On the Performance of Collision Avoidance Protocols in Ad-Hoc Networks

MICHAEL PATERAKIS AND JOHN ATZARAKIS

Department of Electronics & Computer Engineering
Technical University of Crete
Chania, Greece
pateraki, johnny@telecom.tuc.gr

ABSTRACT

In this work a channel access protocol for ad-hoc networks based on topology-dependent transmission scheduling named collision-avoidance time allocation (CATA), first proposed by Tang and Aceves [8], is extended and evaluated. Using packet-level simulations we examine various performance and design issues. Because network configuration plays an important role in system performance, our simulation results are based upon three network characteristics:

- Node population: Two different node populations have been simulated (eight and sixteen nodes).
- Transmission type: Both broadcast and unicast transmissions have been considered.
- Node connectivity: fully connected and partially connected network topologies have been simulated.

Finally we propose a new collision resolution algorithm for this protocol and compare its performance with that of Slotted Aloha for all the above network configurations.

1 INTRODUCTION

A mobile ad-hoc network is a mobile, multihop wireless network with no fixed infrastructure. The multihop topology of an ad-hoc network allows spatial reuse of the time division multiple access (TDMA) slots of the shared channel. Different nodes, which are sufficiently separated from each other, can use the same slot since they do not interfere with each other. The problem of assigning these slots to nodes is commonly referred to as *transmission scheduling*. In a scheduled access method, time is divided into fixed length slots, which are organized in cycles. Each cycle (or frame), contains at least one slot in which a node can successfully transmit or receive. Two broad classes of protocols exist in scheduling medium access by nodes in a wireless network:

- Channel-sensing based schemes (CSMA)
- Dialogue-based schemes (e.g. RTS/CTS dialogue)

One of the most popular MAC protocols in wireless local area networks is the Carrier Sense Multiple Access (CSMA). In CSMA, every node senses the channel before making an attempt to transmit. If the channel is idle, the node transmits otherwise it defers its transmission to avoid the collision with the transmitting node. Unfortunately, wireless networks typically have single hop connectivity with a base station but ad-hoc networks do not. In the ad-hoc environment, not all nodes hear each other and hence collisions may occur in spite of the use of CSMA.

Two types of problems arise using the CSMA protocol in an ad-hoc multi-hop network:

- *Exposed terminals*: Nodes that are out of the range of the receiver, but within the range of the transmitter.
- *Hidden terminals*: Nodes that are out of the range of the transmitter, but within the range of the receiver.

To overcome the hidden terminal problem in CSMA, several MAC protocols have being developed for ad-hoc networks that follow the dialogue-based scheme. Examples of such protocols are: MACA [1], MACAW [2], DBTMA [3], FPRP [5], HRMA [7], CATA [8] and IEEE802.11 [9]. All of these protocols use small control packets as handshakes to reserve the channel slots and avoid collisions in the data packets transmitted between nodes, since data packets are long and their possible destruction due to a collision can be very costly in wireless data resources. On the other hand, collisions in the control packets are not very costly in wireless data resources due to their relative small size. In addition, a timeout/backoff mechanism is generally used, to handle situations in which control packets have not been received correctly (or have not been received at all) due to collisions. This mechanism lowers the probability of future control packets collisions and increases the channel utilization as the channel reservation procedure speeds up.

Generally speaking, from all the above protocols developed for ad-hoc networks, CATA is a simple MAC protocol with the ability to support real time applications and collision free broadcast, unicast and multicast traffic, which makes it much more attractive than other MAC protocols. In the following sections we analyze the performance of CATA protocol for different number of nodes and network topologies. This paper is organized as follows. Network and channel and backoff algorithm issues are examined in section 2. Based on these issues, in section 3 we study the behavior and performance of CATA for eight network populations and various network topologies. The impact in performance of the backoff algorithm is examined in section 4 and finally in section 5 we present our conclusions and some ideas for future work.

2 NETWORK & CHANNEL MODELS AND BACKOFF ALGORITHM

2.1 System model and assumptions

The experimental results presented in the following sections, were obtained using an event driven simulation program build in C++ that simulates the reservation mechanism and behavior of the CATA protocol.

In our experiments both unicast and broadcast transmissions are examined. We assume that new, retransmitted or multihop[1] propagation requests to establish reservations arrive at each network node according to a Poisson process with average arrival rate of λ requests per slot. Each node has an unlimited first in first out (FIFO) buffer where newly arrived messages[2] are stored in. For simplicity we assume that each node can reserve at most one slot for data transmission in each frame. We consider variable message length and assume that messages arriving at a node have sizes according to a Geometric distribution with average message length (called AFL, for Average Flow Length) δ slots. This means that on average, it takes δ slots to transmit all data packets in a message. The communication channel is assumed to be error free, so that collisions of packets are the only source of errors.

2.2 Network topology

Node population plays an important role in the performance of the protocol. In general, as node population, N, increases, the maximum average arrival rate per node that a protocol can support decreases.

Node interconnection is also an important factor because it affects the interference/contention between nodes and the spatial reuse of the communication channel. In a fully connected network topology, all nodes are within transmission range of each other, while in a partially connected network, some nodes are within transmission range of others. Differences between fully and partially connected networks are shown in Table 2-1.

Fully Connected Network	Partially Connected Network
Higher interference / contention between nodes	Lower interference / contention between nodes
Complete channel state information	Partial channel state information
Symmetry	A Symmetry
Load balance	Load Imbalance

Table 2-1: Differences between fully and partially connected networks.

Although a partially connected network usually performs better than a fully connected, in terms of interference and contention between nodes, some times its performance is degraded due to partial channel state information and load imbalance. The following example explains this fact.

Figure 2-1 and Figure 2-2, show a fully connected and a partially connected network with four-nodes, respectively. Assume that node B has reserved the current slot in a previous frame and is ready to transmit a data packet to node A, while at the same time node C wants to transmit a data packet to node D in the same slot.

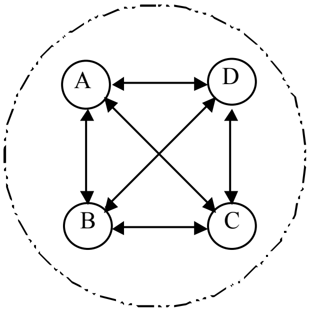

Figure 2-1: A 4-node, fully connected network.

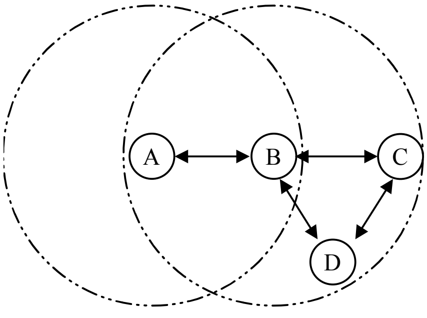

Figure 2-2: A 4-node, partial connected network

---

[1] A node randomly selects a one-hop neighbor node as its destination. Destination nodes outside the one-hop area are supposed to be covered by the transmission request arrival rates within their one-hop areas.

[2] All data packets, that must be transmitted by a node to one or multiple neighbors over a given collision free time slot, are referred to as *flow* or *message.*

Remember that in CATA [8], every node that receives data in the current slot sends a slot reservations packet (SR) in CMS1, which causes noise or is received by its neighbor nodes and prevents them from attempting to reserve the current slot. In addition, every node that sends data in the current slot sends a request to send (RTS) packet during CMS2 and causes interference to all neighbor nodes that did not hear the SR of the receiver node(s) in CMS1 and are trying to reserve the slot.

In our example, node A, which is the receiver, will transmit an SR packet and node B, which is the transmitter, will transmit an RTS packet, to prevent another slot reservation attempt. In the fully connected case, node C will hear the SR from node A and will know that this slot is reserved. Thus it will defer its transmission to the next slot, without reducing its slot reservation attempt probability. In the partially connected case, node C will not hear the SR from node A and sends an RTS that will collide with node's B RTS. Node C will assume then that another node wanted (not reserved) this slot also and will defer its transmission to another slot, reducing at the same time its slot reservation attempt probability.

From the load balancing perspective, in the fully connected case, packets arrive at each node with rate λ and are transmitted (with equal probability) to neighbor nodes with rate $\lambda/3$. In the partially connected case, node B transmits packets to its neighbors with rate $\lambda/3$, nodes C and D with rate $\lambda/2$, and node A with rate λ. Although the total network load is the same in both cases ($G=4\lambda$), load is not balanced among nodes in the partially connected case and this affects the packet waiting and service time for each node.

2.3 Frame length

Frame length is an important parameter for any MAC protocol based on time scheduling, because it directly affects delay and channel reuse. The frame length L for the fixed TDMA protocol in a network with N identical nodes is N slots.

For a node A to broadcast successfully using single-channel half-duplex radios, no node B within two hops from A can broadcast at the same slot as A does. Otherwise, A and B cannot receive the broadcast data packet send by each other if they are one-hop neighbors, or their common neighbors can experience a collision if A and B are two-hop neighbors. Therefore, for every node to broadcast successfully in one slot every frame, the frame length L required in CATA must be larger than the number of nodes in a two-hop neighborhood. This in the worst case equals to $Min\{d^2+1,N\}$ slots (CATA [8]), where d is the maximum node degree (number of neighbors a node has) of the network.

The worst case frame length for every node to unicast successfully in one slot every frame is also $Min\{d^2+1,N\}$ slots. Unicast transmissions can be considered as a special case of broadcast transmissions because a transmitting node A, instead of addressing a transmission to every (broadcast) neighbor node within one-hop, it can address it to a single (unicast) neighbor node.

In this work, all simulations use frame length equal to $Min\{d^2+1,N\}$ slots that is calculated dynamically according to the given network topology.

2.4 Backoff algorithm

CATA does not specify a backoff mechanism to handle control packets collisions. In order to lower the probability of future control packets collisions and increase the channel utilization, we propose a new backoff mechanism, referred to as the "**A**ccumulated **B**ackoff **A**lgorithm" (ABA), which works as follows:

- Every node has a backoff counter (bn) that sets to zero (bn=0) if its message queue is empty.
- When a new message arrives, the node sets its slot reservation attempt probability to one ($P_{reservation}=1$) and tries to make a slot reservation in the next available slot.
- If, and every time, a collision occurs by its slot reservation attempt, the node increases its backoff counter by one (bn=bn+1) and sets its slot reservation attempt probability to $P_{reservation}=(1/2)^{bn}$.
- If one of its competing one-hop neighbor nodes makes a successful reservation, it decreases its backoff counter by one (bn=bn-1), but it does not alter its slot reservation attempt probability. The slot reservation attempt probability is changed only if a collision is experienced during the node's reservation attempt and not by another's node successful reservation.
- When eventually the node makes its slot reservation and completes its message transmission, if its message queue is empty, it sets its backoff counter to zero (bn=0) and waits until a new message arrival occurs. If on the other hand, its message queue isn't empty, then it sets its new slot reservation attempt probability to $P_{reservation}=(1/2)^{bn}$, which is based on the current value of its previous backoff counter. The process starts over again until the message queue empties and the backoff counter is set to zero.

ABA backoff mechanism is based on two key ideas. First, a node with queued messages that just completed the transmission of a message continues its slot reservation based on the system knowledge accumulated in its backoff counter. If, on the other hand, we would let a node, that just completed its message transmission to set its slot reservation attempt probability to one (by setting its backoff counter to zero), then this would be unfair to other competing nodes. In such one hop environments, if all nodes but one have relatively high backoff counters, it is possible that the node with the smallest backoff counter will succeed to transmit a message and thereby will reset its backoff counter to zero. This node will eventually monopolize the channel as it keeps having the smallest backoff counter and will prevent other nodes from making a slot reservation.

Secondly, had we allowed a node to increase its slot reservation attempt probability, after a successful slot reservation by a one hop competitor node, we would only make the nodes more aggressive. Our simulation experiments showed that such policy only increases the percentage of the wasted slots due to collisions and that instead a non-persistent policy is much better.

In the following sections we present the results of our simulation study using the ABA backoff algorithm presented here. In section 4 we compare ABA with the backoff mechanism of slotted aloha for some selected network configurations.

3 EIGHT NODE SIMULATION STUDY

All simulation results presented in this section consider eight-node populations placed in fully and partially connected network topologies. Both unicast and broadcast transmission types are examined with average message length (AFL) of 2, 10 and 20 slots per message. We consider that the system operates within its stable region if for a given node population and a given average arrival rate, at the end of the simulation, the total number of unserviced messages[3] is less than 0.05% of the total number of generated messages. The total number of messages to be serviced is 10^6 (regardless of the average message length).

3.1 Fully connected network topology

In this network topology (Figure 3-1), all nodes are within transmission range of each other. This means that there is the maximum possible competition/interference between nodes, but also complete channel state information and load balance. It should be noted that since there is no spatial reuse in a fully connected network, broadcast and unicast transmission types have the same behavior in this topology and are not presented separately. The frame length L used for this network topology is $L=Min\{7^2+1, 8\}=8$.

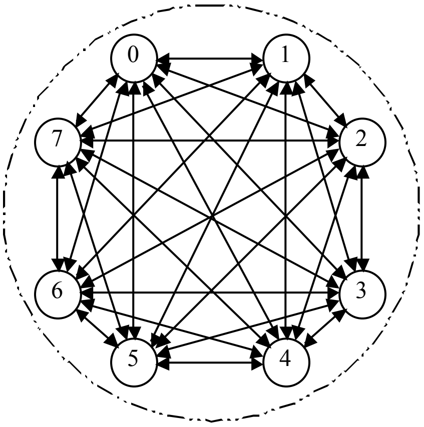

Figure 3-1: An 8-node, fully connected network topology

Figure 3-2 and Figure 3-3 show the average message delay and the average waiting time, respectively, versus offered load. Offered load axis is in logarithmic scale for display purposes, due to large variations in the supported message arrival rates per node caused by the different AFL values. Message delay, represents the time interval (in slots) between a message transmission request arrival and its complete delivery to the destination node. Waiting time represents the time interval (in slots) between a message transmission request arrival and the start of its transmission.

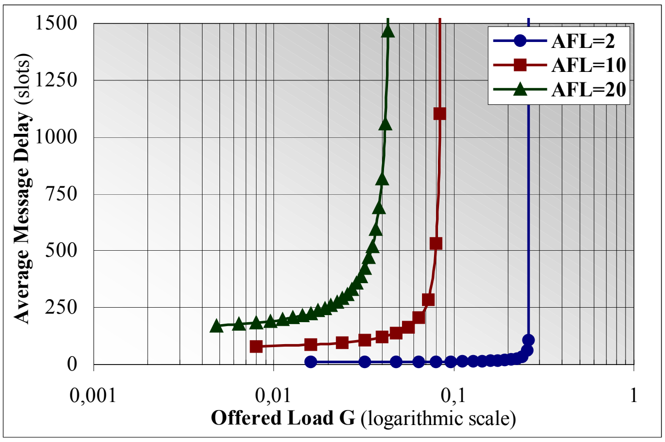

Figure 3-2: Average message delay for 8-node, fully connected network.

The difference between average message delay and average waiting time (for the same AFL) is the time it takes for a node to deliver the message to the destination node (message service time). For example if a message has AFL=20 with frame length L=8, a node will be able to completely deliver the message in approximately (AFL-1)*L+1 = (20-1)*8+1 = 153 slots[4].

---

[3] Unserviced messages are considered messages that their transmission has not yet started or is incomplete.

[4] It takes 19*8=152 slots to deliver the first 19 data packets and 1 slot to deliver the last packet.

For each AFL value, after a certain offered load G, both the average message delay and the average waiting time tend to infinity and the system becomes unstable. Table 3-1 shows the maximum offered load, average message delay and average waiting time values, for which the system is stable.

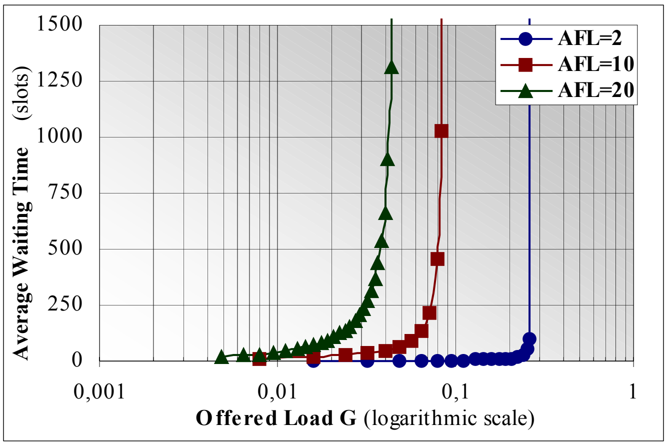

Figure 3-3: Average waiting time for 8-node, fully connected network.

Figure 3-4 shows the channel utilization of the system versus offered load. In Figure 3-4, offered load axis is in linear scale to better display the rising rates of the channel utilization. Channel utilization represents the percentage of used slots (slots in which data transmissions occurred) versus the total number of slots that the system needed to complete the 10^6 message deliveries to its nodes. The difference between total slots and used slots, represents the percentage of slots in which data transmissions did not occur due to collisions, low slot reservation attempt probabilities, or empty message queues.

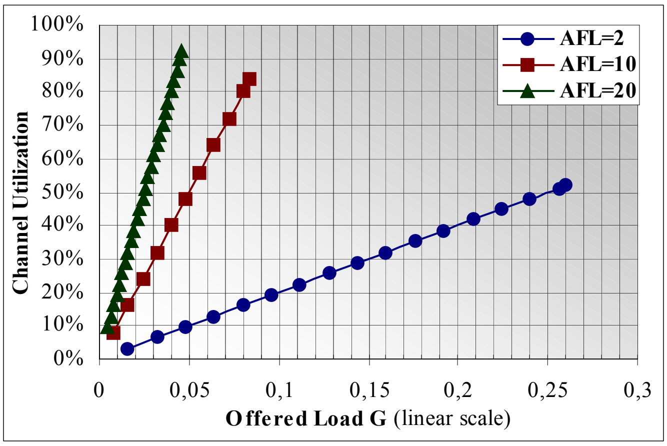

Figure 3-4: Channel utilization for 8-node, fully connected network.

As the AFL value increases, higher channel utilization is achieved more rapidly because nodes have to keep their reservations in more consecutive frames, in order to complete their message transmissions. They also cease the competition to other nodes for a longer time period. Theoretically, as the AFL value tends to infinity, channel utilization tends to 100% and nodes seem to keep their slot reservation forever. On the other hand, if messages are small (low AFL value), nodes keep their reservations for a shorter time period and spend more time in reservation competition with other nodes. Remember that we assume that a node can reserve at most one slot per frame. This means that nodes with messages of average size AFL=20 restart their slot reservation attempts on average every twenty frames while nodes with AFL=2 every two frames. This fact lowers channel utilization by increasing the collision and idle slots, thereby, reducing the slots used for transmission. Table 3-1 also shows the maximum utilization values, for a stable system.

AFL	2	10	20
Offered Load	0.260	0.08	0.046
Average Message Delay (slots)	106.7	531.4	7512.1
Average Waiting Time (slots)	97.7	458.3	7358.8
Channel Utilization	51.997%	80.025%	92.199%

Table 3-1: Maximum metric values for 8-node, fully connected network

3.2 Two-area network topology

In Figure 3-5, nodes are divided into two fully connected sub-areas. More specifically nodes 0, 1, 2, 3 and nodes 4, 5, 6, 7 are within transmission range of each other, respectively. Nodes 3 and 4 are also within transmission range of each other and provide a link between the two fully connected sub-networks. The frame length L used for this network topology is $L=\text{Min}\{4^2+1, 8\}=8$.

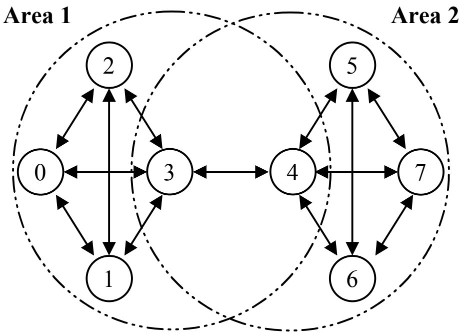

Figure 3-5: An 8-node, two-area network topology

3.2.1 Unicast transmission requests

Spatial reuse of the communication channel is possible in the above network configuration for unicast transmission requests. Any unicast transmission within area 1 is permitted at the same time with a unicast transmission within area 2. The only restriction is that unicast transmissions within area 1 and area 2 are not possible, if node 3 is a receiver and node 4 is a transmitter (or vice-versa) at the same time.

Figure 3-6 shows the average waiting time versus offered load. As explained in the previous section, the difference between the average message delay and average waiting time values, are due to message delivery delays that can be calculated from the network parameters (Number of nodes, AFL, Frame length etc) and thus average message delay graphs will not been shown any further.

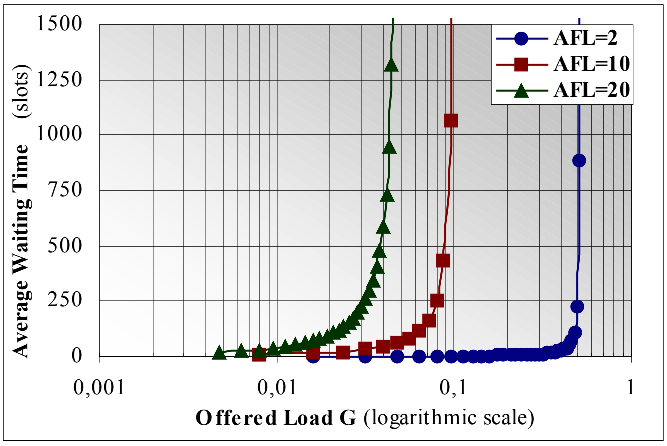

Figure 3-6: Unicast average waiting time for 8-node, two-area network.

Comparing average waiting time with that for the fully connected network (Figure 3-3), we see that for the same offered load, the average waiting time is decreased (see Table 3-2) and that the maximum supported offered load is increased (see Table 3-3), at the two-area network case. This was expected due to the spatial channel reuse in the two-area network case. Great improvement in average waiting time is observed for small messages. As we mentioned earlier, nodes with small messages keep their reservations for a shorter time period and increase the competition with other nodes. This fact holds in this case also, but now the number of the competing nodes has been reduced by at least 37.5% (3/8 of nodes) while at the same time the number of available slots in the channel frame remains the same (L=8). For example, assume that node 3 wants to transmit to node 4 and node 4 wants to transmit to node 3 and that they have already reserved two different slots for their transmissions[5]. For the remaining six slots of the frame, only three nodes have to compete with each other because transmissions within area 1 do not interfere with transmissions within area 2 and can reserve the same slot.

Figure 3-7 shows the channel utilization of the system versus offered load. Because channel reuse is allowed in this case, in many slots more than one data packet transmissions take place causing other slots to become unused. This is the reason that channel utilization seems to "drop" compared to the fully connected case as shown in Table 3-2.

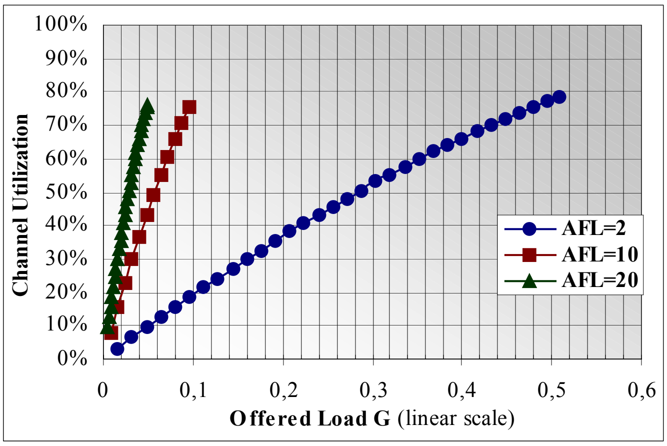

Figure 3-7: Unicast channel utilization for 8-node, two-area connected network.

For this reason, a better metric to compare these network topologies is the system throughput (Figure 3-8). System throughput represents the percentage of data packet transmissions per total slots. In the fully connected case, since only one data packet transmission per slot is allowed, system throughput coincides with channel utilization.

As expected, for the same offered load and AFL values, system throughput is the same in both network topologies. As explained before, the difference between system throughput and channel utilization is attributed to slots that are used to transmit more than one data packets. For this network topology this percentage is at best 37.5% (3/8) while it could be as high as 60% (3/5) if the frame length used, was equal to five slots. The problem is that nodes cannot schedule their intended slot reservations based on others node's reservations. Instead they reserve the first available slot in which they do not experience a collision. For example in this network topology, it is possible that node 1 will reserve

---

[5] Node 3 transmitting to node 4 and vice-versa is a special case for this topology because they do not allow spatial channel reuse.

the first slot for its transmission and node 7 the second slot, while both nodes could instead use either slot to simultaneously make their transmissions without interfering with each other.

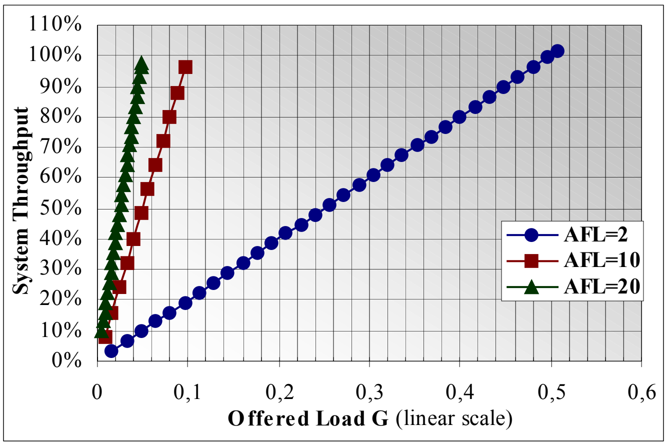

Figure 3-8: Unicast system throughput for 8-node, two-area connected network.

	Fully connected network			Two-area network		
AFL	2	10	20	2	10	20
Offered Load	0.256	0.0720	0.0416	0.256	0.0720	0.0416
Average Message Delay (slots)	61.3	284.8	1057.2	17.4	236.2	882.6
Average Waiting Time (slots)	52.3	211.8	904.0	8.4	163.2	729.7
Channel Utilization	51.2%	72.0%	83.2%	45.8%	60.7%	68.0%
System Throughput	51.2%	72.0%	83.2%	51.3%	72.0%	83.2%

Table 3-2: Unicast comparison between fully connected and two-area network.

AFL	2	10	20
Offered Load	0.508	0.1	0.0488
Average Message Delay (slots)	890.3	2867.8	6637.8
Average Waiting Time (slots)	881.3	2794.9	6484.9
Channel Utilization	78.240%	77.544%	76.311%
System Throughput	101.524%	99.839%	97.513%

Table 3-3: Unicast maximum metric values for 8-node, two-area connected network.

Finally, Figure 3-9 shows the coefficient of variation (***cf***) versus offered load multiplied by message size. The cf value measures the variation of message waiting time around its mean, its definition and means for estimating it are given in Figure 3-10. In Figure 3-9, the horizontal axis typically represents the total workload of the system, which depends of the message size.

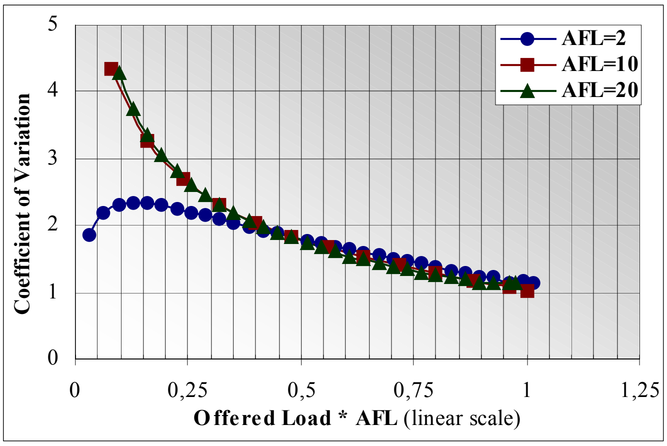

Figure 3-9: Unicast coefficient of variation for 8-node, two-area connected network.

(1) $cf = \dfrac{\sqrt{\mathrm{Var}(X)}}{\mathrm{E}(X)}$

(2) $\mathrm{Var}(X) \approx \dfrac{1}{n-1}\left[\sum_{i=1}^{n} X_i^{2} - \dfrac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^{2}\right]$

(3) $\mathrm{E}(X) \approx \dfrac{1}{n}\sum_{i=1}^{n} X_i$

where $X_i = i^{th}$ Transm. Message Waiting Time and

n = Number of Transm. Messages $= 10^{6}$

Figure 3-10: Coefficient of variation equations.

For small workload values, large messages (large AFL values) can experience four times bigger waiting times than the corresponding average waiting time. This is caused by the Geometric distribution nature of the message size that we assumed in section 2.1. Although large messages might have for example, size of 20 packets per message on average, in fact some messages might have as many as 38 packets and others as few as 2 packets per message[6]. For small offered loads where collisions are rare, the service time of a predecessor message is the main reason for the waiting time of queued messages. As the offered load increases, collisions are more often and the variation of message waiting time starts to drop while at the same time it starts to become independent of its size. Finally for large offered loads, behavior of message waiting time becomes more predictable and the coefficient of

---

[6] There is a larger variation in the geometrically distributed message sizes as the AFL value increases.

variation converges to values around one regardless of the message size.

3.2.2 Broadcast transmission requests

Spatial reuse of the communication channel is possible in the two-area network configuration for broadcast transmission requests. Any broadcast transmission within area 1 is permitted at the same time with a broadcast transmission within area 2 as long as nodes 3 and 4 are not the transmitting nodes.

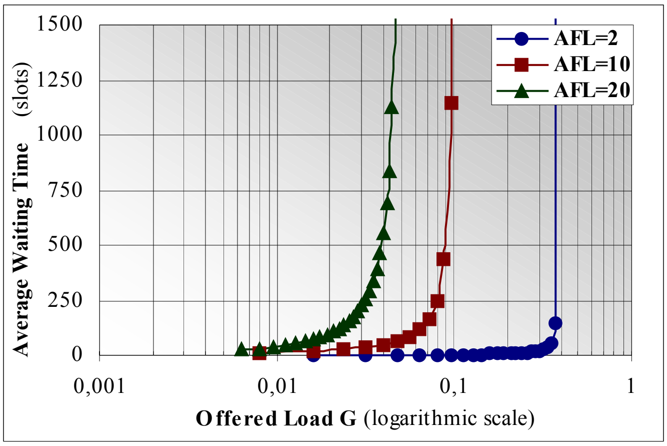

Figure 3-11: Broadcast average waiting time for 8-node, two-area network.

Once again average waiting time is decreased (see Figure 3-11) and the maximum supported offered load is increased (see Table 3-1 and Table 3-5) compared with that for the corresponding fully connected network case, due to the spatial channel reuse in the two-area network case.

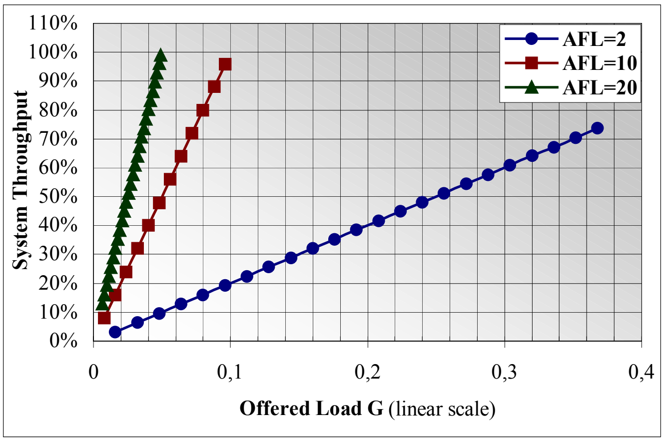

Figure 3-12: Broadcast system throughput for 8-node, two-area connected network.

An interesting phenomenon for this network topology is the difference between unicast transmissions and broadcast transmissions (see Table 3-3 and Table 3-5). For small messages, the maximum supported offered load is much greater for unicast transmissions than for broadcast transmissions and as the message size increases both transmission types support almost[7] the same maximum offered load. The nodes 3 and 4 that connect the two areas are the reason for this difference. Broadcast transmission in a particular slot from node 3 or 4 requires that the remaining seven nodes in the network are able to listen (i.e., that they do not participate in any message transaction) in that slot or else a collision will occur. On the other hand unicast transmission from node 3 or 4 to any destination does not require all other nodes to listen. It only requires the one-hop neighbors of the destination node including the destination node itself to be idle (which in the worst case corresponds to four nodes in this case). Thus in broadcast transmissions, nodes 3 and 4 have almost twice as many competing nodes than in unicast transmissions.

	Fully connected network			Two-area network		
AFL	2	10	20	2	10	20
Offered Load	0.256	0.0720	0.0416	0.256	0.0720	0.0416
Average Message Delay (slots)	61.3	284.8	1057.2	20.8	238.8	839.5
Average Waiting Time (slots)	52.3	211.8	904.0	11.8	165.9	686.5
Channel Utilization	51.2%	72.0%	83.2%	46.6%	62.9%	70.8%
System Throughput	51.2%	72.0%	83.2%	51.2%	72.0%	83.2%

Table 3-4: Broadcast comparison between fully connected and two-area network.

AFL	2	10	20
Offered Load	0.368	0.096	0.0488
Average Message Delay (slots)	156.5	1215.9	6817.3
Average Waiting Time (slots)	147.5	1143.0	6664.4
Channel Utilization	62.627%	78.587%	80.576%
System Throughput	73.721%	95.893%	99.018%

Table 3-5: Broadcast maximum metric values for 8-node, two-area connected network

As the message length increases, nodes keep their slot reservation for a longer time period and a slot reservation pattern tends to be established. Theoretically

---

[7] In fact as the AFL value increases the maximum supported offered load is a little bit higher in broadcast transmissions than in unicast transmissions. The reason, will be explained in section 3.3.2

for very long messages, after the first collisions have been resolved, every node tends to reserve the same slot for many consecutive channel frames for its transmissions. This dramatically reduces competition between nodes and the above-described phenomenon has a negligible performance impact.

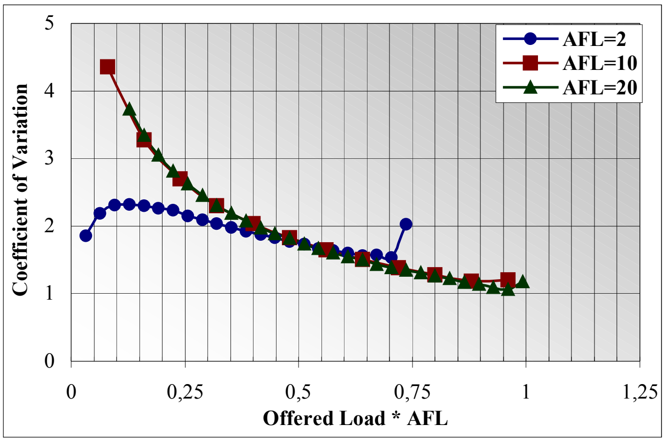

Figure 3-13: Broadcast coefficient of variation for 8-node, two-area connected network.

Finally, Figure 3-13 shows the coefficient of variation versus offered load multiplied by AFL. Once again, for small offered loads, large messages (large AFL values) may experience four times bigger waiting times than the corresponding average waiting time due to the Geometric distribution nature of the message size. As the offered load increases, the variation of the waiting time of a message starts to drop while at the same time it becomes independent of its size. For small messages, after a certain offered load, the cf starts to diverge in contrast to a convergence observed for large messages. This is caused by the peculiarity of the broadcast transmission that was previously explained.

3.3 Eight-area network topology

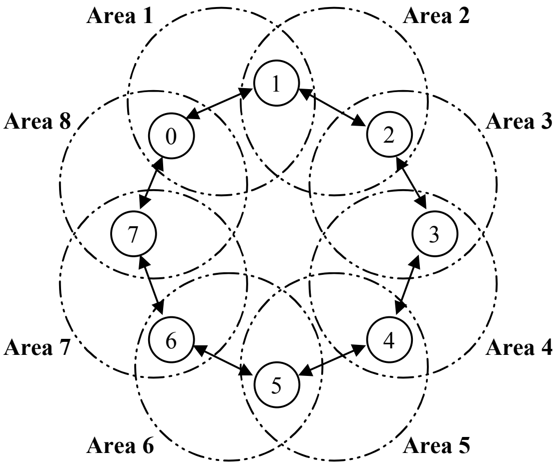

Figure 3-14: An 8-node, eight-area network topology

In Figure 3-14 all nodes are connected in a sequential manner. Each node is within transmission range of two other nodes and the maximum node degree is d=2. Thus the frame length, L, used for this network topology is $L=d^2+1=5$. This topology is selected to demonstrate the effect of smaller frame length compared to the previously examined topologies.

3.3.1 Unicast transmission requests

Spatial reuse of the communication channel is possible in the above network configuration for unicast transmission requests. Any node can transmit in the same slot with another node, if either one of these nodes does not have the other or a common neighbor node as its destination.

Figure 3-15 shows the average waiting time versus offered load. Average waiting time, for the same offered load, has been significantly reduced (see Table 3-7). This is caused not only by the higher spatial reuse but also by the smaller frame size compared to the previous network topologies.

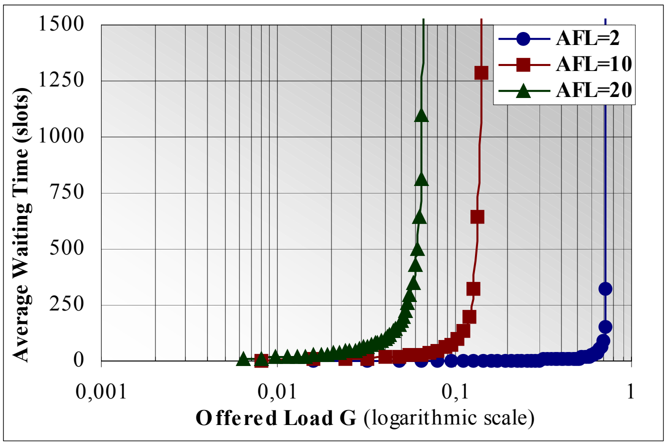

Figure 3-15: Unicast average waiting time for 8-node, eight-area network.

One of the parameters that increase the average waiting time of a message is the service time of its predecessor message. Consider for example the case that two messages arrive at a node very closely in same time and one of them starts its delivery to a destination node. At best, the waiting time of the second message (no delays due to collisions etc.) is the service time of the first message. In section 3.1 we explained that for the fully connected network (Frame length L=8) if a message has AFL=20, a node will be able to completely deliver the message in approximately (20-1)*8+1=153 slots. In this network topology (Frame length L=5) if a message has AFL=20, a node will be able to completely deliver the message in approximately (20-1)*5+1=96 slots. This means that we have a 37% reduction in message service time and therefore a reduction in the average waiting time of the remaining queued messages.

Figure 3-16 shows the system throughput versus offered load. Higher throughput than previous area-networks is achieved due to higher spatial reuse of the communication channel but also due to smaller frame length.

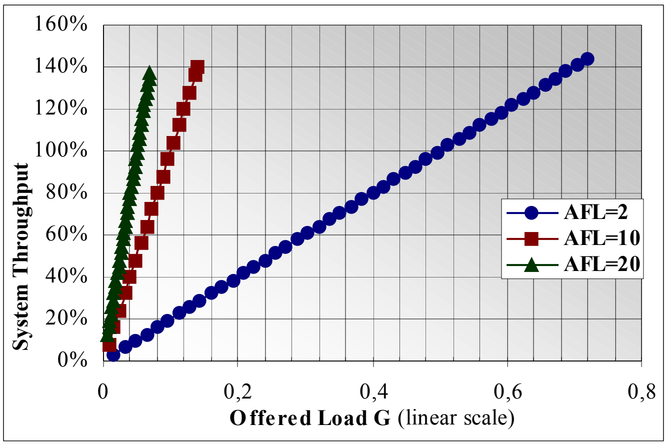

Figure 3-16: Unicast system throughput for 8-node, eight-area connected network.

AFL	2	10	20
Offered Load	0.720	0.14	0.0688
Average Message Delay (slots)	329.6	1332.6	3004.2
Average Waiting Time (slots)	323.6	1286.7	2908.2
Channel Utilization	89.606%	88.796%	88.075%
System Throughput	143.823%	139.853%	137.538%

Table 3-6: Unicast maximum metric values for 8-node, eight-area connected network.

Table 3-6 shows the maximum offered load and the maximum channel slot utilization and throughput supported by this network. As explained in section 3.2.1, the difference between system throughput and channel utilization is attributed to slots that are used to transmit more than one data packets which in this case is more than 50% for all AFL values. This means that we would need at least 50% more slots to deliver all messages to their destination nodes, if spatial reuse was not allowed and only one data packet was transmitted per slot.

	Fully connected network			Eight-area network		
AFL	2	10	20	2	10	20
Offered Load	0.256	0.0720	0.0416	0.256	0.0720	0.0416
Average Message Delay	61.3	284.8	1057.2	9.6	84.0	203.2
Average Waiting Time	52.3	211.8	904.0	3.6	38.0	107.3
Channel Utilization	51.2%	72.0%	83.2%	44.5%	58.5%	64.9%
System Throughput	51.2%	72.0%	83.2%	51.2%	72.0%	83.1%

Table 3-7: Unicast comparison between fully connected and eight-area network.

Finally, Figure 3-17 shows the coefficient of variation versus offered load multiplied by AFL.

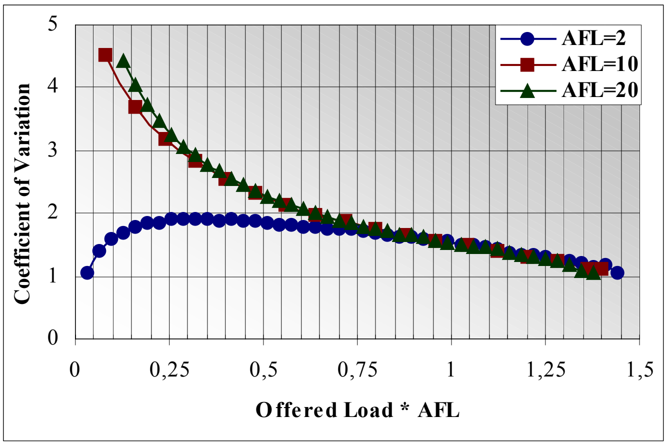

Figure 3-17: Unicast coefficient of variation for 8-node, eight-area connected network.

3.3.2 Broadcast transmission requests

Spatial reuse of the communication channel is possible in the eight-area network configuration for broadcast transmission requests. Any node can transmit in the same slot with another node, if they are at least two-hops away.

As shown in Figure 3-18, higher spatial reuse and smaller frame size has once again reduced significantly the average waiting time (see Table 3-9) and increased the maximum supported offered load (see Table 3-8). System throughput (see Figure 3-19) is also increased significantly compared to the fully connected and the two-area networks.

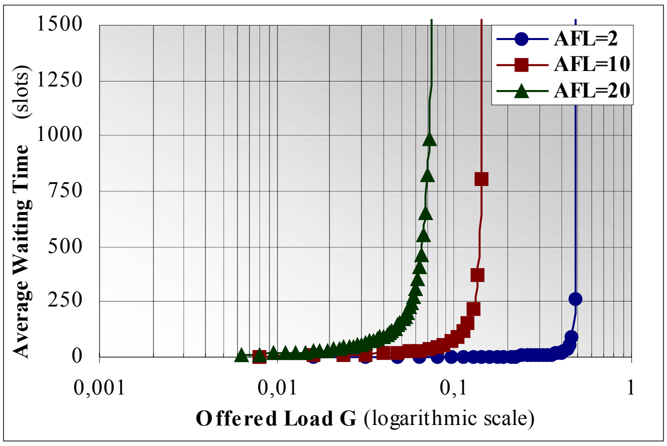

Figure 3-18: Broadcast average waiting time for 8-node, eight-area network.

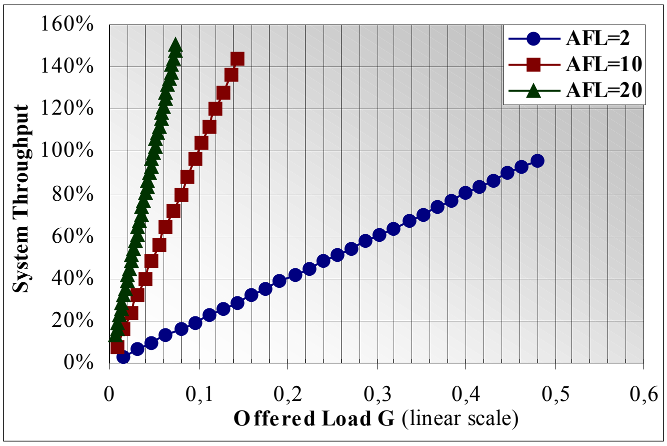

Figure 3-19: Broadcast system throughput for 8-node, eight-area connected network.

A difference in the maximum supported offered load between unicast and broadcast transmissions (see Table 3-6 and Table 3-8) is observed in this topology, like in the two-area network case. For small messages, the maximum supported offered load is much greater for unicast transmissions than for broadcast transmissions. This time, all nodes are responsible for this difference. Every node that wants to make a unicast transmission requires that only one node, its destination, is able to listen (i.e. it does not participate in any message transaction). On the other hand, every node that wants to make a broadcast transmission requires twice as many destination nodes to be able to listen, its left and right neighbors.

AFL	2	10	20
Offered Load	0.48	0.144	0.0752
Average Message Delay (slots)	272.5	852.4	2315.7
Average Waiting Time (slots)	266.5	806.5	2219.6
Channel Utilization	71.84%	92.924%	94.909%
System Throughput	96.088%	143.664%	150.287%

Table 3-8: Broadcast maximum metric values for 8-node, eight-area connected network

Unlike the two-area network, in this topology, this phenomenon not only it does not become negligible as the AFL increases, but it starts to have the opposite effect. For large AFL values the maximum offered load in broadcast transmissions is greater than in unicast transmissions[8]. Remember that as the AFL value increases, slots are reserved in more consecutive frames and that in each reserved slot the destination node informs its neighbors of the on-going transmission by sending an SR packet[9] during CMS1. For this network topology, twice as many nodes are informed of an on-going broadcast transmission than of an on-going unicast transmission. This fact has a little impact for small messages due to short slot reservation time periods, but a great impact for large messages because twice as many nodes are prevented from experiencing a collision in a broadcast transmission for a longer slot reservation time period.

	Fully connected network			Eight-area network		
AFL	2	10	20	2	10	20
Offered Load	0.256	0.0720	0.0416	0.256	0.0720	0.0416
Average Message Delay	61.3	284.8	1057.2	11.5	83.3	199.3
Average Waiting Time	52.3	211.8	904.0	5.5	37.3	103.1
Channel Utilization	51.2%	72.0%	83.2%	45.3%	60.2%	67.4%
System Throughput	51.2%	72.0%	83.2%	51.1%	71.9%	83.4%

Table 3-9: Broadcast comparison between fully connected and eight-area network.

Finally, Figure 3-20 shows the coefficient of variation versus offered load multiplied by AFL. As in the two-area network case, due to the peculiarity of the broadcast transmission, for small messages, after a

---

[8] In the two-area network although this phenomenon existed, it had a little impact because only two nodes experienced it..

[9] The impact of the SR packet to neighbor nodes was explained in section 2.2

certain offered load, the cf starts diverge in contrast to the convergence observed for large messages.

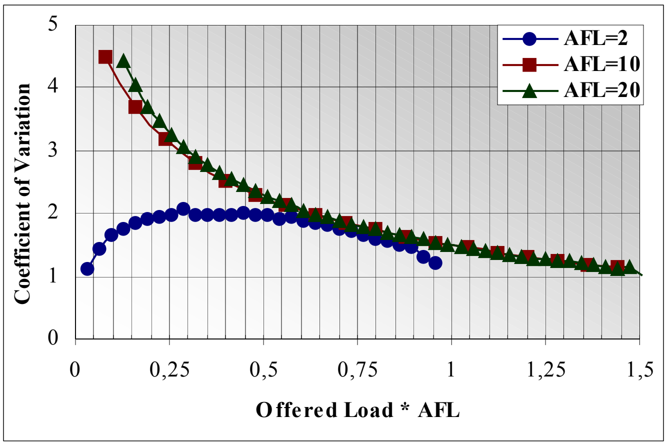

Figure 3-20: Broadcast coefficient of variation for 8-node, eight-area connected network.

4 BACKOFF ALGORITHM EVALUATION

In this section we compare the performance of CATA with the ABA backoff algorithm with that of CATA with the backoff mechanism of slotted aloha. All simulation results presented here, consider eight-node populations placed in fully connected network topology as shown in Figure 3-1 using slotted aloha as the collision resolution mechanism. The fully connected topology is examined due to its maximum competition/interference between nodes. Unicast and broadcast transmission types are not examined separately as they have the same behavior in a fully connected network. Once again we consider that the system operates within its stable region if for a given node population and a given average arrival rate, at the end of the simulation, the total number of unserviced messages[10] is less than 0.05% of the total number of generated messages. The total number of messages to be serviced is 10^6 with average message length (AFL) of 2, 10 and 20 slots per message.

4.1 Slotted aloha backoff algorithm

Slotted aloha backoff is a very simple and popular mechanism that works as follows:

- Every node has a backoff counter (bn) that sets to zero (bn=0) every time it wants to make a message transmission.
- When a new message arrives, the node sets its slot reservation attempt probability to one ($P_{reservation}=1$) and tries to make a slot reservation in the next available slot.
- If, and every time, a collision occurs during its slot reservation attempt, the node increases its backoff counter by one (bn=bn+1) and sets its slot reservation attempt probability to $P_{reservation}=(1/2)^{bn}$.
- When finally the node makes its slot reservation and completes its message transmission, it sets its backoff counter to zero (bn=0) and the process starts over again.

Figure 4-1 shows the average waiting time versus offered load.

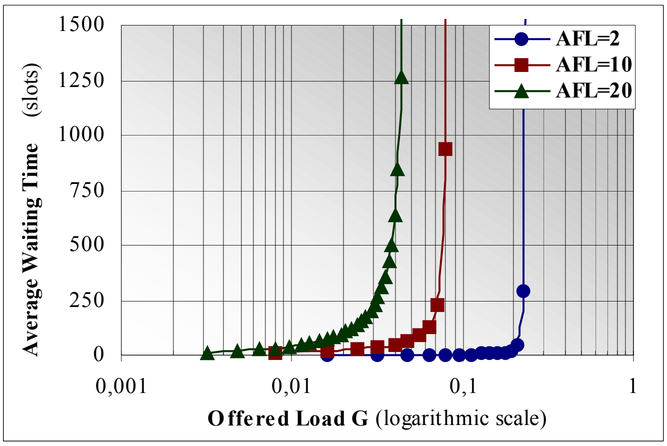

Figure 4-1: Slotted aloha average waiting time for 8-node, fully connected network.

For each AFL value, after a certain offered load G, both the average message delay and the average waiting time tend to infinity and the system becomes unstable. Table 4-2 shows the maximum offered load, average message delay and average waiting time values, for which the system is stable. Table 4-1 shows a comparison between slotted aloha and ABA for various AFL values. For AFL=2 (small messages), average waiting time is almost eleven times greater in slotted aloha than in ABA backoff algorithm. For AFL=10, average waiting time is almost two times greater in slotted aloha than in ABA backoff algorithm and for AFL=20, both backoff algorithms have almost the same message average waiting time. This was expected because as explained in section 3.2.1large messages tend to reduce collisions due to longer slot reservations and makes them almost independent to any collision resolution mechanism. Thus for small messages our backoff algorithm, compared to slotted aloha, significantly reduces the average message waiting time. This can be very important especially when messages

---

[10] Unserviced messages are considered messages that their transmission has not yet started or is incomplete.

(data) are delay sensitive as in audio and video applications.

CATA with backoff mechanism of:	**ABA**			**Slotted Aloha**		
AFL	2	10	20	2	10	20
Offered Load	0.224	0.08	0.0448	0.224	0.08	0.0448
Average Message Delay (slots)	26.0	531.4	2559.3	296.5	1009.9	2598.3
Average Waiting Time (slots)	16.9	458.3	2406.3	287.5	936.9	2445.2
Channel Utilization	44.9%	80.0%	89.6%	44.8%	80.0%	89.7%

Table 4-1: Comparison between ABA and slotted aloha for 8-node, fully connected network

Figure 4-2 shows the channel utilization of the system versus offered load

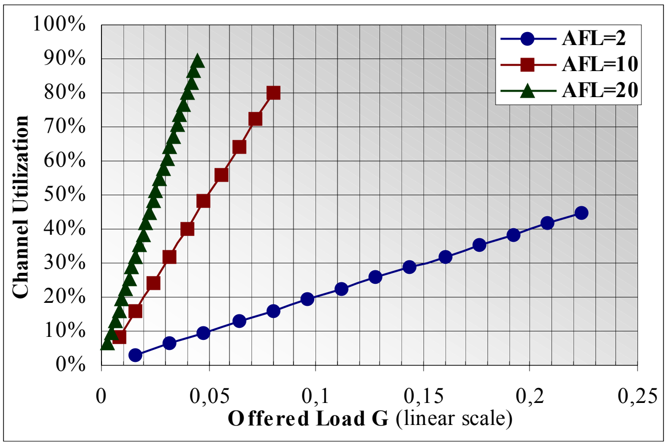

Figure 4-2: Slotted aloha channel utilization for 8-node, fully connected network.

AFL	2	10	20
Offered Load	0.224	0.08	0.0448
Average Message Delay (slots)	296.5	1009.9	2598.3
Average Waiting Time (slots)	287.5	936.9	2445.2
Channel Utilization	44.837%	79.977%	89.675%

Table 4-2: Slotted aloha maximum metric values for 8-node, fully connected network

5 CONCLUSIONS

In this work we focused on a special category of wireless networks, called Ad-hoc networks. The key ideas that make an ad-hoc network very attractive are its support of mobility, the very fast installation of a temporary network without the aid of a central base station and the fact that nodes can join freely.

We focused on the Media Access Control layer in which nodes forming an ad-hoc network, compete with each other to gain access to the medium and make their transmissions. We explained some of the problems that arise by this competition and that various MAC protocols attempt to solve. From these protocols, CATA distinguishes not only for its simplicity, but also for its ability to support real time applications and its explicit support of broadcast and unicast transmission requests.

In the following part of this work, we presented some network and topology issues that affect the performance of a MAC protocol and proposed a new backoff algorithm, named ABA, in order to lower the number of packet collisions and increase the channel utilization. We added this collision resolution mechanism to the CATA reservation mechanism and using an event driven simulation program we evaluated the performance of CATA. Our experimental results examined performance issues based on message delays, message waiting times, channel utilization and system throughput in both unicast and broadcast transmission requests, for various node populations and network topologies.

Finally, we compared the performance of CATA with our ABA backoff algorithm and with the backoff algorithm of Slotted Aloha. Based on our experimental results, ABA significantly improves the performance of CATA especially for small size messages and high arrival rates. As the message size increases, less performance improvement is observed because large messages tend to reduce collisions due to longer slot reservations and makes nodes almost independent to any collision resolution mechanism.

REFERENCES

[1] P. Karn,
"MACA – A new channel access method for packet radio,"
ARRL/CRRL Amateur Radio 9th Computer Networking Conference,
pp. 134-140, ARRL, 1990.

[2] V. Bharghavan, etc.,
"MACAW: A media access protocol for wireless LAN's,"
Proceedings of ACM SIGCOMM '94
pp. 212-225, ACM, 1994.

[3] J. Deng and Z. J. Haas,
"Dual busy tone multiple access (DBTMA): Performance evaluation,"
Proceedings of IEEE VTC '99
Houston, Texas, USA, May 16-20, 1999.

[4] C. L. Fullmer, J. J. Garcia-Luna-Aceves,
"Solutions to hidden terminal problems in wireless networks,"
Computer Communications Review,
vol. 27, no. 4, pp. 39-49, 1997.

[5] C. Zhu and M. S. Corson,
"A five phase reservation protocol (FPRP) for mobile ad-hoc networks,"
Proceedings of IEEE Infocom,
1998.

[6] I. Chlamtac and A. Farago,
"Making transmission schedules immune to topology changes in multi-hop packet radio networks,"
IEEE/ACM Trans. Networking,
vol. 6, pp. 23-29, 1994.

[7] Z. Tang and J. J. Garcia-Luna-Aceves,
"Hop-reservation multiple access (HRMA) for ad-hoc networks,"
Proceedings IEEE Infocom,
New York, March 21-25, 1999.

[8] Z. Tang and J. J. Garcia-Luna-Aceves,
"A protocol for topology-dependent transmission scheduling,"
Proceedings IEEE Wireless Communications and Networking Conference,
New Orleans, Louisiana, September 21-24, 1999.

[9] IEEE,
"IEEE P802.11 draft standard for wireless LAN: Medium access control (MAC) and physical layer (PHY) specifications,"
IEEE, July 1996.

[10] Harshal S. Chhaya and Sanjay Gupta,
"Performance modeling of asynchronous data transfer methods of IEEE 802.11 MAC protocol,"
Wireless Networks,
vol. 3, pp. 217-234, 1997.

[11] M. Zaki, D. Makrakis and J. R. Gallardo,
"Performance analysis of IEEE 802.11-based wireless LANs under realistic traffic loading,"
Proceedings of 11th International Conference on Wireless Communications,
Calgary, Alberta, pp. 214-222, July 1999.

[12] F. Cali, M. Conti, and E. Gregori,
"IEEE 802.11 wireless LAN: Capacity analysis and protocol enhancement,"
in INFOCOM '98, Conference on Computer Communications.
Institute of Electrical and Electronics Engineers, 1998, pp. 142-149.