Optimizing Routing and Storage Node Deployment Together for Consistent Data Storage in Multicultural WSN

Dr. A. Avani

Assistant Professor, Department of CSE, Anu Bose Institute of Technology, Paloncha, India

avanialla@gmail.com

ABSTRACT:

Sensor networks have been widely used in the big data age, yielding a lot of data for several purposes. However, handling the enormous amount of data is a highly difficult problem since sensor networks are often installed in hostile locations. In this research, the topic of data storage reliability in heterogeneous wireless sensor networks, where robust storage nodes are placed in sensor networks and data redundancy is used via coding approaches, is the major emphasis. We create an algorithm to jointly optimize data routing and storage node placement in order to reduce the expenses associated with data delivery and data storage. Since the issue is NP-hard and may be expressed as a binary nonlinear combinatorial optimization problem, it is very difficult to construct approximation techniques. We carefully propose an effective algorithm powered by a continuous-time Markov chain to plan the deployment of the storage node and associated routing strategy by using the Markov approximation framework. We also do in-depth simulations to confirm the effectiveness of our method.

I INTRODUCTION:

Wireless Sensor Networks (WSNs) have rapidly advanced in a variety of applications during the last several decades [1]. A variety of sensors have been widely installed and networked for monitoring and surveying purposes in the big data era[2–5]. Data storage has still grown to be a highly challenging task as a consequence of the development of WSNs, which generates enormous sensory data [6-11]. Difficult problem. Due to their limited resources (such as computation, data processing, and storage capabilities), sensor nodes are often unable to provide professional data storage services. Employing designated storage nodes in a heterogeneous wireless sensor network is one way to increase the capacity of data storage in a WSN. The data detected

by the regular sensor nodes is transmitted to the storage node via multi-hop transmissions.

Nevertheless, fault tolerance should be one of the primary considerations since the sensor nodes may be placed in hostile environments (such as conflict zones and earthquake situations). In particular, all data stored in a storage node is corrupted when it is destroyed. Adding data redundancy is one common solution to this problem. For instance, depending on erasure codes may significantly increase the dependability of data storage[12, 13]. However, the induced data redundancy leads to more data traffic, which energy-constrained sensor networks[14] may find prohibitive. Additionally, the storage nodes incur a significant cost in handling the growing amount of data.

The price of storage should also be considered.

In this paper, we jointly optimize data routing and storage node placement to create an algorithm that reduces the weighted total of the costs associated with data transport and storage. In specifically, we intricately develop a Continuous-Time Markov Chain using the Markov approximation framework [8] (CTMC). By using the technique, we can update the data routing scheme and deploy storage nodes adaptively, enabling us to reduce the induced cost and guarantee the integrity of the data storage. Theoretical analysis and in-depth simulations may be used to demonstrate the effectiveness of our technique.

II PROBLEM FORMULATION & RELATED WORK:

We suppose that j represents the price of storing one unit of data at location sj. To be more precise, the data storage c We suppose that j represents the price of storing one unit of data at location sj. In particular, the cost of data storage comprises not only the energy used for data transfer (such as receiving data from sensor nodes and responding to distant data requests), but also the energy used for keeping and managing the data that has been received.

To communicate with one another and respond to distant data requests, we assume the storage nodes have long-range communication modules. If re-deployment is necessary for adjusting to dynamic network circumstances, we additionally assume that the storage nodes are outfitted with mobile mechanisms [17, 18]. It also takes into account the energy used for handling and keeping the data that has been

received, such as when receiving data from sensor nodes and responding to distant data requests.

To communicate with one another and respond to distant data requests, we assume the storage nodes have long-range communication modules. If re-deployment is necessary for adjusting to dynamic network circumstances, we additionally assume that the storage nodes are outfitted with mobile mechanisms [17, 18].

Replication and coding are two methods that may be used to improve the data storage system's dependability. The first method involves making several copies of the original data and putting them in various storage nodes for each copy. This method is used to duplicate the data. In the latter method, we start by encoding the original data using Maximum Distance Separable (MDS) codes, and after that, we store each individual piece of encoded data in its own individual storage node. Data redundancy serves as the guarantee of dependability in each of these approaches.

For example, accessing just some of the storage nodes in order to partly access the duplicated or encoded data is sufficient to recover the original data. This may be done by following the steps outlined in the previous sentence.

Let's use several coding methods as an illustration, shall we? Let's assume that I represents the pace at which sensor node ג i generates data. To be more specific, during each sensing-storage interval (we will make the assumption that the data storage is carried out on a periodic basis in each individual node). During each period, the data that have been accumulated during the sensing phase are encoded and transferred to the storage nodes.), the total quantity of data that has been perceived by ג i is denoted by the number i.

These original data are capable of being encoded using redundant coding methods, such as MDS codes parameterized byג i/.

Particularly, node ג i encodes the i original data into i. energy on data delivery. More specifically, it is necessary for node µi to send each of its I copies to the given address (for every original data). separate storage nodes.

III SYSTEM EVALUATION:

The framework we use for optimization is easily compatible with. combines replication and coding-based systems via the use of scaling the expenses associated with the

transport and storage of the data; nonetheless, we In this particular research, you should solely concentrate on the second option.

Let the value of the binary variable $X_{i,k} \in \{0,1\}$ 1g indicate whether or not the node $n_i \in N$ transmits one of its encoded data to the storage node that is located at location uk (if any). Additionally, we hypothesise that the value of $y_{i,k} \in \{0,1\}$ 2 f0; 1g indicates whether or not storage node sj is installed at site uk. When we start with the coding scheme$\{(\mu_i,\lambda_i)\}_i$=1,….N, we can then frame our optimization issue of Joint routing and Storage node deployment (JUST) as follows:

$$\min \sum_{i=1}^{|\mathcal{N}|}\sum_{k=1}^{|\mathcal{U}|} c_{i,k}x_{i,k} + \beta \sum_{i=1}^{|\mathcal{N}|}\sum_{k=1}^{|\mathcal{U}|}\sum_{j=1}^{|\mathcal{S}|} \tau_j x_{i,k} y_{j,k} \quad (1)$$

$$\text{s.t.} \sum_{k=1}^{|\mathcal{U}|} x_{i,k} = \mu_i,\ \forall n_i \in \mathcal{N} \quad (2)$$

$$\sum_{j=1}^{|\mathcal{S}|} y_{j,k} - x_{i,k} \geqslant 0,\ \forall n_i \in \mathcal{N}, u_k \in \mathcal{U} \quad (3)$$

$$\sum_{j=1}^{|\mathcal{S}|} y_{j,k} \leqslant 1,\ \forall u_k \in \mathcal{U} \quad (4)$$

$$\sum_{k=1}^{|\mathcal{U}|} y_{j,k} \leqslant 1,\ \forall s_j \in \mathcal{S} \quad (5)$$

$$x_{i,k} \in \{0,1\},\ \forall n_i \in \mathcal{N}, u_k \in \mathcal{U} \quad (6)$$

$$y_{j,k} \in \{0,1\},\ \forall s_j \in \mathcal{S}, u_k \in \mathcal{U} \quad (7)$$

The goal of the quadratic formula Our mission is accomplished by Function (1), which is to reduce, as much as possible, the total amount of money spent on data transmission and storage. In order to strike a balance between the two, we include the parameter into Function (1). The primary limitations are broken down into the following categories:

Each sensor node is unable to transmit its data to a location that is "empty" since there is no storage node present.

We design a CTMC with the function fF.t /gt>0 to drive the adaptive deployment of the storage nodes and the corresponding routing scheme in order to solve the JUST-Approx problem (and by extension, to solve the JUST trouble with averaging optimality). In this function, t stands for time, and F.t / 2 F refers to the storage node deployment that was adopted at time t. This allows us to resolve the JUST-Approx The CTMC experiences state changes, also known as storage node reinforcements, on a continuous basis at any given moment in time, and these transitions are defined by their transition rates. We make the assumption that any two of the states may communicate with one another over a transition link if, and only if, both of the states only have one storage node located in each of their respective sites. It is important

to note that the reachability between the states and, as a result, the irreducibility of the markov chain are not affected by this assumption. The relevance of the following two folds to an explanation of the optimization problems of the CTMC-driven approach may be seen in light of the claims made above: It is more likely to migrate to the lower-cost storage node deployment in each state transition. On the one hand, we choose the storage node distributions that have lower costs for a longer period; on the contrary hand, it is more likely to move to the lower installation number of nodes.

In order to implement the CTMC in a way that is distributed, we suggest a method. The storage nodes are first deployed in a manner that is completely arbitrary by default. A sensor network's lifespan may be broken down into a series of epochs, and the beginning of a new epoch occurs whenever some storage nodes are moved to a different location. At the beginning of each epoch, we first calculate a routing strategy in accordance with the current deployment of storage nodes f by the i-NN policy.

IV EXPRIMENTAL REWVIES:

The difficulty of the method described above is mostly due to based on the computation of cf C, often known as calculating the route approach and the costs associated with it. The quickest routes available from the sensor nodes to the locations and the produced electromagnetic fields The cost may be determined using cutting-edge polynomial-time algorithms (like Dijkstra's technique, which has a time complexity of one). the inherent difficulty of O.jN.j /). Due to the fact that the sensor nodes and The locations of the fixed sites where we install the storage nodes are:

It is possible to preload data delivery channels and their associated costs. inside the storage nodes, particularly considering that the storage nodes have typically enough room for storage in their memory. If a storage facility As a node is redeployed, it communicates to the other nodes its new location. notice of position sent out via the O.log jUj/-length broadcast messages, in order for all of the other storage nodes to be able to locally calculate the equivalent of C. In addition to that, the storage nodes may transmit the alerts to the sensor nodes by means of riding on the coattails of the newly acquired storage node deployment when it comes to the acknowledgments of the data deliveries. In each case, When the notification has been received from the sensor node, the sensor nodes, the storage nodes allow the sensor node to immediately

calculate the cost of transferring a data unit to a previously decommissioned storage location node, and then modify the routing strategy to reflect those changes.

the i-NN guiding principle. According to the configuration described above, each of the storage nodes just needs to put up a timer in the local area. One question is, Is it possible to accomplish CTMC with such a distributed implementation? fF.t/gt>0 by adhering to Propositions 1 and 2, respectively.

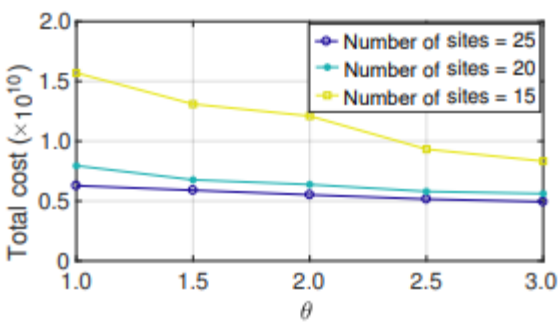

Fig. 1 Total cost of our algorithm under different values of θ

The data collecting has been the subject of investigation by a large number of projects. issue by using a single or several sink nodes [25–28]. Although We steal the concept of enhancing the sinks with mobility, these techniques for collecting data cannot be implemented. Directly applicable to the issue at hand as neither storage nor neither the cost of data storage nor its dependability is taken into account. As combining the two of them forms an The optimization framework is not at all straightforward.

Several of the currently active initiatives have the aim of distributing the data that was detected across all of the sensor nodes for storing. As an example, and in accordance with what is shown in Reference [29], The information collected by sensors organised in a tree-like arrangement information from sensor nodes farther downstream may be kept in a subset of non-root ones to be considered. In Refs. [30, 31], condense Integration was made between sensing and probabilistic broadcasting. to reduce the amount of energy used and the amount of data traffic. In Refs. quorum-based data storage techniques were discussed in [32,33]. suggested, with the reliability of the storage being guaranteed by data replication.

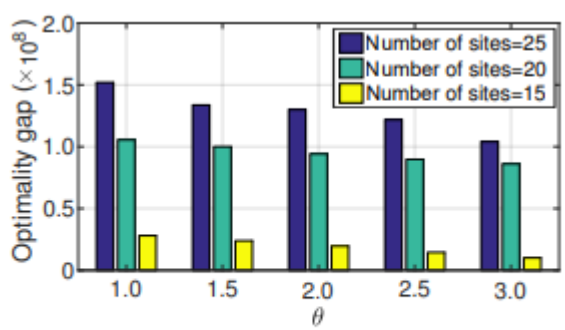

Fig. 2 Optimality gaps of our algorithm under different values of θ.

V CONCLUSION:

In this study, we present a CTMC-based method that makes use of the Markov approximation framework. Our goal is to reduce the costs associated with the transport and storage of data while maintaining a high level of storage dependability. In particular, the deployment of the storage nodes is adaptively planned, and the routing scheme is appropriately updated, all of which is driven by an intricately constructed CTMC. In order to test our approach, we first do theoretical research, and then we run numerical simulations.

We are going to take into consideration dynamic sensor networks, which are networks in which sensor nodes may join or leave the network at any time. In this particular scenario, our goal is to create an online algorithm that is capable of adaptively using the churn of the sensor nodes in order to optimise data routing and the deployment of storage nodes.

REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, Wireless sensor networks: A survey, Computer Networks, vol. 38, no. 4, pp. 393–422, 2002.

[2] C. Y. Ai, F. H. Li, and K. J. Zhang, Detecting isolate safe areas in wireless sensor monitoring systems, Tsinghua Science and Technology, vol. 22, no. 4, pp. 427–436, 2017.

[3] R. Arcucci, C. Pain, and Y. K. Guo, Effective variational data assimilation in air-pollution prediction, Big Data Mining and Analytics, vol. 1, no. 4, pp. 297–307, 2018.

[4] F. Li, J. Luo, S. Q. Xin, and Y. He, Autonomous deployment of wireless sensor networks for optimal coverage with directional sensing model, Computer Networks, vol. 108, pp. 120–132, 2016.

[5] F. Li, J. Luo, W. P. Wang, and Y. He, Autonomous deployment for load balancing k-surface coverage in sensor networks, IEEE Transactions on Wireless Communications, vol. 14, no. 1, pp. 279–293, 2015.

[6] L. Xiang, J. Luo, C. W. Deng, A. V. Vasilakos, and W. S. Lin, DECA: Recovering fields of physical quantities from incomplete sensory data, in Proc. 9th Annu. IEEE Communications Society Conf. Sensor, Mesh and Ad Hoc Communications and Networks, Seoul, South Korea, 2012, pp. 182–190.

[7] S. Y. Cheng, Z. P. Cai, J. Z. Li, and H. Gao, Extracting kernel dataset from big sensory data in wireless sensor networks, IEEE Transactions on Knowledge and Data Engineering, vol. 29, no. 4, pp. 813–827, 2017.

[8] J. Li, S. Y. Cheng, Z. P. Cai, J. G. Yu, C. K. Wang, and D. M. Li, Approximate holistic aggregation in wireless sensor networks, ACM Transactions on Sensor Networks, vol. 13, no. 2, pp. 1–24, 2017.

[9] Z. B. He, Z. P. Cai, S. Y. Cheng, and X. M. Wang, Approximate aggregation for tracking quantiles and range countings in wireless sensor networks, Theoretical Computer Science, vol. 607, pp. 381–390, 2015.

[10] S. Y. Cheng, Z. P. Cai, and J. Z. Li, Curve query processing in wireless sensor networks, IEEE Transactions on Vehicular Technology, vol. 64, no. 11, pp. 5198–5209, 2015.

[11] Q. Y. Meng, K. Wang, X. M. He, and M. Y. Guo, QoEdriven big data management in pervasive edge computing environment, Big Data Mining and Analytics, vol. 1, no. 3, pp. 222–233, 2018.

[12] M. Albano and S. Chessa, Distributed erasure coding in data centric storage for wireless sensor networks, in Proc. 2009 IEEE Symp. Computers and Communications, Sousse, Tunisia, 2009, pp. 22–27.

[13] A. G. Dimakis, V. Prabhakaran, and K. Ramchandran, Decentralized erasure codes for distributed networked storage, IEEE Transactions on Information Theory, vol. 52, no. 6, pp. 2809–2816, 2006.

[14] F. Li, Y. B. Yang, Z. C. Chi, L. Y. Zhao, Y. W. Yang, and J. Luo, Trinity: Enabling self-sustaining WSNs indoors with energy-free sensing and networking, ACM Transactions on Embedded Computing Systems, vol. 17, no. 2, pp. 1–27, 2018.

[15] S. Holzer and R. Wattenhofer, Optimal distributed all pairs shortest paths and applications, in Proc. 2012 ACM Symp. Principles of Distributed Computing, Madeira, Portugal, 2012, pp. 355–364.

[16] M. Ghaffari and J. Li, Improved distributed algorithms for exact shortest paths, in Proc. 50th Annu. ACM SIGACT Symp. Theory of Computing, Los Angeles, CA, USA, 2018, pp. 431–444.

[17] G. Maia, D. L. Guidoni, A. C. Viana, A. L. L. Aquino, R. A. F. Mini, and A. A. F. Loureiro, A distributed data storage protocol for heterogeneous wireless sensor networks with mobile sinks, Ad Hoc Networks, vol. 11, no. 5, pp. 1588– 1602, 2013.

[18] Q. Liu, K. Zhang, X. D. Liu, and N. Linge, Grid routing: An energy-efficient routing protocol for WSNs with single mobile sink, International Journal of Sensor Networks, vol. 25, no. 2, pp. 93–103, 2017.

[19] M. H. Chen, S. C. Liew, Z. Y. Shao, and C. H. Kai, Markov approximation for combinatorial network optimization, IEEE Transactions on Information Theory, vol. 59, no. 10, pp. 6301–6327, 2013.

[20] X. Zheng, Z. P. Cai, J. Z. Li, and H. Gao, A study on application-aware scheduling in wireless networks, IEEE Transactions on Mobile Computing, vol. 16, no. 7, pp. 1787– 1801, 2017.

[21] Z. P. Cai, G. H. Lin, and G. L. Xue, Improved approximation algorithms for the capacitated multicast routing problem, in Proc. of the 11th COCOON, Berlin, Germany, 2005, pp. 136–145.

[22] Z. P. Cai, Z. Z. Chen, and G. H. Lin, A 3.4713- approximation algorithm for the capacitated multicast tree routing problem, Theoretical Computer Science, vol. 410, no. 52, pp. 5415–5424, 2009.

[23] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge, UK: Cambridge University Press, 2004.

[24] R. G. Gallager, Stochastic Processes: Theory for Applications. Cambridge, UK: Cambridge University Press, 2014.

[25] R. L. Deng, S. B. He, and J. M. Chen, An online algorithm for data collection by multiple sinks in wirelesssensor networks, IEEE Transactions on Control of Network Systems, vol. 5, no. 1, pp. 93–104, 2018.

[26] J. Wang, J. Y. Cao, R. S. Sherratt, and J. H. Park, An improved ant colony optimization-based approach with mobile sink for wireless sensor networks, The Journal of Supercomputing, vol. 74, no. 12, pp. 6633–6645, 2018.

[27] G. R. Li, S. C. Peng, C. Wang, J. W. Niu, and Y. Yuan, An energy-efficient data collection scheme using denoising autoencoder in wireless sensor networks, Tsinghua Science and Technology, vol. 24, no. 1, pp. 86–96, 2019.

[28] F. Li, J. Luo, G. T. Shi, and Y. He, ART: Adaptive frequency-temporal co-existing of ZigBee and WiFi, IEEE Transactions on Mobile Computing, vol. 16, no. 3, pp. 662– 674, 2017.

[29] B. Sheng, Q. Li, and W. Z. Mao, Data storage placement in sensor networks, in Proc. 7th ACM Int. Symp. Mobile Ad Hoc Networking and Computing, Florence, Italy, 2006, pp. 344–355.

[30] X. J. Yang, X. F. Tao, E. Dutkiewicz, X. J. Huang, Y. J. Guo, and Q. M. Cui, Energy-efficient distributed data storage for wireless sensor networks based on compressed sensing and network coding, IEEE Transactions on Wireless Communications, vol. 12, no. 10, pp. 5087–5099, 2013.

[31] A. Talari and N. Rahnavard, CStorage: Decentralized compressive data storage in wireless sensor networks, Ad Hoc Networks, vol. 37, pp. 475–485, 2016.

[32] J. Luo, F. Li, and Y. He, 3DQS: Distributed data access in 3D wireless sensor networks, in Proc. 2011 IEEE Int. Conf. Communications, Kyoto, Japan, 2011, pp. 1–5.

[33] C. Zhang, J. Luo, L. Xiang, F. Li, J. C. Lin, and Y. He, Harmonic quorum systems: Data management in 2D/3D wireless sensor networks with holes, in Proc. 2012 9th Annu. IEEE Communications Society Conf. Sensor, Mesh and Ad Hoc Communications and Networks, Seoul, South Korea, 2012, pp. 1–9.

[34] R. F. Zeng, Y. X. Jiang, C. Lin, Y. F. Fan, and X. M. Shen, A distributed fault/intrusion-tolerant sensor data storage scheme based on network coding and homomorphic fingerprinting, IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 10, pp. 1819– 1830, 2012.

[35] J. Tian, T. Yan, and G. L. Wang, A network coding based energy efficient data backup in survivability-heterogeneous sensor networks, IEEE Transactions on Mobile Computing, vol. 14, no. 10, pp. 1992–2006, 2015

Author Details

Dr. A. Avani obtained M. Tech degree from JNTU, Hyderabad, India. She is at present working as professor in Department of CSE of Anu Bose Institute of Technology, Paloncha, Telangana, India. Her area of interest is Data mining.