Congestion Control in IoT

Muhammad Amer Saleem[1], Muhammad Sheraz Arshad Malik[2], Shakeel Waris[1], Sanam Barkat[1], Usman Sadaqat[1]

[1]Ripha International University Faisalabad Pakistan, Department of Computer Science
[2]Government College University Faisalabad Pakistan, Department of Information Technology

Abstract
Internet of things (IoT) is made of both constrained and unconstrained devices. These constraints include power, processing, and memory constraints. To perform well in a constrained environment, a congestion control mechanism must consider these constraints. Congestion in an IoT network causes a degrade in its quality of service (QoS). Quality of service includes different performance measures such as the loss of data packets, decrease in throughput and energy inefficiency etc. To cope with this issue many proposals describe congestion control (CC) techniques for IoT networks. Although enormous CC techniques have been proposed in different researches, but still there exists some issues that are not been fully catered yet. This survey presents a decent summary of different kinds of congestion detection and avoidance techniques been (or being) used in IoT networks. Each CC technique operates at some implicit assumptions about the IoT traffic type. These assumptions cause some CC schemes to perform well in certain scenarios but not in others. To measure performance of a congestion avoidance technique in a scenario, multiple performance metrics are proposed by different authors. We will briefly discuss reasonable number of performance evaluation metrics for CC techniques.

Key words:
Congestion Control in IoT, Congestion Detection and avoidance in IoT, congestion in WSN

1. Introduction

This report briefly summarizes different congestion detection and control schemes being used in internet of things (IoT). Congestion in networks can be defined as the state where a node or link carries so much data that it may affect network service quality. As a result, the issues of long delay in queues, loss of data packets and blocking of new connections can occur. Response time also slows down in congested network with reduced network throughput [1].
Two types of the flows are possible in IoT networks: 1) Bi-directional flow and 2) Uni-directional flow or upstream only. This study considers congestion control protocols for both scenarios. Bi-directional web-based flow of traffic in IoT mainly uses the constraint application protocol (CoAP) and IPv6 over low-power wireless personal area network (6LoWPAN).

There are some applications in which upstream traffic delivery can be, continuous, query-based, hybrid and event-based [4].

1.1 Event Based Application

The network load is less but if the event is found each node sends its samples to base station which may cause congestion.

1.2 Time Based Application

When applications require real time values this require continuous sending of sensed values. If the network is not allowed for such transmissions, at such scenario sensing in periodic manner.

1.3 Applications Based on Query

As event based applications, here is a query instead of event. The query sensing nodes have to answer queries when sink invokes it.

1.4 Multi-type Applications

In such cases, often bulk data is generated in addition to the constantly sensed data. Here congestion occurs when the traffic load exceeds the available capacity on node level (buffer overflow) or link level (interference or contention).

2. Congestion Control Paradigm

2.1Congestion Detection Metrics

Following methods that help to detect congestion in network [1], [4]:

2.1.1 Packet Loss

The average rate of packet loss tells about how network is congested.

Manuscript received February 5, 2019
Manuscript revised February 20, 2019

2.1.2 Queue Length

Buffer length is also a good indicator for congestion detection. In [1] a fixed limit threshold is exploited and the congestion signal is invoked when the length of buffer exceeds the limit, in [10], remaining free buffer size from the overall size is used. In [11], [12], the traffic rate and remaining buffer is used for congestion detection.

2.1.3 Channel Load and Hybrid Queue Length

When collisions increase and result in lot of unsuccessful retransmissions, packets are removed from the queue or buffer, resultantly the size of buffer or buffer occupancy decreases it may means or may be perceived as uncongested scenario. Therefore, the true measure to detect congestion is hybrid approach. In this approach the queue size and channel load is used as congestion indicator. Channel load is the ratio of time quantum's when channel is being used by someone to the total time.

2.1.4 Throughput

Throughput of the channel indicates number of successful transmissions. In fact, it indicates that how much channel remain busy. It can be used for congestion detection.

2.1.5 Packet Service Time

The time interval between arrival packet time and transmission of packet covers the time to resolve collision, MAC transmission time of packet, and waiting time of packet. It is also a useful measure to detect congestion. It can be used to measure node to node delay as well as measurement of end to end delay.

2.1.6 Scheduling Time

Scheduling time indicates the number of packets scheduled in per time unit. The link and node level congestion can be determined by this ratio. The ratio is used instead of time which is known as packet service ratio.

2.1.7 Delay

Delays can also be used for congestion detection.

2.2 How to Propagate Congestion Information

Once congestion occurred there is need to propagate the information about congestion. The information signal can be of different size it may be a single bit or as a rich new data rate information. Often, this congestion information is placed inside the header of outgoing packet.

2.3 Congestion Avoidance and Control

When the congestion has been detected immediate need is to control it. Congestion control refers the methods that are used to maintain the traffic according to capacity of network. The following three techniques are described in [9]:

2.3.1 Hop-by-Hop Back Pressure (Open Loop)

When Congestion occurs, node sends BP in CODA. BP signals are sent to source by upstream nodes. In dense network BP message will be directly sent to source. Local state can impact BP signals. As an upstream node receives message it will not further propagate BP based on its own condition (means there in no chance of congestion at this node).

2.3.2 Multiple Resource Regulation

In CODA the control closed loop congestion control is triggered when the source event rate is greater than throughput of the channel, and if it is less than the throughput of channel, source regulate itself. Whenever, threshold exceeds, the source node requires constant feedback from sink for the sake of maintaining its rate. After receiving acknowledgment, the sources maintain their own rates otherwise they lessen their rates [9].

2.3.3 Hop-by-Hop Flow Control (Network Level)

This technique uses packet header to store congestion bit. This bit is set by each node in outgoing data packet. This comprises of two schemes 1) congestion detection and 2) avoidance.

Detection schemes for this technique includes channel checking or sampling and queue occupancy. Whenever fraction of available limit in the buffer of the downstream node goes below a threshold value, then the value of congestion bit in the header of the outgoing packet is set to 1. This causes immediate upstream node to stops forwarding more data. This allows down-stream node make empty its queues. If whole path is suffering from same condition then this causes BP to reach to the root up-stream node.

Problem: whenever a child node sets congestion bit its immediate parent stops sending more packets. In a result, grandparent of upstream node may not be informed about the bottleneck and keep transmitting data, causing packet loss.

Solution: parent node is allowed to send one extra packet with congestion header. This causes its parent to overhear it and do same as its child that is stop.

Benefits: It lessens the packet loss rate and restrict wasteful transmission of data units which will ultimately dropped at downstream node [1].

2.3.4 Limiting Number of Transmissions

A live-lock problem occurs, when outgoing packet is dropped across multiple hops due to congestion in the network. This is because diameter of WSN increases and causes substantial expanse of energy and bandwidth on a packet that is to be dropped. This technique is based on two impractical (but fair in some scenarios) assumptions. 1) upstream load by each wireless sensor is equal in size and 2) tree is not significantly slopped in one direction.
This technique uses a concept of token count. Whenever a node listens its parent transmissions of N packets it adds one in to its token count. This count continues until maximum number of token count is reached. Each packet transmission costs one token and it only can send its packet if its token count is greater than zero [1].

2.3.5 Layer-2 Prioritization

Problem with the network layer congestion control mechanisms is that they are not well aware of MAC layer and not so quick to alleviate problems packet losses without getting help of MAC layer. Problem occurs when multiple up-streamer to a single node sends packets parallelly and before the down-streamer can take wireless media access its queue size becomes full. This causes packets to be dropped at congested node. To alleviate this Aad and Castelluccia [1] proposed a technique known as prioritizing the MAC. This works on simple rule that “set randomized back-off timer of the congested node one-fourth of the back-off timer of the non-congested node”, so that congested sensor can have more chances or priority of accessing media.

2.3.6 Fusion

It is also proposed by [1] and combines two network layer level and one MAC layer level CC technique. Network layer level techniques are hop-by-hop and rate limiting. While MAC prioritizing operates at layer2. In genuine network loads fusion often performs well by a factor of three than the above three working individually.

2.3.7 Bird Flocking

Bird flocking is one of the solution for Constraint applications. It ensures better results by minimizing data latency, minimizing packet loss ratio and maximizing packet delivery ratio. Congestion control techniques that are used in internet will not be used in IoT. In IoT the chances of congestion at intermediate node are increased due to many to one communication. Moreover, the technique of end-to-end control is also not appropriate as when the node receives the response from end-point, the intermediate node will also suffer from congestion. It works on network layer in the form of hop by hop to control the congestion. Zone of repulsion and Zone of attraction are the two regions. Each node contains buffer which is linked with the focal node (the node over which the reasoning is made) so that the decision can be made on the next hop for its packets. There will be one hop control message in Zone of repulsion and two hop control messages for the nodes in zone of attraction. Qs ZoR and Qs ZoA are the parameters for each successor node S. where Qs ZoR represents the ratio of buffer filling in zone of repulsion and Qs ZoA represents ratio of buffer filling in zone of attraction [5].
Many other techniques have been discussed for congestion control in [3], [6], [8] ; which are discussed in the following text.

2.3.8 Gripping

In griping the receiver sends back-pressure (BP) message to the source that sends the datagram packet, if the queue length of network layer is greater than threshold. Time duration between two consecutive BP messages is K seconds for those that are transmitted by same source. Transmitter works according to additive increase and multiplicative decrease approach. It sends W packets in every time slot and if it receives the backpressure message than it reduces W to half as a result the transmission rate also halved. If no back-pressure message is received during the specific time (T) than it uses additive increase approach for W. The packet is lost at receiver side if the queue of network layer is full whereas the sender discards the message if it is not successfully received during the permissible retransmissions.

2.3.9 Deaf

When the network layer queue length exceeds the threshold, or it is full the receiver stops transmission of layer2 acks. Whereas the transmitter uses back off timer (T wait). It waits for T seconds before retransmitting the packet. After every failure the wait time for retransmitting the same packet is doubled, as a result the transmission rate is halved. The failure occurs only when maximum retransmission count has reached for a specific data packet and despite of these retransmissions, packet is still undelivered.

2.3.10 Fuse

It is a hybrid of both above techniques. The receiver works like gripping until the queue length is less than maximum length of the queue and when the queue of network layer is full than it stops sending layer 2 Acks mimic deaf technique and sends the notification message of congestion that is BP messages (just like a gripping device). The transmitter also uses the additive increase and multiplicative decrease approach.

3. Evaluation Metrics

There several measures that are used to measure the performance of traffic in congested environment in wireless sensor network. These parameters help us to evaluate the network performance under congestion. The list of these parameters is given below.

• Efficiency of Network: It determine the resources(energy) wasted on those packets that are not delivered. As distance from the sink varies this results in varying network efficiency [4].

• Energy Efficiency(EE): EE can be described as the packet delivery ratio [7].

• Energy Tax(ET): In WSN, ET is the proportion of the packets successfully reached at sink and the packets lost in network [2].

• Packet Loss Ratio or Delivery Ratio: Due to buffer overflow and error in packets some packets are lost this determine the packet loss ratio. In literature at some places it is determined through number of retransmissions instead of number of packet dropped [7].

• Fairness: It is about allocation of bandwidth. It represents the changes(variations) in sending rates. It is dependent on fair allocation of bandwidth [1], [7].

• End-to-End Delay: End to end delay is also a good evaluation measure of how our network is congested. End to end delay means total time that has been spent on delivery of packet at based station from the time it was generated [7].

• Control Packet Overhead: Control packet overhead is also a good measure which include the cost of sending such packets that contain the control information due the protocol requirements [12].

• The Total Throughput at the Sink: The total packets that are successfully received at sink node in a time [7].

• Instantaneous Queue Size: It demonstrates the steadiness or variance of buffer. Event based weighted queues also used [4].

• Memory Requirements: The length of the queue, code, and the number of sensing units determined the memory requirements [4].

• Fidelity Index: It is the fraction of the number of packets targeted to be received by the application, to that properly received [4].

CODA and hop-by-hop flow control schemes are useful for minimizing Energy tax (ET) on average and reduces wasteful transmissions of the packets. Rate limiting CC scheme is good at removing serious unfairness to the packets that have to be travel along multiple hops. Fusion is best at improving efficiency and fairness of the network. As bird flocking congestion avoidance and control scheme uses multiple paths for packet flows, it ensures least packet loss ratio. Backpressure CC techniques focus on exploiting current IoT protocol stack for both unidirectional and bidirectional type of communication. Fuse is best performing among the BP congestion control schemes but it requires cross layer design. Deaf is a potential candidate in terms of practical and implementation purposes as does not require any implementation of additional BP massages and requires least modification in current IoT protocol stack at cost of five to ten percent decay in throughput. Backpressure techniques focus on improving throughput and minimizing packet loss ratio.

4. Future Work

First, although the rate limit method presented in [1] is good for better results in terms of fair allocation of bandwidth when network under load, but a variable send rate, and without assumptions there is need for more general approach which can handle variable rates in better way.

Second, when considering RPL/6LoWPAN protocol stack a study is required for congestion avoidance that focus on “parent changing model”. When child’s buffer exceeds a tunable threshold, it should change its parent with second preferred parent. But a comprehensive study is needed to determine a robust algorithm for this threshold value which can optimize average number of parent changes.

Third, BP congestion control schemes further need a robust algorithm for automatic tuning of threshold values according to network type.

References

[1] 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 (pp. 134-147). ACM.

[2] Wan, C. Y., Eisenman, S. B., & Campbell, A. T. (2011). Energy-efficient congestion detection and avoidance in sensor networks. ACM Transactions on Sensor Networks (TOSN), 7(4), 32.

[3] Castellani, A. P., Rossi, M., & Zorzi, M. (2014). Back pressure congestion control for CoAP/6LoWPAN networks. Ad Hoc Networks, 18, 71-84.

[4] Kafi, M. A., 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.

[5] Hellaoui, H., & Koudil, M. (2015). Bird flocking congestion control for CoAP/RPL/6LoWPAN networks. In Proceedings of the 2015 Workshop on IoT challenges in Mobile and Industrial Systems (pp. 25-30). ACM.

[6] Betzler, A., Gomez, C., Demirkol, I., & Paradells, J. (2013). Congestion control in reliable CoAP communication. In Proceedings of the 16th ACM international conference on Modeling, analysis & simulation of wireless and mobile systems (pp. 365-372). ACM.

[7] Yin, X., Zhou, X., Huang, R., Fang, Y., & Li, S. (2009). A fairness-aware congestion control scheme in wireless sensor networks. IEEE transactions on vehicular technology, 58(9), 5225-5234.

[8] Betzler, A., Gomez, C., Demirkol, I., & Paradells, J. (2015). CoCoA+: An advanced congestion control mechanism for CoAP. Ad Hoc Networks, 33, 126-139.
[9] Wan, C. Y., Eisenman, S. B., & Campbell, A. T. (2003). CODA: Congestion detection and avoidance in sensor networks. In Proceedings of the 1st international conference on Embedded networked sensor systems (pp. 266-279). ACM.
[10] Zhang, H., Arora, A., Choi, Y. R., & Gouda, M. G. (2007). Reliable bursty convergecast in wireless sensor networks. Computer Communications, 30(13), 2560-2576.
[11] Wang, G., & Liu, K. (2009). Upstream hop-by-hop congestion control in wireless sensor networks. In Personal, Indoor and Mobile Radio Communications, 2009 IEEE 20th International Symposium on (pp. 1406-1410). IEEE.
[12] Sheu, J. P., & Hu, W. K. (2008). Hybrid congestion control protocol in wireless sensor networks. In Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE (pp. 213-217). IEEE.