T-ANT: A Nature-Inspired Data Gathering Protocol for Wireless Sensor Networks[1]

S. Selvakennedy
School of Information Technologies, University of Sydney, NSW 2006, Australia.
Email: skennedy@it.usyd.edu.au

S. Sinnappan^ and Yi Shang*
^School of Economics & Information Systems, University of Wollongong, Wollongong NSW 2522, Australia.
Email: suku@uow.edu.au
*Dept. of Computer Science, University of Missouri-Columbia, Columbia, MO 65211, USA.
Email: shangy@missouri.edu

***Abstract*—There are many difficult challenges ahead in the design of an energy-efficient communication stack for wireless sensor networks. Due to the severe sensor node constraints, protocols have to be simple yet scalable. To this end, collective social insects' behavior could be adopted to guide the design of these protocols. We exploit the simple heuristics of ant colony in foraging and brood sorting to design a hierarchical and scalable data gathering protocol. Also, we demonstrate how it could exploit data correlations in sensor readings to minimize communications cost in the data gathering process towards the sink. This approach selects only a subset of sensor nodes to reconstruct data for the entire network. A distributed variance estimation algorithm is introduced to capture data correlations with negligible state maintenance. It is shown that this algorithm is able to predict the values rather accurately. Due to the general robustness of any nature-inspired algorithm, our data gathering protocol is reliable. It is fully distributed, and promises scalability and substantial energy savings.**

***Index Terms*—Data Gathering, Data Correlation, Clustering, Swarm Intelligence, Simulation, Sensor Networks**

I. INTRODUCTION

Wireless sensor technology is garnering a lot of interests due to its promises, enabling it to evolve rather rapidly. For instance, in terms of the sensor node hardware, the Mica2 mote has roughly eight times the memory and communication bandwidth as its predecessor, the Rene mote, developed in 1999 for the same power budget [1]. These sensor nodes have found use in many applications such as earthquake monitoring, target tracking and surveillance, and structural monitoring. The nodes are typically less mobile due to their unique application needs, substantially more resource constrained and more densely deployed than mobile ad hoc networks (MANETs). Even though, there have been significant advances in recent years, more energy-efficient solutions are required within the communication stack for the conservation of the battery power. An approach that is likely to succeed is the use of a hierarchical structure [2], which also promotes scalability of wireless sensor networks (WSNs).

Clustering with data aggregation is an important technique in this direction, and it makes the tradeoff between energy efficiency and data resolution. Most clustering algorithms aim at generating the minimum number of clusters and transmission distance. These algorithms also distinguish themselves by how the clusterheads (CHs) are elected. The LEACH algorithm [3] and its related extension, TCCA [4] use probabilistic self-election, where each sensor node has a probability p of becoming a CH in each round. Such role rotation aims to distribute the energy usage for a more load-balanced operation. However, LEACH only works for single broadcast domain networks, and mostly operates in a suboptimally formed hierarchical structure due to its stochastic nature. TCCA overcame the former problem by allowing multihop clusters but still suffers from the latter.

When developing WSN protocols, another crucial design issue to consider is the network reliability. To this end, social insect swarm behavior may provide an ideal model for the design of such less controllable systems. To our knowledge, very few researchers have considered or adopted such nature-inspired approaches for WSN design. However, a number of recent works has been based on different swarm behaviors in the design of routing protocols for MANETs. As there are many important similarities between these two ad hoc technologies, we believe building on these knowledge may be useful for WSNs.

[1] Based on " Data dissemination based on ant swarms for wireless sensor networks", by S. Selvakennedy, S. Sinnappan and Yi Shang which appeared in the Consumer Communications and Networking Conference, 2006. © 2006 IEEE.

In any random network deployment, many sensor nodes may also exhibit data correlations in their sensed data due to their overlapping sensing ranges. This issue may be addressed like a topology control issue where it could be formulated as a minimum *graph covering* problem, or as a data aggregation or compression problem that minimizes the amount of data transmitted to the sink through some in-network processing. In some applications, sensor nodes may also exhibit temporal correlation if the monitored physical characteristic has small variability.

When the correlated data is exploited within the network, the sink gathers snapshots of reduced signal data values measured at the sensor nodes, and uses interpolation to derive the signal value at other points in the monitored region. Here, we exploit the temporal correlations in the sensor data. This correlation at a node is captured entirely based local observations with minimal state, and is used to control the node's participation in the data gathering process towards the sink.

In this paper, we investigate the T-ANT protocol, introduced initially in [5], more in depth and extend it to support some in-network processing to remove redundant data without the sink's control. This protocol adopts the clustered strategy realized using useful principles of the ant colony behavior. It achieves the objective of uniform cluster formation by exploiting two swarm behaviors, namely foraging and brood sorting. T-ANT achieves substantially better performance than that of a flat minimum hop routing strategy, LEACH as well as TCCA.

The rest of the paper is organized as follows. Section 2 presents the perspective of this area of research. Various clustering algorithms, nature-inspired algorithms and data gathering protocols proposed in the literature are discussed. In Section 3, the details of the T-ANT clustering algorithm and the associated data variability estimation algorithm are described. The comprehensive simulator used to experiment with this protocol is described in Section 4. Various experiments and the corresponding results are presented and analyzed in Section 5. The paper concludes with the main findings of this work in the final section.

II. RELATED WORK

Intense research in the field of sensor network technology in recent years has fueled further development in micro-sensor technology and low-power analog/digital electronics. To support scalable data gathering, it is realized that the approach that is likely to succeed to provide an energy-efficient solution is to adopt a hierarchical structure. To this end, various clustering algorithms have been proposed in different context. Generally, clustering algorithms segment a network into non-overlapping clusters comprising a CH each. Non-CH nodes transmit sensed data to CHs, where the data signals could be aggregated and transmitted to the sink. Initially, these algorithms focused on the connectivity problem [6-8] but later energy-efficiency was more of interest in wireless ad hoc and sensor networks [3, 4, 9-13].

Another crucial design aspect of WSNs is the network reliability and fault-tolerance. It has been demonstrated in different context that the collective behavior of social insects has many attractive features, not the least robustness and reliability. However, there are only very limited WSN proposals inspired by such biological behaviors. Due to some parallels to MANETs, we reviewed some nature-inspired algorithms proposed for this domain. The first MANET routing algorithm based on ant colony principles is ARA [14]. It exploited the *pheromone* laying behavior of ants. Pheromone is a quality metric indicating the goodness of a path. Although pheromone evaporates, subsequent ants leave additional pheromone and thus reinforce the path. Over time, ants establish the shortest path between food and their nest in a full-distributed and autonomous manner. Ants are flooded towards destinations while establishing the reverse paths to the ant source. The gradual decay of pheromone introduces a form of negative feedback to prevent old routes from remaining in the forwarding tables when routes fall out of favor with ants. The shortest paths become preferable, and most ants use them. However, longer paths are not entirely lost as some ants may still maintain such routes. Routing schemes based on such ant colony behavior is both robust and adaptable. When the shortest route is lost due to some event, the longer routes provide alternative options. Other nature-inspired protocols were discussed in [5].

The problem of gathering correlated data in WSNs has been recently addressed by means of either compression or topology control-like approaches. The main focus in the compression approach is to reduce the total number of bits transmitted towards the sink using suitable coding techniques. In [15], the authors proposed a distributed compression technique based on the Slepian-Wolf model, and the level of compression is determined centrally by the sink. This node tracks the correlation structure among nodes, and then, individually informs each sensor node the number of bits to be used for encoding. They have however assumed that the network is a single broadcast domain. Moreover, the sink would prove to be a bottleneck in a larger network, and could lead to the scalability problem. Single-input coding strategies were adopted in [16] to encode a node's data based on a neighbor node. As the problem of finding the minimum-energy data gathering tree is NP-complete, they presented approximation algorithms to construct a near-optimal data gathering tree for foreign-coding and self-coding schemes. In [17], the authors proposed efficient approximation algorithms also based on Slepian-Wolf coding to optimize the transmission structure and the rate allocation at each node. In all these approaches, all sensor nodes are required to participate in data transmission at every round even though at reduced number of bits transmissions.

Another type of algorithm that aims to reduce the number of transmissions by making redundant nodes to sleep were proposed in [18] and [19]. In principle, these approaches are similar to topology control algorithms such as SPAN [20] and ASCENT [21]. However, the node redundancy in the data gathering issue is of the application perspective, whereas it is mainly of the routing perspective in the latter. In [18], the authors proposed a scheme to reduce number of transmissions and provided approximate results to aggregate queries through spatial data correlation. Only a subset of nodes disseminates data to the sink. A set of CHs is selected using a simple localized scheme. It uses only the edges of the forwarding tree for selection of CHs and routers. With a similar aim of forcing redundant nodes to sleep, Gupta et al. [19] formulated this problem as finding the minimum dominating set problem, which is a well-known NP-complete problem. Accordingly, they proposed a distributed approximation algorithm and a couple of centralized heuristics to select a small *correlation-dominating* set, which is sufficient to infer data of the remaining nodes. For their distributed algorithm, each node is expected to collect *k*-hop neighborhood information to form the correlation hypergraph. It is also assumed that the correlation structure is fixed. Even though a more complete spatial correlation is achievable here, the computation of the correlation weighting coefficients has energy cost in the order of the transmission cost, and the storage requirement is exponential in the number of neighbors. Thus, in this paper, we extend the T-ANT protocol to exploit temporal data correlation in sensor readings, which only involves local decisions, and has smaller energy and storage costs.

III. The T-ANT Protocol

There are two main components to this protocol. The first aspect is related to the CH election and clustering, whereas the second involves the estimation of the data variance and redundancy detection. These are described in the following subsections, respectively.

A. *The Clustering Algorithm*

T-ANT adopts two-phase clustering process involving the cluster setup and steady state phases. To guide the CH election, we chose to use a swarm of ants. Through the use of a swarm of ants, we could guarantee that the network always maintains an optimal number of clusters.

During the node initialization, the sink releases a number of ants (i.e. control messages). Ramos and Merelo [22] suggest that the ratio of the number of ants to the number of objects (i.e. sensor nodes) should equal 0.1. When the sink releases an ant, it chooses one of its neighbors at random. The ant could travel into the network as deep as restricted by its time-to-live (TTL) field. When an ant arrives at a node, the next node is randomly chosen (excluding the sender) for its subsequent stop if TTL has not expired. If TTL expires, the ant remains at this node. If however the final ant location overlaps with another ant, the former ant must find another location.

The cluster setup (CS) phase is controlled through a CS timer. When this timer expires, a node checks to see whether it possesses an ant. If the node has an ant, it becomes a CH. When a node becomes a CH, it advertises to its neighbors by broadcasting an ADV message with its node id and a TTL field to constrain the ADV propagation. Upon receiving an ADV message, a regular node records the CH id, the sender's id as its parent, the hop distance to this CH, the number of ADV messages received so far and total hop distance to all *seen* CHs, and then rebroadcasts if its TTL permits. A node decides to join a cluster when its join-timer expires. It then computes its *pheromone* level based on its total hop distance (h) to CHs, the number of CHs (n) in its neighborhood, and its normalized residual energy. The pheromone expression is based on the forwarding probability formula used in the ant routing algorithm [23], but expanded as:

$$p = \frac{p + \Delta p}{1 + \Delta p} \tag{1}$$

where Δp is given by:

$$\Delta p = \frac{k}{h_*^2} \times \frac{E_{resi}}{E_{max}} \times \frac{\sum_{i=1}^{n} h_i}{n} \tag{2}$$

h_* is the node's hop distance to the selected CH, E_{resi} is the residual energy, E_{max} is the reference maximum battery energy and k is the learning rate of the algorithm (= 0.1). This expression ensures that Δp is higher when the node is only reachable by fewer CH nodes (smaller n), far from CHs ($\sum h_i$), has higher residual energy (E_{resi}) or is nearer to its selected CH (h_*). A regular node chooses the best cluster to join based on its hop distance to the CH, which would ensure minimal energy dissipation during the data collection rounds. The node joins a cluster by sending a JOIN message with its id, the selected CH id and its pheromone level. If the CH is in range, the message is transmitted directly; otherwise it is forwarded through its parent to the CH. When a CH receives JOIN messages, it finds the member with the highest pheromone level to attract its ant for the following CS phase.

Before the next CS timer expires, the ants wander to the nodes with the highest pheromone level among their neighbors, and these nodes will be the subsequent CHs. Before an ant leaves its current node, an amount of *anti-pheromone* is laid to mimic a rapid decay of pheromone level [5]. This ensures that the ants do not return to the same node too soon, which promotes load balancing.

The given pheromone expression guides the evolution of the swarm to achieve the *separation* behavior between ants in the swarm [5]. It is found empirically that separation is attained rather quickly within 3-5 rounds as an optimal swarm size is used. Another useful swarm behavior is *alignment* [5]. In our context, the *area* served by each ant represents the alignment property. It is reflected by the number of members in a cluster. When the swarm evolves to achieve separation, alignment is also achieved as a side-benefit. The phenomenon due to both behaviors is captured by the following fitness functions, respectively. The CH election fitness function S to capture the separation behavior is:

$$S = \sum_{i=1}^{n_c} \frac{n_i}{\sum_{j=1}^{n_i} h_{ij}} \tag{3}$$

where n_c is the number of CH nodes, n_i is the number of ADVs seen by CH i and h_{ij} is CH i's hop distance to CH j. The clustering fitness function A to represent the alignment behavior is as follows:

$$A = \sum_{i=1}^{n_r} h_i \tag{4}$$

where n_r is the number of regular nodes and h_i is node i's hop distance to its CH.

In the steady state phase, if a regular node is considered redundant, it sends or forwards its sensory data to its CH. It is possible that the foraging ants may die due to the environmental uncertainty or node failure. To avoid a reducing number of ants in the network over time, ants have a finite lifetime. When ants die, the sink re-releases the same optimal number of ants to restart the process. In order to determine the nodes that have redundant information and thus, could be made to sleep, we introduce the following algorithm to capture data correlations.

B. The Variance Estimation Algorithm

In order to decide a node's participation in the data gathering process, we need to determine whether its data is redundant. We choose not to base this selection decision on spatial correlation due to the amount of data to be collected and stored from the k-hop neighbors as well as the subsequent computation cost involved in the value prediction even as a linear combination of the neighbor values.

To reduce the amount of state at each node, only temporal correlation is maintained. If the variability of the monitored value falls below the specified application bound, the node's data is considered redundant, and it locally decides to sleep. This decision is made during the CS phase. Finding a factor's variability problem could be analogized to the round-trip time (RTT) variance estimation problem in setting of the retransmission timeout value in the TCP transport protocol. The timeout algorithm allows a TCP entity to cope with the highly dynamic Internet traffic. This RTT variance estimation is based on the Jacobson's algorithm, and is specified as part of TCP in RFC2988 [24]. Jacobson introduced a variation measure called mean deviation, and used it with the exponential smoothing technique to capture the dynamic nature of Internet traffic. As it is obvious that TCP is successful in adapting to this dynamism, a similar estimation algorithm could prove useful for our purpose to capture the sensor data variability.

To enable such estimation, each node maintains an average value (represented as s_val) that stores the weighted sensor data value based on present and past values as follows:

$$\text{s\_val}^{k+1} = (1\text{-g})\times\text{s\_val}^{k} + \text{g}\times\text{val}^{k+1} \tag{5}$$

where s_val^i is the smoothed value at the i^{th} time instant, val^i is the actual sensed value at the i^{th} time instant and g is a constant, 0<g<1. In order to capture the measure of dispersion of the sensed data, we adopt the mean deviation metric as follows:

$$\text{s\_dev}^{k+1} = (1\text{-h})\times\text{s\_dev}^{k} + \text{h}\times|\text{val}^{k+1} - \text{s\_val}^{k}| \tag{6}$$

where s_dev^i is the smoothed variability of the sensed values at the i^{th} time instant and h is a constant, 0<h<1. Finally, the sensed value for the $(k+1)^{\text{th}}$ time instant can be predicted as follows:

$$\text{val}_*^{k+1} = \text{s\_val}^{k+1} + \text{s\_dev}^{k+1} \tag{7}$$

where val_*^{i} is the predicted value at the ith time instant. If the predicted and actual values deviate lesser than the application bound, this value is uninteresting for the application and could be approximated by the sink from the historic data. Thus, this node should not participate in further data gathering rounds until data variability exceeds the bound. Since redundant nodes are decided during the CS phase, these nodes will not be involved in the clusters formation. Also, the ants will only forage among the active nodes. As sensing the environment is continuously performed by all nodes, an inactive node many rejoin during the next CS phase, if its value falls outside the threshold.

IV. Simulation Framework

The performance of T-ANT clustering is evaluated using a discrete-event simulator. To enable a comprehensive study, the effects of both routing and MAC protocols are integrated. In the description of the simulator, we assume that each sensor node is aware of:

- its neighbors due to the occasional *beaconing* by the sink and the cluster setup phase; and
- the network is synchronized by means of any time synchronization protocols .

The radio model is assumed to follow isotropic propagation. As for the MAC choice, we adopted the CSMA protocol due to its simplicity and its promise of scalability. However, a straightforward application of this protocol in a convergecast scenario is a recipe for failure. In the periodic monitoring type application, when the sensor data timer expires, all nodes capture their sensory value and convert to digital via an analog-to-digital converter (ADC) linked to the sensing hardware. Assuming a time-synchronized clustered network, all nodes generate the sensory message for transmission towards their CHs at the same time. If no precaution was taken, their transmissions would interfere resulting in many collisions and retransmissions. In order to reduce such collisions, Huang and Zhang [25] proposed that a node should delay its transmission relative to its distance to the destination (h) and the node density. As there is likely to be many nodes at each hop, an additional random offset is also included to further reduce the collision probability. We choose to adopt a similar temporal coordination with respect to a node's CH. However, it was modified here to be less conservative to reflect the smaller scope of a cluster rather than an entire network as in [25]. The random wait time function (T) is accordingly given as:

$$T(h) = (1 + r).\tau h \quad (8)$$

Where r is a uniformly distributed random number from 0 to 1 and τ is the average one-hop delay.

At the network layer, we adopted a simple routing mechanism in Greedy Routing Scheme (GRS) to control the network's forwarding behavior. The forwarding objective is to minimize the number of hops between the sink and the other nodes. To establish this minimum hop routing tree, the sink occasionally broadcasts a *beacon* message with a hop count, which is initialized to zero. Upon receiving the beacon, each node records the sender id, increments the hop count by one, and then rebroadcasts it. A node only rebroadcasts if the new hop count is smaller than its stored value. Since we are focusing on a quasi-stationary type of application, the sink node only needs to perform occasional beaconing to avoid significant overhead. This forwarding rule establishes a minimum hop tree rooted at the sink. Finally, the T-ANT scheme is implemented between the application and the network layer, and thus, the overall system framework is as shown in Fig. 1.

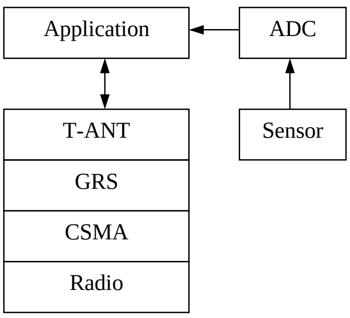

Figure 1. The simulation framework.

Based on the given simulation framework, we investigated T-ANT's performance against LEACH, TCCA and a flat strategy (i.e. the application sits directly on GRS). However, since LEACH can't be applied directly to a multihop network, we modified this algorithm to use a routing protocol to forward messages whenever the destination is not within a node's radio range. We termed this modified algorithm as multihop-LEACH (or m-LEACH). The results from this comparison and other evaluations are presented further.

V. Results and Discussions

For these simulation experiments, we assumed that there are 100 sensor nodes distributed randomly in a square $M \times M$ region with M = 500 m. The transceiver energy parameters are set as: E_{elec} = 50 nJ/bit and ε_{fs} = 10 pJ/bit/m^2. The energy for data aggregation is set to E_{DA} = 5 nJ/bit per signal [3]. The control and data message sizes are fixed at 30 bytes, and sensory data is generated at 2-second interval. Each CH node retains its CH status for 20 seconds. The number of ants is fixed at 10 and the anti-pheromone rate is 0.1.

The performance metrics being investigated are:

- *Clustering fitness*: This metric is based on the fitness function given as (4). It represents the goodness of the cluster formation in terms of alignment involving all regular nodes.
- *CH election fitness*: This metric is based on the fitness function given in (3). It represents the goodness of all the elected CH nodes in terms of CH separation.
- *Average energy per round*: This metric represents the average energy dissipated by all the nodes in a round of data dissemination.
- *Network lifetime*: This metric represents the time period from the instant the network is deployed to the moment when the first sensor node runs out of energy.

Fig. 2 depicts the clustering fitness value at different simulation time. For T-ANT, the initial value is high indicating that the swarm has not yet achieved the alignment behavior as the ants are randomly released into the network. However, as pheromone is laid and anti-pheromone takes effect during CS phases, the swarm alignment improves. Within the third evolution, the swarm is able to align. As for the other schemes, the fitness value varies rather wildly. Unlike T-ANT, TCCA mostly operates in sub-optimal fashion. Also for m-LEACH, the fitness value is always smaller than the other schemes due to the ADV messages being limited to first-hop neighbors. Any u*ncovered* nodes would have to resort to *direct* transmission to the sink. Since m-LEACH and TCCA have probabilistic CH election, it is possible that the CHs may even be clumped. When the CHs are clumped, the disparity among clusters is large in terms of their number of members, as each CH contends for the same regular nodes pool.

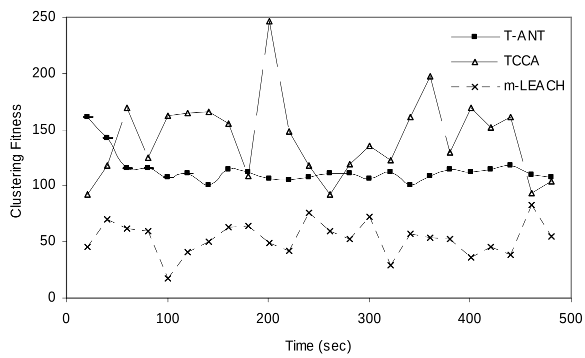

Figure 2. Clustering fitness at different simulation time of T-ANT, m-LEACH and TCCA.

In Fig. 3, the CH election fitness is depicted for the same three algorithms. Again, consistent behavior as above is obtained. For T-ANT, it has a higher function value initially, but it quickly converged somewhat. The ants move to better location based on the computed pheromone level, and within the fifth round, the swarm is able to achieve the separation behavior. This behavior ensures the elected CHs are distributed as uniformly as possible. Even after the uniformity is achieved, the ants keep moving at each round to ensure that the CH role is shared among nodes, and energy-load balancing is attained. As for the other schemes, the topology barely settles and mostly has a lower value than T-ANT. A lower value indicates that the CHs in these schemes are mostly too close to each other. In m-LEACH, the fitness function quite often assumes a zero value compared to TCCA. This is mainly due to its restricted ADV propagation, where a CH is unable to recognize another CH located only two hops away.

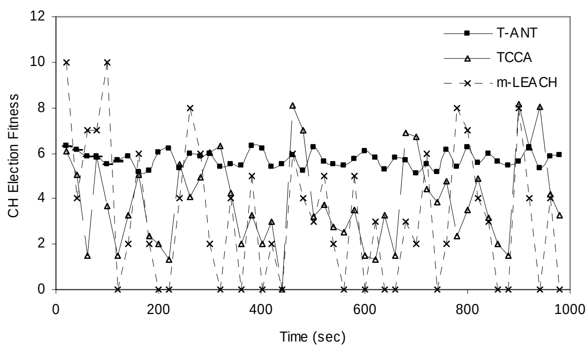

Figure 3. CH election fitness at different simulation time of T-ANT, m-LEACH and TCCA.

Since cluster size was shown to have a significant impact on clustering algorithms [4], we varied ADV's TTL value and compared these algorithms. In Fig. 4, these algorithms exhibit the presence of an optimal cluster size. However, T-ANT achieves significantly more energy savings than m-LEACH and TCCA for cluster sizes up to four. T-ANT achieves energy savings of more than 30% against m-LEACH. When cluster size is two, T-ANT dissipates 27% lesser energy compared to TCCA. This observation is consistent with fitness values reported in Figs. 2 and 3. Since m-LEACH and TCCA mainly operates with sub-optimally formed topology, their energy dissipations are higher. However, for larger cluster sizes, T-ANT's benefit is less apparent. This is mainly caused by the energy expended during the cluster setup phase that is significantly larger as ADV messages are flooded further, and the JOIN messages have to be forwarded many hops before reaching their CHs. Similarly, during the steady state phase, significant intra-cluster traffic is generated negating T-ANT's benefits.

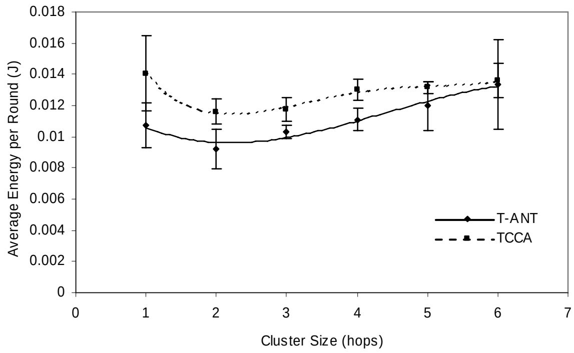

Figure 4. Average energy usage per round against the cluster size of T-ANT and TCCA (and m-LEACH similar to TCCA with cluster size one).

In Fig. 5, the improvement gained through T-ANT is further exemplified by the network lifetime graph. For this investigation, we have fixed the initial battery energy, E_{max} at 0.1J. It is evident that T-ANT exhibits the longest lifetime with all nodes remaining fully functional. It is found that T-ANT achieves almost 3.5 times the lifetime of m-LEACH and almost five times of the flat approach. It also achieves up to 50% longer lifetime than TCCA.

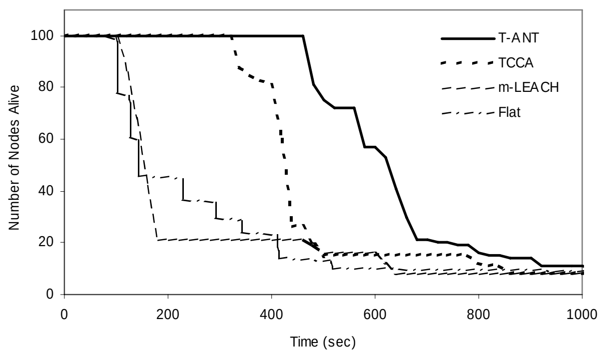

Figure 5. Network lifetime against simulation time of T-ANT, TCCA, m-LEACH and the flat strategy.

To investigate the effectiveness of the proposed estimation algorithm, we gathered daily temperature data (in Fahrenheit) of 100 Australian weather stations for the whole year of 2004 [26]. Using naturally gathered data, we exhibit the extend of correlations inherent in real data. For the variance estimation algorithm, the constants asume these values: g = 0.125 and h = 0.25.

In Fig. 6, the accuracy of the adopted variance estimation algorithm is investigated. The continuous line represents the predicted values, whereas the crosses represent the actual observed data against the left y-axis. The error ratio between these two values is shown using the triangle marker on the right y-axis. As expected, the initial forecasts are quite far from the actual values. However, as more values are used for the smoothing process, the prediction improves significantly. Occasionally, when there are sudden changes in the temperature with steep vertical rises or drops, the error ratio becomes quite large indicating the need for the affected node to report back to the sink. Otherwise, it is evident that the chosen variance estimator algorithm accurately predicts the naturally generated data.

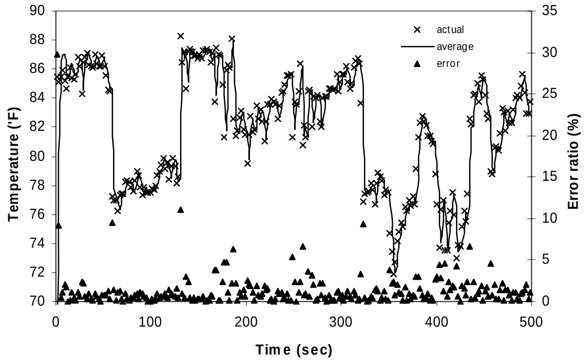

Figure 6. Actual and predicted temperature data values of a single node shown on the primary y-axis against simulation time. The secondary y-axis shows the error ratio between the two values.

To observe the effect of the temporal correlation suppression on the network, a plot of number of sleeping nodes against simulation time is shown with the error ratio in Fig. 7. When the error ratio is higher than the application bound, more nodes are involved in the data gathering process. However, when the prediction is rather accurate, more nodes are made to sleep as evident from the peaks of the continuous line in Fig. 7. For this application data and the chosen bound, the scheme is able to make more than 30% nodes to sleep on average. Even without any expensive spatial correlation exercise, it is evident that we have the potential to make significant energy savings due to the inherent temporal data correlations in this naturally generated data.

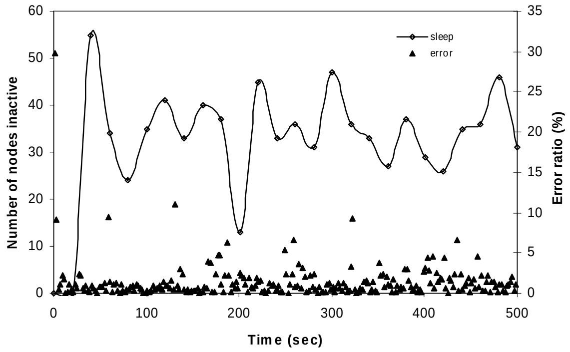

Figure 7. Number of inactive nodes (i.e. sleeping) shown on the primary y-axis against simulation time. The secondary y-axis shows the error ratio between the two values.

Finally, to investigate the effect of data sensitivity on the network performance, we vary the application bound. As expected, Fig. 8 confirms that the energy cost reduces with the increase in the application bound. As the application bound is increased, the energy cost reduces exponentially. When the application is lesser sensitive to minor changes, there would be many more nodes inactive in the network with lesser radio usage. Thus, the proposed scheme exploits the increased insensitivity of the application by making more nodes to sleep.

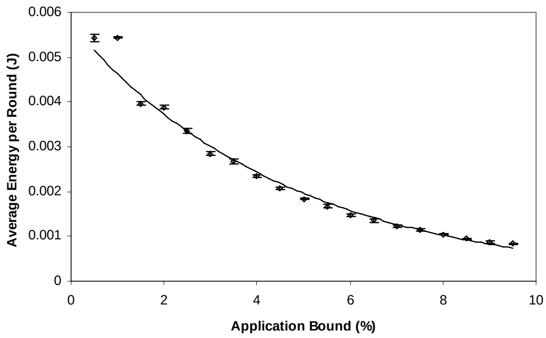

Figure 8. Average energy cost per data collection round against the application bound.

VI. CONCLUSIONS

To our knowledge, the T-ANT is the first nature-inspired approach for data gathering in wireless sensor networks. The algorithm uses a swarm of ants to control the clusterhead election in a distributed manner. It is shown that T-ANT achieves two desirable swarm behaviors, namely separation and alignment. Due to these, a uniform distribution of clusterhead is guaranteed enabling the network to operate in an optimal manner throughout its lifetime. Even though this is possible in a centralized approach as in LEACH-C [3], our algorithm is distributed, robust and does not require position knowledge. T-ANT also stores less state overhead in memory than LEACH or TCCA.

The T-ANT protocol is also able to exploit the inherent data correlations in the sensed data signals. To avoid the amount of state necessary to capture the spatial correlation among neighbors, we resorted to capture temporal correlation only. This involves only local decision-making. The variance estimation algorithm introduced here captures sensor data variability with negligible state maintenance. It is demonstrated that T-ANT with data redundancy detection achieves significant energy savings for periodic monitoring applications.

ACKNOWLEDGMENT

This work was supported in part by grants from ARC DP0664782 and USyd R&D L2844 U3230.

REFERENCES

[1] J. Hill, M. Horton, R. Kling, and L. Krishnamurthy, "The platforms enabling wireless sensor networks," *Comms. of the ACM*, vol. 47, pp. 41 - 46, Jun 2004.
[2] K. Akkaya and M. Younis, "A Survey on Routing Protocols for Wireless Sensor Networks," *Elsevier Ad Hoc Networks*, vol. 3, pp. 325-349, May 2005.
[3] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. on Wireless Comm.*, vol. 1, pp. 660 - 670, Oct 2002.
[4] S. Selvakennedy and S. Sinnappan, "The Time-Controlled Clustering Algorithm for Optimized Data Dissemination in Wireless Sensor Networks," in Proc. IEEE Conference on LCN, Sydney, Australia, 2005, pp. 509-510.

[5] S. Selvakennedy, S. Sinnappan, and Y. Shang, "Data Dissemination Based on Ant Swarms for Wireless Sensor Networks," in Proc. IEEE CCNC, Las Vegas, NV, USA, 2006, pp. 132-136.
[6] M. Gerla and J. T.-C. Tsai, "Multicluster, mobile, multimedia radio network," *Wireless Networks*, vol. 1, pp. 255-265, Aug 1995.
[7] F. Kuhn, T. Moscibroda, and R. Wattenhofer, " Initializing newly deployed ad hoc and sensor networks," in Proc. 10th Int. conference on Mobile computing and networking, Philadelphia, PA, USA, Sept 2004, pp. 260-274.
[8] D. Baker and A. Ephremides, "The Architectural Organization of a Mobile Radio Network via a Distributed Algorithm," *IEEE Trans. on Communication*, vol. 29, pp. 1694-1701, Nov 1981.
[9] G. Gupta and M. Younis, "Load-balanced clustering of wireless sensor networks," in Proc. IEEE ICC, Anchorage, AL, USA, May 2003, pp. 1848-1852.
[10] A. D. Amis and R. Prakash, "Load-balancing clusters in wireless ad hoc networks," in Proc. 3rd IEEE Symposium on Application-Specific Systems and Software Engineering Technology, Richardson, TX, USA, Mar 2000, pp. 25-32.
[11] A. D. Amis, R. Prakash, T. H. P. Vuong, and D. T. Huynh, "Max-min d-cluster formation in wireless ad hoc networks," in Proc. 19th Joint Conference of the ICC, Tel Aviv, Israel, Mar 2000, pp. 32-41.
[12] C.-F. Chiasserini, I. Chlamtac, P. Monti, and A. Nucci, "An energy-efficient method for nodes assignment in cluster-based ad hoc networks," *Wireless Networks*, vol. 10, pp. 223-231, May 2004.
[13] S. Bandyopadhyay and E. J. Coyle, "An energy efficient hierarchical clustering algorithm for wireless sensor networks," in Proc. 22nd IEEE INFOCOM, San Francisco, CA, USA, Mar 2003, pp. 1713 - 1723.
[14] M. Genes, U. Sorges, and I. Bouazizi, "ARA - the ant-colony based routing algorithm for manets," in Proc. ICPP Workshop on Ad Hoc Networks, Vancouver, BC, Canada, Aug 2002, pp. 79-85.
[15] J. Chou, D. Petrovic, and K. Ramachandran, "A distributed and adaptive signal processing approach to reducing energy consumption in sensor networks," in Proc. INFOCOM, 2003, pp. 1054-1062.
[16] P. v. Rickenbach and R. Wattenhofer, "Gathering correlated data in sensor networks," presented at Workshop on Foundations of mobile computing, Philadelphia, PA, USA, 2004, pp. 60-66.
[17] R. Cristescu and M. Vetterli, "Power efficient gathering of correlated data: optimization, NP-completeness and heuristics," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 7, pp. 31-32, 2003.
[18] S. Yoon and C. Shahabi, "Exploiting spatial correlation towards an energy efficient clustered aggregation technique (CAG)," in Proc. IEEE International Conference on Communications, 2005, pp. 3307-3313.
[19] H. Gupta, V. Navda, S. R. Das, and V. Chowdhary, "Efficient gathering of correlated data in sensor networks," in Proc. MobiHoc '05, Urbana-Champaign, IL, USA, 2005, pp. 402-413.
[20] B. Chen, K. Jamieson, H. Balakrishnan, and R. Morris, "Span: an energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks," *Wireless Networks*, vol. 8, pp. 481-494, 2002.
[21] A. Cerpa and D. Estrin, "ASCENT: Adaptive Self-Configuring sEnsor Networks Topologies," *IEEE Trans. on Mobile Computing*, vol. 3, pp. 272 - 285, 2004.
[22] V. Ramos and J. J. Merelo, "Self-organized stigmergic document maps: Environment as mechanism for context learning.," in Proc. Spanish Conference on Evolutionary and BioInspired Algorithms, 2002, pp. 284–293.
[23] Y. Ohtaki, N. Wakamiya, M. Murata, and M. Imase, "Scalable ANT-based Routing Algorithm for Ad-Hoc Networks," in Proc. 3rd IASTED International Conference on Communications, Internet, and Information Technology, St. Thomas, US Virgin Islands, Nov 2004.
[24] V. Paxson and M. Allman, "Computing TCP's Retransmission Timer," *RFC2988*, 2000.
[25] Q. Huang and Y. Zhang, "Radial coordination for convergecast in wireless sensor networks," in Proc. 29th IEEE International Conference on Local Computer Networks, Tampa, FL, USA, Nov 2004, pp. 542-549.
[26] NESDIS, "http://www.nesdis.noaa.gov/datainfo.html," *US National Oceanic & Atmospheric Administration*, 2005.

S. Selvakennedy obtained his PhD in computer communications in 1999 from the University of Putra, Malaysia. Earlier in 1996, he obtained his first-class honors degree in computer science also from the same university.

He is currently serving as a Senior Lecturer at the University of Sydney, Australia. He has published more than 50 articles in journals and conferences mainly in the area of communications protocols for all-optical networking and wireless networking. His current research interests lies in developing algorithms for media access, routing, localization and topology control issues in wireless sensor, ad hoc and mesh networks.

Dr. Selvakennedy is a member of ACM and IEEE. He has served on many conference program committees.

Sukunesan Sinnappan got his doctorate from the University of Newcastle, Australia in 2004. He is currently with the School of Economics and Information Systems, University of Wollongong, Australia. He has published in numerous fields such as online consumer behavior, website analysis and wireless sensor networks. His current research interests include online self-profiling, location-based services, online privacy and trust apart from wireless sensor networks.

Yi Shang was born in China in 1967. He received the PhD degree in Computer Science from the University of Illinois at Urbana-Champaign, Urbana, IL, USA, in 1997; MS in Computer Engineering from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 1991; and BS in Computer Science from University of Science and Technology of China, Hefei, China, in 1988. His research fields include artificial intelligence, nonlinear optimization, wireless networks, and distributed systems.

He is an Associate Professor in the Department of Computer Science at the University of Missouri at Columbia, USA. He was a senior research scientist at the Xerox Palo Alto Research Center at Palo Alto, CA from 2001 to 2003. His current research interests are wireless sensor networks and multi-agent systems. He has published more than 80 technical papers and has 6 patent applications pending.

Dr. Shang is a member of ACM and IEEE. He has been program committee member of many conferences.