EPKF: Energy Efficient Communication Schemes based on Kalman Filter for IoT

Yanqiu Huang, *Member, IEEE,* Wanli Yu, *Student Member, IEEE,* Enjie Ding, and Alberto Garcia-Ortiz, *Member, IEEE*

***Abstract*—The Internet of Things (IoT) has been recognized as the next technological revolution. It faces two challenges: how to achieve energy efficient communication for the battery constrained devices and how to connect a very large number of devices to the Internet with low latency, high efficiency and reliability. To address these problems, this work proposes two methods based on Kalman filter, termed as EPKF (extensions of predicable Kalman filter). They locally reduce the unnecessary transmission (access) of end devices to the network (Internet) utilizing the spatial and temporal correlations with low algorithmic overhead. Each transmitting device (TD) independently controls its transmission using the temporal correlation; and the receiving device (RD) exploits the spatial correlation among the TDs to further improve the reconstruction quality. The reconstruction problem in the RD is nonlinear. To reduce the computation complexity, an in-depth analysis of the local estimate error is conducted and the approximated linear solutions are thereupon obtained. They are fundamental methods applicable to any IoT monitored/controlled physical system that can be modeled as a linear state space representation. The pedestrian-position application is used as a case study to demonstrate the efficiency in the simulation. Remarkably, the EPKF methods using the linear combinations of the local estimates from multiple TDs reduce the transmission rate to 10%, while achieving the same reconstruction quality as using KF in the traditional manner.**

***Index Terms*—Internet of Things (IoT), wireless communication, sensor networks, Kalman filter, state space model, data prediction, state estimate, spatial and/or temporal correlation, multivariate normal distribution.**

I. Introduction

THE emerging Internet of Things (IoT) has been recognized as the next technological revolution, which can play a remarkable role in improving the quality of our daily lives [1]. It refers to uniquely identifiable physical objects and their virtual representations in an Internet-like structure based on sensing, communication, networking and information processing technologies. These objects are typically attached with wireless IoT devices, which enables them to 'see', 'hear' and cooperatively execute tasks [2]. Various applications have already benefited from the IoT, such as industrial automation [3], intelligent agriculture [4], environment monitoring [5], mobile health-care [6], etc.

Y. Huang, W. Yu, and A. Garcia-Ortiz are with the Institute of Electrodynamics and Microelectronics, University of Bremen, 28359 Bremen, Germany (email: huang@item.uni-bremen.de; wyu@item.uni-bremen.de; agarcia@item.uni-bremen.de).

Enjie Ding is with the IoT Perception Mine Research Center, China University of Mining and Technology, Xuzhou 221000, China (email: enjied@cumt.edu.cn).

As the end device is usually powered by a limited battery, energy efficiency is very critical for these applications. For some devices, such as sensor nodes, wireless communication is especially expensive [7]. Besides the energy cost for data packet transmission, additional energy is required by the overhead activities, such as channel listening, communication synchronization, retransmission, etc. Thus, how to achieve energy efficient communication is very important. Moreover, another challenge in IoT is how to connect a very large number of devices to the Internet while satisfying the requirements of latency, energy efficiency and reliability as mentioned in [3]. This is especially critical for the industrial IoT, where WirelessHART and ISA 100.11a standards are designed for the real-world applications in process automation [8]. For addressing these two issues, a promising solution is to appropriately control the transmission of each end device by the local processing techniques.

Kalman filter (KF) has been widely used in many IoT applications for data processing, since the dynamic performance of the physical system monitored/controlled by IoT can be represented mathematically by a state space (SS) model utilizing the physical laws [9]. The available applications include target localization and tracking [10], [11], privacy protection in social network data publishing [12], [13], outliers detection [14], [15], etc. It provides the optimal state estimate from the noisy measurements in the sense of minimum mean square error (MMSE) for a linear dynamical system. However, the traditional way using KF requires the IoT devices to uninterruptedly transmit data, which results in a high communication cost, the competition of the limited bandwidth as well as the transmission latency.

In order to address the above mentioned issues while keeping the functionality of KF, a number of studies have been conducted. For example, the transmitting device (TD) in [16] is required to randomly send the raw data with a Bernoulli distribution and the receiving device (RD) uses a modified KF to estimate the system state. In [17], an improved transmission schedule is proposed, where each TD transmits the raw data following a stochastic decision based on a uniformly distributed random variable. The schedule keeps the Gaussian property and the RD calculates the reconstruction using a variation of KF. As the TD is not required to do any preprocessing in these approaches, the estimate error cumulates due to information loss. To compensate for the inaccurate estimate caused by the missed data, a local KF is suggested to be executed in the TD. Once the local estimate

is received, the lost information can still be inferred. For instance, DKF (dual KFs) is presented in [18], where a pair of KFs are synchronously executed in the TD and RD to predict the output estimate, respectively. The TD uses an additional KF to remove the measurement noise and produce the system output. When the prediction error is larger than a given threshold, the output estimate is transmitted to the RD. This method is further improved by PKF (predictable KF) [19], in which the dual KFs are replaced by a pair of $k-$step ahead predictors with much lower complexity as introduced in Section II. It provides the optimal reconstruction solution with minimum mean square error (MMSE) when using the transmitted information of a single TD [20] and the explicit function of the trade-off between transmission rate and reconstruction quality is derived in [21].

However, the above mentioned techniques have not considered the spatial correlation among the neighboring TDs, which can be utilized to further reduce the number of transmissions of each device (or equivalently the number of active devices at each time step). One common way to exploit the spatial correlation is through the information exchange. For example, in [22], each TD exchanges measurements with its neighbors and estimates the system state using a KF, based on its own and the received data. Although the estimate quality is increased, extra energy is consumed for communication, which is not preferred considering the high wireless communication cost. To avoid intra-communication, a coordinator is usually required. As proposed by [23], each TD executes a local KF and the remote estimator schedules the transmissions of the local estimates. A similar work is [24], where a centralized scheduling is proposed to determine the transmission of each TD to minimize the average estimation error considering multiple linear time-invariant stochastic processes with the constraints of packet length and bandwidth. These approaches increase the computation complexity of the RD. An interesting method is presented in [25]. It utilizes the spatio-temporal correlation without intra-communication and any coordinator. Each TD estimates the system state locally and schedules its transmission independently; the RD improves the estimate quality by using the transmitted data from multiple TDs. However, it requires the RD to calculate a time variant measurement covariance matrix and the reduced order system model online, which is computation intensive.

This work proposes two methods based on KF, termed as EPKF (extensions of predictable KF), to reduce the reconstruction complexity of the RD required by [25]. Each TD executes a PKF [19] to estimate the system state locally and control its own transmission. The RD further exploits the spatial correlation to improve the reconstruction quality. It is proved to be sufficient that the RD only calculates the linear combinations of the local estimates of the TDs with the stored coefficients. The main contributions of the paper are summarized as follows:

a) This work analyzes the effect of applying the spatial correlation on the improvement of estimation accuracy in the general cases using the entropy theory. For a system monitored by multiple devices, the amount of information provided by all TDs represents the spatial correlation. The higher it is, the less estimate error can be achieved.
b) Based on the analysis, two EPKF methods are proposed to reduce the unnecessary transmission (access) of end devices to network (Internet) utilizing the spatial and temporal correlations with low algorithmic overhead. Compared with the traditional way using KF, where the uninterrupted transmission of a single IoT device is required, the EPKF methods reduce the number of transmissions while achieving the same reconstruction quality; different from the KF-based methods for transmission rate reduction, EPKF methods avoid intra-communication and enable each TD to independently schedule the transmission without a coordinator.
c) Moreover, an in-depth analysis regarding the error distribution is conducted, which extends the analysis in [21] to a higher dimension and provides a strong theoretical support for the proposed methods. The error is produced at different Markov states and it satisfies a truncated multivariate normal (MVN) distribution over parallelograms determined region at each state. To the best of our knowledge, this is the first work that provides the calculation of the error covariance by using the Hessian matrix of the probability density function (PDF) of MVN and the matrix transformation.
d) Furthermore, the EPKF methods are the fundamental approaches applicable to any IoT monitored/controlled physical system that can be characterized by a linear state space model. For demonstrating the efficiency, the pedestrian-position application is used as a case study. Remarkably, using the linear combinations of the local estimates is sufficient for reducing the number of transmissions with the guaranteed estimate accuracy.

The rest of this paper is organized as follows. Section II formulates the reconstruction problem when the RD utilizes the preprocessed data of multiple TDs. Section III presents the linear solutions after the deep analysis of the error distribution. The efficiency of the proposed approaches is estimated using an IoT application in Section IV. Finally, we conclude the work in Section V.

II. Problem Description

This section presents the system model and analyzes the effect of using spatial correlation on the improvement of estimate accuracy. After that, the local processing approach of each TD is presented and the reconstruction problem in the RD is formulated.

The state space (SS) representation is widely adopted to characterize the dynamic performance of a physical system monitored/controlled by IoT. Many linear SS models have been derived by utilizing the physical laws of mechanical, electrical, fluid and thermodynamic systems [9], for example, the heat dynamics of a building [26], speed or position of DC motor used in many industrial control systems [27], outlet temperature of the plate heat exchanger in the industrial engineering [28], and pipeline flow [29], etc. When such a

system is monitored by m devices, the SS model satisfies:

$$x_k = A_k x_{k-1} + B_k u_{k-1} + w_{k-1} \tag{1}$$
$$z_k^i = H_k^i x_k + v_k^i \tag{2}$$

where x_k is the state vector, u_k is the system control input and z_k^i is the measurement from device i ($1 \leq i \leq m$, m is the number of monitoring devices); A_k is the state transition matrix, B_k is the control input matrix and H_k^i is the observation matrix; $w_k \sim \mathcal{N}(0, Q_k)$ and $v_k^i \sim \mathcal{N}(0, R_k^i)$ characterize process and measurement noise, respectively. They are mutually independent, i.e., $E[w_k v_j] = 0, \forall k, j$. The system is assumed to be controllable and observable.

As the measurements of the TDs come from the same system, there exists spatial correlation for estimating the system state. The amount of information provided by all TDs represents the spatial correlation. The higher it is, the less estimate error can be achieved. For example, when only the measurements from TD i, $\mathbb{Z}_k^i = [z_1^i, z_2^i, \cdots z_k^i]$, is used for state estimation, it is proved by [30] that the optimal estimate $\hat{x}_k$ based on $\mathbb{Z}_k^i$ minimizes the estimate error entropy, which is determined by the mutual information between x_k and $\mathbb{Z}_k^i$. Mathematically, we have:

$$\begin{aligned} H(\hat{e}_k|\mathbb{Z}_k^i) &= H(x_k) - I(x_k, \hat{x}_k) \\ &= H(x_k) - I(x_k, \mathbb{Z}_k^i) \\ &= H(x_k|\mathbb{Z}_k^i). \end{aligned} \tag{3}$$

where $\hat{e}_k = \hat{x}_k - x_k$; $H(\cdot)$ denotes the entropy of the random variable and $I(\cdot, \cdot)$ is the mutual information between two random variables. They are determined by the probability density functions (PDFs) and the joint PDFs of the random variables, respectively, which are associated with the system parameters from time 1 to k. The second equality holds, because the optimal estimate $\hat{x}_k$ utilizes all of the information provided by $\mathbb{Z}_k^i$ resulting in $I(x_k, \hat{x}_k) = I(x_k, \mathbb{Z}_k^i)$. The larger $I(x_k, \mathbb{Z}_k^i)$ is, the smaller $H(\hat{e}_k|\mathbb{Z}_k^i)$ can be achieved. The uncertainty of the system state is reduced to the conditional entropy $H(x_k|\mathbb{Z}_k^i)$ as depicted in Fig. 1a. Further exploiting $\mathbb{Z}_k^j$, the error entropy is reduced to $H(\hat{e}_k|\mathbb{Z}_k^i, \mathbb{Z}_k^j)$. The value depends on how much additional information is provided by $\mathbb{Z}_k^j$ to the system state when the information provided by $\mathbb{Z}_k^i$ is used. Let $I(x_k, \mathbb{Z}_k^j|\mathbb{Z}_k^i)$ denote this additional information. It can be calculated by:

$$I(x_k, \mathbb{Z}_k^j|\mathbb{Z}_k^i) = I(x_k, \mathbb{Z}_k^j) - I(x_k, \mathbb{Z}_k^i, \mathbb{Z}_k^j).$$

In the general case as shown in Fig. 1b, part of information in $I(x_k, \mathbb{Z}_k^j)$ can be known from $\mathbb{Z}_k^i$ and this information is no more useful to reduce the $H(x_k)$. Then the spatial correlation is the total amount of information reflecting the system state, $I(x_k, \mathbb{Z}_k^i) + I(x_k, \mathbb{Z}_k^j|\mathbb{Z}_k^i)$. Using this correlation, $H(\hat{e}_k|\mathbb{Z}_k^i, \mathbb{Z}_k^j)$ becomes:

$$\begin{aligned} H(\hat{e}_k|\mathbb{Z}_k^i, \mathbb{Z}_k^j) &= H(x_k) - I(x_k, \mathbb{Z}_k^i) - I(x_k, \mathbb{Z}_k^j|\mathbb{Z}_k^i) \\ &= H(\hat{e}_k|\mathbb{Z}_k^i) - I(x_k, \mathbb{Z}_k^j|\mathbb{Z}_k^i). \end{aligned} \tag{4}$$

As $I(x_k, \mathbb{Z}_k^j|\mathbb{Z}_k^i) \geq 0$, we have $H(\hat{e}_k|\mathbb{Z}_k^i, \mathbb{Z}_k^j) \leq H(\hat{e}_k|\mathbb{Z}_k^i)$. The equality holds if and only if $I(x_k, \mathbb{Z}_k^j) = I(x_k, \mathbb{Z}_k^i, \mathbb{Z}_k^j)$, i.e., all information provided by TD j can be known from TD i. Except this degenerate case, the error entropy can be definitely reduced by using spatial correlation.

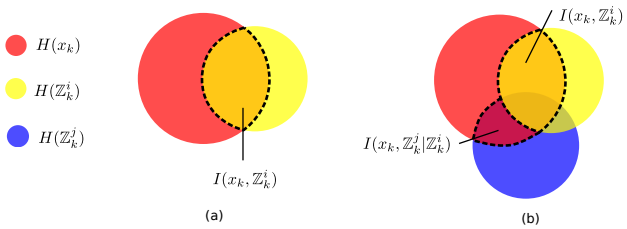


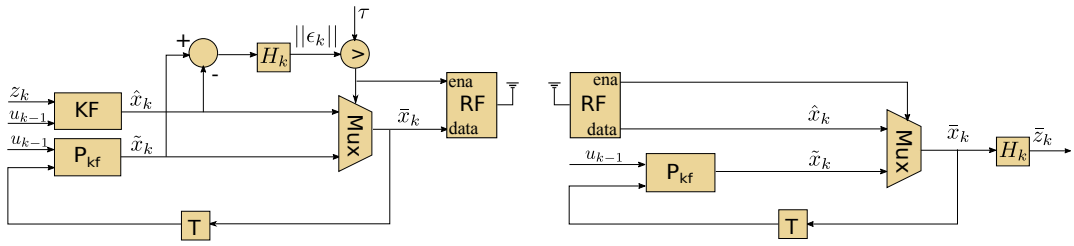
Fig. 1. The estimate error entropy using and without using the spatial correlation: (a) the estimate error entropy without using spatial correlation is $H(\hat{e}_k|\mathbb{Z}_k^i) = H(x_k) - I(x_k, \mathbb{Z}_k^i)$; (b) the spatial correlation equals $I(x_k, \mathbb{Z}_k^i) + I(x_k, \mathbb{Z}_k^j|Z_k^i)$ in the general case and the estimate entropy using the spatial correlation is $H(\hat{e}_k|\mathbb{Z}_k^i, \mathbb{Z}_k^j) = H(x_k) - I(x_k, \mathbb{Z}_k^i) - I(x_k, \mathbb{Z}_k^j|\mathbb{Z}_k^i)$, where $I(x_k, \mathbb{Z}_k^j|\mathbb{Z}_k^i) = I(x_k, \mathbb{Z}_k^j) - I(x_k, \mathbb{Z}_k^i, \mathbb{Z}_k^j)$ is the additional information provided by TD j .

Based on the above analyzed relationship between estimate error entropy and spatial correlation, we are allowed to reduce the transmission rate of TD j, while achieving the same estimate error entropy as using only the information of TD i. Considering the communication cost and limited network resources as mentioned in Section I, each TD can reduce the number of transmissions to the RD; after that, the RD further utilizes the spatial correlation to improve the estimate quality.

In this work, each TD executes the PKF method [19], which combines a predictor with a KF, to control its own transmission. PKF provides the optimal reconstruction solution when using the transmitted information of a single TD as analyzed in [20]. Fig. 2 depicts the block diagrams of the PKF encoder (PKF-en) running in the TD and the PKF decoder (PKF-de) running in the RD. The TD runs a local KF with Eqs. (5) to (9) to produce the state estimate $\hat{x}_k$ while removing the measurement noise. The RD predicts $\hat{x}_k$ with a simple predictor P_{kf}, Eq. (10), and produces the predicted state, $\tilde{x}_k$. To guarantee the prediction quality, the TD follows the prediction of the RD and calculates the prediction error $\epsilon_k = H_k(\tilde{x}_k - \hat{x}_k)$. If $\|\epsilon_k\|$ exceeds a given threshold, τ, the TD sends its local estimate $\hat{x}_k$ and the RD replaces the prediction as formulated in Eq. (11). The reconstructed signal in the RD is Eq. (12).

$$\hat{x}_k^- = A_k \hat{x}_{k-1} + B_k u_{k-1} \tag{5}$$
$$P_k^- = A_k P_{k-1} A_k^T + Q_k \tag{6}$$
$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} \tag{7}$$
$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H_k \hat{x}_k^-) \tag{8}$$
$$P_k = (I - K_k H_k) P_k^- \tag{9}$$
$$\tilde{x}_k = A_k \bar{x}_{k-1} \tag{10}$$
$$\bar{x}_k = \begin{cases} \tilde{x}_k, & \text{if } ||\epsilon_k|| = ||H_k(\tilde{x}_k - \hat{x}_k)|| \leq \tau \\ \hat{x}_k, & \text{otherwise} \end{cases} \tag{11}$$
$$\bar{z}_k = H_k \bar{x}_k \tag{12}$$

After intermittently receiving the local estimates of m ($m \in \mathbb{Z}^+, m > 1$) TDs, the RD further utilizes the spatial correlation to improve the reconstruction quality. More specifically, let $\hat{\boldsymbol{X}}_{\boldsymbol{k}}^{\boldsymbol{i}} = [\hat{x}_1^i, \cdots, \hat{x}_k^i]$ denote the local estimates of TD i. Under



The diagram of PKF-en in the TD

The diagram of PKF-de in the RD

Fig. 2. The block diagrams of PKF encoder (PKF-en) in the TD and PKF decoder (PKF-de) in the RD.

the control of PKF, the RD receives a subset of $\hat{\boldsymbol{X}}_{\boldsymbol{k}}^{\boldsymbol{i}}$ from each TD, i.e., $\hat{\boldsymbol{X}}_{\boldsymbol{j}}^{\boldsymbol{i}} \subset \hat{\boldsymbol{X}}_{\boldsymbol{k}}^{\boldsymbol{i}} \quad (j \leq k, 1 \leq i \leq m)$. Then, the best state estimate that can be produced by the RD is:

$$\bar{\hat{x}}_k = E[x_k | \hat{\boldsymbol{X}}_{\boldsymbol{j}}^{\boldsymbol{1}}, \cdots, \hat{\boldsymbol{X}}_{\boldsymbol{j}}^{\boldsymbol{m}}] \tag{13}$$

and the estimate of system output for TD i at time k is:

$$\bar{\hat{z}}_k^i = H_k^i \bar{\hat{x}}_k. \tag{14}$$

We aim to solve this problem in the next sections.

III. EPKF Methods: Communication Cost Reduction Schemes Based on KF

In Eq. (13), $\hat{\boldsymbol{X}}_{\boldsymbol{j}}^{\boldsymbol{i}}$ consists of two components: the transmitted local estimates and the accuracy indication of the prediction. Each element in the former component is a Gaussian random variable. While the element in the latter is a Boolean indicator: it equals 0 if the RD receives data, indicating that the prediction is inaccurate; otherwise, 1. It means that the reconstruction problem is nonlinear and therefore hard to be solved by the energy constrained devices. This section aims to approximate the nonlinear information hidden in the Boolean indicator and provides linear solutions for Eq. (13).

As known from [31], when TDs randomly transmit the local estimates (with Bernoulli distribution), the MMSE reconstruction solution in the RD can be obtained by using a KF with an augmented system model Eq. (15) and the observation model Eq. (16). This method is termed as Rand-ST. The models are obtained by combining Eq. (1) and Eq. (8).

$$X_k = F_k X_{k-1} + W_k \tag{15}$$

where

$$F_k = \begin{bmatrix} A_k & \cdots & \mathbb{0} \\ \vdots & \ddots & \vdots \\ K_k^m H_k^m A_k & \cdots & (I - K_k^m H_k^m) A_k \end{bmatrix};$$
$$W_k = \left[w_k, \cdots, K_k^m H_k^m w_k^m + K_k^m v_k^m \right]^T.$$

$$Z_k = C_k X_k \tag{16}$$

where $C_k = \begin{bmatrix} \mathbb{0} & I_k \end{bmatrix}$ and I_k is the identity matrix whose size varies to select the transmitted local estimates at time k. If all TDs continuously transmit their local estimates, I_k becomes time invariant I, which is a $nm \times nm$ identity matrix. Then $C_k = C = \begin{bmatrix} \mathbb{0} & I \end{bmatrix}$.

As the untransmitted data of PKF, i.e., Boolean indicator is 1, indicates that the prediction $\tilde{x}_k^i$ is close to $\hat{x}_k^i$. The RD can take $\tilde{x}_k^i$ as the updated $\hat{x}_k^i$ with a deviation. If the deviation satisfies a normal distribution, Eq. (13) can be solved linearly using Rand-ST by adding the Gaussian noise V_k in Eq. (16) and the new model becomes:

$$Z_k = C X_k + V_k. \tag{17}$$

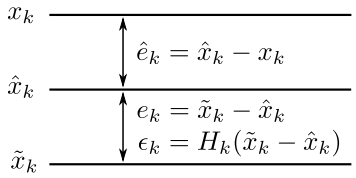


Fig. 3. The schematic diagram of the error definitions.

For this purpose, this section analyzes the deviation between $\bar{x}_k^i$ and $\hat{x}_k^i$. Based on the analysis, it further investigates the possible reconstruction solutions. In the following description, we let m_o denote the original model, i.e., Eqs. (1) and (2), and m_a denote the augmented model, i.e., Eqs. (15) and (17).

A. Analysis of local reconstruction error

Let $\bar{e}_k = \bar{x}_k - \hat{x}_k$ denote the deviation between the local estimate and the reconstruction. We analyze its distribution in two steps: firstly, the distribution of k-step ahead prediction error of the state, $e_k = \tilde{x}_k - \hat{x}_k$, is obtained; after that, we further study the distribution of $\bar{e}_k$ based on the relationship between $\bar{e}_k$ and e_k. The diagram of the notations is shown in Fig. 3. The following analysis extends the one in [21] to a higher-dimension. We assume that the KF enters to the steady state, i.e., $K_{k\to\infty} = K$, $P_{k\to\infty}^- = P^-$ and $P_{k\to\infty} = P$, and the subscripts (time index) of the system parameters are ignored for conciseness. As the analysis is identical for each node, the superscript i is neglected as well.

1) Prediction error: Let $\hat{e}_k = \hat{x}_k - x_k$ denote the *a posteriori* estimate error of KF. Combining Eqs. (1), (2), (5), (8) and (10), we can recursively calculate the k-step ahead prediction error as:

$$\begin{aligned} e_k &= Ae_{k-1} + KHA\hat{e}_{k-1} - KHw_k - Kv_k \\ &= Ae_{k-1} - K\delta_k \end{aligned} \tag{18}$$

where $\delta_k = -HA\hat{e}_{k-1} + Hw_k + v_k = z_k - H\hat{x}_k^-$ is the innovation and satisfies the normal distribution with zero mean and covariance $(HP^-H^T + R)$, i.e., $\delta_k \sim \mathcal{N}(0, HP^-H^T + R)$. This equation indicates that the k-step ahead prediction

error of the state is the linear transform of the error from $k-1$ step with the transition matrix A plus the current innovation weighted by the Kalman gain.

As $e_0 = 0$, e_k can be obtained in a compact form by simplifying Eq. (18) as:

$$\begin{aligned} e_k &= -\sum_{i=0}^{k-1} A^i K \delta_{k-i} \\ &= \alpha_k \vec{\boldsymbol{\delta}}_{\boldsymbol{k}} \end{aligned} \tag{19}$$

where $\alpha_k = [-A^{k-1}K, \cdots, -AK, -K]$ and $\vec{\boldsymbol{\delta}}_{\boldsymbol{k}} = [\delta_1, \cdots, \delta_{k-1}, \delta_k]^T$ is the vector of k innovations. e_k is linearly transformed from k innovations by the matrix α_k. As each δ_k is an i.i.d. normal distributed random variable, their linear combinations are still normally distributed, which indicates that $\vec{\boldsymbol{\delta}}_{\boldsymbol{k}}$ satisfies a MVN distribution [32]. The mean of $\vec{\boldsymbol{\delta}}_{\boldsymbol{k}}$ is a zero vector. The covariance matrix is denoted as $\boldsymbol{\Sigma}_{\vec{\boldsymbol{\delta}}_{\boldsymbol{k}}}$, which is a diagonal matrix with $HP^-H^T + R$ on its main diagonal. The mean of e_k is 0 and its covariance can be calculated using Eqs. (7) and (19) as:

$$\theta_k = E[e_k e_k^T] = \sum_{i=0}^{k-1} A^i P^- H^T K^T (A^T)^i.$$

Thus, when $k = 1$, $e_1 = -K\delta_1$, where the error in each dimension is linearly dependent; when $k > 1$, e_k satisfies a joint normal distribution with zero mean and covariance θ_k, i.e., $e_k \sim \mathcal{N}(0, \theta_k)$. For illustration, we arbitrarily create a system with $n = 2$ and $m = 1$ to depict the distribution of e_k. The system parameters are listed in Table I. In this case, e_k only has two dimensions, which makes it easy to observe the distribution. Figs. 4a and 4b depict the density contour of e_k and $\vec{\boldsymbol{\delta}}_{\boldsymbol{k}}$ when $k = 2$, respectively. Both of them satisfy joint normal distributions, i.e., $\vec{\boldsymbol{\delta}}_{\boldsymbol{k}} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{\vec{\boldsymbol{\delta}}_{\boldsymbol{k}}})$ and $e_k \sim \mathcal{N}(0, \theta_k)$ as analyzed above. The directions of the principal axes of the ellipsoids are determined by the eigenvectors of the covariance matrix. There is no correlation between the two dimensions of $\vec{\boldsymbol{\delta}}_{\boldsymbol{2}}$. The linear transformation of $\vec{\boldsymbol{\delta}}_{\boldsymbol{2}}$ by the matrix α_2, $\alpha_2 \vec{\boldsymbol{\delta}}_{\boldsymbol{2}}$, coincides with e_2. This confirms our former analysis.

2) Reconstruction error: We further analyze the relationship between the k-step ahead prediction error, e_k, and the reconstruction error, $\bar{e}_k$, and then obtain the distribution of $\bar{e}_k$. The idea is to firstly model the process of PKF as a Markov chain, then decompose $\bar{e}_k$ into the prediction error produced at each Markov state. Through the analysis of the error at each state, the distribution of $\bar{e}_k$ can be obtained. The complete description regarding the modeling of PKF using the CDF of ϵ_k as a Markov chain is presented in [20], [21], where the transition probability from state k to $k+1$, $p_{k|k+1}$, the steady state probability over state k, p_k, and the transmission probability of the node have been derived. Therefore, this work only presents a concise formulation to find the relationship between e_k and $\bar{e}_k$.

According to the definition of e_k and Eq. (19), the k-step ahead prediction error of the system output ϵ_k in Eq. (11) can be calculated by:

$$\epsilon_k = He_k = \beta_k \vec{\boldsymbol{\delta}}_{\boldsymbol{k}} \tag{20}$$

where $\beta_k = H\alpha_k$. Depending on whether ϵ_k lies in the interval $[-\tau \quad \tau]$, there are two outcomes of the reconstructed state $\bar{x}_k$: prediction $\tilde{x}_k$ or local KF estimate $\hat{x}_k$. The outcome at time k is independent of the one in other time. More specifically, assuming $\bar{x}_k = \hat{x}_k$, it does not affect the likelihood of $\bar{x}_{k+1} = \hat{x}_{k+1}$ or $\bar{x}_{k+1} = \tilde{x}_{k+1}$. The sequence of reconstructions, $\bar{x}_1, \bar{x}_2, \bar{x}_3, \cdots$, is a random process. The outcome of each random variable $\bar{x}_k$ corresponds to either a *success* when it is $\tilde{x}_k$, or a *failure* otherwise. Let Υ_n denote the number of the most recent consecutive *successes* that have been observed at the nth trial. If the nth trial is a *failure*, then $\Upsilon_n = 0$; if trial numbers n, $n-1$, $n-2, \cdots, n-m+1$ are all *successes* but trial number $n-m$ is a *failure*, then $\Upsilon_n = m$. The collection of $\{\Upsilon_1, \Upsilon_2, \Upsilon_3, \cdots\}$ is thereby a stochastic process. Assuming $\Upsilon_n = k$ at the nth trial, then Υ_{n+1} equals either $k+1$ or 0 at the next trial regardless of the values $\Upsilon_1, \cdots, \Upsilon_{n-1}$. It means that the random process satisfies the Markov property and can be modeled as a discrete-time Markov chain as shown in Fig. 5.

The reconstruction error $\bar{e}_k$ is composed of the error generated at each Markov state, $\bar{e}_k|\Upsilon_n = 1, \bar{e}_k|\Upsilon_n = 2, \cdots$. Finding the distribution of $\bar{e}_k$ is equivalent to study the distribution of the error at each state. From the former analysis, at the Markov state k, the number of consecutive *successes* is k, which indicates that all the past k prediction errors of the system output satisfy the accuracy requirement, i.e., $\boldsymbol{v\epsilon_k} \in \boldsymbol{R_k(\tau)}$, where $\boldsymbol{v\epsilon_k} = [\epsilon_1, \cdots \epsilon_k]$ is the vector of k prediction errors of the system output and $\boldsymbol{R_k(\tau)}$ is the region of k-dimensional space bounded by the threshold $[-\tau \quad \tau]$ in each dimension. Then, $\bar{e}_k|\Upsilon_n = k$ indicates that the k-step ahead prediction errors of the system state satisfy the accuracy restriction, i.e., $e_k\big|\boldsymbol{v\epsilon_k} \in \boldsymbol{R_k(\tau)}$. There is no error at state 0; when $k \geq 1$, the error at state k has the truncated joint normal distribution. Fig. 4a shows the distribution of $e_k\big|\boldsymbol{v\epsilon_k} \in \boldsymbol{R_k(\tau)}$, when $k = 2$ in the former example. To the best of our knowledge, there is no solution so far to calculate the covariance of joint normal distributed variables truncated over a parallelogram determined region. The direct calculation is nontrivial because the bounds are linear functions of the variables. We solve this problem with the help of matrix transformation.

According to Eq. (19) and Eq. (20), we know that e_k and ϵ_k are linearly transformed from $\vec{\boldsymbol{\delta}}_{\boldsymbol{k}}$ by the matrices α_k and β_k, respectively. In order to find the distribution of $e_k|\boldsymbol{v\epsilon_k} \in \boldsymbol{R_k(\tau)}$, we can firstly study $\vec{\boldsymbol{\delta}}_{\boldsymbol{k}}|\boldsymbol{v\epsilon_k} \in \boldsymbol{R_k(\tau)}$, then transform the results by α_k. With the help of Eq. (20), $\vec{\boldsymbol{\delta}}_{\boldsymbol{k}}$ can be represented by $\boldsymbol{v\epsilon_k}$ to eliminate the variable:

$$\boldsymbol{v\epsilon_k} = \boldsymbol{B_k} \vec{\boldsymbol{\delta}}_{\boldsymbol{k}} \tag{21}$$

where $\boldsymbol{B_k} = [\beta_1, \cdots, \beta_k]^T$ is a lower triangular square matrix. For example, when $k = 2$, $\boldsymbol{B_2} = \begin{bmatrix} -HK & \mathbb{0} \\ -HAK & -HK \end{bmatrix}$. It has full rank mk and is invertible. Thus,

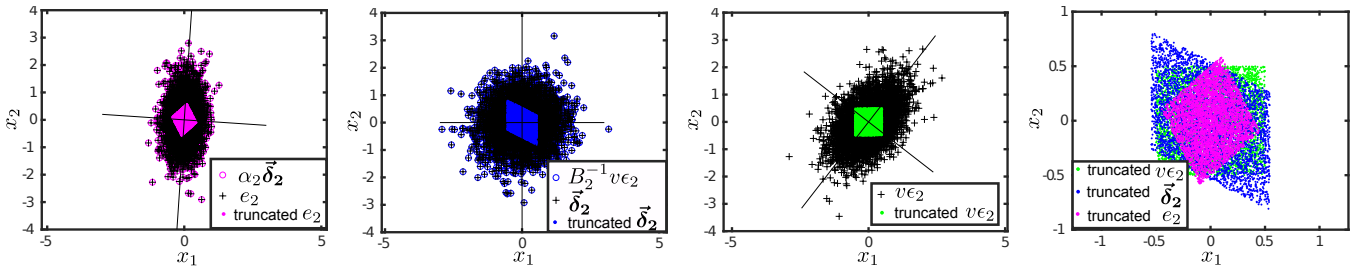
$$\vec{\boldsymbol{\delta}}_{\boldsymbol{k}} = \boldsymbol{B_k^{-1}} \boldsymbol{v\epsilon_k}. \tag{22}$$

Then the problem is further converted to study the distribution of $\boldsymbol{v\epsilon_k}|\boldsymbol{v\epsilon_k} \in \boldsymbol{R_k(\tau)}$.

As analyzed in [20] and [21], $\boldsymbol{v\epsilon_k}$ has the MVN distribution

TABLE I
A RANDOMLY GENERATED SYSTEM WITH THE FOLLOWING PARAMETERS FOR ILLUSTRATING THE ERROR DISTRIBUTION.

A	B	H	R	Q
[0.2237, 0.6533; -0.1705, 1.1203]	[0; 0]	[-0.8660, 1]	0.0403	[0.2471, -0.0052; -0.0052, 0.2527]



(a) (b) (c) (d)

Fig. 4. The joint normal distribution of the errors, their truncated distribution over the region and the linear transformation among them seen from the density contour when $k = 2$ and $\tau = 0.5$ as an example: (a) the distribution of e_k and $e_k|\boldsymbol{v\epsilon_k} \in \boldsymbol{R_k(\tau)}$; e_k is linear transformed from $\vec{\boldsymbol{\delta}}_{\boldsymbol{k}}$ by α_k; (b) the distribution of $\vec{\boldsymbol{\delta}}_{\boldsymbol{k}}$ and $\vec{\boldsymbol{\delta}}_{\boldsymbol{k}}|\boldsymbol{v\epsilon_k} \in \boldsymbol{R_k(\tau)}$; $\vec{\boldsymbol{\delta}}_{\boldsymbol{k}}$ is transformed from $\boldsymbol{v\epsilon_k}$ by $\boldsymbol{B_k^{-1}}$; (c) the distribution of $v\epsilon_k$ and $v\epsilon_k|v\epsilon_k \in \boldsymbol{R_k(\tau)}$; (d) the transformation among three truncated errors.

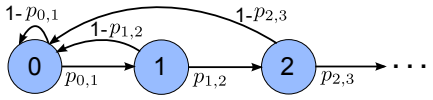


Fig. 5. Description of PKF using success-runs Markov chain.

with zero mean and a new covariance $\boldsymbol{S_k}$, where

$$\boldsymbol{S_k} = \begin{bmatrix} H\theta_1 H^T & \cdots & H\theta_1 A^{k-1^T} H^T \\ \vdots & \ddots & \vdots \\ HA^{k-1}\theta_1 H^T & \cdots & H\theta_k H^T \end{bmatrix}. \tag{23}$$

Fig. 4c depicts the distribution of $\boldsymbol{v\epsilon_k}$ when $k = 2$. The principles axes are determined by the eigenvectors of $\boldsymbol{S}_2$. As illustrated in Fig. 4b, the transformation of $\boldsymbol{v\epsilon_k}$ by the matrix $\boldsymbol{B_k}$ coincides with $\vec{\boldsymbol{\delta}}_{\boldsymbol{2}}$. This is consistent with Eq. (22).

Consequently, $\boldsymbol{v\epsilon_k}|\boldsymbol{v\epsilon_k} \in \boldsymbol{R_k(\tau)}$ has a truncated MVN distribution. Its probability density function satisfies:

$$f_{\boldsymbol{v\epsilon_k}}(\boldsymbol{\chi}, \boldsymbol{R_k(\tau)}) = \frac{\varphi_{\boldsymbol{v\epsilon_k}}(\boldsymbol{\chi})}{\Phi_{\boldsymbol{v\epsilon_k}}(\boldsymbol{R_k(\tau)})} \tag{24}$$

where

$$\varphi_{\boldsymbol{v\epsilon_k}}(\boldsymbol{\chi}) = \frac{1}{\sqrt{(2\pi)^k |\boldsymbol{S_k}|}} \exp\left(-\frac{1}{2}\boldsymbol{\chi}^T \boldsymbol{S_k}^{-1}\boldsymbol{\chi}\right) \tag{25}$$

is the probability density function (PDF) of $\boldsymbol{v\epsilon_k}$; $\Phi_{\boldsymbol{v\epsilon_k}}(\tau) = \Pr\left(\boldsymbol{v\epsilon_k} \in \boldsymbol{R_k(\tau)}\right) = \int_{\boldsymbol{R_k(\tau)}} \varphi_{\boldsymbol{v\epsilon_k}}(\boldsymbol{\chi})\, d\boldsymbol{\chi}$ is the probability that $\boldsymbol{v\epsilon_k}$ locates in the region $\boldsymbol{R_k(\tau)}$. The square area in Fig. 4c depicts the distribution of the truncated $\boldsymbol{v\epsilon_k}$, i.e., $\boldsymbol{v\epsilon_k}|\boldsymbol{v\epsilon_k} \in \boldsymbol{R_k(\tau)}$, when $k = 2$ and $\tau = 0.5$. Transforming it with the matrix $\boldsymbol{B_k^{-1}}$, we can obtain the distribution of truncated $\vec{\boldsymbol{\delta}}_{\boldsymbol{k}}$, $\vec{\boldsymbol{\delta}}_{\boldsymbol{k}}|\boldsymbol{v\epsilon_k} \in \boldsymbol{R_k(\tau)}$, as shown in Fig. 4b; further transforming the results using α_k, the distribution of truncated e_k, $e_k|\boldsymbol{v\epsilon_k} \in \boldsymbol{R_k(\tau)}$, is depicted in Fig. 4a. In Fig. 4d, we compare the distribution of these three truncated errors. For $\boldsymbol{v\epsilon_2}$, the region is restricted by a rectangle in this case; while for δ_2 and e_2, the regions are determined by parallelograms.

The mean of $e_k|\boldsymbol{v\epsilon_k} \in \boldsymbol{R_k(\tau)}$ is a zero vector. The covariance is nontrivial to calculate directly, since the lower and upper bounds of the integral become linear functions of the variables. However, from the above analysis, this problem can be solved with the help of matrix transformation. As reported in [20], [21], the covariance of doubly truncated MVN over rectangles determined region can be derived using the Hessian matrix of Eq. (25). It means we can calculate the covariance of $\boldsymbol{v\epsilon_k}|\boldsymbol{v\epsilon_k} \in \boldsymbol{R_k(\tau)}$, $\bar{\boldsymbol{S}}_{\boldsymbol{k}}$. Combining this result with Eqs. (22) and (19), the covariance of truncated e_k over the parallelogram determined region can be obtained by:

$$\mathrm{cov}\left(e_k|\boldsymbol{v\epsilon_k} \in \boldsymbol{R_k(\tau)}\right) = \alpha_k \boldsymbol{B_k^{-1}} \bar{\boldsymbol{S}}_{\boldsymbol{k}} \boldsymbol{B_k^{-1}}^T \alpha_k^T. \tag{26}$$

This is also the covariance of $\bar{e}_k$ at state k, $\mathrm{cov}(\bar{e}_k|\Upsilon_n = k)$, as mentioned before. Thus, the overall covariance of $\bar{e}_k$ is:

$$\mathrm{cov}(\bar{e}_k) = \sum_{k=0}^{\infty} p_k \,\mathrm{cov}(\bar{e}_k|\Upsilon_n = k) \tag{27}$$

where $p_k = p_0 \Phi_{\boldsymbol{v\epsilon_k}}(\tau)$ is the steady state probability over state k and

$$p_0 = \frac{1}{1 + \sum_{k=1}^{\infty} \Phi_{\boldsymbol{v\epsilon_k}}(\tau)} \tag{28}$$

is the transmission rate.

In summary, the reconstruction error $\bar{e}_k$ consists of the errors generated at each Markov state. At each state, the error $\bar{e}_k|\Upsilon_n = k$ is linearly transformed from a truncated k-variate normal distribution over rectangles determined region to a truncated n-variate normal distribution over parallelograms determined region. The direct calculation of the covariance of such distribution is nontrivial and there is no related solution so far in the literature. Here we solve it by Eq. (26) with the help of matrix transformation and obtain the overall covariance of $\bar{e}_k$ by Eq. (27).

B. EPKF methods

As the deviation $\bar{e}_k^i$ does not satisfy a normal distribution, we aim to approximate it and provide the linear solutions for

Eq. (13) based on the above analysis.

The direct approximation is to take all $\bar{e}_k^i$ together as a normal distribution. The mean is zero and the covariance can be calculated by Eq. (27). Then by using Rand-ST with $\text{cov}(V_k^i) = \text{cov}(\bar{e}_k^i)$, we can obtain the reconstruction solution. However, as known from the analysis, $\bar{e}_k^i$ has different distribution at each state. A more elaborate way is to differentiate the errors at each state. The error generated at each state $\bar{e}_k|\Upsilon_n = k$ is approximated as a normal distributed noise, with zero mean and covariance $\text{cov}(\bar{e}_k|\Upsilon_n = k)$ calculated by Eq. (26). This requires an indicator n_k^i to denote which state is the reconstruction $\bar{x}_k^i$ from. The RD then uses Rand-ST with the time variant covariance $\text{cov}(V_k^i|n_k^i = k) = \text{cov}(\bar{e}_k|\Upsilon_n = k)$ to produce the expanded state estimate $\tilde{X}_k$. The estimate of the original system corresponds to $\bar{\bar{x}}_k = S_l\tilde{X}_k$, where $S_l = [I \quad \mathbb{0}]$ and I is the identity matrix to select the original state. The reconstructed observation for each node can be calculated using Eq. (13). This method is termed as EPKF-norm and its diagram is shown in Fig. 6.

However, the reconstruction process of EPKF-norm can be remarkably simplified in practical scenarios. The idea is to directly let $\bar{x}_k^i = \hat{x}_k^i$, i.e., $\bar{e}_k^i = 0$. This is motivated by three reasons: firstly, the transmission rate decreases exponentially as the threshold increases as indicated from Eq. (28); secondly, $\bar{e}_k$ bounded by the threshold distributes more tightly to the center compared with e_k; thirdly, the KF with augmented model in EPKF-norm is essentially to weight the prediction from the system model and the reconstruction from each TD, respectively. We explain them in detail as follows: as analyzed in Section III-A2, the distributions of e_k and $\boldsymbol{v\epsilon_k}$ are determined by the system model. Given a threshold, the more loose the error distributes, the higher probability that it locates outside the threshold, which results in a higher transmission rate. In other words, for different system models, the threshold has to increase as S_k increases to achieve the same transmission rate, which results in a higher $\bar{e}_k$. In this sense, it seems that $\bar{e}_k$ cannot be ignored anymore. However, it is not the absolute value of $\bar{e}_k$ determines whether it can be neglected but the relative value w.r.t. e_k, as the KF with m_a is essentially to weight the predicted states from the system model and the reconstructions from each TD in EPKF-norm. The relative distribution between e_k and $\bar{e}_k$ determines the weights. According to Eq. (28), the transmission rate decreases exponentially as the threshold increases. A relative small threshold compared with S_k (more specifically S_1) can achieve a high transmission rate reduction. Then $\bar{e}_k$ bounded by the threshold distributes much closer to the center than e_k. This is illustrated by Fig. 4a. Thus, compared with e_k, $\bar{e}_k$ can be neglected.

The above analysis indicates that we can approximate $\text{cov}(V_k) = 0$ in Eq. (17) to reduce the computation complexity. There is no need to have the indicator n_k^i in the RD and the KF with m_a can be replaced by a linear combination of local reconstructions of each node with the corresponding coefficient C_k^i. This simplified method is called EPKF-simp. The main idea behind EPKF-simp is that instead of using the complete KF equations to update the augmented states, it only aims to find the required operations related with the update

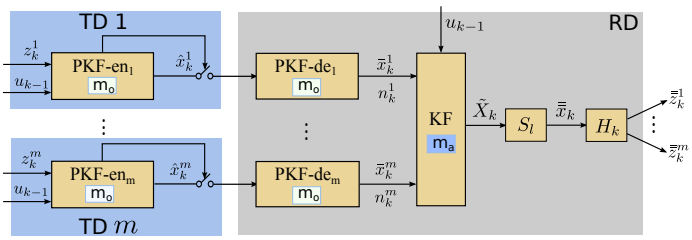


Fig. 6. The diagram of EPKF-norm that $\bar{e}_k^i$ is approximated as normal distribution at each state. Each TD runs a PKF with the original system model m_o, Eqs. (1) and (2), to control the transmission of the local estimates and the RD uses a KF with the augmented models m_a, Eqs. (15) and (17), to improve the estimate, where $\text{cov}(V_k^i|n_k^i = k) = \text{cov}(\bar{e}_k^i|\Upsilon_n = k)$ in Eq. (17) depending on the state indicator n_k^i.

of the original system state. We derive it in the following paragraph.

To differentiate the notations from the former section, let $\ddot{X}_k^- = [\bar{\bar{x}}_k^-, \bar{x}_k^{1-}, \cdots, \bar{x}_k^{m-}]^T$ denote the *a priori* estimate of the augmented system state using Eq. (5). As $\text{cov}(V_k) = 0$, the *a posteriori* estimate is $\ddot{X}_k = [\bar{\bar{x}}_k, \bar{x}_k^1, \cdots, \bar{x}_k^m]$, where the last m elements are the reconstructions of each node. Thus, the updated covariance of $\ddot{X}_k$ can be written as follows:

$$\ddot{P}_k = \begin{bmatrix} \sigma_k & \mathbb{0} \\ \mathbb{0} & \mathbb{0} \end{bmatrix}$$

where σ_k corresponds to the updated estimate covariance of x_k using the augmented model m_a. The *a priori* estimate covariance using Eq. (6) satisfies:

$$\ddot{P}_k^- = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

where

$\Sigma_{11} = A_k\sigma_k A_k^T + Q_k$ is the a priori estimate covariance of the original states;
$\Sigma_{12} = \begin{bmatrix} \Sigma_{11}{H_k^1}^T {K_k^1}^T, & \cdots, & \Sigma_{11}{H_k^m}^T {K_k^m}^T \end{bmatrix}$;
$\Sigma_{21} = \Sigma_{12}^T$;
The ith diagonal elements of Σ_{22} is $\Sigma_{22}(i,i) = K_k^i H_k^i \Sigma_{11} {H_k^i}^T {K_k^i}^T + K_k^i R_k^i {K_k^i}^T$ and the (i,j) entry is $\Sigma_{22}(i,j) = K_k^i H_k^i \Sigma_{11} {H_k^j}^T {K_k^j}^T$.

Because of the special shape of the observation matrix C and the zero measurement noise, the Kalman gain calculated by Eq. (7) is actually:

$$K_k = \begin{bmatrix} \Sigma_{12} & \Sigma_{22} \end{bmatrix}^T \Sigma_{22}^{-1}$$

For updating the original state, only the upper part of K_k is needed, which results in the following coefficient vector:

$$\begin{bmatrix} C_k^1 & \cdots & C_k^m \end{bmatrix} = \Sigma_{12}\Sigma_{22}^{-1} \tag{29}$$

Thus, the *a posteriori* estimate of the original state combining Eq. (5) is:

$$\begin{aligned} \bar{\bar{x}}_k &= \bar{\bar{x}}_k^- + C_k^1(\bar{x}_k^1 - \bar{x}_k^{1-}) + \cdots + C_k^m(\bar{x}_k^m - \bar{x}_k^{m-}) \\ &= \phi_k \ddot{X}_{k-1} + \sum_{i=1}^m C_k^i \bar{x}_k^i \end{aligned} \tag{30}$$

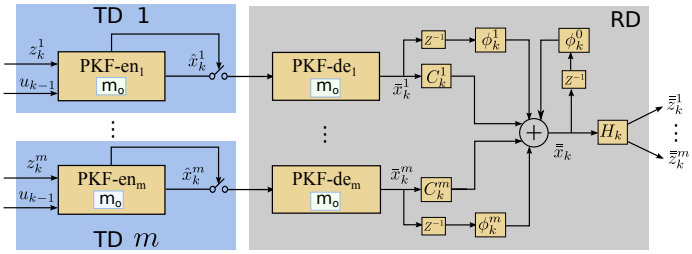


Fig. 7. The diagram of EPKF-simp that $\bar{e}_k^i$ is approximated as zero. Each TD runs a PKF with the original system model m_o, Eqs. (1) and (2), to control the transmission of the local estimates and the RD uses the linear combination of the local reconstructions with different weights C_k^i to improve the estimate.

where $\phi_k = \begin{bmatrix} I & -C_k^1 & \cdots - C_k^m \end{bmatrix} F_k$. The final estimated observation of each node is calculated using Eq. (14). The diagram of EPKF-simp is shown in Fig. 7.

The approaches based on the combined system model are not suitable for quickly change systems. The system matrix F_k requires the local KF gains of each node K_k^i at each time step. It brings synchronization overhead. However, the implementation complexity can be significantly reduced for time invariant systems, where the KF gain and the estimate covariance converge after several steps. The transition matrix F_k is thereby a constant, which can be easily calculated offline. The RD stores this parameter and only needs to receive $\hat{x}_k^i$ and compute Eq. (30) online.

IV. Numerical Analysis

This section estimates the performance of EPKF methods. As mentioned before, they are fundamental approaches applicable to any IoT monitored/controlled physical system that can be modeled as a linear SS model. Here we use the pedestrian-position application presented in [10] as a case study for illustration purpose. As shown in Fig. 8, a number of Bluetooth nodes are deployed on the ceiling of the shopping mall and a smartphone carried by the pedestrian runs the dead reckoning algorithm. In this application, the Bluetooth node is the TD and the smart phone is the RD. The estimate metric is the trade-off between transmission rate and the covariance of the positioning error.

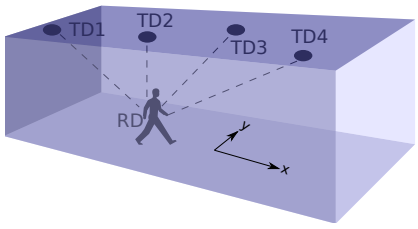


Fig. 8. The schematic for pedestrian positioning application presented in [10].

A Kalman filter with the linear SS model in Eqs. (1) and (2) is used to fuse the results from the inertial sensors of the phone and the measurements from the TDs, considering the limited processing ability and the energy constraints of the smartphone. The step length, L_k, is obtained from the inertial sensors, which acts as the control input, u_k, in Eq. (1); the received signal strength (RSS) from the TD i is used to obtain the coordinates of the pedestrian based on the Bluetooth propagation model, which is the observation, z_k^i, in Eq. (2). The state vector x_k represents the (x, y) coordinates of the pedestrian at time k; the transition matrix A is an identity matrix. Assuming the pedestrian walks towards the x-axis, then the control matrix $B_k = [1,\ 0]^T$ determines the effect of u_{k-1} on x_k; and the inaccuracy of the step length estimate is characterized by the noise w_k whose covariance is 0.04 in our experiment. The observation z_k^i is the measurement of the x-axis from node i and $H_k^i = [1, 0]$. The measurement noise v_k^i represents the uncertainty of the Bluetooth propagation model. The noise covariances of TD1 to TD4 are 0.0625, 0.25, 0.5625 and 1, respectively.

The original algorithm presented in [10] only utilizes the observations from a single TD and this TD needs to transmit uninterruptedly for real time positioning. When the RD receives RSS from multi-nodes, it selects the closest one for updating the predicted coordinates from the system model. Taking the scenario in Fig. 8 for example, the RD selects the RSS from TD2 for positioning. The covariance of the estimate error in this case is 0.0807 m^2. This is the best result that can be achieved by using the data from a single node for estimate, which is called TD2-limit in Fig. 9b. For reducing the transmission rate while compensating the effect of missed data on the estimate accuracy, TD2 could run a PKF-en for preprocessing. It produces the local estimate $\hat{x}_k$ and follows the prediction of RD to guarantee the prediction error within the user defined threshold τ. Here we assign different values to τ and the tradeoff between transmission rate and estimate accuracy is shown in Fig. 9b. When the transmission rate is 1, the estimate accuracy is the same as the result from [10]; as the transmission rate decreases, the estimate accuracy decreases. In other words, the communication cost is reduced at the expense of the decreased estimate quality.

In order to improve the estimate accuracy while keeping the same communication cost of TD2, the RD can further exploit the data from other three TDs. The best estimate is obtained by Rand-ST [31] using all data from four TDs. It is called all-limit in Fig. 9b. For enabling EPKF-norm, we approximate $\bar{e}_k$ at each state as a normal distribution. The covariance is calculated using Eq. (26). To enable EPKF-simp, we approximate the reconstruction error as zero. Fig. 9b demonstrates the efficiency of EPKF methods. Compared with [10], they improve the positioning accuracy by 67.77%, when all data of TDs are used (i.e., transmission rate equals 1), which is equivalent to use Rand-ST [31]. They achieve the same estimate accuracy as [10], but only require around 10% data from TD2. There is no estimate difference between EPKF-simp and EPKF-norm, when the transmission rate is one. However, as the transmission rate decreases, EPKF-norm produces slightly better estimates. This is due to the fact that the distribution of the prediction error is fixed for a given system, while the distribution of $\bar{e}_k$ is determined by the threshold. As the threshold increases, $\bar{e}_k$ satisfies closer to the normal distribution at each Markov state. Thus, the normal approximation is better when the transmission rate is low. However, the superiority of EPKF-norm over EPKF-simp is

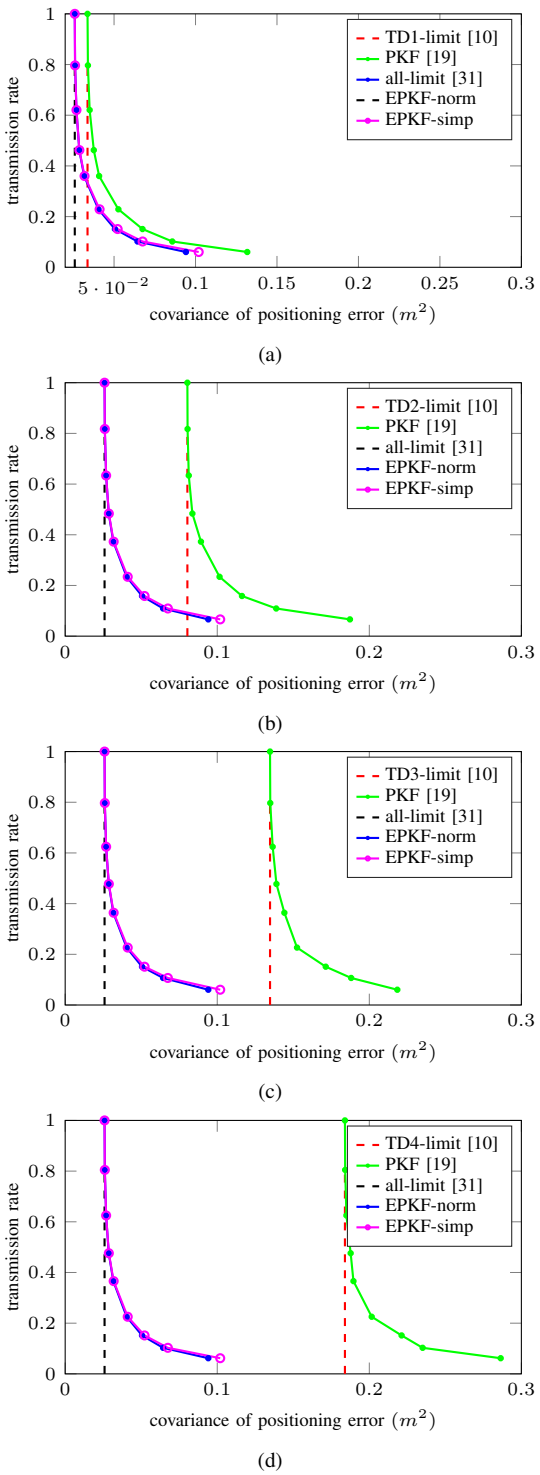


Fig. 9. Comparisons of the trade-off between transmission rate and positioning accuracy among five methods, where TDx-limit is derived by [10] using all data from the single TDx and all-limit is obtained by [31] using all data from four TDs. (a) only the data of TD1; (b) only the data of TD2; (c) only the data of TD3; (d) only the data of TD4, is used by [10] and [31] as the pedestrian moves towards the direction of x-axis.

not very obvious. There is only 8.33% improvement when the transmission rate is 0.0662. This can be explained by Eq. (28): the increase of the threshold results in the exponential decrease of the transmission rate, while it is not big enough to make the error become the normal distribution.

The similar results also hold for TD1, TD3 and TD 4 as shown in Figs. 9a, 9c and 9d, respectively. As the pedestrian moves towards the direction of x-axis, the covariance of the estimate error becomes to 0.1347 m^2 if the smartphone only uses all data from TD3 for positioning. It is 67.64% bigger than using the data from TD2 while 36.55% smaller than using the data from TD4. This is because the increased measurement noise of the TD results in less trustful estimates. Due to the same reason, the gain by using the data from other TDs improves as the measurement noise of the TD increases. Among them, TD1 has the lowest measurement noise. The estimate covariance using [10] is 0.0337 m^2. It can be improved by 23.18% using data from other TDs with Rand-ST, which is 2 times smaller than the gain in TD2. To achieve the same estimate quality as [10], EPKF methods require TD1 to transmit around 35% transmission. The rate increases compared with the one of TD2 due to the decreased gain.

In summary, EPKF methods drastically improve the estimate accuracy by exploiting the data from other TDs. They only require 10% transmission to achieve the same estimate quality as using KF in the traditional manner as in [10]. The required transmission becomes less as the measurement noise increases, which results in a lower level of energy consumption for the device. Compared with EPKF-norm, EPKF-simp is more preferred because of its low complexity and similar estimate accuracy.

V. Conclusion and Future Work

This work proposes EPKF methods to avoid the unnecessary transmissions for energy constrained IoT devices while keeping the functionality of KF. Each transmitting device (TD) independently compresses its transmission based on the temporal correlation; and the receiving device (RD) further exploits the spatial correlation among multiple TDs to improve the reconstruction quality.

The reconstruction problem mentioned above is nonlinear, which is hard to be solved by the energy constrained devices. To provide the linear solutions with reduced complexity, a deep analysis regarding the local estimate error is carried out. The error is produced at different Markov states and it has a truncated MVN distribution over parallelograms determined region at each state. There is no solution so far for calculating the covariance of such a variable. We obtain the error covariance with the help of Hessian matrix of MVN's PDF and the matrix transformation. Based on the analysis, two linear solutions are proposed: EPKF-norm and EPKF-simp. The former one differentiates the error at each Markov state and approximates it as the corresponding Gaussian noise. The RD then uses a KF with augmented system model and a time variant covariance of the measurement noise to produce the reconstruction. The complexity of EPKF-norm is further

reduced by EPKF-simp, where the reconstruction error is simply approximated as zero. The RD then uses the linear combinations of the local estimates with the corresponding weights for reconstruction.

The proposed methods are fundamental approaches applicable to any IoT monitored/controlled physical system that can be modeled as a linear state space representation. The pedestrian-position application is used as a case study to demonstrate the efficiency of EPKF methods in the simulation. EPKF methods decrease the transmission rate to around 10% while achieving the same positioning accuracy as using KF in the traditional manner as in [10]. As the measurement noise increases, the required transmission becomes less, which results in a lower level of energy consumption for the device. Compared with EPKF-norm, EPKF-simp is more preferred due to its low complexity and similar estimate accuracy.

REFERENCES

[1] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," *Future Generation Computer Systems*, vol. 29, no. 7, pp. 1645 – 1660, Sep. 2013.

[2] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2347–2376, Fourthquarter 2015.

[3] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, and M. Gidlund, "Industrial Internet of things: Challenges, opportunities, and directions," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 11, pp. 4724–4734, Nov. 2018.

[4] H. Jawad, R. Nordin, S. Gharghan, A. Jawad, and M. Ismail, "Energy-efficient wireless sensor networks for precision agriculture: A review," *Sensors*, vol. 17, no. 8, p. 1781, Aug. 2017.

[5] A. P. Plageras, K. E. Psannis, C. Stergiou, H. Wang, and B. Gupta, "Efficient IoT-based sensor BIG Data collection-processing and analysis in smart buildings," *Future Generation Computer Systems*, vol. 82, pp. 349 – 357, May 2018.

[6] O. Salem, A. Serhrouchni, A. Mehaoua, and R. Boutaba, "Event detection in wireless body area networks using Kalman filter and power divergence," *IEEE Transactions on Network and Service Management*, vol. 15, no. 3, pp. 1018–1034, Sep. 2018.

[7] Y. Huang, W. Yu, and A. Garcia-ortiz, "Accurate energy-aware workload distribution for wireless sensor networks using a detailed communication energy cost model," *Journal of Low Power Electronics*, vol. 10, no. 2, pp. 183–193(11), June 2014.

[8] J. Åkerberg, M. Gidlund, and M. Björkman, "Future research challenges in wireless sensor and actuator networks targeting industrial automation," in *2011 9th IEEE International Conference on Industrial Informatics*. IEEE, July 2011, pp. 410–415.

[9] R. C. Dorf and R. H. Bishop, *Modern Control System (12th ed.)*. Pearson Education, 2013.

[10] N. Yu, X. Zhan, S. Zhao, Y. Wu, and R. Feng, "A precise dead reckoning algorithm based on bluetooth and multiple sensors," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 336–351, Feb. 2018.

[11] Z. Chen, Q. Zhu, and Y. C. Soh, "Smartphone inertial sensor-based indoor localization and tracking with iBeacon corrections," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 4, pp. 1540–1549, Aug. 2016.

[12] J. Wang, R. Zhu, and S. Liu, "A differentially private unscented Kalman filter for streaming data in IoT," *IEEE Access*, vol. 6, pp. 6487–6495, Jan. 2018.

[13] Q. Wang, Y. Zhang, X. Lu, Z. Wang, Z. Qin, and K. Ren, "Real-time and spatio-temporal crowd-sourced social network data publishing with differential privacy," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 4, pp. 591–606, July 2018.

[14] J. Zhao, M. Netto, and L. Mili, "A robust iterated extended Kalman filter for power system dynamic state estimation," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3205–3216, July 2017.

[15] H. Wang, H. Li, J. Fang, and H. Wang, "Robust Gaussian Kalman filter with outlier detection," *IEEE Signal Processing Letters*, vol. 25, no. 8, pp. 1236–1240, Aug. 2018.

[16] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry, "Kalman filtering with intermittent observations," *IEEE transactions on Automatic Control*, vol. 49, no. 9, pp. 1453–1464, Sep. 2004.

[17] D. Han, Y. Mo, J. Wu, S. Weerakkody, B. Sinopoli, and L. Shi, "Stochastic event-triggered sensor schedule for remote state estimation," *IEEE Transactions on Automatic Control*, vol. 60, no. 10, pp. 2661–2675, Oct. 2015.

[18] A. Jain, E. Y. Chang, and Y.-F. Wang, "Adaptive stream resource management using Kalman filters," in *Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data*. ACM, June 2004, pp. 11–22.

[19] Y. Huang, W. Yu, and A. Garcia-Ortiz, "PKF: A communication cost reduction schema based on Kalman filter and data prediction for wireless sensor networks," in *Proceedings of the 26th IEEE Inernational system-on-chip conference*. CAS, Sep. 2013, pp. 73 – 78.

[20] Y. Huang, "Transmission rate compression based on Kalman filter using spatio-temporal correlation for wireless sensor networks," Ph.D. dissertation, University of Bremen, 2017.

[21] Y. Huang, W. Yu, C. Osewold, and A. Garcia-Ortiz, "Analysis of PKF: A communication cost reduction scheme for wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, pp. 843–856, Feb. 2016.

[22] Z. Wu, M. Fu, Y. Xu, and R. Lu, "A distributed Kalman filtering algorithm with fast finite-time convergence for sensor networks," *Automatica*, vol. 95, pp. 63 – 72, Sep. 2018.

[23] A. S. Leong, S. Dey, and D. E. Quevedo, "Sensor scheduling in variance based event triggered estimation with packet drops," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1880–1895, April 2017.

[24] S. Wu, X. Ren, S. Dey, and L. Shi, "Optimal scheduling of multiple sensors over shared channels with packet transmission constraint," *Automatica*, vol. 96, pp. 22 – 31, Oct. 2018.

[25] Y. Huang, W. Yu, and A. Garcia-Ortiz, "PKF-ST: A communication cost reduction scheme using spatial and temporal correlation for wireless sensor networks," in *Proceedings of the 2016 International Conference on Embedded Wireless Systems and Networks*. USA: Junction Publishing, Feb. 2016, pp. 47–52.

[26] K. K. Andersen, H. Madsen, and L. H. Hansen, "Modelling the heat dynamics of a building using stochastic differential equations," *Energy and Buildings*, vol. 31, no. 1, pp. 13 – 24, Jan. 2000.

[27] R. H. Bill Messner and J. Taylor, "Control tutorials for matlab and simulink: Dc motor position: State-space methods for controller design," online, 2017. [Online]. Available: http://ctms.engin.umich.edu/CTMS/index.php?example=MotorSpeed&section=SystemModeling

[28] Y. Wang, S. You, W. Zheng, H. Zhang, X. Zheng, and Q. Miao, "State space model and robust control of plate heat exchanger for dynamic performance improvement," *Applied Thermal Engineering*, vol. 128, pp. 1588 – 1604, Jan. 2018.

[29] F. Karray, A. Garcia-Ortiz, M. W. Jmal, A. M. Obeid, and M. Abid, "Earnpipe: A testbed for smart water pipeline monitoring using wireless sensor network," *Procedia Computer Science*, vol. 96, pp. 285 – 294, Jan. 2016.

[30] Y. Tomita, S. Ohmatsu, and T. Soeda, "An application of the information theory to estimation problems," *Information and Control*, vol. 32, no. 2, pp. 101 – 111, Oct. 1976.

[31] Y. Huang, W. Yu, M. Usman, S. Wang, and A. Garcia-Ortiz, "Analysis of optimal reconstruction methods based on incomplete information from sensor nodes using Kalman filter," *IEEE Sensors Journal*, vol. 18, no. 16, pp. 6889–6902, June 2018.

[32] Y. L. Tong, *The Multivariate Normal Distribution*. Springer, New York, 1990.