Fault-tolerant Data Aggregation Scheme for Monitoring of Critical Events in Grid based Healthcare Sensor Networks

Center for Informatics and Applied Optimization
Ather Saeed
Andrew Stranieri
Richard Dazeley
University of Ballarat and Melbourne Institute of Technology

asaeed@mit.edu.au, a.stranieri@ballart.edu.au, r.dazeley@ballarat.edu.au

Abstract

Wireless sensor devices are used for monitoring patients with serious medical conditions. Communication of content-sensitive and context sensitive datasets is crucial for the survival of patients so that informed decisions can be made. The main limitation of sensor devices is that they work on a fixed threshold to notify the relevant Healthcare Professional (HP) about the seriousness of a patient's current state. Further, these sensor devices have limited processor, memory capabilities and battery. A new grid-based information monitoring architecture is proposed to address the issues of data loss and timely dissemination of critical information to the relevant HP. The proposed approach provides an opportunity to efficiently aggregate datasets of interest by reducing network overhead and minimizing data latency. To narrow down the problem domain, in-network processing of datasets with Grid monitoring capabilities is proposed for the efficient execution of the computational, resource and data intensive tasks.

Interactive wireless sensor networks do not guarantee that data gathered from the heterogeneous sources will always arrive at the sink (base) node, but the proposed aggregation technique will provide a fault tolerant solution to the timely notification of a patient's critical state. Experimental results received are encouraging and clearly show a reduction in the network latency rate.

Keywords: Data aggregation, grid computing, in-network processing, sensor networks, fault-tolerance

1. INTRODUCTION

In-network monitoring of critical events is a challenge in distributed and decentralized wireless sensor networks, especially when it comes to monitoring patient's critical datasets necessary for survival. Such datasets are content and context sensitive in nature; therefore dissemination of datasets with the specific threshold values need an efficient aggregation and routing scheme. This occurs when the information of interest is pushed or pulled from the heterogeneous data-centric nodes.

Existing data aggregation and routing schemes [1], [2], [3], [4], [5] provide a solution to the data loss and in-network latency problems up to some extent. However, they do not address problems involved with monotonous queries that keep on pinging the data sources for the critical updates.

In [6], a continuous query (CQ) based approach was proposed for information monitoring of content-sensitive datasets. This was done by ignoring the problem that a failed CQ can easily overload network with unnecessary requests if a node is recovering from a previous crash. Secondly, a crashed node will not only affect the routing of a single dataset but will also compromise the quality of information monitored and disseminated at a particular time. In other words data loss will be much higher, if the information is not cached properly at the sink node close to the crashed location. Thus a fault-tolerant solution is required for the computational and data intensive tasks. This will not only minimize the data loss problem, but it will also significantly improve the information monitoring and dissemination of critical patient information to the relevant HP. The purpose is to ensure that datasets are properly stored and the information of interest is efficiently disseminated.

This paper deals with the in-network monitoring of information and proposes a new Grid-based Architecture and data aggregation scheme for the efficient dissemination of information. The approach will minimize the data loss problem by efficiently aggregating information of interest from the neighbour node and communicating it to the base station (GSG Grid).

The proposed approach further extends the data aggregation and routing scheme mentioned in [3] that uses Ad hoc routing protocol, where peer nodes can communicate with other nodes soon after the update is available. Node with high energy consumption will be used as an aggregator. Our approach provides fault-tolerance by introducing a parallel aggregator to minimize network overheads by sharing workload of the peer aggregator.

A new routing scheme HDARS is proposed, which is discussed in section 3. This particular scheme typically deals with datasets aggregated at the sensor node before transferring it to the GSG gateway (sink node).

In TAG [4] queries, during the distribution phase aggregate queries are pushed down into the network and queries can be answered by traversing the node in bottom-up manner. This particular approach has a significant drawback, datasets are aggregated using AVG(), MAX(), MIN(), Group By etc. This approach is not only computational intensive but will also increase the network latency.

The proposed methodology for aggregating dataset has the following fundamental differences.

1. TAG approach has been extended to support query processing in the Grid environment.
2. Processing of monotone queries will be done by materializing results at the Grid node for efficient dissemination of information.
3. Monotone queries will only be routed to the node close to the base station to minimize energy consumption. Proposed in-network processing and grouping approach has been discussed in section 3, where queries can be answered by traversing the DAG for a better throughput.

In synopsis diffusion [4], a grouping scheme is proposed to address the problem of data loss, but this ignored the fact that in-network monitoring of information is not only data-intensive but computational and resource intensive as well. The proposed architecture in [4] also does not provide a solution for queries with large joins and is not suitable for the grid environment, where datasets can be aggregated from a huge collection of nodes and computational intensive tasks are disseminated to the remote nodes for managing the trade-off between network and computation.

The T-DIFF approach proposed in this paper looks at caching and also looks at how the cached query plans will be transferred from the local machine to the remote grid machines. It further explores the quality of retrieved datasets that can be used for data mining purposes, which is crucial for the survival of patients. The proposed approach also addresses problems with existing health sensor devices, where the datasets of interests are only routed when a threshold value becomes true. E-g: A hypertension patient might be already at a critical stage because his BP is at 199.00 but the hard threshold parameter in the sensor device was 200.00 and no alert was sent to the HP. To overcome such problems, the proposed methodology will efficiently minimize data loss and battery energy consumption through a grid based framework for efficient monitoring of patient health. Detail about the proposed data aggregation scheme is discussed in section 3. Section 2 gives an overview of the problem description, while section 3.1, 3.2, and 3.3 discuss the grid-based architecture for data aggregation and dissemination of data streams in Healthcare sensor networks.

Sections 3.1 to 3.3 also propose a new approach for the parallel execution of queries with joins and aggregates by extending the TAG model [4] for processing of computational intensive tasks in the Grid environment. Section 3.2 describes the proposed approach for grouping such queries and discusses the T-Diff algorithm for materializing results of the similar query plans, which might be used for answering similar monotonous queries. This can significantly affect the network bandwidth and energy consumption in interactive wireless sensor networks. Section 3.3 gives a detailed analysis of an enhanced information monitoring model for efficiently storing the materialized results at the base station close to the sensor node. To the best of our knowledge, grid-based monitoring of patients' critical datasets in the wireless sensor networks has never been discussed properly in this context of caching and materializing results of monotonic queries with joins that are computational and data intensive. Section 5 discusses related work, section 4 discusses the experimental results and section 6 discusses conclusions and future work. Section 7 gives the list of references used in this paper.

2. PROBLEM DESCRIPTION

The problem with existing in-network monitoring techniques is that data aggregation schemes give high priority to those sensor nodes where the signals are strong. Routing of information only focuses on those source nodes from where the information of interest can be transferred to the sink (base) using the shortest path algorithms for routing and dissemination of information.

Some of these approaches use the classic flooding technique, where the information will be sent to the sink node as soon as the update is received. Such an approach is widely criticized because of the limited battery problem.

Similarly the ad hoc routing protocol and perfect aggregation scheme for data aggregation looks at the sensor node with a strong signal, which will act as an aggregator or coordinator for sending data to the sink node. The issues that are widely ignored by the existing solutions are the following:

i) The crucial information coming from sensor devices can be lost or the quality of the information can be easily compromised, if the datasets are not stored and disseminated efficiently and effectively.

ii) This approach will significantly increase the network latency if the critical information is not communicated in a timely manner because of battery problems.

iii) Data processing from a variety of sensor devices is also arguable, depending on the topology (star or mesh) and/ or the application domain. For example, a mesh topology is used for a dynamic environment that also depends on the mobility of an object, and the processing of computational and data intensive aggregates is a challenge that can easily compromise the quality of information. This is widely ignored by the existing approaches such as [1-7].

It is also important to note here that the existing data aggregation and dissemination techniques have also ignored the fact that the data gathered from the sensor nodes has a huge amount of similarity. Also, routing of queries to the same data source can easily affect the battery energy consumption and ultimately result in the degradation of the network. Secondly, it will end up with significant data loss because existing solutions do not provide fault-tolerance, which is crucial for a better throughput. A better solution is required to efficiently deal with the data loss problem and monitoring of content sensitive datasets in sensor networks.

3. PROPOSED HYBRID AGGREGATION AND ROUTING SCHEME_ HDARS

The proposed hybrid data aggregation scheme is shown in Figure 1, and the purpose of this scheme is to ensure that datasets generated from the sensor nodes are efficiently aggregated before transmitting to the Grid-Sensor gateway (GSG) or sink node.

The proposed approach extends the Ad-hoc routing protocol and TAG (Tiny Aggregation) scheme mentioned in [3], [4] for in-network processing of datasets and also provides a fault-tolerant aggregation scheme by introducing a parallel aggregator. The purpose of the additional aggregator is to provide fault-tolerance, in case of network congestion, load balancing or if something goes wrong with the first aggregator. The additional aggregator will be automatically activated and start forwarding datasets to the sink node.

The proposed Hybrid Double Ad hoc routing scheme (HDARS) is different in several respects such as:

1. It provides fault-tolerant aggregation by introducing an additional aggregator which will share the workload of the peer aggregator node in the event of a crash. It will also monitor the energy consumption of that node to reduce latency and network overheads, while communicating and disseminating information to the sink.
2. The proposed approach is also different from the aggregation schemes mentioned in [3], [4], [5], [6] for in-network processing of sensor data. Our approach extends the functionality by not only minimizing energy consumption; it also efficiently processes the computational, data, and resource intensive tasks by disseminating information to the Grid Sensor Gateway (GSG).
3. Once the datasets are transferred to the GSG Grid node, datasets will be cached temporarily to answer the monotone queries. In the Healthcare domain, analysing the patient's condition is crucial at several stages. If same query is routed over and over again, it will significantly increase the network latency and battery consumption of the sensor attached to the patient. Therefore most of the monotone queries will be answered by traversing nodes that are temporarily materialized, but in the proposed scheme the aggregator will rout the updated information soon after it is available.
4. The scheme provides fault-tolerance by routing only those queries where update is crucial and necessary to be communicated to the relevant HP. The existing routing schemes have widely ignored that fault-tolerance is crucial for the in-network processing of datasets. That's why a two step aggregation is proposed: i) Aggregation of datasets before in-network processing of sensor data. ii) Efficient processing of monotone queries at the Grid node GSG for minimizing network overheads.

The proposed scheme shown below in Figure 1 uses double aggregators; the nodes with strong signals will be used as aggregators. Datasets of interest will be routed to the peer node, and each node is capable of communicating with the other nodes that sent datasets to the aggregator node. The aggregators finally disseminate relevant updates to the GSG (Grid Sensor Gateway) for in-network processing of critical updates.

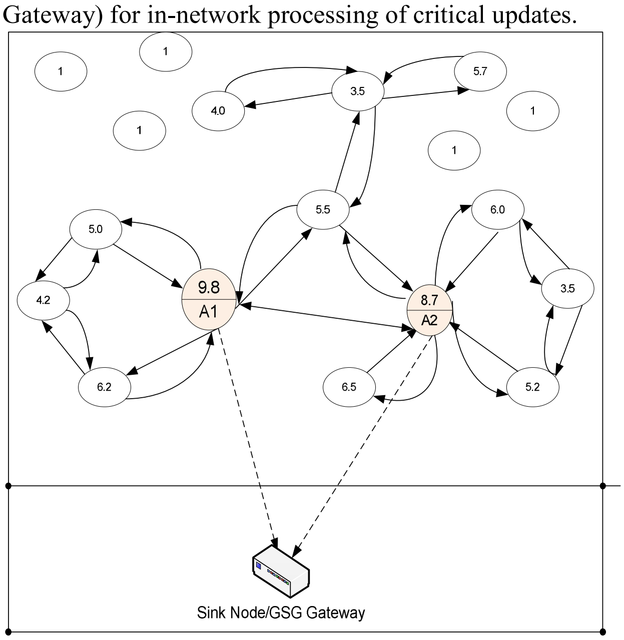


Figure 1. In-network hybrid Aggregation of Sensitive datasets

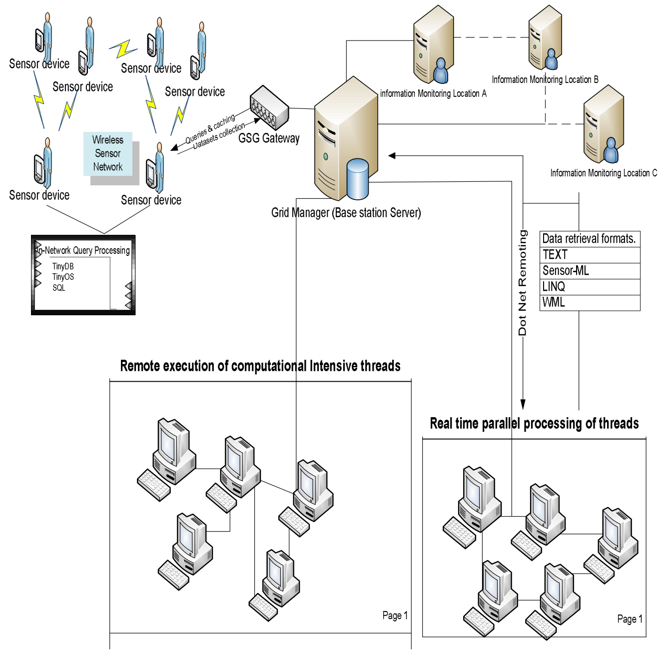


Figure 2. Grid-based Information Monitoring Architecture

3.1. Grid based Architecture to Support HDARS Routing Scheme

The proposed Grid based monitoring of patients' sensitive datasets focuses on reducing data loss in the sensor networks by looking at how to efficiently aggregate datasets in order to reduce the network latency. The approaches mentioned in [12] and [13] are extended so that the quality of information cannot be compromised at any stage. In [12], in-network monitoring is proposed by exploiting commonalities among them. This leads to a cost-effective caching scheme for dissemination of critical events to the sink or base station so that trade-offs between the message and energy consumption of sensor nodes can be reduced.

Patient health monitoring is an important area of research, especially when the healthcare sensor device is attached to a patient's body for monitoring of particular threshold events critical for the survival of the patient. Typical examples are monitoring patients with high blood pressure, heart problems, and other chronic diseases. The patient can die if the sensitive information is not communicated to the sensor networks in a timely manner.

The problem domain is divided into two categories i) patients who are admitted in the hospital and ii) patients who are at home but the sensor devices are attached to them for dissemination of timely information.

A Grid-based framework is proposed to address the caching problem by materializing similar query results in order to improve the monitoring of critical patient data. A Grid-based approach further improves the tiny aggregation (TAG) scheme in two ways. First, it extends the model that exploits commonalities among datasets, as a result of aggregate or group-by functions. Such datasets constitute a large portion of datasets which are scattered across the distributed networks. Second, a materialized caching approach is proposed, which has some similarity with the approach discussed in [9]. This approach was widely used for query optimization but has never been used for addressing the problem of aggregation in wireless sensor networks.

The cached query results and plans will be merged, and a global execution plan will be generated to form a DAG (Directed Acyclic Graph). This will be used for answering only those monotonous queries that can compromise the quality of datasets and increase the network latency. TAG queries are similar to SQL syntax but have a fundamental difference in returning datasets of interest. Normal SQL queries return a value while using aggregate functions but TAG queries return a stream of values, which needs to be managed and stored properly.

In the healthcare domain, the main goal is to retrieve those aggregates that satisfy a range of parameters associated with a particular threshold that are important for monitoring sensitive data. Some of the existing approaches [10], [11] have proposed Continuous Query (CQ) based solutions for the update monitoring but did not address the issue that CQ are content sensitive and time dependent and can easily flood the network with similar queries. This will also increase the network latency.

The proposed architecture for in-network monitoring and dissemination of information in the GSG Grid network is shown below in Figure 2. Components of this particular architecture are explained in the next section.

3.2. Grid-based Information Monitoring Architecture for Healthcare Sensor Networks

The Grid based Architecture for the retrieval and dissemination of datasets is shown in Figure.2. It provides a novel data aggregation and caching solution for duplicate in-sensitive queries. Datasets generated from the sensor devices attached to the patients will be transferred to the GSG gateway. Query plans will be cached temporarily at the local Grid node for answering monotonous queries by traversing the DAG in a bottom-up fashion.

The problem with similar queries with joins is that they are computational and resource intensive. Sensor devices have limited energy consumption, limited I/O, processing and memory problems. Therefore a better framework and aggregation scheme is required for the efficient retrieval and dissemination of information. Details of the components in Figure 2 are discussed below and the data aggregation scheme for transferring datasets from the sensor nodes to the Grid sensor gateway (GSG) will be discussed in section 3.3.

3.3. Wireless Healthcare Sensor Networks

The Grid-based wireless sensor network in Figure 2 shows wearable sensor devices attached to humans. The radio signals transmitted from such devices need to be properly stored and monitored for patients with serious medical conditions.

Typically, only blood pressure (BP) and heartbeat of patients will be monitored using those sensor devices. For the in-network monitoring of such queries TinyOS and TinyDB are used for aggregation and routing of information to the base station or sink node. Heterogeneity in sensor devices also poses research challenges; therefore the proposed model supports XML, SensorML and SQL type queries for the aggregation and dissemination of data streams.

Query processing strategies in such systems are very similar to the group optimization scheme, where the queries will be pushed while the information is stored at a certain depth in a query tree. The dissemination of datasets uses the pull approach for constructing an answer to a particular query by transferring all information to the root node; each node in a tree can communicate with neighbouring nodes and the sensor node with a strong signal will be used to aggregate datasets of interest. Datasets will finally be transferred to the sink for further processing and dissemination of datasets.

3.4. Grid Manager

Minimizing data loss and energy consumption is a challenge in sensor networks. The existing approaches [1], [2], [3] and [4] did not address the problem of queries with large joins, which are not only data intensive and data sensitive, but are computationally intensive as well.

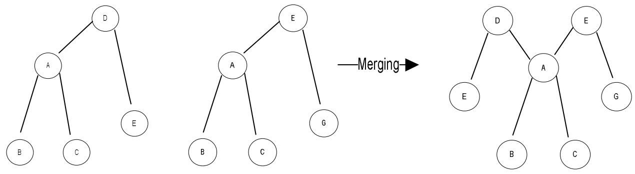
The main role of the Grid Manager (GM) is to execute computationally intensive tasks on the remote machines in order to minimize delays in processing results. It will also act as an interface, where the queries will be sent to the sensors and query plans with large joins will be cached for future reuse. The main idea of introducing such a scheme is to minimize the battery and energy consumption of sensor nodes. Web services based periodic pinging of data sources will be used. The idea is to get the information from the other grid nodes by analysing the previous history of the patient, until the node will be restarted or activated.

The proposed architecture provides a cost effective strategy for minimizing data loss. It also minimizes the routing of similar queries to the sensor node so that energy consumption and network latency can be improved. This leads to efficient retrieval and dissemination of the relevant datasets to the HP, which can then be used to make an informed decision about such patients.

3.5. Remote Executors

Remote executors are responsible for the processing of CPU and data intensive tasks. It is also responsible for disseminating information to the relevant HP connected to the Grid machine.

3.6. Proposed In-network Grouping Approach for Monotone Queries



BP: 120, 140, 200
Node A, B, C
Heartbeat: 60, 80
Node: D, E
Time: t1, t2, t3, t4

BP: 120, 140, 220
Node A, B, D
Heartbeat: 60, 100
Node: E, G
Time: t1, t2, t3, t5

BP: 120, 140, 200,220
Node A, B, C, D
Heartbeat: 60, 80, 100
Node: D, E, G
Time: t1, t2, t3, t4, t5

Figure 3. Proposed merging scheme for queries with joins

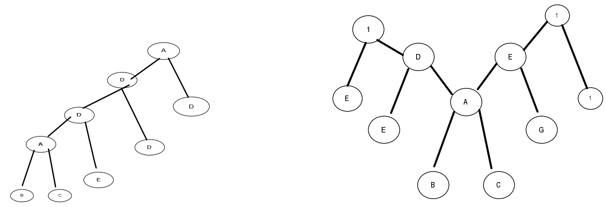


Figure 4. Canonical left-deep tree Aggregated DAG with Materialized results

The proposed grouping technique has some similarity with the approaches mentioned in [4], [7]. In [4] the idea was to come up with a duplicate sensitive plan and minimize energy consumption by grouping the left deep canonical plans. In [7], the approach minimizes the cost of the query plan by exploiting commonalities among the sub-queries so that query optimization cost can be reduced. This particular approach was used for query optimization but was never used for query processing in GSG wireless sensor networks.

In Figure 3 and Figure 4, two simple query plans are merged together to come up with a global execution plan, which forms a DAG. In the diagram TAG queries are used, where the leaf nodes contain values and each node can communicate with the neighbouring node. Traversing the tree in a bottom-up fashion will construct an answer to the query, which will be disseminated through the root. In the diagram, two queries with the left-deep joins are shown with values A, B, C, D, E, and G that contain datasets of patients with high BP and the heartbeat of patients with serious heart problems. BP and heartbeat of two patients are monitored.

It is also important to note that such query plans will be updated with regard to changes in the source information. Such query plans have a lot of commonalities among them, which can easily be exploited. Because of the limited memory, processor and energy consumption of sensor nodes, routing of similar queries will significantly affect the energy and in-network bandwidth.

Therefore, to avoid such problems in the health sensor networks, a grouping approach is proposed to answer the monotonous queries so that information of interest can be monitored and disseminated to the relevant HP in a timely manner. For efficient processing of such plans, queries will be processed and cached at the node close to the base station and computationally intensive and data sensitive tasks will be executed on the remote machines for a better throughput.

Figure 3 shows a left-deep canonical plan, where the query can be answered by traversing a tree in the upward direction. A TAG query example is shown below for the proposed information monitoring of above mentioned patient/s, which are very similar to the SQL type queries.

Query-1:
SELECT COUNT (*)
FROM sensors AS s, patient-Info as p
WHERE p.nodeid = s.nodeid
AND BP < 200 AND heartbeat < 100
SAMPLE PERIOD 10s
Query-2:
SELECT COUNT (*)
FROM sensors AS s, patient-Info as p
WHERE p.nodeid = s.nodeid
AND BP1 > 190 AND BP2 > 210 and heartbeat > 120
SAMPLE PERIOD 20s
Query-3:
SELECT MAX (BP) FROM snsor s, patient-Info as p, doctor_info as d
WHERE p.nodid = s.nodeid
AND p.pid = d.doc_id
And BP1 > 190 and BP2 > 230 and hearbeat >110
AND BP3 > 240 AND BP4 > 260 and heartbeat > 140
Generate_alert()
SAMPLE PERIOD 40s

```
T-Diff Algorithm
Function check-if-query Plan exist
If (sr (sensor reading) is True){
Then broadcast query q to GN (nearest grid node with results)
}else
SR = GN
Participate = true
Generate the DAG (directed acyclic graph)
Materialized the node N
}perform in-network aggregation
Traverse the node N in bottom-up manner
Disseminate result R
Check if the query is monotone
Broadcast result to the Query Q
}
```

3.3 ENHANCED INFORMATION MONITORING AND INFORMATION DISSEMMINATION

Data dissemination and gathering in sensor networks poses significant challenges in the distributed environment. This includes how the data is collected by the sensor nodes and what event needs to be triggered to monitor information of interest. These two things are closely related to understanding the query semantics that were sent to the sensor nodes and gathering of such data that has triggered a particular event.

As discussed above, the existing data gathering and dissemination techniques [1], [2], [3], [5], [6] did not address the need for an architecture that supports parallel execution of computational and data intensive tasks for the queries with large joins. This is required so that computationally intensive tasks can be processed remotely and the aggregated result will be transferred to the relevant node responsible for executing that query.

Our proposed enhanced information monitoring technique is shown in Figure 4. Basic details of the proposed architecture and grouping of queries with joins have been explored in the previous sections, whereas the general overview has been discussed in Figure 1. This section provides in depth analysis of caching query plans on the sink node close to the sensor network.

In the Healthcare sensor networks dissemination, gathering and routing of critical events to the base station or sink are crucial for the survival of patient. To improve fault-tolerance, AJAX (Asynchronous JavaScript and XML) is used with web-services to minimize the partial page refresh problem as well as periodic pinging of the remote data sources.

In the diagram shown in Figure 4 dot-Net remoting is also used for the execution of tasks on the remote machines. Tasks that involve large joins will be parallelised and distributed for execution; results will be aggregated and disseminated to the Grid Manager for the final outcome. Results of those query plans will be merged to come up with a global execution plan by exploiting the common sub-queries that might be used for answering a particular query.

Query plans that are generated will be materialized at the node close to a sensor node temporarily. The purpose of this is to reduce latency and improve dissemination of relevant events. For simplicity, only blood pressure and heartbeat of patients will be monitored and time series datasets are generated to monitor events of interest.

The benefit of using the proposed architecture is that monotonous queries will be routed to the sink before sending it to the sensor node for improving the network latency and better throughput. Sensor devices have heterogeneity, battery and energy consumption problems; therefore monotone queries will be answered by traversing the DAG shown in Figure 2 and Figure 3.

Datasets generated from the sensor nodes are in the form of a stream, therefore using the Grid based approach [9] will provide significant improvement in terms of dealing with heterogeneity, security, and interoperability and scalability and resource discovery problems in such networks. Another reason for using the Grid-based approach is to efficiently handle datasets of interest from variety of data sources such as XML, SQL type and LINQ (Language Integrated Network Query) for the processing of sensitive datasets.

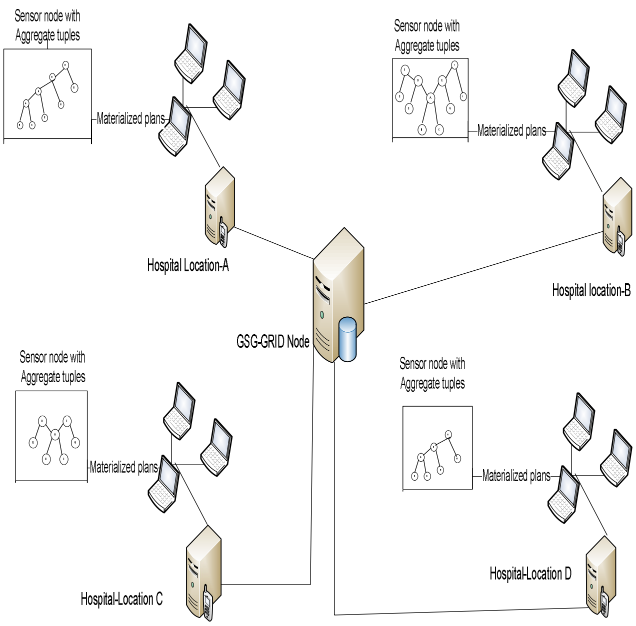


Figure 5. Enhanced Information Monitoring by materializing query plans

4. EXPERIMENTAL ANALYSIS

Experimental results of monotonous queries with joins were performed on running a Grid machine with duo core processor, 4GB RAM and 500GB hard drive. Dot-NET remoting was used for the execution of threads on the remote machines. Alchemi [10] and MyGrid [11] were used to setup the Grid environment and MATLAB parallel computing version was used for the testing performance of proposed T-Diff approach mentioned above. Two types of queries were routed to the sink node, one with grouping approach and the other without a grouping approach.

Queries without grouping were routed and processed without Grid, whereas the grouped queries were executed on the Grid nodes responsible for the remote execution of threads. Figure 5 results clearly show that the grouped approach has outperformed the non-grouped in terms of retrieving results and timely monitoring of sensitive information.

The only thing that needs to be taken into account is the cost of generating the global execution plan once the query will be saved at the sink node. We have also noticed that the proposed approach for communicating query plans results to the nearest sink has also improved the network latency by providing an aggregation scheme T-diff to deal with large joins monotonous queries.

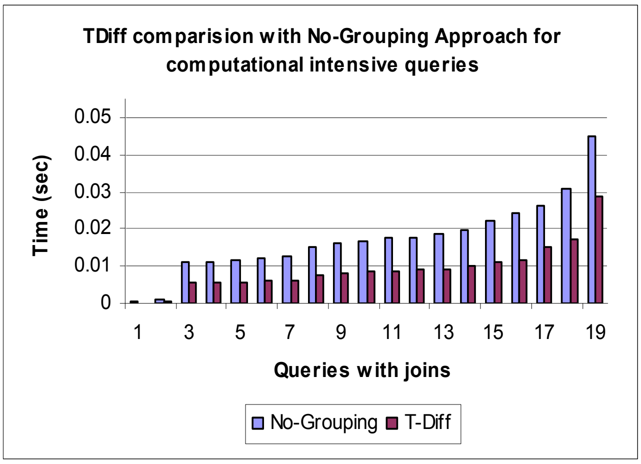


Figure 6. Computational Intensive Join Queries with Grouping and No- Grouping in the Grid environment

In Figure 7, comparison of the proposed hybrid HDARS scheme is performed with AODV (ad-hoc routing scheme) and No Aggregation approaches. It clearly shows that the HDARS approach has outperformed the existing methodologies mentioned above in terms of network latency and bandwidth. From the graph we can clearly see that the No-aggregation approach is not only affecting the battery energy consumption but will also increase the network latency. Although the AODV scheme has performed better as compared to the No-aggregation scheme, it still suffers from the network overheads and latency problem.

The network overhead and latency can be calculated using the similar mathematical model described in the [3], which is shown below.

*Latency (l_HDARS) = min (In-Network, Grid) + S (Overhead) * M*

l = Network latency

M = mobility from low to high

L_HDARS = latency of hybrid double aggregation routing scheme.

S = switching from one scheme to another.

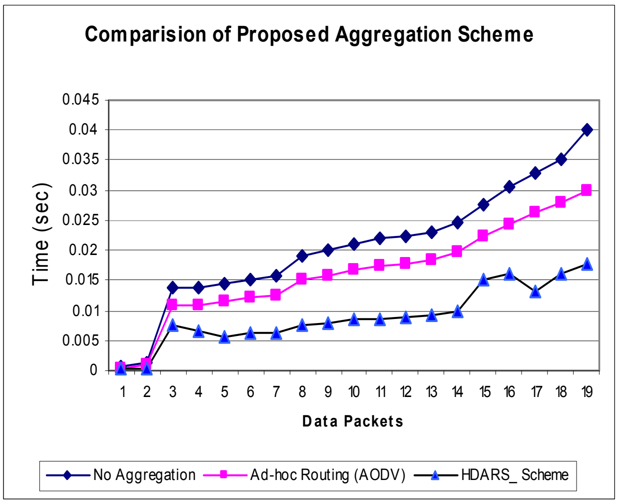


Figure 7. Comparison of proposed Hybrid Aggregation Scheme

Figure 8 shows the workload distribution of the main Grid machine. It shows the number of threads executed on the remote machine for processing computational intensive, data sensitive and resource intensive queries with joins and aggregate queries. The graph clearly shows the trade-off between threads and machine power consumption, if more threads will be running more power will be consumed. But in the Grid environment the idea is to use the power of those CPU's when they are idle to perform the computational intensive tasks.

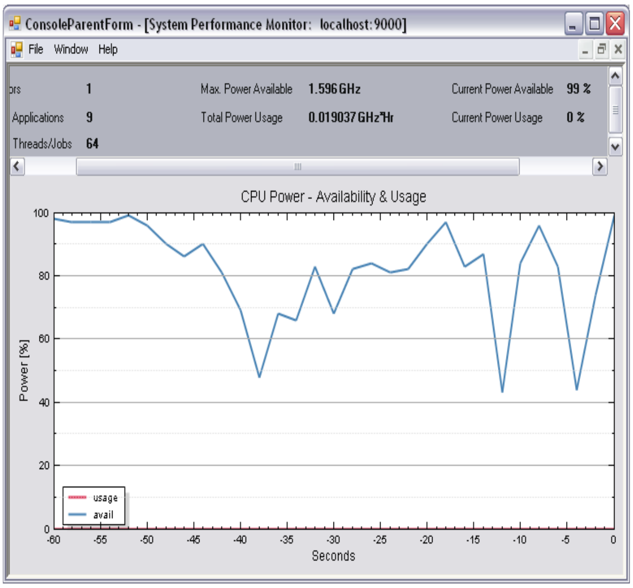


Figure 8. Execution of threads and computational Intensive Tasks on local Remote Grid machines

5. RELATED WORK

Information monitoring of content-sensitive datasets in the healthcare domain while retrieving datasets of interest from wireless sensor networks (WSN) is a challenge. Therefore constantly pinging such data sources will significantly affect network bandwidth by overloading the network with un-necessary requests.

Healthcare GSG is a data and computationally intensive environment, while monitoring datasets of interest for patients with serious health conditions. Device failure can cost the loss of life; therefore dissemination of timely information is crucial for monitoring of such data sources.

The existing in-network aggregation techniques [1], [2], [3], [4], [5], [6], [7], [8], [11] for wireless sensor network, discuss the need for a better routing and aggregation scheme for dissemination of datasets of interests. It also discusses about routing protocols such as LEACH and PEGASIS (Power –Efficient gathering for WSN). In PEGASIS dissemination of datasets to the sensor node will be communicated through the chain to propagate messages to the sink node. The overall idea is to reduce the network overhead by using shortest path and greedy algorithms. In directed diffusion [1] and synopsis diffusion, data sensitive approach looks at the duplicate insensitive and duplicate sensitive approach to reduce overheads by avoiding monotone queries to the sensor nodes.

Most of these approaches use a flooding method for broadcasting messages to the neighbouring node. This particular approach has many disadvantages while routing information of interest to the sink node. The same message can be sent over and over again to the neighbouring node, which is called implosion. This will not only significantly increase the network latency, but it will also affect on the available energy at the sensor node.

In [5], the Tiny aggregation approach is discussed for the aggregation and routing of information to the sink node. Up to some extent this particular approach supports duplicate sensitive transmission. Values will be pushed to the nodes in a tree. It also supports SQL type query for the dissemination and aggregation to datasets. Once the query is constructed on the root node, the tree will be traversed in a bottom-up approach to answer a particular query.

Grid-based information monitoring and processing of computational and resource intensive tasks is discussed in [12], [13]. Grid computing is an emerging area of research, it provides an opportunity to dynamically discover, process and execute tasks that are computational, data and resource intensive.

The idea is to utilize the resources of millions of computers. When the CPU is idle, computationally intensive tasks will be distributed to the remote machines. It is important that only those tasks that can be parallelised need to be distributed to reduce the query execution plan generation cost. Finally, results will be aggregated and disseminated to the relevant node.

6. CONCLUSION AND FUTURE WORK

A Grid based approach is proposed for the efficient monitoring of patient information in the Healthcare sensor networks. A new Hybrid double aggregation routing scheme (HDARS) is proposed for routing datasets of interest to the sink node GSG. A comparison of algorithms has been shown in the experimental section, and this clearly shows that the HDARS scheme has outperformed the ad-hoc routing and hybrid schemes by significantly reducing the network overheads and latency for a better throughput.

The proposed approach has addressed the problem of information monitoring, and it proposes a two step in-network processing of datasets. First the datasets will be aggregated by the first aggregator, and the second aggregator will be activated in case of a problem with the first. This provides fault-tolerance if the aggregator is overloaded with requests. A grouping scheme was also proposed to address the problem of monotone queries, where a query can be answered from the cached results by traversing the DAG in a bottom-up manner.

The proposed methodology discussed in this paper provides a Grid based architecture for efficiently routing and disseminating information. It also provides a novel approach to deal with the monotone queries to minimize the network communication with the sensor node to minimize energy consumption and processing cost.

Our future work in progress will further explore efficient ways of reducing network latency and cost of generating the global query execution plan, which needs to be tested at much broader scale.

7. ACKNOWLEDGEMENTS

I would like to thanks Dr. Elizabeth Kendall and Dr. Savitri Bevinakoppa for their valuable input on research issues discussed in this paper.

8. REFERENCES

[1] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, and F. Silva, “Directed diffusion for wireless sensor networking,” IEEE/ACM Trans.Networking, vol. 11, no. 1, pp. 2–16, Feb. 2002.

[2] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, “An application-specific protocol architecture for wireless microsensor networks,” IEEE Trans. Wireless Commun., vol. 1, no. 4, pp. 660–670, Oct. 2002.

[3] S. Lindsey, C. Raghavendra, and K. M. Sivalingam, “Data Gathering Algorithms in Sensor Networks Using Energy Metrics,” IEEE Trans.Parallel Distributed. System., vol. 13, no. 9, pp. 924–935, Sep. 2002.

[4] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, “TAG: A Tiny AGgregation Service for Ad-Hoc Sensor Networks,” in OSDI 2002, Boston, MA, US, Dec. 2002.

[5] Y. Yao and J. Gehrke, “Query processing for Sensor networks,” in ACM CIDR 2003, Asilomar, CA, US, Jan. 2003.

[6] Y. Xu, J. Heidemann, and D. Estrin, “Geographic-Informed Energy Conservation for Ad Hoc Routing,” in ACM/SIGMOBILE MobiCom 2001, Rome, Italy, Jul. 2001.

[7] S. Nath, P. B. Gibbons, Z. R. Anderson, and S. Seshan, “Synopsis Diffusion for Robust Aggregation in Sensor Networks,” in ACM SenSys 2004, Baltimore, MD, US, Nov. 2004.

[8] A. Manjhi, S. Nath, and P. B. Gibbons, “Tributaries and Deltas: Efficient and Robust Aggregation in Sensor Network Stream,” in ACM SIGMOD 2005, Baltimore, MD, US, Jun. 2005.

[9] I. Solis and K. Obraczka, “The Impact of Timing in Data Aggregation for Sensor Networks,” in IEEE ICC 2004, Paris, France, Jun. 2004.

[10]F. Hu, C. Xiaojun, and C. May, “Optimized Scheduling for Data Aggregation in wireless sensor networks,” in IEEE ITCC 2005, Las Vegas, NV, US, Apr. 2005.

[11] X. Jianbo, Z. Siliang, and Q. Fengjiao, “A new In-network data aggregation technology of wireless sensor networks.” in IEEE SKG’06, Guilin, China, Nov. 2006.

[12] (2010) MyGrid website. [Online].Available: http://www.mygrid.org.au

[13] (2010) Alchemi website. [Online]. Available:http://sourceforge..net/projects/alchemi/