Congestion Control Based on Reliable Transmission in Wireless Sensor Networks

Sun Hua
Yancheng Institute of Technology, Yancheng, Jiangsu, China
Email: sunhuashao@163.com

Abstract—A new congestion control method is proposed in this paper. It makes use of the combination of buffer queue length and its variation rate to estimate the degree of congestion. It divides the node states and adopts various bandwidth allocation strategies according to different states, which ensures the reliable transmission of emergent information. Firstly, this paper introduces the queue model with priorities, divides the services according to different data flows and makes a detailed introduction on congestion adjustment mechanism based on the above information. The simulation experiment shows that the congestion control mechanism in this paper can ensure the reliable transmission of emergency information in monitoring sensor networks.

Index Terms—Reliable Transmission; Congestion Control; The Queue Variation Rate; Wireless Sensor Networks

I. INTRODUCTION

Along with Internet application service being widely and deeply used and its diversity becoming increasingly complex, the problem of internet congestion control becomes more and more serious. Most of LTE network load mainly depend on TCP/IP protocol. If the network load [1] cannot be restrained effectively and the node cannot restore its normal state from the congestion state, this will cause network congestion, even disrupt whole network. Therefore, how to help the TCP congestion control adapt to LTE dynamic network topology and ensure the stability and robustness of the network[2-3] has become an essential question to be solved as soon as quickly.

In literature [4], the AODV protocol based on the dynamic detection scheme applied in mobile networks is studied. In literature [5], the AODV routing recovery mechanism in mobile Ad Hoc network is analyzed. In literature [6], the on-demand routing mechanism based on the directed flooding in wireless Mesh network is analyzed and its effectiveness is also proved, but do not consider the gain produced by the cooperative communication scheme. Relevant protocols for the network congestion control are also deeply analyzed and studied in literatures [7], [8] and [9], such as throughput and delay, considering never the influence of the routing protocol on congestion control.

Wireless sensor network is a wireless network composed of a group of sensors, which are deployed artificially or randomly in a special area, collects, transmits and merges the surrounding information with self-organization and cooperative sensing. It is a large-scale, low-cost, unattended, non-infrastructure, automatic networking, robustness and invulnerability network, and has been well concerned because of its unique technical advantages and good prospects in many monitoring applications such as ecosystem detection, volcanic activity, fire, toxic gas leakage and so on. Especially when there is a urgent event in the monitoring area, the sensor network can synthesize the collected information from various angles and directions, assemble information to policy-makers with reliable data and useful information timely and accurately, and provide the information source of emergency event.

Differing from wireless sensor network application based on period monitoring, the collected information form monitoring wireless sensor network for emergency event has the following characteristics: (a) burstiness. The emergency event usually is occurred instantly, so the generated data has its inherently bursty nature; (b) timeliness. It is important to ensure the information can be broadcast for users in minimum time, so as to take measures timely, thus it is necessary to ensure information transmission timeliness. (c) diversity of reliable demand. Under normal conditions, most of multi-attribute data streams are produced in emergency event, such as temperature, humidity, position, picture, where some information( the location information of emergency event) for users is of special importance, which these information required reliable message transfer service are called the emergency information, while other reliability could be reduced properly. Due to these characteristics, congestion is not inevitable. In order to make sure the reliability of emergency information transmission, it is necessary to design appropriate congestion control mechanism.

In general, congestion control is divided into two parts: congestion detection and rate regulating. At present many studies about congestion control mechanism in wireless sensor network have been performed. Most of studies use the ratio of the length of buffer queues, the channel load or the rate of server node to the arrival rate as the measure of the congestion control, nevertheless, The essence of congestion harm to network is because the overflow of the queue length cause the packet dropout, but these measures cannot correctly reflect the overflow of the node queue length. Moreover, traditional
congestion detection assume the congestion is occurred instantly (namely, exceed a critical value), while neglecting the fact that the congestion is a consecutive, gradual process. Therefore, it is necessary to redesign the congestion detection method. Once the congestion is detected, appropriate rate regulating measures should be taken in order to reduce congestion and avoid the deterioration of the network environment. The traditional congestion control mechanism will adjust the original rate after having detected the congestion, so as to decrease the data quantity transmitting to network for reducing congestion. However, this will reduce the reality of an event in safety monitoring applications. So downstream bandwidth can be redistributed after the congestion is occurred, so as to ensure the reliable transmission of emergency event without decreasing the reality.

The reliable transmission of the congestion control mechanism for emergency information (CCM-UI) is designed in this paper, which adopts queue transmission idea with priorities. Each node judges its degree of congestion by the congestion detection mechanism, determines its work state (normal or emergence) according to different degrees of congestion, and further makes use of various bandwidth allocation strategies. The simulation experiment demonstrates that, compared with traditional congestion control mechanism, the proposed congestion control mechanism in this paper can keep the stability of network throughput and make sure of the reliable transmission of emergence information simultaneously.

II. PROPOSED SCHEME

When the network load of TCP source server and core network SAE is substantially exceeding process ability and radio resource capacity, the network congestion will happen. So the main reasons causing congestion are as follows:

The relay forwarding node with limited storage capacity. When many end-to-end transmission services share the same relay node to realize data forwarding, the node produces a service queue. If the storage capacity of the node cannot satisfy the demand of multi-service transmission, this will easily cause data packet loss. For burst data stream and real-time streaming media data services, it is unable to provide effective protection. Therefore, it is necessary to establish an effective mechanism of relay selection.

The wireless link with limited bandwidth. Link level low bitrate channel can not adapt to upper high-speed data stream service, which will easily cause network congestion. If the rate of data stream from TCP server is larger than the channel capacity of wireless link, the reliability of the wireless TCP p2p communication can’t be guaranteed. Simultaneously, the low-speed links in the relay forwarding will become a “bottleneck” of p2p communication, which cause network congestion. Therefore, it is necessary to establish a relaying decision protocol and routing update mechanism with selection ability.

The radio resource in LTE network with uneven distribution. The distributed network mode and the uneven data stream distribution, because the unbalanced distributions of the LTE radio resource and the unbalance load. Therefore, it is necessary to study a congestion control mechanism based wireless TCP in cooperation mode.

And, furthermore, applying congestion control in high-speed wireless environment based on LTE network has some following problems that need to be solved as soon as quickly.

The p2p delay. When the p2p connection between TCP source server and core network SAE is formed in early period, the wireless channels have not been fully utilized. The users will always be waiting for network application serves, which increase the delay in data transmission.

The transmission rate of the wireless link. The burststones and real-time of data stream make the service rate in LTE network show various dynamic characteristics. Therefore, the data sending rate is increased in TCP source server, while the network is in the congestion state. The resource need redistribution in p2p communication and the communication always is waiting. Or the TCP source server decrease the rate, which cause the Link level buffer queues collapse in relay node, increase their packet error rate, and shapen delay jitter.

The ACK packet flooding. In LTE network, the working frequency of wireless links of the upstream channel and downstream channel is different, in order to ensure the upstream and downstream data stream is rapidly weaken through the different and independent channel. Simultaneously, the transmitted ACK data packet has multiplied in the upstream channel and downstream channel. The more the data stream serves are, the more flooding the data packets are.

The outage probability of wireless links. When mobile node in LTE network applies the roaming to switch community, the base station or the refraction signal can not be connected, which will cause the wireless link is interrupted. If the routing selection can not be applied or roaming area has a large amount of residential users, it is unable to provide effective protection for mobile nodes.

A. Priority-based Queue Model

It is not a new method to adopt the priority-based transmission idea. Protocols of PCCF and QCCP-PS, etc, all use priority queue and provide different Qos guarantees for different data types. To be simple, each node has three kinds of priority queues: High, Median and Low. The maximum queue length is max_queue(i ∈{H,M,L}) . According to the different importance, each queue is given different weight values: \( w_H, w_M, w_L \) \((w_H > w_M > w_L > 0)\). Figure 1 shows the priority-based queue models of nodes. When a data packet arrives, it will be added to the tail of different queues according to the data packet’s types. All queues are FIFO (First-In-First-Out) queue. We set emergent information as the highest priority; therefore, the data of HPQ (High Priority Queue) will get priority service to
make sure of emergent information’s low latency and high reliability. Other queues (MPQ, LPQ) will not wait until the queues of HPQ are empty.

Figure 1. Priority-based queue model

B. Congestion Estimation

It is stated in the previous part of the paper that the most present congestion detection methods are all based on the single metric, for example, the present buffer length, channel occupancy, and the ratio of data packet’s service rate to the arrival rate, etc, which all cannot accurately indicate the load situation of the present network. In addition, the traditional congestion detection methods are all based on 0-1 model, namely, the node only has two work states: congestion and non-congestion. Obviously, it is not suitable for the real situation that different congestion control will be carried out on the basis of the above situation. Hence, a new congestion detection mechanism must be designed for a fine congestion control mechanism.

Because the data packet of HPQ will be served with priority and the High Queue Length (HPL) will change at first in network congestion, we make use of High Queue Length to finish the congestion detection reference. Definition: the queue variation rate is defined as the following:

$$\rho = \frac{HQL_{now} - HQL_{last}}{\text{max}_\text{queue}_{\text{HPQ}} - HQL_{now} + 1}$$

where, $HQL_{now}$, $HQL_{last}$, $\text{max}_\text{queue}_{\text{HPQ}}$ denote the present, past and the maximum queue length of HPQ, respectively. Queue variation rate reflects the queue's variation trend: the larger value (positive) denotes the larger amplification of queue length and the longer length of present queue, which means the possibility of overflow at the next moment in node’s queue length is higher; Otherwise, if the value is negative, it denotes that the queue length becomes smaller (the denominator is always greater than zero.), namely, the network congestion is mitigated.

Based on $HQL_{now}$ and $\rho$, we define the node’s various work states as the following:

State 0:
$HQL_{now} = 0, \quad \rho = 0$.

State 1:
$HQL_{now} \leq (1-a) * \text{max}_\text{queue}_{\text{HPQ}}, \quad 0 < \rho < \Delta_{thr}$

State 2:
$HQL_{now} \leq (1-a) * \text{max}_\text{queue}_{\text{HPQ}}, \quad \rho < 0$

State 3:
$HQL_{now} \leq (1-a) * \text{max}_\text{queue}_{\text{HPQ}}, \quad \rho > \Delta_{thr}$

State 4:
$HQL_{now} > a * \text{max}_\text{queue}_{\text{HPQ}}, \quad \rho > \Delta_{thr}$

State 5:
$HQL_{now} > a * \text{max}_\text{queue}_{\text{HPQ}}, \quad \rho < 0$

where, $a$ is a constant, $0.5 < a < 1$; $\Delta_{thr}$ is a fixed threshold value, $0 < \Delta_{thr} < 1$.

Under the initial state, a node is at state 0. Once the node receives the first data packet, the data’s queue length increases and the node gets into state 1. When data packet is served (forwarded or discarded) and the queue length declines, node gets into state 2. At this time, the node is at Normal State, namely, the state is transferred back and forth between state 1 and 2. When emergent event occurs and a number of data inject the network in a short time, the queue length increases quite fast. When the queue variation rate $\rho$ is greater than $\Delta_{thr}$ ($\rho > \Delta_{thr}$), the node gets into state 3, which denotes that the node gets into Emergent State. Simultaneously, the node has the probability of congestion and it may trigger data packet loss in the next scheduling period, so the Normal Rate Adjustment (NRA) is started. If the queue length continuously increases fast, until $HQL_{now} > a * \text{max}_\text{queue}_{\text{HPQ}}$, the node gets into state 4 now, which denotes that the probability of the node’s congestion in the next scheduling period is higher and the more active rate adjustment (the second order rate control) need to be adopted to mitigate or avoid congestion. State 5 denotes that node’s queue length starts to decline and congestion begins to be mitigated and the rate control is effective. Figure 2 shows the diagram of node state transition.

C. Rate Adjustment

From section B, we can see that node judges its state based on the congestion detection method of queue variation rate and further adopts different rate control mechanisms. According to node’s work state, the rate adjustment methods in this paper are divided into two parts: Normal Rate Adjustment (NRA) and Emergent Rate Adjustment (ERA).

When emergent events don’t occur, network is in periodical data collection and node is at normal work state which is transferred back and forth between state 1 and 2. In order to use the network resource effectively, node should adjust its rate adaptively according to priority, and allocate various service rates according to the different data types with different weight values at the same time. In [14], the author proposes a quite fine rate adjustment mechanism based on the node’s priority,
which is similar to the rate adjustment method in this paper. Suppose the service rate of destination node (Sink node) obeys exponential distribution, and we can get the average service time of Sink node with the method of Exponentially Weighted Moving Average (EWMA):

\[ t_{\text{sink}} = (1 - \alpha) * t_{\text{sink}}' + \alpha * t'_{\text{sink}} \]

where, \( \alpha (0 < \alpha < 1) \) is a constant, \( t_{\text{sink}}' \) is the average service time of the present Sink node, \( t_{\text{sink}}' \) is the average time of the Sink node in the last record. Hence, we can get Sink node's average service rate:

\[ r_{\text{sink}} = \frac{1}{t_{\text{sink}}} \]

Then, from Sink node, each node \( i \) is based on itself and its child node’s overall priorities, calculates and broadcasts each child node \( j \)'s maximum transmission rate:

\[ r^j_{\text{max}} = r^j_{\text{parent}} * \frac{GP^j}{GP_{\text{parent}}} \]

where, \( GP^j_{\text{parent}} \) and \( GP^j \) are the overall priorities of father node \( i \) and child node \( j \), respectively. Suppose \( C^i_{\text{parent}} \) is the child node set of node \( i \), then we can get the following equation:

\[ GP^j_{\text{parent}} = \sum_{j \in C^i_{\text{parent}}} GP^j \]

In each next scheduling period, node \( i \) calculates its input rate \( r^i' : r^i' = \sum_{j \in C^i} r^j \), and then calculates the rate difference \( \Delta r^i : \Delta r^i = \beta * r^i' - r^i \), where \( \beta \) is a constant.

On the basis of this, the child node \( j \) of node \( i \) adjusts its rate again according to \( \Delta r^i : r^i = r^i' + \Delta r^i * \frac{GP^j}{GP^j}, j \in C^i \)

where, \( C^i \) is the child node set of node \( i \).

When node is transferred from state 2 to state 3, \( \rho > \Delta _{\text{max}} \), node’s queue length growth rate increases and the packet loss probability of overflow increases, so we need to adopt the relevant rate adjustment in order to decrease queue’s increase rate. According to node’s various states, we adopt multilevel rate adjustment methods under emergent work state.

1) The first level rate adjustment

Once node gets into state 3, it is predicted that the network traffic increases and at the next moment the data packet may increase quickly, and the first level rate adjustment is started.

node \( i \):

\[ r_i = r_i' * (1 + \lambda_1 * \rho) \]

Sink:

\[ r_{\text{sink}} = r_{\text{sink}}' * (1 + \lambda_2 * \rho_{\text{sink}}) \]

where, \( \lambda_1, \lambda_2 \) are separately the constant factors of the first level rate adjustment. The upper limit of rate adjustment depends on the real application.

When node is at state 3, the rate increases with the growth of \( \rho \). The larger \( \rho \) denotes the more serious network congestion. Hence, the node should be allocated more bandwidth from the father node to mitigate congestion.

The first level rate adjustment can mitigate or avoid congestion in a certain way. However, when the queue length continuously increases to \( \alpha * \text{max queue}_i \), node will get into state 4. At this time, the node’s congestion will become serious. If the stricter measures are not taken, it may produce data packet loss of HPQ as a result of queue overflow. Now the node adopts the second level rate control mechanism:

node \( i \):

\[ r_i = r_i' + r_i * \max(\lambda_3 * \rho, 1) \]

Sink:

\[ r_{\text{sink}} = r_{\text{sink}}' + r_{\text{sink}}' * \max(\lambda_3 * \rho_{\text{sink}}, 1) \]

It is noted that, once rate returns to the normal work state (state 1 and 2) after the congestion is mitigated, the node’s rate will be calculated again according to equation (3) at this time.

III. THE SIMULATION ANALYSIS

This section conducts the instance simulation analysis on protocol. In the process of simulation, we don’t consider the concrete details of MAC layer, so we ignore the congestion in link layer and only consider the spillover congestion of node queue. Based on C language, we develop the simulation software of discrete events, test the performance indexes, including the protocol’s throughput, delay and packet loss rate, etc and make a comparison with the classical CCF, PCCP and QCCP-PS. Figure 3 is the topology of the instance simulation.

Figure 3. The topology of the simulation

Figure 4. Total normalized throughput
In order to better estimate the utilization ratio of idle resources and the transmission performance in the emergency event, we suppose node 4 will be closed in 10s-40s, namely, the source rate becomes zero. However, during the period of 60s-90s, node 4 has detected emergency event, which produces the durative emergent data in large scale. The simulation time is 100s.

From Figure 4, we can see that the throughput of CCM-UI has a moderate fluctuation in the above abnormal time intervals, which shows the stability and has a better performance in throughput than other protocols. CCF [13] shows good performance in the normal time interval and then its throughput decreases obviously in the abnormal intervals of 10s-40s. The primary reason is that it does not adopt the effective utilization mechanism of idle resources, which leads to the resource waste and decreases the throughput.

From Figure 5, we can see that the throughput of CCM-UI has a moderate fluctuation in the above abnormal time intervals, which shows the stability and has a better performance in throughput than other protocols. CCF [13] shows good performance in the normal time interval and then its throughput decreases obviously in the abnormal intervals of 10s-40s. The primary reason is that it does not adopt the effective utilization mechanism of idle resources, which leads to the resource waste and decreases the throughput.

The transmission of emergency information not only needs to ensure the reliability, but also needs the promise of low delay. Figure 6 offers the queue delay of three priority data flows. From the figure, we can see that the delay is high during the initiation period of early protocol in network, and then the network’s delay gets decreased greatly. After the emergency event in 60s, the ERA mechanism is started to decrease the HP’s delay under 0.004.

IV. CONCLUSION

This paper designs a novel congestion control mechanism for the monitoring wireless sensor networks. Comprehensively considering the features of emergency information in monitoring network, the paper proposes the congestion detection mechanism based on the queue and the queue variation rate, divides the nodes into different work states according to the degree of congestion, and further adopts various rate adjustment mechanisms, which makes sure of the reliable transmission of emergency information in the occurrence of emergency event.
transmission of emergency information in the monitoring sensor networks.

REFERENCES


[31] Qian, Qingji, Xuanjing Shen, and Haipeng Chen. "An improved node localization algorithm based on DV-Hop


