CrossMark

A Comprehensive Analysis of Congestion Control Protocols in Wireless Sensor Networks

Mian Ahmad Jan[1] · Syed Rooh Ullah Jan[1] · Muhammad Alam[2] · Adnan Akhunzada[3] · Izaz Ur Rahman[1]

Abstract

Wireless Sensor Networks (WSNs) consist of miniature sensor nodes, capable to operate and capture events in human-inaccessible terrains. The data packets generated by these networks may either be continuous, event-based or query-driven, and are application-specific in nature. These packets need to be transmitted in an energy-efficient manner to the base station. In these networks, congestion occurs when the incoming traffic load exceeds the available capacity of the network. There are various factors that lead to congestion in WSNs such as buffer overflow, varying rates of transmission, many-to-one communication paradigm, channel contention and the dynamic nature of a transmission channel. Congestion leads to depletion of the nodes energy, deterioration of network performance and an increase in network latency and packet loss. As a result, energy-efficient congestion control protocols need to be designed to detect, notify and control congestion effectively. Furthermore, these protocols need to ensure a reliable delivery of data in resource-constrained WSNs. In this paper, we present a review of the latest state-of-the-art congestion control protocols. Depending on their inherent nature of control mechanism, these protocols are classified into three categories, i.e., traffic-based, resource-based and hybrid. Traffic-based protocols are further subdivided, based on their hop-by-hop or end-to-end delivery modes. Resource-based control protocols are further analyzed, based on their route establishment approach and efficient bandwidth utilization techniques. We also discuss the internal operational mechanism of these protocols for congestion alleviation. Finally, we provide a comprehensive analysis of these protocols in terms of various performance metrics to justify in which scenario a particular class of these protocols needs to be deployed. Based on the performance analysis, we conclude that the behaviour of each class of protocols varies with the type of deployed application and a single metric alone cannot precisely detect congestion of the network.

Keywords Wireless sensor networks · Congestion control protocols · Traffic-based congestion control · Resource-based congestion control · Hop-by-hop · End-to-end

✉ Muhammad Alam
  alam@ua.pt

  Mian Ahmad Jan
  mianjan@awkum.edu.pk

  Syed Rooh Ullah Jan
  roohullahsyed@awkum.edu.pk

  Adnan Akhunzada
  a.adnan@siswa.um.edu.my

  Izaz Ur Rahman
  izaz@awkum.edu.pk

[1] Department of Computer Science, Abdul Wali Khan University Mardan, Mardan, Pakistan

[2] Department of Computer Science and Software Engineering, Xi'an Jiaotong-Liverpool University, Suzhou, China

[3] Department of Computer Science, COMSATS Institute of Information Technology Islamabad, Islamabad, Pakistan

1 Introduction

Wireless Sensor Network (WSN) is a collection of nodes capable of sensing, processing and communication in an autonomous manner. These tiny sensor nodes interconnect with each other in an ad hoc fashion to sense and monitor the phenomena of interest. Each node uses a short-range transmitter to route the data packets via intermediate nodes towards a sink node, also known as base station. These miniature sensor nodes are typically deployed in hazardous and human-inaccessible terrains to sense and monitor various applications [10]. These applications include but are not limited to seismic sensing, habitat monitoring, healthcare, intelligent transportation, home automation, industrial automation, agricultural monitoring and target tracking [39].

In WSNs, the data flow from a source node to the sink is application-specific in nature and may either be

periodic or continuous in nature [34]. In these networks, the data flows from sensor nodes toward the base station, i.e., upstream nodes, in a many-to-one fashion. The flow of packets in an upstream fashion may overload both the channel and nodes, and may exceed their handling capacity. An irregular upstream traffic may result in an increased delay, packet loss, energy utilization and an increase in the number of retransmission attempts. As a result, the performance of the underlying network deteriorates which adversely affect the reliability of any monitored application. Apart from these factors, WSNs have scarcity of various resources such as processing power, available memory, battery power and network bandwidth. The scarcity of resources, the overloading of nodes and the presence of error-prone communication links coupled with irregular upstream traffic flows, lead to an increased retransmission attempts. These factors ultimately lead to congestion in the network. Congestion arises when the number of transmitted packets exceeds the packet handling capacity of a particular node [2, 40]. This significantly decreases the performance of the network which results in higher data losses at the node level. In a multi-hop environment, the intermediate nodes suffer from resource exhaustion due to an unfair traffic distribution which is routed towards the base station via them. These nodes consume a considerable amount of resources as compared to the source nodes. As a result, energy-efficient congestion control protocols need to be designed that effectively alleviate congestion, ensure fairness and reliability of the network.

In WSNs, congestion can be controlled mainly via two different mechanisms, i.e., traffic-based and resource-based. Besides the researcher are using hybrid approaches as well by incorporating the distinguishing features of these two mechanisms. In a traffic-based congestion control mechanism, the data rate of incoming flows from the downstream nodes is adjusted against the forwarding capacity of the upstream node(s). Resource-based mechanism, on the other hand, exploits the idle network resources to balance the traffic load whenever congestion arises. The feasibility of these mechanisms may vary from one application to another. For example, traffic-based congestion control mechanism is feasible in situations when the transient overload occurs [30, 44]. Various protocols under this category reduce the effect of congestion by adjusting the data rate of incoming flows. However, these protocols result in packet drop and are not feasible for real-time applications. The resource-based mechanisms are useful to increase the network capacity by exploiting the unused resources in order to avoid congestion [12]. Various protocols under this category achieve a much higher data rate without compromising the network lifetime. The main contributions of this paper are as follow.

1. The latest protocols from literature are considered for this survey, mostly beyond 2013. The existing surveys consider protocols until 2013. The spectrum of our research confined mainly to congestion control protocols. Furthermore, we classify these protocols according to their routing modes and provide a detailed operational mechanism.

2. We design a trade-off mechanism to decide which congestion control mechanism is optimal and applicable in a particular scenario. To back-up this contribution, we analyze these control mechanisms in terms of various performance metrics such as energy efficiency, throughput and end-to-end delay. Based on our analysis, we conclude that a single parameter alone cannot predict congestion precisely in the network.

The rest of this paper is organized as follows. In Section 2, we discuss various causes of congestion, followed by congestion avoidance techniques in Section 3. In Section 4, we examine and classify various congestion control protocols based on their nature of operation. In Section 5, we have presented the design trade-off among these protocols in terms of various performance metrics. Finally, the paper is concluded and future research directions are provided in Section 6.

2 Causes of congestion

WSNs support diverse range of applications that are broadly classified into two types, i.e., monitoring and tracking [13]. The upstream data generated by these applications may either be event-driven, continuous, query-driven or hybrid [3]. In event-driven applications, the traffic load is relatively low and is activated only upon the occurrence of an event. As a result, congestion occurs in these applications because all the nearby nodes generate data simultaneously. In continuous sensing applications, data is sensed and transmitted continuously to the base station. In these applications, the network remain busy all the time in sensing, which results in congestion. In query-driven applications, the network is centrally controlled by the base station. In such applications, the data is generated as a result of a query broadcasted to the nodes by the base station. Hybrid applications, on the other hand, supports bulk and continuous data flows at the same time. These applications suffer from congestion due to varying data rates maintained by these flows.

Irrespective of the underlying application, the type of flow plays a pivotal role in monitoring the network congestion. There are various types of data flows such as, one packet, block of packets and stream of packets. Depending on the data flows, we require a light, medium or tight level of congestion control mechanisms. Whenever the

nodes transmit the data packets simultaneously, these packets pass through various intermediate nodes. The flooding of these packets results in congestion, which reduces the reliability, network performance, throughput and energy. In WSNs, it is relatively difficult to find the exact locations of congestion occurrence. This is due to the topological changes, variation in the report rate, and variation in the quality of radio channel with respect to time [18]. The aforementioned factors may convert uncongested regions within a network to congested regions [16]. The area around these regions may become a hotspot and there is a possibility of either buffer overflow or link interference. In these networks, there are mainly two types of congestion, i.e., node-based congestion [15, 36] and link-based congestion [7, 37].

3 Congestion avoidance techniques

Congestion in WSN can be avoided using three different techniques, i.e., congestion detection, congestion notification and congestion control [44]. Congestion can be detected using various parameters such as, queue length [38, 42], delay [14, 27], packet loss [23], packet service time [33], the ratio between service time and packet inter-arrival time and, channel load [41]. After successful detection of congestion, the next stage is congestion notification. In congestion notification, various methods are used for notifying congestion across the network. In WSNs, congestion maybe notified either implicitly or explicitly by the congested node across the network. In congestion control, various strategies and techniques are used to control congestion. In this survey, we focus mainly on congestion control protocols and algorithms. Unlike the existing surveys [18, 36, 40], we considered the latest novel contributions from literature for our work.
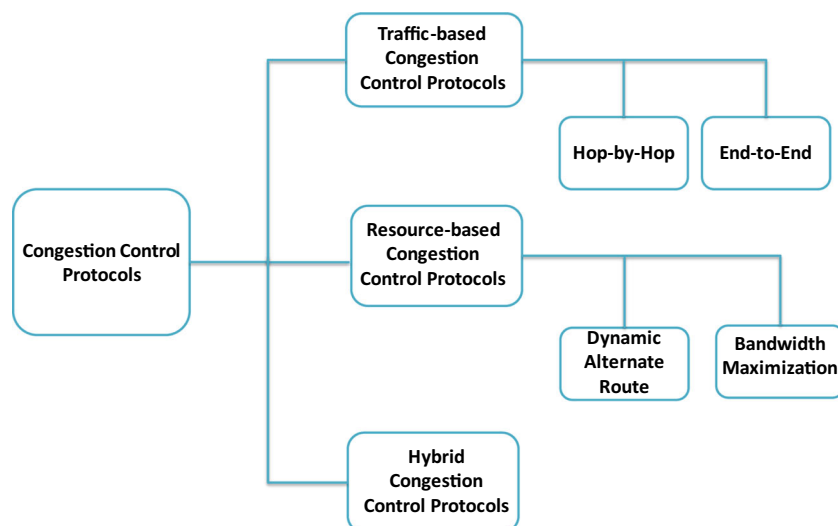
4 Classification of congestion control protocols

The choice of applying a congestion control method is application-specific in nature. Each application has its own requirements for data transfer. Thus, applying the same congestion control method to multiple applications is not appropriate and will have drastic consequences on the throughput and network lifetime of the underlying application. For example, in event-based applications, phase shifting techniques produce much better outcome in comparison to traffic-based congestion control methods, at the source node. This is because, the number of messages per event are relatively low [15]. Contrary to event-based applications, in continuous-based real-time applications, the data packets are sensed continuously and are time-bounded to reach their destination. Thus, resource-control methods serve a better choice as compared to traffic-based congestion control methods. In this section, we present a detailed survey of the latest state-of-the-art congestion control protocols in WSNs. We classify these protocols as either traffic-based, resource-based or hybrid. In case of traffic-based protocols, we further classify them based on their hop-by-hop and end-to-end communication paradigm. In case of resource-based protocols, we use the available bandwidth and idle resource exploitation of the network, as the classifying parameters. Hybrid congestion control protocols, on the other hand, consider the distinguishing features of traffic-based and resource-based protocols for network operation. The classification of various congestion control protocols is shown in Fig. 1.

4.1 Traffic-based congestion control protocols

In traffic-based congestion control methods, congestion is controlled by throttling the traffic rate at which the incoming

Fig. 1 Classification of congestion control protocols

packets are transmitted towards the destination. The traffic rate is adjusted based on the state of congestion. Traffic-based congestion control methods are further subdivided into two categories, which are hop-by-hop and end-to-end. In hop-by-hop traffic control, each relay node along the path stores information related to a packet in its buffer. This technique is energy-efficient as it conserves the energy of a node due to a shorter retransmission distance. Moreover, this technique reduces the required buffer size significantly under high feedback delays [12]. Each intermediate node monitors the overflow of its buffer. Thus, congestion is detected quickly and easily as compared to end-to-end mechanism. Whenever congestion occurs, all the intermediate nodes from source to destination are informed to take appropriate action. In an end-to-end packet recovery, each source node stores the packet information and is responsible for retransmission in case of packet loss. Congestion is signaled whenever there is a timeout or redundant acknowledgement messages are received by the end node (base station). The choice of a particular technique, i.e., hop-by-hop or end-to-end, depends on the reliability, and time-sensitivity of an underlying application. Traffic-based congestion control mechanisms have their own advantages such as, congestion control strategy, is applied to the congested nodes. Thus, congestion is quickly and easily controlled. In this subsection, we consider various traffic-based congestion control protocols which alleviate congestion using either a hop-by-hop or end-to-end approach. A summary of these protocols is provided in Table 1.

A priority rate-based routing protocol (PRRP) was proposed for Multimedia WSNs in [35]. PRRP deploys a hop-by-hop approach for congestion control. There are three phases of the PRRP protocol, i.e., congestion detection, congestion notification and rate adaptation. In congestion detection phase, the PRRP protocol uses buffer occupancy. It uses minimum and maximum threshold values to calculate the level of congestion at each node. There are three possible scenarios. If the queue length is less than the minimum threshold, then there is no congestion and the data rate is maintained. If the queue length is higher than the maximum threshold, the child nodes need to decrease their data rates to avoid congestion and packet loss. If the queue length is in-between the maximum and minimum thresholds, the actual congestion level is calculated and the data rate is adjusted, accordingly. The drawback of PRRP protocol is that it does not select an optimal route from source node to the sink. In [19], a congestion control protocol (CCP) was proposed for congestion alleviation in heterogeneous WSNs. The CCP protocol operates in three phases, which are congestion detection, congestion notification and congestion control. In CCP, congestion is detected via a hop-by-hop approach in order to calculate congestion degree indicator (CDI) at the node level by using buffer-free occupancy (BFO), queue

length and packet transmission time. Once congestion is detected, it is notified using a congestion control acknowledgement, i.e., CDI, by the congested node(s). Finally, during the third phase, congestion control occurs. The disadvantage of CCP is that it consumes too much energy during heavy traffic. In [1], a wireless multimedia congestion control protocol (WCCP) was proposed for alleviating congestion in multimedia WSNs. WCCP is a combination of two protocols, i.e., Source Congestion Avoidance Protocol (SCAP) and Receiver Congestion Control Protocol (RCCP). SCAP is used at the source nodes, whereas, RCCP is used at the relay nodes. WCCP considers the features of frames within the multimedia packets at transport layer. WCCP operates in two stages. During the first stage, SCAP uses group of picture (GOP) size for congestion detection and avoids it by adjusting the sending rate and the distribution of the outgoing packets from the source nodes. During the second stage, RCCP inspects and uses queue length of relay nodes for congestion detection. Once congestion is detected, SCAP protocol at the source node is informed. WCCP consumes a considerable amount of resources due to higher energy consumption. A weighted fairness congestion control protocol (WFCC) was proposed in [21]. WFCC introduces node weight to reflect the importance of each node. It is based on the assumption that the data generated at each node varies significantly due to the difference in the value of their importance. In WFCC, congestion is detected using the ratio of average packet service time to average packet inter-arrival time. WFCC controls congestion by modifying the incoming rate at regular intervals. WFCC incurs excessive overhead caused by the feedback at each interval.

Congestion control based on reliable transmission (CCRT) was proposed in [15]. It uses a priority-based congestion control mechanism for reliable transmission of critical information. Each node has different priority queues, such as high, median and low. Different types of data are classified based on predefined priority criteria. Queues at the node level are maintained based on first-come-first-serve basis. CCRT detects congestion based on the queue length and queue variation rate. A positive queue variation rate indicates that congestion may occur at the next time interval and a negative variation rate indicates that congestion is mitigated. If the queue variation rate is higher than a threshold value, it means that there is a probability of congestion. As a result, a normal rate adjustment needs to be applied. However, if the queue length is continuously increasing at a much faster rate, it means that there is a high probability of congestion at the next moment. As a result, emergent rate adjustment needs to be applied. CCRT offers higher throughput by reducing the delay and packet loss. However,CCRT does not take into account the congestion that occurs at the link layer. CCRT uses an end-to-end approach

Table 1 Traffic-based congestion control protocols

Protocols	Congestion detection	Congestion notification	Congestion control	Strengths	Weaknesses	Evaluating parameters
PRRP	Buffer Occupancy	Implicit	Hop-by-Hop	Higher Throughput, Lower Packet Loss	Optimal Path selection Undefined, Waiting time for low priority packets undefined	Delay, Throughput
CCP	Congestion Degree Indicator (CDI)	Explicit	Hop-by-Hop	Energy-efficient	Performance decreases under heavy traffic	End-End Delay, Energy Consumption
WCCP	Group of Picture (GOP)	Explicit	Hop-by-Hop	Improves the network performance and quality of video received at the sink	High Energy Consumption	Packet Loss Rate, Frame Loss Rate, Peak Signal-to-Noise Ratio (PSNR), End-To-End Delay, Throughput, and Energy Consumption
WFCC	Ratio of average packet service time to average packet inter-arrival time	Explicit	Hop-by-Hop	Throughput of the overall network is maintained	Overhead of feedback at each interval. The AIMD approach is employed for rate adjustment at the sink node however this does not control congestion effectively	Throughput and Fairness
CCRT	Queue Length, Queue Variation Rate, Rate Adjustment method	Explicit	End-to-End	Fair bandwidth allocation	Loss recovery mechanism undefined	Throughput and Buffer Occupancy
PASCCC	Queue Length, Packet Dropping	Piggybacked bits	Hop-by-Hop	Enhance Energy Efficiency, Reliability and other QoS parameters	Excessive delay during network setup phase	Stability Period, Instability Region, Energy Consumption, and Data Transmission

for congestion control. In [17], the authors proposed a priority-based application specific congestion control clustering (PASCCC) protocol for congestion detection in cluster-based hierarchical WSNs. In PASCCC, whenever the reading of a captured packet exceeds a predefined threshold value, each source node turns on its radio, senses the environment, collects the data and forwards it upstream towards the base station. During congestion, time-critical packets get prioritized to make sure their timely arrival at the base station. Thus, reliability of the data traveling across the network is ensured without compromising its quality. Furthermore, the nodes located at far distance from a cluster head get higher priority over the neighbouring nodes. A novel queuing scheduling is employed for each cluster head to ensure coverage fidelity, which ensures that the extra resources consumed by the far distant nodes are efficiently utilized. In PASCCC, the mobility of the nodes during setup phase incurs excessive delay. This is mainly due to the frequent change in the position of a node. Another drawback is that of dropping humidity packets, which has an adverse effect on applications such as, environmental monitoring and weather forecasting.

4.2 Resource-based congestion control protocols

Resource-based control methods are useful when the traffic rate control methods do not meet the requirements of an application. For example, in applications demanding high reliability such as real-time multimedia WSNs. In resource control, idle resources of the network are exploited to balance the traffic load. Traffic load is diverted towards idle nodes to optimize the available bandwidth. It can be accomplished either by turning on idle nodes or by using uncongested routes in case of heavy traffic, to alleviate congestion [7]. As a result, data packets have maximum chances to reach their destination, i.e., the base station. However, with every solution, there are certain associated drawbacks. For example, in resource control methods, loop avoidance, local knowledge, packet travel time and end-to-end topology information of the network is required, which comes with an additional overhead. In this subsection, we discuss various protocols which control congestion based on resource utilization in WSNs. The utilization of resources is controlled either by using dynamic alternative routes or efficient allocation of the available bandwidth. First, we discuss the protocols which use dynamic alternative routes followed by those which use an efficient mechanism for the allocation of available bandwidth. A summary of these protocols is provided in Table 2.

A hierarchical tree alternative path (HTAP) protocol was proposed for event-based sensor applications in [31]. It is a resource control algorithm that uses simple, scalable,

efficient, and low computational approach, by creating dynamic alternative path to the base station for alleviating congestion. HTAP employs resource-control mechanism for congestion avoidance. The concept of HTAP protocol is based on two algorithms, i.e., Alternative Path Creation (APC) and Hierarchical Tree Creation (HTC), and uses network density to choose the optimal one among the two. There are four stages of HTAP protocol, i.e., topology control, hierarchical tree creation, alternative path creation, and handling of powerless nodes. During the first stage, each node creates a table containing entries of the neighbouring nodes, located in its vicinity. In the second stage, the HTC nodes are placed in a hierarchical-leveled tree, starting from 0 at the root for the leaves nodes. Using two-way handshaking method, a receiver node informs the downstream neighbouring nodes about its own congestion level. As a result, these neighbouring nodes refresh their own tables, containing entries of neighbouring nodes, when packets are overheard. A receiver node has the ability to recognize a deadlock using negative acknowledgement by applying backpressure to ease congestion. A transmitter node stops forwarding packets to such nodes and searches for a less congested node, also known as idle node, which leads to alternative path creation. In the final stage, if there are any redundant nodes, i.e., extremely low-powered or energy depleted nodes, the table of a neighbour node is updated by removing such nodes to ease transient congestion. The main advantage of HTAP protocol is that, it is simple in its operational mechanism and as such results in a much lower overhead. The drawback of HTAP algorithm is that the receiver node receives the same data from multiple sensors, resulting in network redundancy. This problem is solved by using Redundancy Aware Hierarchical Tree Alternative Path (RAHTAP) algorithm [22]. In RAHTAP, every node runs a redundancy detection technique. Whenever, a receiver node receives a packet, it checks its queue to see if the packet with the same ID already exists. If the packet with the same ID exists, then that packet is discarded [61]. However, HTAP protocol does not provide fairness results and at the same time, is not energy efficient. A dynamic alternative path selection (DAIPaS) protocol for dynamically selecting the shortest route from source node to the base station, was proposed in [32]. This protocol utilizes the idle resources by routing data on alternative paths to resolve congestion. DAIPaS operates in three different phases, which are setup phase, topology control, and soft and hard stage. During the setup phase, the nodes connect with each other using CSMA/CA mechanism. The sink node, at Level 0, broadcasts hello messages to the nearby nodes, which reply with an acknowledgement. As a result, such nodes are classified as Level 1 nodes. These nodes update their neighbour tables with the sink ID and its level ID. This process is repeated

Table 2 Resource-based congestion control protocols

Protocols	Congestion detection	Congestion notification	Congestion control	Strengths	Weaknesses	Evaluating parameters
HTAP	Buffer Occupancy	Implicit	Dynamic Alternative Route	Minimal overhead, Efficiently create alternative routes, Scalable	Excessive Energy Consumption	Throughput, Avg hop-by-hop delay, energy consumption and increasing number of relay nodes
DAIPaS	Channel Load	Implicit	Dynamic Alternative Route	Simple with improved performance	Does not take into account Propagation Delay, Energy Consumption with hop count	Overall Energy Consumption
CcEbH	Queue Length	Explicit	Dynamic Alternative Route	Effectively deals with Network Congestion and Unbalanced Energy Consumption	Lack of Fairness	Packet Drop Rate, Average End-To-End Delay, Throughput, and Average Residual Energy
CCOR	Queue Occupancy Rate	Explicit	Efficient Allocation of Bandwidth	Energy-efficient, Reduces Packet Loss Rate	Degrade performance under heavy traffic, lack of choosing the shortest path, high delay under heavy traffic load	Packet Loss Rate, Average Routing Hops, Energy Consumption per Packet Received
Flock-CC	Queue Load, MAC Layer Collisions and Retransmissions	Implicit	Efficient Allocation of Bandwidth	Improve packet delivery ratio, Self-adaptive, Reduces packet loss and delay, Improves Scalability and Energy Consumption.	Sink direction discovery and the lack of finding an optimal path in terms of hop count and energy consumption, Does not guarantee fairness	Packet Delivery Ratio, Packet Loss, Delay and, Energy Tax under low, high and extreme traffic loads
TADR	Queue Length	Implicit	Dynamic Alternative Route	Improves overall throughput and lowers overhead	Unable to find sufficient idle nodes for caching all the dropped packets.	Throughput, Energy Consumption per Packet Received.

until the whole network is configured. During topology control phase, each node searches its neighbour table to find the nodes with the lowest level, i.e., nodes located near the sink, for data transmission. Therefore, a dynamic spanning tree is created for data transmission using the shortest path from source to destination. During the first step of final phase, i.e., soft stage, traffic is controlled at the node level and alternative routes are advised for the data flow from source to sink. Backpressure messages are used for reducing traffic rate to avoid any further packet drop and congestion. Finally, in the hard stage, the advised routes are used at the time of congestion. The selection of an alternative route depends on the availability, buffer occupancy and energy level of the next-hop neighbour, and the number of hops towards the sink. In [8], a Congestion control and Energy-balanced scheme based on Hierarchy (CcEbH) was proposed for congestion alleviation in WSNs. Initially, CcEbH, organizes the network into a hierarchical structure, starting from the sink node. In CcEbH, the nodes are classified into three categories, i.e., nodes at the same hierarchy, upstream nodes and the downstream nodes. During congestion, the nodes use lower hierarchical neighbouring nodes to forward data when their downstream nodes are congested. Congestion is detected at the node level using queue length, i.e., whenever the incoming traffic rate exceeds the capacity of the outgoing rate of the node. Each upstream node checks the buffer occupancy of the next-hop node. If the buffer occupancy of the said node is greater by 20% of total buffer size, the upstream node will choose alternative downstream nodes to receive data towards the sink to alleviate congestion. Traffic Aware Dynamic Routing (TADR) is a resource-control protocol that exploits the idle resources of the network [25]. Once congestion occurs in the network, TADR uses alternative routes. The proposed protocol forwards the redundant packets, i.e., packets exceeding the nodes handling capacity, in order to avoid packet drops. These redundant packets are directed toward alternative paths. Although, this protocol effectively tackles congestion for non-real time traffic, it fails to produce similar results for real-time traffic. This is because, TADR may discover a longer path from a congested node towards the sink. As a result, the end-to-end delay increases, which is not the case with most of the real-time applications. Furthermore, this protocol is unable to find sufficient idle nodes for caching all the dropped packets.

Congestion Control protocol based on Optimizing Routing (CCOR) is a novel approach that alleviates congestion while maintaining energy efficiency and fairness, simultaneously [9]. Initially, a queuing network model is built at the node level to detect congestion degree and identify congestion. Next, the CCOR protocol considers two independent functions for routing purposes, i.e., link gradient

and traffic radius. These two functions are based on the location of each node and service rate of the packets. In case of congestion, i.e., whenever congestion degree exceeds a specific threshold, the traffic-radius function ensures that the packets are forwarded toward the least congested areas of the network for an efficient utilization of the available bandwidth. The proposed approach effectively alleviates congestion and subsequently preserves energy of the network. However, the performance of CCOR deteriorates under heavy traffic. Flock-based Congestion Control (Flock-CC) is a scalable, self-adaptive and robust congestion control protocol proposed for event-based applications [5]. Flock-CC is based on the collective behavior of the bird flocks, inspired by swarm intelligence. The proposed algorithm directs the packets (synonym to birds) to form flocks and forwards them towards a sink, i.e., a global attractor, for avoiding obstacles. The movement of a bunch of packets, i.e., packet flock, depends upon attraction and repulsion forces among neighbouring packets, the field view, and the artificial magnetic field, guiding these packets towards the sink. As a result, idle resources of network are efficiently utilized. Flock-CC is easy to implement at node level with minimal exchange of information. The proposed protocol is highly scalable and efficiently handles the redundant nodes. Moreover, Flock-CC improves the packet delivery ratio, packet loss, delay, and energy consumption.

4.3 Hybrid congestion control protocols

In hybrid congestion control protocols, each protocol uses an integrated approach, i.e., by combining the desirable features of both traffic and resource-based congestion control protocols. In most cases, these protocols use a traffic-based congestion control approach. However, if the aforementioned technique is not feasible and optimal, then they use the resource-based congestion control approach. A summary of these protocols is provided in Table 3.

Hybrid Resource and Traffic Control (HRTC) is a hybrid algorithm for efficient congestion control in WSNs [29]. HRTC combines the desirable features of these two congestion handling techniques and provides a suitable solution, based on the network condition. During congestion, a congested node informs a source node using a hop-by-hop communication link to reduce its data rate, using a back-pressure message. When the back-pressure message traverses the affected downstream nodes on its way towards the source node, HRTC protocol examines to see if resource control technique can be applied to the traversed nodes before reaching the source node. If that is the case, HRTC aborts the transmission of back-pressure message. If HRTC is unable to find an alternative route, then the back-pressure message continues its journey towards its destination, i.e., the source node. Once this message reaches its destination,

Table 3 Hybrid congestion control protocols

Protocols	Congestion detection	Congestion notification	Congestion control	Strengths	Weaknesses	Evaluating parameters
HRTC	Buffer Occupancy, Remaining power of the node	Implicit	Hybrid	Enhance Network Lifetime and Throughput.	No Fairness and not energy-efficient in some cases	Throughput
HOCA	Active Queue Mechanism at the relay nodes	Explicit	Hybrid	Minimize End-To-End Delay and maximizes Network Lifetime by energy conservation and Fairness in the network.	Congestion Avoidance for non-sensitive data is missing	Network Lifetime, Remaining Average Energy, End-to-End Delay and Fairness
HTCCFL	Packet Service Ratio and Buffer Occupancy	Explicit	Hybrid	Improved Packet Delivery Ratio, Packet Dropping and Energy Consumption	Excessive Jitter with varying Data Rates	Packet Delivery Ratio, Packet Drop Ratio, Energy Consumption and Delay

it then applies traffic control technique by altering the date rate of the source node. Moreover, the new data rate is adapted by nodes across the network. Healthcare aware Optimized Congestion Avoidance (HOCA) is a data-centric congestion control protocol proposed for healthcare applications of WSNs [26]. It employs the concept of active queue management (AQM) [6]. In HOCA, the data is divided into two categories, i.e., sensitive and non-sensitive. The former requires a higher data rate while the latter requires a lower data rate, depending on the priority levels of the data. HOCA operates in four different stages. In the first stage, data dissemination request is performed by the sink (medical center) to all the nodes (patients) in the network. Patient's type, priority of the data, timing and characteristics of the request, are specified during this stage. During the second stage, the occurrences of events are reported by the nodes located on the patient's body, to the sink. During the third stage, route is established by the sink node using multipath and QoS-aware routing techniques, to mitigate congestion. During the final stage, data is forwarded by adjusting hop-by-hop source traffic rate. This adjustment of traffic rate occurs specifically during congestion. HOCA avoids congestion by lowering end-to-end delay and maximizes the network lifetime by energy conservation. Furthermore, HOCA ensure fair use of the resources in the network. Hierarchical Tree-based Congestion Control using Fuzzy Logic heterogeneous traffic in WSN (HTC-CFL), is a fuzzy-based congestion alleviation protocol [28]. This protocol operates in three phases, i.e. hierarchical tree construction, fuzzy-based congestion detection and priority-based rate adjustment. During the first phase, a hierarchical tree is constructed using a topology control algorithm. During the second phase, congestion is detected using a fuzzy logic technique, based on input parameters such as packet service ratio, number of contender nodes and buffer occupancy. The state of congestion is predicted from the outcome obtained using fuzzy rules. A node experiencing congestion informs all its neighbours about its status by using a control packet. For different classes of traffic, prioritized queues are maintained and weight values are assigned to each queue. Once congestion is detected, the next phase is congestion control. During the final phase, a dynamic rate adjustment is performed. If in case, the rate adjustment is not possible, each source node selects an alternative route from an already existing hierarchical tree for congestion alleviation.

5 Design trade-off of congestion control protocols

In this section, we aim to address the question "which congestion control method is optimal and in which scenario"?

This decision is based on various performance metrics such as energy efficiency, throughput, average delay and packet dropping. Depending on data delivery mechanism, all applications of WSNs can be categorized into event-driven, continuous, query-driven, and hybrid. Therefore, the type of application and location of occurrence of congestion, are considered distinguishing features among various congestion control protocols. Applying the same congestion control technique to all applications and all congestion locations is, by no mean appropriate and will consequently lead to performance degradation.

5.1 Energy efficiency analysis

In WSNs, it is imperative to design transport layer protocols that tackle congestion in an energy-efficient manner, reduce the packet loss and prolong network lifetime. Therefore, transport protocols need to account for congestion control and loss recovery. These factors have a direct impact on energy efficiency of the network and QoS of the application under observation. Researcher in this field are continuously trying to preserve energy of the network using various software-based solutions [11, 20, 24]. A detailed discussion on various techniques used for energy conservation in available in [4]. Traffic-based congestion control protocols reduce the traffic load on the network by reducing the source data rate. Instead of finding alternative routes, these protocols forward the packets on the same route towards destination. This results in much higher energy consumption per unit time which lowers the network life time. Contrary to traffic-based congestion control protocols, resource-based protocols exploit the idle resources of the network whenever a node becomes a hotspot. In case of congestion, these protocols re-route excessive traffic to the base station via alternative routes. Therefore, the traffic load is balanced between congested and uncongested routes of the network, increasing the network throughput and life time. Furthermore, resource control protocols always start using a shortest route to forward their packets. These protocols always employ a topology control scheme [43] to increase the number of alternative routes. This means that several paths can be used which leads to a balanced energy consumption during heavy network load. Thus, maximum lifetime can be achieved because the nodes exhaust their energy uniformly. Hybrid congestion control protocols, on the other hand, first try to control congestion by checking buffer occupancy of the congested nodes. Whenever the buffer of a congested node reaches a threshold value, instead of packet dropping or traffic rate reduction, these protocols choose an alternative route by exploiting idle resources of the network and re-route the outgoing traffic from the congested node towards the base station. Avoiding congestion through hybrid congestion control protocols

seems to be a promising solution which results in higher throughput, lower packet loss and a uniform utilization of energy. However, these protocols suffer from an increased overhead in the form of control messages. A comparison among these various congestion control protocols is shown in Fig. 2.

5.2 Throughput

Traffic-based control protocols frequently use the shortest path from source towards the destination, that result in quicker energy depletion of the network resources. In term of throughput, these protocols avoid congestion by throttling the data rate of source nodes that results in higher packet losses and lower throughput at the base station. Resource-based control protocols use alternative routes by exploiting the idle resources of the network. As a result, these methods have a higher throughput and uniform energy utilization of the network while avoiding packet dropping. However, these protocols incur an additional overhead associated with control messages transmitted across the network. Hybrid congestion control protocols are suitable for continuous event-based applications that require higher reliability, higher throughput and mission-critical real-time transmission. These protocols achieve much higher throughput as compared to resource and traffic-based congestion control protocols, resulting in a uniform energy utilization of the nodes. A comparison among these protocols in term of throughput is shown in Fig. 3.

5.3 Average end-to-end delay analysis

Delay is an important metric to consider while designing a congestion control strategy. The late arrival of packets may have catastrophic affects on various mission-critical applications. For example, even a minor delay in healthcare
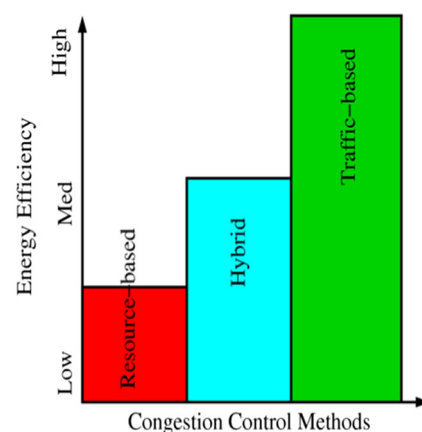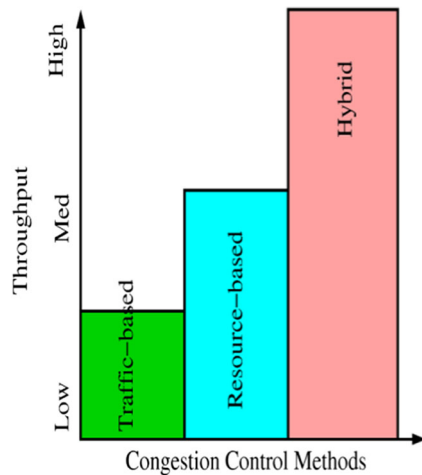

Fig. 2 Energy efficiency analysis

Fig. 3 Throughput analysis

monitoring applications may risk the life of a patient. This metric is an indication of the efficiency of a particular congestion-control protocol to quickly mitigate or avoid congestion. A smaller delay represents an enhanced performance and vice-versa. Resource control protocols experience higher delay as compared to traffic-based protocols. This is because, the packet is forwarded through a larger number of intermediate nodes before reaching its final destination. Furthermore, the retransmission that occur in the congested hotspots also increases delay, i.e, multiple hops form a route from source towards the base station, using alternative routes once congestion is detected. In hybrid applications, delay is higher or at the same degree as in resource control methods due to switching from traffic-based to resource control, using back pressure messages. A comparison among these congestion control protocols is shown in Fig. 4.
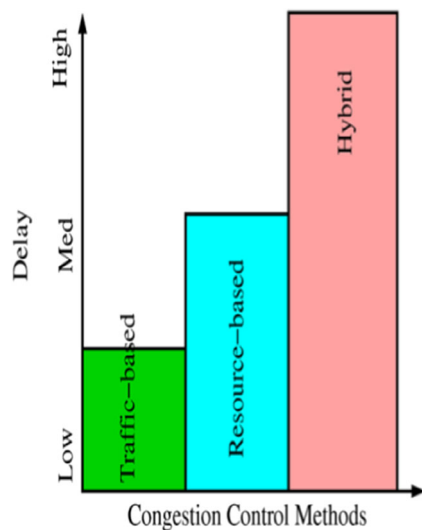
Fig. 4 Average end-to-end delay analysis

6 Conclusion

Wireless Sensor Networks (WSNs) consist of resource-constrained nodes that face various challenges. One such challenge is network congestion which mainly arises when the reception rate of a node exceeds its transmission rate. Congestion also arises due to contention and interference at the link layer. As a result, the performance and reliability of the network deteriorate to a larger extent. In WSNs, congestion needs to be detected by one or more nodes along the upstream path towards the base station. Once congestion is detected, it needs to be notified either explicitly or implicitly across the network in order to take precautionary measures to control it. In this paper, we presented a detailed and comprehensive analysis of the latest congestion control protocols in WSNs. We classified these protocols into three categories, i.e., traffic-based, resource-based and hybrid. The protocols in these three categories have their own strengths and weaknesses, and are application-specific in nature. In this paper, we have examined various performance metrics with the aforementioned congestion control protocols. It is concluded that a single metric is unable to detect congestion with a much higher precision. Therefore, more than one metric needs to be used for precisely detecting congestion. The behaviour of each category of such protocols changes with the change of a performance metric. In WSNs, energy-efficient congestion control protocols are required that control congestion in order to ensure low energy consumption, fairness, and at the same time, achieve a much higher QoS. More sophisticated congestion control strategies may be explored that are based on statistical analysis of the queues, neural networks, fuzzy logic and machine learning. At present, many such solutions exist but the effectiveness of these solutions need to be investigated further for various applications and under various scenarios.

References

1. Aghdam SM, Khansari M, Rabiee HR, Salehi M (2014) Wccp: A congestion control protocol for wireless multimedia communication in sensor networks. Ad Hoc Netw 13:516–534

2. Aguirre-Guerrero D, Marcelín-jiménez R, Rodriguez-Colina E, Pascoe-Chalke M (2014) Congestion control for a fair packet delivery in wsn: from a complex system perspective. The Scientific World Journal, 2014

3. Akyildiz IF, Su W, Sankarasubramaniam Y, Cayirci E (2002) Wireless sensor networks: a survey. Comput Netw 38(4):393–422

4. Anastasi G, Conti M, Francesco MD, Passarella A (2009) Energy conservation in wireless sensor networks A survey. Ad Hoc Netw 7(3):537–568

5. Antoniou P, Pitsillides A, Blackwell T, Engelbrecht A, Michael L (2013) Congestion control in wireless sensor networks based on bird flocking behavior. Comput Netw 57(5):1167–1191

6. Athuraliya S, Low SH, Li VH, Yin Q (2001) Rem: Active queue management. IEEE Netw 15(3):48–53
7. Bachir A, Dohler M, Watteyne T, Leung KK (2010) Mac essentials for wireless sensor networks. IEEE Communications Surveys & Tutorials 12(2):222–248
8. Chen W, Niu Y, Zou Y (2016) Congestion control and energy-balanced scheme based on the hierarchy for wsns. IET Wireless Sensor Systems 7(1):1–8
9. Ding W, Tang L, Ji S (2016) Optimizing routing based on congestion control for wireless sensor networks. Wirel Netw 22(3):915–925
10. Dulman S, Nieberg T, Wu J, Havinga P (2003) Trade-off between traffic overhead and reliability in multipath routing for wireless sensor networks. In: 2003 IEEE wireless communications and networking, 2003. WCNC 2003, vol 3. IEEE, pp 1918–1922
11. Felix Enigo VS, Ramachandran V (2009) An energy efficient congestion control protocol for wireless sensor networks. In: IEEE international advance computing conference, 2009. IACC 2009. IEEE, pp 1573–1578
12. Fang W-W, Chen J-M, Shu L, Chu T-S, Qian D-P (2010) Congestion avoidance, detection and alleviation in wireless sensor networks. Journal of Zhejiang University-Science C 11(1):63–73
13. Ghaffari A (2015) Congestion control mechanisms in wireless sensor networks: A survey. J Netw Comput Appl 52:101–115
14. He T, Stankovic JA, Lu C, Abdelzaher T (2003) Speed: A stateless protocol for real-time communication in sensor networks. In: Proceedings of the 23rd international conference on distributed computing systems, 2003. IEEE, pp 46–55
15. Hua S (2014) Congestion control based on reliable transmission in wireless sensor networks. Journal of Networks 9(3):762–768
16. Hull B, Jamieson K, Balakrishnan H (2004) Mitigating congestion in wireless sensor networks. In: Proceedings of the 2nd international conference on embedded networked sensor systems. ACM, pp 134–147
17. Jan MA, Nanda P, He X, Liu RP (2014) Pasccc: Priority-based application-specific congestion control clustering protocol. Comput Netw 74:92–102
18. Kafi MA, Djenouri D, Ben-Othman J, Badache N (2014) Congestion control protocols in wireless sensor networks: a survey. IEEE Communications Surveys & Tutorials 16(3):1369–1390
19. Kumar KA, Krishna AVN, Chatrapati KS (2016) Congestion control in heterogeneous wireless sensor networks for high-quality data transmission. In: Proceedings of the international congress on information and communication technology. Springer, pp 429–437
20. Kumar KA, Krishna AVN, Chatrapati KS (2016) Congestion control in heterogeneous wireless sensor networks for high-quality data transmission. In: Proceedings of the International Congress on Information and Communication Technology. Springer, pp 429–437
21. Li G, Li J, Yu B (2012) Lower bound of weighted fairness guaranteed congestion control protocol for wsns. In: Proceedings of the IEEE INFOCOM, 2012. IEEE, pp 3046–3050
22. Luha AK, Vengattraman T, Sathya M (2014) Rahtap algorithm for congestion control in wireless sensor network. International Journal of Advanced Research in Computer and Communication Engineering 3(4):6250–6255
23. Messaoud D, Djamel D, Nadjib B (2012) Survey on latency issues of asynchronous mac protocols in delay-sensitive wireless sensor networks. IEEE Trans Communications Surveys Tutorials PP(99):1–23
24. Palattella MR, Accettura N, Vilajosana X, Watteyne T, Grieco LA, Boggia G, Dohler M (2013) Standardized protocol stack for the internet of (important) things. IEEE Communications Surveys & Tutorials 15(3):1389–1406
25. Ren F, He T, Das SK, Lin C (2011) Traffic-aware dynamic routing to alleviate congestion in wireless sensor networks. IEEE Transactions on Parallel and Distributed Systems 22(9):1585–1599
26. Rezaee AA, Yaghmaee MH, Rahmani AM, Mohajerzadeh AH (2014) Hoca: Healthcare aware optimized congestion avoidance and control protocol for wireless sensor networks. J Netw Comput Appl 37:216–228
27. Sankarasubramaniam Y, Akan ÖB, Akyildiz IF (2003) Esrt: event-to-sink reliable transport in wireless sensor networks. In: Proceedings of the 4th ACM international symposium on mobile ad hoc networking & computing. ACM, pp 177–188
28. SAYYADA J, Choudhari NK (2014) Hierarchical tree based congestion control using fuzzy logic for heterogeneous traffic in wsn. International Journal of Current Engineering and Technology 4(6):4136–4143
29. Sergiou C, Vassiliou V (2014) Hrtc: A hybrid algorithm for efficient congestion control in wireless sensor networks. In: 2014 6th international conference on new technologies, mobility and security (NTMS). IEEE, pp 1–5
30. Sergiou C, Vassiliou V, Paphitis A (2013) Hierarchical tree alternative path (htap) algorithm for congestion control in wireless sensor networks. Ad Hoc Netw 11(1):257–272
31. Sergiou C, Vassiliou V, Paphitis A (2013) Hierarchical tree alternative path (htap) algorithm for congestion control in wireless sensor networks. Ad Hoc Netw 11(1):257–272
32. Sergiou C, Vassiliou V, Paphitis A (2014) Congestion control in wireless sensor networks through dynamic alternative path selection. Comput Netw 75:226–238
33. Sharif A, Potdar V, Rathnayaka AJD (2010) Prioritizing information for achieving qos control in wsn. In: 2010 24th IEEE international conference on advanced information networking and applications (AINA). IEEE, pp 835–842
34. Spachos P, Song L, Hatzinakos D (2013) Prototypes of opportunistic wireless sensor networks supporting indoor air quality monitoring. In: 2013 IEEE consumer communications and networking conference (CCNC). IEEE, pp 851–852
35. Tshiningayamwe L, Lusilao-Zodi G-A, Dlodlo ME (2016) A priority rate-based routing protocol for wireless multimedia sensor networks. In: Advances in nature and biologically inspired computing. Springer, pp 347–358
36. Vijayaraja V, Hemamalini RR (2010) Congestion in wireless sensor networks and various techniques for mitigating congestion-a review. In: IEEE international conference on computational intelligence and computing research
37. Vuran MC, Gungor VC, Akan OB (2005) On the interdependency of congestion and contention in wireless sensor networks. In: Proceedings of the SENMETRICS'05, pp 136–147
38. Wan C-Y, Eisenman SB, Campbell AT (2011) Energy-efficient congestion detection and avoidance in sensor networks. ACM Transactions on Sensor Networks (TOSN) 7(4):32
39. Wang C, Sohraby K, Hu Y, Li B, Tang W (2005) Issues of transport control protocols for wireless sensor networks. In: Proceedings of the 2005 international conference on communications, circuits and systems, 2005, vol 1. IEEE, pp 422–426

40. Wang C, Sohraby K, Li B, Daneshmand M, Hu Y (2006) A survey of transport protocols for wireless sensor networks. IEEE Netw 20(3):34–40

41. Wang C, Sohraby K, Li B, Daneshmand M, Hu Y (2006) A survey of transport protocols for wireless sensor networks. IEEE netw 20(3):34–40

42. Woo A, Culler DE (2001) A transmission control scheme for media access in sensor networks. In: Proceedings of the 7th annual international conference on mobile computing and networking. ACM, pp 221–235

43. Yin X, Zhou X, Huang R, Fang Y, Li S (2009) A fairness-aware congestion control scheme in wireless sensor networks. IEEE Trans Veh Technol 58(9):5225–5234

44. Zheng J, Jamalipour A (2009) Wireless sensor networks: a networking perspective. Wiley, New York