A Hybrid Approach for Clustering-based Data Aggregation in Wireless Sensor Networks

Woo-Sung Jung, Keun-Woo Lim, Young-Bae Ko
Graduate School of Information and Communication
Ajou University, Korea
{woosung, kwlim27}@uns.ajou.ac.kr, youngko@ajou.ac.kr

Sang-Joon Park
IT Convergence Technology Research Division
Electronics & Telecommunications Research Institute, Korea
sangjoon@etri.re.kr

Abstract—**In a wireless sensor network application for tracking multiple mobile targets, large amounts of sensing data can be generated by a number of sensors. These data must be controlled with efficient data aggregation techniques to reduce data transmission to the sink node. Several clustering methods were used previously to aggregate the large amounts of data produced from sensors in target tracking applications. However, such clustering based data aggregation algorithms show effectiveness only in restricted type of sensing scenarios, while posing great problems when trying to adapt to various environment changes. To alleviate the problems of existing clustering algorithms, we propose a hybrid clustering based data aggregation scheme. The proposed scheme can adaptively choose a suitable clustering technique depending on the status of the network, increasing the data aggregation efficiency as well as energy consumption and successful data transmission ratio. Performance evaluation via simulation has been made to show the effectiveness of the proposed scheme.**

I. Introduction

Wireless Sensor Networks(WSN) are gaining much reputation as an essential technology for the future ubiquitous world. The current trend of the cost decrease in computation and communication devices allows this technology to be used in many promising areas of applications. WSN consists of smart tiny sensor nodes that can sense various kinds of phenomenon using sensor modules and transmit the specific data wirelessly to a sink node, which will collect and compute all the data and provide various sensing information to specific users. These sensor nodes are usually deployed in large numbers (from a few to thousands) and in areas where it is difficult or impossible to be managed by humans. As a result, sensors have to be randomly scattered, and must use limited energy resources such as batteries. Therefore, sensor nodes must collaborate with each other to create a self-organized network, and must be equipped with energy efficient modules and protocols to decrease energy consumption and ensure long network lifetime.

Data aggregation in WSN is a data transfer technique by which several packets from sensor nodes are combined into one. This technique is essential in a wireless sensor environment because the reduction of data packets may reduce energy consumption, increase network lifetime, and increase successful data transmission ratio. To make the definition of data aggregation more convincing, the table below shows the comparison of the factors in data aggregation with familiar traffic scenarios. Each sensor can be represented as vehicles, each data transmission will be compared to the movement of the vehicles, and the network protocols will be represented as the rules for how the vehicles may move.

As seen from Table 1, creating an ideal WSN using data aggregation method is identical to managing a ideal vehicle traffic scenario, as improving the factors of data aggregation will be the same as trying to improve the traffic situation on the road. Efficient data aggregation protocols can reduce the traffic of data transmissions by ensuring quick and high data aggregation rates, while improving data transmission success ratio (traffic accidents in traffic scenario comparison) in sensor nodes. When a specific target occurs in a certain area of a network, there may be more than one sensor node that detects the target. The number of detecting sensors would increase even more in a high density sensor field. If all the sensor nodes were to send the data to the sink without any specific control protocols, much traffic will be overlapped in the same region. Eventually this can result in packet losses from collisions, delays from waste of bandwidth, and ultimately, waste of energy resources. For solving these problems, data aggregation protocol is required.

Data aggregation is especially important in the area of target tracking, because the movement of a specific target will enable many sensors to detect the same target, creating large amounts of redundant data. However, existing network or medium access protocols for general wireless sensor networks such as Directed Diffusion[1], S-MAC[2], or LEACH[3] are not suited to tracking these movements of targets, because excessive

TABLE 1
Comparison of data aggregation with traffic scenarios

Data Aggregation Factors	Comparison with Traffic Scenarios
Energy Consumption	The total amount of gas consumed by all the vehicles while traveling to its destination
Aggregation Count	The amount of car pooling to reduce the number of vehicles on the road
Data Transmission Success Ratio	The ratio of a vehicle traveling to its destination safely with accidents

"This research was supported by the MKE(Ministry of Knowledge and Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the IITA(Institute for Information Technology Advancement)" (IITA-2008-C1090-0801-0015), (IITA-2008-C1090-0801-0003)

978-0-7695-3526-5/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDS.2009.40

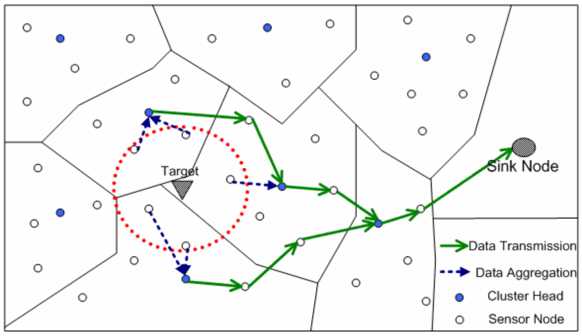


Fig. 1 Static Clustering based Data Aggregation

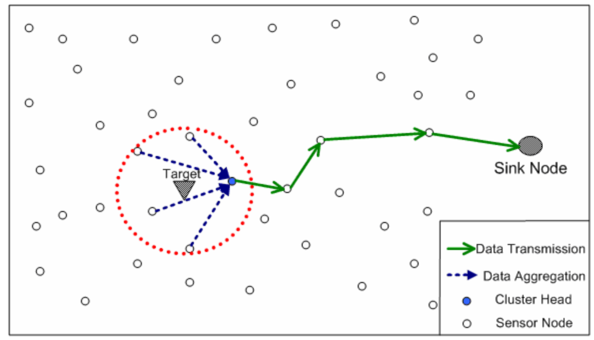


Fig. 2 Dynamic Clustering based Data Aggregation

amount of data generated by moving targets and appropriate sleep schedules cannot be controlled efficiently with the existing protocols.

This paper explains some general data aggregation methods that are used in target tracking applications, and proposes a hybrid clustering based data aggregation approach to improve the performance of the existing protocols. Section II describes some related works of data aggregation in target tracking applications. In Section III, we will introduce our novel hybrid clustering based data aggregation scheme, named "Adaptive cluster based data aggregation." A comprehensive performance evaluation conducted with the Qual-Net simulator will be explained in Section IV. The paper will explain our future work and conclude with the conclusion in Section V.

II. Related Work

Much of the research on data aggregation techniques have been done on the basis of a flat sensor network. For example, a tree-based data aggregation approach [4] creates a simple topology based on a parent and child relationship. This technique is effective in that a complicated protocol is not required, but causes high transmission delays in waiting for data aggregation, and does not ensure high data aggregation rate.

This paper has its motivation from networks with a certain hierarchy, rather from a flat network. Little research has been made to increase data aggregation efficiency in a hierarchical network in terms of target tracking. One popular method of controlling data aggregation through hierarchy is by using the concept of clustering. Two types of clustering methods can be used; one method being static clustering and the other dynamic clustering. Statically clustered networks [5][6][7] divide the network proactively into many clusters. A cluster would consist of general cluster nodes, and a cluster head that can schedule or aggregate data from its general cluster nodes. Any sensing events from a general cluster node are directly sent to the cluster head, which would aggregate several packets and send it towards the sink node. Data aggregation via static clustering uses pre-elected cluster heads for data aggregation, so data can be quickly and easily transmitted to the sink node with relatively low overhead. However, each cluster head will have to flood control packets every time static clusters are periodically created, which will cause great overhead. Also, as seen from Fig. 1, more than one cluster may sense a target at the same time, reducing the data aggregation efficiency. On the other hand, when all the sensing is arranged inside one cluster, static clustering will show best aggregation results. The problem of the static clustering technique can be somewhat alleviated by using multi-hop clusters [12][13]. This method will reduce the overall number of clusters and increase the size of each cluster instead, so the average number of clusters that will sense a target at the same time will decrease. However, additional overhead will occur when data is aggregated towards the cluster head from the general sensor nodes, because it may take more than one hop to send data from the general node to the cluster head.

Dynamically clustered networks [8][9] create a cluster reactively in the vicinity of the event sensing nodes. Upon event detection, a certain sensor node (preferably the one with the most energy or closest to the event) that has sensed a target will be elected as a cluster head, while all the other nodes that also sensed the target will become general cluster nodes that are located in a single hop range of the cluster head. All the data are collected and aggregated by the cluster head, and then sent to the sink. The advantage of this technique is that only the necessary nodes will participate in the data aggregation, considerably preserving energy of the other sensor nodes. Also, the static clustering method, only the selected cluster head will broadcast control packet, so the control packet overhead is relatively low. Furthermore, since all the data aggregation is made near the vicinity of the event, the data does not have to travel many hops to be aggregated. Therefore, aggregation rate for dynamic clustered data aggregation is very high.

However, the dynamic clustering based data aggregation also poses some serious problems. Firstly, clusters are made upon sensing of an event, so additional delay for electing a cluster head will occur before the transmission of the data. As a result, data can only be aggregated after the cluster election phase is complete. Also, unlike static clustering, dynamic clusters have to be frequently created when the velocity of a target is high. This is because whenever a target moves out of the sensing range of the cluster head, new sensor nodes that are sensing the target must restart the cluster election phase. This may result in control packet overhead and data transmission delay. On the other hand, the static cluster method can benefit when there are many targets or the target velocity is high, because proactively created clusters can be used without additional overhead. The general characteristics of the dynamic clustered data aggregation method can be seen in Fig. 2.

As we have seen from the explanation of two clustering

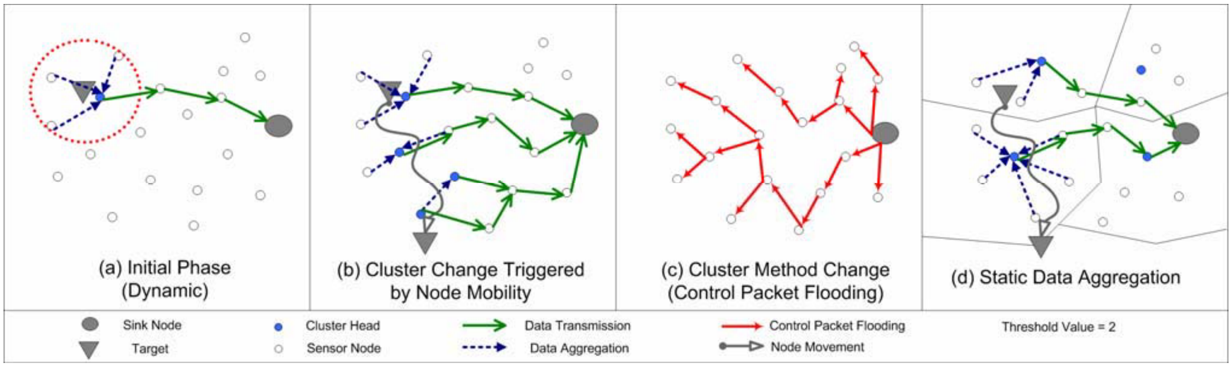


Fig. 3 Adaptive Clustering based Data Aggregation

based data aggregation protocols, it is important to utilize the advantages of the two methods, while alleviating the problems of them at the same time. Our attempt to approach this problem was done by developing the adaptive clustering based data aggregation scheme, which is explained below in section III.

III. PROPOSED SCHEME

The Adaptive clustering based data aggregation protocol is proposed in the section. Before explaining the characteristics of each protocol, we will make some assumptions that are needed for the proposed scheme to function properly.

Firstly, an initial tree topology is created between the sink node and all general nodes at the network initialization phase. This is initialized at the sink node by flooding a control message to all the sensors in the network. The broadcast message will contain the node's ID and the *hop_count* variable which will be initiated at 0. Any sensor nodes that receive this message will take the message sender as its parent, increment the *hop_count*, and then calculate its current distance from the sink node by using the *hop_count* variable. After the process, the sensor node will fill the message with its ID and the *hop_count*, and then broadcast the message again. This procedure is continued until all the nodes in the network join the tree topology, with the sink node as the root node of the tree. Any sensor node that receives a duplicate of the broadcast message will compare the current distance with the *hop_count* variable in the message. If the current distance is larger than the *hop_count*, the current parent and the distance to the sink will be replaced with the new information in the *hop_count* message. Otherwise, the message will be dropped. Using this control message, each sensor node will maintain its hop count to the sink, and be eligible to transmit all its data and also data from other nodes to the sink following this tree topology.

After the setup of the tree topology, a random static cluster head electing method much similar to [3] or [10] will be used to cluster all the nodes in the network. At the start of the static clustering phase, each sensor node will start a timer with a random delay. When the timer of a node expires, that node will declare itself as the cluster head and broadcast this information to its neighbors. Any node that receives this message and does not yet have its timer expired will become a cluster member and send back a reply to the cluster head. If a sensor is already a part of a cluster, it will drop this message. Using this process, the timers of all the sensors will eventually expire and clusters will be formed in the network. After a regular period, all the nodes will deform their clusters and restart the static cluster process again. In this case, the sensors nodes that have been cluster heads in the previous phase will be penalized and have longer timers than other sensor nodes. This will ensure that some nodes do not become exhausted, and the energy consumption of the sensors will be evened out throughout the network.

Adaptive clustering based data aggregation technique is a method that implements both static and dynamic clustering methods explained in the section II. The proposed scheme is designed on the assumptions that the static clustering based data aggregation technique has advantages when there are multiple targets, and when the velocity of those targets is high. On the other hand, dynamic clustering based technique has great advantages when there are only few targets with low velocity. Therefore, the proposed scheme will select the static cluster based aggregation when data traffic is high, and adaptively switch to dynamic cluster based aggregation when the network realizes that the data traffic is low. The threshold for deciding when to switch between the data aggregation methods will be configured and decided at the sink node. The initial clustering method of the network will also be configured at the sink. Details on the scheme are shown in Fig. 3.

The initial clustering method, shown in Fig. 3 (a), is configured to dynamic clustering method. The threshold of the data traffic is also configured. The threshold will be calculated by how many data packets the sink receives over a certain time, and this value will be used to switch between each data aggregation method. In the case of Fig. 3, the threshold value is configured to 2. When the sink node starts receiving data packets above the threshold value, shown in Fig. 3 (b), it decides that the dynamic clustering method is no longer suitable for the network. Therefore, shown in Fig. 3 (c), the sink node will flood the network with a control message and change the whole network to static data aggregation. We can observe in Fig. 3 (d) that the data aggregation method has been switched to static cluster based data aggregation. When the data traffic reduces again below the configured threshold, the sink node will flood the network and change back to dynamic clustering. This technique may alleviate the problems of each data aggregation protocols by effectively switching between them. However, the aggregation switching requires the whole network to be flooded by the sink with a control packet. Therefore, too much aggregation switching may cause the

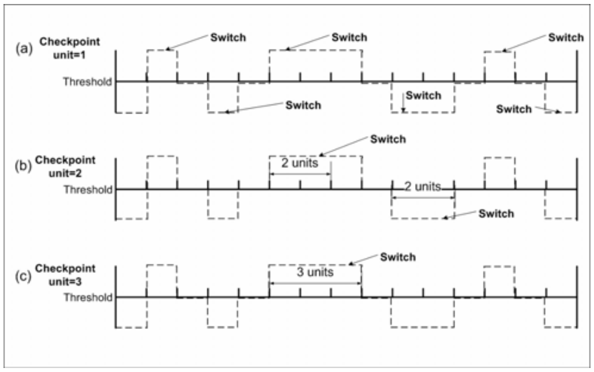


Fig. 4 Threshold Configuration

network performance to slightly degrade. To alleviate this problem, we will reduce the frequency of the switching in the data aggregation method by using the technique described in Fig. 4. Fig. 4 shows the rate of the data traffic the sink is receiving every certain period. It can be observed in Fig. 4 (a) that switching will occur too frequently when the data aggregation method switches every time the rate of data traffic is above or below the threshold.

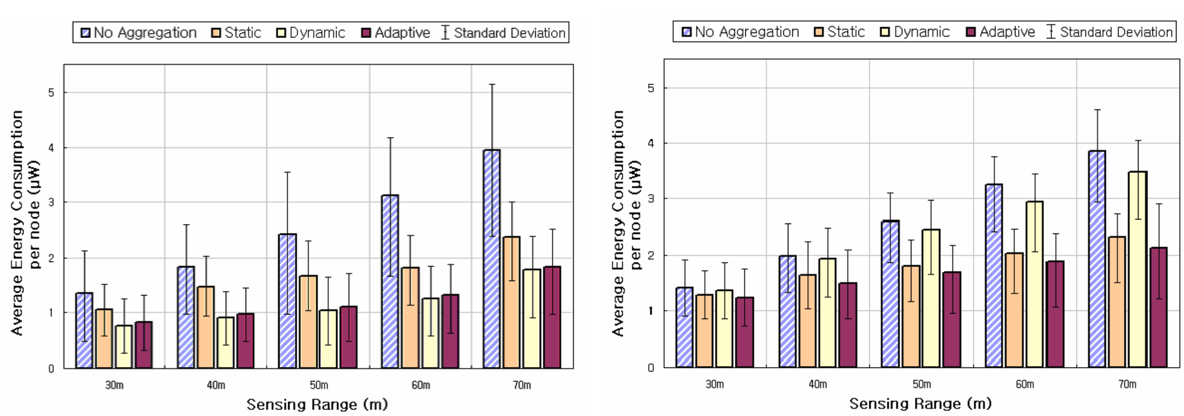
To reduce the data aggregation switching, instead of switching the data aggregation method every unit of data traffic measurement, we can configure the switch to occur once every two units of data traffic measurement, shown in Fig. 4 (b). This means that the data aggregation method will only change if the data traffic is over or under the threshold value for two consecutive time units. Fig. 4 (c) shows the data aggregation switching every three consecutive time units, and as a result, the switching occurs only once. Therefore, by controlling the frequency of the data aggregation switching, we can reduce the overhead that occurs from too much data aggregation switching. However, if the frequency of aggregation switching is too low, the protocol cannot effectively adapt to the changes in the data traffic, causing data overhead. Therefore, more analysis on configuring the appropriate switching frequency is required to optimize the performance of the adaptive clustering based data aggregation protocol.

IV. Performance Evaluation

The performance evaluation on comparing the performance of the adaptive clustering based data aggregation scheme has been carried out by using the Qual-Net 4.0 simulator. The proposed scheme was compared with the no aggregation method, static clustering based data aggregation protocol, and dynamic clustering based data aggregation protocol to prove that there have been enhancements from the existing protocols. We have used the simple one-hop static clustering algorithm and the TODA[4] dynamic clustering algorithm as the protocols for comparison.

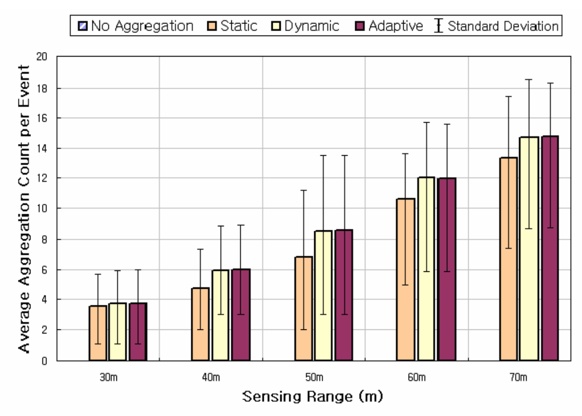
The simulation was conducted on a $500m^2$ area with 100 sensor nodes randomly deployed. The number of targets varies from 1 to 4. The sensing range of each sensor also varies from 30 to 70 meters. Each simulation time was configured to 100 seconds, with each simulation being carried out 10 times on a random topology to normalize the graph. Carrier Sensing Multiple Access (CSMA) protocol supplied by Qual-Net was used as the MAC protocol. An abstract radio type defined in the Qual-Net simulator was used, with the transmission rate configured to 250kbps and the transmission power to 100 meters. CBR traffic was used to transmit data packets with size of 100 bytes and control packets with size of 8 bytes. One sensing event is created by each target every second. The energy consumption of each sensor node follows the specifications of CC2420[14] transceiver, which is a widely used transceiver for various sensor motes. Following the datasheet of CC2420, the energy consumption of tx has been configured to 31.32mW/s, and the rx has been configured to 35.46mW/s.

Several parameters required for the Adaptive data aggregation method is configured. The packet threshold for aggregation switching is set to 5 packets per second. The switching frequency of the adaptive data aggregation has been set to 3 seconds. Also, for each protocol, the data aggregation time is configured to 20msec. This means that after receiving the first data packet, the cluster head will wait and aggregate data for the duration of the aggregation time before sending out its packet towards the sink. We increased the data aggregation time compared to previous works such as [4] to increase data aggregation efficiency. However, we expect that this will lead

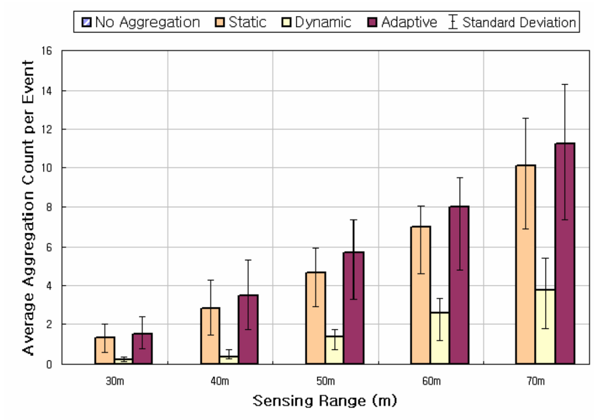


(a) Static Targets (b) Mobile Targets

Fig. 5 Comparison of Average Energy Consumption per node



(a) Static Targets



(b) Mobile Targets

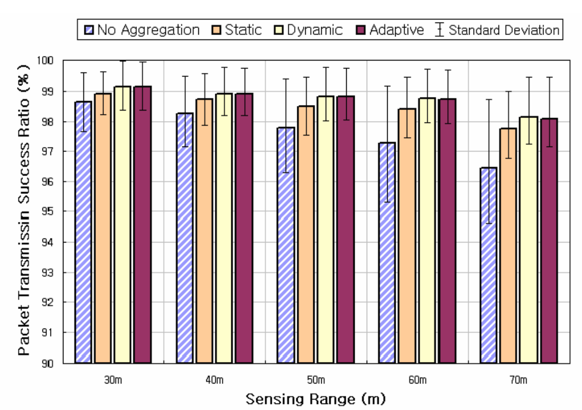
Fig. 6 Comparison of Average Aggregation Count per Event

to longer latency for each protocol.

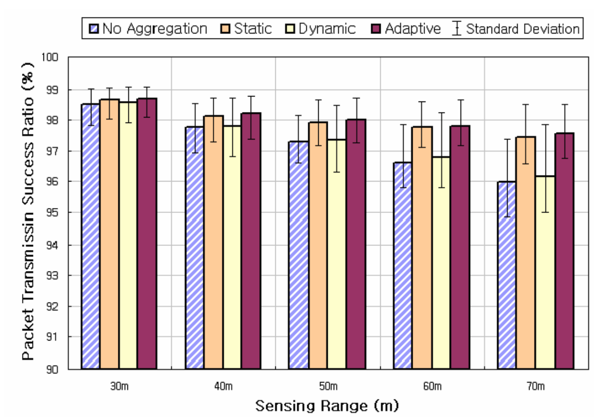
Two types of simulations were conducted, one case where targets are assumed to be static and the other where targets are mobile. We made our performance evaluation on 3 factors, which are average energy consumption per node, data packet transmission success ratio, and average aggregation count per event. The standard deviation of each outcome has also been presented on the graph to show the minimum and maximum values of each evaluation. Firstly, the average energy consumption of each sensor node in no aggregation, static aggregation, and dynamic aggregation were compared with our proposed scheme sensing targets with and without movement. The evaluation is shown in Fig. 5.

We can see from Fig. 5 (a) and (b) that energy consumption of each algorithm will increase as the sensing range increases. This is because the as the sensing range increases, there will be more sensor nodes that will sense an event at the same time, generating more data packets. It can be concluded from Fig.5 (a) that when targets are static, the dynamic clustering based data aggregation method shows the best performance while the static clustering based data aggregation has considerable overhead. This is due to the fact that when targets do not have mobility, the dynamic clustering method does not have to frequently generate clusters, saving energy in the process. The performance of the proposed adaptive clustering based data aggregation scheme tends to follow the dynamic clustering scheme, instead of following the static clustering scheme. This proves that the adaptive clustering method has successfully utilized the dynamic clustering scheme when data traffic in the network is low. On the other hand, Fig. 5 (b) shows that even when targets are mobile, the static clustering scheme has maintained its performance. However, the energy consumption in dynamic clustering based data aggregation scheme has dramatically risen. As already explained from section II, this is because the frequent movement of the target forces the algorithm to create many dynamic clusters, causing great control packet overhead. The proposed scheme does not follow the performance of the dynamic clustering in this case, and instead adaptively chooses the static clustering based data aggregation method successfully. We can conclude from Fig. 5 (a) and (b) that the proposed scheme has maintained its performance regardless of the changes in the environment, while the existing algorithms have weaknesses in certain environments and target behaviors.

Fig. 6 compares the average aggregation count of an event in each protocol. This counts the number of aggregations that has been made when one event has been sensed by a number of nodes. Fig. 6 (a) shows performances of each algorithm when



(a) Static Targets



(b) Mobile Targets

Fig. 7 Comparison of Packet Transmission Success Ratio

targets are static. For no aggregation scheme, no data is aggregated. Therefore, all the sensed data is sent directly to the sink node, and the aggregation ratio is zero. Dynamic and Adaptive clustering based data aggregation protocols have performance have aggregated data packets the most, which also means that these two protocols have highest data aggregation efficiency when targets are static. However, the static clustering method shows poor performance when target is static. This is due to the reason that static clustering method creates multiple clusters and all the data may not be sent to only one cluster. Therefore, the data are spread to more than one cluster head, degrading the aggregation ratio. When target is mobile in Fig. 6 (b), the overall performance of each scheme degrades considerably, but we can observe that the adaptive data aggregation method maintains has the highest aggregation efficiency. We believe that this is due to the method switching algorithm implemented in the adaptive scheme, where the network realizes that the default dynamic clustering method is not the optimal solution when targets are mobile and switches the network to static clustering method. As proven in Fig. 5 already, the dynamic clustering method shows dramatic performance decrease when targets are mobile.

Fig. 7 compares the successful data packet transmissions that occurred in the network of each algorithm. This graph will show the percentage of the data packets that has been successfully transmitted to the next hop node, while a packet loss due to collisions will be rendered as an error. The static clustering based scheme in Fig. 7 (a) shows low transmission success rate when targets are static, because relatively low aggregation ratio makes the algorithm to generate more packets, which creates a higher chance of each packet colliding with each other. On the other hand, since dynamic clustering scheme has a nearly perfect aggregation ratio, the number of packets traveling in the network is very low. Therefore, less data collision will occur and the success of data transmission is very high. However, in Fig. 7 (b) when targets are mobile, the dynamic clustering based scheme engages in frequent cluster head elections, transmitting many broadcast packets in the process. Due to this, the performance of the dynamic scheme degrades significantly, as bad as using no aggregation method at all. The proposed scheme can maintain a high performance, regardless of the movement of the targets.

In overall, by analyzing the graphs, it can be concluded that the adaptive clustering based data aggregation method will have the best performance in target detecting and tracking environments. The ability to switch its aggregation scheme based on the status of the network can keep its performance to a high level. The dynamic clustering protocol shows the best performance when targets are static, but it is rendered useless when targets are mobile. The static clustering method shows reliable performance whether or not the target is moving, but the overall performance is only adequate.

V. CONCLUSION

This paper proposes a hybrid protocol for improving data aggregation efficiency in target tracking applications of wireless sensor networks. By taking a hybrid approach and adaptively selecting the appropriate data aggregation method, considerable improvements were achieved. Simulation results have shown advantages of the proposed scheme, but more accurate evaluations and research needs to be done by changing some parameters in the simulation. For example, a MAC suited for sensor networks, such as S-MAC, B-MAC or IEEE 802.15.4 can replace the CSMA technique that is used in our simulation above. Using multi-hop clusters or more complex static clustering could also increase the performance of each protocol. Further research on configuring the aggregation time to optimize the proposed scheme is also important. Also, the flooding technique used to switch aggregation methods in adaptive clustering based data aggregation protocol may produce large amounts of overhead. Therefore, a localized scheme on changing only some parts of the network instead of using flooding to change the whole network could be a solution. Finally, implementing the algorithm on actual sensor motes and using more realistic parameters are required to test our proposed scheme on real world situations.

REFERENCES

[1] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, F. Silva, "Directed diffusion for wireless sensor networking," In *IEEE/ACM Transactions on Networking,* Vol 11, Issue 1, pp. 2-16, Feb 1999.

[2] W. Ye, J. Heidemann, and D. Estrin. "An energy-efficient MAC protocol for wireless sensor networks," In *proceedings of IEEE INFOCOM*, June 2002.

[3] W. Heinzelman, A. Chandrakasan and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," In *33rd Hawaii Int'l Conference on System Sciences*, Jan 2000.

[4] S. J. Park, "Performance Analysis of Data Aggregation Schemes for Wireless Sensor Networks Static-cluster aggregation and Dynamic-cluster aggregation Section," A dissertation for Doctoral in North Carolina State University, 2006.

[5] M. Busse, T. Haenselmann, W. Effelsberg, "TECA: A Topology and Energy Control Algorithm for Wireless Sensor Networks," In *ACM Proceedings of Modeling analysis and Simulation of Wireless and Mobile Systems (MSWiM)*, Oct 2006.

[6] Y. Chen and S.H. Son, "A fault tolerant topology control in wireless sensor networks," In *ACS/IEEE International Conference on Computer Systems and Applications (AICCSA)*, Jan 2005.

[7] R. Ghosh and S.Basagni, "Napping backbones: energy efficient topology control for wireless sensor networks," In *IEEE Radio and Wireless Symposium (RWS)*, Jan 2006.

[8] W.-P. Chen, J.C. Hou and L. Sha, "Dynamic clustering for acoustic target tracking in wireless sensor networks," In *IEEE Transactions on Mobile Computing*, Vol 3, Issue 3, pp. 258-271, July 2004,.

[9] J.Y. Cheng, S.Ruan, R.Cheng, T.Hsu, "PADCP: Power-Aware Dynamic Clustering Protocol for Wireless Sensor Network," In *IFIP International Conference on Wireless and Optical Communications Networks (WOCN)*, Apr 2006.

[10] D. Kejun, Z. Xingshe, Z. Xingguo and L. Zhigang,"HETCP: A Hierarchical Energy Efficient Topology Control Protocol for Wireless Sensor Networks," In *IEEE Wireless Communications, Networking and Mobile Computing (WiCOM)*, Sep 2006.

[11] T. He, L. Gu, L. Luo, T. Yan, J. Stankovic, S. Son, "An Overview of Data Aggregation Architecture for Real-Time Tracking with Sensor Networks," In *20th International Parallel and Distributed Processing Symposium,* Apr 2006.

[12] Y. Qian, J. Zhou, L. Qian, K. Chen, "Highly Scalable Multihop Clustering Algorithm for Wireless Sensor Networks," In *International Conference on Communications, Circuits and Systems (ICCCAS)*, Jun 2006.

[13] M. Ma, Z. Zhang, Y.Yang, "Multi-channel polling in multi-hop clusters of hybrid sensor networks," In *IEEE Global Telecommunications Conference (GLOBECOM)*, Nov 2005.

[14] CC2420 Datasheet, http://www.ti.com