Secure Data Aggregation Scheme in Wireless Sensor Network

Prof. Kalyani Pendke\textsuperscript{2}, Anshula Dupare\textsuperscript{1}, Ashwini Khadatkar\textsuperscript{1}, Neha Pawar\textsuperscript{1}, Nikita Wadhwani\textsuperscript{1}, Rasika Salodkar\textsuperscript{2}

\textsuperscript{1}BE Students, Department of Computer science and Engineering Rajiv Gandhi College of Engineering & Research, Nagpur, Maharashtra, India

\textsuperscript{2}Assistant Professor, Department of Computer science and Engineering Rajiv Gandhi College of Engineering & Research, Nagpur, Maharashtra, India

ABSTRACT

In wireless sensor networks, data aggregation assumes an essential part in diminishing vitality utilization. As of late, explore has concentrated on secure data aggregation because of the open and unfriendly condition conveyed. The Homomorphic Encryption (HE) conspire is widely used to secure data classification. Be that as it may, HE-based data aggregation plans have the accompanying disadvantages: flexibility, unapproved aggregation, and constrained aggregation capacities. To take care of these issues, we propose a secure data aggregation plot by consolidating homomorphic encryption innovation with a mark conspire. To answer this issue we presented a system speaks to a strategy in that powerful cluster head is picked based on the separation from the base station and remaining vitality. Subsequent to choosing the cluster head, it influences utilization of minor measure of vitality of sensor to network and in addition enhances the lifetime of the network of sensor network. Aggregation of the data got from the cluster individuals is obligation of cluster head in the cluster. Confirmation of data is finished by the cluster head preceding the data aggregation if data got isn’t legitimate at that point got data is disposed of. Just confirmed data is taken for aggregation at cluster head. Encryption is finished by making utilization of homomorphic encryption technique and additionally encoded data send to the cluster head and data decoding is performed by base station (BS) for offering end to end security. An ID based mark system is created for hop by hop authentication. In this paper, we show the technique for recuperating the data which is lost because of the cushion flood. In given system cache memory is given by the cluster head to recuperation of data misfortune. Finally test comes about shows relying upon parameter like time and additionally vitality utilization on Jung test system that system exhibited is great contrasted with the accessible system.

Keywords: Sensor Nodes, Cluster Head, Base Station, Wireless Sensor Networks, Cache Based System, Hop by hop authentication.

I. INTRODUCTION

Wireless sensor networks (WSNs) have been generally sent in numerous applications, for example, ecological screens, social insurance, natural life observation, and mishap reports \cite{3,4}. WSNs, which are as of now thought to be one of the fundamental parts of the Internet of Things comprise of various sensor hubs obliged regarding their storage room, battery control, and computational ability\cite{2}. Along these lines, arrangements intended to drag out the lifetime of the network are broadly looked for.

Data aggregation is known as one of the strategies that are useful to limit the vitality utilization of sensors \cite{1}. With such system, data detected by
different part hubs are totaled into a solitary one by applying some aggregation capacities, for example, Sum, Average, and MAX lastly transmitted to the base station by means of the wireless connection. Subsequently, data aggregation is useful to decrease parcel transmissions and excess. For instance, in an antiquated woodland, sensors are conveyed to report their detected temperature to the base station for flame observing. For this situation, the base station may require the greatest estimation of all the detecting data to trigger alerts. Thusly, each cluster head just needs to choose the most extreme incentive from among various data esteems got from its part hubs and afterward send the outcome to the base station.

Plainly, the correspondence overhead is reduced in light of the fact that exclusive the accumulated outcome is transmitted to the base station. Along these lines, data aggregation is advantageous to drag out the general lifetime of the sensor networks. However, in light of the fact that they are regularly sent in antagonistic and unattended situations, WSNs are presented to different assaults, for example, replay assault, infusion assault, and hardening assault. The asset compelled qualities of WSNs make existing copious security calculations unsatisfactory for WSNs. In this manner, guaranteeing security for data aggregation is a test.

Advances in wireless correspondence made it conceivable to create wireless sensor networks (WSN) comprising of little gadgets, which gather data by collaborating with each other. These little detecting gadgets are called hubs and comprise of CPU (for data handling), memory (for data stockpiling), battery (for vitality) and handset (for accepting and sending signs or data starting with one hub then onto the next). The span of every sensor hub shifts with applications. For instance, in some military or observation applications it may be minutely little. Its cost relies upon its parameters like memory estimate, handling rate and battery. WSNs are customarily executed regions, for example, open or normally un-trusted and even threatening situations that provoke diverse security issues. These join the strategies, similar to key organization, security, get to control, authentication and DoS protection and so forth.

There are a few issues in the sensor network like altering or empowering the hub batteries in light of thick and specially appointed operation in basic condition and also because of in secret nature of WSNs. There would one say one is imperative inquiry emerges that is how to expand the lifetime of the sensor networks? Despite the fact that it gets extremely basic like expanding network lifetime by lessening vitality utilization of hub in WSNs. Test outcomes shows that the exchange of data is particularly exorbitant based on vitality utilization (EC) however on the opposite side data preparing use low vitality. Moreover, a down to earth strategy expected to expand the lifetime of WSN additionally to limit the sensor vitality usage while data exchange. There is one more issue of security of data at the season of sending data from source to goal in WSN.

Sensor hubs with compelled assets are subject to number of assaults; in this way the data encryption is indispensable in WSNs. If data is transmitted without encryption then the assailants will separate the data and fuses false data in the system. In hop-by-hop scrambled data aggregation (EDAs), which is a middle person aggregator having keys of all as for sensor hubs decodes got encoded values, complete all the unscrambled esteems and scrambles the result for sending to a base station (BS). This method needs that middle person aggregators store keys for unscrambling in that a got aggregator would reveal these characterized data.

In this paper, fundamentally concentrate on the three inconveniences which is generally address in the wireless sensor networks. At first enhancing the system lifetime of the sensor system through limiting the vitality use in the system. Second is to give the security while the data transmission from sender to
beneficiary hub or from sender to base station. Third is data misfortune recuperation, when sending the data to cluster head data is lost on account of limit restrain requirement of cluster head. For updating the structure lifetime displayed the methodology in which cluster head is singled out the introduction of vitality, number of neighbors and division to the base station. By picking the cluster head through choosing these three parameters diminishes the imperativeness regard anticipated that would the sensor hub. Homomorphic encryption is utilized for giving the security to the data. Data is sent in the encoded route to the base station, base station unscramble the data resulting to tolerating the data. In like way the strategy of data aggregation is refined in which cluster head total the data which is gathered by the cluster hubs. For data misfortune recuperation we are give cache memory at cluster head. Finally, the outcome is separated for the system lifetime, vitality utilization and for past and proposed structure.

II. LITERATURE SURVEY

This section depicts the different works achieved by the scientists for the data aggregation, improving network lifetime of the sensor hubs.

Kyung-Ah Shim proposed a SDA methodology, SenSDA, which depends on upon the gathering of sensible cryptographic natives in heterogeneous cluster WSNs. To diminish the aggregate length of figure messages and to fulfill end-to-end affirmation, they expect an extra substance HE technique, so only a BS can decode encoded data amassed by the CHs got from part hubs for each party of cluster. To give hop-by-hop confirmation, they use a sorting out free personality based mark (IBS) system, thusly the BS and the CHs can watch the validity of all the transmitted scrambled data. To improve adequacy of various marks confirmations, they require a stamp strategy in which distinctive imprints from different endorsers on different messages can be checked quickly [1].

D. Boneh and M. Franklin propose a completely down to earth personality based encryption strategy. This methodology has figure content security in the subjective prophet show getting a variety of the computational Diffie-Hellman issue. This structure relies upon bilinear maps between clusters. The Weil union on elliptic bend is an instance of such an associate. They give a correct definition to secure personality based encryption orchestrates and give a couple of employments to such structures [5].

C. Castelluccia, E. Mykletun, and G. Tsudik focus on profitable, data transmission chatting security in WSNs. More particularly, they unite unassuming scrambled methods with fundamental aggregation systems to perform altogether useful enormously profitable of encoded data. To survey the sensibility of proposed systems, they assess them also, show to an uncommon ensuring comes about which clearly demonstrate quantifiable data transmission limit protection and insignificant overhead start from both scrambled and aggregation operations [6].

C.- M. Chen, Y.- H.Lin, Y.- C.Lin, and H.- M. Sun [7] show a thought called as Recoverable Concealed Data Aggregation (RCDA). In RCDA, a base station can recover each recognizing data made by all sensors paying little personality to the probability that these data have been totaled by cluster heads or aggregators. With this individual data, two functionalities are given. To start with, the base station can affirm the uprightness and validity of all recognizing data. Next, the base station can play out any aggregation restricts on them. By at that point, they propose two RCDA systems named RCDA-HOMO and RCDA-HETE for homogeneous and heterogeneous WSN self-rulingly. They demonstrate that the proposed procedures are secure under these strike models in the security examination.

J. Domingo-Ferrer addresses one such PH which can be shown secure against known-clear content
ambushes; the length of the figure content space is substantially higher than the reasonable content space. A couple of uses to task of tricky dealing with and data and to e-betting are immediately spoken to[8].

J. Girao, D. Westhoff, and M. Schneider display a strategy that 1) covers recognized data end-to-end by 2) starting at as of late giving valuable and adaptable in-network system data gathering. The accumulating mediatory center points are not basic to work at the perceived plaintext data. They execute a particular class of encoded scrambled and discuss structures for selecting the total limits "normal" and "development discovery." They demonstrate that the procedure is plausible for the class of "going down" steering conventions. They consider the danger of spoiled sensor centers by proposing a key pre-scattering calculation that restrains an aggressors advancement and show up how key pre-appropriation and a key-ID touchy "going down" directing convention builds up the quality and dependability nature of the related spine[9].

E. Mykletun, J. Girao, and D. Westhoff reevaluate the congruity of additively homomorphic open key encryption means certain classes of wireless sensor networks. Finally, they offer proposition to picking the most sensible open key strategies for different topologies and wireless sensor network conditions[10].

A. Shamir introduce another sort of cryptographic game plan, which engages any match of customers to pass on securely and to check each other's etchings without exchanging private or open keys, without keeping key records and without using the relationship of an untouchable. The game-plan perceives the closeness of trusted key age, whose sole setup is to give each customer an adjusted brilliant card when he first joins the structure. The data introduced in this card enables the customer to sign and encode the messages he sends and to interpret and check the messages he gets in an absolutely free way, despite of the identity of the other party. Starting at now issued cards should not to be overhauled when new customers join the structure, and the grouped concentrations don't have to empower their exercises or even to keep a customer list. The concentrations can be closed after every one of the cards are issued, and the system can keep working in a completely decentralized manner for a questionable period[11].

S. Lindsey and C.S. Raghavendra proposed Power-proficient collecting in sensor data systems (PEGASIS), which is a change over the LEACH. It is chain based convention, in which hubs need to talk with their closest neighbors and exchange visiting with BS. Each inside in the structure uses flag quality to discover the closest Neighbor. The chain in PEGASIS consolidates hubs closest to each other that shape a course to the BS. The gathered sort of the data will be sent to the BS by any inside point in the chain and the hubs in the tight spot will exchange sending to the BS. This lessens the power required to transmit data per round in light of the way that the power exhausting is spread reliably finished all hubs. In any case, the suppositions in PEGASIS may not by and large be sensible.

- PEGASIS expect that each sensor focus point can speak with the BS coordinate. In valuable cases, sensor hubs utilize multi-hop correspondence to complete the BS.
- It considers that all hubs keep up an entire database about the domain of each other focus point in the structure; however the system by which the inside point zone are gotten isn't delineated.
- It considers that all sensor hubs have a similar level of vitality and are apparently going to pass on in the interim.

Despite the way that everything considered sensors will be settled or stationary as recognized in PEGASIS, a few sensors may be permitted to move and along these lines influence as far as possible.
A. Manjeshwar and D.P. Agrawal proposed a dynamic clustering based custom passed on for responsive structures in which hubs react in a flicker to sudden and uncommon changes in condition known as TEEN. Cluster strategy and data exchange are done as in beyond what many would consider possible respects close by various qualities - Hard Threshold (HT) and Soft Threshold (ST). These qualities and in like manner the earth are perceived by the hubs interminably. Exactly when the middle point finds that the apparent trademark has accomplished HT, the inside point switches on its transmitter and sends the distinguished data.

The recognized respect is secured in an inward figure SV the middle. In the present cluster time distribution, the middle will next transmit data definitely when the present estimation of the distinguished property is higher than HT and the present estimation of the perceived trademark fluctuates from SV by an aggregate equal to or higher than the ST. The usage of HT and ST will decrease the measure of transmissions in the network and in this way it lessens the general vitality spread in the network. This course of action is suited for time fundamental data perceiving applications.

All hubs in APTEEN sense nature reliably, yet the data transmission happens absolutely when perceived data quality is at or more unmistakable than HT. For an inside point, if a data transmission does not happen in day and age comparing to the number time, it must perceive and transmit the data yet again. In APTEEN, each CH aggregates the data from the part hubs inside its cluster and transmits the amassed data to the BS. The convention perceives that the data got from part hubs are sufficiently related; consequently it reduces a gigantic measure of abundance of the data to be sent to the BS. Moreover, a balanced TDMA setup comprehends the cream system by doling out transmission space to all hubs in a cluster. In addition, APTEEN offers a huge amount of adaptability by enabling the clients to set the CT among time and the edge respects for vitality utilize can be controlled by changing the CT and what's undeniably the most remote point respects.

O. Younis and S. Fahmy proposed an another detectable vitality gifted focus clustering estimation is the Hybrid, Energy - Efficient and Distributed (HEED) clustering approach for extemporized sensor networks. Regard made by is a circulated clustering convention which was proposed with four basic destinations as takes after:

- Extending network future by scrambling vitality usage,
- Completing the clustering structure inside a dependable number of cycles,
- Lessening control overhead (to be prompt in the measure of hubs),
- Creating amazing - dispersed cluster heads and reduced clusters.

Notice now and again picks cluster heads in context of a cross kind of two clustering parameters: The significant parameter is whatever remains of the vitality of each sensor focus point and the optional parameter is the intra-cluster correspondence cost as a piece of Neighbor region or cluster thickness. The
principal parameter is used to probabilistically pick a concealed course of action of cluster heads while the assistant parameter is utilized for breaking ties.

The social event approach at each sensor focus requires a few rounds. Each round is sufficient long to get messages from any Neighbor inside the cluster expand. As in LEACH, a covered rate of cluster heads in the system, Cprob, is predefined. The parameter Cprob is starting late used to constrain the major cluster head disclosures and has no speedy impact on the last cluster structure. In HEED, every sensor focus sets the likelihood CHprob of changing into a cluster head as takes after Where Eresidual is the surveyed current holding up vitality in this sensor focus point and Emax is the most over the top vitality (relating to a completely charged battery), which is normally misty for homogeneous sensor hubs. The CHprob respect must be more noteworthy than a base edge p min. A cluster head is either a transitory cluster - head, if its CHprob is < 1, or a last cluster - head, if its CHprob has achieved 1.

In the midst of each round of HEED, every sensor focus point that never got see from a cluster head lifts itself to twist up surely a cluster head with likelihood CHprob. The beginning late picked cluster heads are added to the present game-plan of cluster heads. In the event that a sensor focus is bent up a social event head, it imparts a disclosure message as an unexpected cluster - head or a last cluster head. A sensor focus point tuning in to the cluster - head list picks the cluster head with the most immaterial cost from this strategy of cluster heads. Each middle by then duplicates its CHprob and goes to the subsequent Step. In the event that an inside point completes the HEED execution without lifting itself to twist up clearly a cluster head or joining a cluster, it verbalizes itself as a last cluster-head. A brief cluster - head focus point can change into a general focus point at a later cycle in the event that it gets see from a lower cost cluster head. Note that an inside can be picked as a cluster head at consecutive clustering between times on the

off chance that it has higher waiting vitality with bring down cost. Since a WSN is acknowledged to be a stationary network, where fixate call attention to pass on all of a sudden, the Neighbor set of each middle point does not change once in a while.

Here HEED does not need to do Neighbor disclosure occasionally. The spread of vitality usage of HEED enlarges the lifetime of the great number of hubs in the network. Hubs also in this manner restore their Neighbor sets in multi-bob networks by intermittently sending and enduring messages. The HEED clustering improves structure lifetime over LEACH clustering since LEACH carelessly picks cluster heads (and from this time forward cluster sizes), which may achieve speedier death of a few hubs. The last cluster heads picked in HEED are especially coursed over the network and the correspondence cost is obliged[15].

**III. EXISTING SYSTEM**

This segment portrays the past system used for secure sending the data.

Working of the past system is as per the following:

1. Network chart created as Graph G (V, E) where; V is vertices/hubs and E is edges.
2. Clustering is done on the quantity of the hubs and delegated number of clusters and pick the cluster head haphazardly.
3. Play out the key dispersion and course ages at every hub through Base Station.
4. Make the data and Encrypt with people in general key of base station at every hub.
5. Process the hash estimation of the scrambled data and Record the timestamp.
6. Forward the individual data to the cluster head from each cluster part in every one of the clusters.
7. Get all data at the cluster head and check the data by its hash esteem and acknowledge the confirmed data or dispose of if not confirmed.
8. Solidify every one of the data and forward the same to the base station.
9. Base station gets the data from each cluster head.
10. Base station confirms the data and unscrambles the data with legitimate key.

IV. PROPOSED SYSTEM

This section portrays the system review in which proposed calculation and scientific model of the proposed system is likewise present.

System Overview

System architecture of the proposed is shown in figure 1 which appears in different advances and steps are provided beneath.

1. Network Generation
At begin network is created where vertices/hubs are related with the edges.

2. Clustering Process
After the network generation, the clustering strategy is executed in that hubs are isolated in various clusters.

3. Cluster Head Selection
In the wake of making the gathering of clusters, from each gathering of clusters, the cluster head is picked based on vitality and separation from base station and neighbor hubs parameters.

4. Key generation and distribution
Base station can achieve key generation and dissemination to each hub. Course ages performed from each hub to the base station.

5. Data Encryption
At every node data is generated and encrypted through the Paillier Encryption.

6. Hash value evaluation
After the data is encoded, hash esteem is evaluated and recorded the timestamp.

7. Data Collection
Consequent to evaluating, the hash regard at every center point, every center advances data to its cluster head. Cluster head have some constrained capacity to store the data if the cluster head stockpiling is overwhelmed then the data is dropped at assemble head. The cluster head merges each one of the data and check the substantial data.

8. Cached Data
In system, to restrain the loss of data at cluster head because of the impediment of capacity limit we are keeping a cache stockpiling that can store the data dropped during the time spent data sending in cluster individuals and cluster head.

9. Data verification
By batch verification method, validate the information by making use of hash value and timestamp. In this we are verifying cached data also data which is stored in cluster head storage.

10. Data aggregation
At last, process of data aggregation is accomplished after verifying the valid data by the cluster head and data forwarded to the base station.

11. Data Decryption
Base station receives the data from every cluster head and decrypts the data by the appropriate key.
Algorithm

Algorithm 1: Proposed Algorithm

Step 1: Generate a network chart as Graph G (V, E) where; V is vertices/hubs and E is edges.
Step 2: Implement clustering calculation over the quantity of hubs and separate the hubs in to number of clusters.
Step 3: based on vitality, number of neighbors and separation to the base station select the Efficient Cluster Head.
Step 4: Perform the key conveyance at each hub by means of Base Station.
Step 5: Perform the course ages from each hub to the base station.
Step 6: Create the data at each hub and scramble the data with general society key of base station.
Step 7: Compute the hash estimation of the scrambled data and Record the timestamp.
Step 8: Send the individual data to the cluster head from each cluster part in every one of the clusters. In the event that capacity limit of cluster head is surpass the farthest point at that point store the data in cache memory.
Step 9: Collect all data at the cluster head. Confirm the data by its hash esteem and acknowledge the checked data or dispose of if hash esteem is invalid.
Step 10: Aggregate every one of the data and send this data to the base station. Base station acknowledges the data from each cluster head.
Step 11: Base station checks the data and unscrambles the data with proper key.

Explanation: Proposed calculation outlines the working stream of the structure. At in the first place, system is made including sensor center points; additionally clustering calculation and number of hubs is isolated in number of clusters. Clusters head is picked in light of parameters; key circulation is executed at every center point by the base station. Course is produced using every center point to the base station. Data encryption is done through the Paillier Encryption with the private key. Hash esteem is surveyed of the scrambled data and timestamp is recorded. Cluster part progresses the data to the cluster head in all clusters and surge data is secured in cache memory. Data check is done on the start of hash esteem; in case it is affirmed then simply recognized for the most part rejects. After that total each data is forward to the base station. Base station decodes the data with the appropriate keys.

Algorithm 2: Paillier Cryptosystem

Step 1: Key Generation:
   a) Select two large prime numbers a and b arbitrary and independent of each other such that gcd(n, Φ (n)) = 1, where Φ (n) is Euler Function and n=ab.
   b) Calculate RSA modulus n = ab
   c) Carmichael's function is given by λ = LCM (a-1, b-1).
   d) Select g called generator where g∈ℤ*n2. Select α and β randomly from a set ℤn* then calculate g = (an + 1) β mod n.
   e) Compute the following modular multiplicative inverse μ = (L(gλ mod n2)-1) mod n. Where the function L is defined as L(u) = (𝑢−1)/𝑛.

The public (encryption) key is (n and g).
The private (decryption) key is (λ and μ).

2) Encryption:
   a. Let mess be a message to be encrypted where mess∈ ℤn.
   b. Select random r where r ∈ ℤ*n2.
   c. The cipher text can be calculated as: Cipher = g^mess.r^n mod n^2.

3) Decryption:
   a. Cipher text c ∈ ℤ*n2

Original message: mess = L (cipher^λ mod n^2).μ mod n.
Experimental Setup
System buildson Java framework (version jdk 8) over Windows platform. For development, the Netbeans (version 8.1) tool is utilized. The network is created utilizing Jung tool with sensor nodes. System doesn’t require any particular hardware to run any standard machine is able to run the application.

V. RESULT AND DISCUSSION
Following are Results generated during the implementation of the system.

![Energy Consumption of the nodes for sending the Data](image)

**Figure 2.** Energy Consumption of the nodes for sending the Data

The results shown in Figure 2 gives the Energy Consumption while sending the data. For better analysis of the results we have shown the results of 5 tries. The energy consumption vs data size graph shows the energy consumption of the nodes in traditional system and proposed system. The blue colored dotted line represents energy consumption of data with respect to data size in proposed system and the red dotted line represents the same in existing system.

VI. CONCLUSION

By utilizing proposed system we can boost the network lifetime of WSN likewise built up the strategy that can choose the cluster head contingent upon three parameters by which network can use vitality effectively and lifetime of the Wireless Sensor Network get moved forward. Proposed strategy additionally built up a system for data recuperation which is lost while broadcasting the data. Finally the result demonstrates that the proposed system will amplify the network lifetime.

VII. REFERENCES


