Sensor Networks And Aggregation Using Hybrid Protocol

Ashok.A, Harikrishnan.N, Thangavelu.V.
Bit Campus, Anna University, Trichy

Abstract

In this paper we introduced a new protocol called HYBRID PROTOCOL which overcomes the disadvantages of earlier methods like tree aggregation, gossip based aggregation and centralized processing. This hybrid protocol can be widely used in large scale applications of sensor networks like military applications, vehicle monitoring, traffic control and harsh environments such as volcanoes or hurricanes where the failure rate is high. The ability to efficiently aggregate information in large networks is crucial for successful employment of sensor networks. This paper addresses the problem of designing truly scalable protocols for computing aggregates in presence of faults, Protocols that can enable million nodes sensor networks to work efficient. We have evaluated the behavior of existing two aggregation methods: tree aggregation and gossip aggregation. The behavior of two protocols depends on the size of the network and probability of failure is high. So, the new hybrid protocol introduced is an optimal mix between the two protocols. The performance of the hybrid protocol is analyzed and it is better compared to other two protocols.

INTRODUCTION

Recent advances in computing technology have led to the production of a new class of computing device: wireless, battery powered, smart sensor. Information aggregation is a common operation in sensor networks. Traditionally, information sampled at the sensor nodes needs to be conveyed to a central base station for further processing, analysis, and visualization by the network users. Information aggregation in this context can refer to the computation of statistical means and moments, as well as other cumulative quantities that summarize the data obtained by the network. Such accumulation is important for data analysis and for obtaining a deeper understanding of the signal landscapes observed by the network.[2]The modern sensors can sense and measure odors, vibration, acceleration, pressure, temperature and other physical
environments around the world.[4]All this information is useful in gathering data about the surrounding world. Silicon scaling has also drastically reduced the cost of communication between the wire and wireless networks.[6]The small size and low cost of the components allow deployment of networks at very large number of such devices in environments that are out of reach of human beings. [11]We envision that sensor networks consisting of hundred/thousand to a million sensors will not be uncommon in these settings and will be dispensable.

The promise of sensor networks is to enable spatial and temporal data collection at a much higher rate than currently prevalent.[5]The various queries for information processing in sensor networks can be classified as point queries and aggregate queries. A point query usually involves routing the query from point of query injection in the sensor network to the sensor node of interest and then efficiently sends the results back.[7]Finding a set of values at a particular node is an example of point query. Aggregate queries are relatively more complicated as they involve all the sensor nodes for processing of information and hence these types of query are more challenging to execute in a fault prone environment.

AGGREGATION METHODS CURRENTLY USED

The main challenge in designing algorithms/protocols for distributed computation of aggregation in sensor networks is to keep the resource utilization to the minimum by reducing the communication among the nodes and computation performed. The most successful technique to significantly reduce the resource utilization for aggregation computation is in-network aggregation is to perform aggregation as the information is transmitted. Depending upon how the in-network aggregation is performed, it is classified into three main techniques. They are

- Centralized processing.
- Tree aggregation.
- Gossip based aggregation.

- Centralized processing: This involves transmitting the values collected at each sensor node directly to the centralized processing unit. So, aggregation can be easily performed .The main drawback in this method is the sheer amount of communication required since each node's information has to be sent into central station. This leads to very large processing time, hence this technique cannot be applied to crucial applications. This method leads to large power consumption since each node has to use a storing signal for the wireless transmission. It is lacking in terms of scalability and reliability since the coordinator can quickly become a bottleneck and is a single point of failure. The total transmission required is very high.

- Tree aggregation: The main idea depicted here is to organize all the sensor nodes into an aggregation tree. The root of the tree is the node where the query is injected and where the aggregation result is retrieved. The request is propagated from parent to children. The leaf nodes sent the value of the measurement to the parent. Intermediate nodes wait for values from the children, do a local aggregation of these values and its own measurement, and sent the aggregate to the parent. The root node presents the overall aggregate to the user. If one of the nodes fails, none of the descendants are reachable.

The expected number of reconstructions for a successful tree aggregation operation is $1/(1-p)^L$ *given that the number of links in the tree is L and the probability of failure is p.*

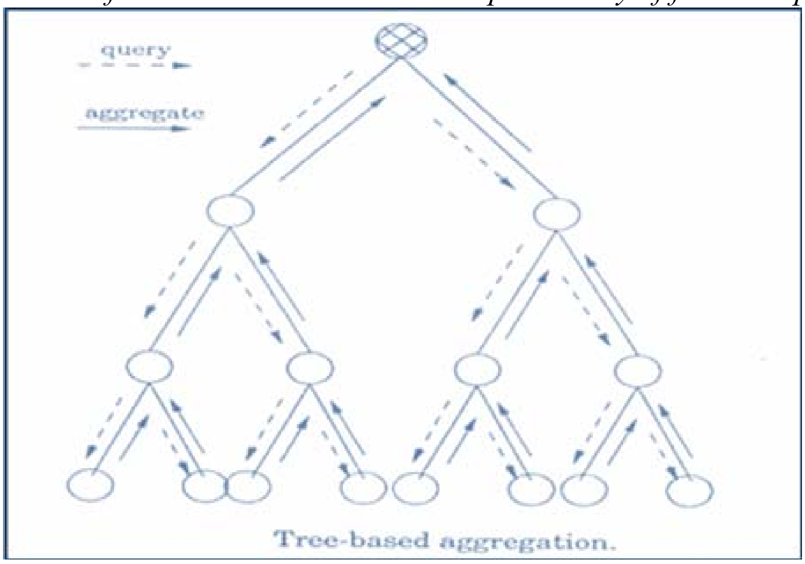


Figure 3

- Gossip based algorithm: Each node contacts its neighbors and exchanges information with these needed. In each round, each node chooses randomly a node in the network and sends this node half the weight and half the value of the aggregate. Information spread resembles the spread of an epidemic. Excels in the network where nodes can fail with significant probability. Immune to faults but slower.

A NEW AGGREGATION MODEL-HYBRID PROTOCOL

A hybrid aggregation could combine a tree and gossip architecture and hopefully, perform significantly better than any other method. Here we propose a hybrid protocol that partitions the

sensors into groups and uses gossip for aggregation within the groups and a tree for aggregation between the groups. The algorithms to determine right combination between tree and gossip aggregation is introduced. The efficiency of the protocol is analyzed.
Consider N - sensor network size.
d(N) - the amount of aggregation time in fault-free tree.

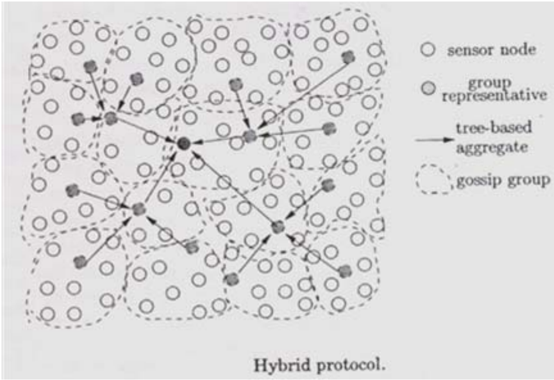


Figure 4

It depends on factors like topology, connections between sensors and tree construction algorithm used. To avoid faults in the network if the probability of failure is large (>>10/ d(N)N),we can use gossip aggregation. If it is large we can use for tree aggregation.

But hybrid protocol can be used for all the networks without considering the probability of failures in the network. The reduction in the number of nodes in the tree can be achieved by partitioning sensors groups, computing aggregate within each group using robust algorithm and tree to compute the aggregate over the groups. In hybrid protocol, first the nodes are organized, based on the local proximity into 'm' groups each of its average size(N/m).Within each group, gossip type aggregation is used to compute aggregate. At the time, a leader is selected for each group (any gossip based selection can be used).The group leaders organize into an aggregation tree and aggregate the group values using the tree aggregation protocol. If the group leader fails, any other node in the group can take over the role and participate in tree aggregation. Normally gossip based aggregation converge to the same value, groups can be reused between queries. So, there is no need to repartition the network for each query. This protocol has robust correctness.

CHOOSING THE GROUP SIZE FOR HYBRID PROTOCOL

A reasonable value for the size of the group is to choose N/m such that pd(N/m)N/m will be approximately equal to one. So, the tree aggregation part of the protocol needs only few restarts. A more principled approach is to formulate and solve an optimization problem that gives the optimal value for the number of groups given the characteristics of the sensor network. This approach can be used for determining the optimal group size in order to minimize the time of completion and the total number of messages transferred. d (N) is the product of maximum depth of the tree and the number of messages required for parent and child communications in tree aggregation. Let g(N) is the convergence time for the gossip protocol as a function of the size of the network. The total time for completion is the sum of the completion times for running gossip protocol within the group and the tree aggregation between the groups.

$$T = T_{gossip} + T_{tree}$$
$$= c1g\,(c2N/m) + d\,(m)\,(1 / (1-p)^{m\,d\,(m)})$$

In the above mentioned equation, the factor c_1>=1 is introduced to account for the fact that g(N) is the expected time and it is not the actual running time of the gossip protocol. Since, the actual run time of the algorithm might require more time. The factor c_2>=1 is introduced to account for the fact that the size of individual groups will be N/m only on average but individual groups could be much larger.

Gossip Speed Factor c_1: Here we propose to determine the appropriate value for the factor c_1 using an empirical approximation of the distribution method obtained by simulating the behavior of gossip on the given network.

Maximum Group Size Factor c_2: The variability due to uneven group size has to be based on the particular grouping algorithm used. Unless the grouping algorithm strives to achieve uniform groups, the grouping can be considered random and each group size modeled by a binomial variable –the group of all sizes form a multinomial distributions.

EMPRICAL EVALUATION OF THE HYBRID PROTOCOL

The optimal combination between the tree and gossip protocol is determined by formulating, based on the characteristics of the application and hence solving an optimization problem. The main goal of this paper is to present the empirical evidence that the hybrid protocol is always as the best compared to the tree and gossip protocols.

IMPLEMENTATION OF HYBRID PROTOCOL

In this section we present a detailed a description of the implementation that is used to produce the results. Now, consider a simple sensor network and for that network the results are produced. The results are qualitatively similar for different type of networks. Consider a sensor network constructed by randomly placing the nodes in a unit square field. For each node, the transmission radius was selected such that the number of direct neighbors was 32.The hybrid protocol that we introduced use gossip protocol for aggregation within group. Here, we use Push-Sum protocol for performing aggregation. Three variations of this protocol are

1) Nodes are allowed to talk only with their immediate neighbors.
2) Nodes are allowed to talk with any other node in the network.
3) Nodes are allowed to talk to a logarithmic number of nodes that are distributed randomly throughout the network.

Algorithm: (Push-Sum protocol):

1) Let { ($\hat{s}_r$, $\hat{w}_r$) } be all pairs sent to *i* in round *t-1*
2) Let $s_{t,i} := \Sigma_r \hat{s}_r$, $w_{t,i} := \Sigma_r \hat{w}_r$
3) Choose a target f_t (i) uniformly at random
4) Send the pair (½ $s_{t,i}$, ½ $w_{t,i}$) to f_t(i) and *i* (yourself)
5) $s_{t,i}$ / $w_{t,i}$ is the estimate of the average in step t

The Push-Sum protocol:

- For computing sums or averages of values at the nodes of a network.
- At all times t, each node i maintains a *sum* $s_{t,i}$, initialized to $s_{0,i} := x_i$, and a *weight* $w_{t,i}$, initialized to $w_{0,i} := 1$. At time 0, it sends the pair ($s_{0,i}$, $w_{0,i}$) to itself, and in each subsequent time step t, each node i follows the above Push-sum protocol.

The algorithm uses the basic property of mass conservation: the average of all sums $\boldsymbol{s_{t,i}}$ is always the correct average, and the sum of all weights $\boldsymbol{w_{t,i}}$ is always n

Our first contribution is a simple and natural protocol Push-Sum for computing sums or averages of values at the nodes of a network. Notice also that the lengths of all messages are bounded by the largest number of bits to encode the xi, plus the number of rounds that the protocol has run. If we are interested in computing the sum instead of the average, then we only need to apply a small change: instead of all nodes starting with weight w0;i = 1, only one node starts with weight 1, while all others start with weight 0. We than obtain exactly the same kind of approximation guarantees. Push-Sum is a very natural protocol, yet the proof of the approximation guarantee is non-trivial and relies crucially on a useful property the term *mass conservation*: the average of all sums st;i am always the correct average, and the sum of all weights wt;i am always n. The approximation stays bounded away from the true average.

Diffusion Speeds:

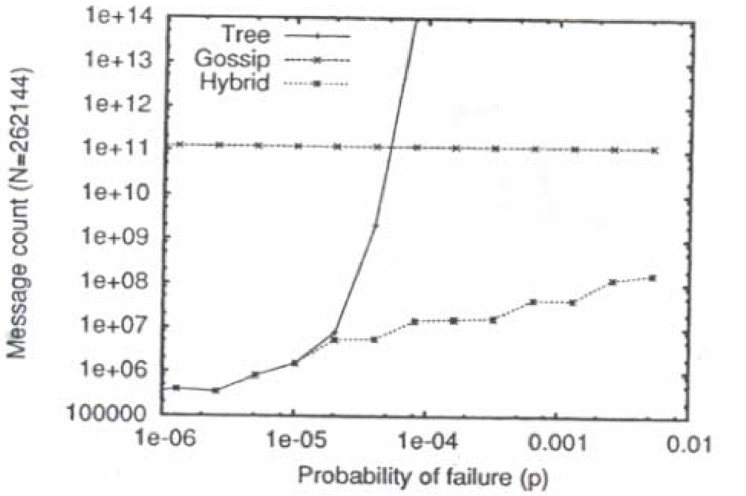
The analysis of Push-Sum builds on an understanding of the *diffusion speed* of Uniform Gossip, characterizing how fast a value originating with any one node diffuses through the network. The groups once formed, can be used for multiple instances of aggregation. In tree aggregation approach, aggregate value is formed by a spanning tree among the participating nodes by starting with the root nodes splitting the space into four regions, selecting a random node in each part to be the child of the node, and repeatedly until no nodes are left. This form of splitting will produce a balanced tree. Hence the spanning tree is formed over respective nodes from the gossip groups. As a result, a parent-child pair of nodes in the spanning tree might be in multiple hops apart from each other.

TYPES OF GOSSIP COMMUNICATION IN HYBRID PROTOCOL

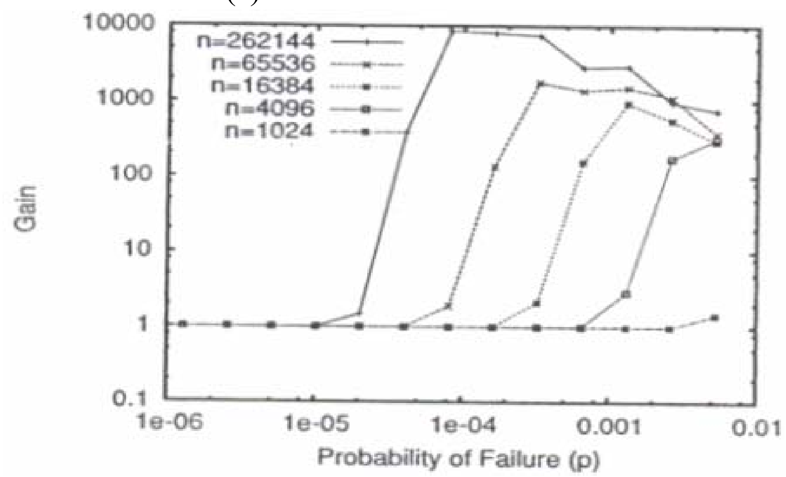
For all types of communication, three plots are used to compare the aggregation methods. They are

- The first plot is the absolute performance of the quantity that is optimized.
- The second plot is the relative gain in performance of the hybrid protocol over the tree and gossip aggregation. On y-axis, the ratio of the minimum of the tree and gossip value to the value of the hybrid protocol.(Higher the ratio, the better the relative performance)
- The third plot is the group size determined as the solution of the optimization problem of the hybrid protocol.

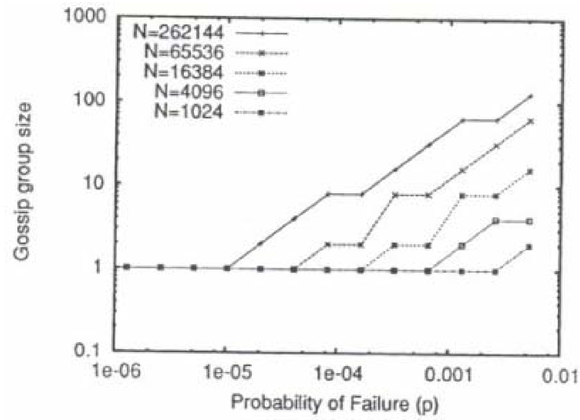
<u>Local Gossip Communication:</u> Here, nodes are allowed to communicate only with their immediate neighbors when running the gossip protocol-no long distance communication is allowed for gossiping but it is allowed in tree



(a) Absolute Performance

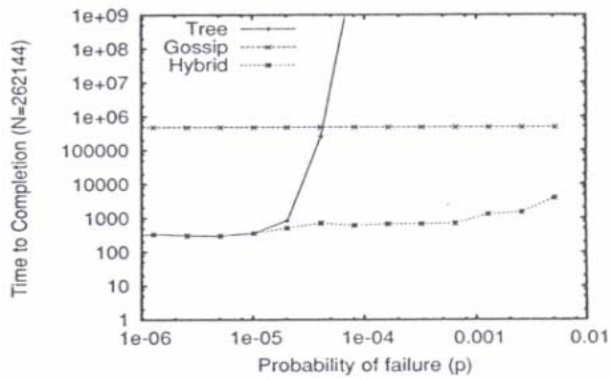


(b) Gain of hybrid protocol

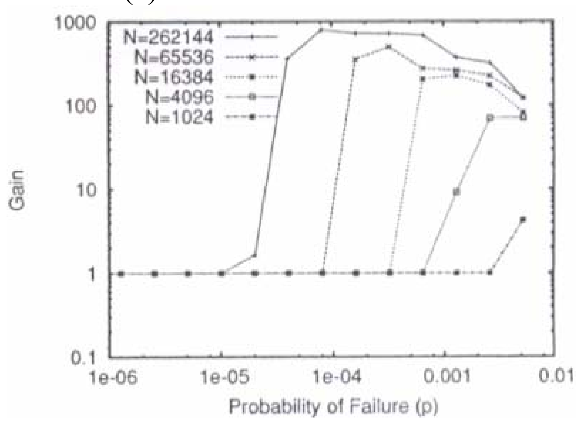


(c)Optimal gossip group size

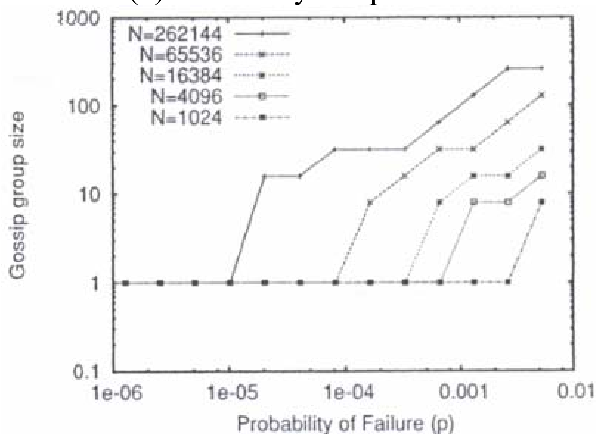
Figure 5: Comparison of message counts when using local gossip communication



(a) Absolute Performance



(b) Gain of hybrid protocol



(c)Optimal gossip group size

Figure 6: Comparison of time of completion using local gossip communication

From the figure 5(a), we observe that,

- For small probabilities of failure, the tree performance is five times greater than that of gossip aggregation.
- As the probabilities of failure increases, the performance of the tree deteriorates exponentially to the point that is worse than the rather performance of the gossip.
- The performance of gossip is mostly immune to failures, but not proved experimentally.
- The performance of the hybrid protocol coincides with the performance of the tree for small failure probabilities. Also, always, the hybrid protocol outperforms the gossip protocol.
- The dependency on the probability of failure of the hybrid protocol is almost linear.

The same is observed in the plot of the absolute values of the time of completion (figure 6(a))

From the figure 5(b) and 6(b), we determine that, except for small probability of failure, the hybrid protocol is more efficient both in terms of total number of messages and total time. For different network sizes, the shape of the curve of the gain is very similar. Hence, this means that the advantage of the hybrid protocol widens for large networks-its use is crucial instead of the pure protocols. From the figure 5(c) and 6(c), for small probabilities of failure, only the tree aggregation is used. The larger the network the sooner-the group size increases linearly with the probabilities of failure.

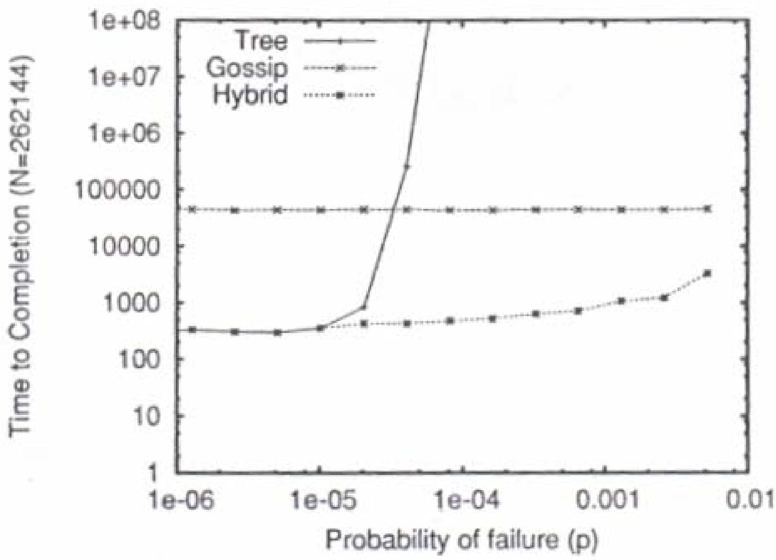
Global Gossip Communication: Nodes are allowed to communicate with any node in the network. The result of global gossip communication is similar to the local gossip communication with the following exceptions:

- The gap between the tree and the gossip communication for small probabilities of failures is smaller compared to the local gossip communication.
- Gains of the hybrid protocol is smaller

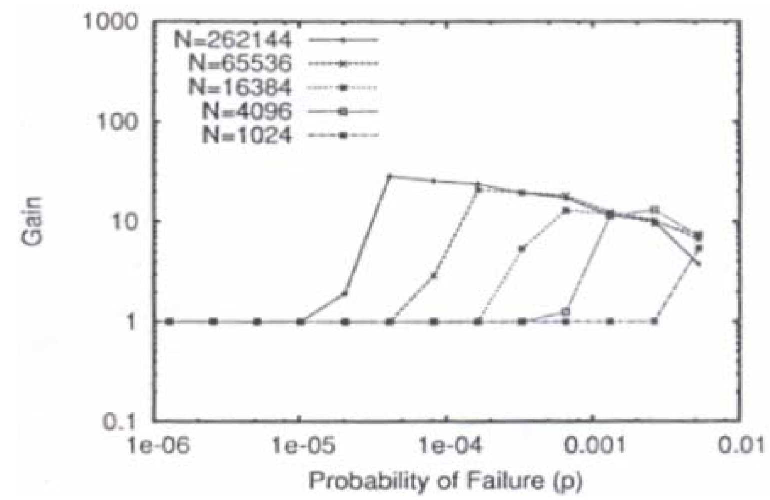
Limited Global Gossip Communication: Here, nodes can communicate with any other node in the network but fixes the number of such nodes for any specific node to the logarithm of the size of the network. In this way, knowledge about only a small number of nodes in the network is needed. It is more practical than global gossip. The performance is similar to the global gossip with the following exceptions:

- Gossip performance s about five times worse both in terms of the number of messages and the time to completion.

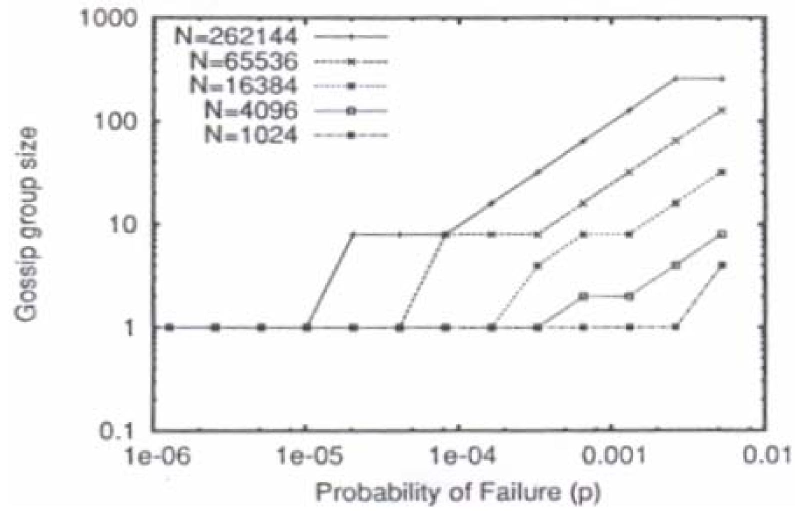
Hence, this is reflected in the performance of hybrid protocol. It is an acceptable compromise between performance and the need to keep every node informed about all the other nodes.



(a) Absolute Performance

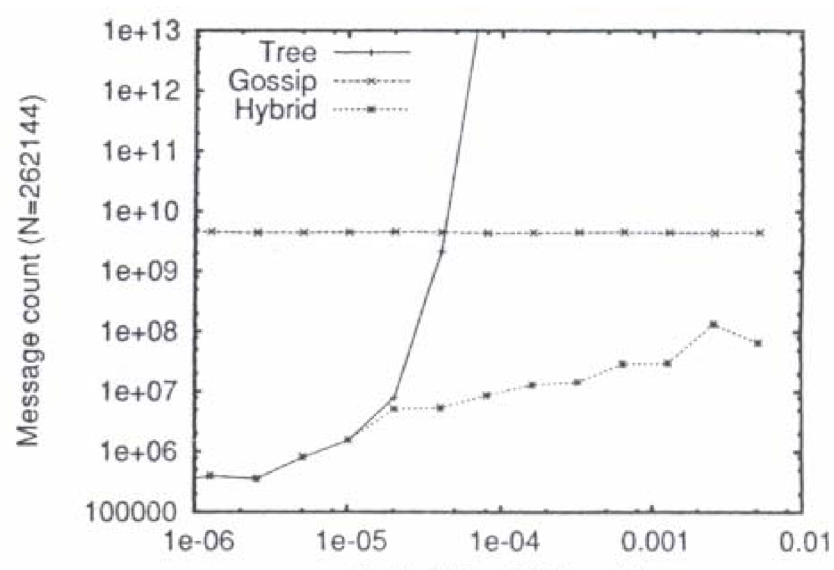


(b) Gain of hybrid protocol

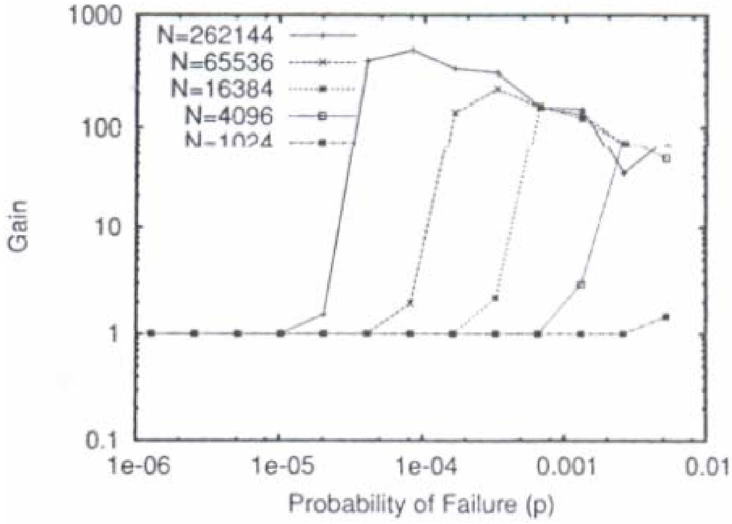


(c) Optimal gossip group size

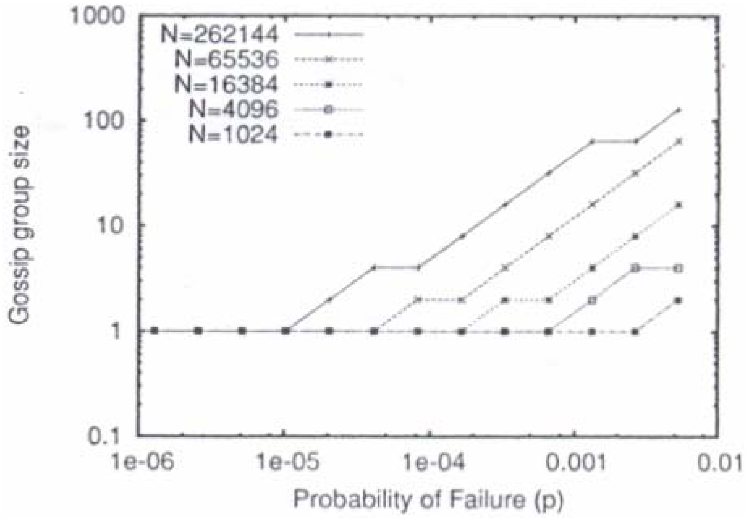
Figure 7: Comparison of time of completion when using limited global gossip communication



(a) Absolute Performance



(b) Gain of hybrid protocol



(c) Optimal gossip group size

Figure 8: Comparison of message counts when using limited global gossip communication

EFFICIENCY OF HYBRID PROTOCOL

By using hybrid protocol in sensor network for data aggregation, we determine that

1) The total number of messages required for an aggregation is minimized.
2) So, the amount of power consumed is reduced.
3) The time of completion of the hybrid protocol when compared to the tree and gossip protocol is reduced.
4) The protocol is independent of the probability of failure.

CONCLUSION:

In this paper we proposed a new protocol, which we called as hybrid protocol for computing aggregates in large-hundreds to thousands of sensor nodes and also works successfully in a highly faulty sensor networks that combines in a principled way two existing protocols: tree and gossip aggregation. Our first approach is to first determine the behavior of two existing protocols to identify when they are/are not acceptable and then to combine them in a way that capitalizes on their strengths and weaknesses. Then, we formulated a protocol that gives a optimal combination between the two existing protocols(tree and gossip aggregation protocol) .We empirically evaluated the performance of the hybrid protocol and observed that it is good and often outperforms, both tree and gossip aggregation. We believe that for large networks, a hybrid protocol as described in this paper is required to achieve good performance.

REFERENCES

1. Amis, A.D., Prakash, R., Huynh, D., and Vuong, T. 2000. Max-min d-cluster formation in wireless ad hoc networks, In Proceedings of IEEE INFOCOM (Tel Aviv, Israel). 32-41.
2. Bandyopadhyay, S and Coyle, E. 2003. An energy efficient hierarchical clustering algorithm for wireless sensor networks. In Proceedings of the 22nd Annual Joint Conference of the IEEE Computer and Communications Societies (Infocom 2003). Vol3.Springer-Verlag, San Francisco, CA, 1713-1723.
3. Bawa, M., Garcia-Molina, H., Gionis, A., and Motwani, R. 2003. Estimating aggregates on a peer-to-peer network. Tech. rep. Computer Science Department, Stanford University, Stanford, CA.
4. Ding, m., Cheng, X., and Xu, D. 2005. Aggregation tree construction in sensor networks. In Proceedings of the Vehicular Technology Conference (Orlando, FL).
5. Fiedler, M, 1973. Algebraic connectivity of graphs. Czechoslovak Math. J. 23,98, 295-305.
6. Girod, L., Elson, J., Cerpa, A., Stathopoulas, T., Ramanathan, N., and Estrin, D. 2004. Emstar : A software environment for developing and deploying wireless sensor networks. In Proceedings of the USENIX Technical Conference (Bostan, MA).
7. Gupta, I., van Renesse, R., and Birman, K.P. 2000. A probabilistically correct leader election protocol for large groups. In DISC'00: Proceedings of the 14th Conference on Distributed Computing. Springer - Verlag, London, UK. 89-103.
8. Jia, L., Noubir, G., Rajaraman, R., and Sundaram, R. 2006. Group-independent spanning tree for data aggregation in dense sensor networks. In Proceedings of the International Conference on Distributed Computing in Sensor Systems (DCOSS). , San Francisco, CA.
9. Kempe, D., Dobra, A., and Culler, D. E. 2003. Gossip-based computation of aggregate information. In Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (Cambridge, MA).
10. Madden, S., Frankin, M. J., Hellerstein, J. M., and Hong, W. 2002. A tiny aggregation service for ad-hoc sensor networks. SIGOPS Oper. Stst. Rev.36, SI, 131-146.
11. Shao, J. 2003. Mathematical Statistics, 2 ed. Springer Texts in statistics, Springer, Berlin, Germany.
12. Zhou, M and Liu, J. 2005. Tree-assisted gossiping for overlay video distribution. In Kluwer Multimedia Tools and Applications, Springer, Amsterdam, The Netherlands.

Article received: 2009-01-19