Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 2, Issue. 9, September 2013, pg.184 – 190

RESEARCH ARTICLE

A Framework Based Integrated Dynamic Data Storage Scheme Based on Network Coding and Homomorphic Fingerprinting

K. Jothimani[1], N.Hema[2]

[1]Research Scholar, Department of Computer Science, Vivekanandha College, Elayampalayam, Tiruchengode, Tamil Nadu, India

[2]Assistant Professor, Department of Computer Science, Vivekanandha College, Elayampalayam, Tiruchengode, Tamil Nadu, India

[1] kjothimani88@gmail.com; [2] hemaguna_80@gmail.com

ABSTRACT: - Recently, distributed data storage has gained increasing popularity for efficient and robust data management in wireless sensor networks (WSNs). The distributed architecture makes it challenging to build a highly secure and dependable yet lightweight data storage system. On the one hand, sensor data are subject to not only Byzantine failures, but also dynamic pollution attacks, as along the time the adversary may modify pollute the stored data by compromising individual sensors. On the other hand, the resource-constrain nature of WSNs precludes the applicability of heavyweight security designs. To address the challenge, in this article we propose framework based integrated dynamic data storage scheme with dynamic integrity assurance. Based on the principle of secret sharing and erasure coding, we first propose a hybrid share generation and distribution scheme to achieve reliable and fault-tolerant initial data storage by providing redundancy for original data components. To further dynamically ensure the integrity of the distributed data shares, we then propose an efficient data integrity verification scheme exploiting the techniques of algebraic signature and spot-checking. The proposed scheme enables individual sensors to verify in one protocol execution the correctness of all the pertaining data shares simultaneously in the absence of the original data. Extensive security analysis shows that the proposed scheme has strong resistance against various data pollution attacks.

Keywords ---- Clustering Algorithm; Aggregation

I.INTRODUCTION

Dynamic data storage and access has recently found increasing popularity due to many reasons. First, new-generation sensor nodes with significant performance enhancements are available. Such enhancements include energy-efficient storage, greater processing capabilities, and data management abilities. It is now possible to equip sensor devices with energy-efficient storage such as the new-generation flash memory with several gigabytes and low-power consumption. Second, distributed data storage has more efficient energy consumption. In the platform, flash memory has less energy efficiency, thereby reducing the energy benefits of local data storage. However, new-generation flash memory has significantly improved its energy efficiency and computation versus communication tradeoff as well. For example, transmitting data over a radio channel consumes 200 times more energy than storing the same amount of data locally on a sensor node; radio reception uses 500 times more energy than reading the same amount of data from local storage. A measurement study in Desnoyers et al. showed that equipping the MicaZ1 platform with NAND flash memory allows storage to be two orders of magnitude cheaper than communication and comparable to computation in cost, which makes local storage and processing more desirable. Last but not least, distributed data storage achieves more robustness. Centralized storage can lead to the single point of failure, and easily attracts attacks. Moreover, it may also cause a performance bottleneck, as all data collection and access have to go through the base station. To the best of our knowledge, distributed data storage and access security in wireless sensor networks (WSNs) as a fairly new area has received limited attention so far. Previous research on WSN security issues has been focused on network communication security, such as key management, message authentication, secure time synchronization and localization, and intrusion detection; suggested moving the stored data constantly among sensors to increase dependability of one particular data item. Zhang et al. [2005] proposed a secure data access approach by using a polynomial-based key management scheme, where the mobile sinks can retrieve the network data following fixed routes. Subramanian et al. [2007] studied the distributed data storage and retrieval problem in sensor networks, and designed an adaptive polynomial-based data storage scheme for efficient data management. However, none of these schemes considered the data dependability and integrity. Rabin [1989] proposed an information dispersal algorithm (IDA) for secure data storage and transmission in distributed systems, where the original information F is dispersed into n by using erasure codes. Chessa et al. [2004] extended the idea of information dispersal in Rabin [1989], and investigated the data storage problem in the context of a redundant residue number system (RRNS).

However, the system has to maintain a large library of parameters together with a big set of moduli. [Subbiah and Blough 2005] developed a novel combination of XOR secret sharing and replication mechanisms, where each share is managed using replication-based protocols for Byzantine and crash fault tolerance. However, while the computation overhead is reduced drastically, additional servers and storage capacities are required. Shows the comparisons between our scheme and some typical data storage schemes with respect to several desired properties. Data integrity and availability is an important and necessary component of secure data storage for distributed sensor networks. Sensor data are vulnerable to random Byzantine failures as well as data pollution attacks, in which the adversary can modify the data and/or inject polluted data into the storage nodes. These attacks prevent authorized users from recovering the original data information correctly. Therefore, in order to ensure the data integrity and availability over the entire data lifetime, any unauthorized data modifications or random data corruptions due to malicious attacks and Byzantine failures should be detected as soon as possible. However, this important and unique security issue has been largely overlooked in most existing designs in WSNs. To address the problem, in this article we propose an efficient and flexible dynamic data integrity checking scheme to verify the consistency of data shares in a distributed manner. In our scheme, the data-originating sensor partitions the original data into multiple shares based on the techniques of erasure coding and secret sharing. This construction drastically reduces the communication and storage overhead as compared to the traditional replication-based techniques, and achieves reliable data storage by providing redundancy for original data components. To ensure data integrity and availability, we utilize algebraic signatures with favorable algebraic properties and a spot-checking approach, which allow the shareholders to perform dynamic data integrity checks in a random way with minimum overhead. Since the

data-originating sensor appends a distinct parity block to each data share, all shareholders can verify the distributed data shares independently in each check. A salient feature of our scheme is that the false-negative probability can be reduced to almost zero. Thus any unauthorized modifications can be detected with high probability in one verification operation. Most importantly, the proposed scheme can verify the integrity of aggregated data shares with great efficiency. Through detailed security analysis and experiments on sensor platforms Tmote Sky and iMote2, we show that the proposed scheme is highly effective and efficient and can be well suited for the resource-constrained WSNs. The rest of the article is organized as follows. Section 2 introduces the system model and attack model, and briefly describes some necessary background for the techniques used in this article. Sections 3 and 4 provide the detailed description of our proposed schemes. Sections 5 and 6 present the security analysis and performance analysis, respectively. Section 7 summarizes related work. Finally, Section 8 concludes the article. The proof of Proposition 5.2 is given in the Appendix.

II.RELATED WORK

Distributed sensor data storage involves storing data reliably on multiple sensor nodes instead of a single Source node, so that the original data can be further Accessible to any authorized data collectors in wireless Sensor networks (WSNs). Compared with the centralized Data storage, distributed data storage is of special benefit. The reliable data management in WSNs is that where individual sensors are vulnerable to failures and various attacks. Nowadays, distributed sensor data storage is universally applied to various scenarios. For instance, sensor networks are deployed in the remote environments where sensing Nodes take measurements and store data on storage nodes over a long period of time. Any authorized data collector May appear at any location to retrieve the useful data from Storage nodes [1], [2]. Despite the benefits of reliable data management, Distributed data storage is susceptible to various threats to the data availability and integrity in WSNs. In practical, Individual nodes are prone to random byzantine failures, which mean some nodes may behave erroneously or fail to behave consistently [1], [6], [31]. In addition, malicious Sensors may deliberately pollute or destroy the stored data by initiating various attacks (e.g., pollution attacks). All These phenomena result in the corruption of data availability and integrity, which correspondingly causes different Data to be recovered from different subsets of Fragments.

Furthermore, if corrupted fragments are not located and updated, the limited resources, such as memory and energy, are abused to store these incorrect fragments or perform computations on them. Therefore, we should provide both the data availability and integrity guarantees for sensor data storage in WSNs. For the data integrity, Wang et al. Argue that dynamic Integrity verification should be utilized to assure data Correctness over the period of storage [1]. Corrupted Fragments need to be identified by dynamic integrity checking, and then data maintenance is performed to replace corrupted fragments for the future data reconstruction. On the other hand, dynamic integrity verification becomes meaningless without data maintenance. Only if Dynamic integrity verification and data maintenance are both provided, can data collectors successfully obtain the stored data. In fig. 1, we illustrate the integrity verification and data maintenance for data availability and integrity Guarantees in sensor data storage. A source node generates Four fragments from three original data pieces (or called Source data pieces) using coding techniques and then stores Encoded fragments at different storages nodes. We first examine the existing data storage schemes. Most previous works focus on data distribution in P2P and wireless networks. For instance, erasure codes (e.g., Reed- Solomon codes [10]), fountain codes (e.g., LT codes [7]), Growth codes [5], and priority linear codes [11] have been Proposed in the recent years. In addition, some schemes are Proposed to provide the integrity guarantee for network Coding or erasure codes [26], [27], [29], [30], but the data Maintenance issue is not addressed in these schemes. There are also some secure data storage schemes which include the secure data access approach with polynomial-key Management, the adaptive polynomial-based data storage Scheme, etc. However, none of them achieves the requirements of data Integrity, high availability, and efficiency at the same time.

Recently, Wang et al. [1] propose a dependable and secure Sensor data storage scheme with erasure codes and Algebraic signature. In summary, distributed sensor data Storage with both data integrity and

availability guarantees has not been studied so far. To tackle the above problem, we propose a distributed Fault/intrusion-tolerant sensor data storage scheme based on network coding and homomorphic fingerprint. Unlike The traditional store-and-forward mechanism, network Coding (NC) [23] allows intermediate nodes to actively Code/mix the input packets. This novel information dissemination Technique can achieve potential throughput Improvement [24], transmission energy minimization [25], Delay reduction [21] as well as robustness enhancement [20] in communication networks. Homomorphic finger printings Proposed for the integrity checking can preserve the Algebraic structure of network coding and allow verifiers to rapidly locate corrupted fragments. It has been demonstrated to be secure, compact, and efficient with low Bandwidth and computational overheads [8]. Our scheme relies on network coding to encode the original data and Distribute encoded fragments with original data pieces and Homomorphism finger printings to storage nodes. Homomorphism Finger printings are used to fast identifies incorrect Fragments and initializes data maintenance. During the data Maintenance phase, network coding is also applied to generate alternative fragments using original data pieces and encoded fragments to guarantee the data availability.

III.METHOD OF PROCESS

The proposed sensor data storage Scheme

A. System Model

We consider a wireless sensor network with a large number of sensor nodes, each of which have a unique ID and may perform different functionalities. These nodes are deployed strategically into areas of interest and continuously sense the environments. Some of them are equipped with sufficient capacity to store the sensed data locally in a distributed manner for a certain period. We assume that these nodes have limited power supply, storage space, and computational capability. Due to the constrained resources, computationally expensive and energy-intensive operations are not favorable for such systems. In addition, for such an WSN, we also assume that basic security mechanisms such as pair wise key establishment between two neighboring nodes are already in place to provide basic communication security [Blundo et al. 1992]. However, individual sensors are not reliable since they can undergo random Byzantine failures and be compromised due to a lack of tamper-proof hardware.

B. Data Storage Model

We consider a general and framework based dynamic data storage model regarding data storage security and dependability. More specifically, we consider an adversary with both passive and active capabilities.

These are interested in modifying or polluting the data stored at the storage sensors without being detected. Once a storage sensor is compromised, an “active” adversary cannot only read the stored data but also pollute it by modifying or introducing its own fraudulent data. Furthermore, the adversary aims to remain stealthy in order to periodically or occasionally visit the network and harvest potentially valuable data. The adversary seeks to compromise as many storage sensors as possible, and as long as it remains in control of that node it reads all of the memory or storage contents and monitors all incoming and outgoing communication. Furthermore, the adversary’s movements are unpredictable and untraceable, and can compromise different sets of nodes over different time intervals. Note that, if the adversary compromises a sensor node and resides there, it can always respond to the “verifier” with the correct data and successfully pass the periodic data integrity checks. In fact, there is no way to detect such a compromised sensor if it is fully controlled by the adversary and behaves properly all the time.

C. Design Goal

Our goal is to provide various mechanisms for ensuring and maintaining the security and dependability of sensed network data under the aforementioned adversary model. Specifically, we have the following goals.

Security: To enhance data confidentiality and integrity by increasing the attacker's cost, that is, decreasing the gain on compromising individual sensors Dependability. To enhance data availability against both sensor Byzantine failures and sensor compromises, that is, minimizing the effect brought by individual sensor failures and compromises.

Dynamic Integrity Assurance**:** To ensure that the distributed data shares are correctly stored over their lifetime, so they can finally be used to reconstruct the original data by authorized users.

Lightweight**:** The scheme design should be lightweight as always in order to fit into the inherent resource-constrained nature of WSNs.

D. Dependable Initial Data Storage

To guarantee the security of the stored data, sensor nodes must encrypt the data for confidentiality. Thus only authorized user can obtain the access privilege and decrypt the data information. In addition, as sensors may exhibit Byzantine behaviors and are attractive for attacks, data dependability should also be ensured to avoid single point of failure. To address these problems and achieve a lightweight design in resource constrained WSNs, we discuss two schemes for initial data storage (a basic scheme followed by an enhanced scheme), which we believe will lead us to the final desirable solution

IV.EXPERIMENT

In this section, we evaluate the performance of our dynamic data integrity checking scheme in terms of storage, computational, and communication overheads. In our scheme, sensor data is generated and distributed stored, and retrieved and decrypted by authorized users. Since sensor nodes are usually resource constrained, they may not be able to efficiently execute expensive operations, which may become the bottleneck of the scheme. Thus, our evaluation focuses on the efficiency of the scheme. In the following section, we first present the numeric results. Then, we present our implementation results on real sensor platforms. The notation of cryptographic operations is summarized in Table I.

Table 1. Notification Summary of Cryptographic Operations

$Hash_l^t$	t hash operations with input size of l
$SymEncr^t$	t symmetric-key encryption operations
$SymDecr^t$	t symmetric-key decryption operations
$PolyEval_m^t$	t polynomial evaluations with polynomial of degree m
$ParGen_k^t$	t parity generations with vector of size k
$AlgSigGen_k^t$	t algebraic signature generations with vector of size k

Dynamic Data Integrity Check

Consider a shareholder w who initiates a data integrity check to verify the integrity of the data. It broadcasts a challenge message $\{w, seqno, \alpha, r\}$ to all the shareholders. In addition, a 1-symbol algebraic signature based on its own data share is generated and included in the challenge. Hence, the communication overhead involved in this broadcast message is $5 \cdot q$ bits. Upon receiving the challenge, each shareholder needs

to compute a 1-symbol algebraic signature and return it to the check initiator. Thus, for each shareholder (including the initiator), the computational cost is just *AlgSigGen*1*k* . The communication overhead involved in every response message *sigα*(*Si*) is *q* bits. After obtaining all *sigα*(*Si*)s (*i* = 1, . . . , *n*), each shareholder can act as a verifier to check the integrity of the data. It is clear that in this step the computational cost at each node is *ParGen*1*k* + *AlgSigGen*1*k*

Table 2. Performance Comparison Under Different Sensor Platforms

	Tmote Sky (8MHz)		iMote2 (13MHz)	
Metric\Parameter	m = 10, n = 20	m = 20, n = 30	m = 10, n = 20	m = 20, n = 30
Coefficient matrix gen. time (ms)	4.7	15.7	0.47	1.4
RS coding time (ms)	135.0	271.0	8.90	17.8

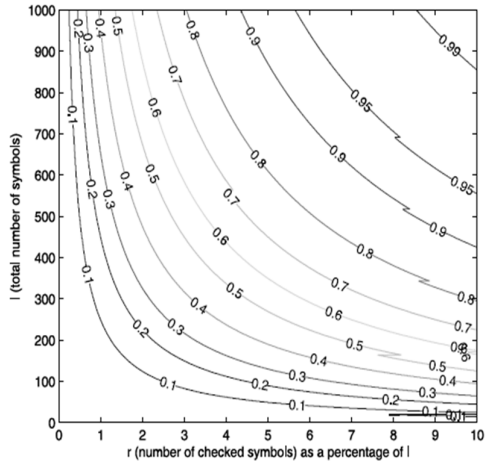

Fig.1. The detection probability of data modification P_d *(z/l = 5%)*

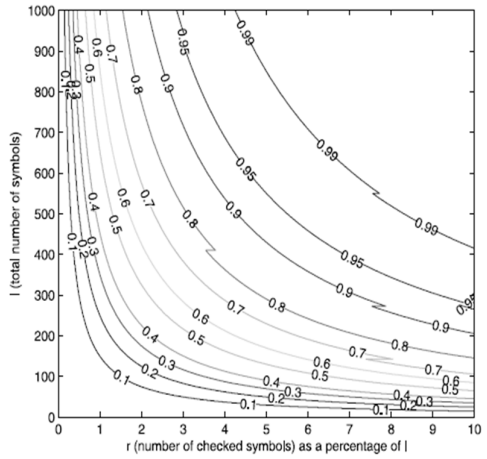

Fig.2. The detection probability of data modification P_d *(z/l = 10%)*

V.CONCLUSION

In this article, we propose a framework based integrated dynamic data storage scheme with dynamic integrity assurance in wireless sensor networks. We utilize perfect secret sharing and erasure coding in the initial data storage process to guarantee data confidentiality and dependability. To ensure the integrity of data shares, an efficient dynamic data integrity checking scheme is constructed based on the principle of algebraic signatures and a spot-checking method. In contrast to the existing approaches, more desirable properties and advantages are achieved in our scheme. Furthermore, through detailed performance and security analysis and experiments on real sensor platforms, we show that the proposed scheme is highly secure and efficient, and thus can be implemented in the current generation of sensor networks.

REFERENCES

[1] Q. Wang, K. Ren, W. Lou, and Y. Zhang, "Dependable and Secure Sensor Data Storage With Dynamic Integrity and Assurance," Proc. IEEE INFOCOM, 2009.
[2] A.G. Dimakis and K. Ramchandran, "Network Coding for Distributed Storage in Wireless Networks," Networked Sensing Information and Control, Signals and Communication Series, V. Saligrama, Springer, 2008.
[3] A.G. Dimakis, P.B. Godfrey, M.J. Wainwright, and K. Ramchandran, "Network Coding for Distributed Storage System," Proc. IEEE INFOCOM, 2007.
[4] R. Rodrigues and B. Liskov, "High Availability in DHTs: Erasure Coding vs. Replication," Peer-to-Peer Systems, 2005.
[5] A. Kamra and V. Misra, "Growth Codes: Maximizing Sensor Network Data Persistence," Proc. ACM SIGCOMM, 2006.
[6] Y. Lin, B. Liang, and B. Li, "Data Persisitence in Large-Scale Sensor Networks with Decentralized Fountain Codes," Proc. IEEE INFOCOM, 2007.
[7] M. Luby, "LT Codes," Proc. 43rd Symp. Foundations of Computer Science (FOCS '02), 2002.
[8] J. Hendricks, G.R. Ganger, and M.K. Reiter, "Verifying Distributed Erasure-Coded Data," Proc. 26th Ann. ACM Symp. Principles of Distributed Computing (PODC '07), 2007.
[9] A.Z. Broder, "Some Applications of Rabin's Fingerprinting Method," Sequences II: Methods in Comm., Security, and Computer Science, pp. 143-152, 1993.
[10] I.S. Reed and G. Solomon, "Rolynomial Codes Over Certain Finite Fields," J. Soc. for Industrial and Applied Math., vol. 8, pp. 300-304, 1960.
[11] Y. Lin, B. Liang, and B. Li, "Priority Random Linear Codes in Distributed Storage Systems," Proc. IEEE Conf. Distributed Computing Systems (ICDCS), 2007.
[12] J. Girao, D. Westhoff, E. Mykletun, and T. Araki, "Tinypeds: Tiny Persistent Encrypted Data Storage in Asynchronous Wireless Sensor Networks," Elsevier Ad Hoc Networks, vol. 5, no. 7, 1073- 1089, 2007.
[13] R.D. Pietro, L.V. Mancini, C. Soriente, A. Spognardi, and G. Tsudik, "Catch Me (If You Can): Data Survival in Unattended Sensor Networks," Proc. IEEE Sixth Ann. Int'l Conf. Pervasive Computing and Comm., 2008.
[14] D. Ma and G. Tsudik, "Forward-Secure Sequential Aggregate Authentication," Proc. IEEE Symp. Security and Privacy, 2007.
[15] S. Chessa, R.D. Pietro, and P. Maestrini, "Dependable and Secure Data Storage in Wireless Ad Hoc Networks: an Assessment of DS2," Proc. WONS, 2004.
[16] W. Zhang, H. Song, S. Zhu, and G. Cao, "Least Privilege and Privilege Deprivation: Towards Tolerating Mobile Sink Compromises in Wireless Sensor Networks," Proc. ACM MobiHoc, 2005.
[17] A. Subbiah and D.M. Blough, "An Approach for Fault Tolerant and Secure Data Storage in Collaborative Work Environments," Proc. Int'l Workshop Storage Security and Survivability, 2005.
[18] N. Subramanian, C. Yang, and W. Zhang, "Securing Distributed Data Storage and Retrieval in Sensor Networks," Proc. IEEE Fifth Int'l Conf. Pervasive Computing and Comm. (PERCOM '07), 2007.
[19] M.O. Rabin, "Efficient Dispersal of Information for Security, Load Balancing, and Fault Tolerance," J. ACM, vol. 36, no. 2, pp. 335-348, 1989.
[20] C. Fragouli, J.Y. Boudec, and J. Widmer, "Network Coding: an Instant Primer," ACM SIGCOMM Comm. Rev., vol. 36, 2006.

