mathrm{PK}_{S_i}, \mathrm{SK}_{S_i}$) and a pair of signing/verifying key ($\mathrm{CK}_{S_i}, \mathrm{VK}_{S_i}$). A user with identity U_i ($i = 1, 2, \ldots$) will receive a smart card embedded with a pair of public/private keys ($\mathrm{PK}_{U_i}, \mathrm{SK}_{U_i}$) and a pair of signing/verifying key ($\mathrm{CK}_{U_i}, \mathrm{VK}_{U_i}$). Moreover, electric vehicles are equipped with a tamper-proof meter (part of the OBU) to carry out battery power level checking. And each user should insert his/her smart card into the OBU before operating. We assume that all parties are consistent with the standard time so they do not have any dispute on timestamps.

4.3.2. Token Obtaining. When a user U_i wants to charge his/her vehicle along the way to a destination, he/she can use the On-Board Unit (OBU) to obtain a token by connecting to his/her bank B_j via RSU or VANET over a cellular network. The token is prepaid with two parts of money by deposit or by credit cards. One is for free-riders compensation; the other is the cost for the wireless charging from a number of charging plate segments. The token is retrieved in three steps as follows.

(1) U_i obtains a random pair of temporary signing/verifying keys ($\mathrm{TCK}_{U_i}, \mathrm{TVK}_{U_i}$) and a hash chain h_0, $h_1 = H(h_0), \ldots, h_n = H(h_{n-1})$. The key pair should be qualified for the selected signature scheme, and h_0 should be selected randomly. The key pair and the hash chain can be generated in leisured time. U_i sends a request message to the bank. The message contains the temporary verifying key TVK_{U_i}, user's identity U_i, an expected expiration time T, the end value of hash chain h_n, and the number n for the requested token. n is the expected number of the charging plate segments that the vehicle needs. If the user wants to drive along the way to a destination without charging for some reason (e.g., the way is a shortcut or other ways are jammed), he/she should set $n = 0$. T should be bounded according to the application, for instance 4 hours later from now. The request message is signed with user's signing key CK_{U_i} and then encrypted with bank's public key PK_{B_j}. The request message may be sent through a protected channel such as a transport layer security protocol (TLS) link. Let $M_1 = (\mathrm{TVK}_{U_i}, U_i, T, h_n, n)$. The request message is denoted as

$$U_i \longrightarrow B_j : C_1 = \mathrm{ENC}_{\mathrm{PK}_{B_j}}\left(M_1, \mathrm{SIG}_{\mathrm{CK}_{U_i}}(M_1)\right). \quad (1)$$

(2) When receiving the request, the bank B_j decrypts the message with its secret key SK_{B_j}. The process can be denoted as $(M_1, \mathrm{SIG}_{\mathrm{CK}_{U_i}}(M_1)) = \mathrm{DEC}_{\mathrm{SK}_{B_j}}(C_1)$ and M_1 is parsed as $(\mathrm{TVK}_{U_i}, U_i, T, h_n, n)$. It verifies the signature of user and the expected expiration time T. It also checks whether the user has enough money in the deposit or enough credit score for the tokens. If all are qualified, the bank generates the token for user and the corresponding amount of money is frozen. The token is the bank's signature to $(\mathrm{TVK}_{U_i}, T, h_n, n)$. The user should refresh to obtain a new token if the token is not spent before the expiration time. Let $M_2 = (\mathrm{TVK}_{U_i}, T, h_n, n)$. The returned message is denoted as

$$B_j \longrightarrow U_i : C_2 = \mathrm{ENC}_{\mathrm{PK}_{U_i}}\left(\mathrm{SIG}_{\mathrm{CK}_{B_j}}(M_2)\right). \quad (2)$$

(3) After receiving the returned message, U_i decrypts with the secret key SK_{U_i} and does the verification with the verifying key VK_{B_j}. If the signature is correct, $\mathrm{SIG}_{\mathrm{CK}_{B_j}}(\mathrm{TVK}_{U_i}, T, h_0, n)$ is stored as the token. As we notice, the token is the bank's signature on the temporary verifying key TVK_{U_i}, which is a randomly generated value and does not cause any privacy leakage of the user. At the same time, the bank knows the connection between U_i and TVK_{U_i}, so it can take method to punish the user if there is any malicious behavior.

4.3.3. Token Using. Before any vehicle enters the wireless charging road, it should be authenticated with a token. There are two cases for the authentication: normal charging and passing by without charging. No matter which case, when an EV reaches the entrance of the charging road, the EV should connect and show its token to the service provider. After that, the EV should periodically answer the service provider's query on battery power level. We explain these separate interactions in detail as follows.

(1) Entering. Each vehicle should finish this interaction at the entrance of charging road, or it will be rejected. The user U_i sends a message $(\mathrm{TVK}_{U_i}, T, h_n, n, \mathrm{SIG}_{\mathrm{CK}_{B_j}}(\mathrm{TVK}_{U_i}, T, h_n, n))$ to the service provider S_k together with the number of segments m ($m \leq n$) that it needs for charging. The user sets $m = 0$ indicating that this vehicle is taking the road without charging. Then, the service provider S_k checks the validity of the signature and the expiration time T and then forwards the message to bank B_j. If the signature is valid, T is not expired and bank B_j shows the token is spent for the first time; the service provider replies with a message (co, S_k, T_e, $\mathrm{SIG}_{\mathrm{CK}_{S_k}}(\mathrm{co}, S_k, T_e, h_n)$) that contains the cost of each segment of the charging plate co, its identity S_k, and a timestamp T_e. The timestamp is served as entering time of this vehicle and also used to prevent reply attacks. co can be varied according to m so as to attract customers by using price strategy. The user computes $\mathrm{SIG}_{\mathrm{TCK}_{U_i}}(m, \mathrm{co}, S_k, T_e)$ as message and sends it to the service provider. With this message, the service provider checks the validity of this message with TVK_{U_i}. If the signature is valid, the service provider now assures the authentication of this vehicle and lets it enter the charging road. This interaction can be denoted as

$$\begin{aligned} &U_i \longrightarrow S_k : \mathrm{TVK}_{U_i}, T, h_n, n, \mathrm{SIG}_{\mathrm{CK}_{B_j}}\left(\mathrm{TVK}_{U_i}, T, h_n, n\right), m; \\ &S_k \longrightarrow U_i : \mathrm{co}, S_k, T_e, \mathrm{SIG}_{\mathrm{CK}_{S_k}}\left(\mathrm{co}, S_k, T_e, h_n\right); \\ &U_i \longrightarrow S_k : \mathrm{SIG}_{\mathrm{TCK}_{U_i}}\left(m, \mathrm{co}, S_k, T_e\right). \end{aligned} \tag{3}$$

(2) Charging. This interaction should be taken between the service provider and the vehicles that need wireless charging. If the token is valid and $m > 0$, the service provider switches on the first charging segment to this vehicle. After receiving a charge from each segment, the vehicle should report the value of the hash chain to the service provider sequentially. The message after the first charge should be $(T_{m-1}, h_{m-1}, \mathrm{TVK}_{U_i})$, where T_{m-1} is current time and the sequential hashing of h_{m-1} should lead to h_n. It means the service provider can verify the following equations $h_m = H(h_{m-1})$, $h_{m+1} = H(h_m), \ldots, h_n = H(h_{n-1})$. If the hash chain is valid, the service provider switches on the next charging segment to this vehicle and sends a confirming message to the vehicle $\mathrm{SIG}_{\mathrm{CK}_{S_k}}(T_{m-1}, h_{m-1}, \mathrm{TVK}_{U_i})$. Then the vehicle sends the second message $(T_{m-2}, h_{m-2}, \mathrm{TVK}_{U_i})$, where the service provider can validate it as $h_{m-1} = H(h_{m-2})$ and switches on the third charging segment. The service provider also returns a signature $\mathrm{SIG}_{\mathrm{CK}_{S_k}}(T_{m-2}, h_{m-2}, \mathrm{TVK}_{U_i})$ on the message containing the new time T_{m-2}. We should remark that the user should reject releasing the next value of the hash chain if the vehicle did not receive a charge from current segment or did not receive the confirming message, where the messages can be used to prove its innocence. Since $H(\cdot)$ is a secure one-way hash function, the service provider cannot figure out the input value from the given sequence of output values. Thus, in order to redeem the money from the bank, the service provider should treat the vehicle fairly to obtain a full hash value chain. This interaction will continue until the user sends $(T_0, h_0, \mathrm{TVK}_{U_i})$ to the service provider and receives a charging and a confirming message from the mth segment. This interaction can be denoted as

$$\begin{aligned} &U_i \text{ gets the first charging} \\ &U_i \longrightarrow S_k : T_{m-1}, h_{m-1}, \mathrm{TVK}_{U_i}; \\ &S_k \longrightarrow U_i : \mathrm{SIG}_{\mathrm{CK}_{S_k}}\left(T_{m-1}, h_{m-1}, \mathrm{TVK}_{U_i}\right); \\ &U_i \text{ gets the second charging} \\ &U_i \longrightarrow S_k : T_{m-2}, h_{m-2}, \mathrm{TVK}_{U_i}; \\ &S_k \longrightarrow U_i : \mathrm{SIG}_{\mathrm{CK}_{S_k}}\left(T_{m-2}, h_{m-2}, \mathrm{TVK}_{U_i}\right); \\ &\vdots \\ &U_i \text{ gets the } m\text{th charging} \\ &U_i \longrightarrow S_k : T_0, h_0, \mathrm{TVK}_{U_i}; \\ &S_k \longrightarrow U_i : \mathrm{SIG}_{\mathrm{CK}_{S_k}}\left(T_0, h_0, \mathrm{TVK}_{U_i}\right). \end{aligned} \tag{4}$$

We notice that it is possible that the user refuses to offer a sequent hash value after receiving the ith ($0 \leq i \leq m-1$) charging from the service provider. In this case, the service provider will treat it as a free-rider. The service provider saves the interactions and the periodical power reporting and reports them to the bank to obtain the compensation. The processing is described in the following *Power Reporting* and *Redeeming* phases.

(3) Power Reporting. This interaction should be taken between the service provider and the vehicles entering the charging road. The service provider periodically (e.g., once per segment or every 20 seconds) queries these vehicles with a message as $(\mathrm{TVK}_{U_i}, T_q)$. Each vehicle should answer its battery power level Pow_q to the query with a form as $(\mathrm{Pow}_q, \mathrm{SIG}_{\mathrm{TCK}_{U_i}}(\mathrm{Pow}_q, \mathrm{TVK}_{U_i}, T_q))$, where Pow_q is generated in a certified form [13] by the tamper-proof module embedded in vehicle's OBU. If Pow_q is detected increasing during the interactions while the vehicle is not in the authenticated state of charging, the service provider decides the vehicle is acting as a free-rider. It can alert this vehicle and save the report as the proof for later use. This interaction can be denoted as

$$\begin{aligned} &S_k \longrightarrow U_i : \mathrm{TVK}_{U_i}, T_q; \\ &U_i \longrightarrow S_k : \mathrm{Pow}_q, \mathrm{SIG}_{\mathrm{TCK}_{U_i}}\left(\mathrm{Pow}_q, \mathrm{TVK}_{U_i}, T_q\right). \end{aligned} \tag{5}$$

4.3.4. Redeeming. In normal circumstances, the service provider builds a message from those received in Entering and Charging interactions of the Token Using phase to the bank B_j to redeem the token. The message is of the form $(M_{E_1}, M_{E_2}, M_{\text{hash}}, M_{\text{sig}})$, where $M_{E_1} = (\text{TVK}_{U_i}, T, h_n, n, \text{SIG}_{\text{CK}_{B_j}}(\text{TVK}_{U_i}, T, h_n, n))$, $M_{E_2} = (m, \text{co}, S_k, T_e, \text{SIG}_{\text{TCK}_{U_i}}(m, \text{co}, S_k, T_e))$, $M_{\text{hash}} = (h_0, h_1, \ldots, h_n)$, and $M_{\text{sig}} = \text{SIG}_{\text{CK}_{S_k}}(M_{E_1}, M_{E_2}, M_{\text{hash}})$. The bank can verify whether all the above signatures are valid and whether it is a transaction between a normal user and the specified service provider. The service provider redeems the money for that charging and the rest will be returned the user's deposit. If the M_{hash} is incomplete (such as $M_{\text{hash}} = h_l, \ldots, h_m, \ldots, h_n$), the service provider can only redeem the money for $m - l$ segments. This interaction can be denoted as

$$\begin{aligned} M_{E_1} &= \text{TVK}_{U_i}, T, h_n, n, \text{SIG}_{\text{CK}_{B_j}}\left(\text{TVK}_{U_i}, T, h_n, n\right); \\ M_{E_2} &= m, \text{co}, S_k, T_e, \text{SIG}_{\text{TCK}_{U_i}}\left(m, \text{co}, S_k, T_e\right); \\ M_{\text{hash}} &= h_0, h_1, \ldots, h_n; \\ M_{\text{sig}} &= \text{SIG}_{\text{CK}_{S_k}}\left(M_{E_1}, M_{E_2}, M_{\text{hash}}\right); \\ S_k &\longrightarrow B_j : M_{E_1}, M_{E_2}, M_{\text{hash}}, M_{\text{sig}}. \end{aligned} \tag{6}$$

In a circumstance where at least one suspected free-rider exists, the service provider builds a message from those received in all interactions of the Token Using phase to the bank. The message is of the form $(M_{E_1}, M_{E_2}, M_{\text{hash}}, M_{\text{rep}}, M_{\text{sig}})$, where $M_{E_1} = (\text{TVK}_{U_i}, T, h_n, n, \text{SIG}_{\text{CK}_{B_j}}(\text{TVK}_{U_i}, T, h_n, n))$, $M_{E_2} = (m, \text{co}, S_k, T_e, \text{SIG}_{\text{TCK}_{U_i}}(m, \text{co}, S_k, T_e))$, M_{hash} is set to $(h_l, \ldots, h_n)$ where h_l $(l \leq n)$ is the last valid hash value received from the vehicle or a null value indicating that the vehicle claims no charging, $M_{\text{rep}} = (\ldots, \text{Pow}_q, \text{SIG}_{\text{TCK}_{U_i}}(\text{Pow}_q, \text{TVK}_{U_i}, T_q), \ldots)$, and $M_{\text{sig}} = \text{SIG}_{\text{CK}_{S_k}}(M_{E_1}, M_{E_2}, M_{\text{hash}}, M_{\text{rep}})$. They are listed as follows.

$$\begin{aligned} M_{E_1} &= \text{TVK}_{U_i}, T, h_n, n, \text{SIG}_{\text{CK}_{B_j}}\left(\text{TVK}_{U_i}, T, h_n, n\right); \\ M_{E_2} &= m, \text{co}, S_k, T_e, \text{SIG}_{\text{TCK}_{U_i}}\left(m, \text{co}, S_k, T_e\right); \\ M_{\text{hash}} &= h_l, \ldots, h_n; \\ M_{\text{rep}} &= \ldots, \text{Pow}_q, \text{SIG}_{\text{TCK}_{U_i}}\left(\text{Pow}_q, \text{TVK}_{U_i}, T_q\right), \ldots; \\ M_{\text{sig}} &= \text{SIG}_{\text{CK}_{S_k}}\left(M_{E_1}, M_{E_2}, M_{\text{hash}}, M_{\text{rep}}\right); \\ S_k &\longrightarrow B_j : M_{E_1}, M_{E_2}, M_{\text{hash}}, M_{\text{sig}}. \end{aligned} \tag{7}$$

The bank can verify whether all the above signatures are valid and whether there is a valid proof of some vehicle being free-rider. If so, the bank accepts the complaint and calls the suspected user to show the proof for its innocence. If the user cannot show the confirming messages holding a sequence of timestamps that cover the complained time period, the user will be blamed for being a free-rider and compensation fee is given to the service provider. If the user can show the confirming messages, the bank decides the service provider is wrong and the user is innocent. How to punish the service provider is out of scope of this paper.

5. Security Analysis and Comparisons

We first discuss the security of our proposed scheme. Then, we compare our method with several schemes proposed for plug-in electric vehicles and wireless charging electric vehicles.

5.1. Security Analysis. According to the security requirements presented in Section 3, we will discuss the double spending avoidance, location privacy infringement, and free-riders resistance. For each of them, we discuss how our method can achieve them.

(1) Security against Double Spending. Double spending may occur when an adversary replays a sniffed token or certain user shows the already-spent token again. We will discuss them in the following two cases, respectively, as follows.

Case 1. The adversary can access the messages exchanged between the owner of the tokens and the service provider. Since the adversary does not possess the secret signing key of that token, he/she cannot show its validly to any service provider.

Case 2. Each user as the owner of a token may spend the token several times through different service providers since he/she holds the token's private key. However, double spending of the same token will be detected when the token is forwarded to local certificate management server setup by the bank. Token matching is fast in local server since each token has a limited validation time period and the server does not need to store large amount of tokens.

(2) Location Privacy. A secure channel is established between EVs and the bank by using signature and encryption schemes, so the adversary cannot obtain any information about token received by the specific EV. The temporary verification key and the hash value in each token are generated randomly by its owner, so the adversary cannot link one to another and the privacy of each EV is enhanced. Even the service provider cannot track the vehicles by analyzing all messages in the interactions unless they use a camera in charging places to record the physical identities of vehicles.

(3) Security against Free-Riders. The battery power level of each vehicle is checked periodically when it is moving. If power level is increasing while the vehicle is not in the authenticated state of being charged, it will be treated as a free-rider. The battery power level is reported by a tamper-proof module embedded in vehicle's OBU, so the user cannot cheat on it. The service provider can complain to the bank to obtain the compensation. If the compensation is confirmed by the bank, the free-rider will have to pay certain amount of money that is more than the battery charging it received. The money was prepaid by the deposit or by credit cards, so the

Table 1: Comparison with previous works.

	Feature	Location privacy	Price flexibility	Detect double spending	Prevent double spending	Avoid fraudulence in charging	Track illegal user	Avoid free-rider
[4]	Plug-in EV	√	not specified	√	√	×	√	Not concern†
[5]	Plug-in EV	√	√	√	√	√	×#	Not concern†
[6, 7]	OLEV	√‡	√	√	√	√	√	×
[8]	OLEV	√	√	×	√	√	×	×
[9]	OLEV	√	√	√	×	√	√	×
Proposed scheme	OLEV	√	√	√	√	√	√	√

†: the schemes for plug-in EV do not care about free-rider.
#: the scheme can track a stolen vehicle with consent from its owner but cannot track illegal user.
‡: weak protection that CP can link all authentications of the same vehicle.

free-rider should pay the compensation or get his/her credit score decreased.

5.2. Comparisons. In this section, we compare different features of our method with the proposed methods of two billing schemes for plug-in electric vehicles [4, 5] and four schemes for wireless charging electric vehicles [6–9]. The comparison is shown in Table 1, which focuses on several capabilities including location privacy, flexibility of charging price, detecting double spending, preventing double spending, avoiding fraudulence in charging, tracking illegal user, and avoiding free-rider.

The scheme [4] is for plug-in electric vehicles. It uses hash function and online authentication to achieve its designing goal. But it cannot avoid fraudulence in charging since the charging station can refuse to charge the vehicle after receiving the pseudonymous public key and the signed request. The scheme [5] cannot track illegal users since each user will be totally anonymous without consent value from the user. In the scheme by Hussain et al. [6, 7], the department of mobile vehicle (DMV) selects a set of pseudonyms for each vehicle and the vehicle authenticates itself to the charging plate (CP) using the hash value X_{OBU} of this set. However, all authentications of the same vehicle use the same hash value X_{OBU}, so CP can link all the charging locations of this vehicle together though it does not know who owns this vehicle. Thus, their scheme achieves a weak privacy protection. In the scheme of [8], zero-knowledge proof is used in all phases so the user is unconditional privacy-preserving. Thus, their scheme cannot track illegal users. In their scheme, the authenticating values sent to CP will be used once and then discarded, so the action of double spending will cause authentication failure in CP but CP cannot know whether the voucher has been spent or not. As a result, their scheme cannot detect double spending though it is very easy to achieve. In the scheme of [9], the service provider authenticates each user in an offline form, so the user can double spend his/her token elsewhere with different service providers. All the above schemes cannot fight against free-riders, which is the main focus of this paper.

6. Conclusion

We present an efficient privacy-preserving billing scheme for wireless charging electric vehicles against free-riders by using compensation-prepaid tokens and detecting battery power levels periodically when the vehicles are moving on the road. We need a tamper-proof device inside OBU, so the vehicle will report the power level of its battery honestly. How to get rid of the tamper-proof device will be an interesting topic for future research in smart grid and vehicle to grid (V2G) network.

Notations

λ:	A security parameter that measures the input size of the computational problem in cryptography
U_i:	The identity of a user with index i
B_j:	The identity of a bank with index j
S_k:	The identity of a service provider with index k
PK:	A public key used in the encryption scheme, for example, PK_{U_i} is the public key of U_i
SK:	A private key used in the encryption scheme, for example, SK_{B_j} is the private key of B_j
CK:	A signing key used in the signature scheme
VK:	A verifying key used in the signature scheme
ES-SETUP(1^λ):	The algorithm to setup the system parameters for the encryption scheme
$\mathrm{ENC}_{\mathrm{PK}}(m)$:	The algorithm that encrypts a message m with a public key PK to generate a ciphertext
$\mathrm{DEC}_{\mathrm{SK}}(c)$:	The algorithm that decrypts a ciphertext c with a private key SK to recover a message
DS-SETUP(1^λ):	The algorithm to setup the system parameters for the signature scheme

$\mathrm{SIG}_{\mathrm{CK}}(m)$:	The algorithm that signs a message m with a signing key CK to generate a digital signature
$\mathrm{VER}_{\mathrm{VK}}(\sigma)$:	The algorithm that verifies a signature σ with a verifying key VK
$H(m)$:	The algorithm that maps data m of arbitrary size to a hash value of fixed size
TCK:	A temporary signing key used in the signature scheme
TVK:	A temporary verifying key used in the signature scheme
$B_j \rightarrow U_i : m$:	An entity B_j sends another entity U_i a message m
h_i:	A hash value with index i
T, T_c, T_e, T_q:	Some timestamps used in different phases
OBU:	On-Board Unit equipped in a vehicle
VANET:	Vehicular ad hoc network
RSU:	Road Side Units deployed along the roadside to help to construct VANET.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work is supported by the Research Fund for the Doctoral Program of Higher Education of China (no. 20130203120003), the National Natural Science Foundation of China (no. U1401251), Major Basic Research Program of Shaanxi Province Natural Science Foundation Research Project (no. 2016ZDJC-04), and the China 111 Project (no. B16037).

References

[1] E. Valsera-Naranjo, A. Sumper, P. Lloret-Gallego, R. Villafáfila-Robles, and A. Sudria-Andreu, "Electrical vehicles: state of art and issues for their connection to the network," in *Proceedings of the 10th International Conference on Electrical Power Quality and Utilisation (EPQU '09)*, pp. 1–3, Łódź, Poland, September 2009.

[2] P. Dutta, "Coordinating rendezvous points for inductive power transfer between electric vehicles to increase effective driving distance," in *Proceedings of the 2nd IEEE International Conference on Connected Vehicles and Expo (ICCVE '13)*, pp. 649–653, December 2013.

[3] N. P. Suh, D. H. Cho, and C. T. Rim, *Design of On-Line Electric Vehicle (OLEV)*, Springer, Berlin, Germany, 2011.

[4] H. Nicanfar, S. Hosseininezhad, P. Talebifard, and V. C. M. Leung, "Robust privacy-preserving authentication scheme for communication between electric vehicle as power energy storage and power stations," in *Proceedings of the 32nd IEEE Conference on Computer Communications (IEEE INFOCOM '13)*, pp. 3429–3434, Turin, Italy, April 2013.

[5] M. H. Au, J. K. Liu, J. Fang, Z. L. Jiang, W. Susilo, and J. Zhou, "A new payment system for enhancing location privacy of electric vehicles," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 1, pp. 3–18, 2014.

[6] R. Hussain, D. Kim, M. Nogueira, J. Son, A. O. Tokuta, and H. Oh, "PBF: a new privacy-aware billing framework for online electric vehicles with bidirectional auditability," https://arxiv.org/abs/1504.05276.

[7] R. Hussain, D. Kim, M. Nogueira, J. Son, A. Tokuta, and H. Oh, "A new privacy-aware mutual authentication mechanism for charging-on-the-move in online electric vehicles," in *Proceedings of the 11th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN '15)*, pp. 108–115, Shenzhen, China, December 2015.

[8] T. Zhao, L. Wei, and C. Zhang, "A secure and privacy-preserving billing scheme for online electric vehicles," in *Proceedings of the IEEE 83rd Vehicular Technology Conference (VTC '16)*, pp. 1–5, IEEE, Nanjing, China, May 2016.

[9] Z. Rezaeifar, R. Hussain, S. Kim, and H. Oh, "A new privacy aware payment scheme for wireless charging of electric vehicles," *Wireless Personal Communications*, pp. 1–18, 2016.

[10] Y. D. Ko, Y. J. Jang, and S. Jeong, "Mathematical modeling and optimization of the automated wireless charging electric transportation system," in *Proceedings of the IEEE International Conference on Automation Science and Engineering: Green Automation Toward a Sustainable Society (CASE '12)*, pp. 250–255, Seoul, Korea, August 2012.

[11] Z. Chen, F. He, and Y. Yin, "Optimal deployment of charging lanes for electric vehicles in transportation networks," *Transportation Research Part B: Methodological*, vol. 91, pp. 344–365, 2016.

[12] B. Neuman and G. Medvinsky, "Requirements for network payment: the NetCheque perspective," in *Proceedings of the Digest of Papers. Technologies for the Information Superhighway (COMPCON '95)*, pp. 32–36, San Francisco, Calif, USA, 1995.

[13] A. Rial and G. Danezis, "Privacy-preserving smart metering," in *Proceedings of the 10th Annual ACM Workshop on Privacy in the Electronic Society (WPES '11)*, Y. Chen and J. Vaidya, Eds., pp. 49–60, Chicago, Ill, USA, October 2011.

[14] C. Rottondi, G. Verticale, and C. Krauss, "Distributed privacy-preserving aggregation of metering data in smart grids," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 7, pp. 1342–1354, 2013.

[15] F. Diao, F. Zhang, and X. Cheng, "A privacy-preserving smart metering scheme using linkable anonymous credential," *IEEE Transactions on Smart Grid*, vol. 6, no. 1, pp. 461–467, 2015.

[16] H. J. Jo, I. S. Kim, and D. H. Lee, "Efficient and privacy-preserving metering protocols for smart grid systems," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1732–1742, 2016.

[17] T. E. Gamal, "A public key cryptosystem and a signature scheme based on discrete logarithms," in *Advances in Cryptology, Proceedings of CRYPTO '84, Santa Barbara, California, USA, August 19-22, 1984, Proceedings*, pp. 10–18, 1984.

[18] H. Cohen, G. Frey, R. Avanzi et al., Eds., *Handbook of Elliptic and Hyperelliptic Curve Cryptography*, Chapman and Hall/CRC, 2005.

[19] Y. Jiang, M. Shi, X. Shen, and C. Lin, "BAT: a robust signature scheme for vehicular networks using binary authentication tree," *IEEE Transactions on Wireless Communications*, vol. 8, no. 4, pp. 1974–1983, 2009.

[20] M. Bayat, M. Barmshoory, M. Rahimi, and M. R. Aref, "A secure authentication scheme for VANETs with batch verification," *Wireless Networks*, vol. 21, no. 5, pp. 1733–1743, 2015.

Hindawi
Mobile Information Systems
Volume 2017, Article ID 2084684, 18 pages
https://doi.org/10.1155/2017/2084684

Research Article

A Window-Based, Server-Assisted P2P Network for VoD Services with QoE Guarantees

Noé Torres-Cruz,[1,2] Mario E. Rivero-Angeles,[1] Gerardo Rubino,[3] Ricardo Menchaca-Mendez,[1] and Rolando Menchaca-Mendez[1]

[1]*Centro de Investigación en Computación, Instituto Politécnico Nacional, Av. Juan de Dios Bátiz, Esq. Miguel, Othón de Mendizábal, Col. Nueva Industrial Vallejo, 07738 Ciudad de México, Mexico*
[2]*UPIITA, Instituto Politécnico Nacional, Av. IPN, No. 2580, Col. Barrio la Laguna Ticomán, 07340 Ciudad de México, Mexico*
[3]*INRIA Rennes, Bretagne Atlantique, Campus Universitaire de Beaulieu, 35042 Rennes Cedex, France*

Correspondence should be addressed to Noé Torres-Cruz; ntorresc@ipn.mx

Received 29 October 2016; Revised 19 January 2017; Accepted 1 February 2017; Published 23 March 2017

Academic Editor: Ben Niu

We describe a Peer-to-Peer (P2P) network that is designed to support Video on Demand (VoD) services. This network is based on a video-file sharing mechanism that classifies peers according to the window (segment of the file) that they are downloading. This classification easily allows identifying peers that are able to share windows among them, so one of our major contributions is the definition of a mechanism that could be implemented to efficiently distribute video content in future 5G networks. Considering that cooperation among peers can be insufficient to guarantee an appropriate system performance, we also propose that this network must be assisted by upload bandwidth from servers; since these resources represent an extra cost to the service provider, especially in mobile networks, we complement our work by defining a scheme that efficiently allocates them only to those peers that are in windows with resources scarcity (we called it *prioritized windows distribution scheme*). On the basis of a fluid model and a Markov chain, we also developed a methodology that allows us to select the system parameters values (e.g., windows sizes or minimum servers upload bandwidth) that satisfy a set of Quality of Experience (QoE) parameters.

1. Introduction

Per-to-Peer (P2P) networks have been widely used to increase the capacity of systems due to the fact that nodes cooperate among them to reduce data traffic at the servers. Unlike conventional client/server systems, which experience a stark performance degradation when the number of clients increases, P2P networks are able to scale much better because their capacity also increases with the number of users. Originally, P2P networks were designed to distribute files whose download times were not very restricted, since those files were supposed to be used only after their download completion. However, in recent years a large body of research has been focused on analyzing these networks for video distribution. In these services it is necessary to consider that the video playback is initiated even if its download is still in progress. Services like Live IPTV and Video on Demand (VoD) have been considered in the context of P2P networks in [1–11].

VoD streaming has become widely popular because of its particular features: subscribers are allowed to select and play back a video as well as to rewind, fast-forward, pause, or stop it; and the playback startup time (also known as initial delay) is considerably reduced since the playback can be initiated even if the video download is not completed. These features have made VoD services very attractive and nowadays represent an important proportion of the current Internet traffic. According to [12], VoD represented approximately 23% of the total traffic in mobile networks in North America, during 2015; and it is expected that this trend will continue in future mobile networks [13].

Considering this, it is highly probable that P2P networks will be a key technology for the distribution of video content

in the next generation of wireless communications, including the fifth generation of mobile systems (5G); indeed, several works have recently addressed this issue by proposing strategies to allocate resources in this context [14, 15]. In this paper, we analyze a P2P network for VoD services by means of both a fluid model and Markov chains. These analytic tools capture the main characteristics of the distribution of a video file among the nodes in a P2P network. Building on this, we obtain different performance metrics that accurately describe the quality of the video streaming service as experienced by a user in terms of initial delay and duration of the pauses along the playback. These parameters are considered in order to obtain a Quality of Experience (QoE) score as detailed further in the paper.

One protocol that has had a lot of impact in the development of P2P networks is BitTorrent. In this protocol, the main idea is to divide a video file into many pieces called *chunks*. The peers download a specific video file by exchanging the corresponding chunks according to some rules. The BitTorrent protocol differentiates two types of peers: *leeches*, which are peers that have a subset (possibly the empty one) of the chunks that compose the file, and *seeds*, which are peers that have downloaded the whole file and remain in the system to share their resources. Both leeches and seeds cooperate to upload the file to other leeches. Whenever a peer joins the system to download the file, it contacts a particular node called *tracker* which has the complete list of peers that have part or all the file's content. Then, the tracker returns a random list of potential peers that might share the file with the arriving peer. At this point, the downloading peer contacts the peers on the list and establishes which chunks it is willing to download from each peer it is connected with.

BitTorrent is not suited for VoD applications since chunks are distributed over the network in no particular order, while VoD services require a specific download order to guarantee a low initial delay. However, BitTorrent can still be used for streaming VoD services by making relatively minor modifications. The window-based peer selection scheme described in [16–18] is an example of such modifications.

The analyzed scenario in [16–18] consists of a group of peers that have required the same video-file download in a VoD-P2P system. In order to efficiently distribute this file, the authors proposed the following procedure. The set of ordered chunks that composes the shared video file is divided into N segments of equal size. These segments are called *windows*; they are denoted by $w_0, w_1, \ldots, w_{N-1}$; and they follow the chunks' order in the stream, as it is shown in Figure 1. The basic idea is that peers must be classified in the system according to the window they are currently downloading and this information must be dynamically updated in the tracker. As a consequence, all leeches can have accurate knowledge about the specific peers in possession of potential chunks to share. Moreover, leeches that are downloading the file at window w_i can download chunks from peers in any group that are downloading the file at window w_j, for $j > i$. Conversely, the peers that are downloading the file in window w_i can serve any other peer that is downloading the file at window w_j, for $j < i$. By enabling this peer selection strategy,

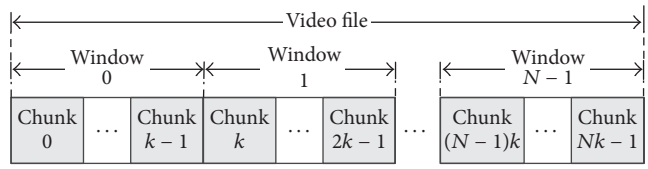


Figure 1: Video file divided into windows.

leeches only have to know the current window of the other peers, rather than the chunks that each peer possesses. This facilitates the task of identifying the peers to connect to.

For this window-based strategy, it is assumed that (a) leeches begin the file downloading process at window zero; (b) leeches at window w_i do not leave the current window until all chunks in it are downloaded; and (c) peers download any chunk in a given window with no predefined order. The information about the individual downloading progress of each peer in the system is registered at the tracker.

In order to guarantee an adequate service in terms of a QoE score, we propose a number of key modifications to the previously described window-based strategy. Specifically, unlike the works presented in [16–18], where windows were assumed to have the same size, here we propose that the size of window 0 can be different from the size of the remaining windows. The rationale behind this modification is to achieve a trade-off between the initial delay and the pauses duration.

We also propose a methodology to select the size of the windows, by developing a mathematical analysis that allows us to calculate the probability distribution of both, the initial delay and the pauses duration.

Through numerical results, we evaluate the analytical expressions derived in this work and we conclude that in order to guarantee an acceptable level of QoE, the system requires the use of additional bandwidth that has to be provided by the network manager. Indeed, relying solely on the bandwidth of peers cannot guarantee an acceptable performance under some conditions that depend on the random nature of arrivals and departures of peers and the level of cooperation of the peers. Specifically, we calculate the exact amount of additional bandwidth required to achieve conditions that guarantee that all leeches in the system are able to download the file at the maximum rate. This is another major contribution of this paper.

Evidently, the utilization of this additional bandwidth increases the implementation cost of the system; therefore, we propose a novel chunk distribution scheme, named prioritized windows distribution (PWD), where the servers provide more bandwidth to leeches in higher windows, rather than a uniform distribution as it was originally proposed. In the uniform distribution scheme, leeches in lower windows are served by seeds and other leeches in higher windows, while leeches in higher windows are only served by seeds and a small amount of leeches. For that reason, a large amount of resources from servers must be provided to leeches in higher windows. The rationale behind PWD is to counteract this disadvantage.

Lastly, we provide strict guidelines on how to select the system parameters such that QoE guarantees can be provided under different system conditions.

The rest of the paper is organized as follows: Section 2 discusses some of the previous work in the area and makes a detailed comparison with our proposal. Then, Section 3 presents in detail the window-based system, including the main assumptions and considerations. Section 4 explains the server-assisted P2P network proposal in order to attain abundance conditions in the system, including the prioritized windows distribution scheme to reduce the amount of extra server capacity. Following this, in Section 5 we derive the probability distributions of initial delay and interruption duration. Then, the implications of the window-based scheme on the QoE level are discussed in Section 6. Building on this, we provide strict guidelines to attain such QoE levels considering different system conditions. We end this paper discussing relevant numerical results and conclusions.

2. Related Work

One of the earliest works that pointed out the advantages of complementing traditional client-server networks with P2P systems in the context of video services was [19]. There, the authors demonstrate how much upload bandwidth from servers can be reduced by implementing a hybrid network. In recent years, a great researching effort from different perspectives has been made on analyzing this kind of systems in order to make them more efficient.

Some works have been focused on defining efficient P2P networks topologies. For instance, in [3] a hybrid tree-mesh topology is proposed. Other researchers have identified that a way to increase the capacity of a P2P network is by implementing strategies that efficiently distribute content (videos) among the population of peers. Examples of this kind of work are [4–6]. Schemes that incentivize cooperation among peers to increase the upload network capacity have been proposed on works such as [6–8, 20]. Finally, the main focus of some other research works has been on defining schemes that determine which peers are the most appropriate to serve a downloader (given the network topology, the video distribution, and a level of cooperation among peers) in order to improve a QoS or QoE parameter; some examples of this kind of works are [7, 9, 10, 14, 15, 20]. The scheme considered in this paper belongs to the last class, because the proposed window-based scheme finds the best peers capable to provide service to a given downloader; however, our contribution is not limited to that, since in addition we propose an efficient scheme to distribute the server resources. To our knowledge, this kind of proposal has not been published so far.

On the other hand, different analytical tools have been used to model P2P networks, including fluid models. In [21] a fluid model is proposed to analyze P2P networks and it is used to calculate performance parameters such as the number of peers in steady state in the system, as well as the average time required to download a file; however, the proposed scenario in that work does not consider the specific features of VoD services. In [16–18] the fluid model was applied to a VoD, P2P network, where the shared video is split into windows in order to simplify the chunks interchange among peers. In this paper we also analyze a windows-based P2P network; but, in order to guarantee an adequate level of service, we additionally consider that the system is assisted by servers' bandwidth and that the size of the initial window is different from the remaining ones. In recent years, some other papers have also reported analyses of VoD services over P2P networks that are based on fluid models, including [5, 11, 14, 20].

Among the works that were mentioned in the two previous paragraphs, [5, 7, 11, 20] are the most related to ours. In [7], a window-based P2P network is also described: the authors consider the existence of three buffers and a different strategy must be used to upload chunks to the network from each of them. The reason to propose such a scheme is the assumption that the peer storage capacity is limited; however, recent advances in hardware technology make low-priced devices increasingly equipped with abundant memory [20], and consequently in our system we propose the existence of only one buffer, which allows us to design a sharing mechanism that is significantly simpler than the one described in [7].

In [20], the authors present a system which achieves scheduling video sharing between peers by adopting a dynamic buffering-progress-based scheme: a downloader receives chunks only from peers with a similar playback progress. Though this proposal has some similarities with our window-based one (the chunk sharing mechanism is based on the download progress of peers), and though in both works a fluid model is used, the analysis perspective is quite different. For instance, they propose that leeches stay in the system until the downloading is finished, while we use a more realistic model in which leeches can leave the system at certain rate (denoted by θ). Additionally, in that work, as well as in ours, one of the main targets is to reduce the required server bandwidth; however, we are interested in reducing it by proposing an efficient distribution among peers (our PWD scheme), rather than incentivizing cooperation, as they suggest. Note that their cooperation scheme can also be applied to our system in order to further reduce the assisted server bandwidth.

In [11], the authors propose a modeling framework to compute the required server bandwidth, which has several similarities with our modeling work: a Poisson arrival process, analysis restricted to only one video, and homogeneous download rate. Though the model that they propose considers a wider range of scenarios than ours (nonstationary traffic, heterogeneous upload rate), our contribution is not limited to compute the required server bandwidth: we also propose the aforementioned PWD scheme and evaluate its effect on the amount of required server bandwidth; hence we consider that these works can be complementary.

Additionally, in [11] the server bandwidth is computed considering that no interruptions occur during the playback process; however, a more significant reduction of those resources could be achieved by allowing the occurrence of initial delays and interruptions, provided that their durations and occurrence probabilities do not degrade significantly the users' experience (measure through QoE parameters); the integration of such a consideration, as well as its analysis, is

another contribution of this paper, since none of the above described works has addressed it and, to our knowledge, little research has been done about this kind of issue. In order to consider the effect of QoE parameters, we use the experimental results reported in [22], where relations between QoS parameters (initial delays and interruptions duration) and QoE parameters (Medium Opinion Score, widely known as MOS) are established.

3. The Model for the Window-Based System

In this section we present the mathematical model for the *conventional* window-based strategy. As we mentioned in Section 1, this strategy was proposed for chunk distribution among peers in a P2P-VoD system. We refer to this system as *conventional* since it relies only on the bandwidth shared by the peers in the network as proposed in [16, 21]. As it is shown below, the QoE guarantees in this system can be achieved only in very particular conditions, outside the capabilities of the network manager, since it can only rely on the cooperation among users. This section also presents the main assumptions and parameters of the system.

In order to facilitate the reading of Sections 3–6, we summarize in Most Relevant Variables Summary section the most relevant variables that are used through the whole paper.

The number of leeches that are downloading window w_i at time t is denoted by $x_i(t)$. The total number of leeches in the system at time t is denoted by $x(t)$; that is,

$$x(t) = \sum_{i=1}^{N-1} x_i(t). \tag{1}$$

There is a single file in the system, assumed for simplicity to be of size 1 as in [21]. The number of seeds in the system at time t is denoted by $y(t)$. Seeds share all chunks with leeches, independently of their current window. Additionally, according to the window-based strategy, a leech can send all its chunks to any other leech currently downloading data in previous windows. We also assume that at any given time there is at least one seed in the system, in order to prevent the starving of the system. We also assume that new leeches arrive at window 0 with constant rate λ. Observing that we are not considering Poisson processes here, the model is a deterministic one. The transition rate from any peer at window i to the next one is denoted by τ_i and the rate at which a leech leaves the system before the completion of the download or the playback process is denoted by θ. A seed leaves the system with rate γ. We assume that all peers have the same physical characteristics. Specifically, they all have the same uploading bandwidth μ (in files/sec). We denote by $1/c$ the (mean) time needed to download the whole file without interruptions working at full capacity; hence, c is the maximal file download rate for any peer, in files/sec, where $c \geq \mu$. The mean time needed to download a window is thus $(1/c)/N$; we denote by $c_w = Nc$ the corresponding window download rate. In the same way, $\mu_w = N\mu$. Recall that the file size is 1.

In this analysis, we consider that peers have complete knowledge of the state of system, namely, that peers know the current group of all the other peers. This is a reasonable assumption, since, in a BitTorrent-like system, peers can obtain this information from the tracker server. Under this assumption, if the number of leeches and seeds is sufficiently high, all leeches download a window at the maximum rate c_w. This condition is referred to as *abundance*. However, when there are not enough peers in the system, the leeches download a window at a rate that is smaller than c_w. This condition is referred to as *penury*.

Finally, it is important to note that, in order to simplify the analysis, it is considered that users always play the video file in order, that is, once some user starts the video playback, the latter is not fast-forwarded. This is due to the fact that all users consider that leeches in window w_i have all previous chunks (from window w_0 to window w_{i-1}). If some user fast-forwards the video, it may not continue the download of the chunks corresponding to the parts of the video that was not played. Recall that since a managed network is considered, this particular simplification can be easily implemented in a commercial system. However, in a future work, the case where users can forward and rewind the video playback will be considered.

From the previous description, the evolution in time of the number of leeches in each window, $x_i(t)$, and of the number of seeds, $y(t)$, for the system satisfies the following equations:

$$\begin{aligned} x_0'(t) &= \lambda - \theta x_0(t) - \tau_0, \\ x_i'(t) &= \tau_{i-1} - \theta x_i(t) - \tau_i, \quad 1 \leq i \leq N-1, \\ y'(t) &= \tau_{N-1} - \gamma y(t), \end{aligned} \tag{2}$$

where

$$\tau_i = \min\left\{ c_w x_i(t), \mu_w x_i(t) \cdot \left(\sum_{k=i+1}^{N-1} \frac{x_k(t)}{\sum_{j=0}^{k-1} x_j(t)} + \frac{y(t)}{x(t)} \right) \right\}, \tag{3}$$

for $i = 0, 1, \ldots, N-1$.

In the case of the last window, that is, when $i = N-1$, we have

$$\tau_{N-1} = \min\left\{ c_w x_{N-1}(t), \mu_w x_{N-1}(t) \frac{y(t)}{x(t)} \right\}. \tag{4}$$

These last equations, which are related to the transition of peers from window i to window $i+1$, can be explained as follows: Note that, in case of abundance, leeches in window w_i download the file at the maximum bandwidth c_w. However, when there are not enough peers in the system, the leeches download a window at a rate which is described as follows. First of all, note that every peer can upload a window at rate μ_w. Secondly, seeds upload the file to all peers in the system. As such, all the upload bandwidth is distributed uniformly among all leeches. Therefore, the proportion of the upload bandwidth for leeches at window w_i is $x_i(t)/x(t)$. Finally, only the leeches in a posterior window k (from $i+1$ to $N-1$) can

send chunks to leeches in window w_i. This upload bandwidth is distributed uniformly for all leeches in windows 0 to $k-1$. As such, the proportion of the upload bandwidth for leeches in window w_i is $x_i(t)/\sum_{j=0}^{k-1} x_j(t)$. Leeches in the last window only receive chunks from seeds.

Let us compute the equilibrium point of this dynamical system, which we denote by $(\overline{x}_0, \overline{x}_1, \ldots, \overline{y})$. This equilibrium point can be obtained by solving the system $\{x_0'(t) = 0, x_1'(t) = 0, \ldots, y'(t) = 0\}$ and by replacing $x_i(t)$ by $\overline{x}_i$ and $y(t)$ by $\overline{y}$ in (2); hence, we obtain a system of equations given by

$$\begin{aligned} \lambda - \theta\overline{x}_0 - \tau_0 &= 0, \\ \tau_{i-1} - \theta\overline{x}_i - \tau_i &= 0, \quad 1 \le i \le N-1, \\ \tau_{N-1} - \gamma\overline{y} &= 0. \end{aligned} \tag{5}$$

Though this system can be solved in abundance or in penury (see (3) and (4)), we focus our analysis in the former one, because a target QoE score can be guaranteed if this abundance condition exists, as we show later in this paper. Considering this, we substitute $\tau_i = c_w\overline{x}_i$ in (5) and by solving the system we obtain

$$\begin{aligned} \overline{x}_i &= \frac{\lambda c_w^i}{(\theta + c_w)^{i+1}}, \\ \overline{y} &= \frac{\lambda}{\gamma}\left(\frac{c_w}{\theta + c_w}\right)^N, \end{aligned} \tag{6}$$

for $0 \le i \le N-1$.

The total number of leeches in equilibrium, $\overline{x}$, is given in the next equation:

$$\overline{x} = \sum_{i=0}^{N-1} \overline{x}_i = \frac{\lambda}{\theta}\left[1 - \left(\frac{c_w}{\theta + c_w}\right)^N\right]. \tag{7}$$

All numerical explorations showed convergence towards equilibrium (see the numerical results that are shown in Section 8). The analysis of the stability of this nonlinear system of differential equations is left for future work.

In the following, the conditions to achieve abundance in the system are identified. This is an important feature for the practical implementation of the window-based strategy. In particular, the seeds departure rate (γ) is an important variable, considering that seeds already have all the chunks of the video file. From (3), it can be observed that in order to have abundance at window w_i, the following condition must be met:

$$c_w x_i(t) \le \mu_w\left(\sum_{k=i+1}^{N-1} x_k(t)\frac{x_i(t)}{\sum_{j=0}^{k-1} x_j(t)} + y(t)\frac{x_i(t)}{x(t)}\right). \tag{8}$$

If we consider $t = \infty$ in (8), we can substitute $x_i(t)$ and $y(t)$ by the definitions of $\overline{x}_i$ and $\overline{y}$ that were given in (6); in other words, we are analyzing the system at the equilibrium point. Hence, the condition that must be met in order to have abundance at window w_i can also be expressed as

$$\begin{aligned} c_w \le \mu_w\Bigg(&\sum_{k=i+1}^{N-1} \frac{c_w^k/(\theta + c_w)^{k+1}}{\sum_{j=0}^{k-1}\left(c_w^j/(\theta + c_w)^{j+1}\right)} \\ &+ \frac{\theta\left(c_w/(\theta + c_w)\right)^N}{\gamma\left[1 - \left(c_w/(\theta + c_w)\right)^N\right]}\Bigg). \end{aligned} \tag{9}$$

Notice that in (9) the abundance condition is given as a function of the system parameters; consequently, by simplifying this expression and by clearing γ, we can find a value γ_i that represents the maximum seeds departure rate that guarantees abundance in window i (i.e., $\gamma \le \gamma_i$, for all i).

$$\begin{aligned} &\gamma_i \\ &= \frac{(\theta + c_w)\, c_w^N/\left((\theta + c_w)^N - c_w^N\right)}{c_w(\theta + c_w)/\theta\mu_w - \sum_{k=i+1}^{N-1}\left(c_w^k/\left((\theta + c_w)^k - c_w^k\right)\right)}. \end{aligned} \tag{10}$$

Rate γ is clearly a *manager* variable that could be controlled according to (10). In Figure 2, the abundance condition for different values of the departure rate for a leech (θ) is presented. It can be seen that as departure rate for a leech increases, it is necessary to increment the time that seeds remain in the system; that is, the seeds departure rate should be also reduced, since there are fewer leeches sharing the file. Hence, by using (10), and by measuring the arrival and departure rates for a leech, the network manager has a tool to offer an acceptable performance of the system by allowing all leeches to download windows at their maximum capacity. In other words, our analysis allows finding an appropriate value for this control variable. However, achieving the adequate value of γ, that is, encouraging peers to remain sufficient time in the system after the complete file download has occurred, is not an easy task; since when the peer has already played or downloaded the file, it has no need to remain longer just to share the file. One possibility of coping with this problem is to introduce penalties or rewards to peers in order to encourage a cooperative behavior (as in [6–8, 16, 18, 20]). However, these mechanisms do not guarantee to achieve an acceptable QoE in the network since they still relay on the behavior of the users. In a managed system, the network manager can decide for users equipment to remain connected sharing chunks to other users. Nonetheless, users are able to shut down or disconnect their equipment. This option is outside the capabilities of the network manager to provide QoE guarantees.

From (10) it can also be observed that abundance condition could be achieved by controlling N, the number of windows. Figure 3 shows that, by reducing N, the required value of γ_{N-1} to guarantee abundance is slightly relaxed (increased); however, reducing N deteriorates the QoE parameters, as it is widely discussed in Sections 5–7 of this paper.

The rest of the variables that are involved in (10) are much harder to control, since the download and upload bandwidths (c_w and μ_w, resp.) are usually fixed by the hardware used and

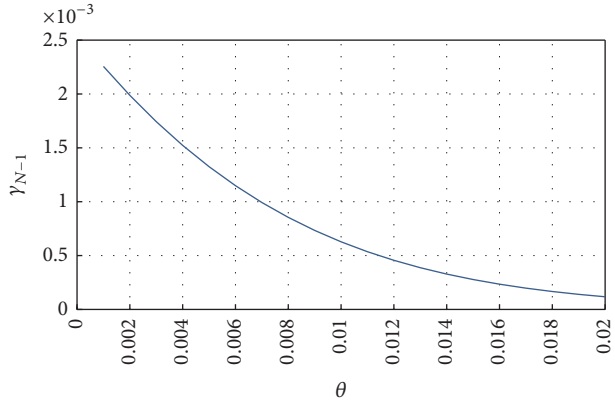


Figure 2: Relationship between minimal departure rate for a seed and departure rate for leeches in order to keep the system in abundance conditions; $\lambda = 0.04$, $c = 0.00407$, $\mu = 0.00255$, and $N = 48$.

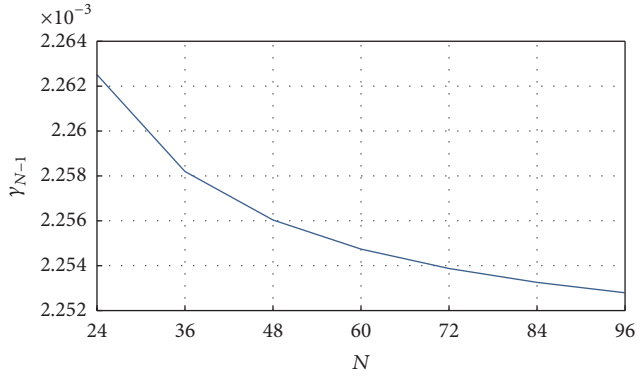


Figure 3: Relationship between minimal departure rate for a seed and number of windows in order to keep the system in abundance conditions; $\lambda = 0.04$, $c = 0.00407$, $\mu = 0.00255$, and $\theta = 0.001$.

the arrival rate (λ) depends on the file popularity, that is, how many users are willing to play the video.

Considering the previous paragraphs, we propose three mechanisms to improve the system's performance by guaranteeing the satisfaction of QoE parameters. First, we propose to use a different size for the first window in order to achieve a trade-off between the initial playback delay and the pauses duration. Second, we propose the use of additional server bandwidth that the network manager can explicitly provide to satisfy QoE parameters. Third, we propose an efficient distribution scheme of these extra servers resources, called *prioritized windows distribution*, which is based on the window of the leech that is downloading the file. These mechanisms can be fully controlled by the network manager and are explained in detail in the following sections.

4. Server-Assisted System and Reduced Initial Window

According to [22] the QoE that a user experiences during an online video playback is highly determined by the initial delay, by probability that a pause occurs during the playback, and by the duration of such pauses. In this regard, it is important to note that there is a trade-off between the values that must be assigned to these parameters, since reducing the initial delay increases the probability that the video playback pauses and vice versa. The rationale behind this is as follows: when the device remains downloading the video file for a long period of time before the playback begins, a larger portion of that file will be available for its future playback; hence, the probability that a pause occurs is reduced.

Among these parameters, it has been identified that the pause probability and the duration of these pauses are much more harmful to the user's perception than the initial delay. In general, users prefer to wait longer times at the beginning of the file playback, if this means that the video will not be paused at all. However, this initial delay cannot be arbitrarily long, since it would negatively impact the QoE of users. In the window-based strategy, the initial delay is directly related to the size of the windows. Note that it is not feasible to reduce the size of all the windows in the system since the bandwidth required to attain abundance conditions is much harder with a large number of windows, as explained later in this section. However, the size of the first window can be modified in order to guarantee an acceptable initial delay, while the rest of the windows can be set to an adequate value.

Considering the previously described trade-off, we propose the following modifications to the conventional window-based system in order to provide an acceptable level of QoE:

(i) Unlike the conventional model described in the previous section, here we assume that the size of the first window can be different from the size of the remaining windows. This is important because these window sizes define a trade-off between initial delay and average pause duration.

(ii) In order to prevent unacceptable levels of initial delay or average pause duration, we propose to enhance the P2P network with a fixed download bandwidth provided by servers.

(iii) Lastly, a nonuniform chunk distribution scheme is proposed in order to avoid the resource starving problem experienced by leeches in upper windows. This scheme effectively reduces the required server bandwidth to reach abundance conditions.

Taking into account the previous considerations, we denote by μ_s the bandwidth provided by servers, and as an initial analysis we assume that this extra bandwidth is uniformly distributed among all the leeches, that is, in the same way that the bandwidth provided by peers is distributed (as previously mentioned, in Section 4.2 we propose a more efficient bandwidth distribution scheme); consequently, the upload bandwidth that servers can provide to leeches in window i is $x_i\mu_s/x$.

On the other hand, we denote by d_0 and d_1 the average time to download window 0 and window i (for $1 \leq i \leq N-1$), respectively. As a consequence, the average download rates for window 0 and window i are, respectively, $c_0 = 1/d_0$ and $c_1 = 1/d_1$. In order to simplify the subsequent equations, we

define the parameter α, which represents the ratio of the size of window i ($i > 0$) to the size of window 0; hence $\alpha = d_1/d_0$. Since the average download time for the whole file must be equal to the sum of the average download times of all the windows, that is, $1/c = d_0 + (N+1)d_1$, we can express c_0 and c_1 as follows:

$$c_0 = c\,(\alpha N + 1 - \alpha)\,,$$
$$c_1 = c\left(N + \frac{1-\alpha}{\alpha}\right). \tag{11}$$

Following an analogous analysis to the previous paragraph, the upload rates for windows 0 and i are, respectively, given by the following expressions:

$$\mu_0 = \mu\,(\alpha N + 1 - \alpha)\,,$$
$$\mu_1 = \mu\left(N + \frac{1-\alpha}{\alpha}\right). \tag{12}$$

The incorporation of parameters μ_s, c_0, c_1, μ_0, and μ_1 in our analysis does not modify the essence of the fluid model described in Section 3. However, the definition of τ_i must be modified since in this case the rates to download or upload a window are given by (11)-(12) and, additionally, the upload capacity of the system must be increased by the server bandwidth. As a result of these new considerations, we have the following expressions (where time t is omitted for the sake of room):

$$\tau_0 = \min\left\{c_0 x_0, x_0\left[\mu_0\left(\sum_{k=1}^{N-1}\frac{x_k}{\sum_{j=0}^{k-1}x_j} + \frac{y}{x}\right) + \frac{\mu_s}{x}\right]\right\},$$
$$\tau_i = \min\left\{c_1 x_i, x_i\left[\mu_1\left(\sum_{k=i+1}^{N-1}\frac{x_k}{\sum_{j=0}^{k-1}x_j} + \frac{y}{x}\right) + \frac{\mu_s}{x}\right]\right\}, \tag{13}$$

for $1 \le i \le N-2$, and

$$\tau_{N-1} = \min\left\{c_w x_{N-1}, x_{N-1}\frac{\mu_1 y + \mu_s}{x}\right\}. \tag{14}$$

Again, working as in Section 3, the new equilibrium point of the system in abundance is now given by

$$\overline{x}_0 = \frac{\lambda}{\theta + c_0},$$
$$\overline{x}_i = \frac{\lambda c_0}{(\theta + c_0)}\frac{c_1^{i-1}}{(\theta + c_1)^i}, \tag{15}$$

for $1 \le i \le N-1$, and

$$\overline{y} = \frac{\lambda}{\gamma}\frac{\alpha}{\beta}\left(\frac{c_1}{\theta + c_1}\right)^N, \tag{16}$$

where $\beta = (\theta + c_0)/(\theta + c_1)$. The total number of leeches in equilibrium, $\overline{x}$, is given by the following expression:

$$\overline{x} = \frac{\lambda}{\theta}\left(1 - \frac{\alpha}{\beta}\left(\frac{c_1}{\theta + c_1}\right)^N\right). \tag{17}$$

It is important to note that even if the parameters related to the video-file uploading (μ_0, μ_1, and μ_s) are not explicit in previous equations, their values are fundamental to guarantee the abundance condition of the system, as it will be shown in the next subsection. Finally, it must be said that (13)–(17) are reduced to their corresponding counterparts given by (3)–(7), when $\alpha = 1$ (i.e., the sizes of all windows are equal).

4.1. Minimum Server Bandwidth Requirement to Guarantee Abundance Conditions. According to (13)–(17), the abundance condition can be expressed by

$$c_0 \le \mu_0\left(\sum_{k=1}^{N-1}\frac{x_k}{\sum_{j=0}^{k-1}x_j} + \frac{y}{x}\right) + \frac{\mu_s}{x} \tag{18}$$

for window 0, by

$$c_1 \le \mu_1\left(\sum_{k=i+1}^{N-1}\frac{x_k}{\sum_{j=0}^{k-1}x_j} + \frac{y}{x}\right) + \frac{\mu_s}{x} \tag{19}$$

for windows $1 \le i \le N-2$, and by

$$c_1 \le \mu_1\frac{y}{x} + \frac{\mu_s}{x} \tag{20}$$

for window $N-1$.

From (18) to (20), the abundance conditions in terms of γ can be found as shown in Section 3. However, since γ is a parameter that highly depends on users' behavior, a complementary way to guarantee abundance in the system can be based on μ_s. As it was previously mentioned, we propose different schemes in order to guarantee abundance with a minimum value of μ_s (nonuniform chunk distribution and different size of the initial window). In this regard, it is important to notice that, as we show in our numerical evaluations, small values of γ significantly help to satisfy abundance conditions. As such, the use of incentives to encourage peers to remain longer times in the system is still an important issue to provide QoE guarantees.

If we define $\mu_{\min}^i$ as the minimum bandwidth that is required from servers in order to preserve the abundance condition in window i, we can say that such conditions are guaranteed in the whole system if $\mu_s \ge \mu_{\min}^i$, for every i. From (18) to (20) we obtain

$$\mu_{\min}^0 = \max\left\{0, x\left[c_0 - \mu_0\left(\sum_{k=1}^{N-1}\frac{x_k}{\sum_{j=0}^{k-1}x_j} + \frac{y}{x}\right)\right]\right\},$$
$$\mu_{\min}^i = \max\left\{0, x\left[c_1 - \mu_1\left(\sum_{k=i+1}^{N-1}\frac{x_k}{\sum_{j=0}^{k-1}x_j} + \frac{y}{x}\right)\right]\right\} \tag{21}$$

for $1 \leq i \leq N-2$, and

$$\mu_{\min}^{N-1} = \max\{0, c_1 x - \mu_1 y\}. \quad (22)$$

In order to understand the intuition behind the max operation in (21)-(22), it must be observed that the network can reach abundance without the need of servers and in that case the second term of this operation will be a negative number. Building from this, we can write

$$\mu_{\min}^{i-1} = \max\left\{0, \mu_{\min}^{i} - \frac{\mu_1 x x_i}{\sum_{j=0}^{i-1} x_j}\right\} \quad (23)$$

for $2 \leq i \leq N-1$.

Since x, x_i, and μ_1 are always nonnegative values, it is clear that $\mu_{\min}^{i} \geq \mu_{\min}^{i-1}$ for $2 \leq i \leq N-1$. This inequality implies that if the bandwidth provided by servers is enough to reach abundance in window $N-1$, then abundance is guaranteed in the remaining lower windows; that is, the bandwidth required from servers to guarantee abundance in the whole system is $\mu_{\min} = \mu_{\min}^{N-1}$.

After substituting (16) and (17) in (22) we have that

$$\mu_{\min} = \max\left\{0, \frac{\lambda c_1}{\theta}\left[1 - \frac{\alpha}{\beta}\left(1 + \frac{\theta\mu_1}{\gamma c_1}\right)\left(\frac{c_1}{\theta + c_1}\right)^N\right]\right\}. \quad (24)$$

It is important to remark that if $\alpha = 1$, then the model of this section is reduced to the conventional window-based model; consequently, x_i and y are simply given by (6).

From this model, we have identified an important issue that directly impacts the performance of the system. Specifically, we have noted that the uniform distribution of resources produces a resource starvation for leeches in upper windows. Indeed, leeches at lower windows are served by leeches in upper windows, seeds, and the extra bandwidth that is provided by servers. However, leeches in upper windows are served by a much lower amount of leeches. Additionally, there is a higher amount of leeches in lower windows than in upper windows, which, according to (23), produces that most resources are being consumed by leeches in lower windows. As such, the amount of resources assigned to peers in the upper windows is drastically reduced. The previous explanation is supported by Figure 4, where the amount of upload bandwidth that a leech in window i can get (UB_i) is shown. It can be noticed that even with large amounts of server bandwidth (e.g., $\mu_s = 3$) the peers in the upper windows can access only a small amount of resources. In order to have a more efficient distribution we propose a novel chunk sharing scheme detailed in the next subsection.

4.2. Server Bandwidth Distribution Scheme with Prioritized Windows. In the scheme that was described in the previous subsection, the server bandwidth is uniformly distributed among all the leeches, which makes its implementation significantly simple. However, as mentioned above, a lot of extra bandwidth is required to provide abundance to leeches in the last window, since they have too few options to download their required chunks.

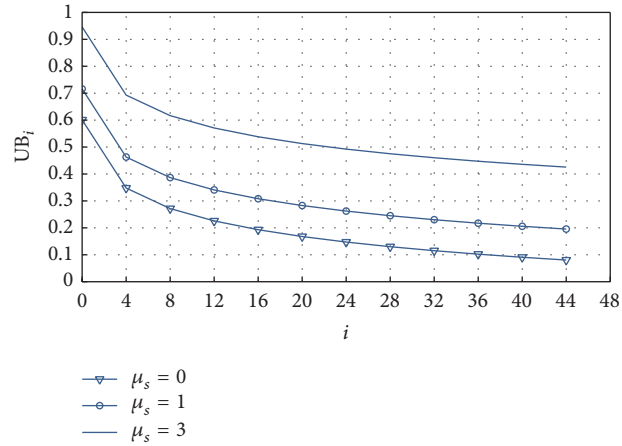


Figure 4: Upload bandwidth per peer in window i, for $\lambda = 0.04$, $c = 0.00407$, $\mu = 0.00255$, $\theta = 0.001$, $\gamma = 0.006$, $\alpha = 1$, and $N = 48$.

Consequently, we propose that the amount of server bandwidth that is assigned to the leeches in window i must be proportional not only to the numbers of leeches in that window, but also to an additional weight that must give priority to leeches in high windows. Specifically, the server bandwidth assigned to leeches in window i will now be proportional to $x_i(i+1)^{\varepsilon}$, since the factor $(i+1)^{\varepsilon}$ will prioritize upper windows over the lower ones, for $\varepsilon > 0$. In the rest of the paper, this strategy is referred to as *prioritized windows distribution* (PWD) scheme, while the one described in Section 4.1 is referred to as *uniform distribution* (UD) scheme.

According to the previous description, the transition rate from window i to window $i+1$ now is given by

$$\begin{aligned}
\tau_0 &= \min\left\{c_0 x_0, x_0\left[\mu_0\left(\sum_{k=1}^{N-1}\frac{x_k}{\sum_{j=0}^{k-1} x_j} + \frac{y}{x}\right) + \frac{\mu_s}{x_\varepsilon}\right]\right\}, \\
\tau_i &= \min\left\{c_1 x_i, x_i\left[\mu_1\left(\sum_{k=i+1}^{N-1}\frac{x_k}{\sum_{j=0}^{k-1} x_j} + \frac{y}{x}\right) + \frac{(i+1)^{\varepsilon}\mu_s}{x_\varepsilon}\right]\right\}
\end{aligned} \quad (25)$$

for $1 \leq i \leq N-2$, and

$$\tau_{N-1} = \min\left\{c_w x_{N-1}, x_{N-1}\left[\frac{\mu_1 y}{x} + \frac{N^{\varepsilon}\mu_s}{x_\varepsilon}\right]\right\}, \quad (26)$$

where $x_\varepsilon = \sum_{j=0}^{N-1}(j+1)^{\varepsilon} x_j$ is a normalization used to guarantee that the sum of the server bandwidth assigned to all the windows has to be equal to μ_s.

From (25) to (26), it can be observed that the priority of window i over window $i-1$ (for $1 \leq i \leq N-1$) is more accentuated, if ε increases. If $\varepsilon = 0$, these equations are reduced to (13)-(14). On the other hand, if $\varepsilon < 0$, the prioritized windows are the lower ones, which entails a system that assigns more resources to peers in low windows and less resources to peers in high windows, accentuating the resource starvation of these upper windows peers. Hence we are interested in analyzing the system only for $\varepsilon > 0$.

It is important to remark that these modifications in the server bandwidth distribution do not alter the essence of the previously described model but only modify the uplink capacity of the system, in the way that was already considered in (25)-(26).

The abundance conditions in terms of the server bandwidth now are given by

$$\mu_0 = \max\left\{0, x_\varepsilon\left[c_0 - \mu_0\left(\sum_{k=1}^{N-1}\frac{x_k}{\sum_{j=0}^{k-1}x_j} + \frac{y}{x}\right)\right]\right\},$$

$$\mu_i = \max\left\{0, \frac{x_\varepsilon}{(i+1)^\varepsilon}\left[c_1 - \mu_1\left(\sum_{k=i+1}^{N-1}\frac{x_k}{\sum_{j=0}^{k-1}x_j} + \frac{y}{x}\right)\right]\right\} \quad (27)$$

for $1 \leq i \leq N-2$, and

$$\mu_{N-1} = \max\left\{0, \frac{x_\varepsilon}{N^\varepsilon}\left[c_1 - \mu_1\frac{y}{x}\right]\right\}. \quad (28)$$

It is important to emphasize that if ε is too large, the leeches in the higher windows would have access to an excessive amount of server bandwidth, while some other windows would become too prone to penury and consequently a lot of bandwidth servers must be installed to guarantee abundance in those windows. Hence, ε must be selected in such a way that, given the parameters of the system, the abundance condition is guaranteed, while the amount of server bandwidth is maintained at the lowest possible value. In order to clarify the problem described above, in Figure 5 we show $\mu^i_{\min}$ for different values of ε. It can be seen that with $\varepsilon = 3$ the minimum assisted server bandwidth to guarantee abundance is $\mu_{\min} = 1.5$ and it is no longer for the last window, $i = N$. On the other hand, with $\varepsilon = 1.5$, the value of $\mu_{\min}$ is now less than 1.

Unlike the UD scheme analysis, in the PWD case it is not easy to determine which window requires the largest amount of server bandwidth and it is not straightforward to find a closed expression for the optimal value of ε (ε_{opt}). However, we can find this optimal value by numerically evaluating the following expressions:

$$i_{\text{crit}}(\varepsilon) = \underset{i}{\text{argmax}}\left\{\mu_i(\varepsilon)\right\}, \quad (29)$$

$$\varepsilon_{\text{opt}} = \underset{\varepsilon}{\text{argmin}}\left\{\mu_{i_{\text{crit}(\varepsilon)}}\right\}. \quad (30)$$

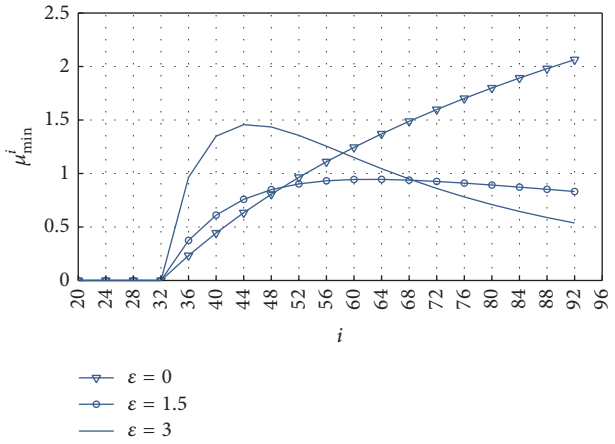


FIGURE 5: Minimum server bandwidth required by window i for $\lambda = 0.04$, $c = 0.00407$, $\mu = 0.00255$, $\theta = 0.001$, $\gamma = 0.006$, and $N = 96$.

Here, $i_{\text{crit}}(\varepsilon)$ represents the index of the window that requires the largest value of $\mu_i^{\min}$ for a given ε. Strictly, ε_{opt} depends on all the parameters of the system; however, after evaluating (29) and (30), we found that it only significantly depends on θ and γ, as shown in Figure 6. Considering this, we propose to calculate ε_{opt} through the approximation given by

$$\varepsilon_{\text{opt}} \approx 4916\theta^2 - 139\theta + 2279\theta\gamma - 49\gamma + 971\gamma^2 + 1.65. \quad (31)$$

The polynomial coefficients in (31) were obtained by applying a second-order linear regression to the results shown in Figure 6. In Figure 7 we show a comparison between the exact evaluation and the corresponding approximation. Lastly, it must be noticed that even if ε is chosen exclusively in terms of θ and γ, $\mu_{\min}$ is still a function of the remaining parameters of the system (e.g., N and α). In Section 8, we show that using the evaluation of ε_{opt} that is defined by (31), the PWD scheme significantly reduces $\mu_{\min}$ in comparison with the UD scheme.

5. Probability Distributions of Initial Delay and Interruption Duration

So far, we have described the operation of the proposed network; we have developed a mathematical model to evaluate the number of peers under abundance conditions and proposed two schemes to distribute additional bandwidth provided by servers. However, one of our major concerns in this paper is to establish a method to select the parameters of the system that satisfy some Quality of Experience (QoE) targets. To this end, we first analyze the behavior of the Quality of Service (QoS) parameters that are related to such QoE targets. For this purpose, we change the viewpoint and interpret now the system probabilistically. Specifically, in this section we model the probability distributions of the initial delay and the interruption duration along the playback of a video which is being downloaded from the described system.

Since the mathematical analysis in the previous sections is valid only if the abundance condition exists, the analysis in this section is also limited to such circumstances; that is,

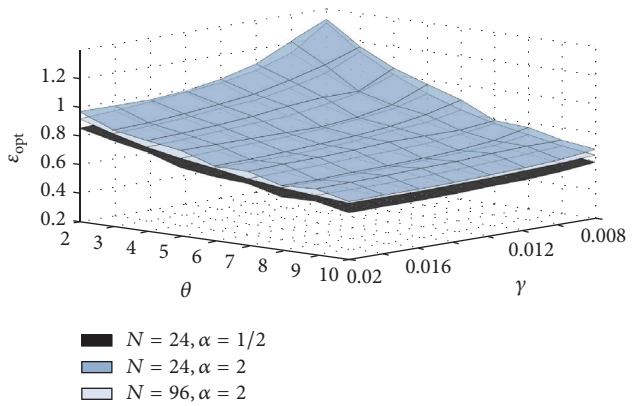


Figure 6: Optimal values of ε for $\lambda = 0.04$, $c = 0.00407$, and $\mu = 0.00255$.

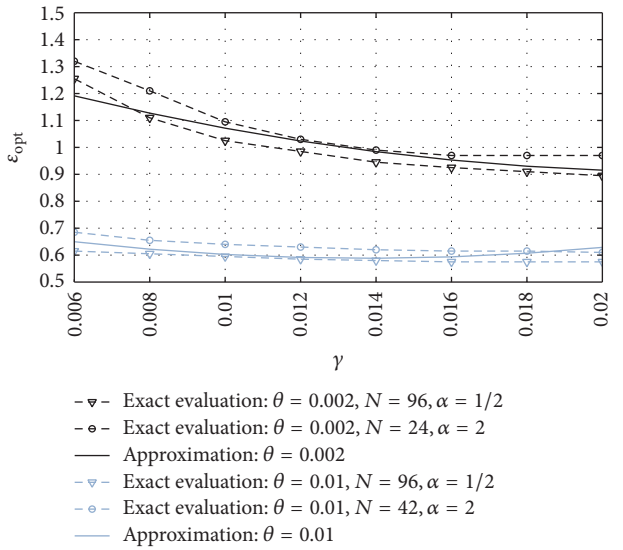


Figure 7: Optimal values of ε: exact versus approximated evaluations ($\lambda = 0.04$, $c = 0.00407$, and $\mu = 0.00255$).

the probability distributions that we find are valid only if the system is in abundance. Additionally, in our analysis we are assuming that the playback of any window starts until that window has been completely downloaded, as it is shown in Figure 8 (e.g., the video playback starts until window 0 has been completely downloaded). This assumption is valid because some buffering is needed to satisfy QoE targets that are related to interruptions. In Figure 8 we also illustrate the meaning of some of the variables that are defined along this section.

5.1. Probability Distribution of Initial Delay and Mean Downloading Time. We assume here that the distributions of the sojourn time of leeches and the time to download window 0 are both negatively exponential, that is, that their densities are given by $f_{\mathbf{u}}(x) = \theta e^{-\theta x} 1 \ (x \geq 0)$ and $f_{\mathbf{v}_0}(x) = c_0 e^{-c_0 x} 1 \ (x \geq 0)$, respectively. This is the natural assumption to keep coherence with the constant rates assumed in the fluid model, which in turn are coherent with the idea of considering the system operating in steady state, far from the initial conditions.

It is important to notice that, according to our model, $\mathbf{u}$ and $\mathbf{v}_0$ are independent random variables and while the former models the users' sojourn time in the system, the latter only models the time to download window 0, with no regard of the user's sojourn. According to that, we define a new random variable, $\mathbf{w}_0$, which represents the time to download window 0, given that the sojourn time was large enough to achieve such download. Hence, $\mathbf{w}_0$ is equal to $\mathbf{v}_0$, given that $\mathbf{v}_0 \leq \mathbf{u}$. Unconditioning, we have that the density of $\mathbf{w}_0$ is

$$f_{\mathbf{w}_0}(x) = (c_0 + \theta) e^{-(c_0+\theta)x} 1 \quad (x \geq 0). \tag{32}$$

Under the assumption that the playback of window 0 will start until it has been successfully downloaded, we can say that (32) also represents the initial delay probability distribution. Furthermore, the mean initial delay is given by

$$T_0 = \frac{1}{c_0 + \theta}. \tag{33}$$

Note that $T_0 < 1/c_0$ because T_0 is an average download time that does not include the cases in which the downloading time is larger than the sojourn time of a leech.

In order to find the distribution of the required time to download window i (for $1 \leq i \leq N-1$), it is necessary to identify the distribution of the remaining sojourn time of a leech that has downloaded the preceding windows. Let $\mathbf{r}_0$ be the remaining sojourn time of a leech that has downloaded window 0. Hence, $\mathbf{r}_0 = \mathbf{u} - \mathbf{v}_0$, given that $\mathbf{u} - \mathbf{v}_0 > 0$. Considering the distributions of $\mathbf{u}$ and $\mathbf{v}_0$, their common density is given next:

$$f_{\mathbf{r}_0}(x) = \theta e^{-\theta x} 1 \quad (x \geq 0). \tag{34}$$

Although this result could seem a contradiction, it is explained by the fact that now we only focus on those leeches whose sojourn time is large enough to successfully download window 0 and by the memoryless property of the negative exponential distribution.

Given the previous result, it is clear that the distribution of the required time to download window 1, given that window 0 was downloaded, can be found by substituting c_0 by c_1 in (32). Furthermore, after applying the previous analysis to every window, the distribution of the required time to successfully download window i, given that all the preceding windows were downloaded (denoted by $\mathbf{w}_i$, for $1 \leq i \leq N-1$), is

$$f_{\mathbf{w}_i}(x) = (c_1 + \theta) e^{-(c_1+\theta)x} 1 \quad (x \geq 0). \tag{35}$$

And the mean time to download window i, given that the leech has not left the system at this point, is simply $T_i = 1/(c_1 + \theta)$. Consequently the mean time to download the whole file, given that the leech did not leave the system is

$$T = \frac{1}{c_0 + \theta} + \frac{N-1}{c_1 + \theta}. \tag{36}$$

It is interesting to note that if $\alpha = 1$, the required time to download the whole file has an Erlang distribution with rate

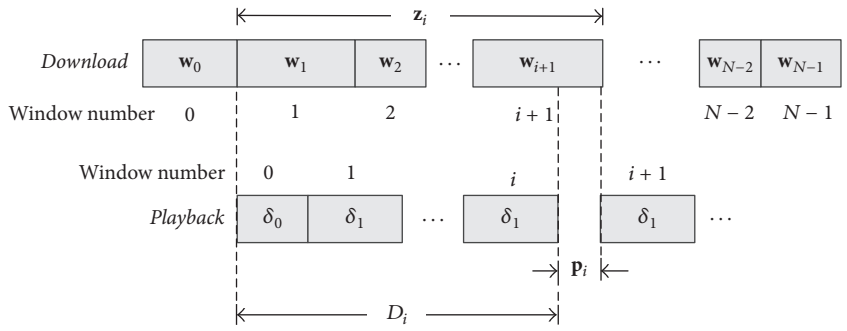


Figure 8: Relationship between download and playback processes.

parameter equal to $c_w + \theta$ and shape parameter equal to N, and (36) is reduced to $T = 1/(c + \theta/N)$. In this expression it is clear that by increasing the number of windows, the mean delay T also increases but only to a maximum value of $1/c$.

5.2. Probability Distribution of Interruption Duration. On the basis of some of the previous results, in this subsection we analyze the probability distribution of the interruption duration that may occur after the playback of window i.

As it was previously established, we are considering that the playback of a window initiates until all the chunks of the window have been completely downloaded. Building from this, an interruption can occur after the playback of window i, if the time required to play back all the windows from 0 to i is smaller than the time required to download windows 1 to $i + 1$.

Due to the different initial window size scheme proposed to guarantee the QoE, we can represent by δ_0 the playback time of window 0 and by δ_1 the playback time of any other window (notice that these variables are proportional to d_0 and d_1, resp.). Hence, the playback time of windows 0 to i can be expressed by $D_i = \delta_0 + i\delta_1$. Additionally, let $\mathbf{z}_i$ be the time to download windows 1 to $i + 1$. Hence $\mathbf{z}_i = \sum_1^{i+1} \mathbf{w}_i$ (see Figure 8). Considering (35), we know that $\mathbf{z}_i$ will follow an Erlang distribution with rate parameter equal to $c_1 + \theta$ and shape parameter equal to $i + 1$.

Now, let $\mathbf{p}_i$ be a random variable that represents the interruption duration after playing window i. From Figure 8, $\mathbf{p}_i$ can be expressed as follows:

$$\mathbf{p}_i = \begin{cases} 0 & \text{if } \mathbf{z}_i \leq D_i, \\ \mathbf{z}_i - D_i & \text{if } \mathbf{z}_i > D_i. \end{cases} \tag{37}$$

Consequently, its probability density function must be given by $f_{\mathbf{p}_i}(x) = F_{\mathbf{z}_i}(D_i)\delta(x) + f_{\mathbf{z}_i}(x + D_i)1$ $(x \geq 0)$, where $F_{\mathbf{z}_i}(x)$ is the cumulative distribution function of $\mathbf{z}_i$ and $\delta(x)$ is Dirac's delta function. Then, the probability density function of the interruption duration after playing window i is given by

$$f_{\mathbf{p}_i}(x) = \left[1 - \sum_{j=0}^{i} \frac{e^{-(c_1+\theta)D_i}\left(\left(c_1+\theta\right)D_i\right)^j}{j!}\right]\delta(x) + \frac{\left(c_1+\theta\right)^{i+1}\left(x+D_i\right)^i}{i!}e^{-(c_1+\theta)(x+D_i)}1 \quad (x \geq 0). \tag{38}$$

Observe that the definition of $\mathbf{p}_i$ implies that no interruptions occurred at the end of the preceding windows (D_i is a deterministic variable that only models the playback time). Additionally, we are considering that the user did not pause or move forward or backward the video. Despite these limitations, the distribution given in (38) can be very useful to define QoE targets as functions of the parameters of the system, as we will see in the following section.

6. QoE as a Function of Initial Delay and Interruption Duration in YouTube Service

Several works have identified that conventional QoS parameters (e.g., bandwidth, jitter, and delay) are not necessarily correlated to users' experience. Hence, significant effort has been made in order to link them to QoE parameters [22–25]. Particularly, in [22] some experiments were conducted in order to define MOS (one of the most used QoE parameters) as a function of the initial delay and the duration of interruptions that occur along the playback of YouTube videos, one of the most popular services of VoD.

In regard of initial delay (w_0 in our paper), [22] performed different measurements and expressed the MOS as $M_{\text{id}}(w_0) = 5 - 0.862 \log w_0 + 6.718$, where the highest MOS that can be reached is 4.287, occurring when $w_0 = 0$. Since in our model the video playback starts until window 0 has been downloaded, we can evaluate the previous equation by substituting w_0 with $\mathbf{w}_0$. Moreover, if we define an initial delay target MOS, denoted by $M_{\text{id,tgt}}$, we can define a corresponding target initial delay as

$$w_{\text{tgt}} = 10^{(5-M_{\text{id,tgt}})/0.862} - 6.718. \tag{39}$$

Finally, we can evaluate the probability that $M_{\text{id,tgt}}$ is not satisfied, since we know the distribution of $\mathbf{w}_0$:

$$\begin{aligned} q_{\text{id}} &= P\left\{M_{\text{id}}\left(\mathbf{w}_0\right) < M_{\text{id,tgt}}\right\} = P\left\{\mathbf{w}_0 > w_{\text{tgt}}\right\} \\ &= e^{-(c_0+\theta)w_{\text{tgt}}}. \end{aligned} \tag{40}$$

In [22] the authors also described the MOS as a function of the interruption duration (p_0), with the following expression: $M_{\text{int}}(p_0) = 1.75e^{-0.334p_0} + 3.19$. In this case, we can also define a target MOS, denoted by $M_{\text{int,tgt}}$. Hence, a target interruption duration can be defined as

$$p_{\text{tgt}} = -\frac{1}{0.334}\ln\left(\frac{M_{\text{int,tgt}} - 3.19}{1.75}\right). \tag{41}$$

At this point, it is important to remark that the aforementioned MOS model was designed considering only one interruption along the video playback. However, the model for interruptions that we developed in Section 5 considers interruptions at the end of every window. This means that only one of the random variables $\mathbf{p}_i$ can be used at one time to evaluate a probability analogous to (40). We select $\mathbf{p}_0$ because it has the advantage that it does not depend on other interruptions (since it is the first one) and, most important, because it is the most probable pause, as it can be seen from (38), under the assumption that $\delta_0 \leq d_0$ and $\delta_1 \leq d_1$.

Considering the previous paragraphs, we can also evaluate the probability that $M_{\text{int,tgt}}$ is not satisfied, since we know the distribution of $\mathbf{p}_0$; that is,

$$\begin{aligned} q_{\text{id}} &= P\left\{M_{\text{int}}(w_0) < M_{\text{int,tgt}}\right\} = P\left\{\mathbf{p}_0 > p_{\text{tgt}}\right\} \\ &= e^{-(c_1+\theta)(\delta_0+p_{\text{tgt}})}. \end{aligned} \tag{42}$$

It is relevant to make a comparison between (40) and (42) in terms of the size of window 0. It can be seen from (40) that when c_0 increases (the size of the window decreases), the probability that $M_{\text{id,tgt}}$ is not satisfied is reduced; under the same assumption, the probability of no satisfying $M_{\text{int,tgt}}$ is increased (since δ_0 and c_1 decrease). In other words, c_0 and c_1 must be selected in such a way that a trade-off between these QoE parameters exists. An alternative interpretation of this trade-off is that when the initial buffering is small, the user has a high probability of perceiving a satisfying small initial delay, but this, inherently, increases the probability of perceiving an unsatisfying interruption.

Having said that, we consider that in order to have a complete set of QoE parameters (and of elements to properly select c_0 and c_1), it is needed to define target probabilities that $M_{\text{id,tgt}}$ or $M_{\text{int,tgt}}$ are not satisfied, which are denoted by $q_{\text{id,tgt}}$ and $q_{\text{int,tgt}}$, respectively. According to (40) and (42), it can be said that we want that $q_{\text{id}} = e^{-(c_0+\theta)w_{\text{tgt}}} \leq q_{\text{id,tgt}}$ and that $q_{\text{int}} = e^{-(c_1+\theta)(\delta_0+p_{\text{tgt}})} \leq q_{\text{id,tgt}}$ which lead us to

$$d_0 \leq \frac{-w_{\text{tgt}}}{\ln\left(q_{\text{id,tgt}}\right) + \theta w_{\text{tgt}}}, \tag{43}$$

$$d_1 \leq \frac{-\left(\delta_0 + p_{\text{tgt}}\right)}{\ln\left(q_{\text{int,tgt}}\right) + \theta\left(\delta_0 + p_{\text{tgt}}\right)}. \tag{44}$$

As it will be shown in the next subsection, (43) and (44) are used to select the number of chunks in every window in such a way that the QoE parameter set be satisfied.

Another important issue that may be observed while selecting d_0 and d_1 is that while $M_{\text{id}}(w_0)$ is not very susceptible to increases in the initial delay, even small increases in the interruption duration may seriously degrade $M_{\text{int}}(p_0)$ ($M_{\text{id}}(w_0)$ is logarithmic, while $M_{\text{int}}(p_0)$ is exponential). According to [22], this is due to the fact that users are more tolerant to long initial delays than they are to long interruptions.

7. Parameter Selection for YouTube Services

In order to evaluate the performance of the proposed system, we are considering some currently implemented features of YouTube service by using the measurements reported by [26, 27]. According to [26], the download strategies that are used to download a video in YouTube depend on the network and the device that are involved. Hence, it is relevant to specify that the measurements they are reporting were obtained with mobile devices downloading videos through a WiFi network.

We now give a brief explanation on how some YouTube service parameters are related to our model. According to [26], the most common video format in YouTube service (MPEG-4 Visual) has an encoding rate between 200 and 275 kb/s, and the authors also identify that the download rate allowed by this service is two times the encoding rate. If we denote by r_{cd} the ratio between the download rate and the encoding rate, we have $r_{\text{cd}} = 2$ for the aforementioned case.

In addition, in [27] it is reported that the average YouTube video duration is 490.5 seconds. Considering this information, we selected as reference for a numerical evaluation a codification rate of 200 kb/s and a chunk size of 64 kBytes, characterizing a video file whose features are illustrated in Figure 9. We denote by M the number of chunks in the file. As such, according to the previous data, for our reference file we chose $M = 192$. Additionally we denote by k_0 and k_1 the number of chunks in window 0 and in any other window, respectively.

From the previous data it can also be established that the normalized download rate for the reference file would be $c = 0.00407$. Because of the usual asymmetry in the users network access links, we consider $\mu < c$. Specifically we select a normalized upload rate $\mu = 0.00255$ (which is equivalent to a data transmission rate of 250 kb/s, for our reference file). Given the reference file, the possible values for N are between 1 and 192; so we consider $N = 24, 48, 60, 72, 84$, and 96 for most of our evaluations.

According to [27], the average playback time of a YouTube video before it is interrupted by the user is 172 seconds. Notice that in our model we identified this variable as the leeches' sojourn time. Hence $\theta = 1/172 = 0.0058$. However, in order to evaluate our system under a variety of scenarios, we set θ for values from 0.002 to 0.01 (the superior limit corresponds to a sojourn time of 100 seconds).

To model the seeds sojourn time, we set γ in the range from 0.006 to 0.02; this accounts for seeds remaining in the system an average of 50 to 167 seconds.

Having introduced the video-file parameters, we can establish relations between them and some model variables.

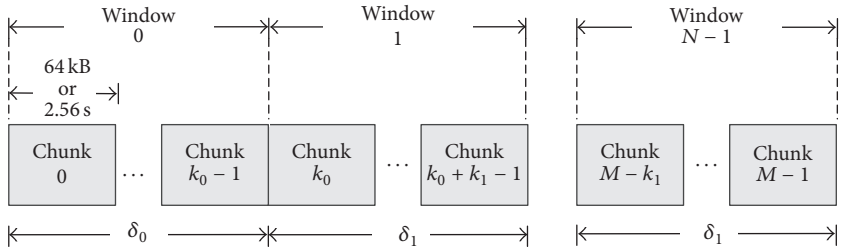


Figure 9: Evaluation reference file.

Specifically, $d_0 = k_0/(cM)$, $d_1 = k_1/(cM)$, $\delta_0 = r_{cd}d_0$, and $\delta_1 = r_{cd}d_1$.

In the following we provide a method to select the design parameters of the system:

(i) Since we are interested in satisfying a predetermined QoE parameter set, we have to select d_0 and d_1 in such a way that inequalities (43) and (44) are simultaneously satisfied. Since (43) only depends on θ and QoE targets we can select the number of chunks in window 0 by means of the following expression:

$$k_{0,s} = \left\lfloor \frac{-w_{tgt}cM}{\ln\left(q_{id,tgt}\right) + \theta w_{tgt}} \right\rfloor. \tag{45}$$

(ii) Once k_0 is selected, a calculation analogous to (45) can be carried out to select k_1 using (44):

$$k_{1,s} = \left\lfloor \frac{-cM\left(r_{cd}k_{0,s} + cMp_{tgt}\right)}{cM\ln\left(q_{int,tgt}\right) + \theta\left(r_{cd}k_{0,s} + cMp_{tgt}\right)} \right\rfloor. \tag{46}$$

(iii) From (45) and (46) we can define the selected values of N and α:

$$\begin{aligned} N_s &= \left\lceil \frac{M - k_{0,s}}{k_{1,s}} \right\rceil + 1, \\ \alpha_s &= \frac{k_{1,s}}{k_{0,s}}. \end{aligned} \tag{47}$$

Notice that $k_{0,s}$ and $k_{1,s}$ are the maximum values that satisfy the QoE parameters set, which means that if lower values were used, N would acquire a larger value than N_s. However, as it is shown in Section 8, the larger the value of N, the larger the value of μ_s required to guarantee abundance.

Assuming that the users' behavior and download/upload parameters are known and using N_s and α_s, it is possible to numerically evaluate (30). This solution provides the optimal values of ε and i. By substituting these values in (27) or (28), depending on the value of i, the minimum value of μ_s that guarantees abundance can be calculated. Notice that the previous operations are needed only if the PWD scheme is used; if the UD scheme is used, μ_{min} can be calculated by simply substituting N_s and α_s, as well as the remaining parameters, in (24).

8. Numerical Results and Discussions

In this section, we provide relevant numerical results that evaluate the performance of the system in terms of the required server capacity to guarantee QoE, the average initial delay, and the average download time for different system parameters. Additionally, we provide the required parameters to guarantee an acceptable level of QoE for different scenarios.

First, we show in Figure 10 the minimum bandwidth that servers must provide in order to achieve abundance in all the windows, while making a comparison between the distribution schemes described in Sections 4.1 and 4.2. The values shown in this figure were found by converting the normalized rates (file length equal to one) into practical data rates, according to the values given in Section 7. As it was previously explained, the PWD drastically reduces the extra capacity required to provide abundance conditions in all the system compared to the uniform distribution scheme. It is also important to mention that, as it was expected, when leeches are more cooperative (θ takes small values), the value of μ_{min} is smaller. Additionally, it is crucial to emphasize that increases in the number of windows significantly increase the required server bandwidth; however, as it is shown later, a very small number of windows have a negative impact on the QoE parameters. In addition to this, Figure 11 shows μ_{min} for PWD and two different values of γ. It can be seen that the existence of cooperative seeds (small values of γ) significantly reduces the need of bandwidth provided by servers in the P2P network.

In Figures 12 and 13 we show the number of leeches ($\overline{x}$) and seeds ($\overline{y}$) in equilibrium, respectively. In Figures 14 and 15 we show the system performance in terms of the average initial delay (T_0) and the average video download delay (T), respectively. Those results were obtained through the evaluation of (17), (16), (33), and (36). We also solved an associated Markov chain (see the Appendix) for cross-correlating the results against those in the fluid model.

These evaluations were carried out using the PWD scheme, as well as a server capacity of 120 Mb/s, which is equivalent to a normalized download file rate of $\mu_s = 1.24$. Also, abundance conditions are guaranteed. As it was expected, a small value of θ entails a large number of leeches, as it is depicted in Figure 12. Since the seeds are peers that finish the file download, the number of them also increases when θ decreases, as it is shown in Figure 13. Indeed, these conditions correspond to a cooperative system which, as previously

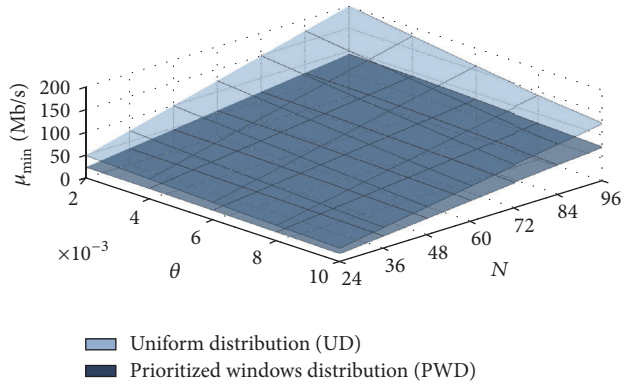


Figure 10: Minimum server bandwidth to achieve abundance, considering $\lambda = 0.04$, $c = 0.00407$, $\mu = 0.00255$, $\gamma = 0.006$, and $\alpha = 1$.

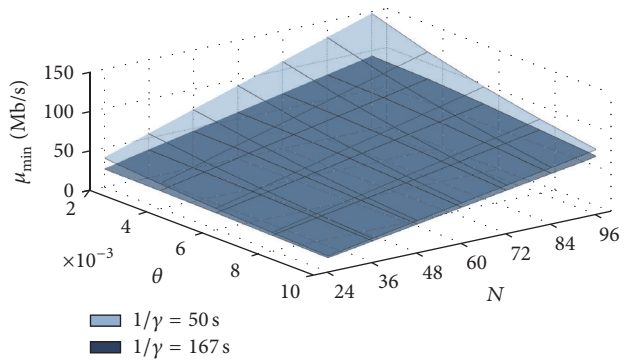


Figure 11: Minimum server bandwidth to achieve abundance, considering the PWD scheme, $\lambda = 0.04$, $c = 0.00407$, $\mu = 0.00255$, and $\alpha = 1$.

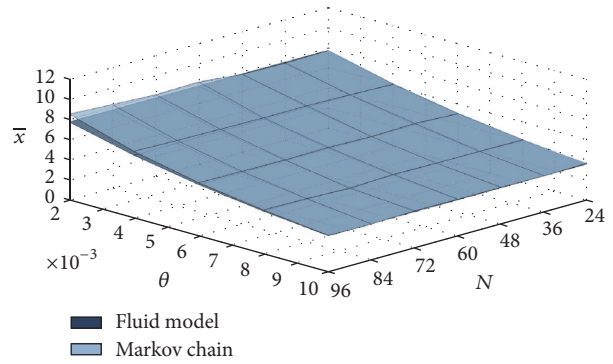


Figure 12: Number of leeches in equilibrium, considering PWD scheme, $\lambda = 0.04$, $\gamma = 0.006$, $c = 0.00407$, $\mu = 0.00255$, $\mu_s = 1.24$, and $\alpha = 1$.

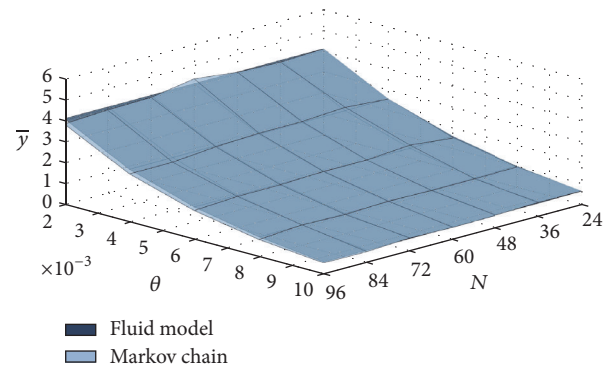


Figure 13: Number of seeds in equilibrium, considering PWD scheme, $\lambda = 0.04$, $\gamma = 0.006$, $c = 0.00407$, $\mu = 0.00255$, $\mu_s = 1.24$, and $\alpha = 1$.

mentioned, reduces the required assisted server bandwidth to guarantee abundance. Though these parameters are not directly related to the QoE levels, they offer an insight into the system performance. For example, it is clear that the larger the number seeds, the larger the capacity in the system.

In Figure 14 we corroborate that in order to reduce the average initial delay, T_0, the size of the first window must be reduced. In these results, the initial window size is reduced by increasing N, although the rest of the windows are also reduced, since we selected $\alpha = 1$. Indeed, by increasing N, the number of chunks per window decreases, effectively reducing the initial delay. However, this has a negative effect in the overall system performance, as shown in Figure 10 where a high value of N entails a higher amount of server bandwidth to maintain the system in abundance conditions. As such, in order to reduce the initial delay efficiently, the value of α should be increased instead. Additionally, it must be noticed that θ has no significant effect in T_0, since it is related to both system capacity and resources demand. This last result exhibits the scalability properties of P2P networks.

Regarding the performance of average download delay, T, Figure 15 presents some interesting results that show that under abundance conditions its value is almost a constant. The reason for this is that, in an abundance situation, the required time to download any window is the same (except for window 0 when $\alpha \neq 1$), according to (36). It is important to note that results in Figures 12–15 directly proved that our analytical solution agrees with the numerical one obtained by the Markov chain.

In Figure 16 we evaluate the relation between the QoE parameters that are associated with initial delay and the design parameters α and N. An analogous comparison is shown in Figure 17 for the QoE parameters that are associated with interruptions. As expected, the higher the target MOS ($M_{\text{id,tgt}}$ or $M_{\text{int,tgt}}$), the higher the probability of no satisfaction, q_{id} and q_{int}, respectively. We can also appreciate that any of these probabilities increases when N diminishes, since a small N means a large window. In these figures the effect of α can also be appreciated: when it increases, the size of the initial window decreases and as a consequence q_{id} is improved, but q_{int} is degraded.

In Tables 1 and 2 we summarize the values that were found for the design parameters ε, N, α, and μ_s by following the proposed methodology described at the end of Section 7. Table 1 simply shows that when QoE targets are high, a large amount of server bandwidth must be provided. On the other hand, Table 2 confirms that a P2P network with cooperative seeds (small values of γ) requires a small amount of extra

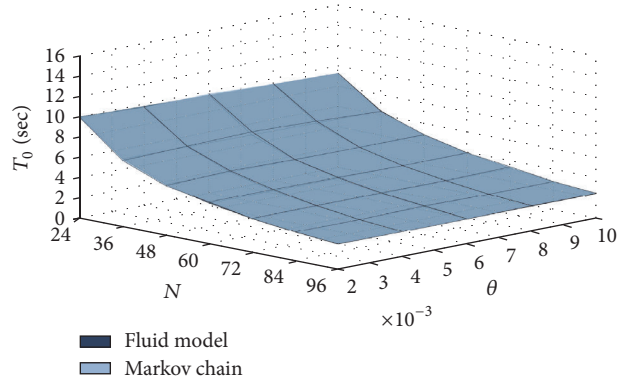


Figure 14: Average time to download window 0 (initial delay), considering PWD scheme, $\lambda = 0.04$, $\gamma = 0.006$, $c = 0.00407$, $\mu = 0.00255$, $\mu_s = 1.24$, and $\alpha = 1$.

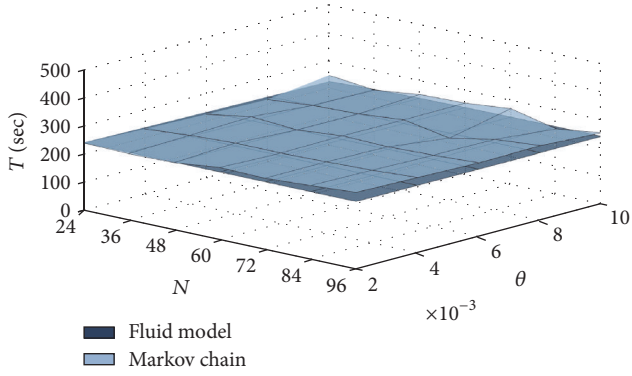


Figure 15: Average time to download the whole file, considering PWD scheme, $\lambda = 0.04$, $\gamma = 0.006$, $c = 0.00407$, $\mu = 0.00255$, $\mu_s = 1.24$, and $\alpha = 1$.

Table 1: Design parameters as functions of the QoE parameters set for $\lambda = 0.04$, $\theta = 0.006$, $\gamma = 0.006$, $c = 0.00407$, and $\mu = 0.00255$.

$M_{\text{id,tgt}}$	$q_{\text{id,tgt}}$	$M_{\text{int,tgt}}$	$q_{\text{int,tgt}}$	ε	N	α	$\mu_{s,pw}$	$\mu_{s,u}$
4.0	5%	4.0	5%	0.78	191	0.5	171.9	319.3
4.0	10%	4.0	10%	0.84	96	1	86.3	159.2
3.9	10%	4.0	10%	0.82	64	0.75	57.7	106.3
3.9	10%	3.9	10%	0.84	48	1	43.2	79.1

server bandwidth but also shows that when θ is large, the required amount of extra server bandwidth is small, since the download demands are reduced, even though this situation implies the existence of noncooperative leeches.

Finally, it must be remarked that we provide the necessary tools and analytical methodology to select the design parameters (ε, N, and α) that satisfy a target set of QoE parameters, given the basic system variables (c, μ, θ, γ, etc.). Furthermore, this is done under a scheme (PWD) that considerably reduces the amount of server bandwidth that is needed to maintain the system in abundance conditions.

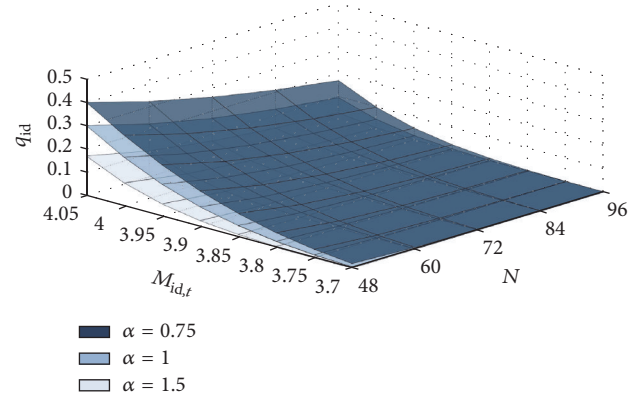


Figure 16: Probability of not satisfying $M_{\text{id},t}$, considering PWD scheme, $\lambda = 0.04$, $\gamma = 0.006$, $c = 0.00407$, $\mu = 0.00255$, and $\mu_s = 1.24$.

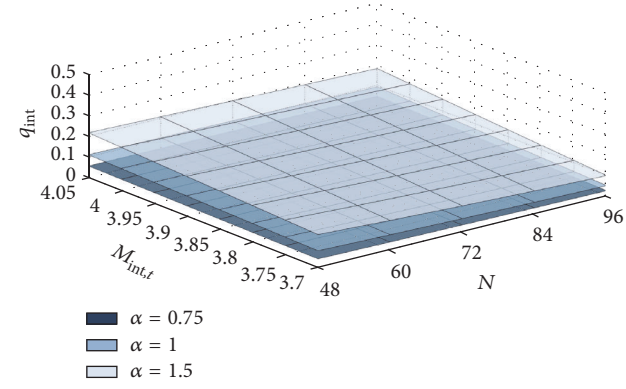


Figure 17: Probability of not satisfying $M_{\text{int},t}$, considering PWD scheme, $\lambda = 0.04$, $\gamma = 0.006$, $c = 0.00407$, $\mu = 0.00255$, and $\mu_s = 1.24$.

9. Conclusions and Future Work

In order to achieve abundance conditions in a P2P network, the existence of cooperative peers is necessary. Since this cooperative scenario does not necessarily exist, it is obviously possible to guarantee an acceptable performance of the system if some servers provide extra upload bandwidth. Since this extra bandwidth represents an additional cost, it is imperative to make an efficient use of it. Building on that, we proposed the PWD scheme. Our numerical evaluations showed that this scheme significantly reduces the amount of extra bandwidth required in the system to satisfy a set of QoE parameters, in comparison with a scheme that uniformly distributes those resources. We conclude that this scheme is one of our major contributions, since it implicitly takes advantage of the peers upload bandwidth, by assigning extra server bandwidth only to leeches that really need it, and, at the same time, making the system less dependent of peers' behavior.

On the other hand, the window-based sharing mechanism not only provides an easy-to-implement and efficient way to interchange video files, but also allows finding a trade-off between the initial delay and the interruption duration,

Table 2: Design parameters in terms of the users' behavior parameters set for $\lambda = 0.04$, $c = 0.00407$, $\mu = 0.00255$, $M_{\text{id,tgt}} = 3.9$, $q_{\text{id,tgt}} = 10\%$, $M_{\text{int,tgt}} = 4.0$, and $q_{\text{int,tgt}} = 10\%$.

θ	γ	ε	N	α	$\mu_{s,pw}$	$\mu_{s,u}$
0.006	0.006	0.82	64	0.75	43.2	79.1
0.006	0.02	0.69	64	0.75	72.3	123.8
0.01	0.006	0.65	48	1	37.0	62.4
0.01	0.02	0.58	48	1	54.2	92.2

by varying the size of the initial window. This is another contribution of this work, since, to the best of our knowledge, this trade-off has not been previously analyzed in the context of QoE parameters (measured by a MOS metric).

In addition, we also developed an evaluation framework that can be used to calculate the design parameters of the system (number and sizes of windows, upload server bandwidth) that satisfy a set of target QoE-based values, given the behavior of the peers (peers' arrival/departure rates) and the network features (peers' upload/download bandwidth). According to the numerical evaluations that we reported (which capture some of the features of YouTube service), we conclude that this framework is a powerful design tool that can be used by VoD servers providers in order to reduce their implementation costs, while controlling the QoE that they give to their users.

From the results obtained in this work, we have identified a number of issues that have to be addressed in a future work. We are particularly interested in including in our analysis the possibility of pause or moving forward or backward in the video while it is being downloaded and shared, as well as considering varying peer's upload/download bandwidth conditions.

Appendix

Markovian Model

With the fluid model proposed in this paper (Section 3), we can associate a Markovian one as follows. We have now a discrete model of the same system. Consider vector $W(t) = (L_0(t), L_1(t), \ldots, L_{N-1}(t), S(t)) \geq (0, 0, \ldots, 0, 1)$, where $L_i(t)$ and $S(t)$ are, respectively, the number of leeches at window w_i, $i = 0, \ldots, N-1$, and of seeds, at time t. With the standard "exponential assumptions" (plus independence), $W(t)$ is a continuous time homogeneous Markov chain, with initial state $W(0) = (0, 0, \ldots, 0, 1)$. Starting from state $(l_0, l_1, \ldots, l_{N-1}, m)$, for any $i \in \{0, 1, \ldots, N-1\}$, $l_i \geq 0$ and $m \geq 1$, the transition rates are described as follows:

(i) λ, to state $(l_0 + 1, l_1, \ldots, l_{N-1}, m)$,

(ii) $l_i\theta$, to state $(l_0, l_1, \ldots, l_i - 1, \ldots, l_{N-1}, m)$ $(l_i \geq 1)$,

(iii) $\tau^*_{i-1,i}$, to state $(l_0, l_1, \ldots, l_{i-1} - 1, l_i + 1, \ldots, l_{N-1}, m)$ $(1 \leq i \leq N-2$ and $l_{i-1} \geq 1)$,

(iv) $\tau^*_{N-1,N}$, to state $(l_0, l_1, \ldots, l_{N-1} - 1, m + 1)$ $(l_{N-1} \geq 1)$,

(v) $(m-1)\gamma$, to state $(l_0, l_1, \ldots, l_{N-1}, m-1)$ $(m \geq 2)$,

where, for $0 \leq i \leq N-2$,

$$\tau^*_{i,i+1} = \min\left\{c_w l_i, \mu_w\left(l_i \frac{\sum_{k=i+1}^{N-1} l_k}{\sum_{j=0}^{k-1} l_j} + m\frac{l_i}{l}\right)\right\},$$
$$\tau^*_{N-1,N} = \min\left\{c_w l_{N-1}, \mu_w\left(m\frac{l_{N-1}}{l}\right)\right\}, \quad \text{(A.1)}$$

with $l = \sum_{k=0}^{N-1} l_k$.

This model can be used to get confidence in the good behavior of the continuous dynamical one, governed by differential equations. The latter has the huge advantage that analytical expressions are often available (the former can only be simulated). The mathematical relations between both approaches are out of the scope of this paper (see [28] for some fundamentals and [29] for the analysis of a P2P case, the one described in [18]).

Most Relevant Variables Summary

c: File download rate ($c \geq \mu$)
c_0: Initial window download rate
c_1: Window i download rate, $i > 0$
$M_{\text{id,tgt}}$: Target MOS as a function of initial delay
$M_{\text{int,tgt}}$: Target MOS as a function of interruption duration
N: Number of windows
q_{id}: Probability that $M_{\text{id,tgt}}$ is not satisfied
q_{int}: Probability that $M_{\text{int,tgt}}$ is not satisfied
T: Required time to download the video file
T_0: Required time to download the initial window
$\overline{x}$: Number of leeches in steady state
$\overline{x}_i$: Number of leeches in window i in steady state
$\overline{y}$: Number of seeds in steady state
α: Noninitial window size to initial window size ratio
γ: Departure rate for a seed
ε: Priority control parameter
θ: Departure rate for a leech
λ: Arrival rate of peers
μ: File upload rate
μ_0: Initial window upload rate
μ_1: Window i upload rate, $i > 0$
$\mu_{\min}$: Minimum servers upload bandwidth that guarantees abundance in the whole system
$\mu^i_{\min}$: Minimum servers upload bandwidth that guarantees abundance in window i
μ_s: Servers upload bandwidth
τ_i: Transition rate from window i to window $i+1$
τ_{N-1}: Transition rate from window $N-1$.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] Y.-W. Huang and Y.-C. Chen, "A study on content delivery strategy for QoE enhancement in P2P IPTV," in *Proceedings of the 14th IEEE International Conference on Computer and Information Technology (CIT '14)*, pp. 520–527, Xi'an, China, September 2014.

[2] D. Kondo, Y. Hirota, A. Fujimoto, H. Tode, and K. Murakami, "P2P live streaming system for multi-view video with fast switching," in *Proceedings of the 16th International Telecommunications Network Strategy and Planning Symposium, Networks 2014*, 7, 1 pages, September 2014.

[3] O. Ojo, A. Oluwatope, and O. Ogunsola, "UStream: ultra-metric spanning overlay topology for peer-to-peer streaming systems," in *Proceedings of the 17th IEEE International Symposium on Multimedia (ISM '15)*, pp. 601–604, Miami, Fla, USA, December 2015.

[4] Y. Zhou, T. Z. J. Fu, and D. M. Chiu, "A unifying model and analysis of P2P VoD replication and scheduling," *IEEE/ACM Transactions on Networking*, vol. 23, no. 4, pp. 1163–1175, 2015.

[5] P. Romero, F. Robledo, P. Rodríguez-Bocca, and C. Rostagnol, "Analysis and design of peer-assisted video-on-demand services," *International Transactions in Operational Research*, vol. 21, no. 4, pp. 559–579, 2014.

[6] W. Wu, R. T. B. Ma, and J. C. S. Lui, "Distributed caching via rewarding: an incentive scheme design in p2p-vod systems," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 3, pp. 612–621, 2014.

[7] M. Faiqurahman and A. I. Kistijantoro, "Implementation of modified Probabilistic Caching Schema on Bittorrent protocol for video on demand content," in *Proceedings of the 16th International Seminar on Intelligent Technology and Its Applications (ISITIA '15)*, pp. 357–362, IEEE, Surabaya, Indonesia, May 2015.

[8] M. Wichtlhuber, S. Dargutev, S. Müller, A. Klein, and D. Hausheer, "QTrade: a quality of experience based peercasting trading scheme," in *Proceedings of the IEEE International Conference on Peer-to-Peer Computing (P2P '15)*, Boston, Mass, USA, September 2015.

[9] T. Rohmer, A. Nakib, and A. Nafaa, "Priori knowledge guided approach for optimal peer selection in P2P VoD systems," *IEEE Transactions on Network and Service Management*, vol. 11, no. 3, pp. 350–362, 2014.

[10] R. Dubin, O. Hadar, Y. Freifeld et al., "Hybrid clustered peer-assisted DASH-SVC system," in *Proceedings of the IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM '15)*, pp. 1651–1656, October 2015.

[11] D. Ciullo, V. Martina, M. Garetto, E. Leonardi, and G. L. Torrisi, "Peer-assisted VoD systems: an efficient modeling framework," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 7, pp. 1852–1863, 2014.

[12] "Global Internet Phenomena. Latin America and North America Report 2015," Internet: https://www.sandvine.com.

[13] Ericsson Mobility Report, June 2016, https://www.ericsson.com/mobility-report.

[14] S. Jia, Y. Ma, Y. Zhang, T. Wang, and M. Zhang, "Modelling of P2P-based video sharing performance for content-oriented community-based VoD systems in wireless mobile networks," *Mobile Information Systems*, vol. 2016, Article ID 1319497, 13 pages, 2016.

[15] C. Xu, S. Jia, L. Zhong, and G.-M. Muntean, "Socially aware mobile peer-to-peer communications for community multimedia streaming services," *IEEE Communications Magazine*, vol. 53, no. 10, pp. 150–156, 2015.

[16] E. Esquivel, M. E. Rivero-Angeles, and G. Rubino, "Priority scheme for window-based video-on-demand transmission on BitTorrent-like Peer-to-Peer networks," in *Proceedings of the IEEE International Conference on Communications (ICC '13)*, pp. 3000–3005, Budapest, Hungary, June 2013.

[17] M. E. Rivero-Angeles, G. Rubino, I. O. O. Torres, and L. A. Martinez, "Window-based streaming video-on-demand transmission on BitTorrent-like peer-to-peer networks," in *Proceedings of the IEEE 10th Consumer Communications and Networking Conference (CCNC '13)*, pp. 1–6, IEEE, January 2013.

[18] M. E. Rivero-Angeles and G. Rubino, "Priority-based scheme for file distribution in peer-to-peer networks," in *Proceedings of the IEEE International Conference on Communications (ICC '10)*, pp. 1–6, May 2010.

[19] S. Tewari and L. Kleinrock, "Analytical model for BitTorrent-based live video streaming," in *Proceedings of the 4th Annual IEEE Consumer Communications and Networking Conference (CCNC '07)*, pp. 976–980, IEEE, Las Vegas, Nev, USA, January 2007.

[20] C. Liang, Z. Fu, Y. Liu, and C. W. Wu, "Incentivized peer-assisted streaming for on-demand services," *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 9, pp. 1354–1367, 2010.

[21] D. Qiu and R. Srikant, "Modeling and performance analysis of BitTorrent-like peer-to-peer networks," in *Proceedings of the 2004 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM '04)*, Portland, Ore, USA, October 2004.

[22] T. Hossfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, and C. Lorentzen, "Initial delay vs. interruptions: between the devil and the deep blue sea," in *Proceedings of the 4th International Workshop on Quality of Multimedia Experience (QoMEX '12)*, pp. 1–6, IEEE, Yarra Valley, Australia, July 2012.

[23] P. T. A. Quang, K. Piamrat, and C. Viho, "QoE-aware routing for video streaming over ad-hoc networks," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM '14)*, pp. 181–186, December 2014.

[24] S. Mohamed and G. Rubino, "A study of real-time packet video quality using random neural networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 12, pp. 1071–1083, 2002.

[25] K. D. Singh, Y. Hadjadj-Aoul, and G. Rubino, "Quality of experience estimation for adaptive HTTP/TCP video streaming using H.264/AVC," in *Proceedings of the IEEE Consumer Communications and Networking Conference (CCNC '12)*, pp. 127–131, Las Vegas, Nev, USA, January 2012.

[26] J. Ramos-Munoz, J. Prados-Garzon, P. Ameigeiras, J. Navarro-Ortiz, and J. Lopez-Soler, "Characteristics of mobile youtube traffic," *IEEE Wireless Communications*, vol. 21, no. 1, pp. 18–25, 2014.

[27] G. Dimopoulos, P. Barlet-Ros, and J. Sanjuàs-Cuxart, "Analysis of YouTube user experience from passive measurements," in *Proceedings of the 9th International Conference on Network and Service Management (CNSM '13)*, pp. 260–267, Zurich, Switzerland, October 2013.

[28] S. N. Ethier and T. Kurtz, *Markov Processes: Characterization and Convergence*, John Wiley & Sons, New York, NY, USA, 1986.

[29] L. Aspirot, E. Mordecki, and G. Rubino, "Fluid limits applied to peer to peer network analysis," in *Proceedings of the 8th International Conference on Quantitative Evaluation of Systems (QEST '11)*, pp. 13–20, IEEE, Aachen, Germany, September 2011.

Hindawi
Mobile Information Systems
Volume 2017, Article ID 7843843, 12 pages
https://doi.org/10.1155/2017/7843843

Research Article

A Novel Secure Transmission Scheme in MIMO Two-Way Relay Channels with Physical Layer Approach

Qiao Liu,[1,2] Guang Gong,[2] Yong Wang,[1] and Hui Li[1]

[1]*State Key Lab of ISN, Xidian University, Xi'an, Shaanxi, China*
[2]*Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada*

Correspondence should be addressed to Qiao Liu; windachilles@gmail.com

Received 20 October 2016; Revised 24 December 2016; Accepted 29 December 2016; Published 2 March 2017

Academic Editor: Jing Zhao

Security issue has been considered as one of the most pivotal aspects for the fifth-generation mobile network (5G) due to the increasing demands of security service as well as the growing occurrence of security threat. In this paper, instead of focusing on the security architecture in the upper layer, we investigate the secure transmission for a basic channel model in a heterogeneous network, that is, two-way relay channels. By exploiting the properties of the transmission medium in the physical layer, we propose a novel secure scheme for the aforementioned channel mode. With precoding design, the proposed scheme is able to achieve a high transmission efficiency as well as security. Two different approaches have been introduced: information theoretical approach and physical layer encryption approach. We show that our scheme is secure under three different adversarial models: (1) untrusted relay attack model, (2) trusted relay with eavesdropper attack model, and (3) untrusted relay with eavesdroppers attack model. We also derive the secrecy capacity of the two different approaches under the three attacks. Finally, we conduct three simulations of our proposed scheme. The simulation results agree with the theoretical analysis illustrating that our proposed scheme could achieve a better performance than the existing schemes.

1. Introduction

As the next evolution of the mobile communication system, 5G (5th-generation mobile network) has become the hottest topic in the academia as well as industry. To meet the high rate requirement in a cost-efficient way, many techniques are designed to be employed into the 5G system, including heterogeneous network (HetNet), massive multiple inputs multiple outputs (Massive MIMO), Device-to-Device (D2D), and millimeter wave (mmWave) techniques.

Meanwhile, security issues have attracted much more attention compared to any other time in the past. Thus, research on 5G security would undoubtedly be of theoretical and practical interest. Besides focusing on the security strategy for the system level, it is also worth investigating security transmission in some basic channel model, especially relevant to the aforementioned key techniques for 5G. With such motivation, we consider the secure transmission in two-way relay channels, which are one of the most basic channel models in HetNet and D2D networks. In addition, the proposed scheme is designed for multiple antenna system. Although this scheme is not particularly designed for Massive MIMO, it can be easily transplanted into it.

The research on TWRC channel has continued for a long time; the early works on this type of channel model are concentrated on efficient transmission [1–8]. The results in these works show that MIMO TWRC channels can drastically improve the transmission performance. After that, some researchers began to think about the security of this type of channel model. Instead of following the upper layer encryption approach, the physical layer approach is believed to be a promising way which enjoys less system complexity compared with the traditional cryptography scheme.

Along with the pioneering study by Wyner [9], physical layer security problems were first introduced into MIMO case by Hero [10] utilizing space-time code at the transmitter to enhance information security and hiding capabilities. After that, the research of multiple antennas mainly focused on MISO (multiple input, single output) [11] or SIMO (single input, multiple output) [12] until Khisti et al. analyzed the

MIMO wiretap channel secrecy capacity in [13], and they gave an upper bound for secrecy capacity under the situation that the transmitter knows the instantaneous channel state information (CSI) about the eavesdropper. After that, a lot of researchers aimed at giving a secrecy capacity bound by different approaches and different constraints [14–16].

The physical layer security in cooperation relaying was first considered in [17]. Depending on the relay adversarial model, the security problems in cooperation relaying system are divided into two parts in [18]: (1) untrusted relay model and (2) trusted relay model.

For the untrusted relay model, the relay itself acts as an untrusted node which may attempt to illegitimately recover the information messages from the users. This is a common case in the HetNet network, since many potential unfriendly devices exist in the HetNet network and some of them are eager to wiretap to the messages by providing fake assistance. In [19], the authors give a joint source and relay secure beamforming design for the one-way MIMO untrusted relay model. Transmitting jamming signals by friendly jammers in [20] is a secure method for TWRC channels, but the selection of friendly jammers will be difficult to realize in practice. The authors in [21] give an approach to achieve secrecy capacity in MIMO two-way untrusted relay channels based on the signal alignment precoding. However, this scheme is power inefficient especially in bad channel condition. So, the optimization of signal alignment is critical in improving the secrecy capacity.

For the trusted relay model, the relay assists the legitimate users to achieve secure transmission. A lot of works have focused on the single antenna system. Securing the trusted relay model for MIMO systems was first introduced in [22], which uses artificial noise alignment to jam the eavesdropper. The authors in [23] present a physical layer network coding design with secure precoding for two-way MIMO trusted relay channels.

After reviewing these existing solutions in the literature, we feel that considerable improvements can be made in terms of transmission efficiency and security for MIMO TWRC channels. Two approaches have been introduced based on different performance requirements: information theoretical approach and physical layer encryption approach.

Motivated by [5], we use *Direction Rotation Alignment* as the key to our information theoretical approach. From the transmission efficiency aspects, Direction Rotation Alignment can overcome the power loss in signal alignment scheme. From the physical layer security aspects, the alignment of the two separated signals causes the received signal to be a signal sum in the view of the intended receivers, relay, and eavesdroppers. However, only the intended receivers can directly decode the information symbols from their communication partners with their self-information serving as the private key, while the relay or eavesdroppers can obtain partial information with the sum signal. Therefore, by finding the ideal transmission rate, the system can achieve information theoretical security.

On the other hand, encryption vector has been nested into precoding matrix in physical layer encryption approach. After physical layer encryption, the signal direction of each user will be distorted. So, these will certainly bring about enough confusion for the adversaries. With such encryption, the system can achieve computational security.

The main contribution and results of this paper are listed below:

(i) A new information theoretical security approach is introduced with key technique Direction Rotation Alignment. This technique can eliminate the power loss caused by signal alignment. At the same time, this technique can conceal the user message to achieve information theoretical security.

(ii) Following information theoretical approach, physical layer encryption approach is presented to achieve better transmission efficiency and security performance.

(iii) We show that our proposed scheme is secure under three different adversary models: untrusted relay attack model, (2) trusted relay with eavesdropper attack model, and (3) untrusted relay with eavesdroppers attack model. To the best of our knowledge, our scheme is the first secure method in all these adversary models

(iv) We analyze the secrecy capacities of the two approaches under each adversary model. With such analysis, ideal transmission rate could be found.

The paper is organized as follows. In Section 2, we introduce the system and adversarial models. Section 3 presents information theoretical approach as well as the capacity analysis under different adversarial models. And the physical layer encryption approach with its capacity analysis is discussed in Section 4. In Section 5, we demonstrate simulation results on our proposed scheme. Finally, we give conclusions and extensions in Section 5.

Notations. $\mathrm{Tr}(\cdot)$, $\varepsilon(\cdot)$, $(\cdot)^{-1}$, and $\det(\cdot)$ denote the trace, expectation, inverse or pseudoinverse, and determinant of matrix, respectively. And $[x]^+$ denotes the max $(0, x)$.

2. System Model

In this section, we will introduce the system model for the proposed scheme which is previously defined in [24].

2.1. Transmission Model

(1) Channel Model. In this subsection, we will describe the TWRC channels system. This is depicted in Figure 1. K communication pairs exchange their information with a relay. The users on the left side in Figure 1 are denoted as A_k ($k \in \kappa \{\kappa = 1, 2, \ldots, K\}$) and the users on the right side are denoted as B_k. Furthermore, we assume each user is equipped with n_T antennas, and the relay is equipped with n_R antennas.

Both the relay and the users work in half-duplex mode and there is no direct link between each pair. We assume that all the channels experience the flat fading and the channel coefficient between user m ($m \in \{A_k, B_k\}$) and relay is $\mathbf{H}_m$ which is an $n_R * n_T$ matrix. The channel coefficient between

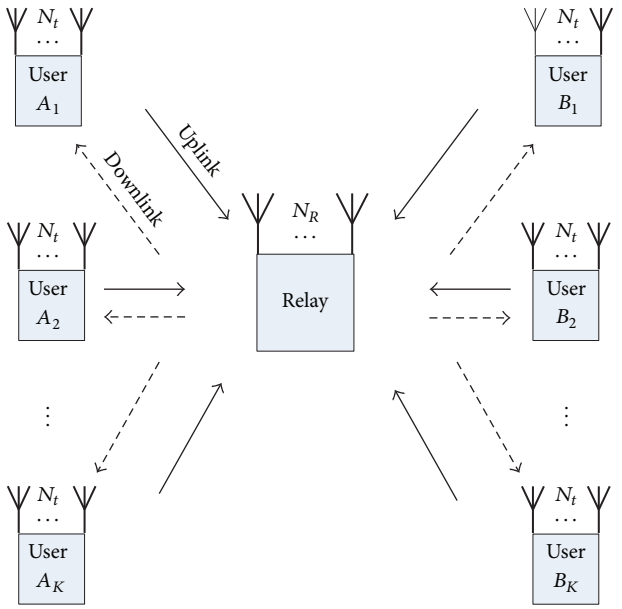


Figure 1: The channel model of multiusers MIMO two-way relay channels.

relay and user m is $\mathbf{G}_m$ with the size of $n_T * n_R$. Since all channels experience flat fading, both $\mathbf{H}_m$ and $\mathbf{G}_m$ are kept constant in each round of information exchange.

Finally, we assume that the channel state information (CSI) is available for the users and relay. The channel state information could be obtained by the channel estimation for the MIMO channel. There are already a lot of relative works focusing on the channel estimation. In [25], the CSI has been estimated with partial channel information. In addition, the authors propose a semiblind estimation method in [26]. Particularly, in [27], the authors consider using fine time synchronization in the estimation which is the most suitable method for our proposed scheme.

The proposed transmission protocol consists of two time slots to accomplish one round of information exchange. In the first time slot, all users transmit their information to the relay simultaneously. Because the users act as a source node in the first time slot, this time slot is called uplink phase or multiple access (MAC) phase. Upon receiving the message, the relay broadcasts its signal in the second time slot with the name of downlink or broadcast (BC) phase. Now, we introduce the two phases separately.

(2) Uplink Phase. In the uplink (MAC) phase, the relay receives the converging signals from all the user nodes as follows:

$$\mathbf{Y}_R = \sum_{k=1}^{K} \left(\mathbf{H}_{A_k} \mathbf{X}_{A_k} + \mathbf{H}_{B_k} \mathbf{X}_{B_k} \right) + \mathbf{Z}_R, \tag{1}$$

where $\mathbf{X}_{A_k}$ is an $n_T * 1$ column vector that represents the transmitted signal vector of user A_k containing the information message $\mathbf{c}_{A_k}$; $\mathbf{X}_{B_k}$ represents the transmitted signal of user B_k; $\mathbf{Y}_R$ denotes the received signal vector by relay, which is an $n_R * 1$ column vector; and $\mathbf{Z}_R$ is an $n_R * 1$ zero mean circularly symmetric complex Gaussian noise vector at the relay node modelled by $\mathbf{Z}_R \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$.

We denote the covariances of the channel inputs in user m as $\mathbf{Q}_m = \varepsilon(\mathbf{X}_m \mathbf{X}_m^H)$. Then, we have the power constraint in uplink phase like

$$\mathrm{Tr}\left\{ \sum_{k=1}^{K} \left(\mathbf{Q}_{A_k} + \mathbf{Q}_{B_k} \right) \right\} \leq P_T. \tag{2}$$

(3) Downlink Phase. In the downlink (BC) phase, the relay broadcasts its signal $\mathbf{X}_R$ to all users, and each user recovers the information message from its communication partner.

The relay is set up as an Amplify-and-Forward (AF) model in the proposed scheme. Thus, the transmitted signal $\mathbf{X}_R$ of the relay is just the same as the received signal $\mathbf{Y}_R$.

Then, we consider the situation in user A_m as a case. In A_m, we have the observer as

$$\mathbf{Y}_{A_m} = \mathbf{G}_{A_m} \mathbf{X}_R + \mathbf{Z}_{A_m}, \tag{3}$$

where $\mathbf{Z}_{A_m}$ is the zero mean circularly symmetric complex Gaussian noise vector at the user A_m modelled by $\mathbf{Z}_{A_m} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. The user A_m then decodes the partner's message $\mathbf{c}_{B_m}$ with the help of its self-information and detection vector.

2.2. Adversary Model. In this subsection, we will discuss the system security model for MIMO TWRC channels. We divide the system security model into two cases: untrusted relay attack model and trusted relay with eavesdropper attack model.

(1) Untrusted Relay Adversary Model. With a pessimistic consideration, we assume the relay itself is an untrusted node. Under such assumption, the relay acts as an eavesdropper to wiretap message from the communication pairs illegitimately. In order to exchange information, each user regulates its transmission rate to guarantee the successful transmission and to resist the untrusted relay attack. In doing so, we could obtain the achievable secrecy channel capacity C_s^{UR} as follows:

$$C_s^{\mathrm{UR}} = \left[\sum_{k=1}^{K} \left(R_{A_k}^{\mathrm{UR}} + R_{B_k}^{\mathrm{UR}} \right) - R_R^{\mathrm{UR}} \right]^+, \tag{4}$$

where $R_{A_k}^{\mathrm{UR}}$ and $R_{B_k}^{\mathrm{UR}}$ are the achievable maximum information rate from users A_k and B_k to their respective partners as

$$\begin{aligned} R_{A_k}^{\mathrm{UR}} &= \frac{1}{2} I\left(\mathbf{Y}_{B_k}; \mathbf{X}_{A_k} \mid \mathbf{Y}_R, \mathbf{X}_R \right), \\ R_{B_k}^{\mathrm{UR}} &= \frac{1}{2} I\left(\mathbf{Y}_{A_k}; \mathbf{X}_{B_k} \mid \mathbf{Y}_R, \mathbf{X}_R \right). \end{aligned} \tag{5}$$

And R_R^{UR} denotes the achievable information rate at the untrusted relay as

$$R_R^{\mathrm{UR}} = \frac{1}{2} I\left(\mathbf{Y}_R; \mathbf{X}_{A_1}, \mathbf{X}_{A_2}, \ldots, \mathbf{X}_{A_k}, \mathbf{X}_{B_1}, \ldots, \mathbf{X}_{B_k} \right). \tag{6}$$

Note here that the achievable secrecy channel capacity above is a general result independent of the transmission

scheme. One of our goals in this paper is to develop a novel scheme to achieve high capacity in an untrusted relay attack model scenario.

(2) Trusted Relay with Eavesdropper Adversary Model. In this subsection, we consider the situation where the communication pair exchange their information message via a trusted relay in the presence of the eavesdroppers. We assume there exists an eavesdropper E_m in between users A_m and B_m. In addition, the eavesdropper has complete knowledge of the channel information and transmission protocol. Furthermore, let the channel coefficient between user and eavesdropper be $\mathbf{H}^E_{A_m}$ and $\mathbf{H}^E_{B_m}$, respectively; then, the received signal by the eavesdropper $\mathbf{Y}_{E_m}$ is

$$\mathbf{Y}_{E_m} = \mathbf{H}^E_{A_m}\mathbf{X}_{A_m} + \mathbf{H}^E_{B_m}\mathbf{X}_{B_m} + \mathbf{Z}_{E_m}, \tag{7}$$

where $\mathbf{Z}_{E_m}$ is the noise at eavesdropper E_m. And upon receiving $\mathbf{Y}_{E_m}$, the eavesdropper tries to recover the information messages $\mathbf{c}_{A_m}$ and $\mathbf{c}_{B_m}$.

The MIMO wiretap channel introduced by [15] can be considered as multiple and parallel single subwiretap channels; each subchannel contains the communication user pair and the potential eavesdroppers. In doing so, we obtain the achievable secrecy channel capacity C_s^{TR} for trusted relay with eavesdropper attack model as

$$C_s^{\mathrm{TR}} = \left[\sum_{k=1}^{K}\left(R^{\mathrm{TR}}_{A_k} + R^{\mathrm{TR}}_{B_k} - R^{\mathrm{TR}}_{E_k}\right)\right]^+, \tag{8}$$

where $R^{\mathrm{TR}}_{A_k}$ and $R^{\mathrm{TR}}_{B_k}$ are the secrecy information rate between users A_k and B_k, respectively. They have identical analysis as (5):

$$\begin{aligned} R^{\mathrm{TR}}_{A_k} &= \frac{1}{2}\left[I\left(\mathbf{Y}_{B_k}; \mathbf{X}_{A_k} \mid \mathbf{Y}_R, \mathbf{X}_R\right)\right], \\ R^{\mathrm{TR}}_{B_k} &= \frac{1}{2}\left[I\left(\mathbf{Y}_{A_k}; \mathbf{X}_{B_k} \mid \mathbf{Y}_R, \mathbf{X}_R\right)\right]. \end{aligned} \tag{9}$$

And $R^{\mathrm{TR}}_{E_k}$ is the information rate at the eavesdropper as

$$R^{\mathrm{TR}}_{E_k} = \frac{1}{2}\left[I\left(\mathbf{Y}_{E_k}; \mathbf{X}_{A_k}, \mathbf{X}_{B_k}\right)\right]. \tag{10}$$

The secrecy channel capacity here is also a general result.

(3) Untrusted Relay with Eavesdropper Adversary Model. In the worst case, the relay itself is an unfriendly node; meanwhile, there exist a considerable number of eavesdroppers between each communication pair. We hold the same assumption as the above two subsections, and then we can obtain the achievable secrecy capacity C_s in such scenario as

$$C_s^{\mathrm{UE}} = \left[\sum_{k=1}^{K}\left(R^{\mathrm{UE}}_{A_k} + R^{\mathrm{UE}}_{B_k} - R^{\mathrm{UE}}_{E_k}\right) - R^{\mathrm{UE}}_{R}\right]^+, \tag{11}$$

where $R^{\mathrm{UE}}_{A_k}$ and $R^{\mathrm{UE}}_{B_k}$ are the secrecy information rate between users A_k and B_k, respectively, and they still have identical analysis as (5):

$$\begin{aligned} R^{\mathrm{UE}}_{A_k} &= \frac{1}{2}\left[I\left(\mathbf{Y}_{B_k}; \mathbf{X}_{A_k} \mid \mathbf{Y}_R, \mathbf{X}_R\right)\right], \\ R^{\mathrm{UE}}_{B_k} &= \frac{1}{2}\left[I\left(\mathbf{Y}_{A_k}; \mathbf{X}_{B_k} \mid \mathbf{Y}_R, \mathbf{X}_R\right)\right]. \end{aligned} \tag{12}$$

And $R^{\mathrm{UE}}_{E_k}$ and R^{UE}_{R} are the information rate at the eavesdropper and untrusted relay, respectively, as

$$\begin{aligned} R^{\mathrm{UE}}_{E_k} &= \frac{1}{2}\left[I\left(\mathbf{Y}_{E_k}; \mathbf{X}_{A_k}, \mathbf{X}_{B_k}\right)\right], \\ R^{\mathrm{UE}}_{R} &= \frac{1}{2} I\left(\mathbf{Y}_R; \mathbf{X}_{A_1}, \mathbf{X}_{A_2}, \ldots, \mathbf{X}_{A_k}, \mathbf{X}_{B_1}, \ldots, \mathbf{X}_{B_k}\right). \end{aligned} \tag{13}$$

We will discuss the capacity analysis of our proposed scheme in the following section.

3. Achievable Secrecy Transmission Scheme with Information Theoretical Approach

In this section, we will present an achievable secrecy transmission scheme using the information theoretical approach for the all three adversary models.

3.1. The Transmission Scheme Based on Direction Rotation Alignment. This scheme is composed of two transmission phases and one relay operation phase. The details are shown below.

(1) Multiple Access Phase. The information symbols $\mathbf{c}_{A_k}$(or $\mathbf{c}_{B_k}$) are modified by a precoding matrix prior to the transmission. The precoding matrix is used to construct equivalent parallel subchannels for different communication pairs. The scenario is identical on either user side A_k or B_k, so we only present the design for A_k. The precoding matrix on user A_k is denoted as $\mathbf{F}_{A_k}$. Then, the transmitted signal could be rewritten as $\mathbf{X}_{A_k} = \mathbf{F}_{A_k} \cdot \mathbf{c}_{A_k}$.

We now move on to investigate the design of precoding matrix $\mathbf{F}_{A_k}$. Using the singular value decomposition (SVD), the channel matrix $\mathbf{H}_{A_k}$ could be represented as follows:

$$\mathbf{H}_{A_k} = \mathbf{U}_{A_k}\mathbf{\Sigma}_{A_k}\mathbf{V}^H_{A_k}, \tag{14}$$

where $\mathbf{U}_{A_k}$ and $\mathbf{V}_{A_k}$ are unitary matrices and $\mathbf{\Sigma}_{A_k}$ is a diagonal matrix with positive diagonal elements. So, we define $\mathbf{F}_{A_k}$ in the following form:

$$\mathbf{F}_{A_k} = \mathbf{V}_{A_k}\mathbf{\Sigma}^{-1}_{A_k}\mathbf{U}^H_{A_k}\mathbf{R}\mathbf{\Psi}_{A_k}\mathbf{L}_k, \tag{15}$$

where $\mathbf{R}$ is an $n_R * n_R$ unitary matrix called *Direction Rotation matrix* and $\mathbf{\Sigma}^{-1}_{A_k}$ denotes the pseudoinverse of $\mathbf{\Sigma}_{A_k}$. $\mathbf{\Psi}_{A_k}$ represents the allocated transmission power for user A_k. In addition, we assume it is identical for all users in this work. $\mathbf{L}_k$ is called *channel allocation matrix* to guarantee that multipair users communicate simultaneously. $\mathbf{L}_k$ allocates the

constructed equivalent parallel subchannels to corresponding users. The design and optimization of $\mathbf{R}$ and $\mathbf{L}_k$ will affect the transmission efficiency of the proposed scheme. Then, we will introduce the design of $\mathbf{R}$ and $\mathbf{L}_k$, respectively.

Design of Direction Rotation Matrix. The authors in [5] have given an approximate design of the Direction Rotation matrix for one-pair TWRC channels. Along with their work, we extend their results into multipair TWRC channels.

Before designing $\mathbf{R}$, we first define a new matrix $\mathbf{G}$ as

$$\mathbf{G} \triangleq \sum_{k=1}^{K} \left\{ \mathbf{U}_{A_k} \boldsymbol{\Sigma}_{A_k}^{-2} \mathbf{U}_{A_k}^{H} + \mathbf{U}_{B_k} \boldsymbol{\Sigma}_{B_k}^{-2} \mathbf{U}_{B_k}^{H} \right\}. \tag{16}$$

Using eigendecomposition, $\mathbf{G}$ could be represented as

$$\mathbf{G} = \mathbf{U}_g \boldsymbol{\Lambda}_g \mathbf{U}_g^{H}, \tag{17}$$

where $\mathbf{U}_g$ is a unitary matrix and $\boldsymbol{\Lambda}_g$ is a diagonal matrix whose nonzero elements are the eigenvalues of $\mathbf{G}$ arranged in the ascending order. Then, we choose $\mathbf{R}$ as

$$\mathbf{R} = \mathbf{U}_g. \tag{18}$$

Design of Channel Allocation Matrix. Before designing $\mathbf{L}_k$, we first define two varieties d_k and D_k. d_k is the independent signal streams of the kth communication pair, and D_k is the sum independent signal streams of the 1st communication pair to the kth pair as $D_k = \sum_{i=1}^{k-1} d_i$. In order to align the signals from the same communication pair to the same directions, we define $\mathbf{L}_{A_k} = \mathbf{L}_{B_k} = \mathbf{L}_k$. And $\mathbf{L}_k$ is designed by collecting the $D_k + 1$ to $D_k + d_k$ column vector from the $n_R * n_R$ unit matrix, so $\mathbf{L}_k$ is in the size of $n_R * d_k$. $\mathbf{L}_k$ is shown as follows:

$$\mathbf{L}_k = \begin{bmatrix} & \overbrace{\quad}^{d_k} \\ D_k \begin{cases} \overbrace{00\cdots0} \\ 00\cdots0 \\ \vdots \\ 00\cdots0 \end{cases} \\ d_k \begin{cases} 10\cdots0 \\ 01\cdots0 \\ \vdots \\ 00\cdots1 \end{cases} \\ 00\cdots0 \\ \vdots \\ 00\cdots0 \end{bmatrix}. \tag{19}$$

(2) Relay's Operation. After the precoding operation, the users transmit their coded signal to the relay. Then, the received signal in relay is

$$\mathbf{Y}_R = \sum_{k=1}^{K} \left(\mathbf{H}_{A_k} \mathbf{F}_{A_k} \mathbf{c}_{A_k} + \mathbf{H}_{B_k} \mathbf{F}_{B_k} \mathbf{c}_{B_k} \right) + \mathbf{Z}_R. \tag{20}$$

For $k \in \{A_k, B_k\}$, we have

$$\mathbf{H}_k \mathbf{F}_k = \mathbf{U}_k \Sigma_k \mathbf{V}_k^{H} \mathbf{V}_k \boldsymbol{\Sigma}_k^{-1} \mathbf{U}_k^{H} \mathbf{R} \mathbf{L}_k = \mathbf{R} \mathbf{L}_k, \tag{21}$$

so the received signal $\mathbf{Y}_R$ could be rewritten as

$$\begin{aligned} \mathbf{Y}_R &= \mathbf{R} \sum_{k=1}^{K} \mathbf{L}_k \left(\mathbf{c}_{A_k} + \mathbf{c}_{B_k} \right) + \mathbf{Z}_R \\ &= \mathbf{R} \begin{bmatrix} \mathbf{c}_{A_1} + \mathbf{c}_{B_1} \\ \mathbf{c}_{A_2} + \mathbf{c}_{B_2} \\ \vdots \\ \mathbf{c}_{A_K} + \mathbf{c}_{B_K} \end{bmatrix} + \mathbf{Z}_R. \end{aligned} \tag{22}$$

As shown in (22), the signals from the same communication pairs have been aligned into the same directions.

Upon receiving $\mathbf{Y}_R$, the relay generates its transmitting signal $\mathbf{X}_R$. In the Amplify-and-Forward (AF) agreements, the relationship between $\mathbf{X}_R$ and $\mathbf{Y}_R$ is simply written as $\mathbf{X}_R = \mathbf{Y}_R$.

(3) Broadcast Phase. In the BC phase, each user receives the broadcast signal from the relay. By considering the received signal at user B_m, we can obtain

$$\begin{aligned} \mathbf{Y}_{B_m} &= \mathbf{G}_{B_m} \mathbf{X}_R + \mathbf{Z}_{B_m} \\ &= \mathbf{G}_{B_m} \sum_{k=1}^{K} \left(\mathbf{H}_{A_k} \mathbf{F}_{A_k} \mathbf{c}_{A_k} + \mathbf{H}_{B_k} \mathbf{F}_{B_k} \mathbf{c}_{B_k} \right) + \mathbf{G}_{B_m} \mathbf{Z}_R \\ &\quad + \mathbf{Z}_{B_k}. \end{aligned} \tag{23}$$

Subtracting the self-interference and the cochannel interference from other communication pairs, the user B_m could obtain the equivalent receiving signal $\widetilde{\mathbf{Y}}_{B_k}$ as follows:

$$\widetilde{\mathbf{Y}}_{B_m} = \mathbf{G}_{B_m} \mathbf{H}_{A_m} \mathbf{F}_{A_m} \mathbf{c}_{A_m} + \mathbf{G}_{B_m} \mathbf{Z}_R + \mathbf{Z}_{A_m}. \tag{24}$$

The interference cancellation will be accomplished by the detection vector $\mathbf{D}_{A_m}$. $\mathbf{D}_{A_m}$ is dependent on $\mathbf{G}_{A_m}$. Consider the case where $\mathbf{G}_{A_m}$ is simply the transpose of $\mathbf{H}_{A_m}$. Then, $\mathbf{D}_{B_m}$ is designed such that $\mathbf{D}_{B_m} = \mathbf{L}_m^{H} \mathbf{R}^{H} \mathbf{U}_{B_m} (\Sigma_{B_m}^{H})^{-1} \mathbf{V}_{B_m}^{H}$. Consequently, the received signal will be

$$\begin{aligned} \widetilde{\mathbf{Y}}_{B_m} &= \mathbf{D}_{B_m} \mathbf{Y}_{B_m} = \mathbf{D}_{B_m} \left(\mathbf{H}_{B_m}^{H} \mathbf{X}_R + \mathbf{Z}_{B_m} \right) \\ &= \mathbf{L}_m^{H} \mathbf{R}^{H} \mathbf{U}_{B_m} \left(\Sigma_{B_m}^{H} \right)^{-1} \mathbf{V}_{B_m}^{H} \mathbf{V}_{B_m} \Sigma_{B_m}^{H} \mathbf{U}_{B_m}^{H} \mathbf{R} \\ &\quad \cdot \sum_{k=1}^{K} \mathbf{L}_m \left(\mathbf{c}_{A_m} + \mathbf{c}_{B_m} \right) + \mathbf{D}_{B_m} \mathbf{Z}_R + \mathbf{Z}_{B_m} \\ &= \left(\mathbf{c}_{A_m} + \mathbf{c}_{B_m} \right) + \mathbf{Z}'_{B_m}, \end{aligned} \tag{25}$$

where $\mathbf{Z}'_{B_m}$ denotes the sum of noises in user B_m.

From (25), we can clearly see that the user B_m could recover the information symbols $\mathbf{c}_{A_m}$ by the XOR operation with its self-information symbols $\mathbf{c}_{B_m}$. Exactly alike, the user A_m could recover the information symbols from B_m in the same way. Thus, one round of information exchange is completed.

3.2. The Achievable Secrecy Channel Capacity Analysis for Untrusted Relay Model. In this subsection, we discuss the achievable secrecy channel capacity of our proposed scheme for the untrusted relay model based on the analysis given in the previous section. From (4), we can see that the capacity is affected by R_R^{UR} and $\{R_{A_1}^{\mathrm{UR}}, R_{A_2}^{\mathrm{UR}}, \ldots, R_{A_K}^{\mathrm{UR}}, R_{B_1}^{\mathrm{UR}}, \ldots, R_{A_K}^{\mathrm{UR}}\}$. We now discuss the impact of these components and obtain the achievable secrecy channel capacity C_s^{UR}.

After one round of transmission, the received equivalent information at user B_m is shown in (25). The information rate from A_m to B_m is

$$\begin{aligned} R_{A_m}^{\mathrm{UR}} &= \frac{1}{2}\log\det\left(\mathbf{I} + \mathbf{F}_{A_m}^H \mathbf{H}_{A_m}^H \mathbf{G}_{B_m}^H \mathbf{K}_{A_m}^{-1} \mathbf{G}_{B_m} \mathbf{H}_{A_m} \mathbf{F}_{A_m}\right) \\ &= \frac{1}{2}\log\det\left(\mathbf{I} + \mathbf{L}_m^H \mathbf{R} \mathbf{G}_{B_m}^H \mathbf{K}_{A_m}^{-1} \mathbf{G}_{B_m} \mathbf{R} \mathbf{L}_m\right), \end{aligned} \tag{26}$$

where $\mathbf{K}_{A_m} = \mathbf{G}_{B_m}\mathbf{G}_{B_m}^H + \mathbf{I}$.

Similarly, the information rate from B_m to A_m is

$$R_{B_m}^{\mathrm{UR}} = \frac{1}{2}\log\det\left(\mathbf{I} + \mathbf{L}_m^H \mathbf{R} \mathbf{G}_{A_m}^H \mathbf{K}_{B_m}^{-1} \mathbf{G}_{A_m} \mathbf{R} \mathbf{L}_m\right). \tag{27}$$

For the untrusted relay model, the adversary tries to recover the message symbols from all the users node. So, the achievable information rate is equal to the maximum sum rate of the uplink multiuser MAC channel:

$$\begin{aligned} R_R^{\mathrm{UR}} = \frac{1}{2}\log\det\Bigg[&\mathbf{I} \\ &+ \sum_{k=1}^{K}\left(\mathbf{H}_{A_k}\mathbf{F}_{A_k}\mathbf{F}_{A_k}^H\mathbf{H}_{A_k}^H + \mathbf{H}_{B_k}\mathbf{F}_{B_k}\mathbf{F}_{B_k}^H\mathbf{H}_{B_k}^H\right)\Bigg]. \end{aligned} \tag{28}$$

From (26), (27), and (28), the achievable secrecy channel capacity for the untrusted relay model can be obtained with matrix operation [28] as follows:

$$C_s^{\mathrm{UR}} = \frac{1}{2}\log\det\left[\frac{\prod_{k=1}^{K}\left(\mathbf{I} + \mathbf{L}_k^H \mathbf{R} \mathbf{G}_{B_k}^H \mathbf{K}_{A_k}^{-1} \mathbf{G}_{B_k} \mathbf{R} \mathbf{L}_k\right)\left(\mathbf{I} + \mathbf{L}_k^H \mathbf{R} \mathbf{G}_{A_k}^H \mathbf{K}_{B_k}^{-1} \mathbf{G}_{A_k} \mathbf{R} \mathbf{L}_k\right)}{\mathbf{I} + \sum_{k=1}^{K}\left(\mathbf{H}_{A_k}\mathbf{F}_{A_k}\mathbf{F}_{A_k}^H\mathbf{H}_{A_k}^H + \mathbf{H}_{B_k}\mathbf{F}_{B_k}\mathbf{F}_{B_k}^H\mathbf{H}_{B_k}^H\right)}\right]. \tag{29}$$

3.3. The Achievable Secrecy Channel Capacity Analysis for Trusted Relay with Eavesdropper Model. In this subsection, we will discuss the achievable secrecy channel capacity of our proposed scheme for the trusted relay with eavesdropper model. Before discussing the capacity, we first consider the received signal by eavesdropper E_m between the users A_m and B_m.

The received signal in general case is shown in (7), and for the proposed scheme we have

$$\mathbf{Y}_{E_m} = \mathbf{H}_{A_m}^E \mathbf{F}_{A_m} \mathbf{c}_{A_m} + \mathbf{H}_{B_m}^E \mathbf{F}_{B_m} \mathbf{c}_{B_m} + \mathbf{Z}_{E_m}, \tag{30}$$

where $\mathbf{Z}_{E_m}$ is noise at eavesdropper E_m. Consequently, we have the achievable information rate as

$$\begin{aligned} R_{E_m}^{\mathrm{TR}} = \frac{1}{2}\log\det\Big[&\mathbf{I} + \mathbf{H}_{A_m}^E \mathbf{F}_{A_m} \mathbf{F}_{A_m}^H \left(\mathbf{H}_{A_m}^E\right)^H \\ &+ \mathbf{H}_{B_m}^E \mathbf{F}_{B_m} \mathbf{F}_{B_m}^H \left(\mathbf{H}_{B_m}^E\right)^H\Big]. \end{aligned} \tag{31}$$

Meanwhile, the information rates $R_{A_m}^{\mathrm{TR}}$ and $R_{B_m}^{\mathrm{TR}}$ are identical to the untrusted relay model; namely,

$$\begin{aligned} R_{A_m}^{\mathrm{TR}} &= \frac{1}{2}\log\det\left(\mathbf{I} + \mathbf{L}_m^H \mathbf{R} \mathbf{G}_{B_m}^H \mathbf{K}_{A_m}^{-1} \mathbf{G}_{B_m} \mathbf{R} \mathbf{L}_m\right), \\ R_{B_m}^{\mathrm{TR}} &= \frac{1}{2}\log\det\left(\mathbf{I} + \mathbf{L}_m^H \mathbf{R} \mathbf{G}_{A_m}^H \mathbf{K}_{B_m}^{-1} \mathbf{G}_{A_m} \mathbf{R} \mathbf{L}_m\right). \end{aligned} \tag{32}$$

As a result, we obtain the achievable secrecy channel capacity as

$$C_s^{\mathrm{TR}} = \frac{1}{2}\log\det\left[\sum_{k=1}^{K}\frac{\left(\mathbf{I} + \mathbf{L}_k^H \mathbf{R} \mathbf{G}_{B_k}^H \mathbf{K}_{A_k}^{-1} \mathbf{G}_{B_k} \mathbf{R} \mathbf{L}_k\right)\left(\mathbf{I} + \mathbf{L}_k^H \mathbf{R} \mathbf{G}_{A_k}^H \mathbf{K}_{B_k}^{-1} \mathbf{G}_{A_k} \mathbf{R} \mathbf{L}_k\right)}{\mathbf{I} + \mathbf{H}_{A_k}^E \mathbf{F}_{A_k} \mathbf{F}_{A_k}^H \left(\mathbf{H}_{A_k}^E\right)^H + \mathbf{H}_{B_k}^E \mathbf{F}_{B_k} \mathbf{F}_{B_k}^H \left(\mathbf{H}_{B_k}^E\right)^H}\right]. \tag{33}$$

3.4. The Secrecy Channel Capacity Analysis for Untrusted Relay with Eavesdropper Model. We will discuss the secrecy channel capacity for the worst case: the relay itself is an unfriendly node; meanwhile, there exist a considerable number of eavesdroppers between each communication pair. The situation under this case is just like a combination of the former two cases, and the general capacity analysis of this model is given by (11). So, we first give all the components based on the former two subsections and then give the achievable secrecy channel capacity.

The information rates $R_{A_m}^{\mathrm{UE}}$ and $R_{B_m}^{\mathrm{UE}}$ are given as

$$R_{A_m}^{\mathrm{UE}} = \frac{1}{2}\log\det\left(\mathbf{I} + \mathbf{L}_m^H \mathbf{R}\mathbf{G}_{B_m}^H \mathbf{K}_{A_m}^{-1}\mathbf{G}_{B_m}\mathbf{R}\mathbf{L}_m\right),$$
$$R_{B_m}^{\mathrm{UE}} = \frac{1}{2}\log\det\left(\mathbf{I} + \mathbf{L}_m^H \mathbf{R}\mathbf{G}_{A_m}^H \mathbf{K}_{B_m}^{-1}\mathbf{G}_{A_m}\mathbf{R}\mathbf{L}_m\right). \tag{34}$$

And the achievable information rate at the untrusted relay is given as

$$R_R^{\mathrm{UE}} = \frac{1}{2}\log\det\left[\mathbf{I} + \sum_{k=1}^{K}\left(\mathbf{H}_{A_k}\mathbf{F}_{A_k}\mathbf{F}_{A_k}^H\mathbf{H}_{A_k}^H + \mathbf{H}_{B_k}\mathbf{F}_{B_k}\mathbf{F}_{B_k}^H\mathbf{H}_{B_k}^H\right)\right]. \tag{35}$$

And the achievable information rate at eavesdropper is given as

$$R_{E_m}^{\mathrm{UE}} = \frac{1}{2}\log\det\left[\mathbf{I} + \mathbf{H}_{A_m}^E\mathbf{F}_{A_m}\mathbf{F}_{A_m}^H\left(\mathbf{H}_{A_m}^E\right)^H + \mathbf{H}_{B_m}^E\mathbf{F}_{B_m}\mathbf{F}_{B_m}^H\left(\mathbf{H}_{B_m}^E\right)^H\right]. \tag{36}$$

With (34), (35), and (36), we obtain the achievable secrecy channel capacity as

$$C_s^{\mathrm{UE}} = \frac{1}{2}\cdot\log\det\left\{\frac{\sum_{k=1}^{K}\left(\left(\mathbf{I} + \mathbf{L}_k^H\mathbf{R}\mathbf{G}_{B_k}^H\mathbf{K}_{A_k}^{-1}\mathbf{G}_{B_k}\mathbf{R}\mathbf{L}_k\right)\left(\mathbf{I} + \mathbf{L}_k^H\mathbf{R}\mathbf{G}_{A_k}^H\mathbf{K}_{B_k}^{-1}\mathbf{G}_{A_k}\mathbf{R}\mathbf{L}_k\right)/\left(\mathbf{I} + \mathbf{H}_{A_k}^E\mathbf{F}_{A_k}\mathbf{F}_{A_k}^H\left(\mathbf{H}_{A_k}^E\right)^H + \mathbf{H}_{B_k}^E\mathbf{F}_{B_k}\mathbf{F}_{B_k}^H\left(\mathbf{H}_{B_k}^E\right)^H\right)\right)}{\mathbf{I} + \sum_{k=1}^{K}\left(\mathbf{H}_{A_k}\mathbf{F}_{A_k}\mathbf{F}_{A_k}^H\mathbf{H}_{A_k}^H + \mathbf{H}_{B_k}\mathbf{F}_{B_k}\mathbf{F}_{B_k}^H\mathbf{H}_{B_k}^H\right)}\right\}. \tag{37}$$

Note here that C_s^{UE} has great probability equal to zero, and it is a common case independent of transmission scheme. So, immediately we have this question: how to optimize the proposed scheme to secrecy transmission even under the worst case. For this reason, we optimize the precoding nested physical layer encryption. The design details will be presented in the following section.

4. Secrecy Transmission Scheme with Physical Layer Encryption Approach

In the above section, we have discussed the secrecy transmission scheme based information theoretical analysis. However, the traditional information theoretical approach achieves secrecy by sacrificing the transmission efficiency. And all the schemes including our proposed information theoretical approach are not constantly valid for the worst case.

Under this motivation, we design an encryption vector nested into the precoding matrix to accomplish the message encryption in the physical layer. With this physical layer encryption, the system security will only depend on the security of the shared secret key rather than the mutual information in eavesdropper or untrusted relay. So, the channel could achieve full capacity instead of part of it. In this section, we will first present the physical layer encryption scheme in MIMO TWRC channels and then present the capacity analysis for physical layer encryption.

4.1. Physical Layer Encryption Scheme. In order to accomplish the encryption, we redesign the precoding matrix as $\mathbf{P}_{A_m}$ and $\mathbf{P}_{B_m}$ for users A_m and B_m. We consider the situation in A_m as a case. Containing two function parts, $\mathbf{P}_{A_m}$ can be artificially divided into two parts as follows:

$$\mathbf{P}_{A_m} = \mathbf{F}_{A_m}\mathbf{S}_{A_m}, \tag{38}$$

where $\mathbf{F}_{A_m}$ is designed same as the above discussion and $\mathbf{S}_{A_m}$ is designed for physical layer encryption. Note that the precoding matrix has been artificially separated into two matrices, that is, transmission matrix and security matrix; however, the precoding matrix will show the effect as a whole.

The encryption precoding matrix $\mathbf{S}_{A_m}$ or $\mathbf{S}_{B_m}$ is generated from a key stream $\mathbf{s}_m$ where $\mathbf{s}_m(i) \in \{1, -1\}$. The generation of $\mathbf{S}_{A_m}$ and $\mathbf{S}_{B_m}$ will depend on the security level of the system. For low security level, $\mathbf{S}_{A_m}$ and $\mathbf{S}_{B_m}$ will just be equal to $\mathbf{s}_m$ as $\mathbf{S}_{A_m} = \mathbf{S}_{B_m} = \mathbf{s}_m$. For high level security, $\mathbf{S}_{A_m}$ and $\mathbf{S}_{B_m}$ will be different by dividing $\mathbf{s}_m$ into two parts. Note here that the key stream must be preshared between users A_m and B_m before the transmission, and $\mathbf{s}_m$ is produced by pseudorandom sequence generators (PRSG).

With $\mathbf{S}_{A_m}$ and $\mathbf{S}_{B_m}$, each user could encrypt its information symbols bit by bit. And, in the next part, we will explore how the proposed encryption scheme promotes the security performance under both untrusted and trusted models.

4.2. Attack Analysis for Physical Layer Encryption Scheme. We first consider the untrusted relay model. Because the untrusted relay could be viewed as the most powerful eavesdropper, if the proposed scheme is secure under untrusted relay adversary model, it will certainly be secure under all adversary models.

We now begin to discuss the attack analysis under untrusted relay adversary model. With the new precoding matrix, the received signal in relay now will be

$$\mathbf{Y}_R^{\mathrm{Enc}} = \mathbf{R}\sum_{k=1}^{K}\mathbf{A}_k\left(\mathbf{S}_{A_k}\mathbf{c}_{A_k} + \mathbf{S}_{B_k}\mathbf{c}_{B_k}\right) + \mathbf{Z}_R = \mathbf{R}\begin{bmatrix}\mathbf{S}_{A_1}\mathbf{c}_{A_1} + \mathbf{S}_{B_1}\mathbf{c}_{B_1}\\ \mathbf{S}_{A_2}\mathbf{c}_{A_2} + \mathbf{S}_{B_2}\mathbf{c}_{B_2}\\ \vdots\\ \mathbf{S}_{A_K}\mathbf{c}_{A_K} + \mathbf{S}_{B_K}\mathbf{c}_{B_K}\end{bmatrix} + \mathbf{Z}_R. \tag{39}$$

Table 1: BPSK data patterns of user transmitting signals and relay receiving signal for physical layer encryption approach.

X_{A_k}	S_{A_k}	X_{B_k}	S_{B_k}	Y_R^{Enc}
1	1	1	1	2
1	1	1	−1	0
1	1	−1	1	0
1	1	−1	−1	2
1	−1	1	1	0
1	−1	1	−1	−2
1	−1	−1	1	−2
1	−1	−1	−1	0
−1	1	1	1	0
−1	1	1	−1	−2
−1	1	−1	1	−2
−1	1	−1	−1	0
−1	−1	1	1	2
−1	−1	1	−1	0
−1	−1	1	−1	0
−1	−1	−1	−1	2

Table 2: BPSK data patterns of user transmitting signals and relay receiving signal for information theoretical approach.

X_{A_k}	X_{B_k}	Y_R
1	1	2
1	−1	0
−1	1	0
−1	−1	−2

By comparison, we also consider the relay receiving signal for information theoretical approach as (22). Considering BPSK modulation as a case, we assume the untrusted relay could get the Direction Rotation matrix **R**. Then, we obtain the data patterns for the two different approaches as shown in Tables 1 and 2.

With Table 2, we can see that the untrusted relay could directly recover some characteristic bit like all 1 or all 0. For this reason, the difficulty degree of message recovery for untrusted relay is significantly reduced. However, as in Table 1, the characteristics will be distorted by the encryption. For example, if $Y_R = 2$, the relay could easily recover the transmitting message pair as $X_{A_k} = 1$ and $X_{B_k} = 1$. However, if $Y_R^{\text{Enc}} = 2$, the transmitting message pair could be all the four cases. So, with the physical layer encryption, the untrusted relay will have no better way rather than guessing each bit of the messages or the keys.

The attack analysis is identical for the eavesdropper, so we will have no specific explanation. With this analysis, we conclude that our proposed physical layer encryption scheme is secure under all three adversary models.

4.3. The Achievable Secrecy Channel Capacity Analysis for Physical Layer Encryption Scheme. In this subsection, we will investigate the achievable secrecy channel capacity for physical layer encryption scheme. The capacity is presented by the following theorem.

Theorem 1. *With physical layer encryption, the achievable secrecy channel capacity for MIMO TWRC channels is given by*

$$C_{\text{Enc}} = \left[\sum_{k=1}^{K}\left(R_{A_k} + R_{B_k}\right)\right]^{+}, \tag{40}$$

where R_{A_k} and R_{B_k} are the secrecy information rate between users A_k and B_k.

The proof of Theorem 1 is given in the Appendix.

By comparing (40) with (29), (33), and (37), we can have the following result: because the *log function* is an increasing function, the proposed physical layer encryption scheme evidently increases the sum capacity of the system. However, due to the complexity of key preshare, there will be an apparent trade-off between transmission performance and key preshare complexity. Depending on different performance requirement, the user could choose physical layer encryption approach with better transmission performance or information theoretical approach with lower system complexity.

5. Simulation Results

In this section, we present three simulations for our proposed scheme using MATLAB. First, we show that our proposed scheme outperforms some of the well-known existing schemes in terms of transmission quality. In the second simulation, we, respectively, show that our proposed scheme has good security by comparing the bit error rate (BER) between the intended receiver and the untrusted relay and the BER between the receiver and the eavesdropper. In the last simulation, we show the capacity of our proposed scheme under the three different adversary models. We assume that four pairs of users communicate at the same time via a relay, and the users, relay, and eavesdroppers are all equipped with four antennas; that is, $n_R = n_T = n_E = 4$. Note that all of these results are obtained by averaging over 10,000 realizations.

5.1. Transmission Performance Evaluation between Different Schemes in MIMO TWRC Channels. To test the transmission performance of our proposed scheme, we first compare our scheme with some existing well-known schemes in MIMO TWRC channels like Zero-Forcing (ZF), Minimum Mean Square Error (MMSE), and Maximum Likelihood (ML). From the simulation results in Figure 2, we can clearly see that the proposed scheme can achieve a better BER performance especially under low SNR condition. This is because the proposed scheme can effectively avoid the power loss caused by the direction alignment.

5.2. Security Performance Evaluation of Information Theoretical Approach and Physical Layer Encryption Approach. We then test the security performance of our proposed scheme by comparing the BER of the intended receiver, the untrusted

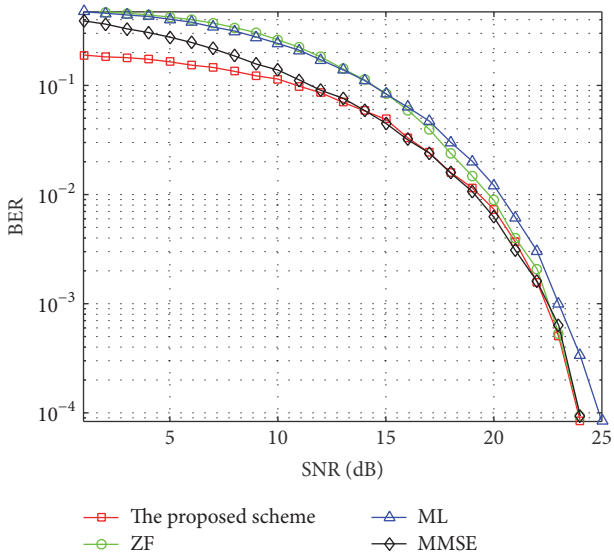


Figure 2: Transmission performance comparison between the proposed scheme and existing schemes.

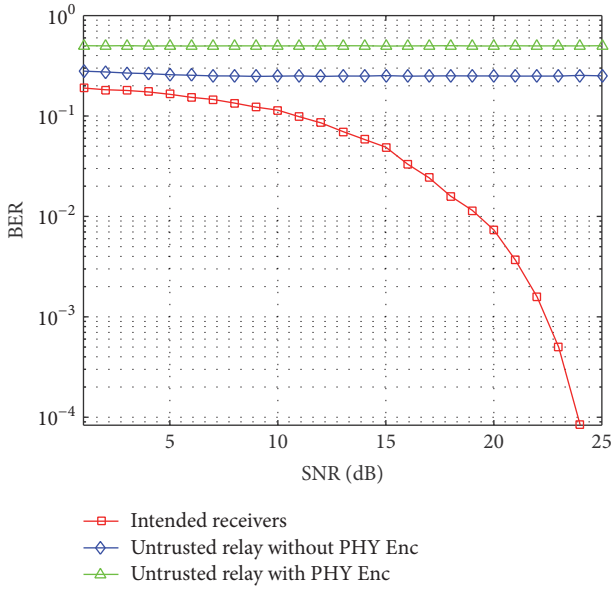


Figure 3: Security performance comparison between the intended receiver and the untrusted relay.

relay, and the eavesdropper. The BER comparison between the intended receiver and the untrusted relay is shown in Figure 3. We assume a stronger adversarial model where the relay can obtain all channel information including the Direction Rotation matrix R. From Figure 3, we can see that the BER at the intended receiver will decrease with SNR by a large proportion; however, the BER at the untrusted relay will stay in the high magnitude constantly. Meanwhile, we can also see that the proposed physical layer encryption will reduce the correct decoding probability at the untrusted relay.

Similar to the untrusted relay adversary model, we compare the BER between the intended receiver and the

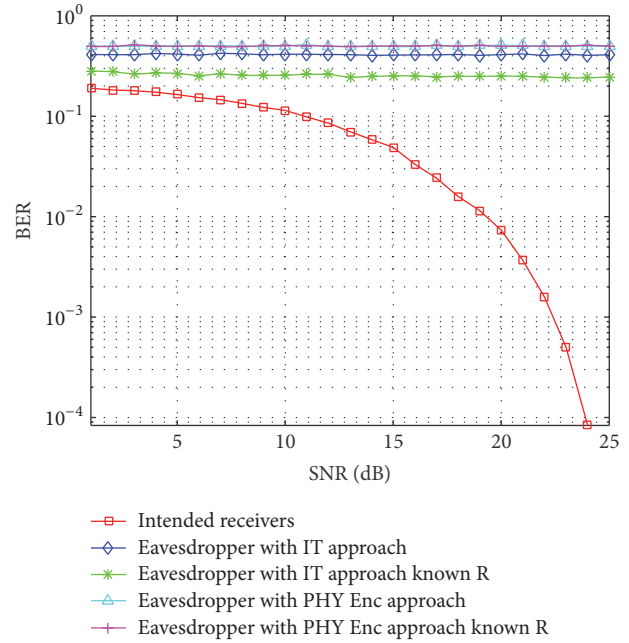


Figure 4: Security performance comparison between the intended receiver and the eavesdropper.

eavesdropper. The results are shown in Figure 4. To test different adversarial levels, we classify four cases to simulate: information theoretical approach with known Direction Rotation matrix, information theoretical approach without known Direction Rotation matrix, physical layer encryption approach with known Direction Rotation matrix, and physical layer encryption without known Direction Rotation matrix. From Figure 4, we can see that the eavesdropper gives the strongest attack under the first case: information theoretical approach with known Direction Rotation matrix. And if the eavesdropper fails to get the Direction Rotation matrix, the eavesdropper will suffer a worse decoding error ratio which is shown as the second case. However, the security performance of information theoretical approach is not as good as the physical layer encryption approach. From the results of the third and fourth cases, we can see that the decoding error bit ratio of these two cases is almost the same which is identical to the error ratio of random guessing.

5.3. Secure Channel Capacity Analysis Evaluation under Different Adversary Models. In this subsection, we will give the secure channel capacity analysis under the three different adversary models. The results are shown in Figures 5–7.

We compare the secure capacities of information theoretical approach and physical layer encryption approach under untrusted relay adversary model in Figure 5. From the comparison result, we can see that, with the SNR increasing, the sum rates of the physical layer encryption approach are almost twice the sum rates of the information theoretical approach. And the simulation result is in accord with (29) and (40).

We then compare the secure capacities of information theoretical approach and physical layer encryption approach

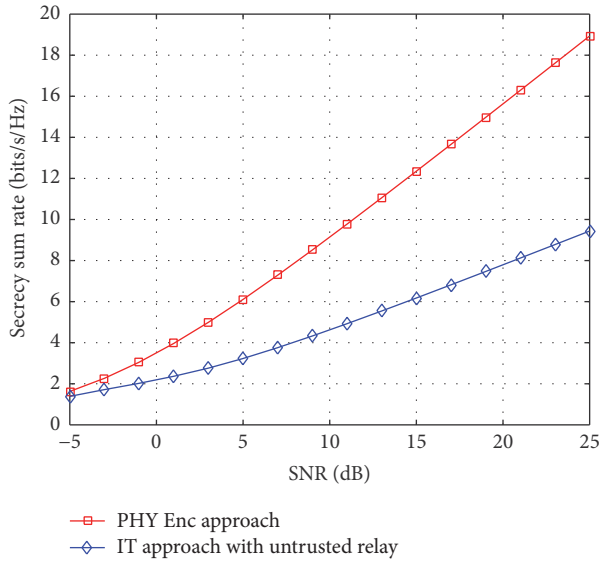


Figure 5: Capacity comparison between physical layer encryption approach and information theoretical approach under untrusted relay adversary model.

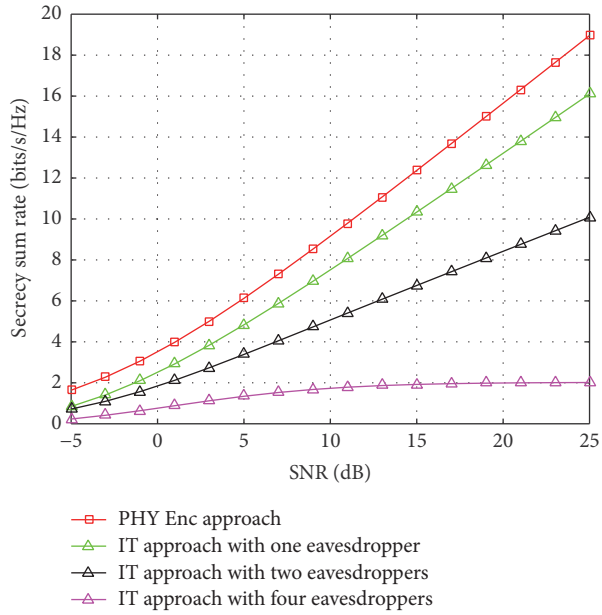


Figure 6: Capacity comparison between physical layer encryption approach and information theoretical approach under trusted relay with eavesdroppers model.

under trusted relay with eavesdroppers adversary model in Figure 6. We test one eavesdropper, two eavesdroppers, and four eavesdroppers cases. And from Figure 6 we can see that the capacity of information theoretical approach suffers a remarkable decline with the increasing of the eavesdroppers.

The secure capacities comparison between information theoretical approach and physical layer encryption approach under untrusted relay with eavesdroppers adversary model is shown in Figure 7. From Figure 7, we can see that the capacity

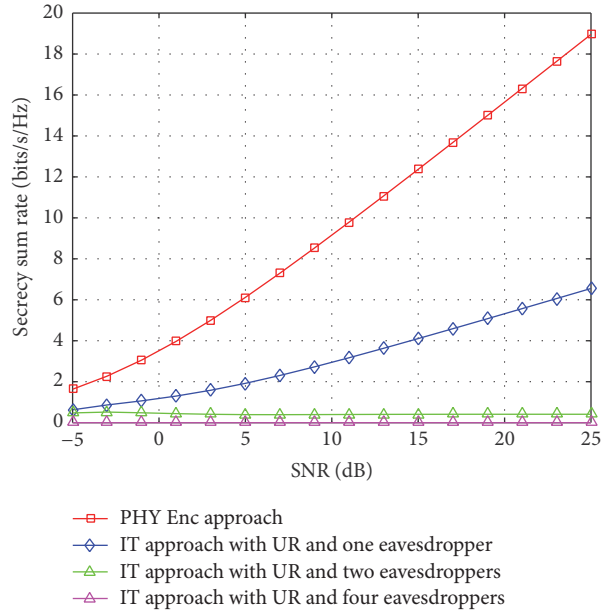


Figure 7: Capacity comparison between physical layer encryption approach and information theoretical approach under untrusted relay with eavesdroppers model.

of physical layer encryption approach shows no change with the increase of the eavesdroppers. However, the information theoretical approach can hardly resist the attack under such scenario, for the capacity has enormous probability equal to zero when there exist more than two eavesdroppers.

6. Conclusion

The security is one of the most important issues for 5G system architecture. Besides designing the security protocol from the view of system level, it is also desirable to consider the secure transmission for some basic transmission channel models, especially relevant to the key techniques for 5G.

In this paper, a novel transmission scheme has been introduced for MIMO TWRC channels, which is the basic channel model in HetNet and D2D networks. We consider three general attack models: untrusted relay adversarial model, trusted relay with eavesdropper adversarial model, and untrusted relay with eavesdropper adversarial model. Two different approaches, that is, information theoretical approach and physical layer encryption approach, have been proposed to achieve transmission efficiency as well as computational security. The key techniques of the proposed scheme lie in Direction Rotation Alignment and physical layer encryption. With alignment, signals from the same communication pair are aligned into the same signal direction. The direction rotation can avoid power loss in bad channel condition. And, with the physical layer encryption, the security of the system only depends on the security of the preshared key rather than the mutual information. Secrecy capacities of our proposed scheme are given for all the three models. Finally, simulation results show that our proposed scheme can achieve better performance in both transmission rate and security.

Comparing the two different approaches, we find that the physical layer encryption approach can get a better performance in transmission and security. However, such performance improvement is obtained by sacrificing the system complexity because the preshared key is needed. Thus, how to balance the trade-off between the system complexity and performance will be one of the future works. In addition, another future work lies in the modification of the proposed scheme fitting Massive MIMO which enjoys more implantation in 5G system.

Appendix

We consider the worst adversary model: untrusted relay with eavesdroppers model. If the physical layer encryption approach can achieve the proposed secrecy capacity like (40) in the worst case, it would certainly achieve the same capacity in the other two adversary models.

The achievable secrecy capacity under untrusted relay with eavesdroppers model has been shown in (11). By comparing (40) and (11), we can see that if we prove $R_R = 0$ and $R_{E_k} = 0$, we can prove (40). And R_R and R_{E_k} are

$$R_R = \frac{1}{2} I\left(\mathbf{Y}_R; \mathbf{X}_{A_1}, \mathbf{X}_{A_2}, \ldots, \mathbf{X}_{A_k}, \mathbf{X}_{B_1}, \ldots, \mathbf{X}_{B_k}\right), \tag{A.1}$$

$$R_{E_k} = \frac{1}{2}\left[I\left(\mathbf{Y}_{E_k}; \mathbf{X}_{A_k}, \mathbf{X}_{B_k}\right)\right]. \tag{A.2}$$

We now start to prove $R_R = 0$. With (A.1), we have

$$\begin{aligned} R_R &= \frac{1}{2} I\left(\mathbf{Y}_R; \mathbf{X}_{A_1}, \mathbf{X}_{A_2}, \ldots, \mathbf{X}_{A_k}, \mathbf{X}_{B_1}, \ldots, \mathbf{X}_{B_k}\right) \\ &\overset{(a)}{=} \frac{1}{2} I\left(\mathbf{X}_{A_1}, \mathbf{X}_{A_2}, \ldots, \mathbf{X}_{A_k}, \mathbf{X}_{B_1}, \ldots, \mathbf{X}_{B_k}; \mathbf{Y}_R\right) \\ &\overset{(b)}{=} \frac{1}{2} \\ &\cdot \sum_{k=1}^{K} \left[I\left(\mathbf{X}_{A_k}; \mathbf{Y}_R \mid \mathbf{X}_{A_{k-1}}, \ldots, \mathbf{X}_{A_1}, \mathbf{X}_{B_K}, \ldots, \mathbf{X}_{B_1}\right)\right. \\ &\left. + I\left(\mathbf{X}_{B_k}; \mathbf{Y}_R \mid \mathbf{X}_{B_{k-1}}, \ldots, \mathbf{X}_{B_1}\right)\right] \overset{(c)}{=} \frac{1}{2} \\ &\cdot \sum_{k=1}^{K} \left[I\left(\mathbf{X}_{A_k}; \mathbf{Y}_R\right) + I\left(\mathbf{X}_{B_k}; \mathbf{Y}_R\right)\right], \end{aligned} \tag{A.3}$$

where (a) is from the basic theorem that $I(A; B) = I(B; A)$, (b) is from the chain rule for mutual information, and (c) is from the fact that all the transmitting signals are independent.

With (A.3), we can see that if we could show that each mutual information part $I(\mathbf{X}_{A_k}; \mathbf{Y}_R)$ or $I(\mathbf{X}_{B_k}; \mathbf{Y}_R)$ is zero, the proposition will be permitted.

So, we move on to the proof of $I(\mathbf{X}_{A_k}; \mathbf{Y}_R) = 0$. We consider the BPSK modulation as a case. So, the transmitting signal in A_k and B_k will be 1 with probability 1/2 and −1 with probability 1/2. And the key streams S_{A_k} and S_{B_k} will also be 1 with probability 1/2 and −1 with probability 1/2.

We have shown the signal pattern for BPSK in Table 1. With Table 1, we can compute the probability distributions of Y_R as shown in Table 3.

Table 3: Probability distributions of Y_R.

Y_R	2	0	−2
p	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Table 4: Joint probability distributions of X_{A_k} and Y_R.

Y_R	X_{A_k}	
	1	−1
2	$\frac{1}{8}$	$\frac{1}{8}$
0	$\frac{1}{4}$	$\frac{1}{4}$
−2	$\frac{1}{8}$	$\frac{1}{8}$

Table 5: Conditional probability distribution between Y_R and X_{A_k}.

Y_R	2	0	−2
$P\left(Y_R \mid X_{A_k} = 1\right)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
$P\left(Y_R \mid X_{A_k} = -1\right)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

And we can also compute the joint probability distributions of X_{A_k} and Y_R as shown in Table 4.

So, we can get the conditional probability distribution between Y_R and X_{A_k} as shown in Table 5.

With Tables 3, 4, and 5, we can compute $H(Y_R)$ and $H(Y_R \mid X_{A_k})$ as

$$\begin{aligned} H\left(Y_R\right) &= H\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right) \\ &= \frac{1}{4}\log_2 4 + \frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 4 = \frac{3}{2} \text{ bit}, \\ H\left(Y_R \mid X_{A_k}\right) &= \sum_{x_{A_k} \in \{1,-1\}} p\left(x_{A_k}\right) H\left(Y_R \mid X_{A_k} = x_{A_k}\right) \\ &= \frac{1}{2} H\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right) + \frac{1}{2} H\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right) \\ &= H\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right) = \frac{3}{2} \text{ bit}. \end{aligned} \tag{A.4}$$

So, we can compute the mutual information $I(Y_R; X_{A_k})$ as

$$\begin{aligned} I\left(Y_R; X_{A_k}\right) &= H\left(Y_R\right) - H\left(Y_R \mid X_{A_k}\right) = \frac{3}{2} - \frac{3}{2} \\ &= 0 \text{ bits}. \end{aligned} \tag{A.5}$$

Exactly alike, the mutual information analysis is identical for the eavesdroppers model. So, we can get the same result where

$$I\left(Y_{E_k}; X_{A_k}\right) = 0. \tag{A.6}$$

With (11), (A.5), and (A.5), we can prove (40).

The analysis of other modulation models is identical to BPSK models, so we omit the details.

Competing Interests

The authors declare that they have no competing interests regarding the publication of this paper.

References

[1] R. Vaze and R. W. Heath, "Capacity scaling for MIMO two-way relaying," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '07)*, pp. 1451–1455, Nice, France, June 2007.

[2] W. Nam, S.-Y. Chung, and Y. H. Lee, "Capacity of the Gaussian two-way relay channel to within 1/2 bit," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5488–5494, 2010.

[3] R. Vaze and J. Heath, "On the capacity and diversity-multiplexing tradeoff of the two-way relay channel," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4219–4234, 2011.

[4] H. J. Yang, J. Chun, and A. Paulraj, "Asymptotic capacity of the separated MIMO two-way relay channel," *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7542–7554, 2011.

[5] T. Yang, X. Yuan, L. Ping, I. B. Collings, and J. Yuan, "A new physical-layer network coding scheme with eigen-direction alignment precoding for MIMO two-way relaying," *IEEE Transactions on Communications*, vol. 61, no. 3, pp. 973–986, 2013.

[6] Z. Fang, X. Yuan, and X. Wang, "Towards the asymptotic sum capacity of the MIMO cellular two-way relay channel," *IEEE Transactions on Signal Processing*, vol. 62, no. 16, pp. 4039–4051, 2014.

[7] G. Zheng, "Joint beamforming optimization and power control for full-duplex MIMO two-way relay channel," *IEEE Transactions on Signal Processing*, vol. 63, no. 3, pp. 555–566, 2015.

[8] Y. Dong, M. J. Hossain, and J. Cheng, "Performance of wireless powered amplify and forward relaying over nakagami-*m* fading channels with nonlinear energy harvester," *IEEE Communications Letters*, vol. 20, no. 4, pp. 672–675, 2016.

[9] A. D. Wyner, "The wire-tap channel," *The Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

[10] A. O. Hero, "Secure space-time communication," *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3235–3249, 2003.

[11] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas I: the MISOME wiretap channel," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3088–3104, 2010.

[12] P. Parada and R. Blahut, "Secrecy capacity of SIMO and slow fading channels," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '05)*, pp. 2152–2155, Adelaide, Australia, September 2005.

[13] A. Khisti, G. Wornell, A. Wiesel, and Y. Eldar, "On the Gaussian MIMO wiretap channel," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '07)*, pp. 2471–2475, Nice, France, June 2007.

[14] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas—part II: the MIMOME wiretap channel," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5515–5532, 2010.

[15] F. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4961–4972, 2011.

[16] T. Liu and S. Shamai, "A note on the secrecy capacity of the multiple-antenna wiretap channel," *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2547–2553, 2009.

[17] Y. Oohama, "Capacity theorems for relay channels with confidential messages," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '07)*, pp. 926–930, IEEE, June 2007.

[18] A. Mukherjee, S. A. A. Fakoorian, J. Huang, and A. L. Swindlehurst, "Principles of physical layer security in multiuser wireless networks: a survey," *IEEE Communications Surveys and Tutorials*, vol. 16, no. 3, pp. 1550–1573, 2014.

[19] C. Jeong, I.-M. Kim, and D. I. Kim, "Joint secure beamforming design at the source and the relay for an amplify-and-forward MIMO untrusted relay system," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 310–325, 2012.

[20] R. Zhang, L. Song, Z. Han, and B. Jiao, "Physical layer security for two-way untrusted relaying with friendly jammers," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 8, pp. 3693–3704, 2012.

[21] J. Mo, M. Tao, Y. Liu, and R. Wang, "Secure beamforming for MIMO two-way communications with an untrusted relay," *IEEE Transactions on Signal Processing*, vol. 62, no. 9, pp. 2185–2199, 2014.

[22] Z. Ding, M. Peng, and H.-H. Chen, "A general relaying transmission protocol for MIMO secrecy communications," *IEEE Transactions on Communications*, vol. 60, no. 11, pp. 3461–3471, 2012.

[23] K. Jayasinghe, P. Jayasinghe, N. Rajatheva, and M. Latva-Aho, "Secure beamforming design for physical layer network coding based MIMO two-way relaying," *IEEE Communications Letters*, vol. 18, no. 7, pp. 1270–1273, 2014.

[24] Q. Liu, G. Gong, Y. Wang, and H. Li, "A novel physical layer security scheme for MIMO two-way relay channels," in *Proceedings of the IEEE Globecom Workshops (GC '15)*, pp. 1–6, San Diego, Calif, USA, December 2015.

[25] O. Longoria-Gandara and R. Parra-Michel, "Estimation of correlated MIMO channels using partial channel state information and DPSS," *IEEE Transactions on Wireless Communications*, vol. 10, no. 11, pp. 3711–3719, 2011.

[26] S. Chen, W. Yao, and L. Hanzo, "Semi-blind adaptive spatial equalization for MIMO systems with high-order QAM signalling," *IEEE Transactions on Wireless Communications*, vol. 7, no. 11, pp. 4486–4491, 2008.

[27] C.-L. Wang and H.-C. Wang, "Optimized joint fine timing synchronization and channel estimation for MIMO systems," *IEEE Transactions on Communications*, vol. 59, no. 4, pp. 1089–1098, 2011.

[28] G. H. Golub and C. F. Van Loan, *Matrix Computations*, vol. 3, JHU Press, 2012.

Hindawi
Mobile Information Systems
Volume 2017, Article ID 3680671, 11 pages
https://doi.org/10.1155/2017/3680671

Research Article

Efficient and Privacy-Aware Power Injection over AMI and Smart Grid Slice in Future 5G Networks

Yinghui Zhang,[1,2,3] Jiangfan Zhao,[1] and Dong Zheng[1,3]

[1]*National Engineering Laboratory for Wireless Security, Xi'an University of Posts and Telecommunications, Xi'an 710121, China*
[2]*State Key Laboratory of Cryptology, Beijing 100878, China*
[3]*Westone Cryptologic Research Center, Beijing 100070, China*

Correspondence should be addressed to Yinghui Zhang; yhzhaang@163.com and Dong Zheng; dzhengcrypto@hotmail.com

Received 20 October 2016; Revised 8 December 2016; Accepted 4 January 2017; Published 30 January 2017

Academic Editor: Jing Zhao

Smart grid is critical to the success of next generation of power grid, which is expected to be characterized by efficiency, cleanliness, security, and privacy. In this paper, aiming to tackle the security and privacy issues of power injection, we propose an efficient and privacy-aware power injection (EPPI) scheme suitable for advanced metering infrastructure and 5G smart grid network slice. In EPPI, each power storage unit first blinds its power injection bid and then gives the blinded bid together with a signature to the local gateway. The gateway removes a partial blind factor from each blinded bid and then sends to the utility company aggregated bid and signature by using a novel aggregation technique called hash-then-addition. The utility company can get the total amount of collected power at each time slot by removing a blind factor from the aggregated bid. Throughout the EPPI system, both the gateway and the utility company cannot know individual bids and hence user privacy is preserved. In particular, EPPI allows the utility company to check the integrity and authenticity of the collected data. Finally, extensive evaluations indicate that EPPI is secure and privacy-aware and it is efficient in terms of computation and communication cost.

1. Introduction

The fifth generation of mobile technology (5G) is positioned to provide a holistic end-to-end infrastructure that will include all aspects of the network. To be specific, the future 5G network is envisioned to provide higher data rates, enhanced end-user experience, and much lower latency and energy consumption. In particular, security and privacy preservation mechanisms are expected to be achieved besides the enhanced performance in 5G networks. Because wireless data services have witnessed an explosive growth driven by mobile Internet and smart devices, the new 5G mobile networks are expected to be deployed around 2020. The 5G architecture should include modular network functions that could be deployed and scaled on demand, to accommodate different use cases in a cost efficient and flexible manner.

As the next generation of power grid, smart grid belongs to a representative use case suggested in the Next Generation Mobile Network (NGMN) association's white paper [1]. Smart grid combines traditional grid with communication and information control technologies. It is expected to be characterized by efficiency, cleanliness, consumer involvement, security, privacy, and so forth. Indeed, as one of the main objectives of smart grid, the reduction of greenhouse gas emissions is greatly meaningful for the lives of the people [2]. This objective can be realized by widely deploying renewable energy generators and adaptively balancing the power demand and supply. Therefore, smart grid should have a large number of power storage units to store the excess power in certain cases, such as strong wind. Then, they inject the excess power to the grid when the utility company begins to collect energy at the period of reduced production. Both the utility company and the power storage units benefit from this process, where the utility company should be able to communicate with the power storage units. As important components of smart grid, smart meters (SMs) are two-way communication devices which are used to record power consumption periodically and collect real-time information on grid operations. The power storage units can be connected to the network of SMs through the existing

network infrastructure Advanced Metering Infrastructure (AMI). SMs can communicate with local gateways based on AMI networks and the communication between the gateways and the utility company could be realized through the 5G smart grid network slice.

Considering the issues of user privacy, communication efficiency, and so forth, it is meaningful for the gateways to aggregate the received requests before sending them to the utility company. Since transactions will be involved during power injection, data security and user privacy are of great importance. For one thing, all the data transmitted in the grid should be authenticated and be secure against unauthorized reading and malicious modifications. For another, user privacy related information must be protected against various attackers. For instance, during the communication process of power injection, individual power injection bids are sensitive and must be hidden. If some power storage units know that the other units do not inject power, they can deny selling power to force the utility company to offer a higher price. Furthermore, computation cost and communication overheads must be taken into account during power injection in smart grid. Therefore, the aggregation technique is important for addressing the above issues. As far as the authors' knowledge, however, most existing solutions cannot tackle the security issues of power injection in smart grid.

To solve the above problems, in this paper, we propose an Efficient and Privacy-aware Power Injection (EPPI) scheme suitable for AMI and 5G smart grid network slice. In EPPI, a novel data aggregation technique, named *hash-then-addition*, is proposed. Specifically, each power storage unit can generate two secret keys based on bilinear pairings and use the hash values of the keys to blind its power injection bid. Based on AMI networks, each power storage unit sends its blinded power bid and the corresponding signature to the local gateway. It also generates a message authentication code based on exponentiation operations and sends the result to the gateway. Upon receiving packets from all the units, the gateway first removes a partial blind factor from each blinded bid. Then it aggregates the bids and signatures to get an aggregated bid and an aggregated signature. The gateway also aggregates all the message authentication codes to achieve an aggregated code. All the aggregated values are sent by the gateway to the utility company through the 5G smart grid network slice. Upon receiving the packet, the utility company can generate some secret keys and get the total amount of injected power at each time slot by removing the sum of the secret keys from the aggregated bid. In the proposed EPPI system, only the utility company can know the total amount of injected power at each time slot, and it is able to ensure the integrity and authenticity of the data. In particular, individual power bids are hidden during the communications in AMI networks and the 5G smart grid network slice. Through extensive evaluations we show that EPPI is secure and privacy-aware under the discrete logarithm assumption and it is efficient in terms of computation and communication cost.

1.1. Related Work. The 5G system deployed initially in 2020 is expected to provide approximately 1000 times higher wireless area capacity compared with the current 4G system [3]. The advanced cloud radio access network (C-RAN) has been presented as a potential 5G solution. C-RAN has attracted intense research interest from both academia and industry [4]. Combined with cloud computing technologies, the 5G network will extend its capability to provide various cloud services, which provides the user a full smart life experience. Cloud computing security has been well studied [5–8] and various cloud services are provided [9–11]. The air interface and spectrum of the 5G system should be combined with the long term evolution (LTE) and WiFi to achieve seamless and consistent user experience across time and space [12]. Nikaein et al. [13] presented a slice-based 5G architecture that efficiently manages network slices. NGMN anticipates countless emerging use cases with a high variety of applications will be supported in 5G. As an important use case, smart grid has attracted many scholars' interest. In smart grid, in order to reduce communication overhead, it is essential to aggregate individual users' data at local intermediate nodes. The trivial method of decrypt-aggregate-encrypt is computationally expensive and is risky when intermediate nodes are not trusted. Castelluccia et al. [14] enabled efficient encrypted data aggregation based on homomorphic encryption techniques. Westhoff et al. [15] proposed a key predistribution scheme that is suitable for the end-to-end encryption in sensor networks. A symmetric homomorphic encryption can be used together with [15] to improve the efficiency and flexibility of data aggregation. In smart grid, each user has energy consumption data of multiple dimensions and each dimensional data is small in size. If homomorphic encryption techniques are used directly on each dimensional data, the communication overhead will be unaffordable. Lin et al. [16] proposed a multidimensional privacy-preserving data aggregation scheme for saving energy consumption in wireless sensor networks by integrating the super-increasing sequence and perturbation techniques into compressed data aggregation. Lu et al. [17] designed a compressed data aggregation scheme under the public key infrastructure to improve efficiency and achieve high reliability. To improve the performance of the power grid, several schemes have been proposed to coordinate power charging [18, 19].

In AMI networks, smart meters periodically send fine-grained power consumption data to the utility company. This data has a relation to the users' activities and hence is sensitive. Tonyali et al. [20] developed a meter data obfuscation scheme to protect consumer privacy from eavesdroppers and the utility company. In order to tackle the scalability of AMI networks, Rabieh et al. [21] divided the AMI network into clusters of SMs and proposed two certificate revocation schemes to identify and nullify the false positives when using bloom filters to reduce the size of the certificate revocation lists. Besides, privacy-aware schemes with various security characteristics have been investigated for different network environment and applications [22–27]. Note that these schemes are not focusing on the security and privacy issues in power injection. Recently, Mahmoud et al. [28] have proposed a power injection querying scheme over AMI and LTE cellular networks. In [28], two aggregation techniques, point addition aggregation and homomorphic encryption based aggregation, are adopted to enable the local gateway to

aggregate individuals' power bids, where the homomorphic encryption [29] is used. However, we found that the scheme [28] fails to achieve privacy protection in that it cannot preserve the power storage unit's bid. In fact, in [28], the utility company recovers the total amount of power by exhaustively computing jQ at different values of $j \in \mathbb{Z}_q^*$ until $jQ = \sum_{i=1}^{n} b_i Q$, where Q is a bilinear group element and b_i is an individual power injection bid. Obviously, this computation contradicts with the discrete logarithm assumption. In order to enable the utility company to get $\sum_{i=1}^{n} b_i$, the authors assume that $\sum_{i=1}^{n} b_i$ is a small number. Unfortunately, if $\sum_{i=1}^{n} b_i$ is a small number, then each b_i is a small number. In this case, any attacker can get b_i and hence violates the bid privacy in that $B_i = b_i Q$ is publicly sent by the power storage unit. More details can be found in [28]. The proposed EPPI realizes privacy preservation by using hash-then-addition aggregation technique and it does not constrain the power amount in any way.

1.2. Organization. The remaining of this work is organized as follows. We first review some preliminaries in Section 2. In Section 3, we present the system architecture and adversary models. We propose an efficient and privacy-aware power injection scheme over AMI and smart grid slice in 5G networks in Section 4. The security analysis and performance evaluations are described in Sections 5 and 6, respectively. Finally, we draw our conclusions in Section 7.

2. Preliminaries

In this section, we give a brief review on some cryptographic backgrounds.

2.1. Cryptographic Background

Definition 1 (bilinear pairings). Let $\mathbb{G}$, $\mathbb{G}_T$ be cyclic multiplicative groups of prime order q. Let $P \in \mathbb{G}$ be a generator. We call $\hat{e}$ a bilinear pairing if $\hat{e} : \mathbb{G} \times \mathbb{G} \to \mathbb{G}_T$ is a map with the following properties [30]:

(1) bilinear: $\hat{e}(aP, bP) = \hat{e}(P, P)^{ab}$ for all $a, b \in \mathbb{Z}_q^*$,

(2) nondegenerate: there exists $P, Q \in \mathbb{G}$ such that $\hat{e}(P, Q) \neq 1$,

(3) computable: there is an efficient algorithm to compute $\hat{e}(P, Q)$ for all $P, Q \in \mathbb{G}$.

We define that $\mathcal{G}(\lambda)$ outputs $(q, P, \mathbb{G}, \mathbb{G}_T, \hat{e})$ where λ is a security parameter.

2.2. Discrete Logarithm Assumption

Definition 2 (discrete logarithm problem [31]). Let $\mathbb{G}$ be a group of prime order q, given two elements P and Q, to find an integer $x \in \mathbb{Z}_q^*$, such that $Q = xP$ whenever such an integer exists.

Definition 3 (discrete logarithm assumption [31]). In group $\mathbb{G}$, it is computationally infeasible to determine x from P and $Q = xP$.

3. System Architecture and Adversary Models

3.1. System Architecture. As shown in Figure 1, the system architecture of power injection over AMI and smart grid slice in 5G networks involves a number of communities and a utility company, and they are connected through a smart grid 5G network slice. In a community, there are many power storage units that are connected to AMI. Each AMI network connects to an access node and eventually to the utility company through a smart grid 5G network slice. The details are given as below.

(i) Power Storage Units. The power storage units can be home batteries or charging stations. They store power energy from the smart grid or other renewable energy sources. Each storage unit can buy power from the grid at a low-price period and inject excess power energy to the grid at a high-price period. Note that a storage unit communicates with a SM based on the IEEE 802.11s protocol.

(ii) AMI Network. The AMI network is an architecture for automated, two-way communication between SMs and a utility company. The goal of an AMI is to provides utility companies with real-time data about power consumption and allow users to make informed choices about energy usage based on the price at the time of use. In this paper, for the sake of efficiency, cleanliness, security, and privacy, smart meters communicate indirectly with the utility company by the gateway. The AMI network corresponding to a community comprises a group of SMs and a gateway. Similar to the work [28], two different AMI network topologies are considered: single hop AMI networks and multihop AMI networks. As shown in Figure 1, the SMs in the multihop AMI network are connected through a multihop wireless mesh topology, where each SM plays a role of relaying packets from other SMs. In the single-hop AMI network, the gateway can directly collect power injection related data from corresponding SMs, and then it aggregates the data and sends the result to the utility company. This process is performed periodically, for example, every 15 minutes. It can also receive the latest power data from the utility company and broadcast it to the SMs in the corresponding AMI network. Note that, similar to [28], the AMI network routes are created using the IEEE 802.11s mesh standard.

(iii) Smart Grid Slice in 5G Networks. A 5G network slice supports the communication service of a particular connection type with a specific way of handling the C- and U-plane for this service [1]. For this purpose, a 5G slice consists of a series of 5G network functions and specific radio access technology (RAT) settings that are combined together for the particular use case. Therefore, a 5G slice can span all domains of the network. Not all slices contain the same functions, and some functions that seem important for a mobile network might even be missing in other slices. After network function virtualization, the radio access network and the core network are called edge cloud and central cloud (or core cloud), respectively. The front haul between the access node and the edge cloud is based on software-define

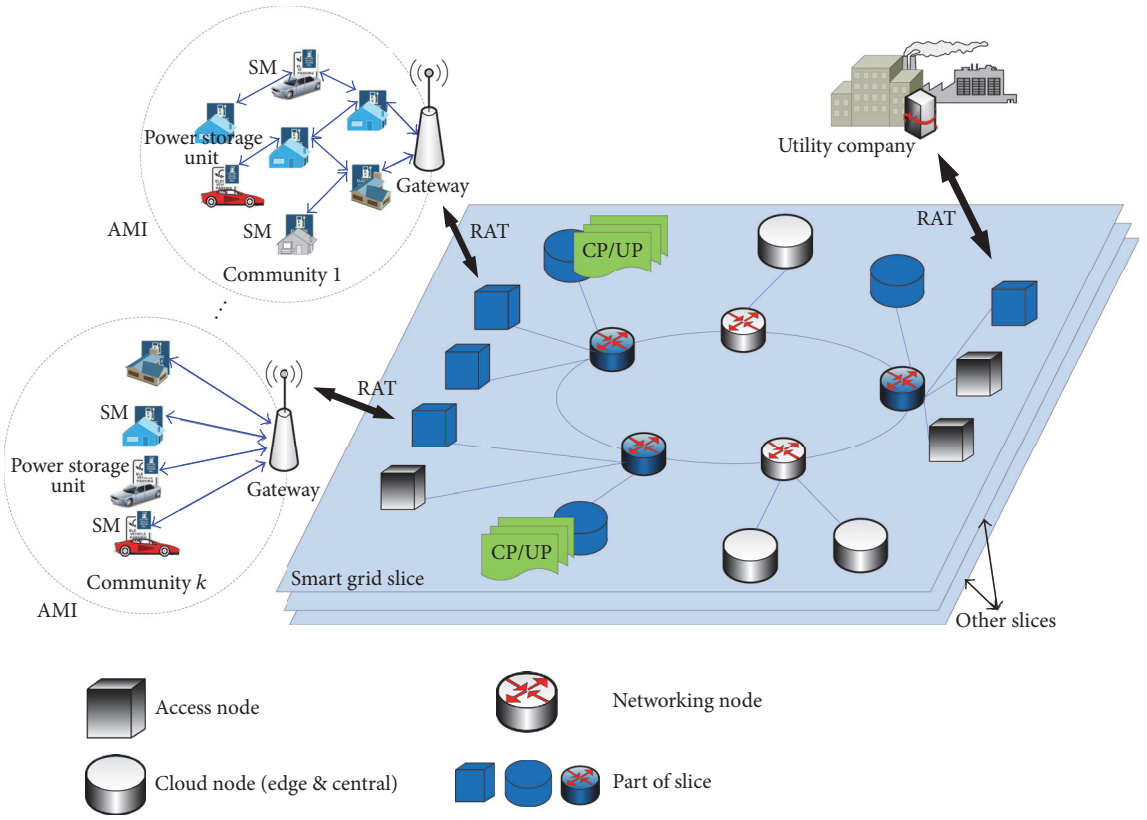


Figure 1: System architecture of power injection in future 5G networks.

networking (SDN). The backhaul between the edge cloud and the core cloud is also based on SDN. For a 5G slice supporting smart grid use case, security, privacy, reliability, and latency are of paramount importance. As shown in Figure 1, to tailor the network functions to suit the smart gird slice, all the necessary functions are instantiated at the cloud edge node.

(iv) Utility Company. If the power energy demand from communities is more than the supply, the utility company should contact electricity vendors or power storage units to buy power. Note that the utility company communicates with the power storage units via the AMI and 5G smart grid slice networks. It connects to the 5G slice through an access node.

3.2. Adversary Models. It is assumed that all the entities are "honest-but-curious." More precisely, they will honestly execute the tasks assigned by legitimate parties but try to find out as much private information as possible. Each power storage unit is curious to know the other units' bids to judge whether it is more profitable to inject power now. We assume that each power storage unit can only send packets in the corresponding community. The AMI network attackers, the smart grid 5G slice attackers, and outsiders are also interested in other's sensitive information, such as the amount and time of the power injection of each power storage unit. Similarly, the utility company does not disrupt the communication, but it tries to get private information on the owners of the power storage units and any other information that can help gain economic benefits. Note that the utility company does not collude with the power storage units in that they have conflicting interests. The utility wants to buy power at low prices but the storage units want to increase revenues. The power storage units will inject the amount of power as committed in their bids because this is more profitable.

3.3. Security Requirements and Design Goals. Considering the practical application environment, security and privacy are significant for the success of a power injection system. In order to prevent aforementioned adversaries from learning power storage units' individual bid and to detect the adversaries' malicious behaviors, the following security requirements should be satisfied in a secure power injection system.

(i) Confidentiality and Privacy Protection. Even if an adversary eavesdrops the communication on the AMI and smart grid slice networks, it fails to achieve the total amount of power injected from the community. The utility cannot know the contents of individual power storage unit's bid. In our scheme, aggregation at gateways is adopted to achieve these goals.

(ii) Authentication and Integrity. The utility company is able to authenticate the received packets to ensure that the packets are really from legal power storage units and have not been altered during the transmission; that is, if the adversary forges and/or modifies a packet, the malicious behavior should be

detected. Besides, the adversary should not impersonate the utility company, the gateway, or the storage units.

In general, under the proposed system architecture and security requirements, our design goal is to design an efficient and privacy-aware power injection scheme based on AMI and smart grid slice in 5G networks. To be specific, the following two objectives should be achieved. Firstly, the security requirements should be guaranteed in the proposed scheme. For one thing, a desirable scheme should provide robust security against various types of attacks including passive eavesdropping, impersonation attack, replay attack, and man-in-the-middle attack. For another, a desirable power injection scheme should enjoy some significant security benefits such as the assurance of session key freshness which enables the forward and backward secrecy. Secondly, the performance-related issue should be taken into consideration. The proposed power injection scheme should enjoy desirable efficiency in terms of the computation cost and the communication overhead.

4. Proposed EPPI Scheme

In this section, we propose an efficient and privacy-aware power injection scheme over AMI and smart grid slice in 5G networks, which comprises the following six phases: system initialization, registration, power collection request, privacy-aware bid generation, privacy-aware bid aggregation, and privacy-aware aggregated bid reading. Figure 2 presents the process of the proposed EPPI system. The details are given in the following.

4.1. System Initialization. The proposed EPPI system is initialized by the utility company. Specifically, in the system initialization phase, given the security parameter λ, the utility company first generates $(q, P_0, \mathbb{G}, \mathbb{G}_T, \hat{e})$ by running $\mathcal{G}(\lambda)$. It computes $Y = \hat{e}(P_0, P_0)$ and chooses two random elements $U, V \in \mathbb{G}$ and four secure cryptographic hash functions H, H_1, H_2, and H_3, where $H : \{0,1\}^* \to \mathbb{Z}_q^*$, $H_1 : \{0,1\}^* \to \mathbb{G}$, $H_2 : \mathbb{G}_T \to \mathbb{Z}_q^*$, and $H_3 : \mathbb{G}_T \to \{0,1\}^*$. Then, the utility company chooses a random element $\mathrm{sk}_u \in \mathbb{Z}_q^*$ as its secret key and calculates $\mathrm{PK}_u = \mathrm{sk}_u P_0$ as its public key. Finally, the utility company keeps sk_u secret and publishes the global public parameters

$$\mathrm{GPK} = \left(\mathbb{G}, \mathbb{G}_T, \hat{e}, q, P_0, U, V, \mathrm{PK}_u, H, H_1, H_2, H_3, Y\right). \quad (1)$$

4.2. Registration. In order to join the EPPI system, each gateway chooses a random element $\mathrm{sk}_g \in \mathbb{Z}_q^*$ as its secret key and calculates $\mathrm{PK}_g = \mathrm{sk}_g P_0$ as its public key. A power storage unit with identity ID_i chooses a random element $\mathrm{sk}_i \in \mathbb{Z}_q^*$ as its secret key and calculates $\mathrm{PK}_i = \mathrm{sk}_i P_0$ as its public key. Similar to [28], in the proposed EPPI system, all the gateways and power storage units should contact the utility company to receive corresponding certificates for public keys. Note that the existing public key infrastructure (PKI) can be used to generate certificates. Besides PKI, in the proposed scheme, the aggregated message authentication code and the aggregated signatures are necessary to ensure the authenticity and integrity of transaction data, which are shown in the privacy-aware bid aggregation phase.

4.3. Power Collection Request. During the peak hours, the utility company can collect power from related communities. To be specific, the utility company sends power collection request (*Power_Req*) packets to corresponding gateways. Upon receiving the packet, each gateway verifies the freshness and validity of the packet. Then the gateway broadcasts the valid packet in its community.

Suppose the utility company wants to collect power in the community corresponding to the gateway ID_g. As shown in Figure 2, the packet contains the identities of the utility company and the gateway, that is, ID_u and ID_g. It also has the power collection information of price per unit in each time slot, that is, $\mathrm{Info}_p = (p_1, p_2, \ldots, p_k)$, where k is the number of time slots. Then the utility company randomly chooses $r_u \in \mathbb{Z}_q^*$, computes $r_u P_0$, and attaches $r_u P_0$ in the packet *Power_Req*. Note that $r_u P_0$ is used by each power storage unit covered by the gateway ID_g in establishing a one-time key shared with the utility company. Besides, the packet contains a timestamp TS and a signature σ_u, where

$$\sigma_u = \mathrm{sk}_u H_1\left(\mathrm{ID}_u \parallel \mathrm{ID}_g \parallel \mathrm{Info}_p \parallel r_u P_0 \parallel \mathrm{TS}\right). \quad (2)$$

Both TS and σ_u will be used by the gateway in verification of the packet.

In fact, after receiving the packet *Power_Req*, the gateway ID_g first checks the freshness of *Power_Req* according to the difference between the current time and the timestamp TS. Then, it verifies the signature by checking if $\hat{e}(\sigma_u, P_0) = \hat{e}(H_1(\mathrm{ID}_u \parallel \mathrm{ID}_g \parallel \mathrm{Info}_p \parallel r_u P_0 \parallel \mathrm{TS}), \mathrm{PK}_u)$ holds. If and only if the equation holds, the gateway ID_g randomly chooses $r_g \in \mathbb{Z}_q^*$, computes $r_g P_0$, and attaches $r_g P_0$ in the packet *Power_Req*. Note that $r_g P_0$ is used by each power storage unit covered by the gateway ID_g in establishing a one-time key shared with the gateway. Then, the gateway broadcasts the packet in its community.

4.4. Privacy-Aware Bid Generation. Upon receiving the power collection request, each power storage unit should prepare a bid with the amount of power it can inject in each time slot. Then, it sends a power request response *Power_Res* packet to the corresponding gateway or its upstream smart meter. The bid format of the power storage unit ID_i is $b_i = (b_{i,1}, b_{i,2}, \ldots, b_{i,k})$, where $b_{i,x}$ represents the number of power units the power storage unit ID_i can inject in the x-th time slot at price p_x for $1 \le x \le k$. As shown in Figure 2, the packet *Power_Res* contains the identities of the gateway and the utility company, that is, ID_g and ID_u. The power storage unit ID_i randomly chooses $r_i \in \mathbb{Z}_q^*$, computes $r_i P_0$, and attaches $r_i P_0$ in the packet *Power_Res*. Note that $r_i P_0$ is used by ID_u in establishing a shared one-time key between ID_i and ID_u. The power storage unit ID_i computes two shared keys as $\hat{k}_i = H_2(\hat{e}(\mathrm{PK}_g, \mathrm{sk}_i r_i r_g P_0))$ and $k_i = H_2(\hat{e}(\mathrm{PK}_u, \mathrm{sk}_i r_i r_u P_0))$, which will be used to mask ID_i's bid and k_i can enable the utility company to ensure the authenticity and integrity of the aggregated bids without needing to read the individual bid.

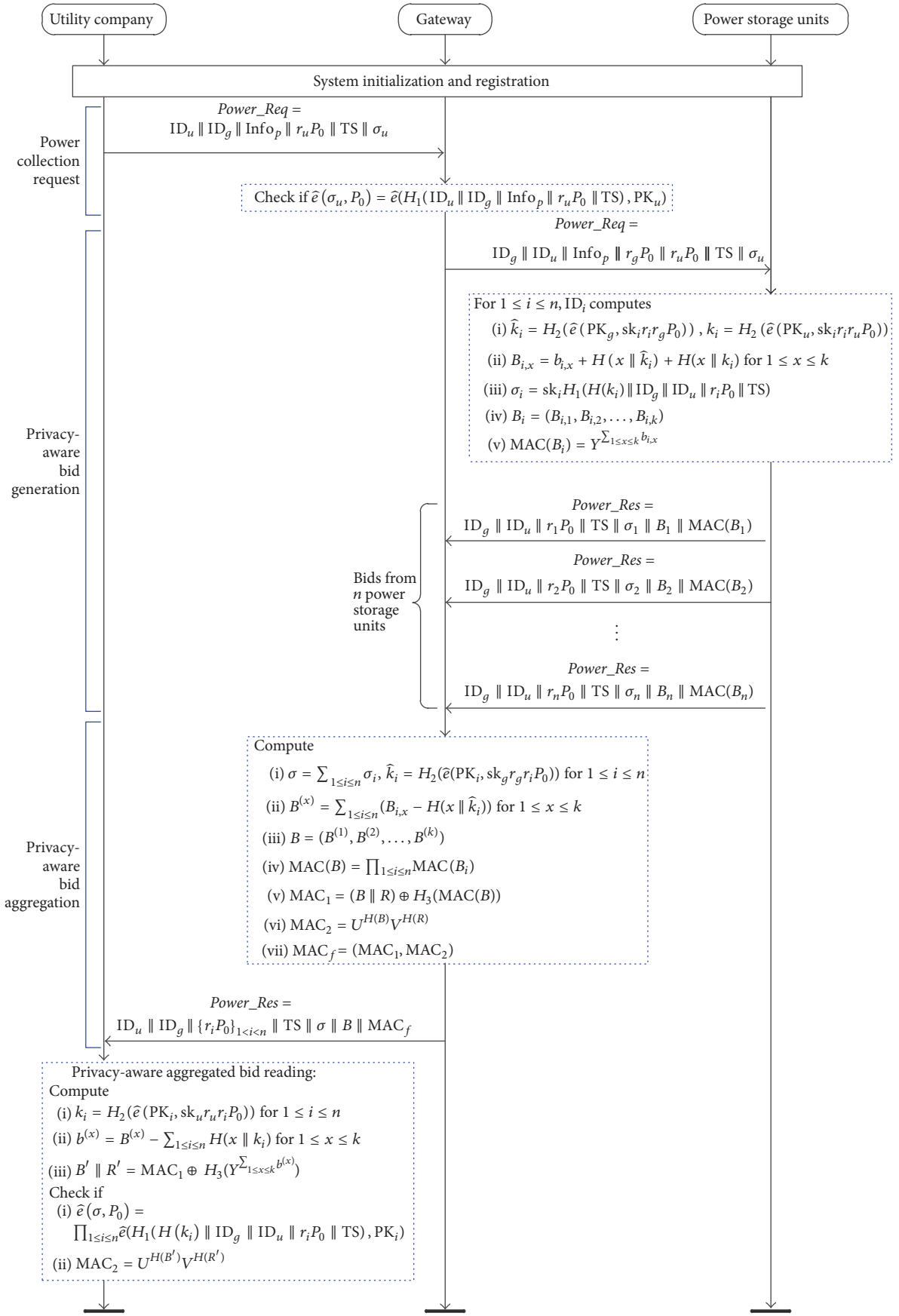


Figure 2: Six phases of the proposed EPPI system.

It is noted that we propose a novel bib aggregation method called *hash-then-addition* in the aggregation phase. Corresponding to this method, the power storage unit ID_i computes its masked bid as $B_i = (B_{i,1}, B_{i,2}, \ldots, B_{i,k})$, where $B_{i,x} = b_{i,x} + H(x \parallel \widehat{k}_i) + H(x \parallel k_i)$ for $1 \le x \le k$. It then calculates a signature $\sigma_i = \mathrm{sk}_i H_1(H(k_i) \parallel \mathrm{ID}_g \parallel \mathrm{ID}_u \parallel r_i P_0 \parallel \mathrm{TS})$ and a message authentication code $\mathrm{MAC}(B_i) = Y^{\sum_{1\le x\le k} b_{i,x}}$. Note that the masked bid can realize privacy protection in the sense of hiding individual bids.

4.5. Privacy-Aware Bid Aggregation. Upon receiving all the power request response packets, the gateway ID_g aggregates these packets and sends an aggregated response packet to the utility company ID_u. The aggregated packet enjoys the following benefits. Firstly, the power storage unit's bid privacy is preserved, which is very important in practical applications. For example, it can prevent the utility company from manipulating the power collection price. In fact, what the utility company needs is not the power storage units' individual power injection data, but the total power amount that can be collected from the community in each time slot. Secondly, the aggregated packet has smaller packet size and hence reduces the required bandwidth for transmitting the data to the utility company. Thirdly, instead of sending one message for each bid, all the bids in different time slots can be collected in one message. In the following, we show how to aggregate the packets considering two different scenarios: a single-hop AMI network and a multihop AMI network.

In the case of a single-hop AMI network, upon receiving all the *Power_Res* packets, ID_g computes a secret key $\widehat{k}_i = H_2(\widehat{e}(\mathrm{PK}_i, \mathrm{sk}_g r_g r_i P_0))$ shared with the power storage unit ID_i for $1 \le i \le n$. We note that $\widehat{k}_i = H_2(\widehat{e}(\mathrm{PK}_g, \mathrm{sk}_i r_i r_g P_0)) = H_2(\widehat{e}(\mathrm{PK}_i, \mathrm{sk}_g r_g r_i P_0))$. Then, ID_g aggregates the signatures, masked bids, and message authentication codes to generate an aggregated signature σ, an aggregated masked bid B, and an aggregated message authentication code $\mathrm{MAC}(B)$. The aggregated signature is $\sigma = \sum_{1\le i\le n} \sigma_i$. The aggregated bid is $B = (B^{(1)}, B^{(2)}, \ldots, B^{(k)})$, where $B^{(x)} = \sum_{1\le i\le n}(B_{i,x} - H(x \parallel \widehat{k}_i))$ for $1 \le x \le k$. The aggregated message authentication code is $\mathrm{MAC}(B) = \prod_{1\le i\le n}\mathrm{MAC}(B_i)$. Additionally, the gateway ID_g randomly chooses $R \in \mathbb{Z}_q^*$ and calculates the final message authentication code $\mathrm{MAC}_f = (\mathrm{MAC}_1, \mathrm{MAC}_2)$, where $\mathrm{MAC}_1 = (B \parallel R) \oplus H_3(\mathrm{MAC}(B))$ and $\mathrm{MAC}_2 = U^{H(B)} V^{H(R)}$. MAC_f will be used by the utility company ID_u to ensure that the aggregated bid in each time slot stems from the intended power storage units and it has not been modified in transit. Note that during the verification process, ID_u does not need to access the individual bid and hence the power storage unit's privacy is preserved.

In the case of a multihop AMI network, the aggregation process of the signatures and masked bids are done by the SMs in a bottom-up way, as shown in Figure 3. Once a SM receives *Power_Res* packets from its downstream SMs, it first aggregates them with its own data and then sends the aggregated packet to its upstream SM. For example, as shown in Figure 3, SM_2 and SM_3 send their *Power_Res* packets to their upstream smart meter SM_4. After receiving packets from SM_2 and SM_3, SM_4 aggregates its signature to the received signatures of SM_2 and SM_3 to generate aggregated signature $\sigma_{2-4} = \sigma_2 + \sigma_3 + \sigma_4$. Then, SM_4 aggregates its masked bid to the received masked bids of SM_2 and SM_3 to generate aggregated masked bid $B_{2-4} = (B_{2-4,1}, B_{2-4,2}, \ldots, B_{2-4,k})$, where $B_{2-4,x} = B_{2,x} + B_{3,x} + B_{4,x}$ for $1 \le x \le k$. SM_4 also aggregates its message authentication code to the received message authentication codes of SM_2 and SM_3 to generate an aggregated message authentication code $\mathrm{MAC}(B_{2-4}) = \mathrm{MAC}(B_2) \cdot \mathrm{MAC}(B_3) \cdot \mathrm{MAC}(B_4)$. Similarly, SM_5 generates $\sigma_{1-5} = \sigma_1 + \sigma_{2-4} + \sigma_5$, $B_{1-5} = (B_{1-5,1}, B_{1-5,2}, \ldots, B_{1-5,k})$, where $B_{1-5,x} = B_{1,x} + B_{2-4,x} + B_{5,x}$ for $1 \le x \le k$, and $\mathrm{MAC}(B_{1-5}) = \mathrm{MAC}(B_1) \cdot \mathrm{MAC}(B_{2-4}) \cdot \mathrm{MAC}(B_5)$. SM_8 generates $\sigma_{6-8} = \sigma_6 + \sigma_7 + \sigma_8$, $B_{6-8} = (B_{6-8,1}, B_{6-8,2}, \ldots, B_{6-8,k})$, where $B_{6-8,x} = B_{6,x} + B_{7,x} + B_{8,x}$ for $1 \le x \le k$, and $\mathrm{MAC}(B_{6-8}) = \mathrm{MAC}(B_6) \cdot \mathrm{MAC}(B_7) \cdot \mathrm{MAC}(B_8)$. Upon receiving response packets from SM_5 and SM_8, the gateway ID_g computes $\widehat{k}_i = H_2(\widehat{e}(\mathrm{PK}_i, \mathrm{sk}_g r_g r_i P_0))$ for $1 \le i \le 8$, $\sigma = \sigma_{1-8} = \sigma_{1-5} + \sigma_{6-8}$, $B_{1-8} = (B_{1-8,1}, B_{1-8,2}, \ldots, B_{1-8,k})$, where $B_{1-8,x} = B_{1-5,x} + B_{6-8,x}$ for $1 \le x \le k$, $B = (B^{(1)}, B^{(2)}, \ldots, B^{(k)})$, where $B^{(x)} = B_{1-8,x} - \sum_{1\le i\le 8} H(x \parallel \widehat{k}_i)$ for $1 \le x \le k$, and $\mathrm{MAC}(B) = \mathrm{MAC}(B_{1-5}) \cdot \mathrm{MAC}(B_{6-8})$. Additionally, the gateway ID_g randomly chooses $R \in \mathbb{Z}_q^*$ and calculates the final message authentication code $\mathrm{MAC}_f = (\mathrm{MAC}_1, \mathrm{MAC}_2)$, where $\mathrm{MAC}_1 = (B \parallel R) \oplus H_3(\mathrm{MAC}(B))$ and $\mathrm{MAC}_2 = U^{H(B)} V^{H(R)}$.

In any cases, the gateway ID_g attaches $\{r_i P_0\}_{1\le i\le n}$ to the aggregated response packet, where n is the number of power storage units covered by ID_g. Finally, the aggregated response packet is sent to the utility company. Note that only the final message authentication code MAC_f is sent to the utility company by the gateway.

4.6. Privacy-Aware Aggregated Bid Reading. After receiving the power request response packet from the gateway ID_g, the utility company computes a secret key $k_i = H_2(\widehat{e}(\mathrm{PK}_i, \mathrm{sk}_u r_u r_i P_0))$ shared with the power storage unit ID_i for $1 \le i \le n$. We note that $k_i = H_2(\widehat{e}(\mathrm{PK}_i, \mathrm{sk}_u r_u r_i P_0)) = H_2(\widehat{e}(\mathrm{PK}_u, \mathrm{sk}_i r_i r_u P_0))$. For $1 \le x \le k$, the utility company computes $b^{(x)} = B^{(x)} - \sum_{1\le i\le n} H(x \parallel k_i) = \sum_{1\le i\le n} b_{i,x}$, which is the power amount the utility company can collect from the community of ID_g in the x-th time slot at price p_x. Then, the utility company ensures the authenticity and integrity of the recovered data by checking if $\widehat{e}(\sigma, P_0) = \prod_{1\le i\le n} \widehat{e}(H_1(H(k_i) \parallel \mathrm{ID}_g \parallel \mathrm{ID}_u \parallel r_i P_0 \parallel \mathrm{TS}), \mathrm{PK}_i)$. Finally, in order to ensure that the recovered aggregated bids stem from the intended power storage units and they have not been modified in transition, the utility company computes $B' \parallel R' = \mathrm{MAC}_1 \oplus H_3(Y^{\sum_{1\le x\le k} b^{(x)}})$ and checks whether $\mathrm{MAC}_2 = U^{H(B')} V^{H(R')}$.

5. Security and Privacy Analysis

In this section, we show EPPI can achieve the expected security and privacy goals.

5.1. Confidentiality. In the privacy-aware aggregated bid reading phase, for $1 \le x \le k$, the utility company computes $b^{(x)} = B^{(x)} - \sum_{1\le i\le n} H(x \parallel k_i) = \sum_{1\le i\le n} b_{i,x}$, which is

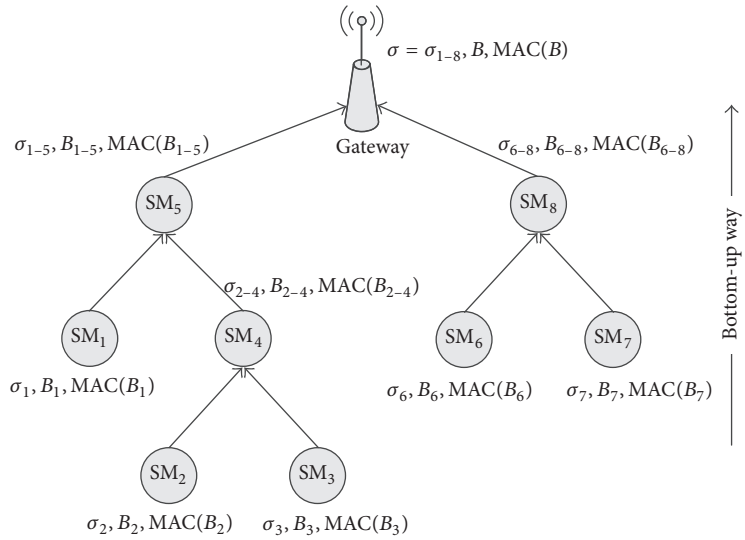


Figure 3: The aggregation way in multihop AMI networks.

the power amount the utility company can collect from the community of ID_g in the x-th time slot at price p_x. Then the utility company can know the total amount of power $\sum_{1\le x\le k} b^{(x)}$ injected from the community of ID_g. Obviously, the secret keys $\{k_i\}_{1\le i\le n}$ are necessary for the computation of the total amount of power. Therefore, adversaries cannot know the total amount of power. On the other hand, based on the discrete logarithm assumption, it is infeasible for attackers to compute $\sum_{1\le x\le k} b_{i,x}$ from $\mathrm{MAC}(B_i) = Y^{\sum_{1\le x\le k} b_{i,x}}$. Also, the gateway fails to recover $\sum_{1\le x\le k} b^{(x)}$ from $\mathrm{MAC}(B) = Y^{\sum_{1\le x\le k} b^{(x)}}$.

5.2. Privacy Protection. In the privacy-aware bid generation phase, the power storage unit ID_i computes its masked bid as $B_i = (B_{i,1}, B_{i,2}, \ldots, B_{i,k})$, where $B_{i,x} = b_{i,x} + H(x \parallel \widehat{k}_i) + H(x \parallel k_i)$ for $1 \le x \le k$. Because the secret keys $\widehat{k}_i$ and k_i are involved in the computation of B_i, any adversaries without knowing $\widehat{k}_i$ or k_i cannot know the original bid $b_{i,x}$ even if B_i is given. Although the utility company has the value k_i, it fails to recover $b_{i,x}$ in that $H(x \parallel \widehat{k}_i)$ cannot be removed from $B_{i,x}$. On the other hand, because the utility company only has the aggregated value $B^{(x)} = \sum_{1\le i\le n}(B_{i,x} - H(x \parallel \widehat{k}_i))$, it just gets the sum $\sum_{1\le i\le n} b_{i,x}$ by computing $b^{(x)} = B^{(x)} - \sum_{1\le i\le n} H(x \parallel k_i)$. In this case, the individual bid privacy is still preserved.

Furthermore, The use of one-time keys $\widehat{k}_i$ and k_i in hash-then-addition aggregation can boost the privacy protection because when the power storage units send the same bids in different cases, the masked bids are completely different and the attacker cannot distinguish the bids. In particular, the time slot parameter x is used in the generation of $B_{i,x}$, which makes it impossible for the attacker to calculate the difference between related bids.

5.3. Authentication and Integrity. The utility company ensures the authenticity and integrity of the recovered data by checking if $\widehat{e}(\sigma, P_0) = \prod_{1\le i\le n} \widehat{e}(H_1(H(k_i) \parallel \mathrm{ID}_g \parallel \mathrm{ID}_u \parallel r_i P_0 \parallel \mathrm{TS}), \mathrm{PK}_i)$. Any modification to a packet content, such as power price, will result in the failure of the signature verification. Signatures can also be used to resist impersonation attacks and external attacks such as denial of service by sending false packets. The attackers cannot impersonate the utility, gateway, or the power storage units because the generation of a valid signature needs a secret key. Based on the discrete logarithm assumption, it is infeasible to compute the secret key sk_i from the corresponding public key $\mathrm{PK}_i = \mathrm{sk}_i P_0$ and the signature $\sigma_i = \mathrm{sk}_i H_1(H(k_i) \parallel \mathrm{ID}_g \parallel \mathrm{ID}_u \parallel r_i P_0 \parallel \mathrm{TS})$. Besides, we developed a message authentication code based on signature techniques. The gateway ID_g randomly chooses $R \in \mathbb{Z}_q^*$ and calculates the final message authentication code $\mathrm{MAC}_f = (\mathrm{MAC}_1, \mathrm{MAC}_2)$, where $\mathrm{MAC}_1 = (B \parallel R) \oplus H_3(\mathrm{MAC}(B))$ and $\mathrm{MAC}_2 = U^{H(B)} V^{H(R)}$. MAC_f can be used by the utility company ID_u to ensure that the aggregated bid in each time slot stems from the intended power storage units and it has not been modified in transit.

5.4. Replay Attacks. In the proposed EPPI system, if attackers record valid packets and replay them in a different community or time slot, these replayed packets will be identified and dropped. For one thing, time stamps are used to protect against this replay attacks. For another, the verification of MAC_f fails if an attacker replays packets associated with old secret keys. In EPPI, we adopt a key management procedure to enable the utility company to share keys with power storage units. The attackers cannot calculate the keys because the secret number r_i is used, which is selected by each power storage unit. It is infeasible to retrieve r_i from $r_i P_0$. Particularly, even if the gateway and some power storage units collude, they cannot achieve the shared secret key between the utility company and a victim because the secret key computation is controlled jointly by the power storage unit and the utility company.

5.5. Man-in-the-Middle Attacks. In the proposed EPPI system, suppose an attacker resides between a power storage unit and the utility company. It tries to establish two secret keys to fool the power storage unit and the utility company to believe that they communicate directly, where one key is shared with the utility company and the other is shared with the power storage unit. The secret key agreement procedure is resilient to this attack because $r_i P_0$ and $r_u P_0$ are signed by the power storage unit and the utility company, respectively.

5.6. Session Key Freshness. It is a very desirable practice to periodically refresh the shared secret keys. In the proposed EPPI system, the secret key management procedure can achieve both forward and backward secrecy, where the attacker cannot derive the previously used session keys nor the future session keys even if the current key is exposed. This is because each time the utility company requests power injection, a new key is computed using one-time random numbers r_u and r_i. Therefore, if an attacker could get one key, this does not help him to know the old or new ones.

6. Performance Evaluation

In this section, we evaluate the performance of the proposed EPPI scheme in terms of the computation complexity and the communication overhead.

6.1. Computation Complexity. As for computation complexity, we will focus on measuring the time required for performing the cryptographic operations in EPPI. Denote the computational costs of a bilinear pairing operation, an exponentiation operation in $\mathbb{G}$, an exponentiation operation in $\mathbb{G}_T$, a multiplication operation in $\mathbb{G}$, a multiplication operation in $\mathbb{G}_T$, and an addition operation in $\mathbb{G}$ by C_p, C_e, C_{et}, C_m, C_{mt}, and C_a, respectively.

In the proposed EPPI scheme, in order to generate a power collection request $Power\_Req = \mathrm{ID}_u \parallel \mathrm{ID}_g \parallel \mathrm{Info}_p \parallel r_u P_0 \parallel \mathrm{TS} \parallel \sigma_u$, the utility company needs $2C_e$ computation cost. In the privacy-aware aggregated bid reading phase, the computation cost for the utility company is $(2n+1)C_p+(n+2)C_e+C_{et}+C_m$. In fact, the computation of the secret key $k_i = H_2(\widehat{e}(\mathrm{PK}_i, \mathrm{sk}_u r_u r_i P_0))$ shared with the power storage unit ID_i involves one C_p and one C_e. The authenticity and integrity of the recovered data based on $\widehat{e}(\sigma, P_0) = \prod_{1\le i\le n}\widehat{e}(H_1(H(k_i) \parallel \mathrm{ID}_g \parallel \mathrm{ID}_u \parallel r_i P_0 \parallel \mathrm{TS}), \mathrm{PK}_i)$ involves $(n+1)C_p$. To compute $B' \parallel R' = \mathrm{MAC}_1 \oplus H_3(Y^{\sum_{1\le x\le k} b^{(x)}})$, one C_{et} is needed. The verification based on $\mathrm{MAC}_2 = U^{H(B')}V^{H(R')}$ involves $2C_e$ and C_m. After receiving the packet *Power_Req*, the gateway verifies the signature by checking if $\widehat{e}(\sigma_u, P_0) = \widehat{e}(H_1(\mathrm{ID}_u \parallel \mathrm{ID}_g \parallel \mathrm{Info}_p \parallel r_u P_0 \parallel \mathrm{TS}), \mathrm{PK}_u)$ holds. Then it computes $r_g P_0$. The computation cost in this process is $2C_p+C_e$. In the privacy-aware bid aggregation phase, the computation cost for the gateway is $nC_p+(n+2)C_e+C_m+(n-1)C_{mt}+(n-1)C_a$. Specifically, the aggregated signature is $\sigma = \sum_{1\le i\le n}\sigma_i$ and it needs $(n-1)C_a$. The computation of the secret key $\widehat{k}_i = H_2(\widehat{e}(\mathrm{PK}_g, \mathrm{sk}_i r_i r_g P_0))$ involves one C_p and one C_e. The aggregated message authentication code is $\mathrm{MAC}(B) = \prod_{1\le i\le n}\mathrm{MAC}(B_i)$ and it needs $(n-1)C_{mt}$. The final message authentication code MAC_f needs $2C_e$ and C_m. In EPPI, the computation cost for each power storage units is $2C_p+4C_e+C_{et}$. We present the computation cost in Table 1, where UC, GW, and PSU represent the utility company, the gateway, and a power storage unit, respectively.

Table 1: Computation complexity of EPPI.

	Computation complexity
UC	$(2n+1)C_p + (n+4)C_e + C_{et} + C_m$
GW	$(n+2)C_p + (n+3)C_e + C_m + (n-1)C_{mt} + (n-1)C_a$
PSU	$2C_p + 4C_e + C_{et}$

Table 2: Communication overhead of EPPI.

	Communication overhead (bytes)
UC-to-GW	$0.5k + 89$
GW-to-PSU	$0.5k + 129$
PSU-to-GW	$20k + 129$
GW-to-UC	$40n + 40k + 109$

6.2. Communication Overhead. In the proposed EPPI system, the communications can be divided into four parts, that is, UC-to-GW communication, GW-to-PSU communication, PSU-to-GW communication, and GW-to-UC communication. We assign two bytes for each identity, four bits for each price p_i, five bytes for TS, 20 bytes for q, and 40 bytes for each group element in $\mathbb{G}$ and $\mathbb{G}_T$. We first consider the UC-to-GW communication, where the utility company generates a power collection request *Power_Req* and delivers the request to the gateway. The *Power_Req* packet is of the form $\mathrm{ID}_u \parallel \mathrm{ID}_g \parallel \mathrm{Info}_p \parallel r_u P_0 \parallel \mathrm{TS} \parallel \sigma_u$. Its size should be $k/2 + 89$ bytes if k time slots are adopted. In the GW-to-PSU communication, the *Power_Req* packet is of the form $\mathrm{ID}_g \parallel \mathrm{ID}_u \parallel \mathrm{Info}_p \parallel r_g P_0 \parallel r_u P_0 \parallel \mathrm{TS} \parallel \sigma_u$ and the size is $k/2 + 129$ bytes. In the PSU-to-GW communication, the power request response *Power_Res* packet is of the form $\mathrm{ID}_g \parallel \mathrm{ID}_u \parallel r_i P_0 \parallel \mathrm{TS} \parallel \sigma_i \parallel B_i \parallel \mathrm{MAC}(B_i)$ for the i-th power storage unit. The packet size should be $20k + 129$ bytes. It is noted that, instead of sending n signatures with a total of $56n$ bytes, the aggregated signature needs only 56 bytes for any number of storage units. In the GW-to-UC communication, the response message is of the form $\mathrm{ID}_u \parallel \mathrm{ID}_g \parallel \{r_i P_0\}_{1\le i\le n} \parallel \mathrm{TS} \parallel \sigma \parallel B \parallel \mathrm{MAC}_f$ and the size is $40n+40k+109$ bytes where n represents the number of power storage units. We present the communication cost in Table 2. As shown in Figures 4 and 5, we plot the communication overhead in terms of the time slot number k and the power storage unit number n.

In general, the proposed EPPI scheme is the first secure and privacy-aware power injection scheme and the above analysis indicates that EPPI is efficient in terms of computation and communication cost.

7. Conclusions

Aiming to tackle the security and privacy issues of power injection over AMI and 5G, we propose an efficient and privacy-aware power injection scheme based on 5G smart

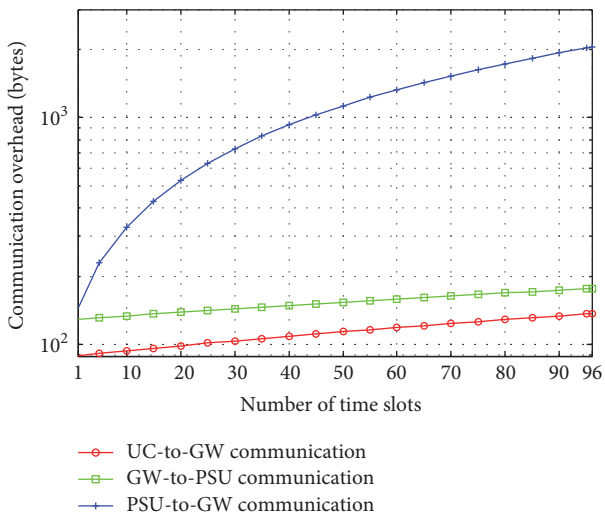


Figure 4: The UC-to-GW, GW-to-PSU, and PUS-to-GW communication overheads of EPPI.

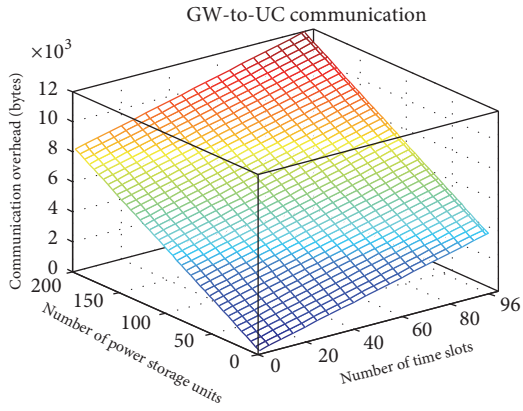


Figure 5: The GW-to-UC communication overhead of EPPI.

grid network slice. The proposed scheme allows the utility company to recover the total amount of collected power and resists any attacker to read individual power injection bid. Each power storage unit blinds its power injection bid, and all the bids will be aggregated by the local gateway based on a novel aggregation technique called *hash-then-addition*. In particular, the utility company can ensure the integrity and authenticity of the collected data. Extensive evaluations indicate that our scheme is secure and privacy-aware and it is efficient in terms of computation and communication cost.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This work is supported by National Natural Science Foundation of China (nos. 61402366, 61472472, and 61272037), Natural Science Basic Research Plan in Shaanxi Province (nos. 2015JQ6236, 2016JM6033, and 2013JZ020), and Scientific Research Program Funded by Shaanxi Provincial Education Department (no. 15JK1686). Yinghui Zhang is supported by New Star Team of Xi'an University of Posts and Telecommunications.

References

[1] N. Alliance, "5 g white paper," White paper, Next Generation Mobile Networks, Frankfurt, Germany, 2015.

[2] G. Locke and P. D. Gallagher, *NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 1.0*, National Institute of Standards and Technology, 2010.

[3] M. Peng, Y. Li, Z. Zhao, and C. Wang, "System architecture and key technologies for 5G heterogeneous cloud radio access networks," *IEEE Network*, vol. 29, no. 2, pp. 6–14, 2015.

[4] M. Peng, S. Yan, and H. V. Poor, "Ergodic capacity analysis of remote radio head associations in cloud radio access networks," *IEEE Wireless Communications Letters*, vol. 3, no. 4, pp. 365–368, 2014.

[5] M. Ali, S. U. Khan, and A. V. Vasilakos, "Security in cloud computing: opportunities and challenges," *Information Sciences*, vol. 305, pp. 357–383, 2015.

[6] J. Li, J. Li, X. Chen, C. Jia, and W. Lou, "Identity-based encryption with outsourced revocation in cloud computing," *IEEE Transactions on Computers*, vol. 64, no. 2, pp. 425–437, 2015.

[7] L. M. Kaufman, "Data security in the world of cloud computing," *IEEE Security and Privacy*, vol. 7, no. 4, pp. 61–64, 2009.

[8] Y. Zhang, X. Chen, J. Li, D. S. Wong, H. Li, and I. You, "Ensuring attribute privacy protection and fast decryption for outsourced data security in mobile cloud computing," *Information Sciences*, vol. 379, pp. 42–61, 2017.

[9] Z. Xia, X. Wang, X. Sun, and Q. Wang, "A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 2, pp. 340–352, 2016.

[10] Z. Fu, K. Ren, J. Shu, X. Sun, and F. Huang, "Enabling personalized search over encrypted outsourced data with efficiency improvement," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 9, pp. 2546–2559, 2016.

[11] J. Li, X. Chen, M. Li, P. P. Lee, and W. Lou, "Secure deduplication with efficient and reliable convergent key management," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 6, pp. 1615–1625, 2014.

[12] J. G. Andrews, S. Buzzi, W. Choi et al., "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[13] N. Nikaein, E. Schiller, R. Favraud et al., "Network store: exploring slicing in future 5G networks," in *Proceedings of the 10th International Workshop on Mobility in the Evolving Internet Architecture (MobiArch '15)*, pp. 8–13, Paris, France, September 2015.

[14] C. Castelluccia, E. Mykletun, and G. Tsudik, "Efficient aggregation of encrypted data in wireless sensor networks," in *Proceedings of the Second Annual International Conference on Mobile and Ubiquitous Systems-Networking and Services (MobiQuitous '05)*, pp. 109–117, July 2005.

[15] D. Westhoff, J. Girao, and M. Acharya, "Concealed data aggregation for reverse multicast traffic in sensor networks: encryption, key distribution, and routing adaptation," *IEEE Transactions on Mobile Computing*, vol. 5, no. 10, pp. 1417–1431, 2006.

[16] X. Lin, R. Lu, and X. S. Shen, "MDPA: multidimensional privacy-preserving aggregation scheme for wireless sensor networks," *Wireless Communications and Mobile Computing*, vol. 10, no. 6, pp. 843–856, 2010.

[17] R. Lu, X. Liang, X. Li, X. Lin, and X. Shen, "EPPA: an efficient and privacy-preserving aggregation scheme for secure smart grid communications," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 9, pp. 1621–1631, 2012.

[18] C. Wu, H. Mohsenian-Rad, and J. Huang, "Vehicle-to-aggregator interaction game," *IEEE Transactions on Smart Grid*, vol. 3, no. 1, pp. 434–442, 2012.

[19] L. Gan, U. Topcu, and S. H. Low, "Optimal decentralized protocol for electric vehicle charging," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 940–951, 2013.

[20] S. Tonyali, O. Cakmak, K. Akkaya, M. M. Mahmoud, and I. Guvenc, "Secure data obfuscation scheme to enable privacy-preserving state estimation in smart grid AMI networks," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 709–719, 2016.

[21] K. Rabieh, M. Mahmoud, K. Akkaya, and S. Tonyali, "Scalable certificate revocation schemes for smart grid AMI networks using bloom filters," *IEEE Transactions on Dependable and Secure Computing*, 2015.

[22] Y. H. Zhang, X. F. Chen, H. Li, and J. Cao, "Identity-based construction for secure and efficient handoff authentication schemes in wireless networks," *Security and Communication Networks*, vol. 5, no. 10, pp. 1121–1130, 2012.

[23] J. Cao, M. Ma, and H. Li, "A group-based authentication and key agreement for MTC in LTE networks," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM '12)*, pp. 1017–1022, Anaheim, Calif, USA, December 2012.

[24] M. M. E. A. Mahmoud and X. Shen, "A cloud-based scheme for protecting source-location privacy against hotspot-locating attack in wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 10, pp. 1805–1818, 2012.

[25] Y. Zhang, X. Chen, J. Li, and H. Li, "Generic construction for secure and efficient handoff authentication schemes in EAP-based wireless networks," *Computer Networks*, vol. 75, pp. 192–211, 2014.

[26] H. Li, X. Lin, H. Yang, X. Liang, R. Lu, and X. Shen, "EPPDR: an efficient privacy-preserving demand response scheme with adaptive key evolution in smart grid," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 8, pp. 2053–2064, 2014.

[27] P. Guo, J. Wang, X. H. Geng, C. S. Kim, and J.-U. Kim, "A variable threshold-value authentication architecture for wireless mesh networks," *Journal of Internet Technology*, vol. 15, no. 6, pp. 929–935, 2014.

[28] M. Mahmoud, N. Saputro, P. Akula, and K. Akkaya, "Privacy-preserving power injection over a hybrid AMI/LTE smart grid network," *IEEE Internet of Things Journal*, 2016.

[29] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Advances in cryptology—(EUROCRYPT '99)*, vol. 1592 of *Lecture Notes in Comput. Sci.*, pp. 223–238, Springer, Berlin, Germany, 1999.

[30] D. Boneh and M. Franklin, "Identity-based encryption from the weil pairing," in *Advances in Cryptology-CRYPTO 2001*, pp. 213–229, Springer, Berlin, Germany, 2001.

[31] W. Stallings, *Cryptography and Network Security: Principles and Practices*, Pearson Education, New Delhi, India, 2006.