Efficient Load balancing in MANETs to Improve Network Performance

Vinh Dien HOANG, Zhenhai SHAO and Masayuki FUJISE

Singapore Wireless Communications Laboratory,
National Institute of Information Communication Technology, Japan
{hvdien, zhshao, fujise}@nict.com.sg

Abstract—**It is expected that future ITS (Intelligent Transport System) network will involve a lot of mobile equipments and form a complex network. Maintaining the good performance of this complex network is a complicated task, and efficient load balancing plays an important role. In this paper, a new load balancing solution is proposed. In this solution, the load balancing is performed by the source node based on the probe packets sent by the destination node. Doing so will reduce the number of probe packets to half. A new formula for available bandwidth estimation of 802.11 ad hoc network is also presented. The simulation results show that the proposed solution has improved the connection throughput significantly.**

Keywords: MANET, Load balancing, bandwidth estimation.

I. INTRODUCTION

Future ITS network will be a network of multiple mobile equipments in a complex hybrid network consists of many mobile equipments. It's a hybrid network mixed between MANETs (Mobile Adhoc Network) and legacy networks. Some mobile nodes have direct connections to the legacy network (Internet, etc), whereas some nodes only have connections with other nodes in an ad hoc manner. In this network, data can be sent on multiple paths from the source node to the destination node. This is called the multi path routing [8-9]. Recent researches [1-4] have shown that the data flow between the source and destination MANET nodes could be speeded up if it is efficiently split on multiple paths between them. The source node will estimate how busy these paths are and send packets to these paths according to its estimation. This is the load balancing and the estimation is done using the probe packet.

The source node periodically sends probe packets on all paths to the destination and waits for these packets to be sent back by the destination. By measuring the information in the probe packet such as the round trip time, the number of congested packet, etc, the source node could estimate on which path it should send more data, and on which path it should send less.

Current load balancing solutions [8-9] require the sender to send probe packets to the destination and waits for these packets to be returned by the destination. It's obvious that by doing so the probe packets will have to travel two times over the same path, consume two times more of network resource and produce two times more overhead. On the other word, it's not efficient.

So, if the probe packets are sent by the destination only, the source node could still correctly estimate how busy the paths are because the information needed is still in the received probe packets. However the overhead and consumed network resources are reduced to half. This is the main idea of the proposed solution. The idea of using destination for sending the probe packet not only could be applied to the MANET but also in other types of network as well.

When the source node receives the probe packet sent by the destination node, it will estimate the available bandwidth on each path to the destination. Paths that have less available bandwidth are busier. From this information, the load balancing decision will be made. Paths that have more bandwidth will receive more packets. Using the available bandwidth information, load balancing will be performed more accurately compared to using the Round Trip Time (RTT) [8] and the delay time [9] since the RTT and the delay time are not as good as the available bandwidth information in judging which path is busier and which path is not. In this paper, a formula to calculate the available bandwidth estimation for 802.11x network is also presented.

This paper is organized as follows. Related works are presented in Session II. The proposed reactive load balancing solution is presented in session III. Session IV is the performance evaluation of the proposed solution based on NS2 network simulator [5]. Finally, session V is the conclusion of the paper.

II. RELATED WORKS

In [8], the authors extend the Dynamic Source Routing (DSR) protocol with the load balancing function. When the source node wants to send packets to the destination node, it will use the DSR protocol to find all possible paths to the destination. It will also measure the number of congested packet on each path based on the information sent by the destination node in the Route Reply (REP) packets. The source node then distributes the data packets on all paths in the way that the total number of congested packets on each path is equal. From time to time the destination node will update the number of congested packet by sending the Load Packet (LP) to the source node and the load balancing decision in the source node is made accordingly with this updated information.

In [9], the authors proposed the Multipath Source Routing (MSR) algorithm also based on the DSR to find all routes

from the source node to the destination node. Real-time information on each path is monitored using a feedback control mechanism. It requires the source node to send probe packets periodically on all paths and wait for these packet returned by the destination. It then measures the Round Trip Time (RTT) of these probe packets. From there, path delay will be calculated using Karn algorithm [10]. If the path has longer delay, it will receive less data packet and vice versa.

III. REACTIVE LOAD BALANCING

The proposed reactive load balancing requires two steps:

- Calculate the available bandwidth on each path using the probe packet

- Perform the load balancing based on the calculated available bandwidth on all paths.

A. Calculate the available bandwidth

The source node initiates the reactive available bandwidth measurement by sending a probe packet to the destination node to start the measurement process. After a timeout period, if no probe packet is received at the source node (the probe packet sent may be lost), it sends the probe packet to the destination node again. When the destination node receives the probe packet sent by the source node. It will reply by sending series of back to back probe packets on all paths to the source node. From now on, the reactive bandwidth measurement has been activated. After a period of time, the destination node – now is the sender of the probe packets, will send a series of probe packets to the source node again. The source node is now the receiver of the probe packets. These packets travel along paths from the sender node to the receiver node and produce gaps between them. These gaps are measured at the receiver node.

All these raw gaps and the estimated number of lost probe packets will be feed into a filtering module. The receiver node can estimate the number of lost probe packets because each probe packet has a unique sequence number. The filtering module will only consider the gap between adjacent probe packets. For example, if the probe packet sequence number (seq.) 3 is lost, the filtering module will only use the gap between the probe packet seq. 1 and the probe packet seq. 2. The gap between the probe packet seq. 2 and the probe packet seq. 4 will not be used because it could wrongly affect the estimation. Next, the filtering module will remove the biggest and the smallest gaps. Doing so will increase the probability of having more accurate estimation. $Gap_{measured}$ is calcualted as the mean of the remaining raw gaps. The number of lost probe packet will also affect the above calculated $Gap_{measured}$ by the percentage point. If 10% of probe packet is lost, the $Gap_{measured}$ will increase by 10%, and so on. From this $Gap_{measured}$, available bandwidth estimation will be calculated.

The frequency to send the probe packets is determined by the sender node based on the throughput of the connection and this frequency could also be changed by the receiver node.

B. Calculate the available bandwidth in 802.11 networks

Let's see how to derive the available bandwidth from the $Gap_{measure}$ in case of 802.11 network.

MANETs such as 802.11 network family share the wireless medium and use four ways handshake RTS/CTS/DATA/ACK mechanism to combat the hidden and exposed terminal problems.

In this mechanism, a node wishes to send data has to wait until the medium is idle and sends a small packet called RTS (Request To Send) which contains information such as data length, source, destination address, etc to inform other nodes that it has data to send. In other words, RTS is used to reserve the medium for the duration stated in RTS. The destination node once receives the RTS will reply with the CTS (Clear To Send) message to inform the source node that it's ready to receive the data for the duration stated in the RTS. Other nodes once heard the RTS and/or CTS will cease to access the medium during the duration stated in the RTS/CTS messages because they know that a transmission is on the way. The source node will then send the Data packet and the destination node will send the ACK to acknowledge that the data has been received successfully.
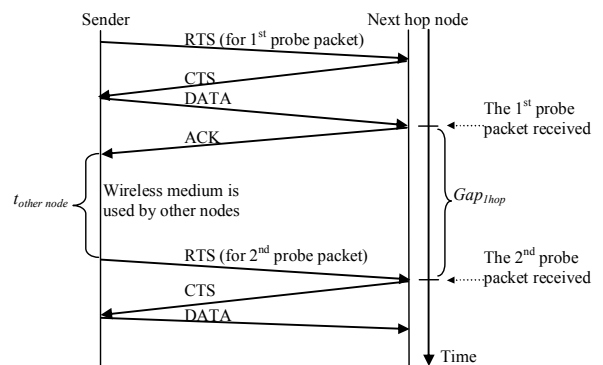

Figure 1: Transmitting probe packets in 802.11 MANET

When a node uses the reactive available bandwidth measurement, it will send the probe packets to the destination to activate the measurement process and the destination will send series of probe packets to the source in a regular interval. These probe packets sent by the destination to the source node are illustrated in Figure 1. The sender is the destination node. The next hop node is the next node on the way to source node.

To send the first successful probe packet, the sender needs to send the RTS to the next hop node, after that the CTS, DATA and ACK packet will follow as in Figure 1. The second successfully sent probe packet may follow the first probe packet immediately. But in general, t here is a delay between these two probe packets when the medium is used by other nodes or there's a contention in the medium. Let's call this delay time the $t_{other node}$.

Total time to transmit a probe packet from the sender node to the next hop node consists of the time to transmit RTS (t_{RTS}), CTS (t_{CTS}), the probe packet itself (t_{DATA}), the ACK (t_{ACK}) for the probe packet and the processing time as depicted in Figure 1. So the gap between the 1st and the 2nd probe packet in the next hop node is:

$$Gap_{1\ hop} = t_{ACK} + t_{other\ node} + t_{RTS} + t_{CTS} + t_{DATA} + t_{proccessing\ time} \quad (2)$$

t_{RTS}, t_{CTS}, t_{DATA}, t_{ACK} is the time to transmit and propagate the RTS, CTS, the probe packet, ACK. $t_{other\ nodes}$ is the duration in which the wireless medium is used by the sender's neighboring nodes. In the best case scenario, $t_{other\ nodes}$ is zero. In this case, the gap between the 1st and the 2nd probe packets in the next hop node is:

$$Gap_{best} = t_{ACK} + t_{RTS} + t_{CTS} + t_{DATA} + t_{proccessing\ time} \qquad (3)$$

$t_{other\ nodes}$ represents how busy the neighboring nodes are. If $t_{other\ nodes}$ is large then the available bandwidth for the sender node is small and vice versa.

If the next hop node is not the destination of the probe packets, the next hop node will forward these packets again until they reach the destination. The $t_{other\ nodes}$ is accumulated during this process and will be measured in the destination node by a gap between the two probe packets. This gap is called the $Gap_{measured}$.

Based on the $t_{other\ nodes}$ which is expressed in $Gap_{measured}$, the available bandwidth for the path could be calculated as follow:

$$Available\ bandwidth = \frac{Gap_{best}}{Gap_{measured}} \times B \qquad (4)$$

B is the currently maximum available bandwidth.

Gap_{best} is the smallest gap, calculates according to (3). For sake of simplicity, $t_{processing\ time}$ could be assumed as 0.

So by measuring the gap between the two back to back probe packets, the available bandwidth on the path could be calculated using equation (4).

C. Perform the reactive load balancing

When the source node already had the available bandwidth information on all paths from the source node to the destination, it will distribute the traffic accordingly to this information. If a path has more available bandwidth, it will receive more packets and vice versa. In the next section, simulation will be used to estimate the effectiveness of the proposed solution

IV. SIMULATION RESULTS

The simulations are done on the NS2 simulator. There're three set of simulations. The first two sets are used to evaluate the correctness of the available bandwidth measurement. The last set of simulation is to evaluate the effectiveness of the proposed reactive load balancing solution.
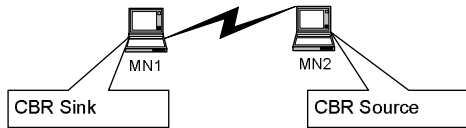

Figure 2: The first measurement simulation scenario

The first simulation scenarios are shown in Figure 2. It consists of just two mobile nodes running the 802.11 in the ad hoc mode (1Mbps ~ 125KB/s). The routing protocol used in the ad hoc network is the proactive routing protocol - Destination Sequence Distant Vector - DSDV [6]. DSDV regularly exchanges routing packets to all nodes in the network. Moreover, there's a constant bit rate - CBR

connection from MN2 to MN1 at the rate of 60pkt/s. Each packet is 512bytes in size. This means the CBR connection uses the bandwidth of around 42KB/s, including the RTS/CTS/ACK packets. So there're two types of traffic on the network: the DSDV routing traffic and the CBR traffic. All simulation lasts for 900 seconds.
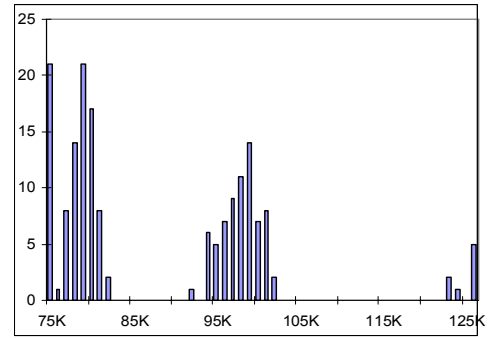

Figure 3: Histogram of reactive available bandwidth measurement in 1 hop with 60pkts/s CBR Connection

Figure 3 shows the histogram of the available bandwidth estimation simulation results for 802.11 network between two mobile nodes next to each other as illustrated on Figure 2. We can see that most of the available bandwidth measurement results are on the 90-100KB/s and 78-82KB/s band. The total bandwidth available is 125KB/s. The CBR connection uses 42KB/s. The remaining available bandwidth is 83KB/s on the average. So there're many measurement results are in the 78-82KB/s band. Moreover, there're moments that no CBR packet is transmitted, only the DSDV routing packets are exchanges in the network. DSDV routing packets consume much less bandwidth. Measurements at these moments will report higher available bandwidth – as shown in 90-100KB/s band on Figure 3. Also there're moments, although not so frequently, that no packet is transmitted. Measurements at these moments will have the maximum bandwidths as shown in the far right side of Figure 3.
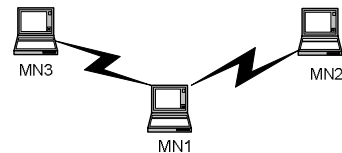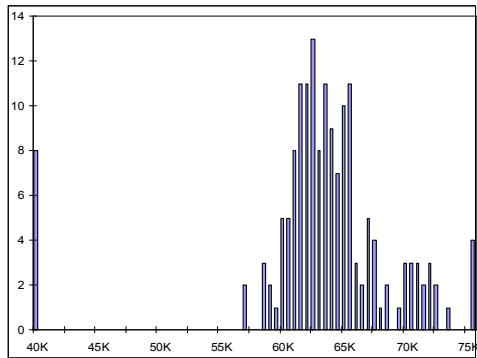

Figure 4: The second measurement simulation scenario

The second simulation scenario is shown in Figure 4. There're there mobile nodes in the 802.11 ad hoc networks. And there's no CBR connection between the mobile nodes. MN3 will measure the available bandwidth with MN2 which is 2 hops away. Other parameters are kept the same as in the first simulation scenario. So there's only the DSDV routing traffic running on the network.

Figure 5 shows the histogram of the available bandwidth estimation simulation results of the second simulation scenario. MN2 and MN3 are two hops away from each other so the maximum available bandwidth between MN2 and MN3 is 62.5KB/s, only half of one hop scenario. This is because the medium is shared between there mobile nodes. If

MN3 is transmitting a packet to MN2 through MN1, MN2 could only keep listening and vice versa. Figure 5 shows that the available measurement results are concentrated around the 62.5KB/s point which is closely matched the expected results. Measurements vary due to the DSDV routing traffic.

(b): Two hops, 0 pkts/s

Figure 5: Histogram of reactive available bandwidth measurement in 2 hops with no CBR Connection
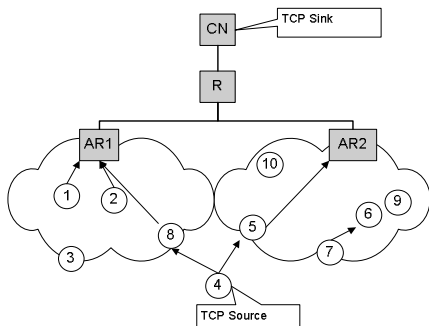

Figure 6: The third measurement simulation scenario

The last simulation scenario is shown in Figure 6. This simulation is used to estimate the effectiveness of the proposed reactive load balancing solution. To achieve a fair evaluation, simulation consists of both wired network (4 nodes) and wireless 802.11 network (10 nodes). Moreover, DSDV is used on the wireless part in the first simulation run. In the second run, reactive routing protocol – Ad hoc on-demand distance vector (AODV) [7] is used. DSDV and AODV represent the proactive and reactive routing protocols in MANET. Measuring the available bandwidth in both type of routing protocols will evaluate the proposed solution more accurately. Available bandwidth measurement module was integrated in the routing module. Mobile nodes move in a 450m x 450m area.

Figure 7 plots the throughput of a TCP connection from mobile node 4 to the correspondent node –CN in the fixed network. To make the simulation more realistic, there're 5 more TCP connections between mobile nodes. In Figure 7.a, Normal DSDV is the TCP throughput when using normal DSDV routing protocol. Multipaths DSDV is the TCP throughput when using two paths between the source node and destination node. The load is shared equally between two paths. Multipaths with LB DSDV is the TCP throughput when using two paths with load balancing. The load shared between these paths is based on their estimated available bandwidth.

Figure 7.b uses the same notions, except that AODV is used in the simulation.

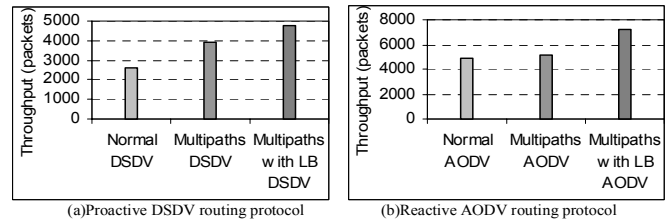(a)Proactive DSDV routing protocol   (b)Reactive AODV routing protocol

Figure 7: Throughput of TCP connection

Figure 7 shows that using the proposed load balancing solution in both proactive DSDV and reactive AODV routing protocol significantly increased the connection throughput.

V. CONCLUSION

In this paper, a new load balancing solution in MANET is proposed. The main idea in this solution is the probe packets used for bandwidth estimation are sent by the destination node. By doing so, these packets only have to travel one time on the path, which will reduce consumed network resource and increase the accuracy of the estimation. A new formula for available bandwidth estimation in IEEE 802.11 network based on the gaps between probe packets is also presented.

Simulation shows that the available bandwidth estimation matches the expected results. It also shows that using the proposed reactive load balancing solution, the connection throughput is increased significantly, in both proactive (DSDV) and reactive (AODV) routing protocols cases.

REFERENCES

[1] V.D. Hoang, M.H. Nguyen, Z.H. Shao, and M. Fujise, "A novel solution for global connectivity in MANET", Proceedings of IEEE VTC 2004 Fall, USA, Sep. 2004.

[2] A. Elwalid, C. Jin, S. Low, and I. Widjaja, "MATE: MPLS adaptive traffic engineering," Proceedings of INFOCOM 2001, Anchorage, Alaska, Apr. 2001, pp. 1300 – 1309.

[3] P. P. Pham, and S. Perreau."Performance analysis of reactive shortest path and multipath routing mechanism with load balance", Proceedings of INFOCOM 2003, Apr. 2003, pp. 251 – 259.

[4] D.S. Phatak, and T. Goff, "A novel mechanism for data streaming across multiple IP links for improving throughput and reliability in mobile environments", Proceedings of INFOCOM 2002, June 2002, pp. 773 - 781.

[5] K. Fall and K. Varadhan. ns Manual. The VINT Project. http://www.isi.edu/nsnam/ns/doc/.

[6] Charles E. Perkins and Pravin Bhagwat, "Highly dynamic Destination-Sequenced Distance-Vector routing (DSDV) for mobile computers", Proceedings of the SIGCOMM '94 Conference on Communications Architectures, Protocols and Applications, August 1994, pp. 234–244.

[7] C. E. Perkins, E. M. Belding-Royer, and S. R. Das, "Ad hoc on-demand distance vector (AODV) routing", IETF Internet Draft. http://www.ietf.org/internet-drafts/draft-ietfmanet-aodv-11.txt, June 2002.

[8] P. Pham, S. Perreau, "Multi-path routing protocol with load balancing policy in mobile ad hoc network", 4th International Workshop on Mobile and Wireless Communications Network, Sept 2002, pp. 48 - 52.

[9] L. Zhang, Z. Zhao, Y. Shu, L. Wang, O. W.W. Yang; "Load Balancing of Multipath Source Routing in Ad Hoc Networks", IEEE International Conference on Communications, May 2002, pp. 3197 – 3201.

[10] Kleinrock L. "Queueing systems", Vol.2. New York: Wiley, 1976