Future Generation Computer Systems 61 (2016) 85–96

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

A kernel machine-based secure data sensing and fusion scheme in wireless sensor networks for the cyber-physical systems☆

Xiong Luo [a,b,*], Dandan Zhang [a,b], Laurence T. Yang [c], Ji Liu [a,b], Xiaohui Chang [a,b], Huansheng Ning [a,b]

[a] School of Computer and Communication Engineering, University of Science and Technology Beijing (USTB), Beijing 100083, China
[b] Beijing Key Laboratory of Knowledge Engineering for Materials Science, Beijing 100083, China
[c] Department of Computer Science, St. Francis Xavier University, Antigonish, NS B2G 2W5, Canada

HIGHLIGHTS

- A novel data sensing and fusion scheme GM–KRLS is proposed in WSNs for the CPSs.
- GM–KRLS develops a prediction mechanism to reduce redundant transmissions in WSN.
- GM–KRLS improves the prediction accuracy with a kernel machine learning algorithm.
- Blowfish algorithm is employed to guarantee the confidentiality in our scheme.

ARTICLE INFO

Article history:
Received 15 June 2015
Received in revised form
8 September 2015
Accepted 30 October 2015
Available online 2 December 2015

Keywords:
Secure data sensing and fusion
Wireless sensor networks
Kernel recursive least squares
Cyber-physical systems

ABSTRACT

Wireless sensor networks (WSNs) as one of the key technologies for delivering sensor-related data drive the progress of cyber-physical systems (CPSs) in bridging the gap between the cyber world and the physical world. It is thus desirable to explore how to utilize intelligence properly by developing the effective scheme in WSN to support data sensing and fusion of CPS. This paper intends to serve this purpose by proposing a prediction-based data sensing and fusion scheme to reduce the data transmission and maintain the required coverage level of sensors in WSN while guaranteeing the data confidentiality. The proposed scheme is called GM–KRLS, which is featured through the use of grey model (GM), kernel recursive least squares (KRLS), and Blowfish algorithm (BA). During the data sensing and fusion process, GM is responsible for initially predicting the data of next period with a small number of data items, while KRLS is used to make the initial predicted value approximate its true value with high accuracy. The KRLS as an improved kernel machine learning algorithm can adaptively adjust the coefficients with every input, while making the predicted value more close to actual value. And BA is used for data encoding and decoding during the transmission process due to its successful applications across a wide range of domains. Then, the proposed secure data sensing and fusion scheme GM–KRLS can provide high prediction accuracy, low communication, good scalability, and confidentiality. In order to verify the effectiveness and reasonableness of our proposed approach, we conduct simulations on actual data sets that are collected from sensors in the Intel Berkeley research lab. The simulation results have shown that the proposed scheme can significantly reduce redundant transmissions with high prediction accuracy.

1. Introduction

A cyber-physical system (CPS) as an integration of sensors networks with cyber resources responds intelligently to dynamic changes in physical world, where the wireless sensor networks (WSNs) as one of the key components collect sensor data from physical environment [1]. With the increasing presence and adoption of WSNs on the deployment of CPS, there has been a growing demand in data sensing and data fusion to utilize intelligence

☆ This research is funded by the National Natural Science Foundation of China under Grants 61174103 and 61272357, the National Key Technologies R&D Program of China under Grant 2015BAK38B01, the Aerospace Science Foundation of China under Grant 2014ZA74001, and the Fundamental Research Funds for the Central Universities.

* Corresponding author at: School of Computer and Communication Engineering, University of Science and Technology Beijing (USTB), Beijing 100083, China. Tel.: +86 10 6233 2873.
E-mail address: xluo@ustb.edu.cn (X. Luo).

http://dx.doi.org/10.1016/j.future.2015.10.022
0167-739X/

Nomenclature

κ	=	the kernel function
$\boldsymbol{u}, \boldsymbol{v}$	=	the input vector of kernel function
ζ	=	the kernel parameter
λ	=	the regulation parameter of kernel recursive least squares (KRLS)
β	=	the forgetting factor of KRLS
P	=	P-array of Blowfish algorithm (BA)
S	=	S-box of BA
F	=	F-function of BA
ε	=	the threshold of prediction error
$e(i)$	=	the prediction error at time i
n	=	the length of the original data sequence
q	=	the number of the data sequences
$\boldsymbol{x}_S(i)$	=	the data sequence of the ith period for sensor node
$\boldsymbol{x}_K(i)$	=	the data sequence of the ith period for sink node
$\widehat{\boldsymbol{x}}_S(i)$	=	the predicted data sequence of the ith period

properly of CPS. Then, by integrating WSNs from different domains, CPS has emerged as a promising direction to enrich the interactions between physical and virtual worlds [2,3]. A WSN is composed of spatially distributed autonomous sensors used to cooperatively monitor physical or environmental data, such as temperature, humidity, light, noise, pressure, speed, and many others [4]. WSNs can be used to data sensing, disposing, and transmitting [5]. While WSNs are employed for real-time monitoring, plenty of sensor nodes sense the data of fluctuant monitored objects within a valid range and send those data to the sink node and end-users [6,7]. More recently, WSNs have demonstrated many successful applications across a wide range of domains, such as military affairs, national security, national defense, environment monitoring, energy management, and so on [8]. Thus, they are one of the key technologies to support sensing and actuation of CPS. WSNs are becoming a multidisciplinary research area attracting researchers from different fields especially industrial area.

In WSNs, the power module provides energy for nodes and once the nodes are deployed in many applications, it is almost impossible to recharge them. It is known that the process of wireless communication consumes most of the energy [9]. Since the data generated by sensor nodes during continuous sensing periods usually are of high temporal coherence, some data in the sustaining data sequence may be redundant, while causing unnecessary data transmission and wasting energy. The prediction-based data sensing and fusion schemes therefore have been proposed to process original data in the sensor nodes and reduce unnecessary transmissions [10]. To achieve the goal of extending the lifetime of the whole network, those schemes fully utilize the high temporal coherence of the sensed data to lessen the redundant transmissions and save the energy of sensor nodes [11,12]. In addition, since some problems like information leakage exist during the data transmission, the security of data is also one of the key issues in WSNs.

Among the known data sensing and fusion methods, a delay-aware network structure for WSNs with in-network data fusion was proposed in [13]. The proposed structure organizes sensor nodes into clusters of different sizes so that each cluster can communicate with the fusion center in an interleaved manner. However, it cannot achieve the best effectiveness without knowing the minimum achievable compression ratio between the sizes of incoming and outgoing data. For different data sensing and fusion topologies (e.g., star, chain, and tree), the optimal solutions were provided while computing the number of transmissions for each node in [14]. The distributed approximation algorithms were also presented for chain and tree topologies, but the model may be more complex as the size of the network increases. In [15], a distributed sensor fusion method was designed using a tree-based broadcasting strategy to improve the estimation efficiency and accuracy in WSN. Through the use of genetic machine learning algorithms, an implementation for data fusion techniques in WSNs was developed [16]. A quality-based multiple-sensor fusion approach was proposed in WSN [17]. However, there is also some improvement for these methods when the data set in WSN is more complex. In [18], the authors proposed a prediction-based temporal data fusion technique through the use of a first-order autoregressive model. However, the prediction accuracy of this model is poor when the time series data set is few or nonlinear [19].

In order to avoid those limitations that existed in the above data fusion approaches, some novel prediction-based data sensing and fusion schemes were presented with the help of grey model (GM) [20]. For instance, in [21], the authors presented a scheme GM-LSSVM. It implements the initial prediction using GM, and then utilizes the powerful nonlinear mapping capability of least squares support vector machine (LS-SVM) to improve the prediction accuracy. LS-SVM is established using the structural risk minimization principle rather than the empirical error commonly implemented in the neural networks. Then, LS-SVM can achieve higher generalization performance and higher precision accuracy than neural networks. But almost all of nonlinear series system identification by LS-SVM is offline, and its model is trained periodically. It imposes a challenging obstacle while using LS-SVM to conduct the prediction for nonlinear time series online. Moreover, in GM-LSSVM, it just employs the prediction mechanism in all sensor nodes, then the sink node could not get any data at some sampling periods. Therefore, this scheme cannot guarantee that the end-users are able to obtain the sensed data in every sampling point, and it is infeasible in most real-time monitoring applications. Motivated by [22], a novel scheme GM-OP-ELM through the combination of GM and optimally pruned extreme learning machine (OP-ELM) was proposed in [23]. Compared with GM-LSSVM, the computing speed of GM-OP-ELM improves greatly. With this scheme, prediction time can be saved immensely, but in some situations its accuracy may be lower than GM-LSSVM.

To improve the prediction accuracy and guarantee the transmission confidentiality, a novel prediction-based secure data sensing and fusion scheme using GM, kernel recursive least squares (KRLS), and Blowfish algorithm (BA) is proposed in this paper to reduce the redundant transmission in WSNs. This scheme is called GM–KRLS. During the data sensing and fusion process, both the sink node and the sensor nodes must use the same small number of recent data items and prediction mechanism to predict the data of the next period, while guaranteeing that the data sequences in the sink node and the sensor nodes are synchronous. Then the end-users can get the accurate data of all sensor nodes from the sink node in every sampling period. When the prediction error is under the threshold defined by end-users, the sensor node does not need to send the sensed data to the sink node, then both the sink node and sensor node will consider the predicted data as the sensed data in this period, otherwise the transmission between the sensor nodes and sink node will happen while encoding and decoding data based on BA [24,25]. In this way, unnecessary transmission is canceled to achieve the goal of secure data sensing and fusion. Moreover, in order to reduce the computational complexity and improve the accuracy of the prediction algorithm with small number of data items, the proposed scheme employs GM to obtain the initial predicted value, then with the help of KRLS learning algorithm [26], our scheme makes the predicted value approximate its actual value with high accuracy.

Through kernel function, KRLS puts original and nonlinear inputs into high-dimensional space to make them linear. It provides a generalization of linear adaptive algorithm. As KRLS exhibits a