Journal of Zhejiang University-SCIENCE C (Computers & Electronics)
ISSN 1869-1951 (Print); ISSN 1869-196X (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Congestion avoidance, detection and alleviation in wireless sensor networks*

Wei-wei FANG[†1], Ji-ming CHEN[2], Lei SHU[3], Tian-shu CHU[1], De-pei QIAN[1]

(*[1]Sino-German Joint Software Institute, School of Computer Science, Beihang University, Beijing 100191, China*)

(*[2]State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China*)

(*[3]Digital Enterprise Research Institute, National University of Ireland, Galway, Ireland*)

[†]E-mail: fangvv@gmail.com

Received Apr. 13, 2009; Revision accepted Aug. 10, 2009; Crosschecked Sept. 29, 2009

Abstract: Congestion in wireless sensor networks (WSNs) not only causes severe information loss but also leads to excessive energy consumption. To address this problem, a novel scheme for congestion avoidance, detection and alleviation (CADA) in WSNs is proposed in this paper. By exploiting data characteristics, a small number of representative nodes are chosen from those in the event area as data sources, so that the source traffic can be suppressed proactively to avoid potential congestion. Once congestion occurs inevitably due to traffic mergence, it will be detected in a timely way by the hotspot node based on a combination of buffer occupancy and channel utilization. Congestion is then alleviated reactively by either dynamic traffic multiplexing or source rate regulation in accordance with the specific hotspot scenarios. Extensive simulation results under typical congestion scenarios are presented to illuminate the distinguished performance of the proposed scheme.

Key words: Wireless sensor network (WSN), Congestion control, Correlation, Traffic multiplexing, Rate regulation
doi:10.1631/jzus.C0910204 **Document code:** A **CLC number:** TP393

1 Introduction

Technological advances in microelectromechanical systems and wireless communications have motivated the development of wireless sensor networks (WSNs) (Akyildiz *et al.*, 2002; Zhang *et al.*, 2007). In a WSN, a large number of sensor nodes are densely deployed in the field of interest to sense the physical phenomenon and to report the event through wireless links to one or more sinks. Once a target event occurs, a sudden surge of data traffic will be triggered by all sensor nodes in the event area, which may easily lead to network congestion when offered traffic load exceeds practical network capacity (Gupta and Kumar, 2000).

Congestion in WSNs has negative impacts on network performance and application objective, i.e., indiscriminate packet loss, increased packet delay, wasted node energy and severe fidelity degradation. However, some unique characteristics of WSNs, such as constrained resources, interference-coupled paths and the lack of centralized coordination, make the congestion problem in WSNs more challenging than in traditional networks. Moreover, congestion control in WSNs has to consider not only the network capacity, but also the application requirements on information fidelity (Zhou *et al.*, 2005; Chen *et al.*, 2008).

To address these challenges, a simple yet effective scheme CADA (congestion avoidance, detection and alleviation) is proposed in this paper. As its name implies, the scheme contains relevant mechanisms for avoiding congestion proactively, detecting congestion timely and alleviating congestion reactively. Control

* Project supported by the National Natural Science Foundation of China (Nos. 60673180, 90412011 and 90612004), the International Science and Technology Cooperative Program of China (No. 2006DFA11080), the Research Program of Federal Ministry of Education and Research of Germany (No. 01BU0680), and the Lion Project of Science Foundation of Ireland to Lei Shu (No. SFI/08/CE/I1380)

operations are performed by sensor nodes in a distributed manner without requiring for sinks' participation. Simulation results validate the distinguished performance in several aspects of the scheme.

2 Related works

In literature, many congestion control schemes have been proposed for wireless ad-hoc networks (Chen *et al*., 2006; Yu and Giannakis, 2009). However, these works involve complicated computation for resource allocation and utilization, which renders their application impractical for resource-constrained WSNs. Meanwhile, they did not take the data fidelity requirements into consideration when performing the control operations.

In recent years, several new congestion control solutions have been studied for WSNs (Fang *et al*., 2008). Among them, some prior work focuses on traffic reduction for congestion control in the WSN. Congestion detection and avoidance (CODA) (Wan *et al*., 2003) is the first detailed investigation on congestion control in WSNs, which combines local backpressure techniques and centralized sink-to-sensors notifications but is not specifically concerned with different classes of traffic flows. Congestion control and fairness (CCF) (Cheng and Bajcsy, 2004) controls congestion in a hop-by-hop manner and each node adjusts rate based on its available service rate and child node number. However, the rate adjustment in CCF relies only on packet service time, which could lead to low utilization when some nodes do not have enough traffic or the packet error rate is high. In event-to-sink reliable transport (ESRT) (Akan and Akyildiz, 2005), the sink is required to regulate the source reporting rate in an undifferentiated manner by broadcasting control messages to all source nodes. The underlying assumption is that a sink can reach all nodes via a high-energy one-hop broadcast, which is not practical for a large-scale sensor network. Congestion control for sensor networks (CONCERT) (Galluccio *et al*., 2005) uses the adaptive data-aggregation technique to reduce the amount of information traveling throughout the network. This leverages a unique characteristic of WSNs to handle the congestion problem. Interference-minimized multipath routing (I2MR) (Teo *et al*., 2008) integrates source rate adaptation and multipath routing for congestion control in WSNs. Learning automata-based congestion avoidance scheme (LACAS) (Misra *et al*., 2009) avoids congestion by using a learning automata based approach. It adaptively makes the data packet arrival rate in the nodes equal to the transmitting rate, so that the occurrence of congestion in the node can be seamlessly avoided. However, it must be noted that pure traffic reduction could impose a negative impact on data fidelity.

Apart from the schemes based on traffic control, there have been attempts to explore other mechanisms for congestion alleviation in WSNs. Siphon (Wan *et al*., 2005) proposes to add multi-radio virtual sinks to the network as a means of dealing with congestion. When congestion occurs, Siphon redirects traffic off the primary low-power radio network and onto the overlay network with long-range radio. The cost for adding and using the long-range radio is not negligible. Biased geographical routing (BGR) (Popa *et al*., 2006) is a geographic routing that reactively split traffic during congestion. However, the bias, which determines how far the trajectory of splitting traffic will deviate from the original path, is randomly chosen and could make congestion worse under some situations. In contrast, topology-aware resource adaption (TARA) (Kang *et al*., 2007) adopts different traffic multiplexing strategies depending on specific topologies. It requires knowledge about the local and the end-to-end topology for capacity estimation, which causes too high an overhead for a large-scale network. Congestion aware routing (CAR) (Kumar *et al*., 2008) proposes to use a priority aware routing protocol with data prioritization to alleviate congestion. Congestion zones are dedicated for high priority data while other data can only be routed out of the congestion zones. Multiple sinks are required to be deployed on the area border for gathering data with different priorities.

3 Congestion control in WSN

A WSN composed of a large number of sensor nodes and a set of sinks in an interested area is considered in this work. Sensor nodes are randomly distributed in the area and remain stationary after deployment. All of them have similar capabilities and equal significance. However, they are constrained in memory space, processing capability, communication bandwidth and energy storage. Sensor nodes and the sink communicate via bidirectional multihop wireless links. Sinks are uniformly scattered across the sensing field. They have more powerful resources to perform data gathering and network management tasks. For the same reason as described in (Wan *et al.*, 2003; Eisenman and Campbell, 2007), all nodes are assumed to implement a carrier sense multiple access (CSMA)-like medium access control (MAC) protocol for data transmission. According to some preset rules (such as event signal type and event spot location), the source node can assign a corresponding priority to its packets based on the importance of the event to be reported (Kumar *et al.*, 2008). Traffic flows initiated will remain active long enough to contribute to changes in the network load.

3.1 Congestion scenarios in WSNs

Consider a WSN consisting of a relatively large number of sensor nodes that collectively transmit sensory data to a small number of sinks. If no control mechanism is adopted in the transmission, congestion hotspots would probably form in the network when offered traffic load exceeds practical network capacity. Generally, there exist three typical hotspot scenarios in WSNs, which are described as follows:

1. Source hotspot (Fig. 1a): The occurrence of an event will be detected by all nodes whose sensing ranges cover the event spot. If all these nodes perform as sources to transmit data towards the sink, a traffic hotspot may quickly form around that event spot. Furthermore, uncontrolled initiation of traffic from sources in this manner could also increase the probability of formation of the following hotspots.

2. Convergence hotspot (Fig. 1b): A sink can subscribe sensing services provided by sensor nodes at different locations in the area. The multi-hop and many-to-one nature of data traffic in the WSN lead to increased traffic intensity as data packets move closer towards the sink. Thus, congestion may be caused by traffic convergence near the sink (Wan *et al.*, 2005).

3. Intersection hotspot (Fig. 1c): The presence of multiple sinks (Das and Dutta, 2005) can result in traffic intersection in the network. Due to the excess of the aggregate traffic load, the intersection nodes can also become congested hotspots.

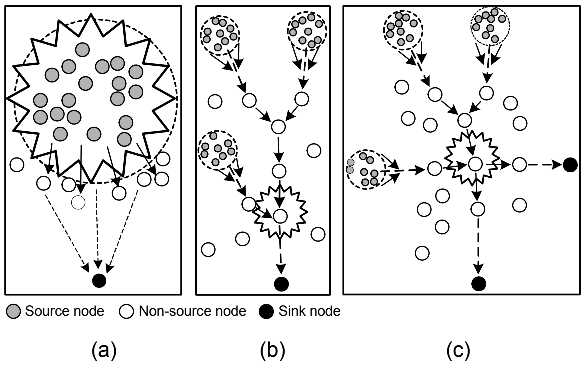


Fig. 1 The hot spot scenarios
(a) Source hotspot; (b) Convergence hotspot; (c) Intersection hotspot

Each of these scenarios is harmful to the operations of a WSN. In the following subsections, we will elaborate on the CADA design for handling congestion problems in WSNs.

3.2 Congestion avoidance in CADA

The main objective of the WSN is to reliably obtain event information from the collective reports provided by relevant sensor nodes in the event area. To achieve this goal, it may not be necessary for every sensor node that detects the same event to report to the sink independently. Instead, a subset of nodes, referred to as representative nodes, can be chosen to become data sources as long as acceptable data fidelity can be achieved at the sink (Vuran and Akyildiz, 2006). That is, the other nodes generating redundant or inaccurate data are suppressed from reporting to the sink. Therefore, the overall traffic load from the event area can be properly reduced, which in turn reduces the congestion probability (Kang *et al.*, 2007).

Before describing design details, the criteria for choosing representative nodes are investigated.

Criterion 1 (Data accuracy) The sensor works by measuring the energy of signal emitted by the target. Suppose sensor node i is d_i away from the event spot. The signal power it measures can be given by

$$X_i = S \cdot d_i^{-k} + n_i,\ 1 \le i \le N, \quad (1)$$

where S is the (unknown) target signal power, k is the (known) attenuation coefficient, and n_i is the white Gaussian noise with zero-mean and variance σ^2 (Chen *et al.*, 2004). From Eq. (1), the farther a node is from the target, the more likely its data accuracy would be affected by the noise. Therefore, sensor nodes that are far away from the event spot should be prevented from reporting data to the sink (Zhao *et al.*, 2007). The threshold distance, D_t, can be determined through experiments before network deployment.

Criterion 2 (Data redundancy) Due to high density in the network topology, spatially proximal sensor observations are highly correlated, which results in considerable data redundancy in the network. The spatial correlation can be modeled as a non-negative, monotonically decreasing function of the internode distance d, with limiting values of 1 at d=0 and of 0 at $d=\infty$ (Berger *et al.*, 2001). Instead of letting all nodes in the event area report simultaneously, representative nodes that are farther apart from each other can be chosen as data sources. The threshold of the correlation degree, C_t, should be preset according to the fidelity requirement. A smaller threshold value increases the observation fidelity, but it also increases the number of representative nodes as well as incurs higher traffic load and more energy consumption.

To sum up, a node i in the representative node set P must fulfill the following requirements:

$$i \in P \text{ iff } \forall j \in P, d_i \le D_t, C(i, j) \le C_t. \tag{2}$$

Due to resource constraints and scalability issues, the centralized solutions for representative node selection (Vuran and Akyildiz, 2006; Liu *et al.*, 2007) are impractical in WSNs. Hence, we propose a distributed node selection algorithm in CADA, as shown in Fig. 2.

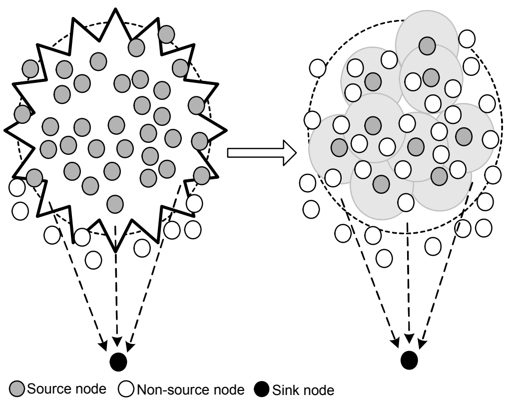


Fig. 2 Representative node selection in the event area

1. When an event occurs, all sensor nodes located in the range of [0, D_t] keep on sensing for a short period of time t_1. Each of them generates a time series of sampling data with series length L.

2. Each node sets a back-off timer for announcing itself to be a representative node. Ideally, a node has a higher priority than its neighbors, if it is closer to the event spot and has more residual energy. Thus, the back-off time of node i, T_i, can be given as

$$T_i = \left[W_d \cdot \frac{d_i}{D_t} + W_e \cdot \left(1 - \frac{e_i}{e}\right) + W_r \cdot r \right] \cdot t_2, \tag{3}$$

where the weights satisfy

$$W_d + W_e + W_r = 1,\ W_d > W_e > W_r. \tag{4}$$

In Eq. (3), e_i is the current residual energy at node i, e is a reference maximum energy (corresponding to a fully charged battery), and r is a random value uniformly distributed in the range (0, 1).

3. When the back-off timer expires, the sensor node announces itself to be a representative node by broadcasting a message containing the sampled data series. If by the time the timer expires this node overhears a broadcast message from its neighbors, it calculates the correlation degree between the received data series and that of its own. For two data series X and Y, the correlation degree is

$$C_{XY} = \left| \frac{\sum_{i=1}^{L} \left(X_i - E(X)\right) \cdot \left(Y_i - E(Y)\right)}{L \cdot \sqrt{E(X^2) - E^2(X)} \cdot \sqrt{E(Y^2) - E^2(Y)}} \right|. \tag{5}$$

Eq. (5) is derived from the textbook definition of the correlation between two series (Casella and Berger, 2001; Gedik *et al.*, 2007). Note that $0 \le C_{XY} \le 1$, with 0 and 1 representing the weakest and strongest correlations, respectively. If the node finds $C_{XY} \ge C_t$, it cancels its timer and will not be a data source afterwards; otherwise, the overheard message is ignored by this node. To reduce the computational overhead, C_{XY} is calculated only if the distance between two nodes is greater than a given distance threshold (Liu *et al.*, 2007).

4. Each sensor node in the event area maintains another timer for updating the set of representative

nodes periodically. When this update timer expires, the above steps are repeated to select a new collection of representative nodes. The node updating balances the workload distribution among all sensor nodes in the event area, and can also guarantee the observation fidelity in case of node failure or event change.

The node selection algorithm presented above is completely distributed. Suppose there are n nodes in an event area. This algorithm has a worst-case time complexity of $O(n)$ and a worst-case message exchange complexity of $O(1)$ per node, i.e., $O(n)$ for all nodes in the event area.

By choosing a number of representative nodes to be data sources, the source traffic can be suppressed to avoid potential congestion without degrading the fidelity achieved at the sink. Meanwhile, the traffic reduction decreases not only interference, contention and collision in the wireless medium, but also energy consumption for data packet transmission. However, it is only a proactive method for congestion avoidance. Multiple simultaneous traffic flows originated from different event areas might occasionally induce congestion due to gradually traffic merging on the routing paths (Kang *et al*., 2007). More work should be done for congestion detection and alleviation.

3.3 Congestion types in WSNs

Congestion in WSNs is slightly different from that of wired network. Two types of congestion can generally occur in WSNs (Cheng and Bajcsy, 2004):

1. Node-level congestion: Within a particular node, the buffer used to hold packets to be transmitted overflows. This is the conventional definition of congestion in a wired network. This type of congestion can result in increased queuing delay, indiscriminate packet loss and additional retransmission overhead.

2. Link-level congestion: In a particular area, severe collisions could occur when multiple active sensor nodes within range of one another attempt to transmit simultaneously. Packets that leave the buffer might fail to reach the next hop as a result of collision. This type of congestion decreases both link utilization and overall throughput, while increasing both packet delay and energy waste.

It is possible to have node- and link-level congestion occurring at the same time in a WSN. Both of them have direct impacts on network performance and application objective. Thus, the onset of congestion must be predicted in advance or detected in time so that it can be alleviated thereafter.

3.4 Congestion detection in CADA

For effective and efficient congestion control, it is necessary to accurately measure the congestion level in hotspot areas. In CADA, a node's congestion level is measured by an aggregation of the two metrics: buffer occupancy and channel utilization (Hull *et al*., 2004; Kang *et al*., 2007).

1. Buffer occupancy monitoring: A sensor node periodically checks for the buffer occupancy and its growth trend. Let b_k and b_{k-1} be the buffer occupancy at the end of kth and $(k-1)$th intervals respectively. The increment Δb_k can be calculated by an exponential weighted moving average (EWMA) approach with b_k and b_{k-1}, that is

$$\Delta b_k = (1-w)\cdot\Delta b_{k-1} + w\cdot(b_k - b_{k-1}), \qquad (6)$$

where $0<w<1$ is the EWMA factor. If $b_k+\Delta b_k$ exceeds the high watermark of buffer size, the sensor node infers that it will experience the node-level congestion in the next interval.

2. Channel utilization sampling: A sensor node is triggered to measure the channel loading when its buffer is not empty, as was done in CODA (Wan *et al*., 2003). If the node finds that the packet delivery ratio decreases drastically while the local channel loading reaches the maximum achievable channel utilization, it infers the onset of congestion. Interested readers can also refer to Wan *et al*. (2003) for further details.

3.5 Congestion alleviation in CADA

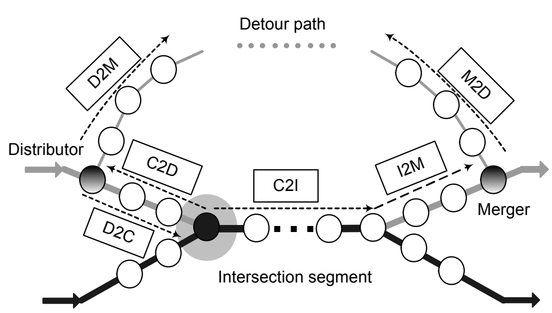
Intuitively, there are two general methodologies for alleviating congestion in WSNs: resource control or traffic control (Kang *et al*., 2007). The first method tries to alleviate congestion by exploiting available network resources to transiently accommodate the traffic surge. For example, detour paths can be established to redirect some traffic away from the traffic hotspot. This method does not affect data reporting at source nodes, so the information fidelity may be well maintained at the sink. However, it is not suitable for the case of convergence hotspot, since congestion may reappear somewhere nearer to the sink. Unlike resource control, traffic control implies that it will alleviate congestion by adjusting the traffic rate at the source nodes or the intermediate nodes. This method is more effective and feasible when resource control is not applicable, but might degrade the information fidelity at the sink to some extent.

The above two methodologies are both adopted in CADA for congestion alleviation. Each node keeps track of the state information of the incoming traffic, denoted as $\sum f_{i\to j}$ ($i\in$\{source\}, $j\in$\{sink\}), by which the node can judge the hotspot type when in congestion. Generally, the congestion in an intersection hotspot ($|\{j|\exists f_{i\to j}\}|>1$) is mitigated by resource control, while the congestion in a convergence hotspot ($|\{j|\exists f_{i\to j}\}|=1$) is mitigated by traffic control.

3.5.1 Resource control in CADA

Building multiplexing paths for redirecting traffic to bypass the hotspot is a challenging task. If the detour path is too close to the congestion area, it can further interfere with the already bad congestion. In the wireless network, the interference range of a node is typically at least twice as large as its transmission range. Therefore, the detour path has to be located at least two hops away from the original intersecting traffic (Jain *et al.*, 2003). However, if the detour path is too far away, the packet delivery latency and the energy consumption will increase accordingly.

The traffic multiplexing process is described as follows and also illustrated with the example scenario in Fig. 3. Initially, the congested node selects the upstream neighbor that injects the most packets, and sends a congestion-to-distributor (C2D) control message to that neighbor node. The C2D message is propagated until it reaches the upstream node two hops away from the congested node. That upstream node will then become the traffic distributor, and send a distributor-to-congestion (D2C) control message back to the congested node. The D2C message contains the ID s of the sink to which traffic packets with the lowest priority are destined. If the D2C message fails in transmission, the congested node will retry for several times. Otherwise, it will start to locate a traffic merger by sending a congestion-to-intersection (C2I) control message to the sink s.



C2D: congestion-to-distributor; D2C: distributor-to-congestion; C2I: congestion-to-intersection; I2M: intersection-to-merger; M2D: merger-to-distributor; D2M: distributor-to-merger

Fig. 3 Illustration of resource control in CADA

The C2I message is propagated along the traffic intersection segment, which includes all the nodes the intersecting flows have in common on their routing paths. This message records the total hop count of the segment h, which is updated by all forwarding nodes. When the node at the end of the intersection segment receives this message, it broadcasts the message and stops forwarding. Then the node sends an intersection-to-merger (I2M) control message to the downstream node that is located two hops away on the split routing path to the sink s. That downstream node will become the traffic merger.

The traffic merger tries to establish the detour path by locally flooding a merger-to-distributor (M2D) control message towards the distributor. The TTL of this message is set as $\alpha\cdot h$, where $\alpha>1$ is preset according to the node density. To avoid the interference between the detour path and the original path, it is required that all nodes overhearing the C2I message transmission broadcast a special C2I message with TTL=0. Then the nodes that have received the C2I message or are already on the original path ignore the received M2D messages, so they will not be involved in the detour path. When an M2D message reaches the distributor, a candidate detour path can be established. The distributor chooses a best path from multiple candidate paths by sending a distributor-to-merger (D2M) control message towards to merger through the selected path. To alleviate congestion, the distributor can split the traffic and routed packets with relatively low priority via the detour path.

To improve control efficiency for the congestion, it is necessary to require that all the control messages with the highest priority be processed ahead of the other packets by the intermediate nodes.

All the information maintained by each node for resource control is based on soft-state with the control timer. The timer at each node silently expires except for the distributor. In some cases, if the network fails to find an available detour path due to some reason, the distributor will send a D2C message to notify the congested node of this failure. In some other cases, the aggregate traffic load may be so high that the congestion cannot be resolved completely by traffic multiplexing. In these two cases, the network can resort to only the traffic control approach, which is presented in the next subsection.

3.5.2 Traffic control in CADA

Whenever a congested node determines to apply traffic control to current data sources, it piggybacks a binary feedback C_g and the ID of the target sink in the acknowledgement (ACK) frame. The notification information will be kept on being propagated to upstream nodes until it arrives at all source nodes. This implicit notification ensures that the feedbacks will be successfully received by the source nodes. The AIMD-like source reporting rate regulation policy is defined as

$$R_i(t+\Delta t_i)=\begin{cases} R_i(t)+\Delta r_i, & R_i(t)<R_{\max}, C_{\mathrm{g}}=0, \\ R_i(t), & R_i(t)=R_{\max}, C_{\mathrm{g}}=0, \\ R_{\min_i}, & C_{\mathrm{g}}=1, \end{cases} \quad (7)$$

where $R_{\min_i}$ is the tolerable minimal reporting rate for the node with priority i, i is the priority level selected by the node for its packets, Δr_i is a constant for rate increment, Δt_i is a constant for update interval, and $R_{\max}$ is the default reporting rate. For the priority set P, if $\forall i, j \in P, i<j$, it is required that $\Delta r_i \le \Delta r_j$, $\Delta t_i \ge \Delta t_j$, $R_{min_i} \le R_{min_j}$. This indicates that higher important packets will receive preferential control treatment as compared with lower important packets. All these values can be determined based on a few initial experiments via simulation and can be set before the deployment is in full-fledged operation.

Since such a traffic control approach may degrade the application fidelity to some extent, we can use the piecewise linear approximation technique, such as the PLAMLiS algorithm (Liu *et al*., 2007), to approximate the originally collected data series. This operation is applied by each source node to the latest sensory data of its own. More representative data samples will be selected by the node as event reports sent to the sink to prevent the loss of so much fidelity.

4 Simulation and analysis

We evaluated the performance of the CADA by means of simulation performed with the ns-2 simulator (http://nsnam.isi.edu/nsnam/index.php/Main_Page). In the simulation, we adopted the energy model given in Heinzelman *et al*. (2002) for sending and receiving data with length l bits when d is the distance between transmitter and receiver:

$$E_{Tx}(l,d)=\begin{cases} l\cdot(E_{\mathrm{elec}}+\varepsilon_{\mathrm{fs}}\cdot d^2), & d<d_0, \\ l\cdot(E_{\mathrm{elec}}+\varepsilon_{\mathrm{mp}}\cdot d^4), & d\ge d_0, \end{cases} \quad (8)$$

$$E_{Rx}(l)=l\cdot E_{\mathrm{elec}}, \quad (9)$$

where E_{elec} is the electronics energy, $\varepsilon_{\mathrm{fs}}\cdot d^2$ and $\varepsilon_{\mathrm{mp}}\cdot d^4$ are the amplifier energy in the free space and in the multi-path fading environment respectively, and d_0 is a distance threshold that depends on the environment. The data correlation between nodes i and j was set according to the power exponential model given in Vuran and Akyildiz (2006):

$$C(i,j)=\exp((-d_{ij}/\theta_1)^{\theta_2}),\ \theta_1>0,\ \theta_2\in(0,2]. \quad (10)$$

The fidelity provided by H data source nodes for the same event (Zhao *et al*., 2007) was calculated as

$$f=1-\left[\left|S-\sum_{m=1}^{H}\left(d_m^k\Big/\sum_{n=1}^{H}d_n^k\right)\cdot X_m\cdot d_m^k\right|\Big/S\right]. \quad (11)$$

The simulation parameters are listed in Table 1.

4.1 Performance evaluation

The effect of reducing source traffic by representative node selection was studied in the first set of experiments. A total of N sensor nodes were assumed to be uniformly deployed in a 500 m×500 m square area, with the target event in the center. As shown in Fig. 4, about half of the nodes (i.e., nodes in the event area) worked as data sources if no control operation was applied (C_t=1). By executing the proposed node selection algorithm, the total number of data source nodes could be greatly reduced with the decrease of C_t, especially when the node density is relatively high.

To study the effect on the data fidelity, we randomly chose a representative node in the experiments

Table 1 Common parameters used in all the experiments

Parameter	Value
Sensing range (m)	200
Transmission range (m)	100
MAC protocol	CSMA
MAC header size (bytes)	8
Buffer capacity (packets)	10
Bandwidth (Mbits/s)	2
Packet size (bytes)	30
Traffic type	Constant bit rate
E_{elec} (nJ/bit)	50
ε_{fs} (pJ·m^2/bit)	10
ε_{mp} (pJ·m^4/bit)	0.0013
D_t (m)	200
d_0 (m)	75
θ_1	100
θ_2	1
σ^2/S	1E−20

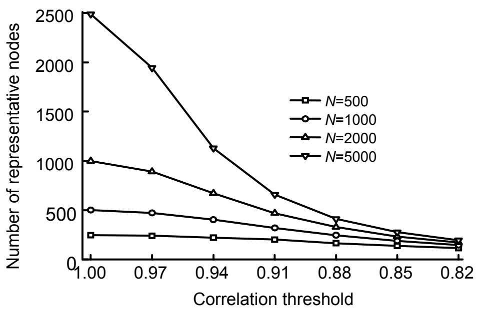


Fig. 4 Number of representative nodes vs. correlation threshold

with network size N=2000. Fig. 5 shows the comparison of data fidelity provided by a chosen node without and with its correlated neighbors (C_t=0.82). It is easy to know from Fig. 5 that source traffic can be reduced properly without much degradation of data fidelity. Due to the uncertainty of signal noise, a single representative node may even provide more accurate data results in some cases. The cost incurred by this algorithm is negligible compared to that saved from reduced source traffic.

An important step in congestion detection is to determine the congestion threshold at which control mechanisms should be started. It is easier to predefine a threshold for the buffer (e.g., 70% in this paper) than for the channel, since the latter is highly dependent on system capacity. Thus, the second set of experiments was performed on a simple 5-node cross topology (Wan *et al.*, 2003), as shown in Fig. 6. In the simulations, we found that a channel utilization of about 80% is where the channel saturates (Fig. 7) and suffers from frequent collisions between the neighboring nodes. Considering the time required for the network to take some actions to alleviate congestion, a channel utilization threshold of about 70% was used (Wan *et al.*, 2005) in the subsequent network simulations.

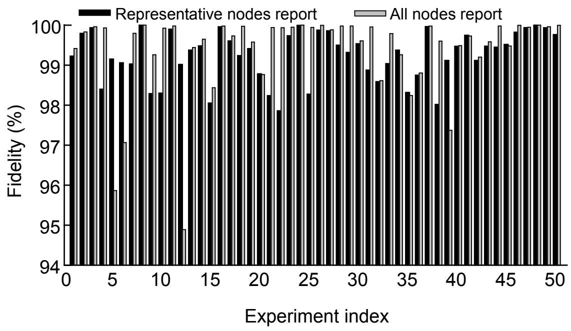


Fig. 5 Comparison of data fidelity achieved

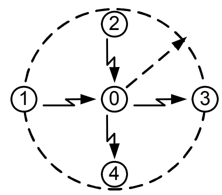


Fig. 6 A wireless network topology for measuring channel utilization

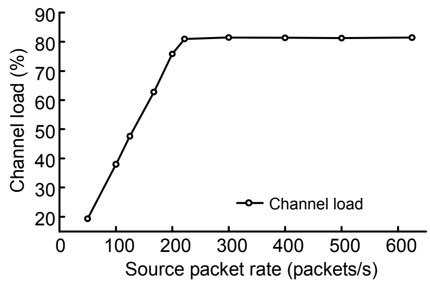


Fig. 7 Channel load measured at node 0

In the third set of experiments, different congestion scenarios were created by manual routing. Event spots were located at least 15 hops away from the target sink. The volume of source traffic was adjusted by tuning C_t. Traffic hotspots were formed at an intermediate node by traffic flows from 4 different event areas. Firstly, intersection congestion scenarios were formed by deploying 2 sinks at different locations. We compared CADA with TARA (Kang *et al.*, 2007) and NOCC (no congestion control). Fig. 8a plots the delivery ratio under a different offered load for each strategy. Both CADA and TARA can alleviate the congestion effectively. Since only one detour path was created in the simulations, both schemes' performance degraded somewhat when the load was so high that the original path got saturated again. However, CADA delivered much more packets than TARA. This is because a multiplexing path is created in TARA mainly based on the candidate node's congestion level, which cannot completely eliminate the interference between the original and the detour paths. The inter-path interferences lead to more packet loss and frequent link-layer retransmission, which in turn reduces the link transmission capacity. By contrast, CADA minimizes the inferences by setting up detour paths two hops away from the traffic hotspot. Thus, CADA also consumed the lowest energy (Fig. 8b) and achieved the lowest average per-hop delay (Fig. 8c) compared to TARA and NOCC. Secondly, only one sink was used to create convergence congestion scenarios in the network. The traffic priorities were set as high ($R_{\min S1}$=67 packets/s), middle ($R_{\min S2}$=$R_{\min S3}$=50 packets/s) and low ($R_{\min S4}$=40 packets/s). The throughput results for each traffic flow are plotted in Fig. 9a. We can observe that despite the fact that sources reduced their reporting rates for alleviating congestion, the rate never fell below the preset minimum rate. Meanwhile, CADA tries to provide differential delivery service for the different traffic by stimulating a proactive action at each source. In this way, smooth network operation is maintained without many energy-costly drops (Fig. 9b) or low delays for all packets are achieved (Fig. 9c). Though in CADA the total network throughput may be reduced to some extent by the traffic control method, we argue that it is worth while performing congestion alleviation operations as stated above. Besides, an end user should remember not to subscribe to so many data services that the overall traffic load may exceed the practical network capacity. Otherwise, he or she must suffer some

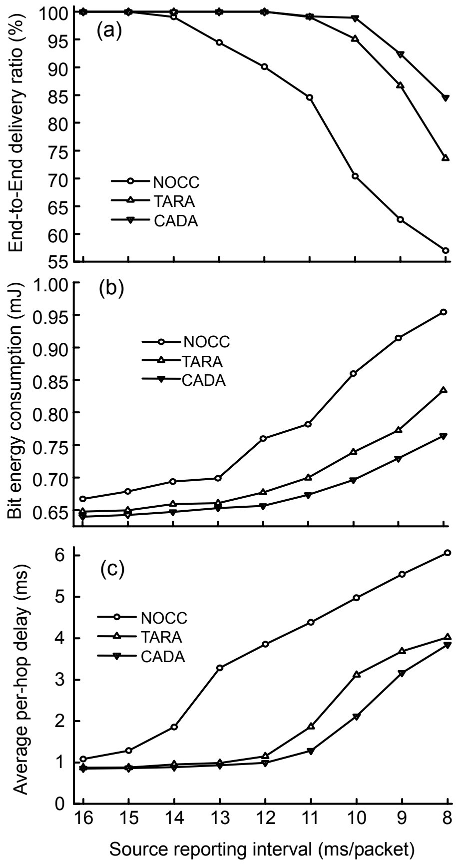


Fig. 8 (a) Delivery ratio, (b) bit energy consumption, and (c) average per-hop delay vs. source reporting interval

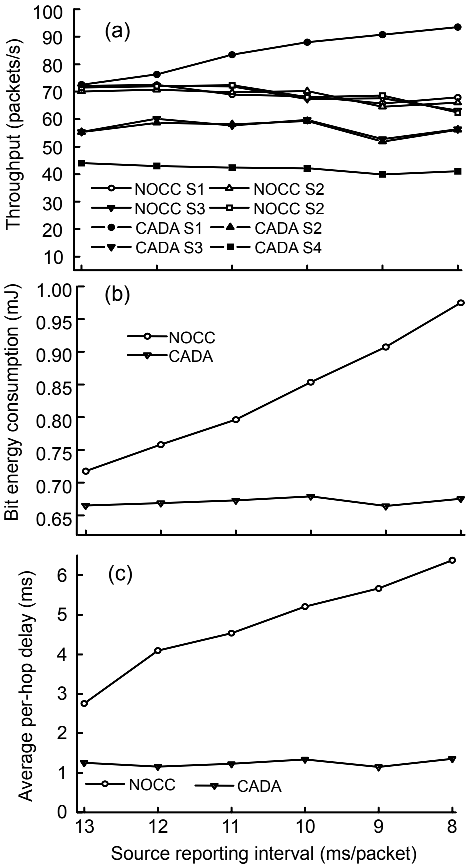


Fig. 9 (a) Throughput at sources, (b) bit energy consumption, and (c) average per-hop delay vs. source reporting interval

possible data fidelity degradation due to the network congestion. We have also proposed to use the piecewise linear approximation technique to reduce the effect of such degradation of data fidelity so as to present more accurate data to end users.

5 Conclusion

In this work, we propose a congestion control scheme CADA for congestion avoidance, detection and alleviation in wireless sensor networks. The key objective is to provide high transmission quality for the data traffic under conditions of congestion. The scheme comprises three main mechanisms. Firstly, it attempts to suppress the source traffic from event area by carefully selecting a set of representative nodes to be data sources. Secondly, the onset of congestion is indicated in a timely way by jointly checking buffer occupancy and channel utilization. Lastly, the network attempts to alleviate congestion in the traffic hotspot by either resource control or traffic control, which is dependent on the specific congestion condition. The ns-2-based simulation has confirmed the advantages of CADA and demonstrated significant performance improvements over existing schemes.

In the future, we hope to extend the proposed CADA scheme to take into account some dynamic behaviors of the network, such as node mobility and link failure. We also plan to extensively investigate CADA performance on our sensor network test-bed ProNet (Du *et al*., 2007).

References

Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E., 2002. A survey on sensor network. *IEEE Commun. Mag.*, **40**(8):102-114. [doi:10.1109/MCOM.2002.1024422]

Akan, O.B., Akyildiz, I.F., 2005. Event-to-sink reliable transport in wireless sensor networks. *IEEE Trans. Network.*, **13**(5):1003-1016. [doi:10.1109/TNET.2005.857076]

Berger, J.O., Oliveira, V.D., Sanso, B., 2001. Objective Bayesian analysis of spatially correlated data. *J. Am. Statist. Assoc.*, **96**(456):1361-1374. [doi:10.1198/016214501753382282]

Casella, G., Berger, R.L., 2001. Statistical Inference. Duxbury Press, CA, USA, p.139-203.

Chen, L., Szymanski, B.K., Branch, J.W., 2008. Quality-Driven Congestion Control for Target Tracking in Wireless Sensor Networks. Proc. 5th IEEE Int. Conf. on Mobile Ad Hoc and Sensor Systems, p.766-771. [doi:10.1109/MAHSS.2008.4660115]

Chen, L.J., Low, S.H., Chiang, M., Doyle, J.C., 2006. Cross-Layer Congestion Control, Routing and Scheduling Design in Ad Hoc Wireless Networks. Proc. 25th Int. Conf. on Computer Communications, p.1-13. [doi:10.1109/INFOCOM.2006.142]

Chen, W.P., Hou, J.C., Sha L., 2004. Dynamic clustering for acoustic target tracking in wireless sensor networks. *IEEE Trans. Mob. Comput.*, **3**(3):258-271. [doi:10.1109/TMC.2004.22]

Cheng, T.E., Bajcsy, R., 2004. Congestion Control and Fairness for Many-to-One Routing in Sensor Networks. Proc. 2nd Int. Conf. on Embedded Networked Sensor Systems, p.148-161. [doi:10.1145/1031495.1031513]

Das, A., Dutta, D., 2005. Data acquisition in multiple-sink sensor networks. *ACM Mob. Comput. Commun. Rev.*, **9**(3):82-85. [doi:10.1145/1094549.1094561]

Du, Z.G., Tong, J., Huang, J.H., Su, Y.M., Yang, X.B., Qian, H.H., Fang, W.W., Wu, J.F., Liu, Y., Qian, D.P., 2007. ProNet: A Wireless Sensor Network Testbed Supporting Performance Measurement and Background Traffic Generation. Proc. 2nd Int. Conf. on the Latest Advances in Networks, p.185-190.

Eisenman, S.B., Campbell, A.T., 2007. E-CSMA: Supporting Enhanced CSMA Performance in Experimental Sensor Networks Using Per-Neighbor Transmission Probability Thresholds. Proc. 26th Int. Conf. on Computer Communications, p.1208-1216. [doi:10.1109/INFCOM.2007.144]

Fang, W.W., Qian, D.P., Liu, Y., 2008. Transmission control protocols for wireless sensor networks. *J. Software*, **19**(6):1439-1451 (in Chinese). [doi:10.3724/SP.J.1001.2008.01439]

Galluccio, L., Campbell, A., Palazzo, S., 2005. CONCERT: Aggregation-Based Congestion Control for Sensor Networks. Proc. 3rd Int. Conf. on Embedded Networked Sensor Systems, p.274-275. [doi:10.1145/1098918.1098951]

Gedik, B., Liu, L., Yu, P.S., 2007. ASAP: an adaptive sampling approach to data collection in sensor networks. *IEEE Trans. Parall. Distr. Syst.*, **18**(12):1766-1783. [doi:10.1109/TPDS.2007.1110]

Gupta, P., Kumar, P.R., 2000. The capacity of wireless networks. *IEEE Trans. Inf. Theory*, **46**(2):388-404. [doi:10.1109/18.825799]

Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H., 2002. *IEEE Trans. Wirel. Commun.*, **1**(4):660-670. [doi:10.1109/TWC.2002.804190]

Hull, B., Jamieson, K., Balakrishnan, H., 2004. Mitigating Congestion in Wireless Sensor Networks. Proc. 2nd Int. Conf. on Embedded Networked Sensor Systems, p.134-147. [doi:10.1145/1031495.1031512]

Jain, K., Padhye, J., Padmanabhan, V.N., Qiu, L.L., 2003. Impact of Interference on Multi-Hop Wireless Network Performance. Proc. 9th Annual Int. Conf. on Mobile Computing and Networking, p.66-80. [doi:10.1145/938985.938993]

Kang, J., Zhang, Y.Y., Nath, B., 2007. TARA: topology-aware resource adaption to alleviate congestion in sensor networks. *IEEE Trans. Parall. Distr. Syst.*, **18**(7):919-931. [doi:10.1109/TPDS.2007.1030]

Kumar, R., Crepaldi, R., Rowaihy, H., Harris, A.F., Cao, G.H., Zorzi, M., Porta, L.T., 2008. Mitigating performance degradation in congested sensor networks. *IEEE Trans. Mob. Comput.*, **7**(6):682-697. [doi:10.1109/TMC.2008.20]

Liu, C., Wu, K., Pei, J., 2007. An energy-efficient data collection framework for wireless sensor networks by exploiting spatiotemporal correlation. *IEEE Trans. Parall. Distr. Syst.*, **18**(7):1010-1023. [doi:10.1109/TPDS.2007.1046]

Misra, S., Tiwari, V., Obaidat, M., 2009. LACAS: learning automata-based congestion avoidance scheme for healthcare wireless sensor networks. *IEEE J. Sel. Areas Commun.*, **27**(4):466-479. [doi:10.1109/JSAC.2009.090510]

Popa, L., Raiciu, C., Stoica, I., Rosenblum, D., 2006. Reducing Congestion Effects in Wireless Networks by Multipath Routing. Proc. IEEE Int. Conf. on Network Protocols, p.96-105. [doi:10.1109/ICNP.2006.320202]

Teo, J.Y., Ha, Y.J., Tham, C.K., 2008. Interference-minimized multipath routing with congestion control in wireless sensor network for high-rate streaming. *IEEE Trans. Mob. Comput.*, **7**(9):1124-1137. [doi:10.1109/TMC.2008.24]

Vuran, M.C., Akyildiz, I.F., 2006. Spatial correlation-based collaborative medium access control in wireless sensor networks. *IEEE Trans. Network.*, **14**(2):316-329. [doi:10.1109/TENT.2006.872544]

Wan, C.Y., Eisenman, S.B., Campbell, A.T., 2003. CODA: Congestion Detection and Avoidance in Sensor Networks. Proc. 1st Int. Conf. on Embedded Networked Sensor Systems, p.266-279. [doi:10.1145/958491.958523]

Wan, C.Y., Eisenman, S.B., Campbell, A.T., Crowcroft, J., 2005. Siphon: Overload Traffic Management Using Multi-Radio Virtual Sinks in Sensor Networks. Proc. 3rd Int. Conf. on Embedded Networked Sensor Systems, p.116-129. [doi:10.1145/1098918.1098931]

Yu, Y.Q., Giannakis, G.B., 2009. Cross-layer congestion and contention control for wireless ad hoc networks. *IEEE Trans. Wirel. Commun.*, **1**(7):37-42. [doi:10.1109/TWC.2008.060514]

Zhao, M., Chen, Z.G., Zhang, L., Ge, Z.H., 2007. HS-Sift: hybrid spatial correlation-based medium access control for event-driven sensor networks. *IET Commun.*, **1**(6):1126-1132. [doi:10.1049/iet-com:20060128]

Zhang, Q., Yang, X.L., Zhou, Y.M., Wang, L.R., Guo, X.S., 2007. A wireless solution for greenhouse monitoring and control system based on ZigBee technology. *J Zhejiang Univ. Sci A*, **8**(10):1584-1587. [doi:10.1631/jzus.2007.A1584]

Zhou, Y.F., Lyu, R.M., Liu, J.C., Wang, H., 2005. PORT: A Price-Oriented Reliable Transport Protocol for Wireless Sensor Networks. Proc. 16th IEEE Int. Symp. on Software Reliability Engineering, p.117-126. [doi:10.1109/ISSRE.2005.32]