ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 2, August 2013

Distributed Query Processing via Secure Routing in Wireless Sensor Network

R. Jebakumar, Dr. P. Vivekanandan

Abstract*— Query processing and its data transmission is a very energy consuming operation in Wireless sensor networks (WSNs). Distributed query processing is a challenging task via secure communication from remote users and the base station (BS) because security is an issue to consider for secure query transmission. To minimize the transmission costs and reduce the time for query processing, the Distributed data server (DDS maintains all transaction and reduce the overhead of centralized DDS and increase the query performance assigned in each BS as a virtual DDS in distributed environment. A randomize secure key is generated periodically by a key distribution center and that is shared to all the secure nodes in this network. In this proposed architecture, a DDS is connected with the BSs and each BS has their own data with their data mart and maintains the distributed meta data overall and its controlled by DDS. This system implements the distributed and parallel query processing among virtual DDSs. It is cared by BS based on the availability of resources between them for secure query processing.

Index Terms—**Distributed data server, Distributed query processing, Key distribution center, Wireless sensor networks.**

I. INTRODUCTION

In wireless sensor networks, query processing is a critical problem due to the huge volume of real time physical data collected from sensors. Query processing is based on the centralized approach, when the base station wants to disseminate a query for execution in distributed environment. The time consumption is more for searching and query redirection for query processing in WSN. In some special cases, WSNs don't have base station at all. Such networks include a military WSN in a battlefield and a WSN situated in a remote area to monitor illegal activities, the centralized approach is not applicable to such WSNs. If that kind of situation arises, we modify the network and use some special nodes which is called storage node to use as BS. It contains more capacity than regular sensors that transmits sensed data to the distributed database. If a WSN having a single centralized server, it affects the quality of service. In this paper, the DDS is introduced among the BS and each BS contain their own region data to serve the query in efficient way and also have the frequently used data set also available in it to overcome such problem arises the query transmission with in a WSN. Here no need to interact the centralized server for each and every time when it is required for some information regarding the other region data through this server. After sending the request the base stations can interact between them. The server utilization can be avoided and also encouraged the distributed communication and sharing environment for the query execution. In WSN query processing systems such as Tiny Database [1],[2] are promising for data acquisitional applications of WSNs. In these systems, a user injects SQL queries into the network through a remote PC. The networked sensor nodes then work together to process the queries and send results back to the PC. This query processing paradigm is more efficient and flexible than centralized query processing [3]. Nevertheless, power consumption remains a critical issue in these systems [4],[5]. In this paper, a new architecture is proposed for distributed data processing and each BS acts like a DDS for query processing and it contains the fact data of overall network and maintaining by the DDS for query processing to reduce the overhead of centralized system, improve query performance and utilize the resources in efficient manner with secure routing too. Random key distribution was proposed by many researchers for secure communications within WSNs. The key distribution center which distributes the keys to all secure nodes randomly chosen from a large pool of keys. After this, neighboring nodes use these keys to establish a pairwise key between them. Communications between neighboring sensors in each hop are encrypted/decrypted using these pairwise keys. Many key management protocols have been proposed based on key pre-distribution [6]..[10], etc., each one list out more features about the same for secure communication. In this paper the group key is generated by the DDS randomly and distributes it to all secure nodes for secure communication and the key generation done based on specific time intervals.

II. RELATED WORK

A centralized scheduling algorithm was proposed for data distribution and collection in WSNs [11]. The sink allocates the transmission timing to all nodes participate in a WSN. This centralized nature, the sink to know the current network topology, which is difficult in practice. It is often costly to disseminate schedules in a network. To avoid the problem, some of the distributed scheduling schemes have been proposed. A distributed on-demand power-management protocol for tree networks [12]. Here, a parent randomly chooses reserved slots and broadcasts the reserved slots. And a child sends a request for a specific reserved slot if it has some message to send, and the parent confirms the request if the slot has not been requested by other children. The problem is that it does not reduce the collisions between neighbors. The distributed scheduling is a scheme proposed [13], in this scheme, a source node first broadcasts a special route setup packet to set up a route and a temporary schedule with a

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 2, August 2013

neighbor. If the route setup packet finally arrives at the sink, the nodes along the path will set their temporary schedules to be permanent schedules; otherwise, the temporary schedule of the source node will be removed. If collision occurs during this setup process, the route setup packet will be postponed to the earliest time when the node does not transmit or receive. The nodes are synchronized under the scheduling scheme and their tasks are similar, collisions may occur frequently. Consequently, even after the route setup transmission of a node is postponed, collisions may still occur when the transmission is started again. The query results accuracy in sensor query processing is low if the number of dead nodes is large. There are many works proposed for key management. In [14], cluster heads are assumed to be equipped with a fast encryption/decryption algorithm to protect their supplementary keys from compromise. Their proposed idea is only for pair-wise key establishment [14]. A key management scheme [15] is proposed, the keys are divided in different categories such as cluster key, intermediate key and private key of each sensor, it may suffers from poor resilience under node compromises. And a protocol proposed for key additions and revocations under network dynamics [16] via one-way hash functions. Hash functions used for key revocation and refreshment however can be energy consuming. In [17] proposed two resilience aware routing protocols: data centric and location centric protocols for differentiated keys among sensors and do not use different cluster head keys, but rather use same key pools that reduces complexity during key pre-distribution and subsequent pair-wise set-up.

III. NETWORK MODEL

A sensor network consists of a large number of sensor nodes and each sensor nodes is connected to other nodes via a wireless network. The sensor nodes have limited computation and storage capabilities. In this paper a special type of node called a storage node is connected with sensors for communication with BS of the sensor network and minimizes the complexity of centralized system. All storage nodes are connected with BS and all BSs are directly connected with DDS. A typical WSN for query processing is shown in Fig 1. In this WSN, the DDS is used to distribute data to the sensor network and to receive data from the network via the BS connected with the network.

The BS forwards commands and queries to the sensor nodes to receive sensed data. The BS processes the queries to produce the output via distributed query processing. Finally, query results are forwarded towards the BS to the user. The storage nodes are same as sensors and network is selected to represent as a storage node which contains more capacity than normal sensors, it used for collecting the sensed data from other sensors and do the aggregation processes if required. If aggregation is over for the same it is sent to the BS. It reduces the number of packet sent to the BS and reduces the overall network utilization too. The BSs are mimicry as the centralized one for the data processing because all BSs shared by the same fact tables maintained by the DDS for known about all data available in the network.

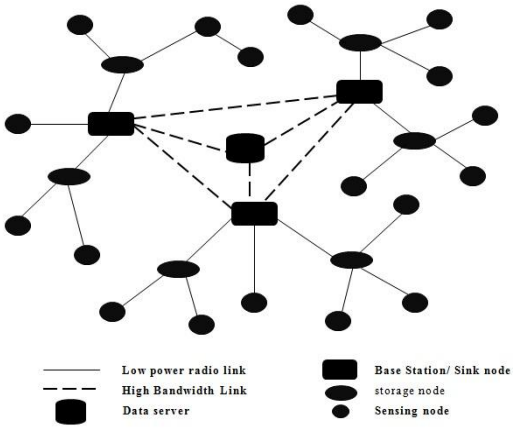

Fig. 1 sample architecture of WSN

IV. METHOD

A. Key distribution center

Group Key is a globally shared key that is generated by the DDS and shared by the base stations for encrypting and decrypting of the messages that transmit to the whole network. A DDS will act like a key distribution center, which will generate a random key based on the time interval and distributes it to all secure nodes in the network. The fast RSA algorithm is chosen for protecting the data because the computation is very limited and easily distributes the keys to all secure nodes. The digital signature is added for more security to shield the data after encryption.

B. Data extraction and loading

Here the data is extracted from DDS when it is required or automatically updated when such event occur in the network. And it is loaded to all BSs based on its region or part of their clustered data among the whole network. The fact data is distributing to all BS and maintain by the DDS. The data clustering is done by the origins of the received data from sensors. This data clustering is used for retrieving the data in efficient way and also distribution of the data is not complicated by DDS.

C. Frequently used data set

Frequently used data set also placed in all BS to reduce the computation and easy way to process the information from the network and satisfy the user in time. And the frequently used data sets are varying from BS to BS; some common data set may be placed in all BSs. Most frequently posted queries by users, that related data set is available in particular BS an also available in DDS also. It will dynamically change based on user queries, a counter is maintaining for this purpose, it contains the query sets and how many number of times the query was posted from present time and look back up to that specific time interval.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 2, August 2013

D. Distribute and query processing

Every base station known very well about the whole data in the network because the fact data is available in all base station and it is maintained by DDS. The system analyzes the query Q_i and generates the query execution plan for the same when the query Q_i is posted by the user U_i to the network and the authentication is successfully over. The process hierarchy is shown in fig 2 for distributed query process.

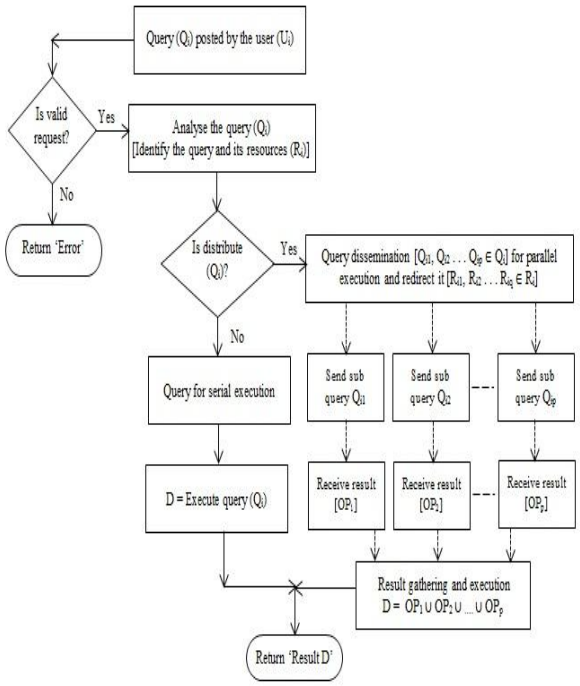

Fig. 2 Distributed query processing

In the first half of the process is checked for authentication and query analysis which generate query execution plan for distribute query process. The second half of the process is redirected the query for distributed computation and received the output for the same and merges the result to produce the output D for the query Q_i. The complete process of distributed query process is explained in the next chapter.

V. ALGORITHM AND DISCUSSION

The users post the query to the network if they want to retrieve any data from the network. When any request is received from the user or any other part of the network to execute the query Q_i, the query execution system will be applied for multilevel authentication and query validation before processing the query. For multilevel authentication, first it checks whether the query received from secure node or not that is $S_i \in SN$, here the i^{th} query Q_i is received from node S_i and SN is the secure node list maintains by the DDS in the network to secure routing process. If it is secure, then it will check the user, $U_i \in UL$ and mapping the query Q_i with U_i, here the user U_i must belong to a member of UL and the user U_i having the access permission with Q_i (check the relations with the particular user). When the above conditions are satisfied, it permits to process the query else it will raise the error and discard the request.

Symbol	Description
S_i	Source Node for i^{th} query
U_i	User posted the i^{th} query
Q_i	i^{th} input query
Q_{ij}	j^{th} sub query from Q_i
SN	Secure Node list
BL	Block List
UL	Set of authenticated user list
R_{ik}	K^{th} Resource assigned for sub query from Q_i
D	Buffer data set
EK	Encryption process by random key
ED	Encrypted Data
OD	Output Data
OP_j	Output for Q_{ij}

Table. 1 Notations used in algorithms

```
Algorithm User_Query_Process (Si, Ui, Qi)
1.  If (Si ∈ SN)
2.     If !(Ui ∈ UL) Return Error 'Authentication Failure'
3.     If (IsNotExists (Qi, Ui)) Return Error 'Access permission failure'
4.     Qi1, Qi2 . . . Qip ∈ Qi
5.     If (IsDataAvail (Qi))
6.        Execute Qi based Query execution plan
7.     Else
8.        Redirect_Query (Qi)
9.        D = Receive_Output (Qi)
10.       ED = EK (D)
11.       OD = ED + Signature
12.       Return (OD, Si)
13.    End if
14. Else
15.    If (Si ∈ BL)
16.       Discard Request
17.    Else
18.       Send request to detect misbehaviour for Si
19.    End if
20. End if
```

Algorithm 1

In algorithm 1, after the authentication is over, the query processing system will check the feasibility for query dissemination during the query execution plan generation for parallel execution, if it is feasible, then the query Q_i is distributed $Q_{i1}, Q_{i2} \ldots Q_{ip} \in Q_i$ or it won't be distributed to the query Q_i. Then it will check the data availability for the given query Q_i, if the desired data is available within node or BS the system will execute the query within it without query forwarding, if it is not, the query Q_i is redirected to where the data will be available based on it query execution plan. In algorithm 2, it defines for redirecting the query based on the availability of data from various resources within the network, let K^{th} resource (R_{ik}) is assigned for sub query Q_{ij} (j^{th} sub query from Q_i) and redirect all sub queries $Q_{i1}, Q_{i2} \ldots Q_{ip} \in Q_i$ to resources $R_{i1}, R_{i2} \ldots R_{iq} \in R_i$ for parallel execution.

```
Algorithm Redirect_Query (Qi)
1. Assign Ri1, Ri2 . . . Riq ∈ Ri for Qi1, Qi2 . . . Qip ∈ Qi
2. For j = 1 to p do
3.    Let k is Resource ID for Qij
4.    Send Qij to Rik
5. End for
```

Algorithm 2

In algorithm 3, it defines how the node or BS receives the

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 2, August 2013

outputs for each and every sub queries Q_{i1}, Q_{i2} . . . $Q_{ip} \in Q_i$ from the distributed resources R_{i1}, R_{i2} . . . $R_{iq} \in R_i$. During this process, the output OP_j will be merged with each one based it execution plan hierarchy with output data set D ($D = D \cup OP_j$), if all the sub queries are merged, then it will return to called one (algorithm 1) for the query result to the user or the requested one among the network.

Algorithm Receive_Output (Q_i)

1. D = NULL
2. For j = 1 to p do
3. Receive (OP_j)
4. $D = D \cup OP_j$
5. End for
6. Return (D)

Algorithm 3

When the query execution is successfully completed, then it will go for encrypt the output data by using the group key and finished the encryption process EK(D) on data D and get updated data ED. Fast RSA algorithm is used for encryption and decryption because the minimum computation is required for accomplishing this encryption process and this is more secure too. The digital signature will be added after the encryption process is over for ensuring the confirmation on the output data (OD). When the authentication fails ($S_i \in SN$) for the query Q_i, then the system will check the input node S_i with BL for reporting the entry of misbehaving node and discard the query request. When the both conditions are failed (either SN or BL) it will send a request for detecting the misbehaving node to the misbehavior detection process in the network.

VI. SUMMARY AND CONCLUSION

A modified WSN architecture was proposed for distributed query processing via the secure communication. The algorithms are defined on how the query dissemination, redirection for distributed process and generate the output for the user query in distributed manner. Various resources in the network are used efficiently for distributed and parallel execution. A key distribution center used for generating random group keys periodically and maintains it from DDS and no specific node is required. The multilevel authentication used for providing more security and allowed only for secure nodes for communication within the network. To execute the query in distributed environment, it will identify the resources and disseminate the query based on query plan with secure routing and collect the result vice versa. The new architecture is used to speed up the query execution process and easy way to share the secure key and supported for distributed processing. Distributed block list is used for secure routing and data transmission too. In future, this network structure will be upgraded based on it desired applications and availability.

REFERENCES

[1] Samuel Madden, Michael J. Franklin, Joseph M. Hellerstein, and Wei Hong, “TAG: a Tiny Aggregation Service for Ad-Hoc Sensor Networks”, OSDI, 2002.

[2] Samuel Madden, Michael J. Franklin, Joseph M.Hellerstein, and Wei Hong, “The Design of an Acquisitional Query Processor for Sensor Networks”, SIGMOD, 2003.

[3] Philippe Bonnet, Johannes Gehrke, and Praveen Seshadri, “Querying the Physical World”, IEEE Personal Communications, vol. 7, no. 5, pp. 10-15, October, 2000.

[4] Athanassios Boulis and Mani B. Srivastava, “Node-Level Energy Management for Sensor Networks in the Presence of Multiple Applications”, Per Com, 2003.

[5] Phil Buonadonna, Joseph Hellerstein, Wei Hong, David Gay, and Samuel Madden, “TASK: Sensor Network in a Box”, EWSN, 2005.

[6] H. Chan, A. Perrig, and D. Song, “Random key predistribution schemes for sensor networks," in Proc. IEEE Symp. Research Security Privacy, May 2003.

[7] W. Du, J. Deng, Y. S. Han, and P. K. Varshney, “A pairwise key predistribution scheme for wireless sensor networks," in Proc. 10th ACM Conf. Comput. Commun. Security, Oct. 2003.

[8] J. Lee and D. R. Stinson, “Deterministic key predistribution schemes for distributed sensor networks," in Proc. 11th Workshop Sel. Areas Cryptography, Aug. 2004.

[9] J. Lee and D. R. Stinson, “A combinatorial approach to key predistribution for distributed sensor networks," in Proc. IEEE Wireless Commun.Netw. Conf., Mar. 2005.

[10] D. Liu and P. Ning, “Establishing pairwise keys in distributed sensor networks," in Proc. 10th ACM Conf. Comput. Commun. Security, Oct. 2003.

[11] Cédric Florens and Robert McEliece, “Packet Distribution Algorithms for Sensor Networks”, INFOCOM, 2003.

[12] Barbara Hohlt, Lance Doherty, and Eric Brewer, “Flexible Power Scheduling for Sensor Networks”, IPSN, 2004.

[13] Mihail L. Sichitiu, “Cross-Layer Scheduling for Power Efficiency in Wireless Sensor Networks”, INFOCOM, 2004.

[14] P. Traynor, H. Choi, G. Cao, S. Zhu, and T. L. Porta, “Establishing pair-wise keys in heterogeneous sensor networks," in Proc. 25th IEEE Conf. Comput. Commun., Apr. 2006.

[15] A. Poornima and B. Amberker, “Tree-based key management scheme for heterogeneous sensor networks," in 16th IEEE International Conf. Netw., 2008.

[16] Y. Zhang, W. Yang, K. Kim, and M. Park, “An AVL tree-based dynamic key management in hierarchical wireless sensor network," in Proc. International Conf. Intelligent Inf. Hiding Multimedia Signal Process., pp. 298-303, 2008.

[17] Wenjun Gu, Neelanjana Dutta, Sriram Chellappan, and Xiaole Bai, “Providing End-to-End Secure Communications in Wireless Sensor Networks” in IEEE Transactions on networks and service management, Vol. 8, No. 3, September 2011.