Received: January 11, 2022. Revised: March 14, 2022.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Energy Aware Compressive Sensing Assisted Secure Data Survivability and Data Collection in Unattended Wireless Sensor Networks

Nischaykumar Hegde[1]* **Linganagouda Kulkarni[2]**

[1]*Visvesvaraya Technological University, Belagavi, India*
[2]*KLE Technological University, Hubli, India*
* Corresponding Author's Email: meetnischay@gmail.com

Abstract: Mobile sink-based sensor networks deployed in unattended environment are characterized by infrequent sink visits. In these applications, ensuring the survivability and confidentiality of data is important till it is collected by the sink. The data must also be prevented from collection by agents other than sink and leakage by neighbours. Data collection in unattended environment must also be energy efficient to prolong the life time of the network. This work proposes a novel Energy Aware Adaptive Compressive Sensing (EA-ACS) assisted secure data survivability scheme which is able to optimize energy consumption at cost of data accuracy level desired by applications. The solution introduces three novelties: adaptive compressive sensing for reducing the foot print of data and energy consumption during transmission, a novel bi party authentication scheme for secure data collection by sink and low energy overhead piggy bagging for data replication. Through performance in NS2, the proposed solution is found to provide at least 6% higher data survivability and 20% lower energy consumption compared to existing works.

Keywords: Data survivability, Data confidentiality, Mobile sink, Unattended environment, EA-ACS.

1 Introduction

Wireless sensor network deployed in unattended environment has battery powered sensors distributed over large geographical area. The sensors cache the sensed parameters locally till a mobile sink arrives to collect it [1]. Multi hop routing is generally restricted in these networks as it can deplete the energy of the sensing nodes [2]. Un-attended wireless sensor networks find application in military, forest and ocean monitoring [3]. These applications demand the network to be energy efficient with prolonged life time, the data to be secure, data to survive node failures to an extent and data from being collected by attackers.

The concept of un-attended wireless sensor networks (UWSN) was first introduced in [4] to collect data in hostile environments. In these networks, sink visits the nodes at regular intervals to collect data. Since the data is cached in the node till sink collects it, attackers attempt to steal the data, corrupt it or even destroy nodes. Thus, UWSN's has following core issues of data confidentiality, survivability and integrity.

Cryptography can be used to secure the data and provide integrity. Survivability was ensured through distribution of encrypted data to multiple nodes. Many solutions have been proposed along the lines of cryptography and distribution [5]. Though the existing works have addressed the constraints of energy consumption in providing three requirements of confidentiality, survivability and integrity, very few have addressed the coverage of survivability. It denotes the volume of information that can be survived in the presence of failures. This paper addresses the problem of providing confidentiality, survivability and integrity satisfying the constraints of minimal energy consumption and maximum coverage of survivability. Since the data storage capability of node is limited, it becomes necessary to provide maximum survivability under constraints of data storage. Sensing nodes in UWSN are energy constrained and it is necessary to meet requirements of confidentiality, survivability and integrity with

constraint of minimal energy consumption. This work proposes an energy aware compressive sensing assisted secure data survivability scheme satisfying the constraints of minimal energy consumption and maximum coverage of survivability.

Compressive sensing [6] with adaptive transform coding is applied to maximize the coverage of survivability. Transform coding in frequency transformation domain along with adaptive Gaussian transformation matrix is applied to provide data confidentiality. A novel low complexity bi party mutual authentication is enforced to secure data collection. Even if data is leaked due to compromise, it becomes difficult to reconstruct without the knowledge of the transform coding and transformation matrix parameters. Following are the novel contributions of this work.

1. Adaptable Compressive sensing along with transform coding scheme to provide data confidentiality and increased coverage of survivability for varied level of data accuracy desired by the applications.
2. Low complexity bi party mutual authentication for secure data collection.
3. Frequent re-keying without necessity of key exchange between sensor nodes and sink to prevent from inference attacks.
4. Data replication to far away nodes in energy efficient manner using piggy bagging.

The solution presented in this work has strong resilience against data confidentiality attacks due to representation in compressed form and only sink can learn original data with its keys. Due to representation in compressed form, the volume of data that can be survived is also high in proposed solution. Due to bi-party authentication before data collection, the proposed solution is secure against data theft attacks. Since data is replicated, the proposed solution is also secure against node failure attacks. There are no existing data survivability schemes considering all the requirements of data confidentiality, data theft, node failures and higher survivability volume and the proposed solution is first of this kind.

The rest of paper is organized as follows. Section 2 details the existing works on data survivability in WSN. The proposed energy aware adaptive compressive sensing scheme for data survivability in WSN is detailed in section 3. Analysis of the various features of the proposed solution is presented in section 4. The simulation results and comparison to existing works are detailed in section 5. Conclusion of the paper is presented in section 6.

2 Related work

Aliberti et al [7] proposed a data survivability solution for WSN based on susceptible (S), Infected (I) and susceptibles (S) model. The data is replicated to a minimum number of nodes probabilistically in the network. The replication is done in such a way to minimize the resource consumption for replication and fast collection time. The work considered only survivability and did not consider confidentiality and integrity of the data. Bahi et al [8] proposed an efficient distribution algorithm for ensuring the data survivability in the presence of attackers. The attack considered in this work is denial of service based node failures. By using two epidemic models for data distribution, the data is replicated in multiple places. But the work did not consider the energy and storage limits of the sensor nodes. Elsafrawey et al [9] proposed a cooperative hybrid self-healing scheme to ensure security and confidentiality of data in UWSN. Sensors generated hash based forward keys and distributed to each node. Each node encrypts the data using the hash key and split the encrypted data to parts using Reed Solomon (RS) code. The parts are then distributed to sensor whose attack vulnerability is less than a threshold. Each sensing node is evaluated in terms of their compromise tendency which creates a higher energy overhead and decreases the life time of the network. Sen et al [10] proposed a data confidentiality and survivability scheme using data replication and key distribution. The sensor node generates the key and encrypts the data. This encrypted data is distributed to neighbour nodes. The key used for encryption is split into shares using Shamir secret sharing algorithm and shares are distributed to sensor nodes. Sink collects the shares of keys and encrypted data from the nodes. Sink reconstructs the key and decrypt the data. Since no attacker can collect the keys distributed across many nodes, it becomes difficult to decrypt the data by the attacker. The energy consumption is very high in this approach due to communication of both data and key parts. Cheng et al [11] proposed a secure distribution scheme based on erasure coding. The data is encrypted using a forward secrecy key agreed with the sink. The encrypted data is split into shares and distributed to random two hop neighbours. Sink collect the shares and reconstructs the data. The method addressed reliability and confidentially but did not consider survivability. Lim et al [12] used fragmentation to ensure confidentially of the data. The sensor data is fragmented and distributed to other nodes. The approach distributed each fragment to several multi hops away. Though the approach was able to provide confidentially and attack resistance,

the energy consumption is very high in this work due to unconstrained multi hop routing. Choi et al [13] proposed a scheme for energy efficient fragment distribution by placing constraints on node selection for distribution. The nodes are located in the path and distribution is done among several nodes in same path. But the approach is vulnerable to guessing attacks. Santos et al [14] analysed different cryptographic algorithms in real sensor platforms and found that re-encryption based algorithms provided stronger defence against attacks but their computation cost is very high due to use of exponential operations. Liang et al [15] provided data confidentiality using regeneration codes. The data is encoded using regeneration codes and distributed as fragments to other nodes. With the use of regeneration codes, the data is secure against tampering and collusion attack. Tang et al [16] proposed secure data collection scheme called aggregation signature based trust routing. A light weight aggregation signature is calculated for aggregated data and used for integrity verification. The approach addressed integrity and did not consider confidentiality and survivability. Maia et al [17] proposed a distributed data storage protocol called ProFlex for heterogeneous wireless sensor nodes. The data distribution is handled by powerful nodes in the network. These powerful nodes use their long range communication to distribute data at far off locations, so that the data is secure against guessing attacks. But the method does not consider data confidentiality and compromise of the powerful nodes. Talari et al [18] proposed a distributed compressive data storage technique. Each node does a probabilistic broadcast of compressed samples of data. Data collector collects these compressed samples and reconstructs the original. There is no protection against any unauthorized data collection and data corruption in this work. Albano et al [19] used erasure coding with probabilistic data distribution to ensure data confidentiality. But the method has low coverage of survivability. Cuevas et al [20] proposed a data centric storage system with long term storage facility. The home node for storing data is found through a node specific localization algorithm and its frequency shifted every epoch. The solution supports anonymous storage but it does not address confidentiality and home node compromise. Zhang et al [21] proposed a distributed data storage and data collection scheme based on compressive sensing. Nodes broadcast their packets to neighbourhood locations. Sink samples only certain nodes using compressive sensing. The method does not support data confidentiality. Nguyen et al [22] proposed a data storage scheme based on compressive sensing on clustered wireless sensor network. Node sends their sensing measurements to their cluster heads. Cluster head does compressive sensing and stores the compressive sensing data. Sink collects the compressive sensing data and reconstruct the original data. Compressive sensed data can be collected by any attacker and reconstructed. The method does not provide any security against any unauthorized data collection. Gong et al [23] proposed a distributed data storage scheme, exploiting the spatiotemporal correlation between the sensor nodes. The sensor readings are collected from nodes in an energy efficient manner using compressive sensing. Though this method reduces the communication energy, there is no protection for data being collected by attacker. Alrashed et al [24] proposed a solution for data confidentiality in UWSN with use of forward secrecy and co-operative data distribution. Key evolution using forward hashing function is done to encrypt the data. The encrypted data is distributed with cooperative data distribution to randomize the distribution every time, so that it is secure against any guessing attacks. The method is insecure against node compromise attacks. Wang et al [25] addressed data confidentiality and integrity for multi-dimensional data from sensing nodes. The data is encoded using bucket partitioning scheme. The encoded data is then encrypted using sequential encryption. The advantage in this scheme is that any query can be executed on data without need for decryption. The scheme also ensures integrity of data. Though this work addressed data confidentiality, survivability is not considered. Monika et al [28] proposed a mobile sink based data gathering solution for under water sensor networks. Mobile sink stops at rendezvous points and collect data. The solution focussed only on energy efficiency in data collection without consideration for security and survivability. Goyal et al [29] proposed a clustering based data collection strategy where aggregated data at cluster heads are sent securely to sink. The data aggregation is protected and data collection is authenticated. But the solution did not consider survivability in case of cluster head failure. From the survey, it can be seen that there is no existing solution which meets requirements of confidentiality, survivability, integrity and increased coverage of survivability with constraint of minimal energy consumption in unattended wireless sensor networks. This paper work is designed to address this problem.

Algorithm 1: Transformation matrix generator

Input: Length of signal N and number of measurements M, binary session key of 16 bits
Output: Array of Sensing Matrix (M*N)
$m \cong Q/P$
m_1 = integer value below m
m_2 = integer value above m
$Nm_2 = Q - P \times m_1$
$Nm_1 = P - Nm_2$
$Rpm_2 = r_1 and r_M$
$Rpm_1 = [r_1, r_2, r_3, \ldots r_M] - Rpm_2$
$rowt_1 = \{1_1, 1_2, 1_3, \ldots 1_{m1}, 0_1, 1_2, 1_3, \ldots 1_{N-m1}\}$ //$m1$ ones and $N - m1$ zeros
$rowt_2 = \{1_1, 1_2, 1_3, \ldots 1_{m2}, 0_1, 1_2, 1_3, \ldots 1_{N-m2}\}$ //$m2$ ones and $N - m2$ zeros
For k=1 to M do
  If $r_k \in Rpm_1$ then
    $row_k = rowt_1$
    $rowt_1 = circularshiftrowt_1 rightbym_1 times$
    $rowt_2 = circularshiftrowt_2 rightbym_1 times$
  Else
    $row_k = rowt_2$
    $rowt_1 = circularshiftrowt_1 rightbym_2 times$
    $rowt_2 = circularshiftrowt_2 rightbym_2 times$
  End if
End for
Diagonalblock, $D_b = \{row_1^T, row_2^T, \ldots . row_M^T\}^T$
For i=1:16
  If sessionkey(i)==1
    Circular left shift D_b
  Else
    Circular right shift D_b
  End
end
$M_s = matrixwithUnumberofD_bs$
End if
allM = []
L=$m1$
For i=1: $m1$
  $M_{temp} = M_s$
  For j=1:rows in M_{temp}
    $M_{temp}[j][L] = 0$
  End
  L=L-1;
allM[i] = M_{temp}
End for
Return allM

3 Energy aware adaptive compressive sensing (EA-ACS)

The proposed energy aware adaptive compressive sensing modifies the conventional compressive sensing algorithm with joint consideration of survivability and confidentiality to achieve two important goals of maximizing the coverage of survivability and minimizing the energy consumption. The proposed solution is based on adaptation of following two important functionalities

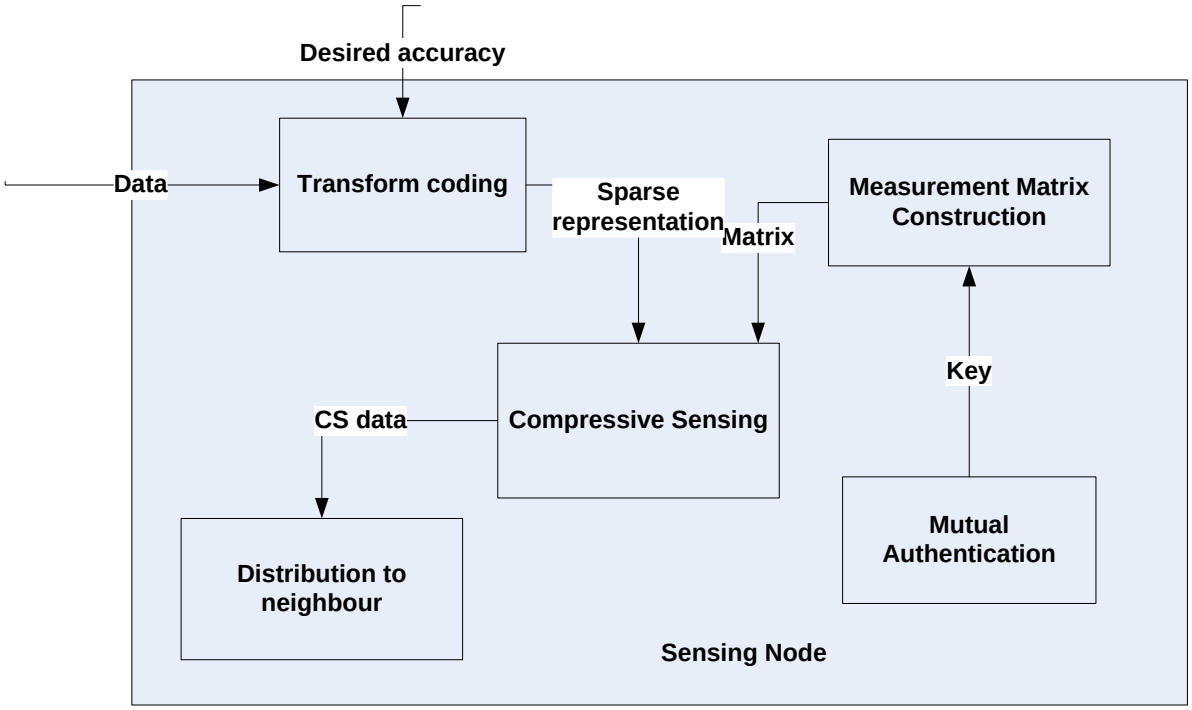

Figure. 1 Architecture of sensing node

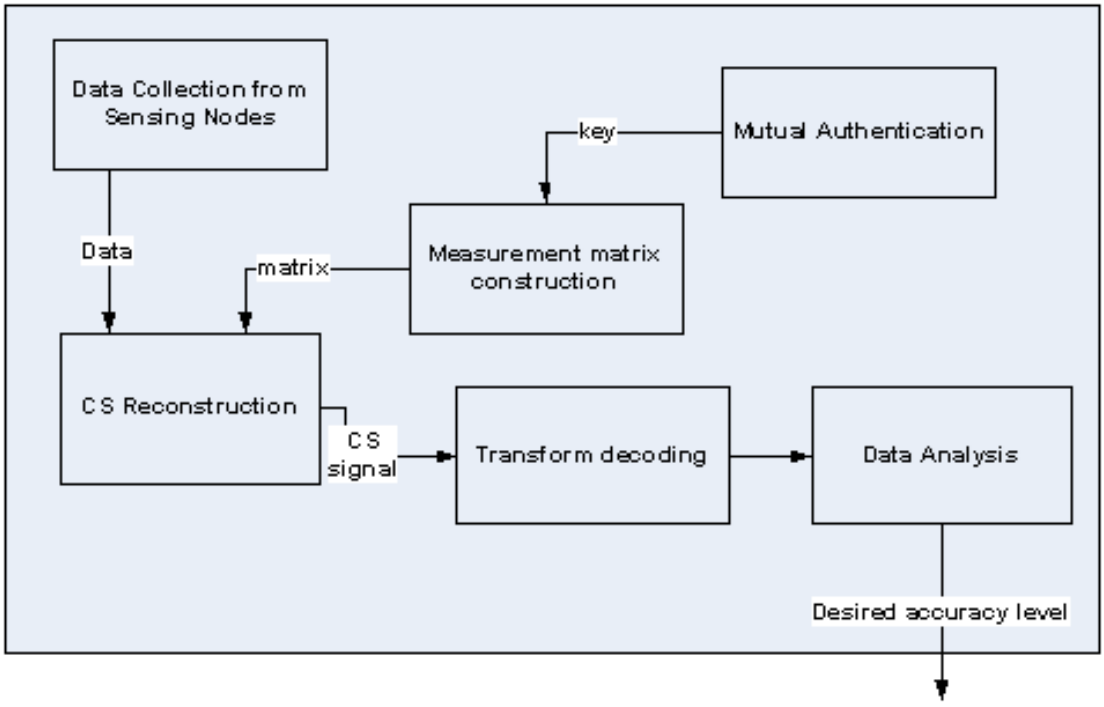

Figure. 2 Architecture of sink

1. Compressive sensing
2. Transform coding

Compressive sensing (CS) is a technique to reconstruct the original signal from a fewer number of observations. By exploiting the sparsity in the Nyquist limit and CS allows to represent the compressed signal below Nyquist rate. The encoding is fast and it effectively preserves the structure of the signal due to use of non-adaptive linear projections. Compressive sampling acquires only the important information of the signal. Reconstruction of original signal is done from the projections using different optimization techniques. The compressive sensing is able to achieve higher data rate by sampling below the Nyquist rate. This is possible because of working on sparse representation of the signal. The compression

effectiveness is proportioned to sparsity in the signal.

Let x be the original signal and its sparse representation in some orthogonal basis is given as

$\varphi = \{\varphi_1, \varphi_2, .. \varphi_N\}$ where the length of the signal is N. The signal x can be represented in term of K linear combination of basis functions ($K \ll N$) as

$$x = \sum_{i=1}^{K} \emptyset_{ni}\, \varphi_{ni} \tag{1}$$

Where $\varphi_{ni} \in \varphi$. Let $\emptyset = [\emptyset_1, \emptyset_2, ... \emptyset_N]^T$ be the vector of coefficients of the signal x in φ. The random measurement of the signal x is given as

$$\begin{gathered} y = \phi\emptyset \\ \phi: M \times N \\ K < M \ll N \end{gathered} \tag{2}$$

ϕ is the uniform random measurement matrix, y is measurement vector of signal x. $\emptyset$ is the coefficients for the signal x and $M = cK(c < 1)$ is the number of measurements to be done for a successful reconstruction of the signal. The reconstruction can be done with higher accuracy when all entries of ϕ is taken from a Gaussian distribution.

The transformation matrix used for compressive sensing is a randomly generated matrix which is agreed between the encoding and decoding side. When there is a mismatch in the transformation matrix between the encoding and decoding side, reconstruction fails or become erroneous. In this work, this feature of transformation matrix is used to provide data confidentiality. Instead of complete random generation of transformation matrix, a session key is used for generating the transformation matrix. This session key is generated by mutual authentication between sensor node and sink.

Most compressive sensing methods use Gaussian random matrix as the sensing matrix. The Gaussian random matrix is denser due to non-zero non-integer values. It results in higher computational and storage complexity. Hardware implementation costs are also increased due to Gaussian random matrix. Work in [27] proposed a spare binary matrix as a replacement for Gaussian random matrix. Compared to Gaussian random matrix, the sparse binary matrix has a smaller number of non-zero values and due to it; the computation complexity, delay and storage requirements are reduced. In this work we adapt the work in [27] to generate the sparse binary matrix based on input key given as Transformation matrix generator (Algorithm 1). For M measurements in a signal of length N, a sensing matrix of dimension M*N is generated. The sensing matrix is generated as in Fig. 3. The dimension of diagonal block is taken as $\frac{Q}{P} \cong m$. An integer value below and above m, m_1 and m_2 is assigned. The total rows in D are divided into two categories of rows with m_1 ones and rows with m_2 ones. The number of rows with m_2 ones is given as

$$Nm_2 = Q - P \times m_1 \tag{3}$$

Where $D = Q \times P$. The number of rows with m_1 ones is given as

$$Nm_1 = P - Nm_2 \tag{4}$$

Diagonal block D is formed by circular shifting of rows with m_1 and m_2 ones and stacking them. The diagonal block is placed in a zero matrix of dimension M*N to create the sensing matrix. This sensing matrix is taken as base matrix and for a iteration equals to continuous ones, each one is replaced with zero in all rows and a subsequent sensing matrix is generated. The session key needed as input for the transformation matrix generation is created through mutual authentication between sensor node and sink. The authentication process between sensing node and sink happens through a bi party mutual authentication process and key is generated at each end without communicating the key.

Sensing node initiates authentication with the sink

It calculates two numbers $S_1$and S_2 as

$$S_1 = a.\left(P_{pub} + h_1(G).P\right) \tag{5}$$

$$S_2 = h_1([e(P,P)]^a, G) \oplus (ID, b) \tag{6}$$

ID is the identifier of the sensing node.

Where a, b are two random number. $S_1$and S_2 are sent to sink. Sink finds the ID of the sensing node as below

$$k = e\left(\frac{1}{S_1 + h_1(G)}.P, S_1\right) \tag{7}$$

$$v = h_1(k, G) \tag{8}$$

$$(ID, b) = S_2 \oplus v \tag{9}$$

At the sink, it computes two number R_1 and R_2 as follows

$$SK = b.S_1 \tag{10}$$

$$R_1 = b.\left(h_1(G).P + P_{pub}\right) \tag{11}$$

$$R_2 = b.\left(h_1(h_1(ID), h_1(G), b, ID, G, SK)\right) \tag{12}$$

R_1 and R_2 are sent back to sensing node. On receiving it, sensing node computes $\bar{R}_2$ as

$$SK = H_1(b.R_1) \quad (13)$$

$$\bar{R}_2 = H_1\big(h_1(h_1(ID), h_1(G), b, ID, G, SK)\big) \quad (14)$$

If the computed $\bar{R}_2$ is equal to R_2, the sink is authenticated. Next, sensing node computes,

$$S_3 = \big(b + H_1(ID, SK, R_2, S_1)\big) \times \left(\frac{P}{S_1 + h_1(ID)}\right) \quad (15)$$

and send it to sink

On receiving S_3, sink verifies the validity of S_3 by checking following relation

$$e\left(S_3.\left(P_{pub} + h_1(ID).P\right)\right) = k.g^{h_1(ID,SK,R_2,S_1)} \quad (16)$$

If the relation is true, then sensing node is authenticated at the sink

The random number b generated by sensing node is available at both sink and sensing node, the session key is created for communication as

$$sk = h_1(ID) \oplus h_1(G) \oplus h_1(b) \quad (17)$$

The effectiveness of compressive sensing depends on sparsity distribution in the input. When data is not sparse in time domain, it is converted to frequency domain and thresholded to introduce sparseness. This process is called as transform coding. But a higher value for threshold can create the error between original data and reconstructed data but it provides the scope for increasing the coverage of survivability. Lower value for threshold reduces the error in reconstruction but reduces the scope for increasing the coverage of survivability. Threshold is adapted in this work based on application desired accuracy.

The transform coding adapted in this work first converts the data in time domain into frequency domain using discrete fourier transform (DFT). The output of DFT is coefficients. A threshold is fixed and coefficients with values less than the threshold are made as 0. By this way signal is converted to sparse in frequency domain. The sparse representation is then multiplied by the measurement matrix to get the observation vector of length M. Sparse representation has helped compression of signal of interest. The accuracy of reconstruction depends on two factors of measurement matrix ϕ and measurement vector y. When the matrix ϕ, φ has near ortho normal restricted isometric property it is possible to recover all K coefficients from the M measurement of y. There are many optimization techniques to reconstruct the sparse signal and out of it l_1 norm minimization and convex optimization are most used. l_1 norm minimization attempts to find the vectors with smallest l_1 norm

$$min||x||_n \; subject \; to \; \phi x = y \quad (18)$$

The sparsity is decided by the threshold used for transform coding. The threshold is made adaptive to application desired accuracy in this work. The threshold (T) is calculated as

$$T = T_b * \max(X_i) \quad (19)$$

Where X_i is DFT transformation $x_i(k)$ given as

$$X_i(k) = DFT\big(x_i(n)\big) = \sum_{n=0}^{N-1} x_i(n) e^{-j\frac{2\pi}{N}kn} \quad (20)$$

A threshold base (T_b) is calculated by conducting a test run with different values of T_b and measuring the reconstruction accuracy. A linear correlation is established between the reconstruction accuracy and T_b as

$$T_b = \alpha + \beta_0 D_A \quad (21)$$

where α is the bias in linear regression fit and β_0 is the coefficient and D_A is the application desired accuracy. The architecture of sensing node and sink in EA-ACS is given in Fig. 1 and Fig. 2. The behaviour of sensor node and sink in proposed solution is detailed below.

3.1 Sensor node behavior

The sensor node caches the sensed data. The threshold for transform coding is calculated based on the application desired accuracy level. Transform coding is done on the cached sensed data. Bi-party mutual authentication between sensed node and sink results is a session key. This session key is used to calculate the transformation matrix as in Algorithm 1. Compressive sensing is done to transform data with the generated transformation matrix. The compressed sensed data is then replicated as follows.

From the survey, there are two general approaches for data replication. (a) Distribution in neighbourhood (b) Distribution in far off places

Distribution in neighbourhood exposes the data to guessing attacks and risk to data survival by node destruction. Distribution in far off places, involves

Table 2. Simulation parameters

Parameters	Values
Number of Nodes	250
Communication range	100m
Area of simulation	1000m*1000m
Node distribution	Random distribution
Simulation time	15 minutes
Interface Queue Length	50
MAC	802.11
Compromised node percentage	5% to 25%
Initial energy	100 Joules

multi hop routing increasing the energy consumption. In this work, a piggy bagging scheme is proposed for data distribution.

A sensor node distributes the data to its 1 hop neighbourhood with a hop count values as T. The 1 hop neighbouring nodes piggy bags this data to their own data, reduce the T value by one and forward to next neighbourhood. The data spreads through piggy bagging till T reduces to zero. By this way distribution to far off places is achieved without much energy consumption but by compromising on the time for spread.

3.2 Sink behavior

Sink node initiates bi-party mutual authentication with sensor node before collecting any data from it. Through this way, the data collection is authenticated and risk of data being collected by un-authorized entity is restricted.

Sink collects the data from node after authentication. The transformation matrix is generated for the corresponding data of the sensor node based on the past session keys stored for that node in the sink. Compressive sensing reconstruction process is done with this transformation matrix. After reconstruction, inverse DCT is done to get back the original data.

Sink propagates the desired accuracy level needed for applications to node on its visit to the nodes.

4 Analysis

Data confidentiality attacks: Data can be collected from the sensor nodes only after a mutual bi party authentication with sink. Even if a sensor node is compromised and data is stolen, without the transformation matrix, it becomes difficult to reconstruct the data back. By this way the proposed work is secure against data theft attacks.

Data survivability attacks: Node neighbourhoods can be destroyed and data survivability can be affected. But the proposed solution distributes the data in piggy bagging manner to far off places and the pattern for distribution is influenced by neighbourhoods generated data. Thus, it is difficult to learn the replication behaviour and destroy neighbourhood to risk data survivability in proposed solution.

Energy consumption: In general, energy consumption in compressive sensing solution is better than Erasure coding or Shamir secret sharing methods, due to ability to compress large volume of data. In the proposed solution, compressive sensing is further improved with adaptive transform coding. Due to this, in the same energy cost for replicating data, more data can be replicated. The reconstruction part and transform decoding part are computationally intensive and consume more energy, but this is moved to sink in the proposed solution.

Coverage of survivability: Due to use of combined transform coding and compressive sensing, the volume of data that can be hidden in a packet is increased. Thus, more amount of information can be survived in packet in the proposed solution compared to erasure coding and Shamir secret sharing schemes. Due to this the proposed solution has higher coverage of survivability.

5 Results

The proposed solution was simulated in NS2 and performance is compared against rendezvous point-based data gathering proposed by Monica et al (2022) [28], secure authentication with data aggregation proposed by Goyal et al (2020) [29] and secure data collection scheme proposed by Miao et al (2021) [7]. The performance is compared in terms of data survival probability, difficult level of confidentiality, storage and energy cost. The simulation is conducted in following setup.

Survival probability is measured for different percentage of compromised nodes and the result is given in Fig. 3. From the results, it can be seen that even for an increase in attacker percentage from 5% to 25%, the data survival probability has reduced only by 4% in proposed solution compared to 12% in Monica et al , 8% in Maio et al and 12% in Goyal et al (2018). Distributing the data to far locations using piggy bagging has provided higher survival probability in

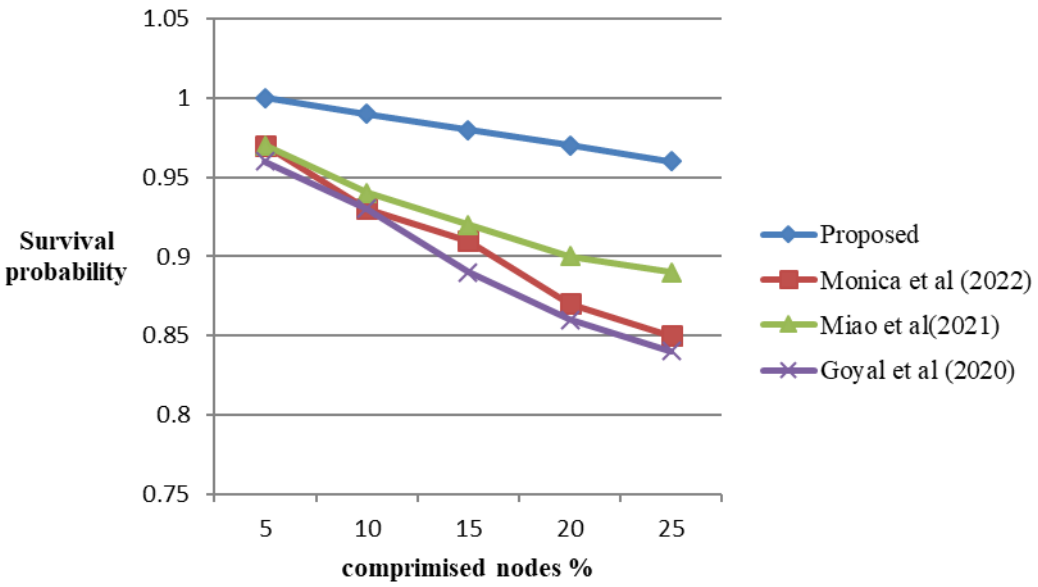

Figure. 3 Compromised nodes vs survival probability

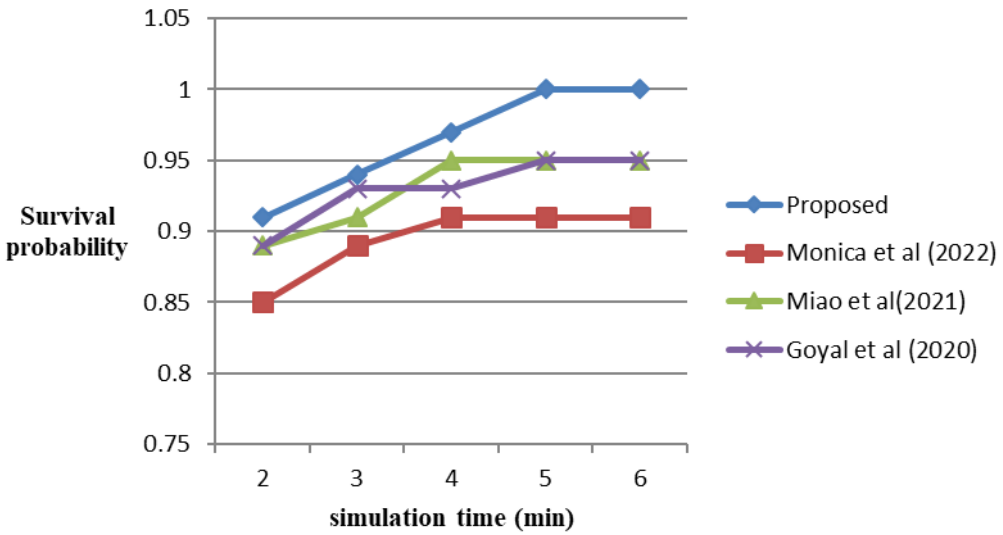

Figure 4 Simulation time vs survival probability

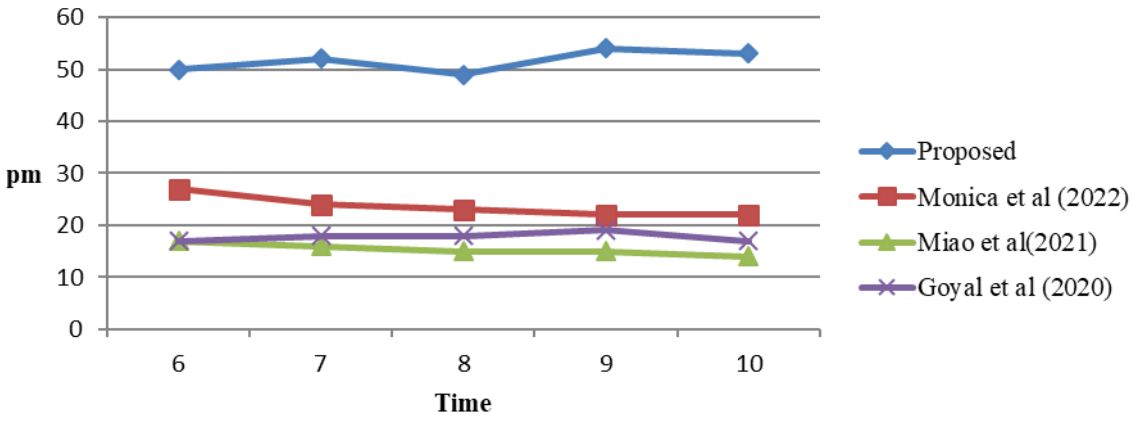

Figure. 5 Data confidentiality pm metric

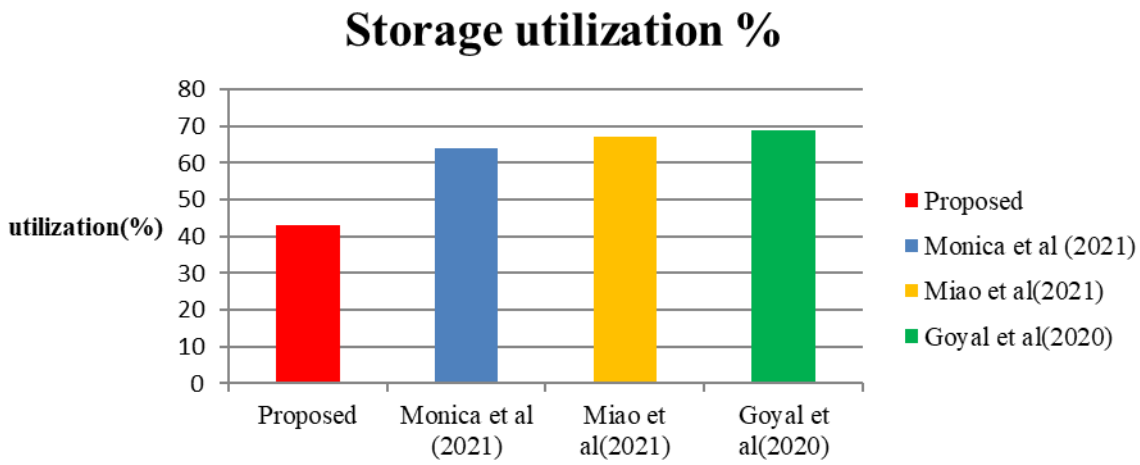

Figure. 6 Storage utilization %

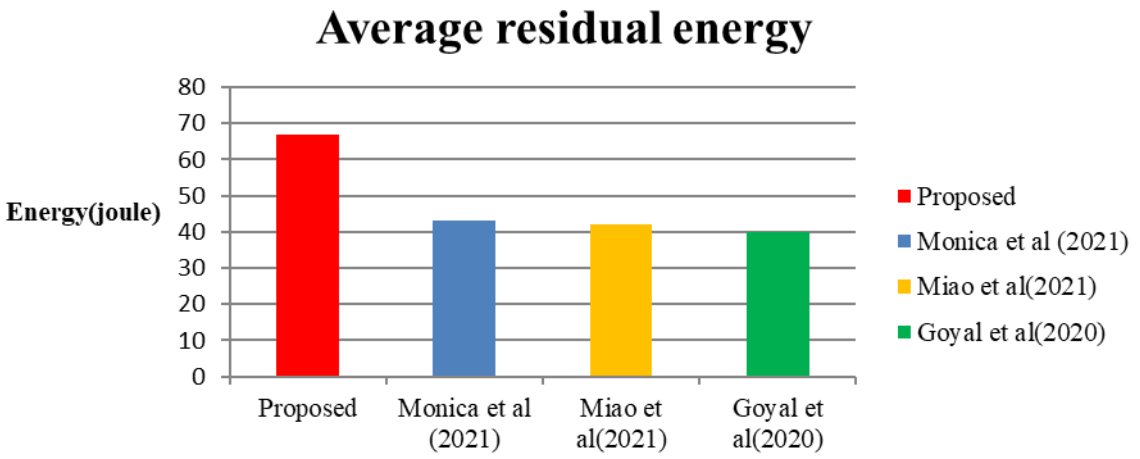

Figure. 7 Average residual energy

proposed solution compared to others. The survival probability is measured for different time periods and the result is given in Fig. 4. The results shows that survival probability increases in the proposed solution by 9% over the period of 6 minutes compared to 6% in all existing works. Piggy bagging-based distribution takes certain time but increases the survival probability to maximum value of 100%, but existing works were constrained in reaching the maximum value. Monica et al and Goyal et al did not consider data survival, so in case of failure of node which cached the data fails, the data is lost. Miao et al used replication for data survivability only in neighbourhood nodes. Due to this when a spot near to data node is damaged, the replicated data in that spot too is lost. But proposed solution replicated data at far off places, so there is a higher chance of data survivability as revealed in results.

The difficult level of confidentiality is measured in terms of variance of difference (VOD) between the original data and predicted data by compromising certain nodes. Let X_i be a random variable representing the data from sensor at time i, X_i' be the estimated result of X_i and difference $D_i = X' - X$. Let mean of D be $E(D_i)$ and variance be $Var(D_i)$. VOD for column i is $Var(D_i)$. VOD is measured for data over a period of 10 minutes and average VOD is given as privacy measure(pm)

$$pm = \frac{\sum_{i=1}^{N} VOD_i}{N} \tag{22}$$

A guess is launched for 10 minutes to predict sensor data and privacy measure (pm) is measured for every 1-minute interval and plotted in Fig. 5. The average privacy measure in proposed solution is 54% higher compared to Monica et al, 70% higher compared to Miao et al and 65% higher compared to Goyal et al. The higher value of pm metric in proposed solution

indicates that even if guessing attack is launched by compromising certain nodes, the difference between the actual and predicted data is very high in the proposed solution. This demonstrates higher data confidentiality in proposed solution. The pm is higher in proposed solution due to two reasons of transform coding and making the measurement matrix known only between sensor node and sink. Due to this, two level of security which becomes difficult to infer data in proposed solution. But existing works of Monica et al, Goyal et al relied only on authentication and when authentication is compromised, data is no longer secure. Miao et al used Shamir secret sharing mechanism. When the attacker is able to get minimal shares, the attacker can reconstruct the shares and know the original data.

The overall storage consumption is measured in terms of percentage of total storage utilized in sensor network for same volume of data generated by sensor nodes across the solutions at the end of simulation time and the result is given in Fig. 6.

The storage utilization is at least 20% lower in proposed solution compared to existing works. Use of combined transform coding with compressive sensing has reduced the storage utilization in the proposed solution. Reduced storage utilization signifies more coverage for survivability in the proposed solution. Monica et al, stored data as in generated form. Goyal et al used aggregation but it is less effective compared to compressive sensing for storage. Miao et al used Shamir based replication but its memory requirement is very high compared to compressive sensing format. Due to this, existing work's storage consumption is higher compared to proposed solution. The energy consumption is measured in terms of average residual energy of the nodes at the end of simulation for same volume of data generated. The result is given in Fig. 7. The average residual energy has dropped from initial energy only by 33% in proposed solution compared to about 60% in existing solutions. Thus, proposed solution has lower energy consumption for replication compared to existing works. This is due to combined transform coding and compression sensing along with piggy backing based data distribution. The existing work, Monica et al and Goyal et al used multi hop transmission to RP point or cluster head, due to which its energy consumption increased. Miao et al used Shamir secret share based data replication. Due to this number of shares for data has increased and it has also increased the energy consumption for replicating the shares to neighbouring nodes.

6 Conclusion

An energy aware adaptive compressive sensing scheme is proposed in this work. The solution is able to provide higher coverage of survivability with minimal energy consumption. Also, the data confidentiality is strong in proposed solution due to use of adaptive transformation matrix for compressive sensing. Piggy bagging based data distribution has provided higher data survivability in the proposed solution with minimal energy consumption costs. Overall, the proposed solution has provided at least 6% higher data survivability and 20% lower energy consumption compared to existing works. The storage consumption is 20% lower compared to existing works, thereby providing more survivability volume. The data survival probability is also 3% higher than existing works due to replication in far off nodes. With two level of security using transform coding and measurement matrix anonymity, data confidentiality in term of pm metric is at least 54% higher compared to existing works.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

Paper background work, conceptualization, methodology, implementation, result analysis and comparison, preparing and editing draft, visualization have been done by the first author. Supervision, review of work and monitoring the implementation have been done by the second author.

References

[1] R. DiPietro, V. Mancini, and C. Soriente, "Data Security in Unattended Wireless Sensor Networks", *IEEE Transactions on Computers*, Vol. 58, No. 11, pp. 1500-1511, Nov. 2009.

[2] S. Rani and H. Ahmed, "Multi-hop Routing in Wireless Sensor Networks: An Overview, Taxonomy, and Research Challenges", *Springer Briefs in Electrical and Computer Engineering*, 2015.

[3] D. Kandris, C. Nakas, D. Vomvas, and G. Koulouras, "Applications of Wireless Sensor Networks: An Up-to-Date Survey", *Appl. Syst. Innov.*, Vol. 3, No. 14, 2020.

[4] D. Pietro, V. Mancini, and C. Soriente, "Catch me (if you can): Data survival in unattended sensor networks", In: *Proc. of International Conference on Pervasive Computing and Communications*, pp. 185–194, 2008.

[5] X. Du and H. Chen, "Security in Wireless Sensor Networks. Wireless Communications", *IEEE.*

Wireless Communication, Vol. 15, No. 4, pp. 60-66, 2008.

[6] I. Orović, V. Papić, C. Ioana, and L. Xiumei, "Compressive Sensing in Signal Processing: Algorithms and Transform Domain Formulations", *Mathematical Problems in Engineering*, Vol. 2016.

[7] G. Aliberti, D. Pietro, and S. Guarino, "Epidemic data survivability in Unattended Wireless Sensor Networks", *J. Netw. Comput. Appl*., Vol. 99, 2017.

[8] .M. Bahi, C. Guyeux, M. Hakem, and A. Makhoul, "Epidemiological approach for data survivability in unattended wireless sensor networks", *J. Netw. Comput. Appl*., Vol. 46, pp. 374-383, 2014.

[9] S. Elsafrawey, S. Hassan, and M. Dessouky, "Cooperative hybrid self-healing scheme for secure and data reliability in unattended wireless sensor networks", *IET Inform. Secur*., Vol. 9, pp. 223-233, 2015.

[10] A. Sen, S. Ghosh, A. Basak, P. Puria, and S. Ruj, "Achieving data survivability and confidentiality in unattended wireless sensor networks", In: *Proc. of the 29th International Conference on Advanced Information Networking and Applications (AINA)*, pp. 239246, 2015.

[11] W. Cheng, Y. Li, Y. Xipeng, "Secure Data Distribution Scheme with Two-Hop Survival Strategy for Unattended WSNs", *International Journal of Distributed Sensor Networks*., Vol. 11, No. 10, pp. 1-9, 2015.

[12] W. Lim, K. Kapusta, G. Memmi, and S. Jung, "Multi-hop Data Fragmentation in Unattended Wireless Sensor Networks", *ArXiv, abs/1901.05831*, 2019

[13] B. Choi, B. Ko, and W. Lim, "Energy-Aware Distribution of Data Fragments in Unattended Wireless Sensor Networks", In: *Proc. of International Conference on Security of Smart Cities, Industrial Control System and Communications*, pp. 1-8,2018.

[14] S. Mateus, B. Margi, A. Simplício, P. Geovandro, and T. Oliveira, "Implementation of data survival in unattended Wireless Sensor Networks using cryptography", In: *Proc. of IEEE Local Computer Network Conference*, pp. 961-967,2010.

[15] W. Liang, Z. Ruan, Y. Wang, and X. Chen, "RESH: A Secure Authentication Algorithm Based on Regeneration Encoding Self-Healing Technology in WSN", *Journal of Sensors*, Vol. 2016, 2098680, 2016.

[16] J. Tang, A. Liu, M. Zhao, and T. Wang, "An Aggregate Signature Based Trust Routing for Data Gathering in Sensor Networks", *Secur. Commun. Netw*., Vol. 2018, 2018.

[17] G. Maia, D. Guidoni, A. Viana, A. Aquino, R. Mini and A. Loureiro, "A distributed data storage protocol for heterogeneous wireless sensor networks with mobile sinks", *Ad Hoc Netw*., Vol. 11, pp. 1588–1602, 2013.

[18] A. Talari and N. Rahnavard, "CStorage: Decentralized compressive data storage in wireless sensor networks", *AdHoc Netw*, Vol. 37, pp. 475–485, 2016.

[19] M. Albano and S. Chessa, "Replication vs. Erasure coding in data centric storage for wireless sensor networks", *Comput. Netw*., Vol. 77, pp. 42–55, 2015.

[20] A. Cueva, M. Urueñ, G. Veciana, R. Cuevas, N. Crespi, "Dynamic data-centric storage for long-term storage in wireless sensor and actuator networks", *Wirel. Netw*., Vol. 20, pp. 141–153, 2014.

[21] C. Zhang, O. Li, G. Liu, and M. Li, "A Practical Data-Gathering Algorithm for Lossy Wireless Sensor Networks Employing Distributed Data Storage and Compressive Sensing", *Sensors*, Vol. 18, No. 10, p. 3221, 2018.

[22] T. Nguyen, A. Teagu, and N. Rahnavard, "CCS: Energy-efficient data collection in clustered wireless sensor networks utilizing block-wise compressive sensing", *Comput. Netw*., Vol. 106, pp. 171–185, 2016.

[23] B. Gong, P. Cheng, Z. Chen, L. Ning, and D. Hoog, "Spatiotemporal compressive network coding for energy-efficient distributed data storage in wireless sensor networks", *IEEE Commun. Lett*., Vol. 19, pp. 803–806, 2015.

[24] A. Alrashed, F. Bagci, and E. Alquraishi, "A key management approach for forward and backward secrecy in unattended WSNs", *Computer Engineering*, Vol. 4, No.4, pp.24-45, 2016.

[25] W. Lei, Z. Meng, and J. Chen, "A novel privacy- and integrity-preserving approach for multidimensional data range queries in two-tiered wireless sensor networks", *International Journal of Distributed Sensor Networks*, Vol. 15, 2019.

[26] S. Stanković, "Compressive sensing: Theory, algorithms and applications", In: *Proc of 4th Mediterranean Conference on Embedded Computing (MECO)*, pp. 4-6, 2015.

[27] S. ArunSankar and S. Sathidevi, "A scalable speech coding scheme using compressive sensing and orthogonal mapping based vector quantization", *Heliyon*, Vol. 5, No. 5, 2019.

[28] C. Monika and N. Goyal, "A rendezvous point-based data gathering in underwater wireless sensor networks for monitoring applications", *International Journal of Communication*

Systems , p. e5078, 2022.
[29] N. Goyal, M. Dave and K. Verma, "SAPDA: Secure authentication with protected data aggregation scheme for improving QoS in scalable and survivable UWSNs", *Wireless Pers. Commun*., Vol. 113, pp. 1–15, 2020.
[30] C. Miao, Y. Fan, and H. Li, "Secure data collection method of WSN based on mobile Sink", *Chinese Journal of Network and Information Security*, Vol. 7, No. 1, pp. 121-129, 2021.