Implementation of Clustering For Data Aggregation Using Evolutionary Algorithm

Ms.Sheetal V.Chate Prof.S.B.Javheri
Department of Computer Engineering Rajarshi
Shahu College of Engineering, Pune, India.

***Abstract*— Network is a collection of different nodes which co-operatively send detected information in Local Area Network. The essential target of information accumulation calculations is to assemble a total information in a vitality efficient way. One such approach is information package which fetching strategy for information gathering in broadcast framework architectures and dynamic access via wired connectivity. Genetic algorithm is a class of developmental algorithm. In this work, we have proposed clustering for data aggregation using evolutionary algorithm in Network. In which Intermediate Node (GH) are selected on basis of node connectivity which act as a data aggregator (DAG).Then the clustering process is executed using evolutionary algorithm such as Genetic Algorithm (GA). When group member want to send a data to Intermediate Node or Group Head (GH) which act as a data aggregator then they use data encryption which offers data confidentiality, data integrity, and authenticity. Group Head is also called as Intermediate node .But it is important that group member must be accepted by Group Head(GH) so that they can transmit data to Intermediate Node or group head(GH).**

Index Terms*— *Clustering, Evolutionary Algorithm, Data Aggregation

I. INTRODUCTION

Recently, Networks are becoming an essential part of many application environments that are used in different application. network is composed of a large number of nodes, which co-operatively send detected information in Local area Network. In day to day life we read about network everywhere like on TV or Newspaper etc. Clustering means connecting two or more computers together in such a way that they behave like a single computer. Clustering is used for parallel processing, load balancing and fault tolerance. Clustering a network involves partition of nodes into independent cluster. There are various methods in clustering such as K-means and Fuzzy C-means clustering. Below Figure 1 shows example of cluster network. Data aggregation is process of collecting Information from different sources and reduces redundancy. Security plays a important role in data aggregation process. It is important to protect aggregated data various attack and ensures data integrity, data confidentiality and data freshness. [6][15].To preserves privacy, end-to-end data encryption is able to protect communications between source node and destination node .Data aggregation has 3 types In-Network Data Aggregation, Cluster based data aggregation and Tree based data aggregation. In this paper we are using cluster based data aggregation. In cluster-based, nodes are collected into clusters and each cluster consists of a Intermediate Node (GH) and some other member. Every Intermediate Node (GH) collects and aggregates the data of its members, then, transmits the fused data in the LAN (local area network). To perform clustering we are using Evolutionary algorithm. In this approach we used genetic algorithm for clustering formation which is a class of evolutionary algorithm. So, Evolutionary algorithms encompass genetic algorithms, and more. Although genetic algorithms are the most frequently encountered type of evolutionary algorithm, there are other types, such as Evolution Strategy. Genetic algorithms use crossover and mutation to search the space of possible solutions. Genetic Algorithm is better than conventional Artificial Intelligence in that it is more robust. Unlike older Artificial Intelligence systems, they do not break easily even if the inputs changed slightly, or in the presence of reasonable noise. Genetic algorithm may offer significant benefits over more searches of optimization techniques. [17]

II. LITERATURE REVIEW

In clustered environments, there are some approaches for data aggregation. The first approach is known as the Group Head(GH) method. The idea behind the approach is that One node in the cluster will be selected as the GH based on node connectivity. The remaining nodes in the cluster will Send data to the GH. The GH will collect all the data and will forward these data to the destination node. [3]

Patil N.S [2] has proposed data aggregation framework on wireless sensor networks is presented. The framework works as a mediator for aggregating data which is measured by a number of nodes within a network. The main goal of the proposed work is to compare the performance of TAG.

Hevin Rajesh, D. Paramasivan [11] has proposed a fuzzy-based secure data aggregation technique. Initially, the clustering is performed with the selection of Group Head(GH) based on the node connectivity. Based on the parameters estimated, Fuzzy logic technique was used to select the secure and Non-faulty node members for data aggregation. Then, aggregated data from the GH was transmitted to the sink.

Nan Guofang [12] has proposed an algorithm to select a Group Head(GH) ,which performs data aggregation in a partially connected networks, it works only for a partially connected network. This algorithm reduces traffic flow within network by selecting a shortest path for routing of data to Group Head(GH).

Xue Liu [14] have proposed two privacy-preserving data
Aggregation schemes 1) Cluster-based Private Data Aggregation (CPDA) and 2) Slice -Mix-Aggregate (SMART) – which focuses on additional data aggregation function. Their main goal is to fill the gap between collaborative data collection by WSN and data privacy.

Mehrjoo et al [15] have proposed a hybrid genetic algorithm (GA) and artificial bee colony (ABC)-based clustering algorithm. GA was used to select the GHs and their number and ABC to select the cluster members. The drawback of this approach is that they use intelligent algorithms like ABC which increases energy consumption.

Alok Chakrabarty [16] have proposed multi-level data aggregation among Group Head(GH)s to reduce the packet size which reduces the transmission and receiving energy of node. They also performed GH selection and cluster formation based on residual energy criterion.

M. H. Yaghmaee [17] has proposed Genetic Algorithm (GA) to optimize sensor nodes energy consumption. Main focus is on clustering technique as an efficient way for reducing energy consumption of a sensor node as well as the cost of transmission.

III. IMPLEMENTATION DETAILS

A. System Overview

The system work as follows:

1. Group Development.

Clustering or grouping is performed using generic algorithm. An adaptive technique that helps in solving search and optimization problems corresponds to GA. It is based on the genetic processes.

2. Developing data aggregation.

Initially, the GH or Intermediate nodes are chosen based on the node connectivity, which acts as a data aggregator (DAG). Data aggregation aims at eliminates redundant data transmission. An improvement over the above approach would be clustering where each node sends data to Intermediate node(GH) and then Intermediate node(GH) perform aggregation on the received data and then send it to Server.

3. Data Encryption and Data Decryption.

This technique provides the secure communication framework that verifies the Data and drops the false Data from malicious nodes. This technique requires each cluster Member and DAG_i to store some information in its cache that includes node ID.

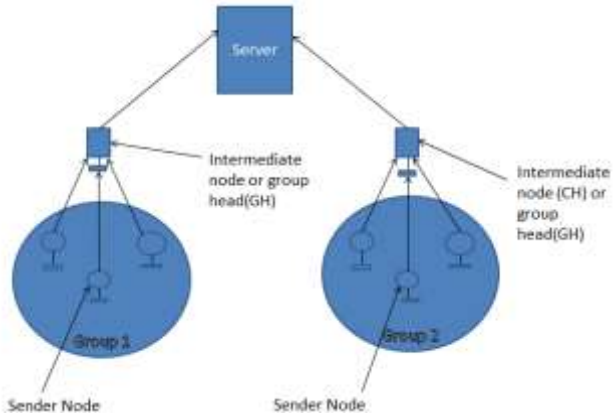

Fig. System Architecture

B. Algorithm Used

Algorithm 1: In this approach we used genetic algorithm for group or cluster formation which is a class of evolutionary algorithm. So, Evolutionary algorithms encompass genetic algorithms, and more. Although genetic algorithms are the most frequently encountered type of evolutionary algorithm, there are other types, such as Evolution Strategy. Genetic algorithms use crossover and mutation to search the space of possible solutions.

- Genetic Algorithm:

1) Begin
2) Creation of an initial population
3) Computing of fitness of each individual
4) While (not stopping condition)
5) do
6) Select parents from population
7) Execute crossover to produce offspring
8) Perform mutations
9) Compute fitness of each individual
10) Replace the parents by the corresponding
11) End if
12) End if

Algorithm 2: RC7

B = B + S [0]
D = D + S [1]
F = F + S [2]
for i = 1 to r do
f t = (B x (2B + 1))_lg w
u = (D x (2D + 1))_lg w
v = (F x (2F + 1))_lg w
A = ((A N t)_u) + S [2i+1]
C = ((C N u)_t) + S [2i+ 2]
E = ((E N v_t) + S [2i+ 3]
(A, B, C, D, E, F) = (B, C, D, E, F, A)g
A = A + S [2r - 1]
C = C + S [2r]
E = E + S [2r + 1]

Mathematical Model:
Let S is the Whole System Consists:
S = {N, G, GH, DA, DE, DD}.

1) $N=\{N1,N2...N_n\}$.

2) $G=\{G1, G2....G_n\}$.

3) $GH=\{GH1, GH2....GH_n\}$

4) $DA=\{DA1, DA2....DA_n\}$

5) $DE=\{DE1, DE2....DE_n\}$

6) $DD=\{DD1, DD2....DD_n\}$

Step 1: All the nodes in a network are combined in group or cluster.
G= {G1, G2….G_n}.
Step 2: on the basis of node connectivity Intermediate Node or Group Head(GH) are choose which work as a data aggregator.
GH= {GH1, GH2….GH_n}
Step 3: perform data aggregation on a cluster.
DA= {DA1, DA2….DA_n}.
Step 4. While sending data, data encryption is done on sender side.
DE= {DE1, DE2….DE_n}
Step 5.on the receiver side data decryption is done.
DD= {DD1, DD2….DD_n}

C. Memorization Parameter:

Table: Memorization Parameter

Symbol	Meaning
N	Set Of Node in network
G	Set Of Cluster or Group
GH	Set Of Intermediate Node or group head(GH)
DA	Data Aggregation
DD	Data Decryption
DE	Data Encryption

D. Experimental Setup

The system is built using Java and J2EE framework (version J2SDK 1.5), web technologies (JSP, HTML, CSS), Apache Tomcat, MySQL database on Microsoft Windows platform. The experiments have been performed on the machine with the following specifications: Intel Core 2 dual processor with 2.5 GHz CPU, 2 GB RAM, 180 GB Hard disk and running Microsoft Windows 7 Home Basic x64 OS.

IV. RESULT AND DISCUSSION

In this work, we are using genetic algorithm for development of clustering which require less energy and also ensures security. Also we are going to compare performance of proposed technique with existing technique. Below figure (4.2) Nodes vs delivery ratio shows delivery ratio of our system is higher than existing system for different nodes scenario. figure 4.1 shows the file or data delivery ratio when the number of attackers is increased.

B. Results

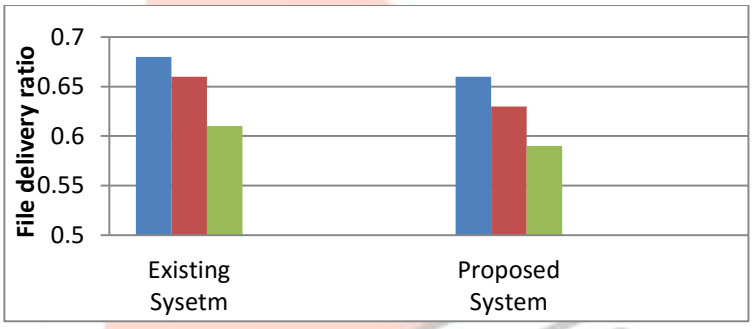

Figure 4.1

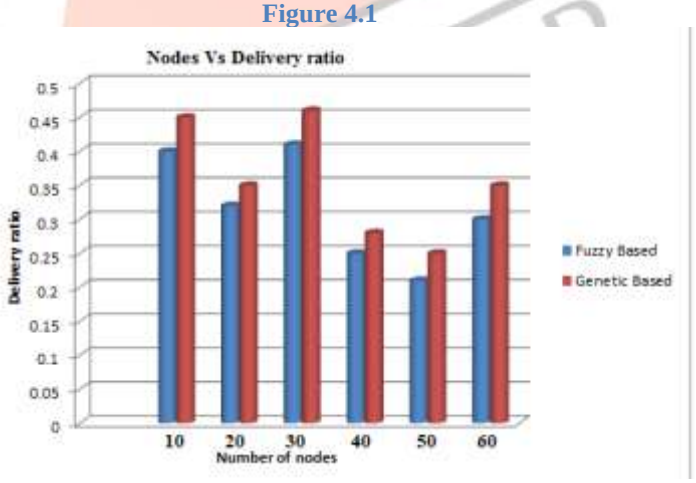

Figure 4.2

V. CONCLUSION

In this approach we propose a protocol which ensures security and secures data transmission. Then, clustering process is executed using the genetic algorithm. This technique highly minimizes energy consumption. The Clustering is performed with the selection of Intermediate node (GH) based on the node connectivity, which acts as a DAG. When the cluster member wants to transmit data to the aggregator, a data encryption technique is utilized, which offers confidentiality to the Data, thus ensuring the authenticity and integrity of the data.

VI. REFERENCES

[1] Lathies Bhasker: 'Genetically derived secure cluster-based data aggregation in wireless sensor networks'.Dept.of. Comp.Science and Engineering, Manonmaniam Sundaranar University, Tirunelveli, India

[2]N.S. Patil, P.R.Patil: 'Data aggregation in wireless sensor network'.IEEE Int. Conf. Computational Intelligence and Computing Research, 2010

[3]Roy, S., Conti, M., Setia, S., Jajodia, S.: 'Secure data aggregation in wireless sensor networks' , IEEE Trans. Inf. Forensics Sec., 2012, 7 , (3)

[4]Bhoopathy, V., Parvathi, R.M.S.: ' Securing node capture attacks for hierarchical data aggregation in wireless sensor networks', Int. J. Eng. Res. Appl., 2012, 2, (2), pp. 466– 474

[5]Jha, M.K., Sharma, T.P.: 'A new approach to secure data aggregation protocol for wireless sensor network', Int. J. Computer. Sci. Eng. (IJCSE), 2010, 02, (05), pp. 1539– 1543

[6]Makin, B.A., Padha, D.A.: ' A trust-based secure data aggregation protocol for wireless sensor networks', IUP J. Inf. Technol., 2010, VI ,(3), pp. 7

[7]Sen. J: 'A survey on wireless sensor network security', Int. J. Communication Network. Information. Security. (IJCNIS), 2009, 1, (2), pp. 55– 78

[8]Ozdemir,S., Xiao, Y.: 'Secure data aggregation in wireless sensor networks: A comprehensive overview', Comput. Network. 2009, 53, (12), pp. 2022–2037

[9]Jha, M.K., Sharma, T.P.: 'Secure data aggregation in wireless sensor network: a survey', Int. J. Eng. Sci. Technol. (IJEST), 2011, 3, (3),pp. 2013– 2019

[10]Kwon, T., Hong, J.: 'Secure and efficient broadcast authentication in wireless sensor networks', IEEE Trans. Computer., 2010, 59, (8), pp. 1120– 1133

[11]Hevin Rajesh, D., Paramasivan : 'Fuzzy based secure data aggregation technique in wireless sensor networks' , J. Computer. Sci., 2012, 8, (6),pp. 899–907

[12]Mozumdar, M.M.R., Nan, G., Gregoretti, F., Lavagno, L., Vanzago, L.:'An efficient data aggregation algorithm for cluster-based sensor network', J. Netw., 2009, 4 , (7), pp. 598– 606

[13]Perez-Toro, C.R., Panta, R.K., Bagchi, S.: 'RDAS: reputation-based resilient data aggregation in sensor network'. IEEE SECON 2010 Proc.

[14]He, W., Liu, X., Nguyen, H., Nahrstedt, K., Abdelzaher, T.: 'PDA: privacy-preserving data aggregation in wireless sensor networks' .IEEE ICWMC, 2010

[15] Mehrjoo, S., Aghaee, H., Karimi, H.: 'A novel hybrid GA–ABC based energy efficient clustering in wireless sensor network', Can. J. Multimedia Wirel. Netw. 2011, 2, (2), pp. 41–45

[16]Thakkar, H., Mishra, S., Chakrabarty, A.: 'A power efficient Cluster-based data aggregation protocol for WSN (MHML)' Int. J. Eng. Innov. Technol. (IJEIT), 2012, 1 , (4), pp. 241– 246

[17] Zahmatkesh, A., Yaghmaee, M.H.: 'A genetic algorithm-based approach for energy- efficient clustering of wireless sensor networks'. Int. Conf. Network Communication and Computer (ICNCC), 2011.