CSI Amplitude Fingerprinting-Based NB-IoT Indoor Localization

Qianwen Song, Songtao Guo, *Member, IEEE*, Xing Liu, and Yuanyuan Yang, *Fellow, IEEE*

***Abstract*—With the proliferation of mobile devices, indoor fingerprinting-based localization has caught considerable interest on account of its high precision. Meanwhile, channel state information (CSI), as a promising positioning characteristic, has been gradually adopted as an enhanced channel metric in indoor positioning schemes. In this paper, we propose a CSI amplitude fingerprinting-based localization algorithm in Narrowband Internet of Things system, in which we optimize a centroid algorithm based on CSI propagation model. In particular, in the fingerprint matching, we utilize the method of multidimensional scaling (MDS) analysis to calculate the Euclidean distance and time-reversal resonating strength between the target point and the reference points and then employ the *K*-nearest neighbor (KNN) algorithm for location estimation. By conjugate gradient method, moreover, we optimize the localization error of triangular centroid algorithm and combine the positioning result with MDS and KNN's estimated position to get the final estimated position. Experiment results show that compared to some existing localization methods, our proposed algorithm can effectively reduce positioning error.**

***Index Terms*—Channel state information (CSI), fingerprinting, indoor localization, Narrowband Internet of Things (NB-IoT).**

I. Introduction

RECENTLY, the demands for low-power wide-area (LPWA) machine-type communications (MTCs) have increased dramatically. It is expected that LPWA connections will reach 2 billion in 2020, exceeding the number of traditional cellular users. Narrowband Internet of Things (NB-IoT) released by the Third Generation Partnership Project 3GPP, is a sustainable and emerging 5G radio access technology for connecting billions of devices from a great range of utilities, logistics to the industrial applications. NB-IoT system with "in-band" mode will occupy some sub-bands out of the long term evolution advanced (LTE-A) band. By using as narrow as 180 kHz bandwidth, NB-IoT user equipment (UE) bears only 15% of complexity compared with the normal LTE-A UE. NB-IoT supports super coverage extension, massive number of connections and long user lifetime with low power/cost and low device complexity. As an ultralow complexity and low power consumption technology with super coverage extension and massive number of connections, NB-IoT can be applied in a large range of IoT scenarios including location-based service (LBS), such as smart parking and smart tracking, smart home, etc.

In practice, in LBS, besides NB-IoT technology, there are some short-range communication technologies, such as Bluetooth, Wi-Fi, and Infrared to be applied. However, these short-range communication technologies have still some difficulties in large coverage and massive connections. For example, Bluetooth has slow transmission rate and low transmission range, thus it is difficult to deal with requirements of multi-terminal, high-density and low power. Wi-Fi performs well only under the condition of high power consumption and high equipment costs. Meanwhile, with 2.4 GHz band interference and short transmission distance in the 5 GHz band, Wi-Fi is hard to finish the task of building a smart city. Infrared technology requires the directivity, i.e., the transmitter must be aligned with the receiver without obstruction in the middle [1]. Overall, the characteristics of these short-range communication lack the capacity to connect massive billions of Internet of Things (IoT) devices. On the other hand, these short-range wireless signals are inefficient in a harsh environment, where large coverage is required.

Compared with the above short-range communication technologies, LPWA supports narrowband data transmission with low communication cost, which can either utilize the unlicensed spectrums (e.g., LoRa and SigFox), or 2G/3G/4G cellular licensed resources (e.g., EC-GSM, LTE enhanced MTC, and NB-IoT). LoRa is the first low-cost wide-area implementation for commercial usage, which utilizes chirp spread spectrum as well as Gaussian frequency shift keying modulation to prevent in/out-band interference. The bandwidth required by LoRa can be flexible, varying from 7.8 to 500 kHz [1]. However, LoRa spread spectrum technology may not have sufficient capacity when working as a wide area network. SigFox provides connectivity without deploying specific network infrastructures for each application and the demanding bandwidth is 200 kHz. However, its security is a key issue for

Manuscript received August 11, 2017; revised November 19, 2017; accepted December 7, 2017. Date of publication December 12, 2017; date of current version June 8, 2018. This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant XDJK2016A011, Grant XDJK2015C010, Grant XDJK2015D023, Grant XDJK2016D047, and Grant XDJK 201710635069, in part by the National Natural Science Foundation of China under Grant 61402381, Grant 61503309, Grant 61772432, and Grant 61772433, in part by the Natural Science Key Foundation of Chongqing under Grant CSTC2015JCYJBX0094, in part by the Natural Science Foundation of Chongqing under Grant CSTC2016JCYJA0449, in part by the China Post-Doctoral Science Foundation under Grant 2016M592619, and in part by the Chongqing Post-Doctoral Science Foundation under Grant XM2016002. *(Corresponding author: Songtao Guo.)*

Q. Song, S. Guo, and X. Liu are with the Key Laboratory of Networks and Cloud Computing Security of Universities in Chongqing, College of Electronic and Information Engineering, Southwest University, Chongqing 400715, China (e-mail: songtao_guo@163.com).

Y. Yang is with the College of Electronic and Information Engineering, Southwest University, Chongqing 400715, China, and also with the Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook, NY 11794 USA.

Digital Object Identifier 10.1109/JIOT.2017.2782479

narrowband connection in addition to cost and complexity issues. LTE enhanced MTC as a new type of data communications has relative large data transmission rate ($\leq$ 1Mb/s), low mobility, large coverage range (5 km in urban and 17 km in suburban).

NB-IoT has gradually emerged and highlighted its superiority among LPWA technologies [2]. NB-IoT has the features of low bandwidth, low power consumption, wide coverage, and high network capacity. First, it works in the 800 MHz ultralow frequency band and provides a comprehensive coverage of indoor cellular data connection. Thus, it can cover almost every indoor corner even in the place, where ordinary wireless signals are difficult to reach, such as underground garage and underground shopping malls [3]. This feature enables the NB-IoT-based positioning system to not only improve the positioning accuracy, but also expand the scope of the application. Second, NB-IoT has lower module cost and lower power consumption with the ability to support massive connections compared with short-range communication technologies. At the same time, NB-IoT based on the cellular network can be directly deployed in the existing LTE network, which reduces operator deployment costs. Overall, these inherent features of NB-IoT make it more suitable for the deployment and application of indoor localization since it can improve the robustness, durability and positioning accuracy of localization system.

Considering the advantages of NB-IoT, in this paper, for large indoor scenes with multiple signal transmitters like office buildings, parking lots, shopping malls, we propose an NB-IoT fingerprinting positioning algorithm based on channel state information (CSI) propagation model [4]. For fingerprinting positioning part, first in the offline stage, we utilize CSI amplitudes as fingerprints and store them in the database. In the online stage, we employ a machine learning method, i.e., multidimensional scaling (MDS) analysis [5], [6], to reduce the dimension of CSI and convert the similarity into relative coordinates. In the process of MDS, we take Euclidean distance and time-reversal resonating strength (TRRS) [7] as distance matrix. Euclidean distance is widely used in classical MDS algorithms, whereas TRRS is a physical quantity typically measuring the similarity in CSI amplitudes. TRRS is a deformation of the correlation coefficient and it is superior to the other similarity on the aspect of reflecting the similarity of fingerprints.

After dimension reduction, we choose a deterministic algorithm, i.e., K-nearest neighbor (KNN) algorithm [8] to match target point positions with fingerprints. By KNN algorithm, we obtain the reference point closest to the target point and take the point as the target estimation position. Moreover, we select a small part of targets to obtain another target position to be estimated via an optimized triangular centroid algorithm based on CSI propagation model. Furthermore, we take the average of positioning results based on MDS and centroid algorithm as the final target position. Finally, we implement the proposed MDS-KNN localization system in a representative indoor NB-IoT environment under the current commercial network standard.

In summary, the main contributions of this paper are as follows.

1) We combine fingerprinting system with CSI model to improve the robustness and accuracy of MDS-KNN system. In particular, in the case, where the narrowband signal transmitter (NBST) information is lost or the target has enough line of sight (LOS) information, we select a small number of points to acquire another estimated position.
2) In the fingerprint matching algorithm, we take CSI from NBST in NB-IoT as fingerprints. In order to adapt to the large area positioning scene with multi-NBSTs, we perform dimensionality reduction on the data to reduce computational complexity by MDS.
3) Experimental results reveal that our fingerprinting-based indoor localization system is able to achieve robust and accurate location estimation. Compared with FILA [9] and FIFS [10], our localization scheme has lower 42.9% and 29.1% mean positioning error, respectively.

The rest of this paper is organized as follows. We briefly describe related work on indoor localization in Section II and give preprocessing about CSI in Section III. In Section IV we establish the model of the indoor localization problem and provide the architecture of fingerprinting localization. Afterward, the entire fingerprinting localization system is described in Section V. Experimental results are discussed in Section VI. Finally, we draw the conclusions in Section VII.

II. Related Work

In the existing location-based services, global positioning system (GPS) plays an important role in precise positioning and navigation. However, GPS is no longer the most favorite due to its expensive cost of installation and signal fading in nonline-of-sight (NLOS) transmission path under common indoor environment [11], [12]. Accordingly, the researchers have proposed many GPS-free indoor positioning algorithms [13]. From a methodology perspective, the indoor localization approaches can be commonly classified as two categories: 1) fingerprinting-based approaches [14], [15] and 2) model-based ones [16], [17]. Fingerprinting-based approaches leverage received signal pattern differences for indoor localization, while model-based ones utilize geometrical metrics relative to several known anchors for localization, like the received signal strength (RSS) [18], direction-of-arrival [19], and angle-of-arrival [20].

The model-based approaches mainly focus on the signal propagation in the LOS path, which are more suitable for extensive and open environment, whereas the fingerprinting-based methods consider the fact that multipath propagations are different at each position. Therefore, the database is designed with "multipath-rich" raw signal information in NLOS and takes the advantages of multipath diversity from each position. In the fingerprinting positioning, the researchers generally yield the following two methods: 1) deterministic algorithms [21] and 2) probabilistic algorithms [22]. The main

idea of deterministic algorithms is to calculate several reference points closest to the target and accordingly obtain the coordinates of the target. The main idea of probabilistic algorithms is to utilize probability statistics to acquire the probability that the target appears at each reference point. The representative algorithm is Bayesian probability algorithm.

In the previous fingerprinting positioning system, RSS is commonly utilized as fingerprints owing to its simplicity and low requirement for hardware [14]. For instance, Radar [15], a deterministic algorithm, whose fingerprints are based on RSS, is the first fingerprinting system utilizing radio-frequency (RF) to locate. Following this, Horus [23], a probabilistic method for indoor localization, achieves higher precision than Radar. However, such RSS-based fingerprints have inevitable weakness: they all cannot avoid the effect of multipath fading and shadows. At the normal indoor environment, RSS value changes from -88 dBm to -58 dBm, whose fluctuation range is up to 30 dBm. Even after filtering, the value of the fluctuations can reach 10 dBm [24]. This shows that even a slight change of the signal on a certain path will generate a large effect on the superposition of all links. Besides, RSS contains coarse information so as to not fully utilize the abundant channel information in each subcarrier. Therefore, RSS-based fingerprinting localization methods may have poor positioning performance.

As a new metric of measuring, CSI, a fine-grained physical layer (PHY) information [25] that provides detailed CSI at the subcarrier level, has attracted researchers' attentions. Some existing works have been done to address CSI-based indoor positioning problem. Those studies based on CSI have shown an obviously enhanced accuracy over RSS for indoor location estimation [26], [27]. For example, Jin *et al.* [26] presented how to use a vector of approximated channel impulse response (ACIR) amplitudes as the fingerprint. The radio map stores both ACIR vectors and the kernel function. However, the scheme is entirely based on emulation other than practical condition, which could not guarantee the reliability of their system. In [27] CUPID converts the CSI to the time domain and then approximates the first rectangle in the time domain energy distribution as LOS energy. When the bandwidth is 20 MHz, the first rectangle corresponds to the energy received in the first 50 ns.

Nevertheless, in view of the uncertain time lag, the LOS energy of the real data does not always follow this rule. Wu *et al.* [9] proposed an indoor propagation model, called FILA, which determines the LOS signal components based on a threshold value by CSI instead of RSS. Afterward, the final position is determined by trilateral positioning. However, in this paper, the authors only used the simplest trilateration. In [10], on the frame of CSI-based fingerprinting system, Xiao *et al.* presented a probabilistic algorithm, called FIFS, for enhancing coherent bandwidth along with a correlation filter to match objective with the fingerprints. The probability density function of normal distribution employed by FIFS is suitable for the ideal environment, nevertheless, it does not apply to all scenes due to the complexity of the indoor scenes.

III. Preprocessing of CSI Data

Due to the aforementioned shortcomings of RSS, it is desirable to find a new physical feature to avoid poor performance of RSS in indoor localization. Ideally, the new physical characteristics should have the superiority of two aspects.

1) It should have good resistance with interference of 2.4 GHz frequency band signal and enough stability in the static environment once the disturbed CSI can respond immediate reflections. Consequently, the fingerprint obtained from certain devices at interested positions can be taken as location reference.
2) It should be as fine as possible to distinguish multipath components.

Fortunately, the orthogonal frequency division multiplexing (OFDM) technology assists us to find the new features in NB-IoT meeting the above requirements [28]. The main idea of OFDM is to divide the data stream into multiple substreams and transmit them in parallel through multiple subchannels with different frequencies. Each subchannel uses independent subcarriers and these subcarriers are orthogonal to each other. Through OFDM, we can deduce the CSI from NB signal transmitter to the receiver. Specifically, in view of the frequency domain, CSI is the PHY layer information with a fine-grained attribute value that describes the amplitude and phase of each subcarrier. In the field of narrowband communications, CSI represents the channel property of the communication link, which describes the attenuation of the signal in the process of transmission between the transmitter and receiver, including scattering, environmental attenuation, distance attenuation, and other information.

In a smooth fading channel, the received signal at the receiver in the frequency domain can be expressed as follows:

$$Y = HX + \xi \tag{1}$$

where X and Y denote the narrowband signal vectors at the transmitter and receiver, respectively, ξ indicates the Gaussian white noise vector and H is the complex matrix of CSI [4].

Each group of CSIs describes the amplitude and phase of an OFDM subcarrier

$$H_k = \|H_k\| e^{j\angle H_k} \tag{2}$$

where $\|H_k\|$ and $\angle H_k$ is the amplitude and phase of kth subcarrier, respectively.

Currently, according to IEEE Standard, the number of subcarriers working at 20 MHz is 30. As a result, in the 20 MHz HT transmission mode, CSI is divided into 30 subcarrier groups

$$H = [H_1, H_2, \ldots, H_k, \ldots, H_{30}]^T. \tag{3}$$

In the indoor environment, researchers have established a model for CSI's indoor signal attenuation [8]. Before describing this model, we first describe the key steps of the preprocessing CSI signals. After that, we can obtain CSI data with better stability.

1) Calibrate CSI Data: To be applied to the indoor positioning, the collected CSI data should exclude multipath

interference. In time domain, the overall response of multipath channel $h(\tau)$ is expressed as

$$h(\tau) = \sum_{k=0}^{L_p-1} \alpha_k \delta(\tau - \tau_k) \tag{4}$$

where L_p is the number of multipath channel components, α_k and τ_k denote the amplitude and the time delay of the kth path, respectively [18].

By performing inverse fast Fourier transform conversion on the CSI signal, we can obtain $h(\tau)$ in time domain, which reflects the signal strength under different time delay and thus can roughly distinguish LOS path and multipath components. We set a threshold and filter out excessive impulses in order to get an approximate LOS path.

2) Set Up CSI Model: According to [4], given a packet with 30 groups of subcarriers, the effective CSI of the packet is expressed as

$$\mathrm{CSI_{eff}} = \frac{1}{S}\sum_{s=1}^{S} \frac{f_s}{f_0} \times |H_s|, s \in (-15, 15) \tag{5}$$

where f_s is the frequency of the sth subcarrier, H_s denotes the amplitude of the sth subcarrier CSI, and f_0 indicates the center frequency.

Yang *et al.* [4] gives a new power function representing the relationship between $\mathrm{CSI_{eff}}$ and d as

$$d = \frac{1}{4\pi}\left[\left(\frac{c}{f_0 \times |\mathrm{CSI_{eff}}|}\right)^2 \times \sigma\right]^{\frac{1}{n}} \tag{6}$$

where c is the speed of light, f_0 is the central frequency of 30 groups of subcarriers in a packet, σ is the environment variable and n is the path fading index.

The values of σ and n are different in different indoor environments so that they should be measured according to unique experiment environments. The environment variable σ represents the synthesis including the gain of the transmitter baseband to the RF band, the antenna gain and so on. As σ varies in different environments, n changes in the range [2, 4].

IV. Localization System Architecture

In this section, we first give an overview of our localization system and then describe the critical steps of the proposed localization algorithm.

We consider the square localization area shown in Fig. 1 which represents the indoor environment. This area is divided into N grids, and one reference point will be set in the bottom-left corner of the grid. For ease of understanding, we come up with one reference point as an example on the map. In the localization area, there are L NB signal transmitters (NBSTs) with known locations. There is one user in the whole area whose location is unknown. The user is equipped with an intelligent device. The blue link represents the signal transmission between NBSTs and reference points while the yellow link indicates the signal transmission between NBSTs and users.

Our fingerprinting system is outlined in Fig. 2, whose working process is composed of two phases: 1) offline and 2) online. The two phases correspond to fingerprint establishment and position estimation, respectively.

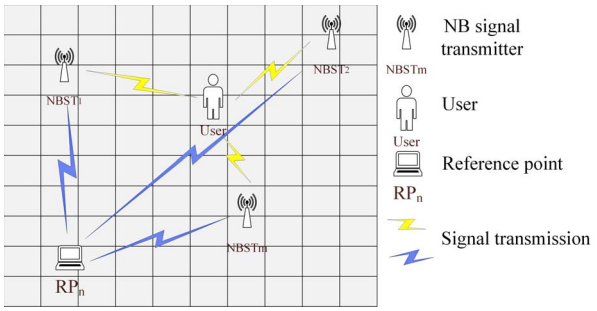

Fig. 1. System model.

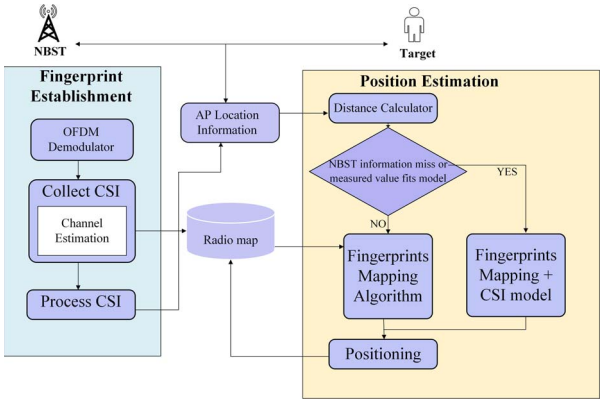

Fig. 2. System architecture.

1) *Fingerprint Establishment:* In the offline phase (fingerprint establishment), on the one hand, we measure CSI data at each reference point and store in the fingerprinting database. On the other hand, in fingerprint processing module, all raw CSI data (tagged with their corresponding known position) are preprocessed: tick the bad data, filter the signal and so on. The fingerprint establishment component serves as the premise of the estimation component.

2) *Position Estimation:* During the online phase (position estimation), when a target is detected, the positioning server receives the location query containing a user's currently measured CSI using mobile device, and an incident report will produce. After receiving the event report, the positioning server will response the user with the estimated location. Specifically speaking, first of all, the processed CSI will be utilized to calculate the distance between NBSTs and the target, preparing for the judging condition, such as whether the target is in LOS path. If CSI fits its model well or the NBST's information is missing, we will combine the CSI model to assist fingerprints mapping algorithm. Otherwise the fingerprints mapping algorithm will work alone. Then the localization engine inquires the database and maps the fingerprint with current real-time collection of CSI using MDS. After that, the target's location can be estimated by KNN and is returned back to the user.

V. CSI-Based Fingerprinting Indoor Localization Algorithm

A. CSI Database Establishment

In this section, we briefly describe how to achieve fingerprint database construction.

In a building with L stationary NBSTs working in NB-IoT and preselected K points with known positions, we obtain J samples (vectors) of CSI measurements for each known position $p \in \mathcal{P}$ ($\mathcal{P} = [p_1, p_2, \ldots, p_k]$). For each sample, we have

$$\mathbf{c}_p(j) = \left[c_p^1(j), c_p^2(j), \ldots, c_p^L(j)\right] \tag{7}$$

where each NBST is identified by the corresponding MAC address and $c_p^l(j)$ denotes the jth CSI of the lth NBST.

For each location $p \in \mathcal{P}$, using the samples $\mathbf{c}_p(j), j = 1, 2, \ldots, J$, we can obtain an average value of CSI from NBST l, $l = 1, \ldots, L$, that is

$$q_p^l(\gamma) = \frac{1}{J}\sum_{j=1}^{J}\left\{c_p^l(j) = \gamma\right\}, k \in (1, 30). \tag{8}$$

Then we can utilize $\{q_p^l, l = 1, \ldots, L, p \in \mathcal{P}\}$ to construct the fingerprint database.

B. Fingerprint-Based Position Estimation

In the online phase, the mobile terminal user at a certain locations collects CSI samples from L different NBSTs. We denote CSI in online phase by $\mathbf{q}(i), i = 1, 2, \ldots, I$, i.e.,

$$\mathbf{q}(i) = \left[q^1(i), q^2(i), \ldots, q^L(i)\right] \tag{9}$$

where $q^l(i)$ denotes the ith CSI at the lth AP.

Considering that for indoor positioning in relatively large scenes, the dimensions of the fingerprints will increase with the number of signal transmitters. In order to reduce the amount of computations and meet the demand of real-time positioning, we first reduce the dimension of fingerprints by MDS algorithm. In the algorithm, we need to construct the matrix consisting of all pairwise measured distances in the network, and the distances, e.g., Euclidean distances, can be used to measure the similarity between fingerprints. In MDS, therefore, we first calculate the similarity of CSI information between each reference point and target point, and then convert the similarity into relative coordinates. Furthermore, we determine the location of the target by the similarity matrix between the online CSI $\mathbf{q}$ and the offline CSI fingerprint $\mathbf{q}_p$ at each known position $p \in \mathcal{P}$.

Therefore, the core step of MDS is how to choose the similarity matrix. In this paper, we adopt Euclidean distance and TRRS [7] to reflect the degree of similarity between CSIs. Euclidean distance is usually used to construct the similarity matrix in the classic MDS, and TRRS is employed to define the similarity between two CSI amplitudes.

1) *Euclidean Distance:* It represents the difference between two items. The closer the distance, the larger the similarity degree. MDS tries to find a subspace R_n, where several objects are embedded, and the similarity of each other is retained as much as possible. The Euclidean distance is defined as

$$d_E = \sqrt{\sum_{l=1}^{L}\left(q_{p_1}^l - q_{p_2}^l\right)^2}. \tag{10}$$

2) *TRRS:* CSI in frequency domain is called channel impulse response (CIR). The TRRS $\eta(\mathbf{h}_1, \mathbf{h}_2)$ between two CIRs $\mathbf{h}_1 = [h_1[0], h_1[1], \ldots, h_1[J-1]]$ and $\mathbf{h}_2 = [h_2[0], h_2[1], \ldots, h_2[J-1]]$ is defined as follows:

$$\eta(\mathbf{h}_1, \mathbf{h}_2) = \frac{\max_i |\mathbf{h}_1 * \mathbf{g}_2(i)|}{\sqrt{\sum_{i=0}^{J-1}|\mathbf{h}_1[i]|^2}\sqrt{\sum_{i=0}^{J-1}|\mathbf{g}_2[j]|^2}} \tag{11}$$

where $\mathbf{g}_2 = [g_2[0], g_2[1], \ldots, g_2[J-1]]$ is defined as the time-reversed and conjugated version of $\mathbf{h}_2$ as follows:

$$g_2[k] = h_2^*[J-1-k], k = 0, 1, \ldots, J-1. \tag{12}$$

In essence, TRRS expresses the cross-correlation of two complex CSIs. But it is more appropriate than the traditional correlation coefficient due to that TRRS is more robust to tolerate the error of channel estimation. In most cases, the positioning system is not fully synchronized, and the correlation coefficient lacking max operation is not fully capable of revealing the real similarity between two fingerprints. With TRRS, the true similarity between fingerprints is better portrayed.

Next, we make a brief introduction of the MDS process. Because there exist L NBSTs, our fingerprint database has L dimensions. Since there are K reference points, the point set with L dimensions to be processed is $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_k]$. Assume that we reduce $\mathbf{Q}$ to n dimensions ($n < L$), labeled as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k]$.

Let $\mathbf{D} = [\hat{d}_{ij}]$ denote the distance matrix of $\mathbf{Q}$, where $\hat{d}_{ij}$ indicates the distance between points $\mathbf{q}_i$ and $\mathbf{q}_j$ for $i \neq j$, and $\hat{d}_{ij} = 0$ for all $i = j$ as follows:

$$\mathbf{D} = \begin{pmatrix} 0 & \hat{d}_{12} & \cdots & \hat{d}_{1L} \\ \hat{d}_{21} & 0 & \cdots & \hat{d}_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{d}_{L1} & \hat{d}_{L2} & \cdots & 0 \end{pmatrix}. \tag{13}$$

The purpose of MDS is to minimize the difference between the distance matrix generated by $\mathbf{Q}$ and $\mathbf{X}$, i.e.,

$$\mathbf{X} = \text{argmin}_X \sum_{i=1}^{L}\sum_{j=1}^{L}\left(\hat{d}_{ij} - \|\mathbf{x}_i - \mathbf{x}_j\|\right)^2 \tag{14}$$

where $\|\cdot\|$ is a vector norm representing the distance between the elements of $\mathbf{X}$. In classical MDS, the distance is denoted by the Euclidean distance.

We define a matrix $\mathbf{T} = \mathbf{X}\mathbf{X}^T$ containing terms like $\hat{d}_{ij} = \mathbf{x}_i\mathbf{x}_j$. By derivation, we get the elements of $\mathbf{T}_{L\times L}$ as

$$t_{ij} = -\frac{1}{2}\left[\hat{d}_{ij}^2 - \frac{1}{L}\sum_{k=1}^{L}\hat{d}_{ik}^2 - \frac{1}{L}\sum_{k=1}^{L}\hat{d}_{kj}^2 + \frac{1}{L^2}\sum_{k=1}^{L}\sum_{l=1}^{L}\hat{d}_{kl}^2\right]. \tag{15}$$

Algorithm 1 Online Location Algorithm

Input: At target position, we collect CSI sequence for each NBST;
The database $\{q_p^l, l = 1, \ldots, L, p \in \mathcal{P}\}$;
$i \in L$: L stationary NBSTs;
$p \in \mathcal{P}$, $\mathcal{P} = [p_1, p_2, \ldots, p_k]$: preselected K RPs;
Output: Estimated location (x_M, y_M) of the target;
set parameters: $\mathbf{q}(i), \mathbf{D}, \mathbf{T}, \mathbf{U}, \Lambda$ and $U', \Lambda', \mathbf{X}, d_E, d_B$;
for $i = 1$ to I **do**
Obtain $\mathbf{q}(i) = [q^1(i), q^2(i), \ldots, q^L(i)]$ by (8);
end for
Construct the matrix $\mathbf{Q}$ containing fingerprints and target's measurement, $\mathbf{Q} = [q_p^l\{l = 1, \ldots, L, p \in \mathcal{P}\}, \mathbf{q}(i)]$;
Compute d_E and $\eta(\mathbf{h}_1, \mathbf{h}_2)$;
Construct the distance matrix $\mathbf{D}$ by d_E and $\eta(\mathbf{h}_1, \mathbf{h}_2)$;
Construct the double centered squared distance matrix $\mathbf{T}$ by (14);
Execute EVD on $\mathbf{T}$ and obtain the eigenvalues $\Lambda = diag[\lambda_1, \lambda_2, \ldots, \lambda_m]$ and the eigenvector matrix $\mathbf{U}$;
Construct the n-dimensional ($n < m$) relative coordinate matrix $\mathbf{X} = \mathbf{U}'\Lambda'^{\frac{1}{2}}$;
Execute the KNN algorithm using $\mathbf{X}$ and estimate the target point position;
Return (x_M, y_M);

As $\hat{d}_{ij}$ is known as the input, t_{ij} can be computed by (15). Note that $\mathbf{T}$ is a positive definite matrix with L positive eigenvalues. We can acquire $\mathbf{X}$ from $\mathbf{U}$ and Λ that can be obtained by the eigen value decomposition (EVD) on $\mathbf{T}$ as follows:

$$\mathbf{T} = \mathbf{U}\Lambda\mathbf{U}^T = \mathbf{X}\mathbf{X}^T \tag{16}$$

where Λ is the diagonal eigenvalues matrix of $\mathbf{T}$, i.e., $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_L)$. $\mathbf{U}$ is the eigenvector matrix.

In order to convert $\mathbf{X}$ from L dimensions to n dimensions ($n < L$), we take n largest eigenvalues $[\lambda_1, \lambda_2, \ldots, \lambda_n]$ from eigenvalue matrix Λ and construct Λ'. Meanwhile, the corresponding eigenvectors in $\mathbf{U}$ is named as $\mathbf{U}'$. Then, we have

$$\mathbf{X} = \mathbf{U}'\Lambda'^{\frac{1}{2}}. \tag{17}$$

After obtaining the n dimension point set $\mathbf{X}$, KNN algorithm will sort these points and achieve the classification by measuring the similarity or distance between different fingerprints. The nearest neighbor calculate the Euclidean distance between the target's CSI from lth NBST and CSI from lth NBST at each fingerprint in database. By averaging the sum of L distances, KNN can obtain K fingerprints closest to the real-time CSI. Then we compute the average of the coordinates of K points whose similarity are closest to the target and take the average coordinate as the estimated position of the target. The steps of fingerprint-based position estimation are listed as Algorithm 1.

C. Improved Localization Algorithm

After completing the positioning of the targets as shown in above sections, we analyze the possible problems in the above algorithm to further improve the positioning accuracy. We consider the following cases.

1) When a positioning area is large, and has quite a few narrowband signal transmitters, a transmitter may not be heard by other transmitters in the area that are remote from the transmitter. For instance, a transmitter is corrupted or becomes invalid due to the location changes of the transmitter. In this case, the reference point in this area may have a similar CSI value, and in the process of MDS, the CSI distance matrix in this area may lead to some estimated errors due to the absence of information about the transmitter.
2) On the other hand, the nearest neighbor found by KNN has the possibility of being scattered everywhere, since the signal fading of each transmitter is not only connected with the distance but also related to the indoor environment factors [29]. Therefore, in the process of KNN, there may exist the estimated error caused by CSI distance matrix.

Fingerprint database stores full NLOS signal information, whereas LOS path information is much enough in larger and open scenes. The subsequent problem is how we can make full use of the LOS information from CSI. To address the problem, we combine an optimized centroid algorithm based on CSI model with fingerprinting algorithm. For the first case without NBST information, the optimized centroid algorithm only needs three NBSTs to obtain positioning information. Thus, we can ensure that the whole positioning system still has good sudden handling capacity and universality, and enhances the robustness of the system. For the second case, the combined CSI model is also able to reduce the error in KNN, for the reason that the error is related to indoor environment factors, whereas environment factors are embodied in the CSI model. Last but not least, CSI model is based on LOS path so that we can take advantage of the abundant LOS information.

In order to reduce the complexity of the algorithm, we refine part of the most suitable targets instead of all the targets to optimize positioning accuracy. In the following, we first describe how to obtain the distance between the target and NBSTs, and then describe the optimized triangular centroid algorithm.

In the preprocessing stage as given in Section III, we have obtained the relationship between CSI_{eff} and the propagation distance d. In the CSI propagation model, we need to measure the indoor environment variables σ and path fading index n according to our environment. After each NBST is fixed, we measure a number of CSIs under the known locations with different propagation distances and calculate the current environment variable σ and path fading index n by (6).

Fig. 3 illustrates our refined propagation model. We can observe from Fig. 3 that the computation result by (6) is basically consistent with the measurement result of the relationship between CSI_{eff} and d. Furthermore, we can find that the model is built on the abundant LOS information, which has complementary advantages with fingerprint system. In order to judge whether the target is in LOS path, we give the following choosing condition (18) to judge how far each measurement value

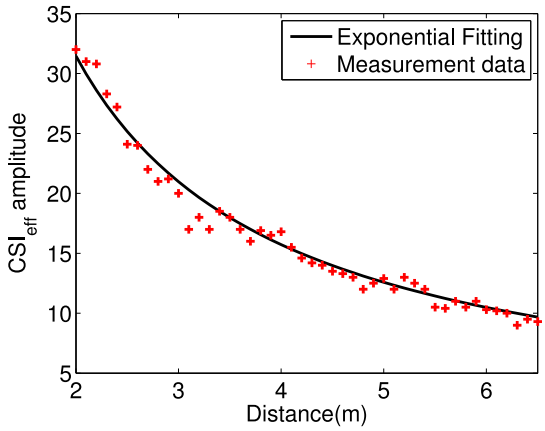

Fig. 3. Relation between $\mathrm{CSI}_{\mathrm{eff}}$ and distance.

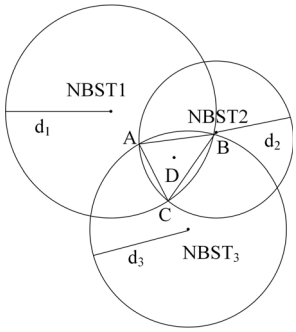

Fig. 4. Triangle centroid algorithm.

deviates from the computation value

$$\sum_{i=1}^{L}\left(\hat{d}_i - d_i\right) \le \varepsilon \tag{18}$$

where d_i is the measured distance between the target and ith NBSTs, $\hat{d}_i$ is the computed distance by (6), L is the number of NBSTs and ε is a threshold. We choose those points satisfying the condition (18) as the ones to be optimized. Depending on precision requirement, ε is different. On the other hand, the NBST points that are not heard are also taken as the ones to be optimized. In the following, we present the optimized triangle centroid algorithm.

Given the distances from the target point to three NBSTs, which are labeled as d_1, d_2, d_3, respectively, we utilize the triangular centroid algorithm [30] to estimate the target position. Fig. 4 shows three circles that NBST1, NBST2, NBST3 are taken as centers and d_1, d_2, d_3 as radius, and their intersection points are denoted as A, B, C. The basic idea of the triangular centroid algorithm is to calculate the coordinates of A, B, C and regard the centroid of ΔABC as the target D. Therefore, the coordinate of target D is given by

$$(x_D, y_D) = \left(\frac{x_A + x_B + x_C}{3}, \frac{y_A + y_B + y_C}{3}\right) \tag{19}$$

where (x_A, y_A), (x_B, y_B), (x_C, y_C) represent the coordinates of A, B, C, respectively.

To reduce the position error in the triangular centroid algorithm, we define the measurement error as the difference between the estimated distance and the measurement distance as follows:

$$f(p)_{p_i} = \sum_{i=1}^{L}(\|p - p_i\| - d_i)^2 \tag{20}$$

where $p = [x, y]^T$ is the estimated position of the target, $p_i = [x_i, y_i]^T$ is the physical position of the ith NBST, d_i denotes the measurement distance between the target and the ith NBST and L indicates the number of NBSTs.

Thus, the optimal position of target p with the minimum measurement error $f(p)$ can be expressed as

$$p = \mathrm{argmin}_p f(p). \tag{21}$$

Then Taylor expansion is carried out at p_0 for each item in (20) to yield

$$\begin{aligned} f(p)_{p_i} &= \|p - p_i\| \\ &\approx f(p_0)_{p_i} + \nabla f(p_0)_{p_i}(p - p_0) \\ &= \nabla f(p_0)_{p_i} p - \left[-f(p_0)_{p_i} + \nabla f(p_0)_{p_i} p_0\right] \end{aligned} \tag{22}$$

where

$$\begin{aligned} \nabla f(p)_{p_i} &= \nabla\left[(p - p_i)^2\right]^{\frac{1}{2}} \\ &= \frac{1}{2}\left[(p - p_i)^2\right]^{-\frac{1}{2}} \times 2(p - p_i) \\ &= \frac{(p - p_i)}{\|p - p_i\|}. \end{aligned} \tag{23}$$

Therefore, the optimal position of target p is simplified as

$$p \approx \mathrm{argmin}_p \sum_{i=1}^{L}\left[\nabla f(p_0)_{p_i} p - \left(-f(p_0)_{p_i} + \nabla f(p_0)_{p_i} p_0\right)\right]^2. \tag{24}$$

We rewrite the above optimization equation in matrix form

$$p^* \approx \mathrm{argmin}_p \sum_{i=1}^{L} \|Ap - B\|^2 \tag{25}$$

where

$$A = \begin{bmatrix} \nabla f(p_0)_{p_1} \\ \nabla f(p_0)_{p_2} \\ \vdots \\ \nabla f(p_0)_{p_L} \end{bmatrix} \tag{26}$$

$$B = \begin{bmatrix} -f(p_0)_{p_1} + \nabla f(p_0)_{p_1} p_0 \\ -f(p_0)_{p_2} + \nabla f(p_0)_{p_2} p_0 \\ \vdots \\ -f(p_0)_{p_L} + \nabla f(p_0)_{p_L} p_0 \end{bmatrix}. \tag{27}$$

We leverage the conjugate gradient method [31] to solve (25), i.e., find a fastest descending path on gradient and converge to a local minimal value eventually in solution space. The conjugate gradient method only needs to use the first derivative information, which not only overcomes the shortcoming of slow convergence, but also avoids the Newton method that needs to store and calculate Hessen matrix.

Algorithm 2 Refinement Algorithm

Input: (x_M, y_M), CSI_{eff};
Output: Estimated location (x, y) of the target after refinement;
Initialize parameters: ε, (x_T, y_T), σ, n, β_k^{HS} and $\delta > 0$, $k = 0$, $p^{(0)} = (x_D, y_D)$;
Compute the distances d_i between the target and ith NBSTs by (6);
if d_i satisfies (18) or a certain NBSTs are not heard **then**
Construct the objective function as (25);
Take the initial direction $s^{(0)} = -\nabla f(p^{(0)})$,
where $\nabla f(p^{(0)})$ represents the gradient of the objective function at point $p^{(0)}$;
$k := 0$;
while $\| \nabla f(x^{(k)}) \| > \delta$ **do**
Determine step size α_k by exact linear search;
$p^{(k+1)} = p^{(k)} + \alpha_k s^{(k)}$;
Determine $s^{(k+1)}$ by

$$s^{(k)} = \begin{cases} -\nabla f(p^{(0)}) & \text{if } k = 0 \\ -\nabla f(p^{(k)}) + \beta_k s^{(k-1)} & k \geq 1 \end{cases}$$

where $\beta_k = \frac{\nabla f(p^k)^T [\nabla f(p^k) - \nabla f(p^{k-1})]}{s^{(K-1)T} [\nabla f(p^k) - \nabla f(p^{k-1})]}$;
$k = k + 1$;
end while
Compute $p^{(k)} = (x_T, y_T)^T$;
Compute $(x, y) = ((x_M, y_M) + (x_T, y_T))/2$
end if
Return (x, y);

Meanwhile, we use the centroid of ΔABC as the initial value p_0. Substituting the objective function into the conjugate gradient, we can obtain (x_T, y_T) and a relatively stable positioning result by averaging (x_T, y_T). Furthermore, we can obtain the positioning result (x_M, y_M) by the MDS algorithm. The processes of the improved localization algorithm are listed as Algorithm 2.

Although we introduce the optimized centroid algorithm based on CSI model, the complexity of the localization algorithm is not increased. This is because: 1) this optimization is not for all positioning points, but only for the points that meet the choosing condition, or that certain NBST information is missing. We can observe from (18) that these points only are a small part of all positioning points by reasonably controlling the threshold and 2) the centroid algorithm can converge after a few iterations. Thus, it does not affect real-time performance of localization algorithm.

VI. Performance Evaluation

In this part, we present the experiment scenario and algorithm evaluation. First, we illustrate the experimental implementation. Then we give an evaluation on the performance of the system compared with FILA [9] and FIFS [10] that act as baselines.

A. System Implementation

The experiments are conducted in a laboratory that lies in a building of Southwest University. We deploy the system in an 12 m × 15 m research laboratory, where computers and books are crowded. This scene is a complicated indoor environment due to most of the LOS paths are sheltered. Fig. 5 shows the layout of the laboratory including the reference/test points. The proposed localization algorithm is implemented on three NBSTs working in NB-IoT whose coordinates are known in advance. By the NB-IoT standard, we let the bandwidth of NB-IoT be 180 kHz and its downlink use OFDMA with 15 kHz of subcarrier spacing. The uplink supports two modes: 15 kHz of multicarrier and 3.75 kHz of single carrier.Meanwhile an intelligent device serves as the receiver equipped with three antennas. We set 42 reference points (red solid circles) to build up the database and 20 test points (green solid triangles). Meanwhile, we collect 100 packets of CSI at each reference point and test point, respectively. During the experiment, there are several students getting up or moving at the same time, which aims to reflect whether the system is robust due to changes of the environment.

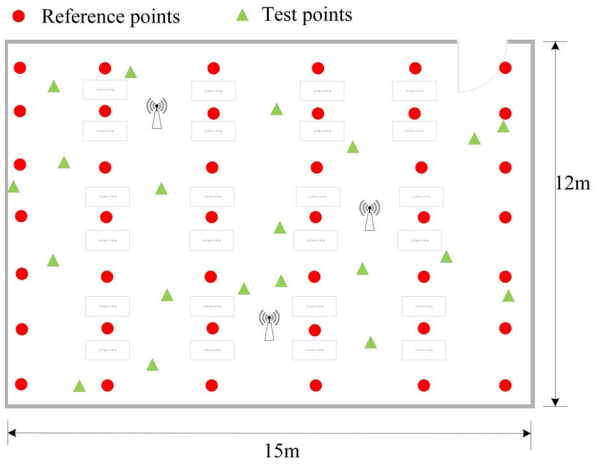

Fig. 5. Layout of the laboratory.

1) Collect CSI Data: The original CSI data is a plural matrix of $m \times n \times s$, where m and n denote the number of antennas of the transmitter and the receiver and s is the number of subcarriers, respectively. Thirty is the number of subcarrier groups and each subcarrier group corresponds to a plural matrix of $m \times n$. During acquiring CSI data, the antennas 1 and 2 of intelligent device are retained at connected state, and the antenna 3 is not connected. Accordingly, the original CSI data is a $2 \times 3 \times s$ complex matrix. We remove the first dimension and get a $3 \times s$ complex matrix. After matrix transpose, we can obtain a $s \times 3$ matrix, where each column data are corresponding to the data from receiving antennas 1–3, respectively.

Fig. 6 depicts CSI's amplitudes from three antennas, which are measured in the laboratory filled with many computers and books with 12 students seating or walking. From the perspective of the antenna, the antenna's frequency, gain, and direction will affect the performance of the antenna. The surrounding environment, including the electromagnetic interference, the surrounding base stations and electronic equipments with the frequency interference will have an impact on the received signal. Based on the consideration, the RSS from the three

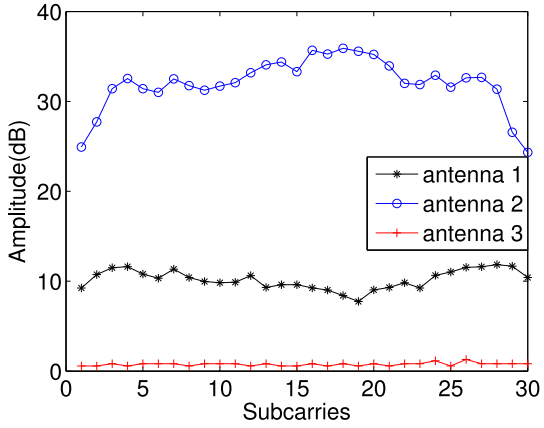

Fig. 6. CSI's amplitudes from three antennas.

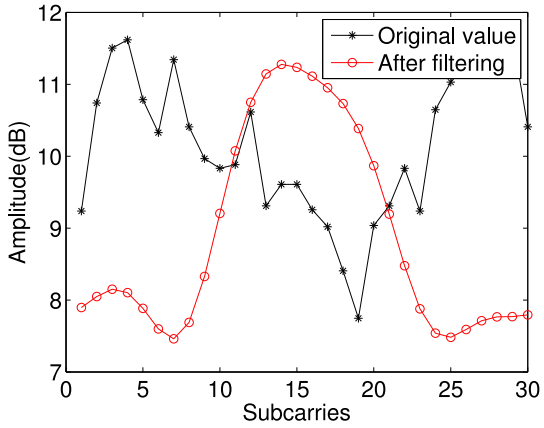

Fig. 7. Amplitude of CSI in frequency domain.

antennas are not the same. Besides, only the first two antennas of intelligent device are connected, and the third antenna is not connected. Thus, there is basically no signal from antenna 3. In the experiment, we collect the data from antenna 1.

In the case of strong interference, the range of CSI amplitudes from three antennas fluctuates between [7 dB, 12 dB], [23 dB, 35 dB], and [0 dB, 1 dB], respectively, while the collected RSS varies from −87 dB to −57 dB. As a consequence, CSI is more robust in comparison with RSS. Fig. 7 shows the amplitude of the CSI signal in frequency domain before and after preprocessing by Section III. After calibrating, CSI data are able to be used in CSI model.

2) Benchmarks and Performance Metric: For comparison purpose, two existing methods are implemented, including FIFS [23] and FILA [10] which are introduced in Section II. We regard laboratory environment as an NLOS scenario because there exist strong blockings like a wall or computers on the paths between NBSTs and the target.

To compare different localization algorithms, we adopt the mean sum error (MSE) as performance metric. Assume that the estimated location of an unknown device is $(\hat{x}, \hat{y})$ while the actual position of the user is (x, y), the MSE of N estimated locations is given by

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(\hat{x}-x)^2+(\hat{y}-y)^2}. \tag{28}$$

TABLE I
PERFORMANCE COMPARISON BETWEEN DIFFERENT ALGORITHMS

algorithm	Mean error(m)	Std.dev.(m)
proposed	1.203	0.8042
FIFS	1.698	0.8366
FILA	2.106	1.1401

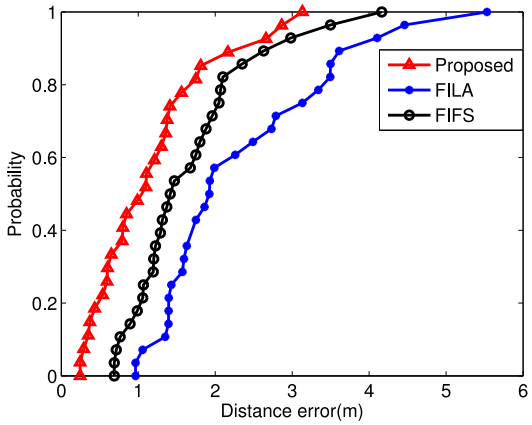

Fig. 8. CDF of distance error with overall experimental results.

In particular, in this section, we use MES as the distance error to evaluate the algorithm performance.

B. Performance Evaluation

The complexity of the laboratory environment has brought a negative impact on the positioning. For instance, abundant multipaths and shadowing effect. Table I shows the MES and the standard deviations (STD) of the proposed system, FILA [9] and FIFS [10]. We can see that our system achieves an MSE of 1.203 m and an STD of 0.8042 m, while FIFS achieves an MSE of 1.698 m and an STD of 0.8366 m and FILA achieves an MSE of 2.106 m and an STD of 1.1401 m. Clearly, the proposed scheme outperforms the other schemes on accuracy and has better robustness with the smallest STD.

Given the same CSI fingerprint database, we evaluate the overall performance of our algorithm, FILA and FIFS. Fig. 8 depicts the cumulative density functions (CDFs) of distance errors for three algorithms. We can observe from Fig. 8 that our algorithm has approximately 50% of target points with the estimation error of less than 1 m, while FIFS is 27% of target points and FILA is 9% of target points. On the whole, our algorithm can ensure that 88% of target points achieve the estimation error under 2 m, while FIFS has 83% of target points to achieve the less than 2 m estimation error. Meanwhile, FILA has only 38% of target points with the estimation error under 2 m.

C. Effect of Different System Parameters

1) Impact of Different Inputs of MDS: Fig. 9 plots the CDF of mean error with different inputs of MDS, i.e., taking Euclidean distance or TRRS as input of MDS, respectively. We can see from Fig. 9 that Euclidean distance or TRRS brings forth different results. For TRRS, more than 75% of

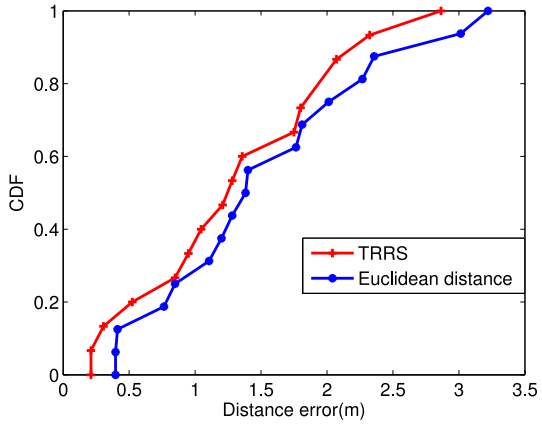

Fig. 9. CDF of localization error with different inputs of MDS.

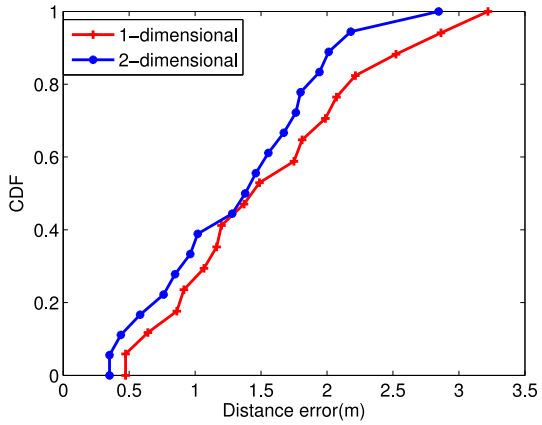

Fig. 10. CDF of localization error versus different outputs of MDS.

target points are within the estimation error of 1.7 m, while Euclidean distance guarantees 75% of target points to have less than estimation error of 2 m. Therefore, TRRS is slightly better than Euclidean distance. This is because that TRRS takes the maximum amplitude between CSIs into account and better characterize the relationship between different sets of CSIs.

2) Impact of Different Outputs of MDS: Fig. 10 illustrates distance localization error versus different outputs of MDS. We reduce the high-dimensional data to 1-D and 2-D data, respectively. It can be seen from Fig. 10 that fewer than 80% of target points achieve a distance error below 2 m, if the data are reduced to 1-D. On the other hand, when the data is reduced to 2-D, approximately 89% of target points have a distance error below 2 m. It is obvious that 2-D data can better reflect the structure and characteristics of the original high-dimensional data while the lower dimension data may lose very important features. Therefore, 2-D data are more accurate and robust than 1-D data.

3) Impact of Different K: To study the impact of K in KNN algorithm, Fig. 11 shows the distance error versus different values of K, i.e., $K = 2, 3, 4$, and 5. We can observe from Fig. 11 that when $K = 3$, our proposed algorithm has over 60% of target points to achieve the estimation error of 1 m and the case of $K = 3$ has the least estimation error than other three cases. When $K = 4$, our algorithm guarantees that about

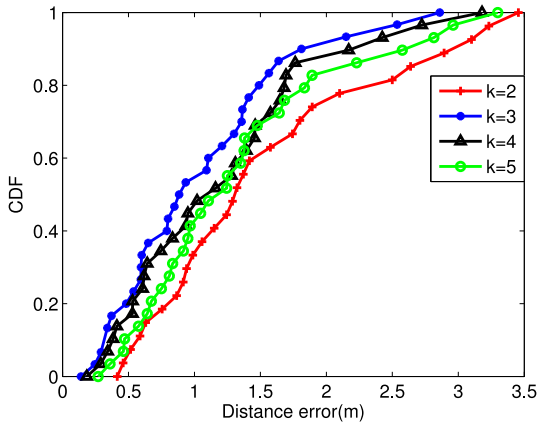

Fig. 11. CDF of distance error with different K.

50% of target points have 1 m localization error, which is similar to $K = 5$. With $K = 2$, the system performs worst due to that 30% of the test points have the estimation error of 1 m. It is noticed that the larger the value of K is, the more errors will be accumulated. In contrast, if K is too small, positioning results will not integrate sufficient effective information so that the estimated position is unstable. Thus, the mean estimated localization error reaches the minimum for the case of $K = 3$.

VII. Conclusion

In this paper, we propose a CSI fingerprinting-based indoor localization algorithm in NB-IoT system. We construct fingerprints with CSI values and further process them by utilizing MDS algorithm so as to reduce their dimensionality. Then we leverage KNN to get the estimated target position. Moreover, we explore the CSI model and the triangular centroid algorithm to further improve the MDS positioning result. In order to reduce the positioning error, we set up an optimization object and utilize the conjugate gradient method to obtain an optimized estimated position. By averaging the results of MDS and optimized estimation, we obtain more stable target position. By real experiment, we validate the proposed localization algorithm in the indoor environment, and experimental results show that our algorithm outperform other two schemes based on CSI, FILA [9], and FIFS [10]. We also examine the effect of different parameters on the positioning algorithm and find that our algorithm can achieve lower positioning errors when TRRS is taken as the input of MDS method and the high-dimension output of MDS can be reduced to 2-D one in the case of $K = 3$ in KNN algorithm.

Nevertheless, there are still some shortcomings in our proposed localization system. For example, the number of NBSTs and RPs is not enough to obtain CSI so that the positioning precision is not too high. Moreover, we do not consider the effect of the diversity of NB-IoT device chips and interference among the NB-IoT devices on positioning precision so that our proposed localization system may be limited in multitarget localization and tracking. In the future, we will further reduce the computational complexity, implement activity recognition, especially, further improve the multitarget localization precision.

REFERENCES

[1] J. Chen *et al.*, "Narrowband Internet of Things: Implementations and applications," *IEEE Internet Things J.*, vol. 4, no. 6, pp. 2309–2314, Dec. 2017.

[2] H. Zhang, J. Liu, and N. Kato, "Threshold tuning-based wearable sensor fault detection for reliable medical monitoring using Bayesian network model," *IEEE Syst. J.*, to be published.

[3] Z. Song *et al.*, "A low-power NB-IoT transceiver with digital-polar transmitter in 180-nm CMOS," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 64, no. 9, pp. 2569–2581, Sep. 2017.

[4] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," *ACM Comput. Surveys*, vol. 46, no. 2, p. 25, 2013.

[5] W. Ni, W. Xiao, Y. K. Toh, and C. K. Tham, "Fingerprint-MDS based algorithm for indoor wireless localization," in *Proc. 21st Annu. IEEE Int. Symp. Pers. Indoor Mobile Radio Commun.*, Sep. 2010, pp. 1972–1977.

[6] G. B. E. Carlos and L. G. Sison, "ALESSA: MDS—Based localization algorithm for wireless sensor networks," in *Proc. 6th Int. Conf. Elect. Eng./Electron. Comput. Telecommun. Inf. Technol.*, vol. 2, May 2009, pp. 856–859.

[7] Z.-H. Wu, Y. Han, Y. Chen, and K. J. R. Liu, "A time-reversal paradigm for indoor positioning system," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1331–1339, Apr. 2015.

[8] L. Demidova and Y. Sokolova, "A novel SVM-kNN technique for data classification," in *Proc. 6th Mediterranean Conf. Embedded Comput. (MECO)*, Jun. 2017, pp. 1–4.

[9] K. Wu, J. Xiao, Y. Yi, M. Gao, and L. M. Ni, "FILA: Fine-grained indoor localization," in *Proc. IEEEINFOCOM*, 2012, pp. 2210–2218.

[10] J. Xiao, K. Wu, Y. Yi, and L. M. Ni, "FIFS: Fine-grained indoor fingerprinting system," in *Proc. 21st Int. Conf. Comput. Commun. Netw. (ICCCN)*, Jul. 2012, pp. 1–7.

[11] S. Adler, S. Schmitt, and M. Kyas, "Path loss and multipath effects in a real world indoor localization scenario," in *Proc. 11th Workshop Positioning Navigation Commun. (WPNC)*, Mar. 2014, pp. 1–7.

[12] B. Han *et al.*, "Shadow fading assisted device-free localization for indoor environments," in *Proc. 8th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2016, pp. 1–5.

[13] P. Yang, "PRLS-INVES: A general experimental investigation strategy for high accuracy and precision in passive RFID location systems," *IEEE Internet Things J.*, vol. 2, no. 2, pp. 159–167, Apr. 2015.

[14] L. Chen, K. Yang, and X. Wang, "Robust cooperative Wi-Fi fingerprint-based indoor localization," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 1406–1417, Dec. 2016.

[15] P. Bahl and V. N. Padmanabhan, "Radar: An in-building RF-based user location and tracking system," in *Proc. 9th IEEE INFOCOM*, vol. 2. Mar. 2000, pp. 775–784.

[16] S. Sioutas *et al.*, "Survey of machine learning algorithms on spark over DHT-based structures," in *Proc. 2nd Int. Workshop Algorithmic Aspects Cloud Comput. (ALGOCLOUD)*, 2016, pp. 146–156.

[17] S. Bozkurt, G. Elibol, S. Gunal, and U. Yayan, "A comparative study on machine learning algorithms for indoor positioning," in *Proc. Int. Symp. Innov. Intell. SysT. Appl. (INISTA)*, Sep. 2015, pp. 1–8.

[18] J. Wang, P. Urriza, Y. Han, and D. Čabrić, "Performance analysis of weighted centroid algorithm for primary user localization in cognitive radio networks," in *Proc. Rec. 44th Asilomar Conf. Signals Syst. Comput.*, 2010, pp. 966–970.

[19] S. Tervo and A. Politis, "Direction of arrival estimation of reflections from room impulse responses using a spherical microphone array," *IEEE/ACM Audio, Speech, Language Process.*, vol. 23, no. 10, pp. 1539–1551, Oct. 2015.

[20] F. O. Akgul and K. Pahlavan, "AOA assisted NLOS error mitigation for TOA-based indoor positioning systems," in *Proc. Mil. Commun. Conf. (Milcom)*, 2007, pp. 1–5.

[21] J. Hou, H. Gao, Q. Xia, and N. Qi, "Feature combination and the kNN framework in object classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1368–1378, Jun. 2016.

[22] Z. Wu *et al.*, "Passive indoor localization based on CSI and Naive Bayes classification," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published.

[23] M. Youssef, "The Horus location determination system," *Wireless Netw.*, vol. 14, no. 3, pp. 357–374, 2008.

[24] L. Xin-Di, H. Wei, and T. Zeng-Shan, "The improvement of RSS-based location fingerprint technology for cellular networks," in *Proc. Int. Conf. Comput. Sci. Service Syst.*, Aug. 2012, pp. 1267–1270.

[25] G. Liu, Y. Li, D. Li, X. Ma, and F. Li, "RoMD: Robust device-free motion detection using PHY layer information," in *Proc. 12th Annu. IEEE Int. Conf. Sens. Commun. Netw. (SECON)*, Jun. 2015, pp. 154–156.

[26] Y. Jin, W.-S. Soh, and W.-C. Wong, "Indoor localization with channel impulse response based fingerprint and nonparametric regression," *IEEE Trans. Wireless Commun.*, vol. 9, no. 3, pp. 1120–1127, Mar. 2010.

[27] S. Sen, J. Lee, K.-H. Kim, and P. Congdon, "Avoiding multipath to revive inbuilding WiFi localization," in *Proc. Int. Conf. Mobile Syst. Appl. Services*, 2013, pp. 249–262.

[28] T. Jiang and Y. Wu, "An overview: Peak-to-average power ratio reduction techniques for OFDM signals," *IEEE Trans. Broadcast.*, vol. 54, no. 2, pp. 257–268, Jun. 2008.

[29] B. Li, J. Salter, A. G. Dempster, and C. Rizos, "Indoor positioning techniques based on wireless LAN," in *Proc. IEEE Int. Conf. LAN*, 2006, pp. 13–16.

[30] Y. Shang, Z. Liu, J. Wang, and X. Xiao, "Triangle and centroid localization algorithm based on distance compensation," in *Proc. IET Int. Conf. Inf. Sci. Control Eng. (ICISCE)*, Dec. 2012, pp. 1–4.

[31] X. Yan *et al.*, "Research on preconditioned conjugate gradient method based on EBE-FEM and the application in electromagnetic field analysis," *IEEE Trans. Magn.*, vol. 53, no. 6, pp. 1–4, Jun. 2017.

Qianwen Song received the B.S. degree in telecommunications engineering from Southwest University, Chongqing, China, in 2015, where she is currently pursuing the master's degree in signal and information processing.

Her current research interest includes indoor localization.

Songtao Guo (M'12) received the B.S., M.S., and Ph.D. degrees in computer software and theory from Chongqing University, Chongqing, China, in 1999, 2003, and 2008, respectively.

He was a Professor with Chongqing University from 2011 to 2012. He is currently a Full Professor with Southwest University, Chongqing. He was a Senior Research Associate with the City University of Hong Kong, Hong Kong, from 2010 to 2011, and a Visiting Scholar with Stony Brook University, Stony Brook, NY, USA, from 2011 to 2012. He has published over 80 scientific papers in leading refereed journals and conferences. He has received many research grants as a Principal Investigator from the National Science Foundation of China and Chongqing and the Post-Doctoral Science Foundation of China. His current research interests include wireless sensor networks, wireless ad hoc networks, and parallel and distributed computing.

Xing Liu received the B.S. degree in computer school from China West Normal University, Nanchong, China, in 2015. She is currently pursuing the master's degree in signal and information processing at Southwest University, Chongqing, China.

Her current research interests include application partitioning and offloading in mobile computing cloud.

Yuanyuan Yang (F'09) received the B.Eng. and M.S. degrees in computer science and engineering from Tsinghua University, Beijing, China, in 1982, and the M.S.E. and Ph.D. degrees in computer science from Johns Hopkins University, Baltimore, MD, USA, in 1989 and 1992, respectively.

She is a Professor of computer engineering and computer science with Stony Brook University, Stony Brook, NY, USA. Her current research interests include wireless networks, data center networks, optical networks, and high-speed networks. She has published over 300 papers in major journals and refereed conference proceedings and holds seven U.S. patents in the above areas.

Prof. Yang has served as an Associate Editor-in-Chief and an Associate Editor for the IEEE TRANSACTIONS ON COMPUTERS and the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS. She has also served as a general chair, program chair, or vice chair for several major conferences and a program committee member for numerous conferences.