Congestion Avoidance in Source Routed Ad Hoc Networks

Bryan J. Hogan, Michael Barry, Sean McGrath
Wireless Access Research Group
University of Limerick, Ireland
bryan.hogan@ul.ie michael.barry@ul.ie sean.mcgrath@ul.ie

ABSTRACT:* *As well as suffering from the same congestion causes as fixed networks, wireless ad hoc-networks introduce several new congestion factors due to both the medium used and the mobility of the user. Congestion control in an ad-hoc network must be local, fast and adaptive. Congestion management attempts to predict when congestion is about to occur and reduces the transmission rate at this time. In an ad hoc network, congestion occurs locally and must be dealt with in a localised fashion. When a node becomes congested, it alerts its neighbouring nodes and these backoff according to the volume of data they are forwarding through the congested node. This paper demonstrates the effectiveness of such congestion control techniques in ad-hoc networks, showing improvement over data delivery rates provided by standard Ad Hoc techniques.

INTRODUCTION

Mobile Ad Hoc network users share resources, applications and information quickly and without the need or possibility of any centralised infrastructure. Applications of ad hoc networks include the emergency services and military sectors. Mobile ad hoc networks come together to form a network without any centralised control. Because of the limited transmission range of wireless devices, data hops across the network from source to destination, making use of intermediate nodes. Ad hoc network nodes accept responsibility for managing the network and routing of data. Routing of data, the most intensive task, is performed in a distributed manner. To send data to a destination who is not an immediate neighbour, a node must first find a route to that destination. Intermediate nodes co-operate to forward packets from the source to the destination.

Congestion occurs in a network or a portion of a network when the total amount of data sent into the network exceeds the available capacity. Congestion manifests itself in excessive delay of packets and dropping of packets due to lack of buffer space. There are a variety of conditions that can contribute to congestion. These include but are not limited to volume of traffic being placed on the network, the underlying network architecture, the specifications of the devices in the network (buffer size, processing power, etc), packet size and the transport protocol in use.

ROUTING IN AD HOC NETWORKS

Routing protocols in ad hoc networks create and maintain routes between pairs of communicating nodes. Routing must deal with unpredictable node mobility patterns, radio transmission errors, and the entrance and exit of nodes. The routing protocol must be reliable and robust. It should also be bandwidth efficient since there may be limited availability in wireless environments. All this must be accomplished in a distributed and uncoordinated manner. The main types of routing protocol used are table driven, on-demand (source routed and hop-by-hop) and hybrid routing [1].

DSR

Dynamic Source Routing (DSR) [2][3] is an on-demand source routed protocol designed for mobile multi-hop wireless ad hoc networks. DSR provides two main functions – route discovery and route maintenance. Route Discovery is performed when a node wishes to send a packet to a destination for which it does not have a route. The route maintenance function identifies link failure on an active route. The source, whose packet precipitated the discovery of the link failure is informed of the link failure and updates its routing cache appropriately. Both route discovery and route maintenance operate on an on-demand basis. No data is exchanged between nodes in a periodic manner. DSR is not capable of monitoring or handling congestion. Because of this, should congestion occur, packet losses are dealt with by a higher layer, such as TCP.

Impact of Mobility

Ad hoc networks consist of mobile devices which are free to enter and exit the network at any time and move at any rate. This has implications for the performance of the network, most notably with regard to stability of routes and delivery of packets. Higher levels of mobility also lead to higher overhead as the routing protocol attempts to discover and maintain routes to ensure delivery of packets. In ad hoc networks with high mobility some routing protocols yield poor (long) routes [4].

Several routing protocols use the fewest number of hops on a route as the sole metric for selection of routes. As such, the shortest route is likely to comprise of the longest links. As the range of a transmission increases, the quality of the wireless radio channel will decrease giving slower and less reliable transmission, leading to more packets losses and retransmissions. Longer range transmissions lead to interference over a wider area, affecting a larger number of nodes [5]. These long range transmissions coupled with the low transmission rates mean that interference persists over an extended period of time.

Mobility and Route Selection

Node mobility has the greatest impact on available routes. As nodes move, certain routes become unavailable and new routes are formed. In a network where mobility is low, routes that are in use may become widely known and used by much of the network. Sniffing of data packets and routing packets, results in a strong bias towards existing known routes. Some nodes or areas of the network appear on many routes and are responsible for forwarding a large volume of data. Over a period of time, these become more and more widely known and used, and finally become congested. Existing protocols, including DSR, do not monitor congestion levels and make no allowance for this in route selection.

In a network where mobility is high, routes in use will change frequently. In this situation, routes cannot become heavily used or widely know before breaking. This can lead to the use of invalid, or stale, routes and a general reduction in throughput. In extreme cases during route discovery the route may no longer be valid by the time the route reply has reached the originator. Similarly, the originator may also never receive a route reply due to the collapse of routes [5-7].

CONGESTION

There are two general solutions to the problem of congestion, avoidance and control. Congestion avoidance attempts to predict when congestion is about to occur and reduces the transmission rate at this time. The algorithm should operate in such a manner to keep response time vs load and throughput vs load operating to the left of the location of the knee in figure 1 below.

Congestion control attempts to take fuller advantage of the network resources by transferring data at a rate close to the capacity of the network. The capacity of the network can be viewed as the point at which any increase in traffic will increase the delay but not the throughput. Congestion control algorithms, like that of TCP, attempt to increase traffic until the capacity of the network is reached, and then slow the transmission rate. Thus these algorithms attempt to operate to the left of the cliff in figure 1.

Congestion Control Mechanisms

Some of the more common congestion control mechanisms function as a demand reduction scheme incorporating [8]:

- Service denial: preventing further resource allocation.
- Service degradation: forcing users of the network to reduce the load they are placing on the network.
- Scheduling: users of the network schedule their requests to use network resources so as not to overload the network.

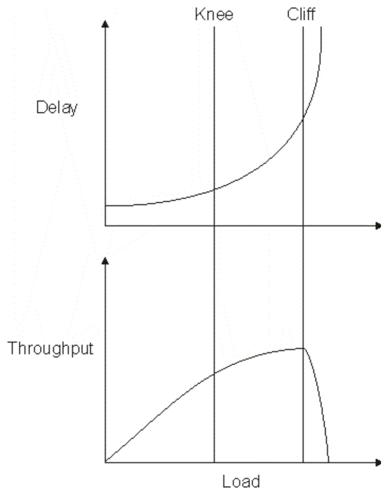

Figure 1 – Load vs Delay and Throughput

To allow each of these mechanisms to function, information relating to status of the network is required. Various approaches have been proposed and developed to facilitate this:

- Observation for evidence of congestion such as packet loss and delay [9].
- Delivery of feedback messages from point of congestion to source.
- Rejection of incoming traffic by congested node causing congestion to ripple through the network to the source.
- Congestion probe packets sent through the network.
- Packets can have feedback fields which indicate to the source the congestion status of the network.

For all of these mechanisms to function effectively they have to operate quickly, have a low overhead, attempt to be fair and allow the network to be used to its limit, i.e. not waste available bandwidth by implementing congestion rules that are too conservative.

Congestion Issues in Wireless Networks

Within wireless ad hoc networks a number of issues further complicate the identification and control of congestion including: [10]

- Interference from other nodes.
- Route failures.
- Variable quality of radio signals.
- Congestion.
- Transmitter power.

In the event of packet loss, appropriate action is not easily taken, as identifying the cause of the loss is difficult. There have been various mechanisms proposed to help classify the reason for packet loss, but all add extra complexity, may not be compatible with existing protocols and none seem to cover all possible causes.

CONGESTION CONTROL IN AD HOC NETWORKS

Dynamic Load-Aware Routing uses the congestion level of the nodes along the route as a primary route selection metric [11]. Its route discovery operates in a similar fashion to that of DSR. Unlike DSR, the destination determines the best route. Medium Time Metric may be used by any shortest path routing protocol [5]. Each link is assigned a weight that is based on the length of time it takes to transmit a packet on that link. These individual link weights are summed to give the weight for a route. The route with the lowest aggregate weight is selected. This method of route selection chooses links with short delay and high bandwidth, alleviating many of the transmission related congestion issues. Transmission times are shorter, require less power and cause less interference.

The quality of the communication channel is primarily determined by the range between the communicating devices. The longer the range a device transmits over, the lower the quality of the channel and consequently the speed at which communication may occur [5]. Long range links inhibit other nodes from communicating if they are within interference range of the transmitter/receiver pair. Negative effects are further compounded as these links remain active far longer than short range ones.

DSR’s route discovery procedure allows intermediate nodes to reply to route requests leading to a small number of routes becoming overused throughout the network. These routes tend to be composed of links that cover very long distances. As a consequence a small number of routes made up of long range, low throughput links are used throughout the network.

Localised Congestion Control Solution

Given that congestion is local, it should be handled locally – data should be pre-emptively routed around the site of congestion by the routing protocol, before the congestion causes loss of data and without reducing the transmission rate.

Nodes monitor the number of packets waiting in a queue to be passed to the network interface. If over a period of time the average number exceeds a certain level, the node considers itself to be congested. It immediately broadcasts, using its priority queue, a congestion notification packet to all neighbours.

All nodes monitor and record the volume of data that they are forwarding through each neighbour. If the average volume of data over a period of time for any individual neighbour exceeds a certain level, this node records that it has exceeded the forwarding limit for that neighbour. No further action is taken at this point.

On receipt of a congestion notification, the forwarding node checks if it has exceeded the forwarding limit for the congested neighbour. If so, the node forwarding must attempt to find an alternate route to the destination, avoiding the congested node.

These same rules are applied when a packet is originated from a source to a destination, with one caveat, if all routes available contain nodes who have recently broadcast congestion notifications, and the originator has exceeded the forwarding limit for these nodes, the originator must wait until an alternate route becomes available or until the node is no longer congested or the forwarding limit is no longer exceeded.

The purpose of the forwarding limit allows the algorithm to target those nodes that are forwarding large volumes of traffic through a congested node. As such, on receipt of a congestion notification, not every node will stop forwarding packets through the congested node. The aim is to gracefully alleviate congestion and prevent global synchronisation from occurring.

Simulation Model

Simulations use the ns-2 [12] simulator incorporating both the ns-2 provided implementation of DSR and a modified version using the congestion control algorithm. Nodes follow a random waypoint mobility model [3], travelling at a variety of speeds. Constant bit rate (CBR) flows are setup between random node pairs. The congestion control parameters in the model are the queue limit (*queueLimit)* and forward limit (*fwdLimit*) in a given node. In the standard version of DSR there is no *fwdLimit* variable and the maximum queue size is 50 packets. If the queue size is reached subsequent packets are dropped.

The congestion control algorithm requires a route with at least four nodes to become fully active. In a three node route the algorithm may partially operate. To facilitate routes consisting of at least four hops, the transmission range of the nodes and the size of the area in which the simulations are taking place must be considered. As shown in by the upper set of nodes figure 2, an appropriate width might be three times the transmission radius. But as can be by the lower set of nodes, this is not necessarily required.

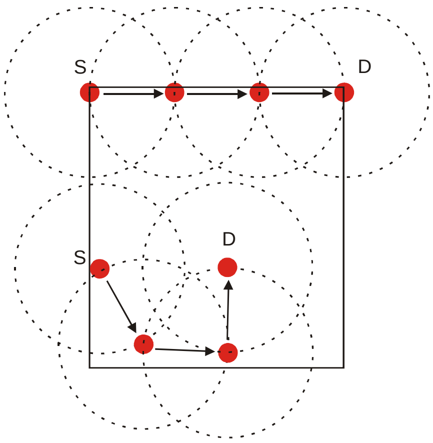

Figure 2 – Four hop configurations

RESULTS

A simulation consisting of 10 nodes, CBR flows with transmission rates of 100kbps (512 byte pkts) are setup between 5 source/destination pairs. The nodes move according to the random waypoint model at speeds of max 5m/s, pausing at each waypoint for 2s in a 600m x 600m plane. The *queueLimit* is set to 35 pkts and the *fwdLimit* set to 50,000 bytes. The route with the fewest number of hops is chosen, after applying the congestion rules. Figure 3, below, shows the number of times that nodes recorded congestion

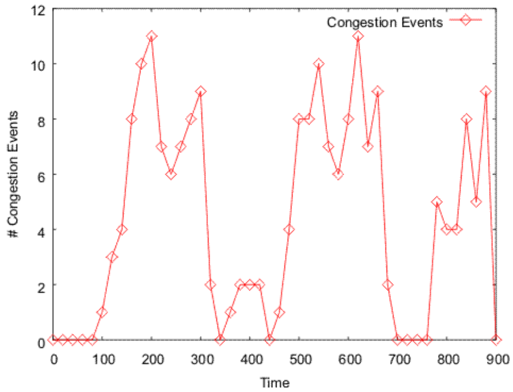

Figure 3 – Occurrence of congestion

Figure 4 shows the percentage of packets received when congestion control is active and inactive. It can be seen that the congestion aware version performed better. Note also that the percentage of packets received correlates with the occurrence of congestion. The congestion aware network is responding by routing packets along uncongested routes and by preventing the sources from adding further packets to already congested routes.

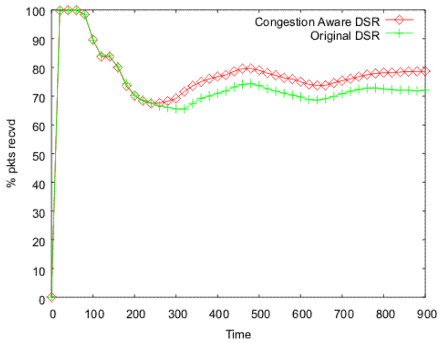

Figure 4 – Packet delivery rate

Figure 5 shows the percentage of packets received for a simulation with 15 nodes, 8 source/destination connections transmitting at a rate of 40kbps with a maximum movement rate of 10 metres/sec in a 700m x 700m plane. The simulation was run with two versions of the congestion aware algorithm and once with no congestion management. The first congestion aware scenario, DSR #1, has a *queueLimit* of 45 packets and a *fwdLimit* of 30,000 bytes. The second scenario; DSR #2, has a *queueLimit* of 45 packets and a *fwdLimit* of 70,000 bytes.

All three implementations perform approximately equally for the first 700 seconds. The second scenario DSR #2 shows a marked decrease in the percentage of packets received in the last 200 seconds. The first scenario aware DSR #1 slightly outperforms the standard implementation with no congestion management.

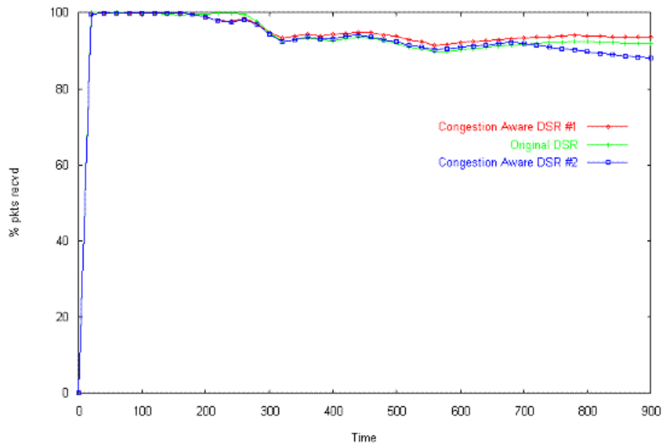

Figure 5 – Packet delivery rate

Figure 6 shows the percentage of packets received for the same simulation in its last 300 seconds.

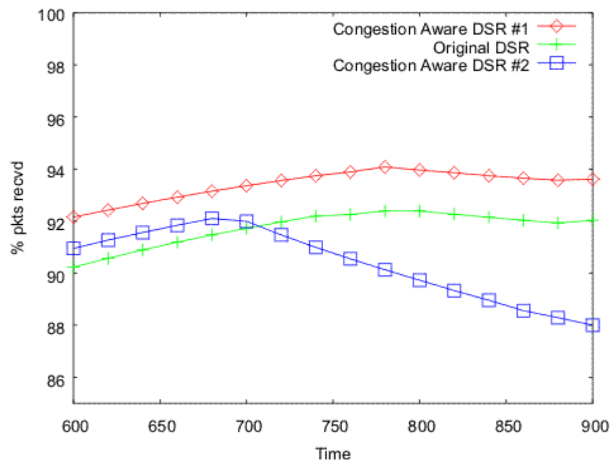

Figure 6 – Packet delivery rate (final 300 sec)

Slightly before the 700 second mark, the percentage of packets received by scenario DSR #2 begins to decline, while scenario DSR #1 and the standard version show no such tendency.

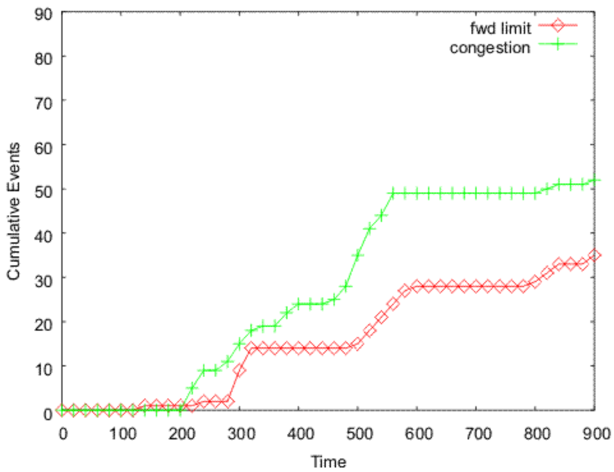

Figure 7 – Congestion events

Figure 7 above shows the cumulative occurrence over time of congestion events and the *fwdLimit* being exceeded for congestion aware DSR #1. There is a close correlation between the frequencies of occurrence of these events.

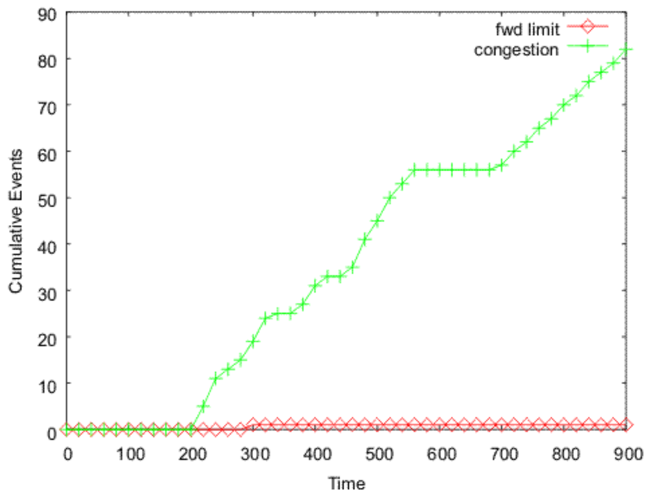

Figure 8 – fwdLimit exceeded events

Figure 8, above, shows the cumulative occurrence over time of congestion events and the *fwdLimit* being exceeded for congestion aware scenario DSR #2. Because of the high level of the *fwdLimit*, 70,000 bytes, it is hardly exceeded at all. It can also be seen that the number of congestion events are higher throughout the simulation and rise quite sharply in the last 200 seconds. From figure 6, it is apparent that it is also in the last 200 hundred seconds that there is a corresponding drop in the percentage of packets received by DSR #2. For this simulation, the *fwdLimit* of 70,000 bytes is too high a setting to allow the algorithm to operate and alleviate the congestion. For a similar simulation with a higher data rate, more favourable results would be obtained with this setting.

CONCLUSIONS

The approach presented in this paper uses a combination of local and source control. Initially congestion is handled locally and packets are re-routed using an appropriate uncongested route. In the event that this does not resolve the issue, congestion is dealt with by preventing source nodes from adding packets to the network using routes which are known to contain congested nodes.

In almost all simulations using the algorithm congestion is successfully managed and improvements in the percentage of packets received can be observed. The actual improvement varied depending on the *queueLimit* and *fwdLimit* variables selected.

Future extensions to the algorithm include mechanisms to dynamically alter the values of the *queueLimit* and *fwdLimit* to suit network conditions taking into account factors such as the average number of neighbours, known alternate routes and congestion history. This information would assist the algorithm to act in a more measured and appropriate manner, weighing up network conditions each time congestion occurs rather that following hard rules.

The congestion solution described in this paper is applicable to all ad-hoc network types. However, in hop-by-hop routed ad hoc networks, such as those using the AODV protocol, it may work more efficiently as routing decisions are made locally for each packet. By using this more up to date information in combination with local congestion knowledge a better routing decision can be made.

ACKNOWLEDGEMENTS

This work is conducted as part of the REALM project funded by Enterprise Ireland. Ericsson Ireland has supported Michael Barry's work on this project.

REFERENCES

[1] Elizabeth M. Royer and Chai-Keong Tok, "A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks," 1999.

[2] David B. Johnson and David A. Maltz, "Dynamic Source Routing in Ad Hoc Wireless Networks," in *Mobile Computing*, T. Imielinski and H. Korth, Eds.: Kluwer Academic Publishers, 1996.

[3] David B. Johnson, David A. Maltz, and Josh Broch, "DSR: The Dynamic Source Routing Protocol for Multi-Hop Wireless Ad Hoc Networks," in *Ad Hoc Networking*, C. E. Perkins, Ed.: Addision-Wesley, 2001.

[4] Josh Broch, David A. Maltz, David B. Johnson, Yih-Chun Hu, and Jorjeta Jetcheva, "A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols," presented at Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom), Dallas, Texas, 1998.

[5] Baruch Awerbuch, David Holmer, and Herbert Rubens, "High Throughput Route Selection in Multi-Rate Ad Hoc Wireless Networks," presented at to appear in the First Working Conference on Wireless On-demand Network Systems, 2004.

[6] Samir R. Das, Charles E. Perkins, and Elizabeth M. Royer, "Performance Comparison of Two On-demand Routing Protocols for Ad Hoc Networks," *IEEE Infocom* . 2000.

[7] Elizabeth M. Royer and Chai-Keong Tok, "A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks," 1999.

[8] R. Jain, "Congestion Control in Computer Networks: Issues and Trends," *IEEE Network Magazine*, vol. 4, pp. 24-30, 1990.

[9] Van Jacobson, "Congestion Avoidance and Control," *Computer Communication Review*, vol. 18, August 1988.

[10] Zhenghua Fu, Xiaoqiao Meng, and Songwu Lu, "How Bad TCP Can Perform In Mobile Ad Hoc Networks," presented at Proc. IEEE International Symposium on Computers and Communications (ISCC'02), Taormina, Italy, July 2002.

[11] Sung-Ju Lee and Mario Gerla, "Dynamic Load-Aware Routing in Ad hoc Networks," presented at The Third IEEE Symposium on Application-Specific Systems and Software Engineering Technology, 2000.

[12] Kevin Fall and Kannan Varadhan, "The ns Manual," 2003