Congestion Control Protocols in Wireless Sensor Networks: A Survey

Atieh Rezaei[1], Marjan Kuchaki Rafsanjani[2]

[1]Department of Computer, science and Research Branch, Islamic Azad University of Kerman, Kerman, Iran
[2]Department of Computer science, shahid Bahonar University of Kerman, Kerman, Iran
atieh.rezaei@gmail.com; kuchaki@uk.ac.ir

Abstract: A Wireless Sensor Network (WSN) is deployed with large number of sensor nodes. Transfer packets in this networks present a range of challenges to protocol designers due to resources constrain, limited battery power, processing power, memory and storage capacity of sensor nodes in WSN. When a large number of sensor nodes transfer their packets, there is a possibility of packet loss due to congestion in sensor nodes. When sensor nodes are densely distributed and/or input packet flow rate exceeds the packet process rate, congestion may occur. Congestion causes decrease overall channel quality and QOS, increased transmission latency and loss rates, leads to buffer occupy and increased delays. If transmission packets to the network are not controlled, congestion status can arise. Therefore, in order to increase QOS and prolong system lifetime, we need various congestion control techniques. Different congestion control protocols have been proposed for wireless sensor networks which are reviewed in this paper.

[Rezaei A, Kuchaki Rafsanjani M. **Congestion Control Protocols in Wireless Sensor Networks: A Survey**. *J Am Sci* 2012;8(12):772-777]. (ISSN: 1545-1003). http://www.jofamericanscience.org. 108

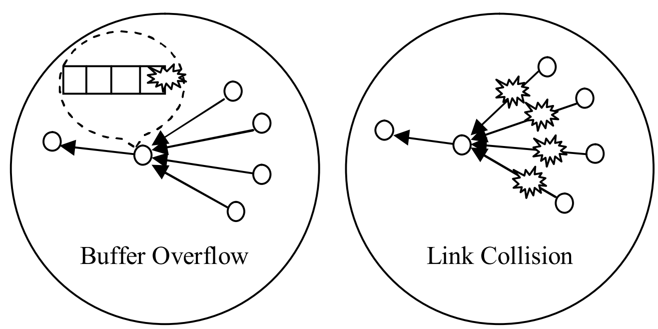
Keywords: Wireless sensor network; congestion control; reliable; priority

1. Introduction

A WSN consist of one or more sink and may ten or thousand of sensor nodes deployed in a zone. In WSN, often packets are sent from a set of sensors to a sink node, such communication terms as many to one traffic pattern. We term the nodes that are near the sources as upstream nodes and the nodes near the sink as downstream nodes. The major issue in wireless sensor network is congestion. In recent years, congestion control methods are used in several area of networking research. The main focus of congestion control protocols in WSNs is to provide tolerable level of reliability at the sink node with minimize the overall energy consumption in sensor nodes.

In this paper, we summarize various congestion control protocols for WSNs, specifically considering the data that are funneled from a subset of sensors toward an observer interested in collecting the data (many to one traffic pattern). Most of these algorithms are about avoiding or mitigating congestion in WSNs. Two types of congestion could occur in WSNs[1][2]: 1) node-level congestion and 2) link-level congestion. The first one is caused by buffer overflow in the node and the second one is happened when wireless channels are shared by several nodes and collisions occur when multiple active nodes try to seize the channel at the same time. This type of congestion can be eliminated by CSMA, FDMA, TDMA and CDMA in MAC layer.

Congestion control generally follows three steps [3][4]: congestion detection, congestion notification, and rate adjustment. All congestion control techniques have the same basic manners: they investigate the network to detect congestion, notify the other nodes of the congestion status, and reduce the congestion and/or its impact using rate adjustment methods.



a) Node-level congestion b) link-level congestion

Figure 1. Congestion in wireless sensor networks

In order to detect congestion, there are three different ways [5]. They are buffer occupancy, channel load and reporting rate. Buffer occupancy-based methods are the simplest while channel load and reporting rate provides more accurate information in some cases. However, monitoring and calculation of channel load and reporting rate can be costly in terms of power consumption but by controlling channel occupancy and reporting rate, we can reduce the congestion.

The congestion control protocols can be categorized into three groups [6]: rate-based, buffer-based and priority-based schemes. Rate-based scheme is to measure the average rate at which packets can be sent from the node and the average aggregate incoming rate. To implement buffer-based, we have to check whether sufficient buffer space is available at the downstream node. The priority-based scheme introduce node priority index to reflect the importance of each sensor node, which means that the important sensors have higher priority and can gain higher throughput.

We present a survey of existing congestion control approaches and classify them based on two parameters such as priority and reliability. These congestion control protocol differ in the way that they recognize congestion, notice congestion information and adjust traffic rate.

The reminder of this paper is organized as follows. In section 2, we explain a brief survey on the congestion control protocols with priority support. In section 3 we give an overview of various congestion control protocols with reliability support in wireless sensor networks. In section 4 we compare mentioned protocols and in section 5 concludes the paper.

2. Congestion control protocols with priority support

In WSN, sensor nodes may be scattered in different locations and they come with different hardware and capacity. They also have different sensing events and functions. Therefore, the priorities of sensors may differ. The important sensors have higher priority and therefore can get higher throughput.

We have presented three basic congestion control protocol with priority support in this section, for the survey. In each subsection below, different protocols are surveyed for controlling congestion in the WSN. In all these protocols in this paper, we assume upstream congestion control for a WSN that supports single-path routing, as Fig 2.

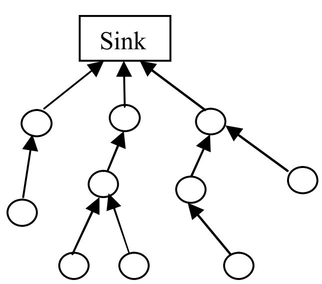


Figure 2. Network topology

2.1. Priority-based Congestion Control (PCCP)

PCCP [7] tries to reduce packet loss in congestion state while achieving the weighted fairness transmission for single-path and multipath routing. This protocol comprises of three units: intelligent congestion detection (ICD), implicit congestion notification (ICN) and priority-based rate adjustment (PRA).

ICD uses packet inter-arrival time (t_a) and packet service times (t_s) in order to produce a measure parameter defined as congestion degree in each sensor node i as follows:

$$d(i) = {t_s}/{t_a} \tag{1}$$

Congestion degree ($d(i)$) can represent current congestion level and can inform the offspring nodes about the traffic level to be increased or decreased by adjusting their transmission rate. ICN uses the piggybacking congestion information in the header of data packets at each sensor node. The piggybacked information at a sensor node includes mean packet service time (t_s), mean packet inter-arrival time(t_a), global priority (*GP),* and the number of offspring nodes. The global priority refers to the relative important of the total traffic at each node. Congestion could be avoided or mitigated through adjusting the scheduling rate in PRA. In order to congestion control, PRA needs to adjust the scheduling rate and the source rate at each sensor node after overhearing congestion notification from its parent node. Congestion degree and priority index at each sensor node provides more information and enables exact rate adjustment. Therefore, the rate adjustment has the following properties: (1) nodes with the same source traffic priority index get the same source rate; (2) nodes with a larger source priority index get higher source rate and higher bandwidth. (3) A node with sufficient traffic gets more bandwidth than one that generates less traffic.

2.2. Queue based Congestion Control Protocol with Priority Support (QCCP-PS)

QCCP-PS [8] protocol also uses the same congestion control mechanism as described in section 2.1. This congestion control protocol focuses more on queue structure. In the QCCP-PS the sending rate of each sensor node is increased or decreased depending on its congestion condition and its priority index. This protocol comprises of three units: Congestion Detection Unit (CDU), Congestion Notification Unit (CNU) and Rate Adjustment Unit (RAU). In the proposed protocol, in each sensor node we use separate queues to buffer arrival traffic from each child node in a disjoin queue.

The CDU use current queue length to calculate the congestion index$(I_x(i))$. This parameter varies from 0 to 1. For this purpose, two different fixed thresholds max and min are defined. When the queue length (q) is less than min, congestion index is very low and the source node could increase its rate. When the queue length is greater than max, congestion index is high and the source node should decrease its rate to avoid congestion. In the case that queue length is between max and min the congestion index is related to queue length linearly.

Suppose that sensor node i has N_i child nodes. So it has $N_i + 1$ queues (N_i queues for its child nodes and one queue for its local traffic source). In this unit, for each queue k in sensor node i, the congestion index $I_x^k(i), k = 1,2,\dots,N_i + 1$ is calculated as follows:

$$I_x^k(i) = \begin{cases} \varepsilon & q^k(i) \le min_{th}^k(i) \\ \frac{max_p(q^k(i) - min_{th}^k(i))}{max_{th}^k(i) - min_{th}^k(i)} - \varepsilon & min_{th}^k(i) \le q^k(i) \le max_{th}^k(i) \\ max_p + \varepsilon & q^k(i) \ge max_{th}^k(i) \end{cases} \quad (2)$$

Where ε and max_p are small numbers less than 1.Then $\overline{I_x^k}(i)$ is calculated as:

$$\overline{I_x^k}(i) = \frac{\sum_{j=1}^{N_i+1} I_x^j(i) - I_x^k(i)}{N_i . \sum_{j=1}^{N_i+1} I_x^j(i)} \quad (3)$$

RAU calculates the new rate of each child nodes based on the current congestion index and the source traffic priority. In order to define the priority of each node, we suppose $SP(i)$ denote the source priority at sensor node i. We define the total priority, as the sum of priorities of all nodes in the subtree rooted at node i. Let $C(i)$ be the set of node i's child nodes. Then the total priority, $TP(i)$, is calculated as follow:

$$TP(i) = \sum_{j \in C(i)} TP(j) + SP(i) \quad (4)$$

Then

$$p_i^1 = \frac{SP(i)}{TP(i)}, \ p_i^k = \frac{TP(k)}{TP(i)}, k = 2,3,\dots,N_i + 1 \quad (5)$$

If a node doesn't have any child, then its total priority is equal to its source priority. In each queue k in node i, the weight W_i^k and the input rate r_i^k are calculated as:

$$W_i^k = \frac{p_i^k \overline{I_x^k}(i)}{\sum_{j=1}^{N_i+1} p_i^j \overline{I_x^j}(i)} \quad (6)$$

$$r_i^k = W_i^k . r_i \quad (7)$$

Note that the rate r_i^k is calculated for all active sources in its subtree. When a sensor node is not active, then its congestion index is set to infinity. In this case the allocated rate to all inactive nodes will be equal to zero.

In this stage, the new rate that calculated in previous unit is sent to the CNU unit which is responsible for notifying all the child nodes of the new rate. The child node adjusts its traffic rate accordingly to the new rate.

2.3. Prioritized Heterogeneous Traffic-oriented Congestion Control Protocol (PHTCCP)

In this protocol [9], we focus on efficient mechanism so that congestion could be controlled by ensuring adjustment transmission rates for different type of data that generated by the sensors have various priorities. We assume that the sink node assigns individual priority for each type of sensed data and each node has n number of equal sized priority queues for n types of sensed data. Heterogeneous applications can reflect the number of queues in a node.

In congestion detection method, congestion level at each sensor node presented by packet service ratio $r(i)$ as follow:

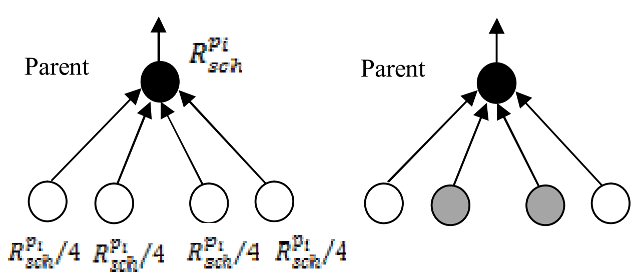
$$r(i) = R_s^i / R_{sch}^i \quad (8)$$

R_s^i is the ratio of average packet service rate and R_{sch}^i is the packet scheduling rate in each sensor node. Here, the packet service rate R_s^i is the inverse of packet service time, t_s^i. The packet service time t_s^i is the time interval when a packet arrives at the MAC layer and when it is successfully transmitted towards the next hop. In equation (8), to obtain R_s^i, t_s^i is calculated using exponential weighted moving average formula (EWMA). By using EWMA, t_s^i is updated each time a packet is forwarded as:

$$t_s^i = (1 - w_s) \times t_s^i + w_s \times inst(t_s^i) \quad (9)$$

Where, $inst(t_s^i)$ is the instantaneous service time of the packet that has just been transmitted and w_s is a constant where, $0 < w_s < 1$.The scheduling rate (R_{sch}^i) is defined as how many packets the scheduler forward per one time from the queues to the MAC layer from which the packets are delivered to the next node along the path towards the base station.

Congestion notification uses implicit congestion notification. Each node i piggybacks its packet scheduling rate (R_{sch}^i), total number of child nodes $(C(i))$, number of active child nodes at time t $(A_t(C(i)))$ and the weighted average queue length of its active child nodes in its packet header. By broadcasting this notification, all the child nodes of node *i* overhear the congestion notification information from its parent and adjust their rate. This protocol guarantees that heterogeneous data reach to the base station at their desired rates by hop-by-hop rate adjustment mechanism. In rate adjustment method, at first each node *i* calculates its packet service ratio $(r(i))$. When this ratio is equal to 1, it means that the incoming rate of packets to the MAC layer is equal to the rate at which packets are forwarded from the MAC layer so that no congestion occurs and R^i_{sch} remains unchanged. When the packet service ratio $r(i)$ is less than the specified threshold value, in such a case, packets would be queuing up at the MAC layer buffer and might cause buffer overflow indicating congestion. In such case, in order to controlling congestion, scheduling rate is decreased to the value of packet service rate. When $r(i)$ reache upper 1, the scheduling rate is increased using the equation, $R^i_{sch} = \beta \times R^i_s$. Here the value of β is chosen to a value smaller than but close to 1. In this case, each child node adjusts its own scheduling rate according to the scheduling rate of its parent node. When node *i* determines that all the child nodes of its parent are active at time t, $A_t(C(p_i)) = C(p_i)$, then node *i* makes adjustment in its scheduling rate. In this case, If the scheduling rate of the parent node is $R^{p_i}_{sch}$, each child node has the scheduling rate, $R^{p_i}_{sch}/4$. This ensures that the total scheduling rate of all child nodes is not greater than the scheduling rate of their parent node.



(a) (b)

Figure 3. Any of the child nodes is named as i and the black node is the parent of i, (a) all child nodes are active (white nodes) (b) two child nodes are idle (grey nodes)

When node *i* determines that some of the child nodes of its parent are idle that is when $A_t(C(p_i)) < C(p_i)$, it again adjusts its scheduling rate. After calculating the scheduling rate, each node *i* update their originating rate R^i_{or} according to priority for each type of data that assigned by the base station as follow:

$$R^i_{or} = \frac{R^i_{sch}(t) * \alpha_i}{\alpha_1 + \alpha_2 + \ldots + \alpha_n} \quad (10)$$

Where α_i is the priority for the *i*th queue of node *i* at time t.

3. Congestion control protocols with Reliability support

The problem of transport control protocols for wireless sensor networks is how to effectively provide congestion control and how to guarantee reliability while simultaneously conserving max energy. In this section, three reliable congestion control protocols namely RCRT, ESRT and STCP are reviewed. All three protocols provide guaranteed reliability for data transfer and energy conservation while avoiding network congestion.

3.1. Rate-Controlled Reliable Transport protocol (RCRT)

RCRT [10] uses end-to-end explicit loss recovery, but places all the congestion detection and rate adaptation functionality in the sinks. In this protocol, more than one sink or flow of data can be running concurrently. RCRT sink has four logical components: end-to-end retransmission, congestion detection, rate adaptation and rate allocation. In end-to-end retransmission, RCRT uses a NACK-based end-to-end loss recovery to insure 100%reliablity. The sink node detects missing packets and recovers them by requesting end-to-end retransmission packets from the source node. When each source node transmits its packets to the sink, the retransmission buffer of the source saves a copy of non acknowledgment packets. The sink node keep a list of sequence numbers of the lost packets then a NACK message containing sequence numbers are sent to the source. Upon receiving a NACK, the source node retransmits the requested packets. For congestion detection technique ,RCRT uses per-flow list of out-of-order which present the numbers of packets have been received after the first unrecovered packet loss and how much time has passed since the first unrecovered loss. The average time taken to repair a loss to be around trip time (RTT) hence the expected number of packets received during one RTT is$r_i RTT_i$. Ideally $1RTT$ should be the time required to recover a loss therefore, if the length of out-of-order packet list at the sink is equal to $r_i RTT_i$, it means 1 RTT has passed since the first recovered loss. The sink recognizes that the network is congested if the time to repair a loss is significantly higher than $1RTT$. In rate adaptation component, RCRT uses AIMD to adapt the transmission rate of each sources flow. The RCRT sink additively increase the rate when congestion is not detected and multiplicatively decrease the rate when sink determines the network is congested and calculates the new rates as follow:

Increase: $R(t+1) = R(t) + A$ (11)

Decrease: $R(t+1) = M(t)R(t)$

Where A is a constant and *M* (*t*) is computed based on loss rate, $M(t) = \frac{p_i(t)}{2-p_i(t)}$ and $p_i(t)$ is the loss rate value of the source *i* at the instant *t*. RCRT' rate allocation component assigns new rates to each flow in maintaining with the rate allocation policy. This policy allocates rate to each sources proportional to its demand.

3.2. Event to sink reliable transport (ESRT)

The ESRT protocol [11] considers reliability at the application level and achieves reliable delivery of packets from sensors to the sink with minimum energy expenditure. The congestion control mechanism in ESRT is designed for this purpose. The algorithms of ESRT mainly run on the sink node. The motivation of ESRT is that in some applications the sink is only interested in reliable detection of event features from the collective information provided by numerous sensor nodes. If the event reporting frequency at the sensors is too low, the sink may not be able to collect enough information to detect the events reliably. On the other hand, if the reporting frequency is too high, it may make congestion and cause the endangering event transport reliability in the WSN. ESRT adjusts the reporting frequency such that the observed event reliability is satisfied value while avoiding congestion. The event reliability is defined as the number of received data packets in a decision interval at the sink. The congestion detection in ESRT is through the local buffer level monitoring in sensor nodes. The sensor nodes set the Congestion Notification (CN) bit in a packet's header if congestion is detected. When the sink receives packets with CN bit marked, it realizes that congestion has occurred in the WSN.

In ESRT, the WSN can stay in one of the 5 states: (No Congestion, Low Reliability) (NC, LR), (No Congestion, High Reliability) (NC, HR), (Congestion, High Reliability) (C, HR), (Congestion, Low Reliability) (C, LR) and Optimal Operating Region (OOR). The sink derives a reliability indicator with a congestion indicator at the end of decision interval. This can help the sink determine in which of the above states the network currently resides. These two indicators can change with dynamic topology.

Depending on the current state, the sink calculates the updated reporting frequency and then broadcasts this information to the source nodes. In (NC, LR) state, the reliability is lower than required, so the reporting frequency is increased. In both (NC, HR) state and (C, HR) state, the reporting frequency is decreased with respective factors. In (C, LR) state, the reporting frequency is decreased more aggressively. In OOR, the required reliability is attained with minimum energy expenditure without any congestion. Thus, the sink informs source nodes to maintain the current reporting frequency for the next decision interval. The primary goal of ESRT is to obtain and maintain operation in state OOR where the required reliability is achieved without network congestion.

3.3. Sensor transmission control protocol (STCP)

STCP [12] is a generic end-to-end upstream transport protocol for WSN applications. STCP offers both congestion control and flexible reliability and puts functionalities and computational tasks in the sink node. For different applications, STCP provides different control policies in a way to guarantee application requirements. STCP uses three types of packets: session initiation, data, and ACK.

The sensor nodes use the session initiation packet to establish an association with the base station which constitutes a number of flows originating from it, type of data flow, transmission rate and required reliability. STCP data packets take an important role in maintaining the congestion information. STCP used NACK-based end-to-end retransmission for applications producing continuous flows. In this application, as the base station knows the rate of transmissions from the sources, the expected arrival time for the next packet can be found.

If the base station does not receive a packet within the expected time, it maintains a timer and sends a NACK packet to the source node.

STCP uses ACK-based end-to-end retransmission for event-driven applications. In this application, the base station sends a positive acknowledgement (ACK), if it has successfully received the packets. Each packet is kept in the source node's buffer until an ACK is received from the base station. Intermediate nodes detect congestion based on queue length and notify the base station by setting a bit in the data packet headers.

4. Protocols comparison

In this section, we compare existing congestion protocols based on priority and reliability. These protocols have different mechanisms in congestion detection, congestion notification and rate adjustment. Table 1 shows the comparison based on these three criteria as mentioned.

Table 1. Congestion control comparison

Features Protocols	Congestion detection	Congestion notification	Rate adjustment
PCCP	Packet inter-arrival time & packet service time	implicit	Exact rate adjustment
QCCP-PS	queue length	implicit	adjusts rate based on congestion index and the source traffic priority
PHTCCP	packet service rate & packet scheduling rate	implicit	adjusts rate based on the scheduling rate of its parent node
RCRT	Time to recover loss	Implicit or explicit	Additive Increase Multiplicative Decrease(AIMD)
ESRT	Packet sending success	implicit	Exact rate adjustment
STCP	Queue length	implicit	Additive Increase Multiplicative Decrease(AIMD)

5. Conclusion

In this paper, we had done a survey on the existing congestion control protocols. The most important functions of congestion control protocol should be focused including provide a reliable data transfer and source traffic priority mechanisms. Both factors of priority and reliability will helps in fairness transmission and reducing packet loss which result in an energy efficient for operation of the network. Furthermore the proposed protocol will increase the lifetime of the sensor network. Although some progress has been achieved, more research efforts are needed to continue to improve congestion control in WSNs.

References

1. Heikalabad S. R, Ghaffari A, Hadian M. A., Rasouli H. DPCC: Dynamic Predictive Congestion Control in Wireless Sensor Networks. International Journal of Computer Science (IJCSI), 2011; 8(1): 1694-0814.
2. Wei-wei Fang, Ji-ming Chen, Lei Shu, Tian-shu Chu, De-pei Qian. Congestion avoidance, detection and alleviation in wireless sensor networks. Journal of University-Sciencevol, 2009; 11: 63-73.
3. Premalatha N, Dr. Natarajan A. M. Congestion Control in Wireless Sensor Networks. International Journal of Computer Science and Network Security (IJCSNS), 2010; 10(3): 246-255.
4. Hull B, Jamieson K, Balakrishnan H. Mitigating Congestion in Wireless Sensor Networks. Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems (ACM SenSys), 2004: 134–147.
5. Helonde J, Wadhai V, Vivek D, Shiv S. EDCAM: Early Detection Congestion Avoidance Mechanism for wireless sensor network. International Journal of Computer Applications, 2010:7(2).
6. Jang-Ping Sheu, Li-Jen Chang, Wei-Kai Hu. Hybrid Congestion Control Protocol in Wireless Sensor Networks. Journal Of Information Science And Engineering, 2009;25: 1103-1119.
7. Wang C, Li B, Sohraby K, Daneshmand M, Hu Y. Upstream congestion control in wireless sensor networks through cross-layer optimization. IEEE Journal On Selected Areas In Communications, 2007; 25(4): 786-795.
8. Yaghmaee M. H, Adjeroh D. A New Priority Based Congestion Control Protocol for Wireless Multimedia Sensor Networks. Proceeding of the 2008 International Symposium on a world of Wireless Mobile and Multimedia Networks, 2008.
9. Monowar M, Rahman M, Pathan A, Hong C. Congestion control protocol for wireless sensor networks handling prioritized heterogeneous traffic. Proceeding of SMPE'08 with MobiQuitous, Ireland, 2008: 21-25.
10. Paek E, Govindan R. RCRT: Rate-Controlled Reliable Transport for wireless sensor networks. Proceedings of the 5th International Conference on Embedded Networked Sensor Systems (SenSys'07), 2007: 305-319.
11. Akan O.B, Akyildiz I.F. Event-to-sink reliable transport in wireless sensor network. IEEE/ ACM Transaction on Networking, 2005; 13(5).
12. Iyer Y, Gandham S, Venkatesan S. STCP: A Generic Transport Layer Protocol for Wireless Sensor Networks. Proceedings of the 14th IEEE International Conference on Computer Communications and Networks (ICCCN), San Diego, CA, USA, 2005.

08/12/2012