Efficient Spatial Data Recovery Scheme with Error Refinement for Cyber-physical Systems

Naushin Nower
Institute of Information
University of Dhaka

Yasuo Tan
Japan Advanced Institute of
Science and Technology

Yuto Lim
Japan Advanced Institute of
Science and Technology

ABSTRACT

Feedback data loss can severely degrade the overall system performance and as well as it can affect the control and computation of the Cyber-physical Systems (CPS). CPS hold enormous potential for a wide range of emerging and time critical applications including different traffic patterns. Therefore, incomplete feedback makes a great challenge in any uncertain condition to maintain the real-time control of the CPS. In this paper, we proposed a data recovery called Efficient Spatial Data Recovery with an error refinement (ESDR/ER) procedure for CPS to minimize the error estimation and maximize the accuracy of the scheme. In this scheme, we present an algorithm with Pearson Correlation Coefficient (PCC) to efficiently solve the missing data for both deterministic and stochastic traffic patterns. We also present an error refinement procedure to refine the error thus to maintain high accuracy. Numerical results reveal that the proposed ESDR/ER outperforms both WP and STI algorithms regardless of the increment percentage of missing data in terms of the root mean square error, mean absolute error and integral of absolute error.

Keywords
Data recovery, Cyber-physical systems, error refinement, deterministic patterns.

1. INTRODUCTION
Cyber-physical systems (CPS) are a new generation of communication, control and computation that have received a great deal of attention recently [1]. CPS enable the virtual world to interact with the physical world in order to monitor and control the intended parameter in the real-time basis. In CPS, technologies such as communication, control, computation, cognition and sensing converge to create new technologies for smarter society. The area of CPS represents the intersection of several systems trends, such as real-time embedded system, distributed systems, control system and networked wireless system.

To facilitate communications between cyber and the physical world, wireless sensor and actuator network (WSAN) is an essential ingredient of CPS. This is because the traditional wireless sensor network (WSN) is limited in its ability to monitor the physical world [2]. However, CPS achieves this requirement by facilitating the system to sense, interact and change the physical world in real-time by using feedback control loop. In typical CPS application, sensor nodes collect information from the physical world as a source of CPS inputs. Upon receiving the information of inputs, a controller makes a corresponding decision based on the inputs by computing and actuators perform the action in the physical world through the closed-loop feedback.

Since, CPS exploit the physical information collected by WSANs; it also faces the wireless contention problem which makes a challenging issue to control in real-time. Wireless channels have many adverse properties like path loss, fading, adjacent channel interference, node/link failure, etc. Besides these, wireless signal can be easily affected by noise, physical obstacles, node movement, environmental change and so on [3-4]. Because of this unpredictable and dynamic nature, the sensing data loss is a common phenomenon, which makes hamper in controlling decision. Since, the applicability of CPS is found in numerous time-critical applications including smart house to the smart grid, data loss makes the system unstable. Emerging applications of CPS include medical devices and systems, aerospace systems, transportation vehicles and intelligent highways, defense systems, robotic systems, process control, factory automation, building and environmental control, smart spaces, intelligent home and so on [5]. In all of these applications, CPS has to monitor and control the state of the physical phenomenon in real-time. In particular, for time-critical applications, feedback data must have to arrive on time, to make a decision. In many cases, re-transmission cannot provide an appropriate solution because of the unpredictable network behavior, which can cause high delay.

To maintain uninterrupted control, we need a data recovery scheme that can handle insufficient feedback control information. In our paper [6], we proposed a highly Efficient Spatial Data Recovery (ESDR) scheme that deals with CPS. In this paper, we enhance our proposed ESDR scheme for both deterministic and stochastic traffic patterns and also propose a refinement procedure to refine the estimated error. To do this, we design a framework structure for the control view of the CPS with our data recovery scheme. The designed framework incorporates the proposed ESDR scheme, which is based on the spatial correlation of neighboring sensors by using the Pearson correlation coefficient (PCC). Since sensor data is highly correlated with space and time, we utilize the spatial relationship to recover the lost data. After each estimation, we calculate the mean error and make refinement on that mean error. The refined mean error is adjusted from the estimated value.

One of our contributions is that the enhance ESDR/ER scheme ensures timely data recovery for both deterministic and stochastic traffic because of minimum computation. This algorithm always gives the result within threshold because of the error refinement. Second, our proposed ESDR/ER scheme is used to examine the smart home environment with CPS approach in order to maintain desired room temperature at different locations. Thus, the feedback measured room temperature is very important to keep the desired room temperature steadily at all the times. Another advantage is that, the proposed scheme ensures scalability. Since it uses only one-hop neighbors, thus the scheme can be applied in both small and large network.

The rest of the paper is organized as follows. Section II summarizes some state-of-the-art research works that are

related to this paper. In Section III, the proposed ESDR/ER is presented. We describe the experimental scenario and the evaluation parameters in Section IV. Simulation results and discussions are presented in Section V. Section VI concludes with the conclusion and future works.

2. RELATED WORK

Data recovery and evaluation of estimation is a part of most research and there exist several methods to handle this. Although there exist several methods, the recovery of data loss for CPS still poses an open problem because of its unique requirement. The whole recovery process for CPS must be held in real-time and invisible to the outside world.

Missing data is a well-studied subject in statistics. Little and Rubin provide an introduction to statistical missing data imputation techniques, such as least squares estimates, Bartlett's ANCOVA and likelihood-based approaches in [7]. Maximum likelihood (ML), multiple imputations (MI) and expectation maximization (EM) are widely used method for missing data imputation. ML[8] calculates the likelihood function for given set of data, which is a hypothetical probability that uses past event with the known outcome. Then, by using iterative steps, ML makes the likelihood function maximum. EM [9] also uses an iterative step to maximize the likelihood function but here, the model depends on unobserved or latent variables. Based on mean and covariance matrix of multivariate normal distribution, expectation (E) step initializes the expected values for latent variables. Maximization (M) step plugs the expected values into the log-likelihood function and maximizes the log-likelihood function by repeating the E and M steps. However initialization step directly impact the performance of EM based imputation. On the other hand, in MI [10], missing data are filled by m different times to generate m complete data sets. Generated m data sets are analyzed by standard procedure and then combined for inference. But these well known techniques for missing data imputation are not suitable for WSNs, due to their high space and/or time complexities.

Imputation methods based on machine learning are sophisticated procedures that use a predictive model to estimate values. These approaches model the missing data estimation based on information available in the data set. If the observed data contain useful information then, imputation procedure maintains high precision [11]. At each iteration, the error is calculated and adjusted to the following step. Multi-layer perceptron (MLP), self-organizing map (SOM), k-nearest neighbors (k-NN) are examples of imputation techniques based on learning. MLP is multi-layer computational unit, which is connected by a feed-forward way. It estimates the missing data by training an MLP to learn incomplete data by using complete data [12]. On the other hand, in SOM, a set of nodes is organized in a 2D grid, where each node has a specific position and weight. Iterative training steps initialize the weight, and then it is used to estimate missing data [13]. Both of these methods require all data to trained and estimate the missing value. But in k-NN [14], to impute missing data, only k nearest neighbor's data is considered. These techniques are used in WSN to impute data but for real-time CPS, these are not suitable.

Compressed sensing (CS) [15] is widely used scheme for signal processing to acquire and reconstruct a signal, based on underdetermined linear systems. This takes advantage of the signal's sparseness or compressibility in some domain, allowing the entire signal to be determined from relatively few measurements. The main difference between the missing data recovery problem and the conventional CS is that in conventional CS, the sampling scheme can be determined by the users, and usually random linear projections are preferred, while in the missing data recovery problem the sampling matrix cannot be controlled by the user since it is determined by the missing events, e.g., locations of missing nodes in the network which is completely uncertain [16].

Guo, et al. [17] design an algorithm considering spatial-temporal correlations of sensor nodes, which is more suitable with WSNs due to nature of WSNs. Their algorithm first checks if a neighbor sensor node is within the missing sensor's sensing range. Then the observation from the neighbor is used for filling in the missing values. This generates a spatially correlated replacement. If there are multiple neighbors within the sensor's range and they do not have the same readings, the majority reading is chosen. But in real life, there is no guarantee that all the sensors within one-hop neighbor are spatially and temporally correlated.

In the existing literature, there are other two ways to investigate the spatial correlation for missing data recovery, which is inverse distance weighted averaging (IDWA) [18] and Kriging [19]. Assuming the spatial correlation in adjacent sensors is uniform, IDWA tries to estimate the values of missing data in the form of some linear combination of neighboring sensor's data. IDWA will work well if the values of missing sensors are expected to be similar to values of the neighboring sensors. However, this assumption affects the estimation accuracy in many practical situations, where a physical phenomenon varies rather than uniformly increasing or decreasing in magnitude. The averaging process in IDWA has the tendency to smoothen the data, which is not suitable for the situation when data change fast in the area of interest. And, none of these data recovery method include error refinement.

Kriging is another way to estimate the missing samples using the combination of available measurements. It defines a semi-variogram by calculating the spatial correlation between sensors. From the semi-variogram, the weight for the linear combination is determined. As a result, these weights vary spatially and depend on the correlation [19]. However, the spatial interpolation may not be right if the semi-variogram varies a lot in the temporal dimension [16]. And, there is no estimation error correction mechanism followed by this data recovery algorithm.

In [20], the authors evaluate the performance of the different estimation refinement techniques including RMSE, average Euclidean error, geometric average error, mean and variance of RMSE. RMSE is dominated by its large individual terms. This amounts to a more serve penalty on large errors than on the small errors. However, RMSE is a measure for performance evaluation and, in particular, it is intended to serve as a metric for average error in magnitude (at least interpreted this way by many). As far as the average error in magnitude is concerned, for many problems it might be reasonable to expect that any large error should be possibly balanced by a sufficiently small error. In this research, we utilize this advantage of the mean error to refine our estimation error.

Besides these, many researchers combine genetic algorithm (GA) with the artificial neural network (ANN) [21], GA with Bayes algorithm [11] and many more to estimate the missing value.

Xia, et al. [22] first propose a solution for CPS over WSANs to cope with packet loss. They illustrate three prediction algorithms and show a comparison between them. The first algorithm based on the assumption that the state of the physical system does not change during the last sampling period. So, the previous sample is used to replace the missing value. The second algorithm computes a moving average of the previous *m* samples to restore the lost data. Thus it treats every previous measurement equally. In third algorithm weighted average of all previous samples is taken to replace the missing one. Simulation result shows that third algorithm works well compared with others.

Choi, et al. [23] exploit an exponentially weighted moving average (EWMA) based value estimation algorithm to reduce the impact of packet. When some packets are randomly dropped in the wireless network environment, the EWMA algorithm filters an abrupt increase or decrease by exponentially smoothing commands or data based on the past value profile.

3. EFFICIENT SPATIAL DATA RECOVERY SCHEME

In this section, we enhance our proposed efficient spatial data recovery (ESDR) scheme for deterministic and stochastic traffic pattern of CPS. To reduce the error in estimation, we propose an error refinement procedure, which makes the estimation more accurate. Before doing this, we classify the pattern and types of CPS data traffic. We classify three traffic patterns for CPS applications: deterministic, stochastic and time-critical. The deterministic traffic pattern always maintains a stable state. On the other hand, any traffic pattern, which involves random change and indeterminacy, is defined as a stochastic traffic pattern. We concentrate these two traffic patterns in this paper. And, these traffic patterns can be transmitted by four different traffic types [24]: fixed, periodic, bursty and arbitrary rate. In this research, we design ESDR/ER scheme to mitigate the problem of periodic traffic type. Figure 1 shows the control view of CPS with our proposed ESDR scheme.

To deploy our proposed ESDR/ER scheme, we propose a flowchart with the ESDR/ER scheme for CPS as depicted in Fig. 2. The following assumptions have been considered. First, the historical dataset for one-hop neighbor is available up to window size to perform the ESDR/ER scheme. Second, the error offset e_0 of the measured data and estimated data is initially computed and known. Third the maximum number of consecutive missing data C is fixed at the initialization stage. The parameter C is also used for terminating the entire system to indicate the estimated data cannot be produced anymore because of the long consecutive missing data.

Fourth, the mean error always is computed and available for refinement.

In the flowchart, the ESDR/ER scheme will compute the estimated data when there is an input measured data from the sensors. If there is no missing data, then the measured data is used as a feedback data. At the same time the difference between the measured and estimated data is computed and if the difference is greater than error offset, ESDR/ER scheme is refined to reduce the error. When there is a missing data, the
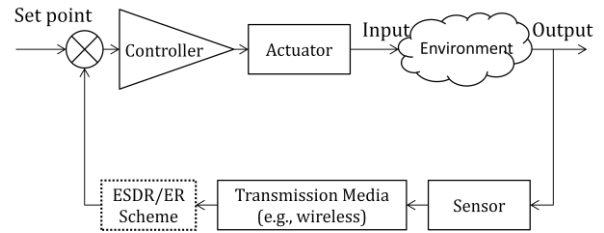

Fig 1: Proposed data recovery scheme for control view of CPS.

consecutive missing data is evaluated and the estimated data is used as a feedback data.

If the sampling interval *I* of the measured data is less than 2 min and the traffic pattern is deterministic, then sensor's previous measurement is used as a replacement. But for the stochastic traffic pattern, since it measured data changes randomly, we always incorporate spatial correlation to estimate the missing data. As far as we are concerned, most of the spatial correlation for data recovery scheme is focusing on the data correlation that based on the difference between the nearest neighbors. In our ESDR/ER scheme, we consider the most spatial correlation among the neighboring sensors based on the Pearson correlation coefficient (PCC) [25]. In PCC, if an environment is highly correlated in space, then the spatial information can be used to estimate missing data and the estimation function can achieve a high accuracy. PCC is a common measure of the linear correlation between two random variables i and j. It reflects the degree of association between two variables. Therefore, the coefficient correlation degree of PCC ρ in between two random variables *i* and *j* in specified window size (*W*) can be computed as follows

$$\rho_{ij} = \frac{\sum_{w=1}^{W}\left(i(w)-\bar{i}\right)\left(j(w)-\bar{j}\right)}{\sqrt{\sum_{w=1}^{W}\left(i(w)-\bar{i}\right)^2}\sqrt{\sum_{w=1}^{W}\left(j(w)-\bar{j}\right)^2}}$$

Table I shows the association degree of the ρ. The range from -1.0 to 1.0 shows that the ρ has a degree of correlation. The negative value of ρ indicates the negative linear relationship, whereas the positive value of ρ indicates the positive linear relationship.

TABLE 1. CORRELATION DEGREE OF PEARSON CORRELATION COEFFICIENT

No Correlation		$0.1 > \rho > -0.1$ and $1.0 > \rho$ and $\rho < -1.0$
Correlation Degree	Small	$0.1 \leq \rho < 0.3$ and $-0.1 \geq \rho > -0.3$
	Medium	$0.3 \leq \rho \leq 0.5$ and $-0.3 \geq \rho \geq -0.5$
	Large	$0.5 < \rho \leq 1.0$ and $-0.5 > \rho \geq -1.0$
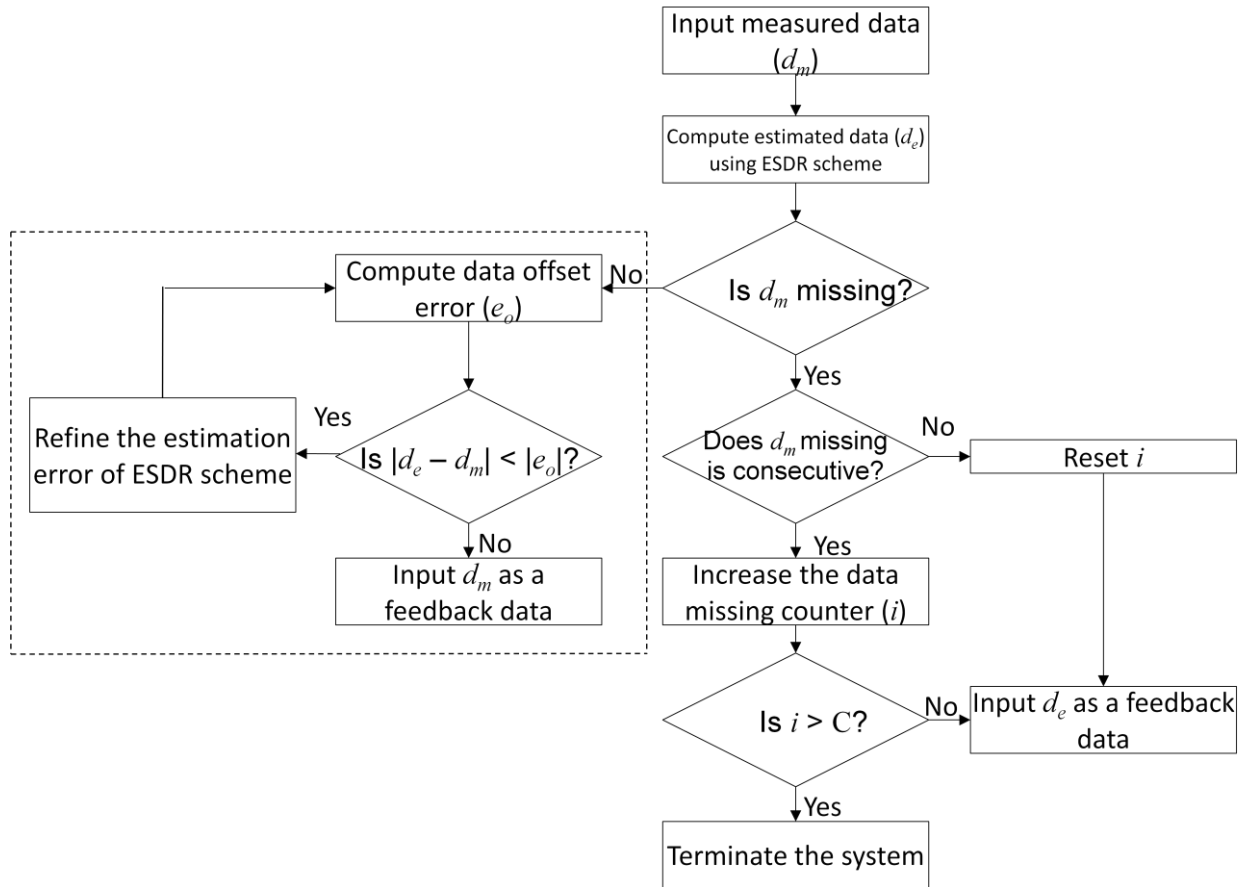
Fig 2: Proposed flowchart with ESDR scheme for CPS

Figure 3 shows the ESDR/ER algorithm, which is used to produce an estimated data from time to time. In this algorithm, we assume that the threshold value of estimation counter c_{th} is used to optimize the estimation function of the algorithm. Once the ESDR/ER algorithm cannot use the PCC, we recommend that the estimated data be produced based on the nearest neighbor data. When the number of estimation counter for the corresponding of the sensor c_k is above the threshold value, the new corresponding of sensor will be computed again. To maintain high accuracy in estimation, we select the value of ρ is in between 0.5 to 1.0.

Algorithm: Efficient Spatial Data Recovery (ESDR)

1: **if** c_k=0 **then**

2:     **for** each input sensor d_i **do**

3:         **for** all sensors d_j within one-hop neighbour of d_i **do**

4:             Compute ρ_{ij} with the specified window size, W

5:             **if** $0.5 < |\rho_{ij}| \leq 1.0$ **then**

6:                 $k \leftarrow \operatorname{argmax}\{|\rho_{ij}|\}$; $c_k \leftarrow 1$

7:             **end if**

8:         **end for**

9:     **end for**

10: **else if** $c_k > c_{th}$ **then**

11:     $c_k \leftarrow 0$

12: **Else**

13:     **if** I<2min and traffic pattern=deterministic

14:         $d_e(t) \leftarrow d_i(t) = d_i(t-1)$

15:     **Else**

16:         Compute ρ_{ik} with the specified window size, W

17:         **if** $0.5 < |\rho_{ik}| \leq 1.0$ **then**

18:             $d_e(t) \leftarrow d_i(t) = d_k(t) + [d_i(t-1) - d_k(t-1)]$

19:         **Else**

20:             $k \leftarrow \operatorname{argmin}\{distance_{ij}\}$

21:             $d_e(t) \leftarrow d_i(t) = d_k(t) + [d_i(t-1) - d_k(t-1)]$

22:             $c_k \leftarrow c_k + 1$

23:         **end if**

25:     **end if**

26: **end if**

Algorithm: Error Refinement

---

    if $d_m - d_e < e_0$

        for each estimated error_i do

            E_mean=Average (error_i)

If error $_{-j}$< E-mean do

Refinemed_error_mean=Average(error $_{-j}$)

Data_after_refinement=d_e-Refined_error_mean

   end_for

  end_if

 end_if

Fig 3: Pseudo code for ESDR/ER algorithm.

To refine the error in the estimated data, we proposed an error refinement procedure. In order to do this, we consider the error that is minimum in between the estimated and measured one and tested value to recover the measured data. We calculate the mean error and define the good error which is less than or equal to the mean error. After selecting the good error, the mean error is refined by computing the average of the selected good error. Then, to make the estimation more accurate, we subtract the refined error mean from the estimated data, which makes the data more accurate.

4. SIMULATION SCENARIOS AND PARAMETERS

In this section, we conduct the simulation studies to evaluate our proposed ESDR/ER scheme on deterministic and stochastic traffic pattern of CPS. We compare our proposed ESDR/ER scheme to the ESDR, WP algorithm [22] and the STI approach [17]. For deterministic traffic pattern, we conduct an experiment to measure the inside temperature of the master bedroom in the iHouse facility which is in situated Nomi city in Japan. The room is equipped with eight sensors at eight corners. The measurement is taken in every two minutes for twenty-four hours. All the sensors forward their data to reach the base station in single radio hop through the simplest spanning tree topology routing protocol. For stochastic traffic pattern, we create a simulation environment with five sensor and generate data series using autoregressive (AR) model in MATLAB simulator. We also add some random noise with each series to it make more realistic.

Based on the collected information, we investigate the performance of our proposed scheme using a MATLAB. In this simulation, we assume that the single sensor produces a missing sensed data when it transmits its packet to the base station. We randomly delete the data according to the percentage of missing data from the original set and recover them using the aforementioned data recovery algorithms. We use the root mean square error (RMSE), mean absolute error (MAE) and integral of absolute error (IAE) to evaluate the performance of the said algorithms.

The RMSE is a frequently used measure of the difference between values estimated by an algorithm and the values actually measured from the real environment. The RMSE of algorithm estimation with respect to the estimated value, d_e is defined as the square root of the mean squared error as written as

$$RMSE = \sqrt{\frac{\sum_{n=1}^{N}\left(d_m(n) - d_e(n)\right)^2}{N}}$$

The MAE is another statistical measurement that used to measure how close the estimated values are to the measured values. The MAE is given by

$$MAE = \frac{1}{N}\sum_{n=1}^{N}\left|d_e(n) - d_m(n)\right|$$

The MAE measures the average magnitude of the errors in a data set, without considering their direction. It is also an average of the absolute error, $e=d_e-d_m$. In other words, it measures the accuracy of the continuous variables. The MAE and the RMSE can be used together to analyze the variation in the errors of the dataset. The RMSE will always be larger or equal to the MAE. The greater difference between them, the greater the variance in the individual errors in the sample [26]. If the RMSE is equal to the MAE, then all the errors are the same magnitude. In [26], Wilmott, et al. indicate that the MAE is the most natural and unambiguous measure of average error magnitude.

On the other hand, the IAE is a widely used performance metric in control community, which is recorded to measure the performance of the control application. The IAE is calculated as follows

$$IAE = \int_0^t \left|d_e(t) - d_m(t)\right|dt$$

where, t denotes total simulation time. In general, the larger the IAE values imply the worse the performance of the control algorithm.

# 5. SIMULATION RESULTS AND DISCUSSIONS

In this section, we present our simulation results and make some discussions on the performance of algorithms. The aim of this simulation is to examine the potential of the proposed algorithm with error refinement in coping with the data missing for the deterministic and stochastic traffic pattern of CPS. In our ESDR/ER scheme, we measure the PCC in between the sensors from time to time by specified the window size (*W*) and if the estimated error is above the error offset, the error is refined using error refinement procedure. For deterministic traffic pattern we set window size (*W*) as ten data samples. Since, stochastic traffic pattern changes randomly, thus to understand the correlation properly we recommend using lower window size, which is three data sample for this experiment. In our simulation, we investigate the impact of increasing percentage of missing data on the data recovery algorithm with and without error refinement performance. The percentage of missing data is varied from 30% to 60% in steps of 10%.
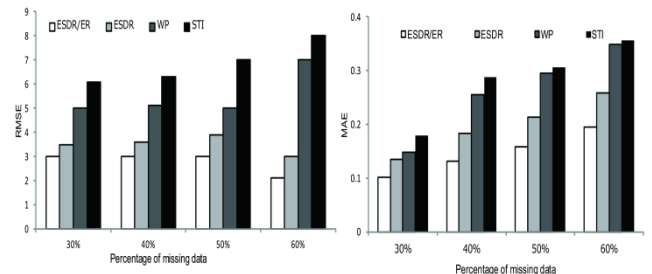
Fig 5: The comparison of RMSE and MAE for stochastic traffic pattern of all the data recovery algorithms as the percentage of missing data changes from 30% to 60%.

Fig. 5 depicts the RMSE comparison among data recovery algorithms for stochastic traffic patterns. As the percentage of data missing increases, the proposed algorithm with refinement always shows better performance that is compared

to the ESDR algorithm and other existing two algorithms. At the 40% data missing, the ESDR/ER scheme performs slightly better than the ESDR algorithm. In this case RMSE of the ESDR/ER is 3 on the other hand RMSE of the ESDR WP and STI is 4,5 and 6.1 respectively. At the 60% data missing, the proposed ESDR/ER scheme reduces almost one-third of the RMSE than the WP and STI algorithm where the RMSE of ESDR/ER is 2 compared with that of STI, which gives the 8 as a RMSE. The reason for this dramatic improvement is because the ESDR scheme does not incorporate refinement and the WP algorithm cannot cope with the long consecutive missing data. Through this simulation, we can observe that this problem also can be found at the STI algorithm. Both WP and STI algorithm use the combination of previous measurements only. Thus, they unable to cope with long consecutive missing and frequent changes in the environment of the conducted experiments.
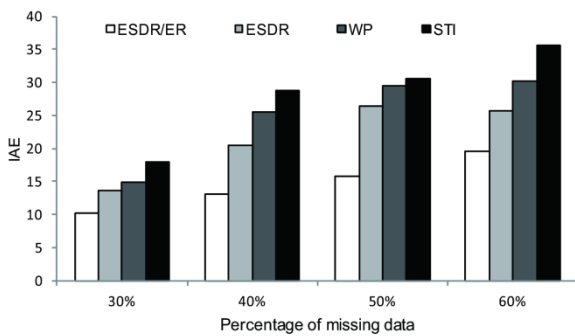

Fig 6: The accumulated IAE comparison for stochastic traffic pattern of all the data recovery algorithms as the percentage of missing data changes from 30% to 60%.

The MAE comparison for stochastic data traffic among four data recovery algorithms is shown in Fig. 3. We can see that the proposed ESDR/ER scheme outperforms the ESDR, the WP algorithm and the STI algorithm. Besides that, the proposed ESDR/ER scheme can steadily maintain a small value of MAE regardless of the increment of missing data. This also means that the distance between the real measured data and estimated data of the proposed ESDR/ER scheme is always stable. Since, the error is refined in each step to keep the error below the error offset, the ESDR/ER scheme always shows better performance. As a result, after the error is refined, the proposed ESDR/ER scheme always outperforms ESDR scheme.

In Fig. 6, the accumulated IAE comparison for stochastic data traffic of all the data recovery algorithms is plotted. The simulation results demonstrate that the proposed ESDR/ER scheme outperforms the ESDR scheme, WP algorithm and the STI algorithm. This is because of the error of the estimation function in the proposed ESDR/ER scheme is minimized by using the refinement approach.

In Fig. 7 the RMSE, MAE and IAE comparison for deterministic traffic is shown. In this case, error, in between the estimated data and measured data always remains below the error offset (where error offset is 0.5). Moreover, the deterministic traffic pattern always maintains a stable condition and thus the estimated data from the most correlated sensor always gives the accurate result. Since, the estimated data never crosses the error offset using ESDR algorithm, thus we do not need to refine the error using refinement procedure.
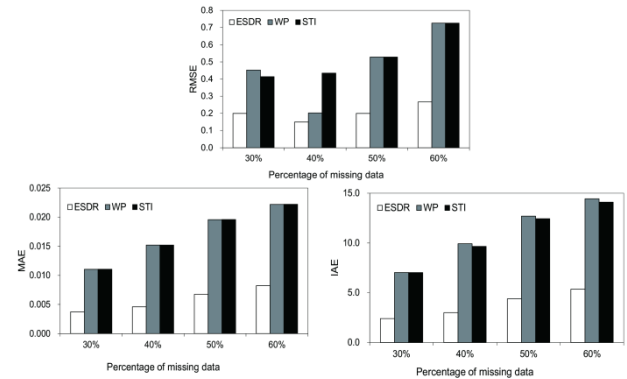
Fig 7: The RMSE, MAE and IAE comparison for deterministic traffic pattern as the percentage of missing data changes from 30% to 60%.

At 30% data loss the RMSE of the ESDR is almost half compare with that of the WP and STI algorithm. And at the 60% data loss the RMSE of the proposed scheme is one-third of the WP and STI algorithm. For MAE, simulation results reveal that the proposed ESDR scheme is about two times smaller than the WP algorithm and the STI algorithm. The simulation result for IAE is also one-third of the WP and STI algorithm. Thus, as the percentage of missing data increases, the deterministic data traffic always shows the better performance compare with the other two data recovery algorithm in terms of RMSE, MAE and IAE.

6. CONCLUSION

In this paper, we have enhanced the ESDR/ER scheme for different data traffics of CPS and proposed a error refinement procedure to reduce the estimation error. In this research work, we also identified that the stochastic data is more difficult to estimate and thus, we need to incorporate error refinement to maintain accuracy. But for the deterministic data, the estimation maintains high accuracy without refinement. Our simulation results reveal that the proposed ESDR/ER scheme is very beneficial and outperforms the ESDR, the WP and the STI algorithms regardless of the increment of missing data because of incorporating ER scheme.

This paper only concentrates on the ESDR/ER algorithm for deterministic and stochastic data and its evaluation on the real-life. Further research is required for the time-critical traffic pattern. Moreover, in our future work we need to consider the other traffic type like fixed, bursty and arbitrary for CPS data transmission. Besides that, a future work will focus on examining the real-time recovery using the proposed ESDR/ER scheme.

7. REFERENCES

[1] A.L. Edward, "Cyber physical systems: Design challenges," IEEE Symp. on Object Oriented Real-Time Distributed Computing, pp.363–369, 2008.

[2] A.L. Edward, "CPS foundations," ACM/IEEE Design Automation Conf. (DAC), pp.737–742, 2010.

[3] F.J. Wu, Y.F. Kao, and Y.C. Tseng, "From wireless sensor networks towards cyber physical systems," J. Pervasive and Mobile Comp., vol.7, no.4, pp.397–413, 2011.

[4] F. Martincic and L. Schwiebert, Introduction to wireless sensor networking, Handbook of Sensor Networks—

Algorithms and Architectures. John Wiley & Sons; New York, USA, 2005.

[5] A.L. Edward, "Towards a science of cyber-physical system design," ACM/IEEE Conf. on Cyber-physical System, pp.99–108, April 2011.

[6] N.Nower, T.Yasuo, A.O. Lim, "Efficient Spatial Data Recovery Scheme for Cyber-physical System," IEEE Int. Conf. on Cyber-Physical Systems, Networks and Applications, pp.72-77,2013.

[7] R.J.A. Little and D.B. Rubin, Statistical Analysis with Missing Data, 2nd Ed., Wiley-Interscience, New York, 2002.

[8] D.C. Howell. University of Vermont. (2009). *Treatment of missing data* [Online]. Available: http://www.uvm.edu/~dhowell/StatPages/More_Stuff/Missing_Data/

[9] J.G. Ibrahim, H. Zhu and N. Tang, "Model selection criteria for missing-data problems using the EM algorithm," J. American Statistical Association. pp.1648–1658, 2008.

[10] H.Y. Chen, H. Xie and Y. Qian, "Multiple imputation for missing values through conditional semi parametric odds ratio models," J. Biometrics vol.67, no.3, pp.799–809, 2011.

[11] J.M.I. Molina, P.J. Garcia-Laencina, E. Alba, N. Ribelles, M. Martin and L. Franco, "Missing data imputation using statistical and machine learning methods in a real breast cancer problem," J. Artificial Intelligence in Machine, Elsevier Science Publishers, vol.50, no.2, pp.105–115, 2010.

[12] C.M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer Science Business Media, 2007.

[13] T. Kohonen, Self-organizing Maps, Springer Series in Information Sciences, Springer-Verlag; 3rd Ed., 2001.

[14] C.C. Huang and H.M. Lee, "A grey-based nearest neighbor approach for missing attribute value prediction, "J. Applied Intelligence, vol.20, no.3, pp.239–252, 2004.

[15] W. Bajwa, "Compressive wireless sensing," ACM Conf. on Inf. Processing in Sensor Networks, pp.134–142, 2006.

[16] D. Guo, X. Qu, L. Huang and Y. Yao.,"Sparsity-based spatial interpolation in wireless sensor networks," J. Sensors, vol.11, no.3, pp.2385–2407, 2011.

[17] Y.Y. Li and L.E Parker, "Classification with missing data in wireless sensor network," IEEE Southeastcon, pp.533–538, April 2008.

[18] G.Y. Lu and D.W. Wong, "An adaptive inverse-distance weighting spatial interpolation technique," J. Comput. Geosci.,vol.34, pp.1044–1055, 2008.

[19] M. Umer, L. Kulik and E. Tanin, "Kriging for localized spatial interpolation in sensor networks," Int. Conf. on Scientific and Statistical Database Management, pp.525–532, 2008.

[20] X. Rong Li, Z. Zhao, "Measures of performance for evaluation of estimators and filters", Signal and Digital Processing Conf., pp.1-12, 2001.

[21] A. Azadeh, S.M. Asadzadeh, R.J. Marandi, S.N. Shirkouhi, G.B. Khoshjhou and S. Talebi, "Optimal estimation of missing values in randomized complete block design by genetic algorithm," Knowledge-Based Systems, Elsevier, vol.37, pp.37–47, 2013.

[22] F. Xia, X. Kong and Z. Xu, "Cyber-physical control over wireless sensor and actuator networks with packet loss," Wireless Networking Based Control, Springer, pp.85–102, 2011.

[23] K. Chen and S. Lien, "M2M Communications: Technologies and challenges", Elsevier Ad Hoc Networks, 2013.

[24] R.H. Choi, S.C. Lee, D.H. Lee and J. Yoo, "WiP abstract: Packet loss compensation for cyber-physical control systems," IEEE/ACM Int. Conf. on Cyber-Physical Systems (ICCPS), pp.205, 2012.

[25] Y. Ke, J. Cheng and J.X. Yu, "Efficient discovery of frequent correlated subgraph pairs," IEEE Int. Conf. on Data Mining (ICDM), pp.239–248, 2009.

[26] C.J. Wilmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over root means square (RMSE) in assessing average model performance, "Climate Research, vol.30, pp.79–82, 2005.