Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 50-1 (2017) 2609–2614

A Secure Data Assimilation for Large-Scale Sensor Networks Using an Untrusted Cloud

Zhiheng Xu and Quanyan Zhu

Tandon School of Engineering, New York University, Brooklyn, 11201, USA. E-mail: {zx383, qz494}@nyu.edu.

Abstract: Cloud computing technologies (CCTs) enable a large-scale sensor network (LSN) to outsource the computations of data assimilation to improve its performance. However, the cyber-physical nature of cloud-enabled LSNs (CE-LSNs) introduces new challenges. Outsourcing the computations to an untrusted cloud may expose the privacy of the sensing data. To address the security issues, we proposed a secure approach to achieve data confidentiality in the outsourcing process. We develop our mechanism by combining a conventional homomorphic encryption and a customized encryption scheme. We present theorems to characterize the correctness of the encryption and investigate the estimation performance and the security of the proposed method. We also analyze the impacts of the quantization errors on the estimation performance. Finally, we present numerical experiments to consolidate our analytical results.

Keywords: Homomorphic Encryption, Cloud Computing, Sensor Networks, Data Assimilation.

1. INTRODUCTION

Large-scale sensor networks (LSNs), such as power grids (see Shnayder et al. (2004)), pollution sensing system, and transportation system, often require extensive sensing information from remote sensors. Given the measurements, the fusion center of a LSN needs to reconstruct the entire state information, known as data assimilation. The performance (e.g., efficiency and accuracy) of the data assimilation is critical to the LSNs. However, with the increasing size of the sensing data, new challenges arise in LSNs. Firstly, the growing number of the sensors makes the fusion center difficult to assemble the data due to the limited bandwidth of the networks. Secondly, the complexity of the data assimilation problem grows with the dimensions of the state of the system. The increasing number of sensors will make the computation cost impractical for the fusion center in the future.

The advent of cloud computing technologies (CCTs) makes it possible to tackle the computational issue. The integration of CCTs with control and sensing systems brings revolutionary features, such as massive computation resources and real-time data processing (see Arumugam et al. (2010)). Similarly, the integration of LSNs with CCTs, leading to a cloud-enabled LSNs (CE-LSNs), can also bring substantial benefits to the systems, especially by enhancing the efficiency of solving the large-scale data assimilation. Fig. 1 illustrates the new architecture of the CE-LSNs, which aims to improve the effectiveness of the computations in LSNs. On the one hand, all the sensors can use their local networks to upload the data to the cloud, instead of directly sending to the fusion center, whose network has a limited bandwidth. On the other hand, with massive computation resources, the cloud can solve the data assimilation problem efficiently.

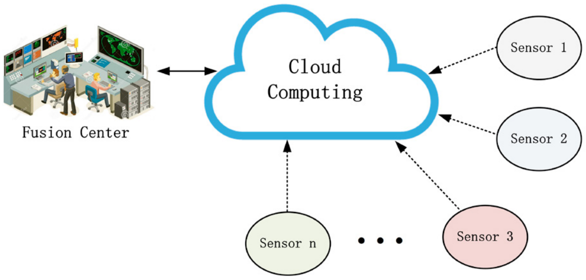

Fig. 1. The new architecture of outsourcing data assimilation for a large-scale sensor network: all the sensor directly upload data to the cloud, and the cloud return the desired results to the fusion center.

Despite the advantages of CE-LSNs, the cyber-physical security issues arise due to the untrusted cloud. Directly sending the sensing information to the cloud will expose the privacy of the sensitive information. Adversaries can steal the valuable information by hacking into the cloud or eavesdropping the communications between the cloud and sensors (see Zhao et al. (2010)). One possible approach is to apply the homomorphic encryption to encrypt data before sending to the cloud. The ideal homomorphic encryption allows the cloud to process operations, such as addition and multiplication, on the ciphertexts, generating an encrypted results, which can be decrypted to the desired results. A fully homomorphic encryption can achieve homomorphic property in multiplications and additions, but standard homomorphic encryptions can only achieve one property (see Lagendijk et al. (2013)). Partially homomorphic encryptions is easy to realize, i.e., the encryption can only achieve homomorphic property either in multiplication or addition. For example, the RSA and *ElGamal* can achieve homomorphic multiplication (see Kogiso and

2405-8963 © 2017, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2017.08.135

Fujita (2015)), while Benaloh and Paillier cryptosystems can achieve homomorphic addition.

In this paper, we aim to design a secure mechanism for LSNs to outsource the data assimilation to an untrusted cloud securely. Since the fully homomorphic encryption is inefficient to realize, we combine the traditional and customized encryption scheme to achieve fully homomorphic property. The main issue of the customized encryption is the requirement of changing the key for each mission. The high frequency in changing the key may introduce new challenges for the system. Hence, in our mechanism, only the fusion center applies the customized encryption, while the sensors use the standard encryption. We present the correctness of the and present analysis of the security and efficiency of the mechanism. Besides, the encryption introduces quantization errors in the estimation problem. We analyze the impact on the performance of the estimation problem. The main contributions of this paper are summarized as follows:

- We develop a mechanism to securely outsource the data assimilation by combining the conventional and customized encryption scheme.
- We analyze the impact of quantization on the estimation performance. The estimation performance can be tuned by selecting an appropriate quantization step.

The remainder of this paper is organized as follows. Section 3 presents the problem formulation, the design objective, and the framework. Section 4 presents fully homomorphic encryption based on the customized and the standard encryption, respectively. Section 5 provides an analysis of the quantization impacts on the estimation performance. Section 6 presents numerical experiments to evaluate the results under the proposed mechanisms. Finally, Section 7 concludes the paper.

2. RELATED WORK

The privacy issues of control systems have gained attention in recent years. Farokhi et al. have developed an algorithm to preserve the privacy of traffic estimation when the agents broadcast the sensitive information through a network (see Farokhi and Shames (2016)). Jerome Le Ny has proposed an approach to protect data privacy for nonlinear observers (see Le Ny (2015)). Different from their problems, our work focuses on preserving data privacy of a cloud outsourcing process.

The issues investigated in this work are also related to the recent literature on secure outsourcing problems. People have developed customized cryptographic mechanism to achieve fully homomorphic properties, e.g., secure outsourcing linear programming problems by Wang et al. (2011), large matrix inversion problems by Lei et al. (2013). In our previous work (see Xu and Zhu (2015b)), we have designed a customized encryption to outsource a model predictive control problem to a cloud. The primary challenge of the customized encryption scheme is the key management since it requires changing the key for every mission, increasing the computational complexity.

Besides designing customized encryption, Kogiso et al. (Kogiso and Fujita (2015)) and Farokhi et al. (Farokhi et al. (2016)) consider using standard homomorphic encryption, such as *ElGamal* and Paillier, to enhance the security level of networked control systems. However, the standard encryption can only achieve partial homomorphic property, so it is challenging to directly apply them to our problem.

3. PROBLEM FORMULATION

In this section, we first design a Kalman filter for an LSN. To improve its efficiency, the LSN outsources the computations of the Kalman filter to a cloud. However, outsourcing the computations to an untrusted cloud introduces new security challenges. To this end, we present the design objectives of the proposed secure mechanism.

3.1 System Model and the Outsourcing Kalman Filter

To describe the dynamics of an LSN, we use the following linear discrete-time model, given by

$$x_{k+1} = A_k x_k + w_k, \quad y_k = C_k x_k + v_k,$$

where $x_k \in \mathbb{R}^{n_x}$ is the state of the LSN with a given initial condition $x_0 \in \mathbb{R}^{n_x}$; $y_k \in \mathbb{R}^{n_y}$ is the sensing information, and $n_y(n_y < n_x)$ is the number of the sensors; matrix $A_k \in \mathbb{R}^{n_x \times n_x}$ describe the dynamic movement of the state x; matrix $C_k \in \mathbb{R}^{n_y \times n_x}$ describe the topology of the sensors; $w_k \in \mathbb{R}^{n_x}$ and $v_k \in \mathbb{R}^{n_y}$ are the additive Gaussian noise with zero means and covariances Σ_w and Σ_v, respectively.

In an LSN, each sensor i, for $i = 1, \ldots, n_y$, uploads its sensing data $y_i(k)$ to a fusion center. The fusion center applies Kalman filter to estimate x_k based on the feedback information y_k. The Kalman filter is given by

$$\begin{aligned} \hat{x}_{k+1} &= A_k \hat{x}_k + H_k(y_k - C_k \hat{x}_k) \\ &= A_k \hat{x}_k + H_k(y_k - z_k) = A_k \hat{x}_k + \phi_k, \end{aligned} \tag{1}$$

where $\hat{x}_k \in \mathbb{R}^n$ is the estimated state of x_k; $z_k := C_k \hat{x}_k \in \mathbb{R}^{n_y}$ is the estimated output; $\phi_k := H_k(y_k - z_k) \in \mathbb{R}^{n_x}$ is the regularized term. The explicit form of the matrix $H_k \in \mathbb{R}^{n_x \times n_y}$ is given by

$$H_k = A_k P_k C_k' (\Sigma_v + C_k P_k C_k')^{-1}, \tag{2}$$

$$P_{k+1} = \Sigma_w + (A_k - H_k C_k) P_k A_k'. \tag{3}$$

Hence, based on the sensing information y_k from each sensor, the fusion center can estimate the state x_k by computing (1)-(3), iteratively.

3.2 The Challenges and Design Objectives

In an LSN, the numbers n_x and n_y are significant. Hence, computing the equations (1)-(3) incurs complicated computations to the fusion center. To this end, the fusion center aims to outsource the computations (1)-(3) to a cloud. Besides improving the efficiency, another significant advantage of the cloud computing is the data sharing. Given the data-sharing property, each sensor can directly upload the the sensing information to the cloud though its local networks. This new architecture illustrated in Fig. 1 significantly reduces the computation and transmission cost of a large-scale data assimilation problem.

Despite the advantages, the integration with an untrusted cloud inevitably introduces new security challenges to the system. An adversary, who intrudes the cloud, can observe the sensitive information y_k, A_k and C_k in the cloud.