Ad Hoc Networks 70 (2018) 73–84

Contents lists available at ScienceDirect

Ad Hoc Networks

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

A secure data collection scheme based on compressive sensing in wireless sensor networks

Ping Zhang, Shaokai Wang, Kehua Guo, Jianxin Wang*

School of Information Science and Engineering, Central South University, Hunan, Changsha, China

ARTICLE INFO

Article history:
Received 13 January 2017
Revised 5 November 2017
Accepted 20 November 2017
Available online 22 November 2017

Keywords:
Secure data collection
Compressive sensing
Wireless sensor network

ABSTRACT

The compressive sensing (CS) based data collection schemes can effectively reduce the transmission cost of wireless sensor networks (WSNs) by exploring the sparsity of compressible signals. Although many recent works explained CS as a symmetric cryptosystem, CS-based data collection schemes still face security threats, due to the complex deployment environment of WSNs. In this paper, we first propose two feasible attack models for specific applications. Then, we present a secure data collection scheme based on compressive sensing (SeDC), which enhances the data privacy by the asymmetric semi-homomorphic encryption scheme, and reduces the computation cost by sparse compressive matrix. More specifically, the asymmetric mechanism reduces the difficulty of secret key distribution and management. The homomorphic encryption allows the in-network aggregation in cipher domain, and thus enhances the security and achieves the network load balance. The sparse measurement matrix reduces both the computation cost and communication cost, which compensates the increasing cost caused by the homomorphic encryption. We also introduce a joint recovery model to improve the recovery accuracy. Experimental evaluation based on real data shows that the proposed scheme achieves a better performance compared with the most related works.

1. Introduction

Wireless sensor networks (WSNs) have been deployed in various civilian, commercial and military areas. It consists of lots of low cost and battery powered nodes, which are often distributed in unattended environments. Secure data collection has been proposed as a crucial approach to solving the energy and security challenges.

The compressive sensing (CS)-based data collection schemes [1], which combine data acquisition with data compression, are able to surpass the bounds of the traditional Shannon–Nyquist theory by exploring the sparsity of compressible signals. It achieves high precision data recovery from less sampling data, and thus reduces the data collection cost and prolongs the life cycle of WSNs.

Many recent works [2–4] pointed out that the randomized measurement of CS also implicitly provides some kind of privacy preservation. The CS was first explained as a symmetric cryptosystem (CS-based encryption, CSE) in [2], where the pseudo-random measurement matrix was utilized as the key, and the measurement result is the ciphertext of original signals. It is a light-weight encryption scheme. The encoding step and the recovering step can be interpreted as the encryption and the decryption respectively, and thus no extra computational cost is needed. They investigated its security by considering the brute force and structured attacks, and concluded that these attacks are infeasible in practice due to the computational complexity. In their security analysis, only the encoding result is accessible to the adversary, and both the measurement matrix and signal were unknowable, which means their conclusion is based on a ciphertext-only attack model. They potentially assumed that the measurement matrix (secret key) was only used once, i.e., one-time sensing (OTS) assumption. Cambareri et al. [3] presented a quantitative analysis of CSE against the known-plaintext attack. Both the standard CS-based and the multiclass encryption schemes were exemplified, where the antipodal random matrices were adopted for encoding. They argued that their schemes feature a noteworthy level of security against known-plaintext attacks. They also admitted that their schemes are not perfectly secure. They implicitly assumed that only a plaintext-ciphertext pair is available to the adversary, which is in fact an OTS assumption. Bianchi et al. [4] focused on the ciphertext-only attack (COA) scenario. They presented a consequence that CS provides information theoretic secrecy in a particular setting and does not provide the same secrecy in a generic setting. They also hold an OTS assumption.

* Corresponding author.
E-mail address: jxwang@mail.csu.edu.cn (J. Wang).

https://doi.org/10.1016/j.adhoc.2017.11.011
1570-8705/

However, due to the complex deployment environment of WSN, CS-based data collection schemes still face security threats. Previous security analyses of CS are based on the OTS hypothesis (i.e. the secret key / measurement matrix is used only once), which is difficult to be satisfied in many situations. For example, in typical CS-based data collection scheme of WSN [1], the data acquisition, encoding and aggregation are processed at each node in a distributed model, while the data recovery is finished at the sink node independently based on measurement results received from other nodes. Each sensing node shares its measurement matrix with the sink, respectively. To realize OTS, these secret keys should be synchronously updated in each measurement round. There exist several challenges. First, it significantly increases the communication cost. The size of the measurement matrix ($N \times M$ elements) is much larger than the size of the measurement result (M elements), and thus the communication cost for the key updating is too high, especially updated frequently. Second, not all of the secret key can be used as a measurement matrix. A measurement matrix that does not satisfy restricted isometry property (RIP) may yield a biased measurement, which will significantly increase the CS recovery error. Third, characteristics of WSN also increase the difficulty of OTS. On one hand, the multi hop transmission mechanism of AdHoc network will increase the communication cost of OTS, especially when a single key distribution process contains several round of data transmission. On the other hand, most sensor nodes cannot resist attacks and are easy to capture, so nodes are not suitable for storing private information.

Hu et al. [5] presented two statistical inference attacks for CS-based data collection schemes in WSN. In their attack models, the adversary compromised a node first, and then implemented the attack to obtain the measurement matrix (secret key). The controllable event triggering attack (CETA) only changed the value of one node, while the random event triggering attack (RETA) changed all nodes value in the influenced region. They assumed that the secret key can be reused. As a result, without OTS assumption, the CSE is not secure enough.

Hu et al. also designed a new scheme to prevent these attacks, where all nodes are assumed to be synchronized, and each node adopts a distinct secure one-way hash function to generate the keys at each sample round. It is still an OTS solution. In their paper, the safe channel for the synchronous update of secure key is built on the one-way hash function and corresponding initial random numbers. In order to ensure the update of key, they assume that hash functions themselves are secure. This assumption violates the classical Kerckhoffs' principle of cryptography which stated that everything about the system, except the key, is public knowledge. At the same time, the low cost sensing node can be compromised easily, and all stored information including the hash function and current secret key (which also used as the input of the hash function for the next generation of key) will be disclosed. Although the one-way hash function can prevent the adversary to recover the historical keys and thus guarantee the security of the historical data, it cannot prevent them to generate keys used in the future measurement.

In this paper, we first analyze the security of CS-based data collection schemes, and propose two attack models for a specific application scenario, which are the known plaintext attack and the chosen-plaintext attack. Then, we present a secure data collection scheme based on compressive sensing (SeDC). We introduce an additive homomorphism encryption mechanism to enhance data privacy, and adopt a sparse matrix to mitigate the cost. We also introduce a joint recovery model to improve the recovery accuracy with the assistance of the historical measurement results.

We adopt the additive homomorphism encryption here, because the CS measurement process contains a large number of addition operations. By using additive homomorphism encryption, the intermediate nodes can directly carry out the data aggregation in the cipher domain. On the one hand, the intermediate nodes do not need to decrypt the received data, which means the private key is not necessary to be stored at the intermediate nodes, and thus the security is guaranteed. At the same time, the computation cost for decryption is also not necessary. On the other hand, the in-network data aggregation reduces the total communication cost dramatically. Each intermediate node aggregates its received packets directly in cipher domain and sends the aggregated result to it parent. The network load is near balance which can avoid the emergence of a network bottleneck.

Both encryption and decryption are computationally intensive operations, and thus the introduction of the homomorphism encryption will also increase the computation cost. The encryption operation appears in each sensor node, and the decryption process is performed by the sink. In order to reduce the computation cost of sensor nodes, we use a sparse random matrix in the proposed scheme. Because a large number of elements in the sparse matrix are zero, the number of the elements to be encrypted is significantly reduced, and the computational cost is reduced accordingly. The decryption cost is also acceptable, because the sink often has powerful computational ability and only need decrypt a few numbers of aggregated results.

A joint recovery model is introduced to improve the recovery accuracy by fully exploiting the spatiotemporal correlation existed in measurement results. We jointly recover the current measurement result together with the historical results. As these historical measurement results have been received in previous round, no extra communication cost is needed.

The major contributions of this paper are summarized as follows.

- We analyze the security problem of CS-based data collection scheme in WSN, and present two attack models for specific application scenarios.
- We propose an enhanced CS-based data collection scheme, which achieves higher security with still acceptable cost.
- We evaluate its performance with most related scheme, both in analytical and experimental comparison. The evaluation results demonstrate that the new scheme achieves good performance.

The remaining parts of this paper are organized as follows. Section 2 introduces the background knowledge. Section 3 present two attack models. Section 4 introduces the proposed scheme. Section 5 is a experimental evaluation. Section 6 briefly reviews the related work. Section 7 is a summary.

2. Background

2.1. Compressive sensing

Compressive sensing (CS) is a novel signal acquisition technology, which can surpass the bounds of the classical Shannon–Nyquist theory by combining signal sampling with data compression. $x \in R^{N\times 1}$ is signal to be sensed. According to the CS theory, the signal x should be sparse or compressible on an orthogonal basis $\psi \in R^{N\times N}$. The measurement matrix of CS is $\phi \in R^{M\times N}$, and the measurement result is $y \in R^{M\times 1}$. $y = \phi x = \phi\psi'\theta = A\theta$, while $A \in R^{M\times N}$ is the sensing matrix or recovery matrix. The measurements number M is related to K. Data compression is achieved because $M \ll N$.

The goal of CS is to recover the original data x from less measurement results y. Since the transformation of signal x (i.e., θ) is K-sparse or K-compressible. θ can be recovered from $\hat{\theta} = argmin||\theta||_0,\ s.t., y = A\theta$. $min||\theta||_0$ is a classical L_0 problem. As the L_0-norm function is a non-convex, non-smooth, discontinuity, global non-differentiable function, $min||\theta||_0$ is hard to solve.