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

M.Sc. Thesis

JOHN ATZARAKIS

A thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE
Department of Electronics & Computer Engineering

EXAMINING COMMITTEE

PROF. M. PATERAKIS
PROF. V. DIGALAKIS
PROF. A. MARAS

Chania February 2003
Technical University of Crete
Department of Electronics & Computer Engineering

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. CATA allows nodes within a two hops range to contend for and reserve time slots by means of a distributed reservation and handshake mechanism. CATA ensures that no collisions occur in successfully reserved time slots, even when hidden terminals exist.

Using packet-level simulations we examine various performance and design issues. Data messages arrivals are assumed to occur according to a *Poisson* process and vary in length according to a *Geometric* distribution. 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.

ACKNOLEDGMENTS

After this work has finally reached to its end, I feel the obligation to express my utmost gratitude to a series of people that with their own unique way helped me with everything I have accomplished so far.

Many thanks to:

My parents Giorgos and Panagiota and my sister Tasia. It is very important for anyone to have the unlimited patience, support, inspiration and guidance of his family, in every step of his life. I will always be grateful to them.

Professor Michael Paterakis, supervisor of this thesis, for his constant guidance and support. I consider myself very lucky that I had the opportunity to work with not only such a brilliant scientist but also with such an honest and sincere man. I hope to be lucky in the future too and have the opportunity to work with people just like him.

Professors Vasilis Digalakis and Andreas Maras for participating in my examine committee. Mr. Maras I enjoyed very match chattering with you.

My friends and fellow-students Aris Meletiou and Chrissa Tsinaraki for helping me in various aspects of this work. Their help was very valuable to me.

My best friends and fellow-students Michael Balantinos, Georgia Kyriakaki, Nikos Papas, Thodoris Margazas and George Terezakis. During my ten-year staying in Chania they were always beside me, in good and bad times, and for that I thank them a lot. We had a lot of fun together during these years and I hope that we will keep on having fun in the years to come.

My co-workers in Network Operations Center of Technical University of Crete

All my non-academic friends that helped and supported me with their own way. I feel privileged of knowing them.

John Atzarakis
Chania February

DEDICATION

The things you taught me I will always know.
How could I not? The roots have sunk so deep:
All lessons of the heart that I will keep
No matter who I am or where I go.
Kids learn from what their parents are, and so
You are my book of life, the thoughts I reap;
Only in your arms I quiet sleep;
Under my words your voice sings soft and slow.
From you I learned the rules of right and wrong
Against which I at times had to rebel,
Though with regret I carry with me still.
How lucky I am to have been loved so well,
Even as I pushed against your will,
Relying on a father fair and strong.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 AD-HOC NETWORKS

In recent years, a wide variety of mobile computing devices have emerged, including portables, palmtops and personal digital assistants (PDAs). While the first portables were designed as stand-alone machines, many of these new devices are intended to work as full network citizens. Consequently, a new generation of wireless network technology is needed to provide adequate network connectivity for these mobile users. A special category of wireless networks is that of **ad-hoc** networks (also referred to as MANETs, for Mobile Ad-hoc NETworks).

An ad-hoc network is a collection of wireless mobile nodes (stations or packet radios) forming a temporary network without the aid of any centralized administration or standard support services regularly available on a wide area network. Nodes communicate with each other either directly or, due to the limited propagation range of each mobile node's wireless transmissions, through intermediate nodes[1] without relying on any preexisting network infrastructure. Ad-hoc networks are mainly intended for situations in which it cannot, or does not make sense, to install a fixed network. Soldiers in a battlefield exchanging tactical information, rescue teams coordinating themselves in a disaster situation, company members sharing information in a meeting and students using laptops to participate in an interactive class, are examples where ad-hoc networks can be necessary.

[1] This is called multihop propagation. Each node that forwards a packet from one node to another is called a "hop".

The advantages of ad-hoc networks compared to fixed networks are:

- *Fast installation:* Ad-hoc networks can be installed quickly in places without previous infrastructure.
- *Fail tolerance:* The malfunctioning or disconnection of a node can easily be encountered with the dynamic reconfiguration of the network. In a fixed network however, if a failure in a router occurs, the traffic redirection is a complex operation.
- *Connectivity:* If two nodes are within range, a communication channel can be established between them. In a fixed network, even if two nodes are side by side, it is necessary that these nodes have a guided mean to communicate with each other.
- *Mobility:* In contrast to fixed networks, in ad-hoc networks, nodes can be mobile, therefore their location can change with time.

The disadvantages of ad-hoc networks compared to fixed networks are:

- *Communication speed:* Wireless communication channels usually have lower speeds compared to wire communication channels. In a wireless network, communication speeds are limited to a few Mbps while in a fixed network communication speeds can be up to a few Gbps.
- *Channel errors:* Usually, errors occurring in a wireless link are in the range of one erroneous bit to every 10^5 or 10^6 transmitted bits, while in a fiber optic link the corresponding range is one erroneous bit to every 10^{12} or even 10^{15} transmitted bits.
- *Location:* In a fixed network the physical location of a node can easily be found from its address. In an ad-hoc network there is no geographical information and the address of a node does not necessarily relates to its position.
- *Routing:* In a fixed network the topology hardly changes. Nodes normally have the same positions and routing paths from one node to another can be known a priori. In an ad-hoc network, due to the non-deterministic movement of its nodes, a routing path from one node to another cannot always be easily obtained.

The self-configuring, dynamic connectivity, multihop-propagation and fully distributed nature of ad-hoc networks, makes them very attractive for many new applications. However, the above attributes also introduce difficult problems at the link and network layer. In this work we focus on the medium access control layer (MAC) for ad-hoc networks, where nodes compete with each other to gain access to the medium and transfer their data to other nodes.

The remainder of this work is organized as follows. In Chapter 2 we give a brief description of various MAC protocols that have being developed for ad-hoc networks. Emphasis is given in CATA protocol in which this work is based on. Network and channel model and backoff algorithm issues are examined in Chapter 3. Based on these issues, in Chapter 4 and Chapter 5 we study the behavior and performance of CATA for eight and sixteen node populations respectively for various network topologies. The impact in performance of the backoff algorithm is examined in Chapter 6 and finally in Chapter 7 we present our conclusions and some ideas for future work.

CHAPTER 2

MAC PROTOCOLS FOR AD-HOC NETWORKS

2.1 MAC PROTOCOL CONCEPTS

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:

1. Channel-sensing based schemes (CSMA)
2. 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:

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

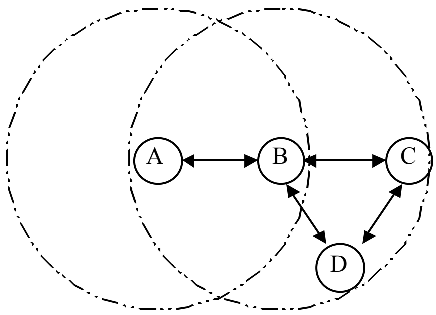

Figure 2-1: A 4-node interconnection scheme.

For example, in Figure 2-1 node A is within range of node B, nodes B, C and D are within range of each other, but nodes A, C and A, D are not within range of each other. If node B is transmitting to node A and node C tries to transmit to node D, then node C is an exposed terminal because it is out of the range of the receiver (node A) and within the range of the transmitter (node B). If node A is transmitting to node B and node C tries to transmit to node B, then node C is a hidden terminal because it is out of the range of the transmitter (node A) and within the range of the receiver (node B).

Although a possible access to the channel by an exposed terminal will not destroy the data packets being received by a node during an on-going transmission, such an access is prevented when CSMA is used. On the other hand, without proper notification, hidden terminals cannot have information about the on-going transmissions and a possible access to the channel by them during a transmission will destroy the data packets being received by a receiving node.

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.

2.2 BRIEF DESCRIPTION OF MAC PROTOCOLS

In the following sections we give a brief description of various MAC protocols that have been developed for wireless ad-hoc networks. The final section of this chapter gives a detailed description of the CATA protocol in which the work in this thesis is based on and a small comparison between MAC protocols that will be presented here.

2.2.1 MACA

MACA uses two types of short, fixed-size control packets. When a node A (Figure 2-1) wishes to transmit to a node B, it sends a request to send (RTS) packet to node B. The RTS packet contains the node that is addressed to (B) and the length of the data to be transmitted. When node B hears the RTS packet, it replies with a clear to send (CTS) packet also containing the node that is addressed to (A) and the length of the data to be transmitted. When node A receives the CTS packet from node B it immediately sends its data. All other nodes detecting the CTS packet from node B will avoid colliding with the returning data transmission by appropriately rescheduling their intended transmissions (taking into account the length of the transmitted data by node A). All nodes detecting the RTS packet from node A will avoid colliding with the returning CTS packet by rescheduling their intended transmissions so that node A can receive a correct CTS from node B. Nodes that hear an RTS but not a CTS because they are within range of the transmitter but out of range of the receiver can commence transmission after the CTS has been sent, without harm since they can infer that they are not within range of the receiver and they cannot collide with the data transmission.

In MACA if a node does not hear a CTS as a response to his RTS, it assumes that a collision has occurred and reschedules its packet transmission. The retransmission time selection is based on a binary exponential backoff algorithm (BEB). The algorithm works as follows: Each node has a backoff counter (BO) that is set to BO_{min} = 2 each time it has a correct transmission. If a node detects a collision it sets its BO to *min [2BO, BO_{max}]*, where BO_{max} = 64, and reschedules its next transmission attempt after BO slots.

2.2.2 MACAW

MACAW is based on MACA mechanism for making the transmission reservations. Some modifications have been made on MACA however, to increase the network utilization and provide a more fair access to the medium. One of the undesirable features of MACA is that it produces rather large variations on the backoff counter in each node (MACAW [2]). To prevent such large variations, MACAW adopts a gentler backoff algorithm named multiplicative increase and linear decrease algorithm (MILD). According to MILD, a node upon a successful transmission decreases its BO to *max [BO-1, BO_{min}]* and upon a collision increases its BO to *min [1.5BO, BO_{max}]*. This backoff algorithm still provides reasonably quick escalation in the backoffs when contention is high but by not resetting BO to BO_{min} after a successful transmission it avoids having to repeat the escalation in backoff counters after every successful transmission.

MACAW makes another modification to MACA by including an acknowledgment packet (ACK) to the RTS-CTS-DATA exchange scheme to help error or collision recovery and a data send (DS) packet to eliminate the exposed terminal problem. The ACK packet is send to the transmitter upon successfully reception of the data. If the sender does not receive the ACK, then the data packet is scheduled for retransmission. If during the RTS for the scheduled retransmission the data packet were correctly received but the ACK packet was not, then the receiver sends the associated ACK packet instead of an CTS. The sender increases its BO if, after sending the RTS, no CTS or ACK arrives and he decreases its BO when the ACK arrives. The sender's BO is not changed if there is a successful RTS-CTS exchange but the ACK packet does not arrive.

In MACA the exposed terminal is free to transmit because even though it is within the range of the sender, it is out of the range of the receiver and will not collide with the data received by the receiver. However the exposed terminal can benefit from its transmission attempt only if it can hear the corresponding CTS packet (which is not always true), otherwise the only result will be the increase of its BO. In addition, in MACAW it is possible that an exposed terminal that did not hear a CTS and started a transmission could cause a collision to the ACK packet of the corresponding CTS. For the above reasons, in MACAW a node before sending a data packet, it sends a DS packet. Every node hearing the DS knows that a successful RTS-CTS exchange has been made and data transmission will occur, so they reschedule their transmissions after the end of the ACK packet.

2.2.3 DBTMA

DBTMA uses the RTS-CTS scheme to establish communication between two nodes. It also uses two narrow-band busy tones, transmit busy tone (BT_t) and receive busy tone (BT_r), to notify neighbor nodes of the on-going transmission. Since the busy tones occupy narrow bandwidth, their bandwidth consumption is very small and considered negligible. If a node A wants to transmit its data packet to a node B, it first senses the BT_r signal. If no signal is sensed it transmits an RTS to node B. Upon reception of the RTS from node A, node B senses the BT_t signal. If no signal is sensed, node B raises its BT_r signal and replies with a CTS packet to node A. After node A receives the CTS packet, it sets up its BT_t signal and transmits its data packet to node B. Both BT_t and BT_r signals will be reset after the transmission of the data packet is completed.

Because busy tones are raised during an on-going transmission, neighbor nodes can monitor the channel continuously. Even if the CTS packet transmitted was not heard correctly by a node (hidden terminal problem), the sensing of the BT_r prevents it from accessing the channel. In the case that the RTS packet transmitted was not heard correctly (exposed terminal problem), a node sensing the BT_t signal could decide that it can transmit but not receive.

2.2.4 FPRP

The FPRP is a contention-based broadcast scheduling protocol, which uses a five-phase reservation process to establish TDMA slot assignments that are non-conflicting with high probability. Contention is limited among nodes within two-hop of one another, which provides a very efficient spatial reuse of the available bandwidth.

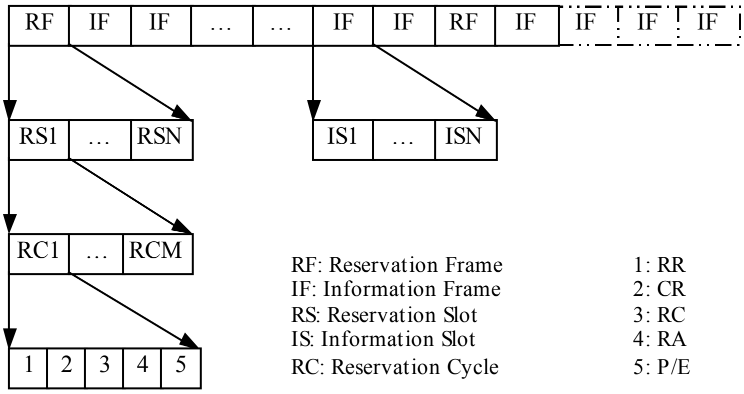

Figure 2-2: Slot and frame structure of FPRP

As shown in Figure 2-2, in the protocol's frame structure there is a reservation frame (RF) followed by a sequence of information frames (IF). There are N information slots[2] (IS) in an IF and N corresponding reservation slots (RS) in each RF. A TDMA schedule is generated in the RF and is used in each of the subsequent IF frames until the next RF, where the schedule is regenerated. An RS is composed of M reservation cycles[3] (RC). Within an RS, a reservation is made through a sequence of five-phase dialogues made in each RC, between a contenting node and its neighbor nodes. A reservation cycle consists of the following five phases:

1. *Reservation request phase:* In this phase, a node that wants to make a reservation sends a reservation request packet (RR) with probability *p*. The node that is sending an RR is referred as a requesting node (RN).
2. *Collision report phase:* If a node receives multiple RR packets from phase 1, it transmits a collision report packet (CR) to report the collision that just occurred otherwise it remains silent. If the RN from phase 1 receives one or more CR packets it assumes that a collision occurred, otherwise it assumes that its RR packet reached every node safely. In such a case it becomes a transmission node (TN) and will proceed and make its reservation in phase 3. It is clear that in this phase the hidden terminal problem is eliminated. If two RN nodes are hidden from each other then their requests will collide and both will receive a CR packet.
3. *Reservation confirmation phase:* A TN sends a reservation confirmation packet (RC) in this phase, informing all nodes within one hop that the slot has been reserved. All nodes that received the RC packet will not content further for this slot.

[2] The value of the parameter N is the node population of a given network.
[3] The value of the parameter M must be determined heuristically for a given network.

4. *Reservation acknowledgment phase:* In this phase, a node acknowledges the RC packet received in phase 3 by sending a reservation acknowledgment packet (RA). This tells a TN that its reservation has been established. If a TN is not connected with any other nodes ("isolated" deadlock), it does not receive any RA and becomes aware of its isolation and no longer considers itself as a TN (otherwise it would always remain a TN). This phase also informs nodes, which receive the RA packet, that they are two hops away from the TN. These nodes mark this slot as reserved and will not contend further by becoming blocked (B) in this slot.
5. *Packing and elimination phase (P/E):* In this phase two kinds of packets are transmitted. Every node that is two hops away from a TN, which has made its reservation since the last P/E phase, sends a packing packet (PP). A node receiving the PP packet learns that there is a recent success three hops away and some of its neighbors cannot contend further for this slot. It can take advantage of this and adjust its contention probability *p* accordingly. In the same phase, each TN sends an elimination packet (EP) with probability of 0.5. This is intended for another TN, which would be potentially adjacent, in an attempt to resolve a *non-isolated* deadlock[4]. If a TN does not transmit, but receives an EP in this phase, it learns there is a deadlock. In this case it will re-label the slot as reserved by the other TN (the one that sent the EP) and will receive, rather than transmit, in this slot. It will contend further in other slots. There is no need to inform its neighbors about this re-labeling event.

In FPRP the contention probability *p* for each node is calculated using an algorithm called pseudo-Bayesian. Each node keeps two estimates, one for the number of nodes (n_c) that contend within two hops and one for the number of nodes (n_b) within two hops that need reservation but cannot contend in the current slot due to a nearby reservation. Three parameters (estimated independently with another program) give the portion of its neighbor contenders that cease to contend in the current slot due to a success reservation. Those parameters are: R_1 if the node is one hop away from the success, R_2 if the node is two hops away and R_3 if the node is three hops away. The algorithm works as follows:

1. At the beginning of a reservation slot, a node sets: $n_c=n_b$ and $n_b=0$
2. After every reservation cycle if a node hears an:
    a. **Idle** - it sets: $n_c=n_c-1$
    b. **Collision** - it sets: $n_c=n_c+(e - 2)^{-1}$

[4] Because nodes cannot receive while transmitting during phase 1, it is possible that two adjacent transmitting nodes do not detect the collision because they do not have a common neighbor to inform them. Thus, both will become TN during phase 2 and a deadlock will be formed.

c. **Success one hop away** - it sets: $n_c=n_c*(1-R_1)-1$ and $n_b=n_b+n_c*R_1$

d. **Success two hops away** - it sets: $n_c=n_c*(1-R_2)-1$ and $n_b=n_b+n_c*R_2$

e. **Success three hops away** - it sets: $n_c=n_c*(1-R_3)$ and $n_b=n_b+n_c*R_3$

3. It calculates the next cycle contention probability by setting $p=1/n_c$.

2.2.5 HRMA

HRMA is based on simple half-duplex, very slow frequency hopping spread spectrum (FHSS) and it can be viewed as a time slot reservation protocol in which in a time slot is assigned a separate frequency channel.

One of the *L* available frequencies (f_0) is used as a dedicated synchronizing channel on which the nodes exchange synchronization information. The rest of the frequencies are divided into M=[(L-1)/2] frequency pairs (f_i, f_i^*), i=1,2,3…M. For any i, f_i is used for sending or receiving hop reservation packets (HR), request to send (RTS) packets, clear to send (CTS) packets and data packets while f_i^* is used for sending or receiving acknowledgments to data packets sent on f_i.

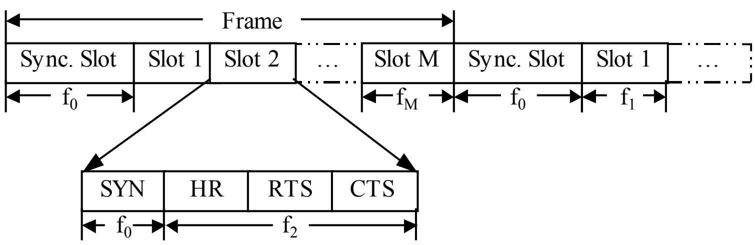

Figure 2-3: Slot and frame structure of HRMA

As shown in Figure 2-3 an HRMA slot consists of one synchronizing period, one HR period, one RTS period and one CTS period. Each slot is assigned a frequency hop, which is one of the M frequency hops in the common hopping sequence. All the nodes that are not transmitting or receiving data packets (idle nodes) hop together to the synchronizing frequency f_0 and exchange synchronizing messages[5] during the synchronization period of each slot. During the HR, RTS and CTS periods of each slot, all idle nodes must hop on the common frequency hop assigned to each slot.

[5] The synchronization messages allow nodes to agree on the beginning of frequency hop in the common hopping sequence, the current frequency hop, etc.

When an idle node has a data packet to transmit before the RTS period of a given slot has started, the node backs off (a random number of HRMA slots) if the HR period contains an HR packet. Otherwise the node sends an RTS to the intended receiver and waits for the CTS. When the intended receiver node receives an RTS it replies with a CTS packet during the CTS period of the same slot, and it stays on the same frequency hop waiting for the data packet. If the sender node receives no CTS it backs off and retries during another slot. If however it receives the CTS packet, it then remains on the same frequency hop of the current slot and starts sending the data packet to the receiver. Both transmitter and receiver nodes stay on the same frequency hop until the end of the data packet transmission. After the CTS period of a slot, all other nodes that are not transmitting or receiving hop to f_0 to synchronize and then hop to the next frequency hop in the common hopping frequency pattern. If an idle node has a data packet to transmit after the end of the HR period of a given slot, it backs off and tries in another slot.

When the data that need to be exchanged between sender and receiver require multiple HRMA frames for their transmission, the sender notifies the receiver in the header of the data packet and the receiver sends an HR packet during the HR period of the same slot of the next frame. This informs the neighbors of the receiver (hidden terminal problem) that they cannot attempt to use this slot (or frequency hop) due to the data transmission. The sender sends an RTS packet during the RTS period of the slot to jam any possible RTS addressed to its own neighbors (exposed terminal problem), which may not hear the HR from the receiver. Both sender and receiver keep silent during the CTS period of the slot, and more data are transmitted after the end of the CTS period. The hop remains reserved in a similar fashion until the end of the data transmission. After the source sends a data packet, it hops to the corresponding acknowledgment frequency and the receiver sends an acknowledgment back to the source on this acknowledgment frequency.

2.2.6 IEEE 802.11

The MAC layer of the IEEE 802.11 standard defines two different access methods with which nodes can access the medium channel, the distributed coordination function (DCF) and the point coordination function (PCF). The PCF is an optional function that is used to implement time-sensitive services like voice or video transmissions. The PCF is not used in an ad-hoc network implementation and thus we will not presented it here.

The DCF, which is the basic access mechanism, is basically a carrier sense multiple access with collision avoidance mechanism (CSMA/CA).

If a node wants to transmit (Figure 2-4) it first senses the medium. If the medium is busy then it defers its transmission, otherwise if the medium is free for a specific time (called distributed inter frame space (DIFS) in the standard), then the node transmits an RTS[6] packet to the intended receiver. If the medium is free, the destination node responds to the source node with a CTS packet. After receiving[7] the CTS packet, the source node starts the transmission of its data packet (called MPDU, for Mac Protocol Data Unit). All other nodes receiving either the RTS and/or the CTS set their **Virtual Carrier Sense** indicator (called NAV, for Network Allocation Vector) and they defer long enough (for the given duration) until the transaction is completed. At the end of the data packet transmission, an ACK packet is send to the transmitter from the receiver as an acknowledgment of the correct transmission of the data.

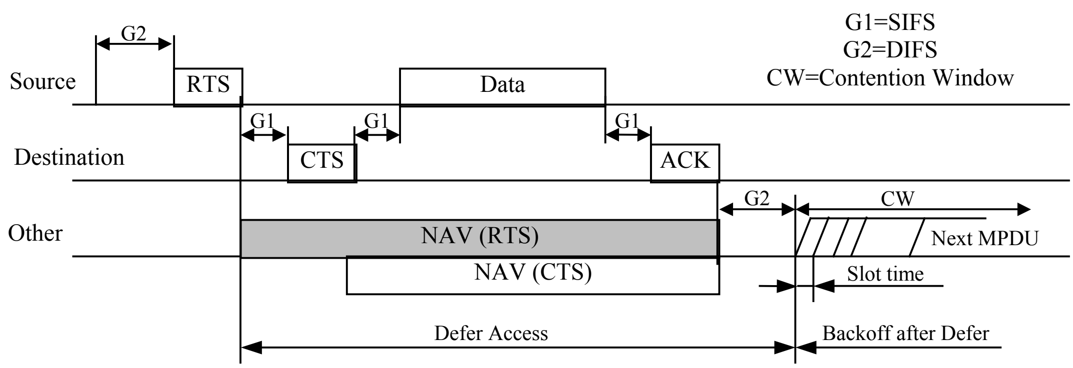

Figure 2-4: Transmission of an MPDU in IEEE802.11

If the sender node has more than one packet to transmit to the same destination node, after the reception of the ACK packet from its first packet transmission, it only waits SIFS time before sensing the medium. This gives it an advantage regarding the other nodes since it only waits SIFS time that is less than DIFS time which other nodes

6 The RTS, CTS and ACK packets in IEEE 802.11 standard include the source, destination and the duration of the following transaction.
7 The time that is used to separate transmissions belonging to a single dialogue (e.g. RTS-CTS) is called short inter frame space (SIFS). The duration of SIFS is the time it takes for a transmitting node to switch back to receive mode and be capable of decoding an incoming packet. The DIFS time is longer than the SIFS time.

have to wait before sensing the medium again. Thus by gaining again access to the channel it continues to transmit its data packets until the last packet has been sent.

The collision avoidance portion of CSMA/CA used in IEEE 802.11 is performed through a random backoff procedure. As mention earlier, if a node wants to transmit a packet and senses the channel to be busy, it defers long enough until the ongoing transaction is completed. After that, the node waits until the channel becomes idle for a DIFS time and it then computes a random backoff time. The slot time in IEEE 802.11 is much smaller than the time required transmitting an MPDU (data packet) and is used to determine the backoff time of a node. The random backoff time is an integer multiple of time slots. Initially, the node randomly computes a backoff time in the range [0-7]. After the medium becomes idle for a DIFS period, nodes decrement their backoff timer until the medium becomes busy again or the timer reaches zero. If the timer has not reached zero and the medium becomes busy, the node freezes its timer. When the timer is finally decremented to zero, the node transmits its MPDU. If two or more nodes decrement to zero at the same time, a collision will occur, and each node will have to randomly generate a new backoff time in the range [0-15]. For each retransmission attempt, the backoff time range grows exponentially as $[0,2^{2+i}-1]$ where i is the number of consecutive times a node attempts to send an MPDU. The idle period after a DIFS period is referred to as the contention window (CW).

2.2.7 CATA

CATA is based on dynamic topology-dependent transmission scheduling and employs similar handshake procedures as those used in collision-avoidance MAC protocols. Contention is limited among nodes within two hopes of one another, which provides a very efficient spatial reuse of the bandwidth available. Reservations in CATA support unicasting, multicasting and broadcasting simultaneously and adapt to dynamic service time. After a successful reservation, a node is able to transmit data packets collision-free on the reserved time slots in the subsequent frames, until the reservation is terminated, thus CATA is able to support real-time applications. CATA assumes that radios used are half-duplex and the physical links are bi-directional. The receiver of a radio is always on while it is not transmitting. The data transmitted by a node over a reserved collision-free time slot are called *messages* and the end of a reservation is notified to the receiving nodes by data packets within the message.

The operation of CATA is based on the following basic principles:

1. Data from a source must flow without interference from other sources over a reserved slot. Because of possible hidden terminals, the receiver(s) of a message must be the one(s) telling the potential sources that the slot is reserved while the sender of a message must be responsible for telling the potential receiver(s) that there exists interference in the slot.
2. The sender of a broadcast or multicast flow should not have to receive explicit feedback on the reservation from each neighbor. In CATA, this is accomplished with what amounts to negative acknowledgments to reservation requests, and by each node sending a control packet at the beginning of a slot in which it is busy receiving data.

To accomplish slot reservations according to the above principles, CATA divides a slot into five mini-slots. The first four mini-slots are intended for control packets and are called control mini-slots (CMS1 to CSM4). The last mini-slot is meant for data and is called data mini-slot (DMS). In practice, the DMS should be much longer than any CMS to reduce the protocol overhead.

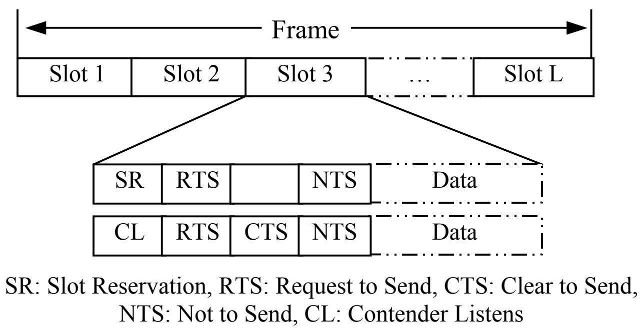

Figure 2-5: Slot and frame structure of CATA

Figure 2-5 shows how slots are identified as reserved and how collision-free data are sent over reserved slots. CMS1 is used to provide a "busy tone" to senders attempting to establish transmissions. Every node(s) that receives data during the DMS of 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 during the DMS of the current slot sends a request to send (RTS) packet during CMS2. This action 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. Both the sender and the receiver node(s) remain silent during CMS3 and the sender sends a not to send (NTS) packet during CMS4. With this mechanism CATA ensures that after a successful reservation, the same slot in the subsequent frames will remain collision free until the end of the message transmission and the termination of the reservation, eliminating all exposed and hidden terminal problems.

Figure 2-5 also shows how nodes content and reserve slots. A node that wants to make a reservation in the current slot sends a request only if it is not receiving data in this slot and there was no busy tone during CMS1. There is a slightly different reservation mechanism when a node is sending an RTS[8] for broadcast, unicast and multicast.

In the case a node wants to make a reservation for unicast transmission, it sends a RTS during CMS2. If the intended receiver received the RTS correctly (no jamming occurred by another RTS) it responds with CTS during CMS3 otherwise it remains silent. The sender of the RTS detects a successful reservation with the reception of correct CTS (no jamming occurred by another CTS or no CTS received at all). It then can start transmitting data in the DSM of the current slot and in the same slot in all subsequent frames until the termination of the message and the end of the reservation.

In the case a node wants to make a reservation for multicast or broadcast transmission, it sends a RTS during CMS2. If the intended receivers received the RTS correctly, they remain silent during CMS3 and CMS4 otherwise they send a NTS at CMS4 as a negative acknowledgment to the intended multicast or broadcast reservation. The sender node of the multicast or broadcast RTS detects the reservation failure either if it receives a NTS or noise (due to multiple NTS) during CMS4. Otherwise it can start transmitting data in the DSM of the current slot and in the same slot in all subsequent frames until the termination of the message and the end of the reservation.

2.3 SUMMARY

All of the MAC protocols presented in this chapter attempt to solve the hidden terminal problem from which the CSMA protocol suffers which degrades its performance to that of the pure ALOHA protocol (i.e., without carrier sensing). In addition, most of the above protocols introduce a backoff algorithm to resolve collisions

[8] The protocol assumes that RTS packets have information about the type of the transmission request (unicast, broadcast and multicast) and the node(s) that the RTS is addressed to.

and increase channel utilization by speeding up the slot reservation mechanism. Two key performance limitations of most collision avoidance MAC protocols are that:

- They do not support real time applications. Real time applications use delay sensitive data that usually need some channel bandwidth allocation (or continuous time slot reservations) in order to be delivered in time. A MAC protocol in order to support real time applications, it must be able to reserve time slots for continuous collision free data packets transmissions.
- They lack explicit support of unicast, multicast or broadcast transmissions. Most MAC protocols support either unicast or broadcast transmissions but not both. If they support unicast transmissions, in order to support multicasting they must transmit the multicast packet multiple times, once to each member of the multicast transmission group. If on the other hand they support broadcast transmissions, in order to support unicasting they must transmit the unicast packet as a broadcast packet, which degrades the performance of the protocol.

As shown in Table 2-1 only IEEE802.11 and CATA explicitly support unicast/ broadcast transmissions and real time applications. In addition, CATA supports multicast transmissions but its disadvantage compared to IEEE802.11 is that it does not have a backoff algorithm to increase channel utilization. Generally speaking, 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.

	MACAW	DBTMA	FPRP	HRMA	IEEE 802.11	CATA
Real Time Application Support	No	No	Yes	Yes	Yes	Yes
Transmission type	Unicast	Unicast	Broadcast	Unicast	Unicast, Broadcast	Unicast, Broadcast, Multicast
Backoff Algorithm	Yes	No	Yes	No	Yes	No

Table 2-1: Comparison of MAC protocols

In the following chapters we analyze the performance of CATA protocol for different number of nodes and network topologies. Unlike CATA, the performance of IEEE802.11 has been examined extensively (Chhaya and Gupta [10], Zaki, Makrakis and Gallardo [11], Cali, Conti and Gregori [12]).

CHAPTER 3

NETWORK & CHANNEL MODELS AND BACKOFF ALGORITHM

3.1 SYSTEM MODEL AND ASSUMPTIONS

The experimental results presented in the following chapters, 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[9] 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[10] 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.

[9] 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.
[10] 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*.

3.2 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 3-1.

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 3-1: 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 3-1 and Figure 3-2, 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.

Remember that in CATA (section 2.2.7), 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.

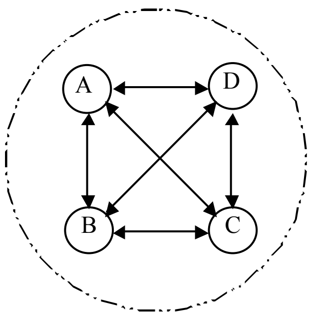

Figure 3-1: A 4-node, fully connected network.

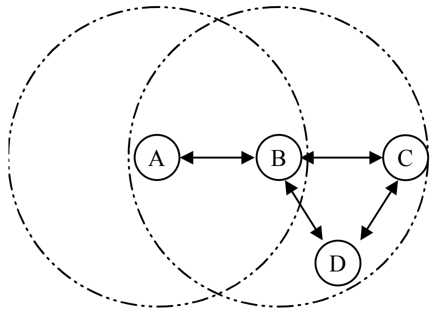

Figure 3-2: A 4-node, partial connected network

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

3.3 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.

3.4 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 Chapter 4 and Chapter 5 we present the results of our simulation study using the ABA backoff algorithm presented here. In Chapter 6, we compare ABA with the backoff mechanism of slotted aloha for some selected network configurations.

CHAPTER 4

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[11] 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).

4.1 FULLY CONNECTED NETWORK TOPOLOGY

In this network topology (Figure 4-1), 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.

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

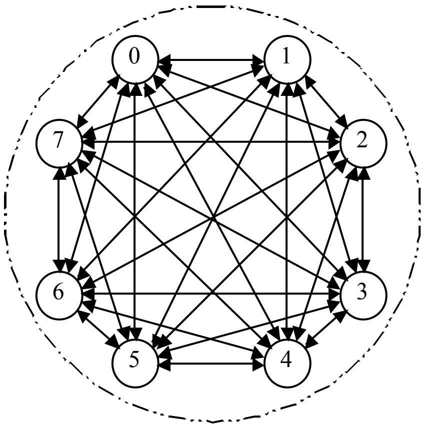

Figure 4-1: An 8-node, fully connected network topology

Figure 4-2 and Figure 4-3 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.

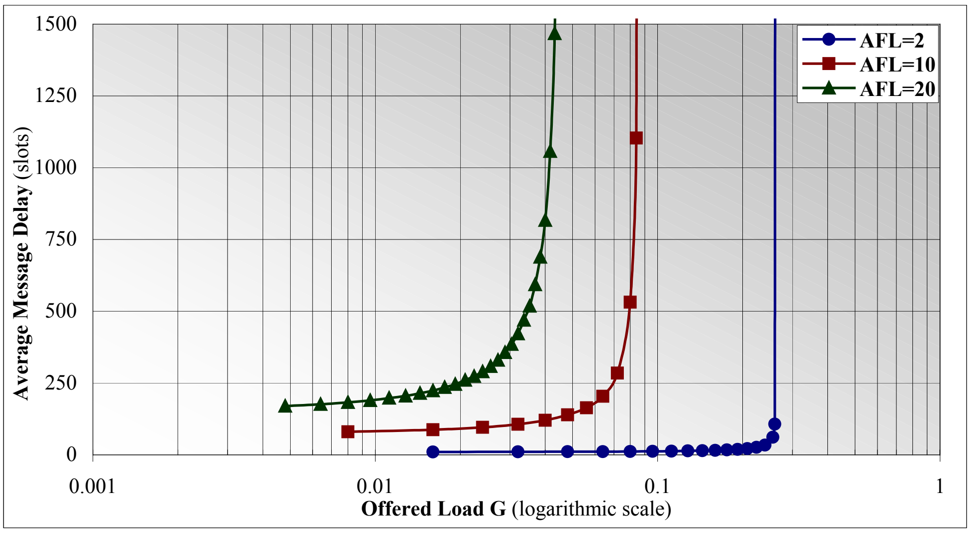

Figure 4-2: Average message delay for 8-node, fully connected network.

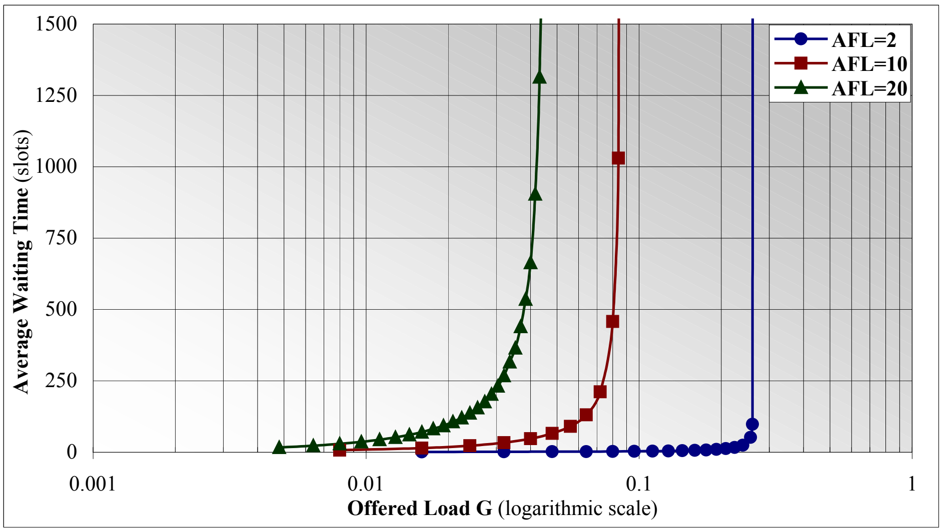

Figure 4-3: Average waiting time for 8-node, fully connected network.

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[12].

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 4-1 shows the maximum offered load, average message delay and average waiting time values, for which the system is stable.

Figure 4-4 shows the channel utilization of the system versus offered load. In Figure 4-4, 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.

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

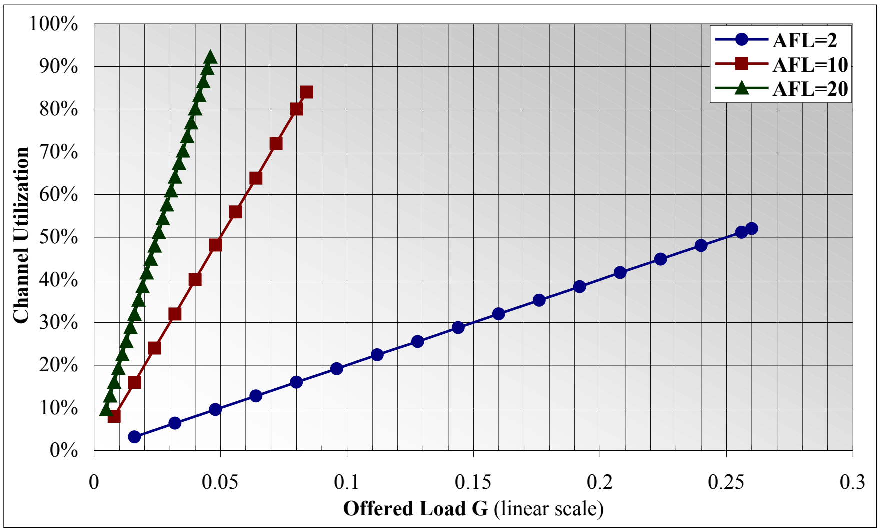

Figure 4-4: Channel utilization for 8-node, fully connected network.

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 4-1 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 4-1: Maximum metric values for 8-node, fully connected network

4.2 TWO-AREA NETWORK TOPOLOGY

In Figure 4-5, 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=Min{4^2+1, 8}=8.

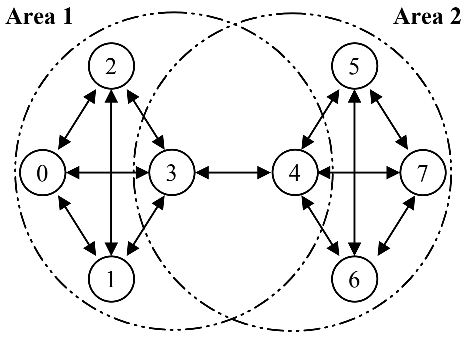

Figure 4-5: An 8-node, two-area network topology

4.2.1 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 4-6 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.

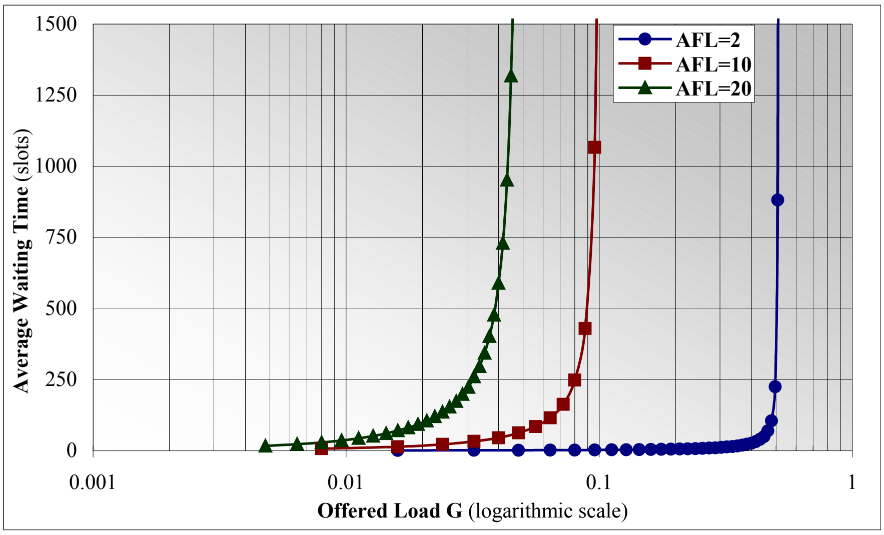

Figure 4-6: Unicast average waiting time for 8-node, two-area network.

Comparing average waiting time with that for the fully connected network (Figure 4-2), we see that for the same offered load, the average waiting time is decreased (see Table 4-2) and that the maximum supported offered load is increased (see Table 4-3), 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 25% 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[13]. 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 4-7 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 4-2.

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

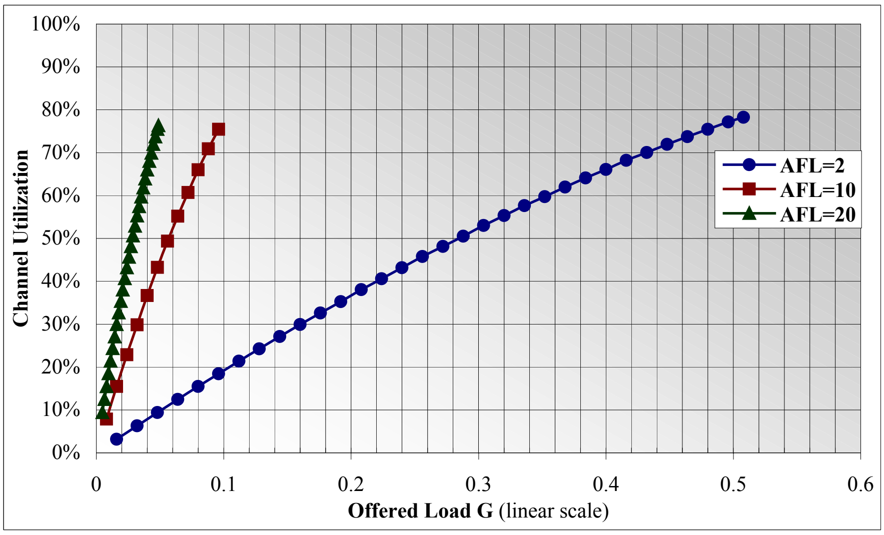

Figure 4-7: Unicast channel utilization for 8-node, two-area connected network.

For this reason, a better metric to compare these network topologies is the system throughput (Figure 4-8). 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.

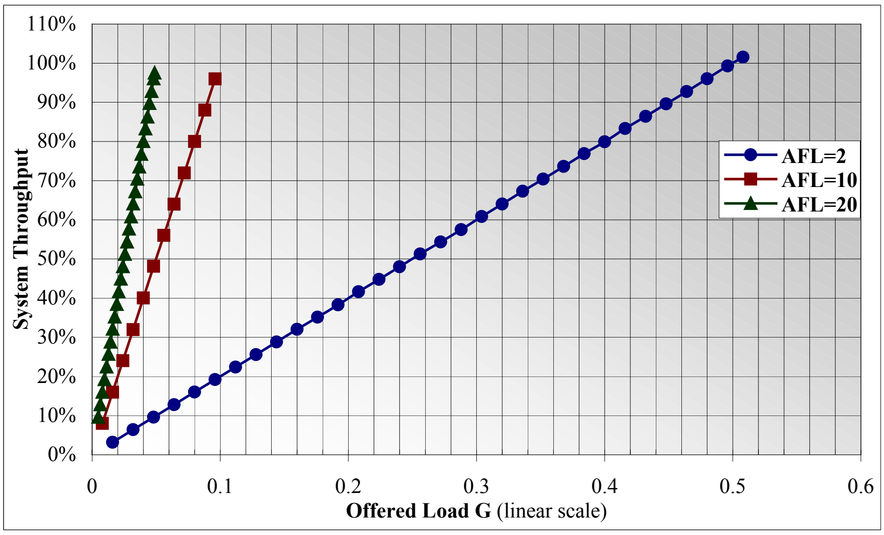

Figure 4-8: Unicast system throughput for 8-node, two-area connected network.

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 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.

	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 4-2: 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 4-3: Unicast maximum metric values for 8-node, two-area connected network.

Finally, Figure 4-10 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 4-9. In Figure 4-10, the horizontal axis typically represents the total workload of the system, which depends of the message size.

$$(1)\ cf = \frac{\sqrt{\mathrm{Var}(X)}}{\mathrm{E}(X)} \qquad (2)\ \mathrm{Var}(X) \approx \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^{\,2} - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2\right] \qquad (3)\ \mathrm{E}(X) \approx \frac{1}{n}\sum_{i=1}^{n} X_i$$

where $X_i = i^{th}$ Transm. Message Waiting Time and n = Number of Transm. Messages $= 10^6$

Figure 4-9: Coefficient of variation equations.

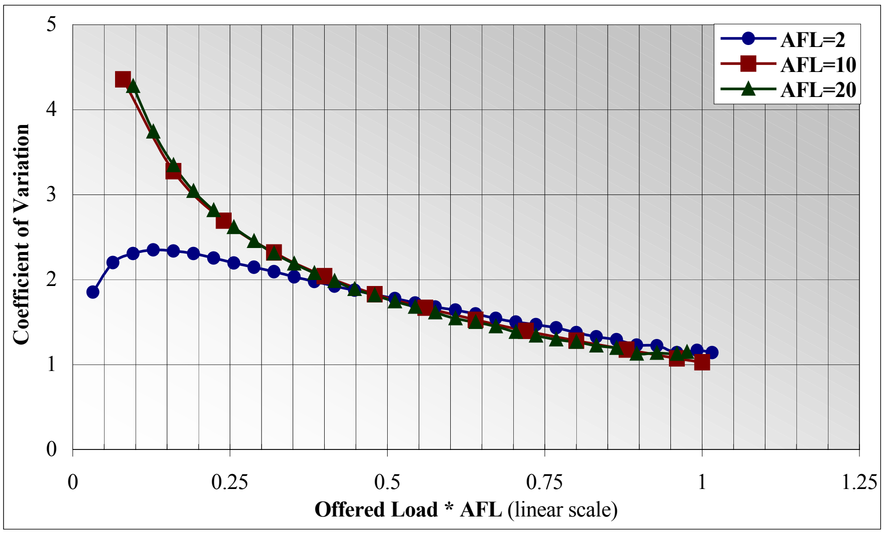

Figure 4-10: Unicast coefficient of variation for 8-node, two-area connected network.

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 3.1 . 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[14]. 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 variation converges to values around one regardless of the message size.

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

4.2.2 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.

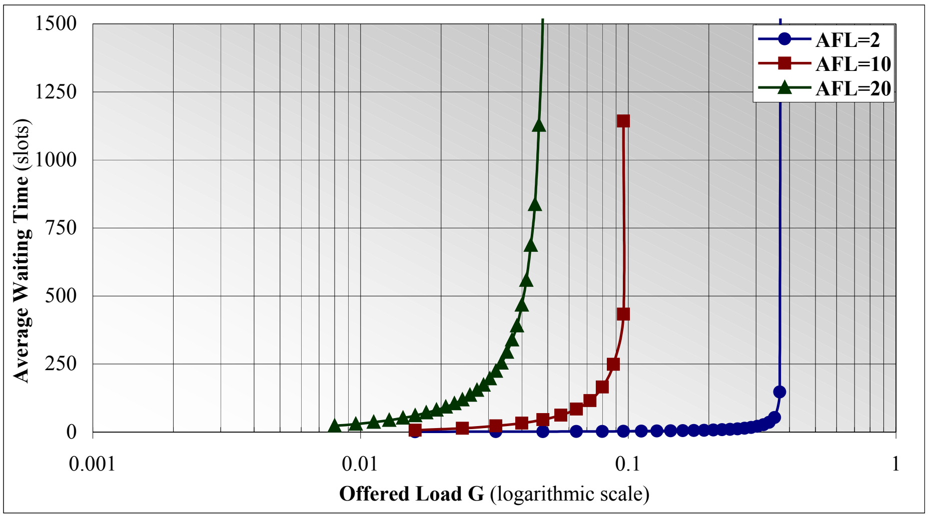

Figure 4-11: Broadcast average waiting time for 8-node, two-area network.

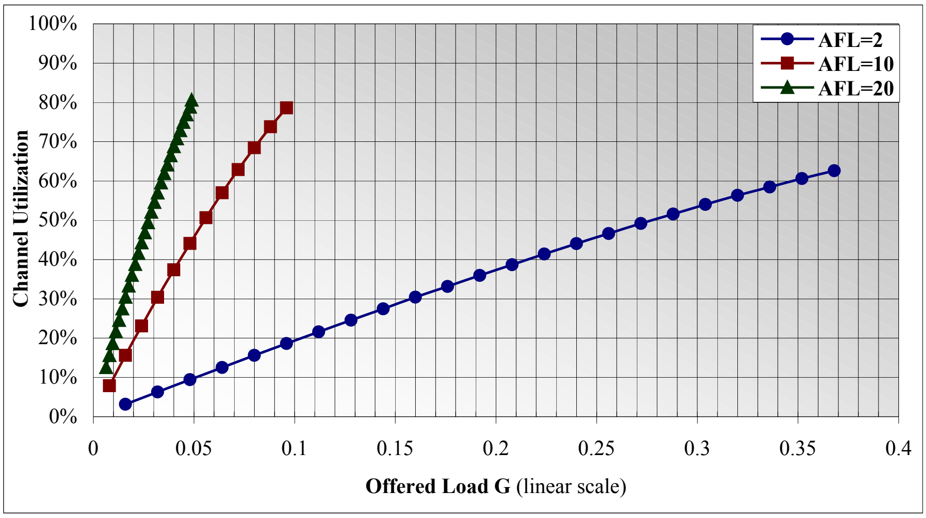

Figure 4-12: Broadcast channel utilization for 8-node, two-area connected network.

Once again average waiting time is decreased (see Figure 4-11) and the maximum supported offered load is increased (see Table 4-1 and Table 4-5) compared with that for the corresponding fully connected network case, due to the spatial channel reuse in the two-area network case.

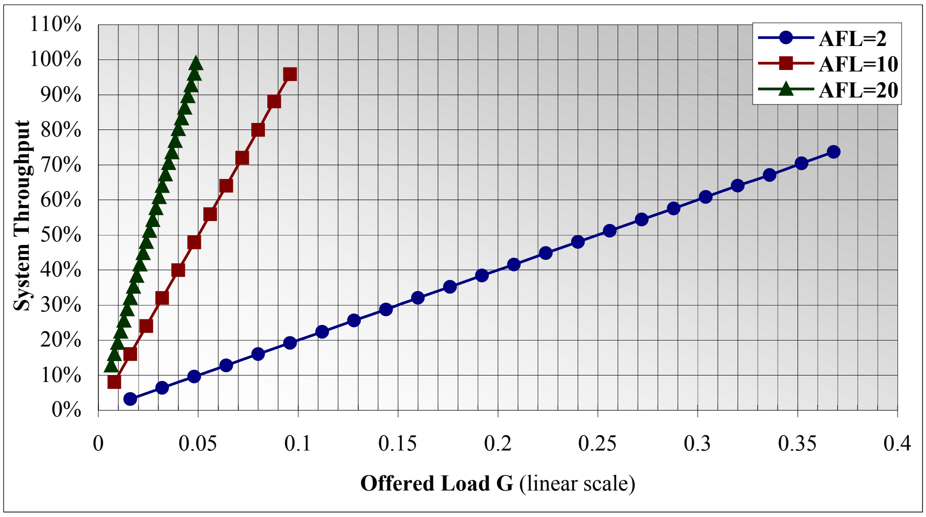

Figure 4-13: Broadcast system throughput for 8-node, two-area connected network.

	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	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 4-4: 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 4-5: Broadcast maximum metric values 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 4-3 and Table 4-5). 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[15] 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.

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 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.

Finally, Figure 4-14 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.

[15] 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 4.3.2

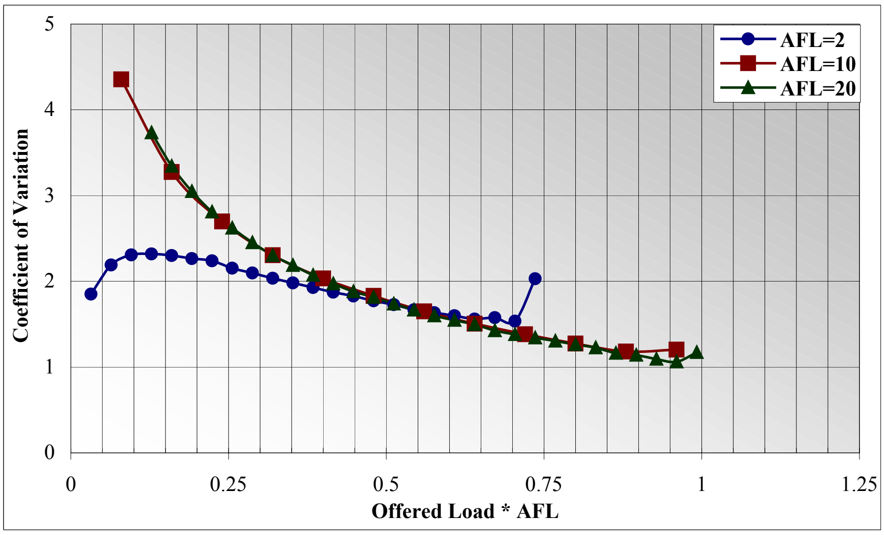

Figure 4-14: Broadcast coefficient of variation for 8-node, two-area connected network.

4.3 EIGHT-AREA NETWORK TOPOLOGY

In Figure 4-15 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.

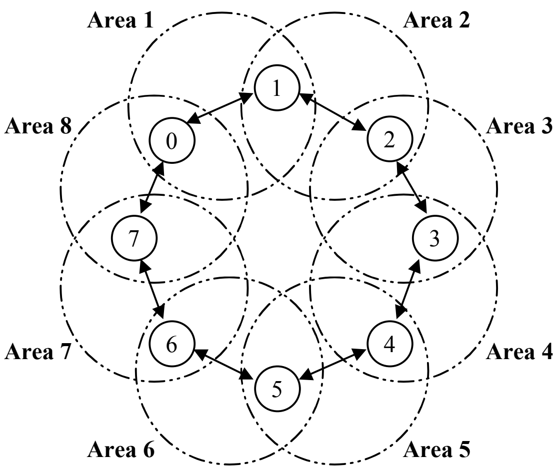

Figure 4-15: An 8-node, eight-area network topology

4.3.1 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 4-16 shows the average waiting time versus offered load. Average waiting time, for the same offered load, has been significantly reduced (see Table 4-7). This is caused not only by the higher spatial reuse but also by the smaller frame size compared to the previous network topologies.

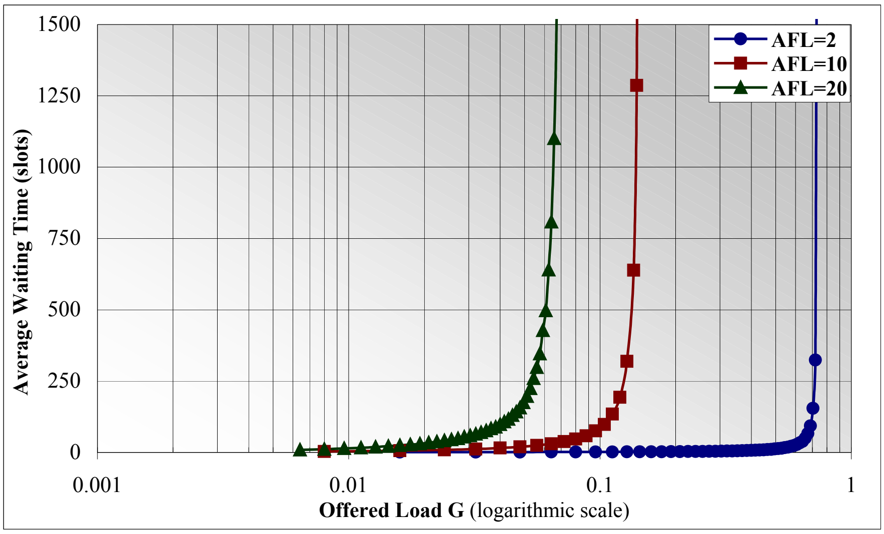

Figure 4-16: Unicast average waiting time for 8-node, eight-area network.

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 4.1 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 4-17 and Figure 4-18 show the channel utilization and system throughput versus offered load, respectively. Higher channel slot utilization and higher throughput than previous area-networks is achieved due to higher spatial reuse of the communication channel but also due to smaller frame length.

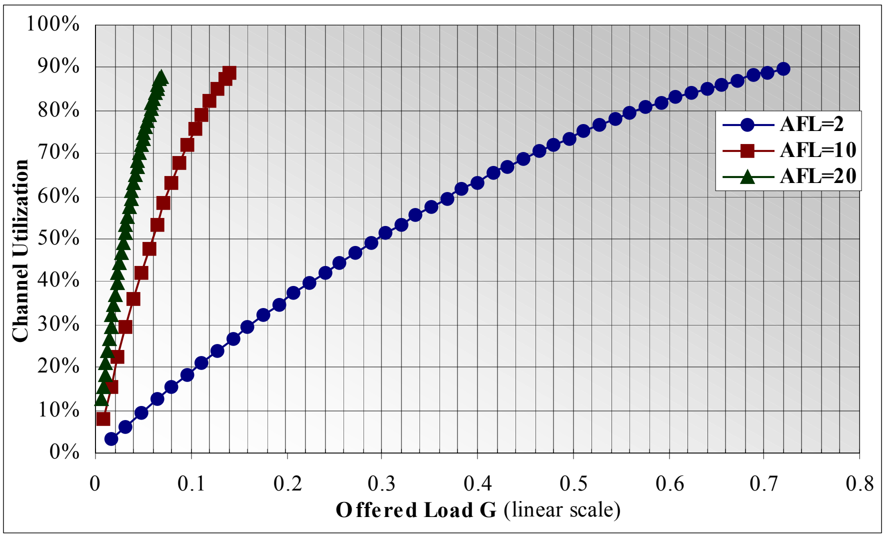

Figure 4-17: Unicast channel utilization for 8-node, eight-area connected network.

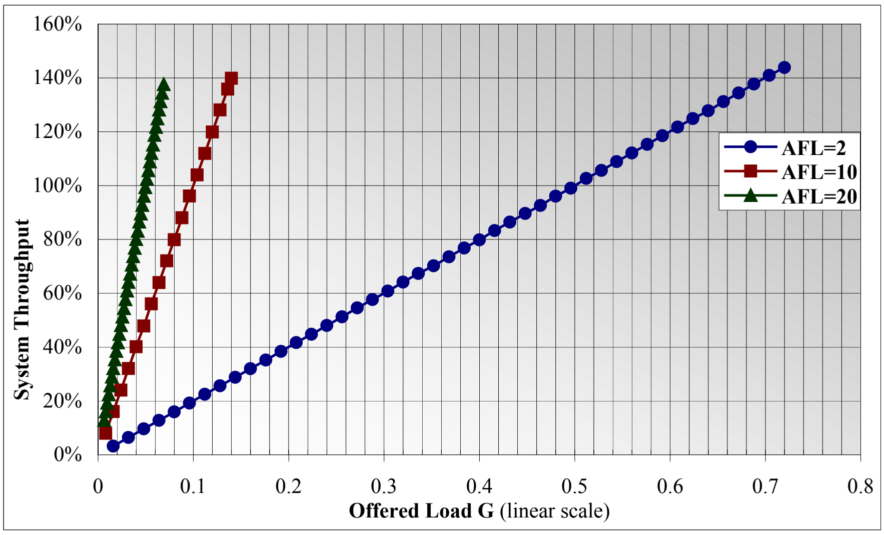

Figure 4-18: Unicast system throughput for 8-node, eight-area connected network.

Table 4-6 shows the maximum offered load and the maximum channel slot utilization and throughput supported by this network. As explained in section 4.2.1 , 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.

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 4-6: Unicast maximum metric values for 8-node, eight-area connected network.

	Fully connected network			Two-area network			Eight-area network		
AFL	2	10	20	2	10	20	2	10	20
Offered Load	0.256	0.0720	0.0416	0.256	0.0720	0.0416	0.256	0.0720	0.0416
Average Message Delay	61.3	284.8	1057.2	17.4	236.2	882.6	9.6	84.0	203.2
Average Waiting Time	52.3	211.8	904.0	8.4	163.2	729.7	3.6	38.0	107.3
Channel Utilization	51.2%	72.0%	83.2%	45.8%	60.7%	68.0%	44.5%	58.5%	64.9%
System Throughput	51.2%	72.0%	83.2%	51.3%	72.0%	83.2%	51.2%	72.0%	83.1%

Table 4-7: Unicast comparison between fully connected, two-area and eight-area network.

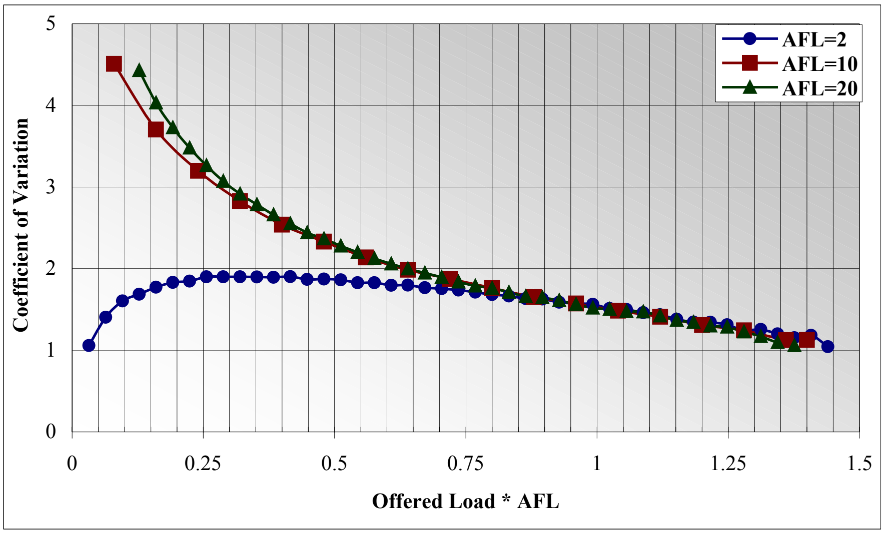

Figure 4-19: Unicast coefficient of variation for 8-node, eight-area connected network.

Finally, Figure 4-19 shows the coefficient of variation versus offered load multiplied by AFL.

4.3.2 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.

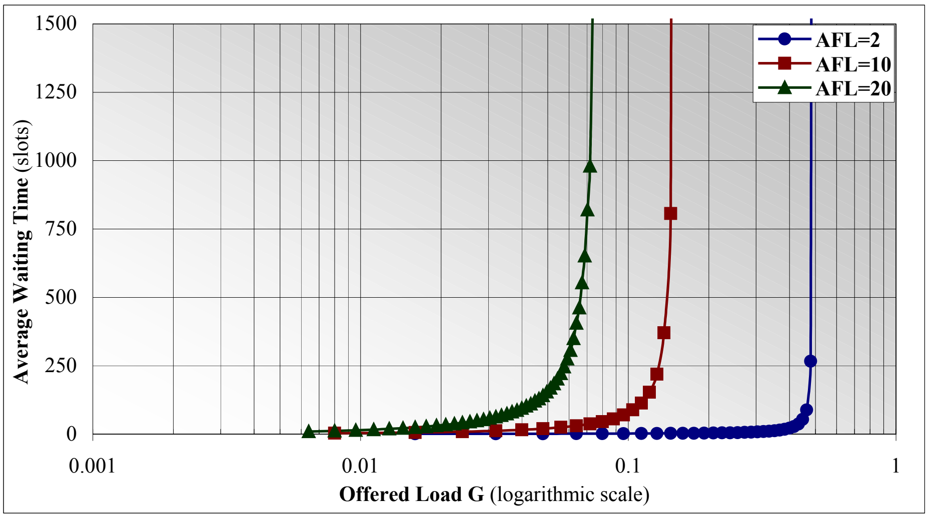

Figure 4-20: Broadcast average waiting time for 8-node, eight-area network.

As shown in Figure 4-20, higher spatial reuse and smaller frame size has once again reduced significantly the average waiting time (see Table 4-9) and increased the maximum supported offered load (see Table 4-8). Channel utilization and system throughput (see Figure 4-21 and Figure 4-22, respectively) is also increased significantly compared to the fully connected and the two-area networks.

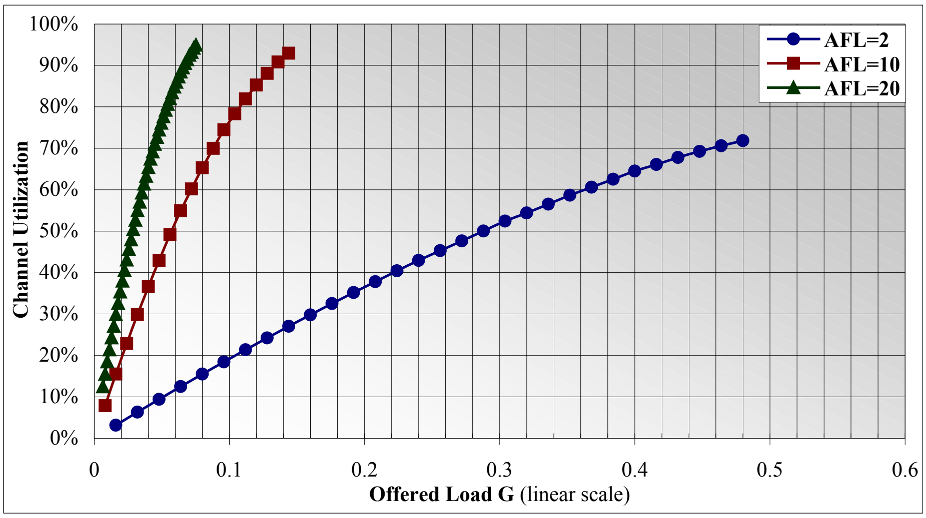

Figure 4-21: Broadcast channel utilization for 8-node, eight-area connected network.

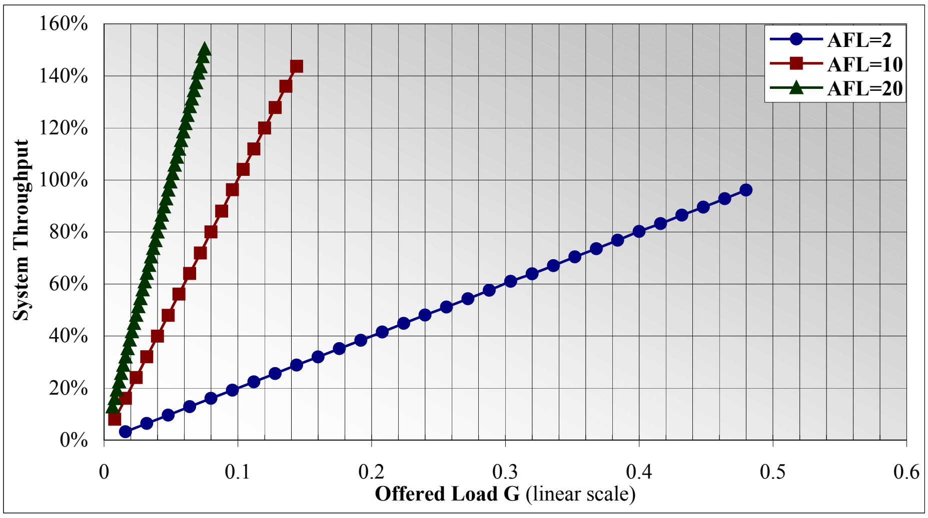

Figure 4-22: Broadcast system throughput for 8-node, eight-area connected network.

A difference in the maximum supported offered load between unicast and broadcast transmissions (see Table 4-6 and Table 4-8) 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 4-8: 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[16]. 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[17] 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			Two-area network			Eight-area network		
AFL	2	10	20	2	10	20	2	10	20
Offered Load	0.256	0.0720	0.0416	0.256	0.0720	0.0416	0.256	0.0720	0.0416
Average Message Delay	61.3	284.8	1057.2	20.8	238.8	839.5	11.5	83.3	199.3
Average Waiting Time	52.3	211.8	904.0	11.8	165.9	686.5	5.5	37.3	103.1
Channel Utilization	51.2%	72.0%	83.2%	46.6%	62.9%	70.8%	45.3%	60.2%	67.4%
System Throughput	51.2%	72.0%	83.2%	51.2%	72.0%	83.2%	51.1%	71.9%	83.4%

Table 4-9: Broadcast comparison between fully connected, two-area and eight-area network.

Finally, Figure 4-23 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

[16] In the two-area network although this phenomenon existed, it had a little impact because only two nodes experienced it..
[17] The impact of the SR packet to neighbor nodes was explained in section 3.2

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

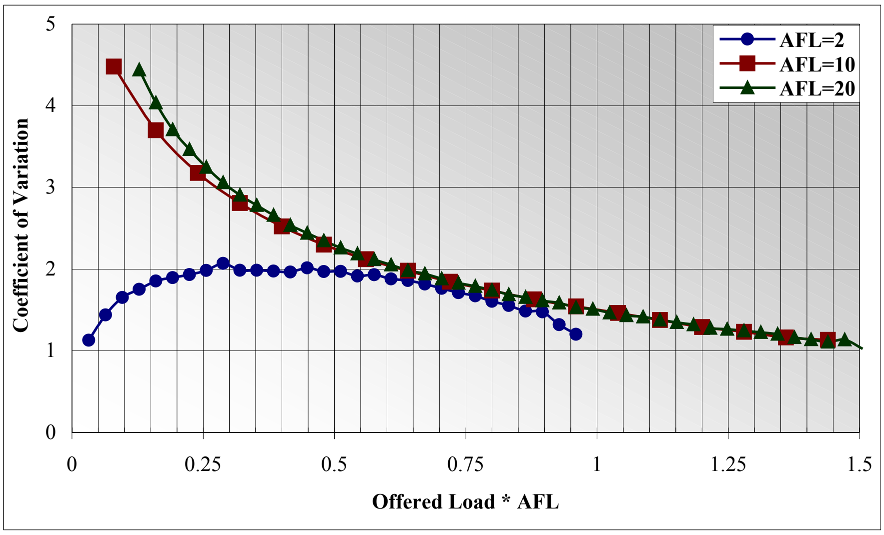

Figure 4-23: Broadcast coefficient of variation for 8-node, eight-area connected network.

CHAPTER 5

SIXTEEN NODE SIMULATION STUDY

All simulation results presented in this section consider sixteen-node populations placed in fully and partially connected network topologies. Sixteen-node population is selected because it has twice as many nodes than the eight-node population network topologies presented so far and a comparison between them can be easily performed. In addition, an ad-hoc network with sixteen-node population is considered "heavy loaded" and network load directly affects the performance of the MAC protocol. Both unicast and broadcast transmission types are examined with average message length (AFL) of 2, 10 and 20 slots per message. 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[18] 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). Channel utilization and coefficient of variation graphs are not shown because they are very similar to the corresponding 8-node network graphs and their presentation would only unnecessarily increase the size of the thesis.

5.1 FULLY CONNECTED NETWORK TOPOLOGY

In this network topology (Figure 5-1), 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. Once again,

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

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{15^2+1, 16}=16.

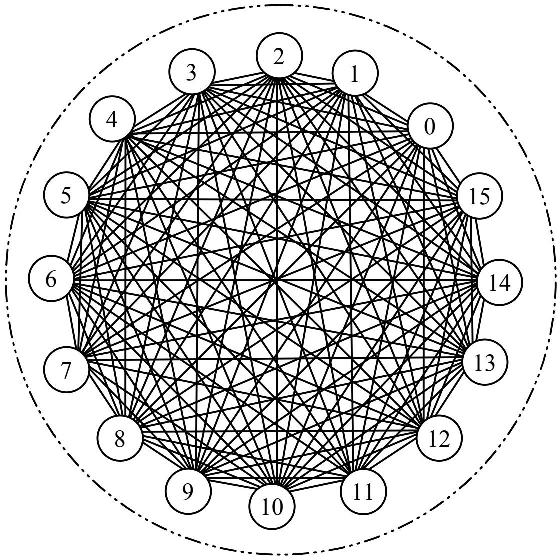

Figure 5-1: A 16-node, fully connected network topology

Figure 5-2 shows the average waiting time versus offered load (offered load axis is in logarithmic scale). As explained in section 4.1 , 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). For this network topology message delivery delays have almost been doubled compared to the 8-node, fully connected network, because the frame length is doubled. For example if a message has AFL=20, in this network topology, a node will be able to completely deliver the message in approximately (20-1)*16+1=305 slots where in the 8-node, fully connected network it would only take 153 slots.

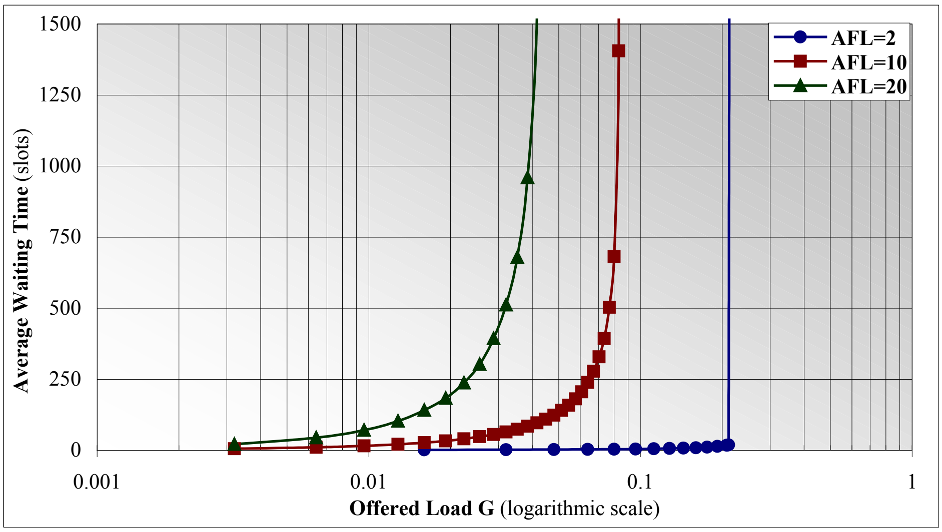

Figure 5-2: Average waiting time for 16-node, fully connected network.

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 5-1 shows the maximum offered load, average message delay and average waiting time values, for which the system is stable.

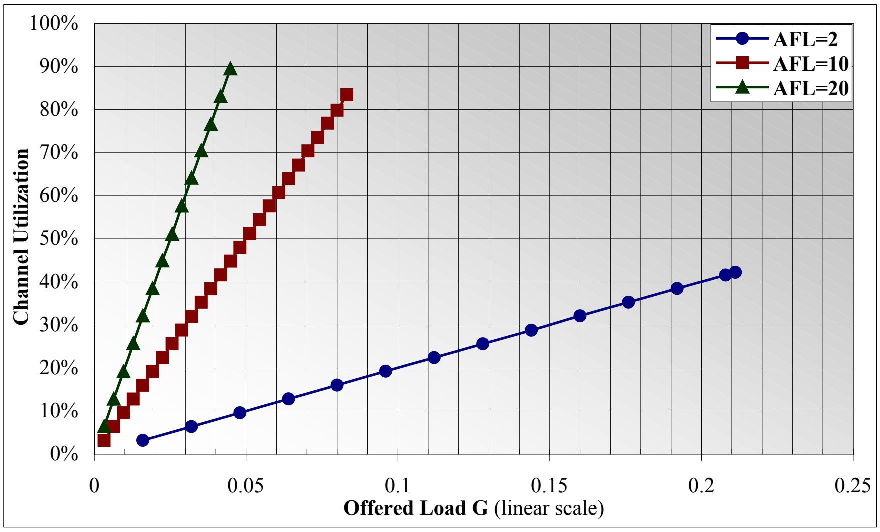

Figure 5-3: Channel utilization for 16-node, fully connected network.

Figure 5-3 shows the channel utilization of the system versus offered load (offered load axis is in linear scale). Table 5-1 also shows the maximum utilization values, for a stable system.

AFL	2	10	20
Offered Load	0.211	0.0832	0.0448
Average Message Delay (slots)	36.11	1550.51	4145.97
Average Waiting Time (slots)	19.11	1405.47	3840.76
Channel Utilization	42.187%	83.423%	89.466%

Table 5-1: Maximum metric values for 16-node, fully connected network

5.2 TWO-AREA NETWORK TOPOLOGY

In Figure 5-4, nodes are divided into two fully connected sub-areas. More specifically nodes 0, 1, 2, 3, 4, 5, 6, 7 and nodes 8, 9, 10, 11, 12, 13, 14, 15 are within transmission range of each other, respectively. Nodes 7 and 8 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=Min{8^2+1, 16}=16.

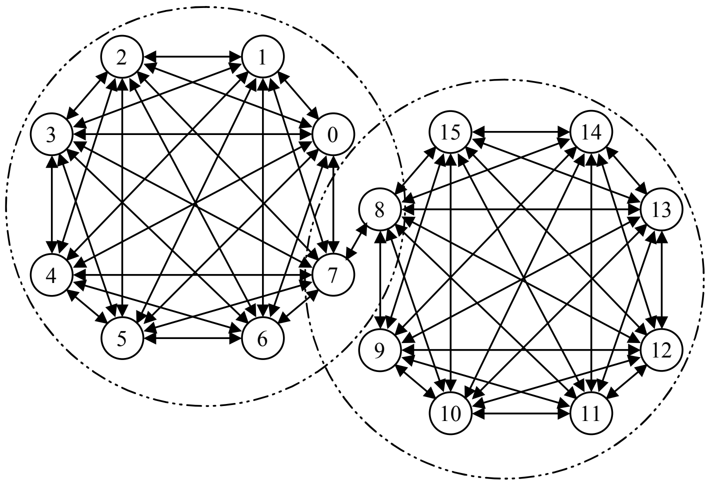

Figure 5-4: A 16-node, two-area network topology

5.2.1 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 7 is a receiver and node 8 is a transmitter (or vice-versa) at the same time.

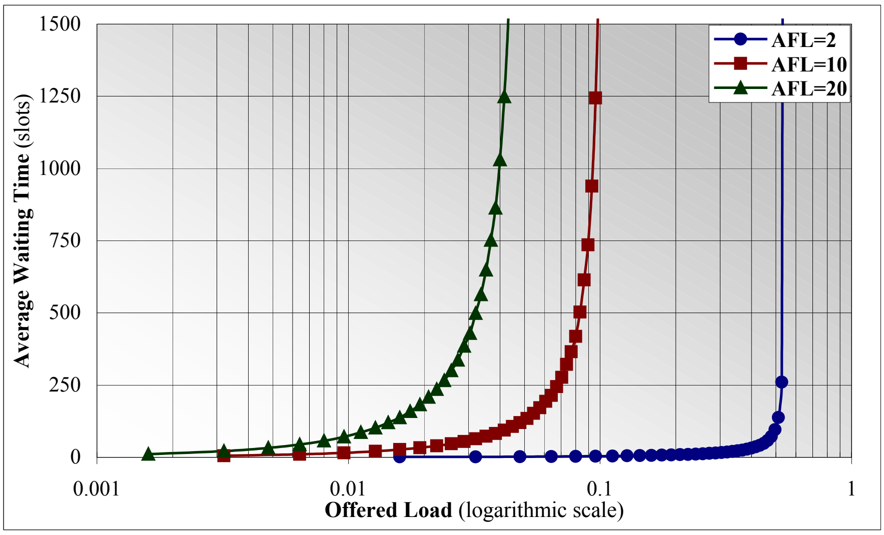

Figure 5-5: Unicast average waiting time for 16-node, two-area network.

Figure 5-5 shows the average waiting time versus offered load. Comparing average waiting time with that for the fully connected network (Figure 5-2), we see that for the same offered load, the average waiting time is decreased (see Table 5-2) and that the maximum supported offered load is increased (see Table 5-3), at the two-area network case. As with the 8-node network topologies, this was expected due to the spatial channel reuse in the two-area network case.

Figure 5-6 shows the system throughput versus offered load. Higher system throughput is achieved compared to the fully connected network (especially for small messages), due to higher spatial reuse of the communication channel.

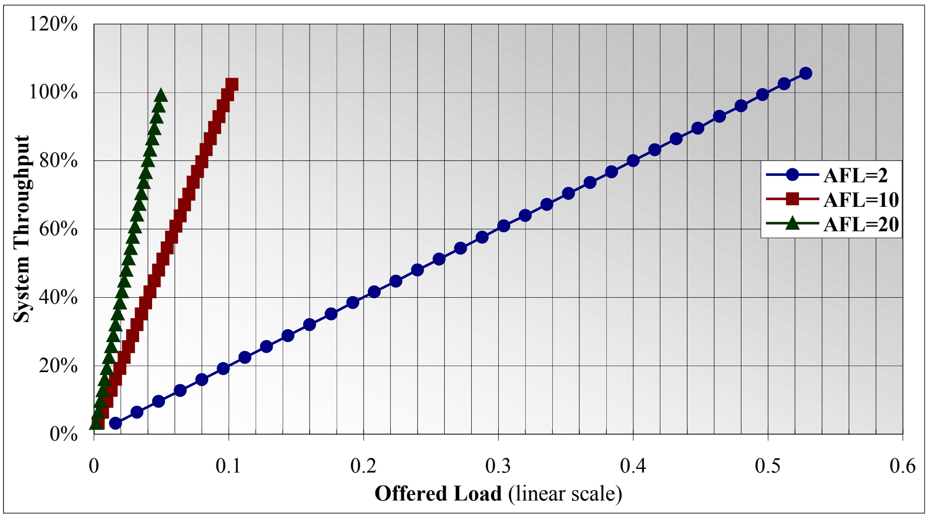

Figure 5-6: Unicast system throughput for 16-node, two-area connected network.

	Fully connected network			Two-area network		
AFL	2	10	20	2	10	20
Offered Load	0.208	0.08	0.0416	0.208	0.08	0.0416
Average Message Delay (slots)	35.1	826.2	1839.4	26.1	563.4	1553.5
Average Waiting Time (slots)	18.1	681.2	1534.9	9.1	418.9	1248.3
Channel Utilization	41.6%	79.8%	83.0%	37.5%	64.3%	66.4%
System Throughput	41.6%	79.8%	83.0%	41.6%	79.7%	83.1%

Table 5-2: Unicast comparison between fully connected and two-area network.

AFL	2	10	20
Offered Load	0.528	0.1024	0.0496
Average Message Delay (slots)	277.62	3041.79	7249.27
Average Waiting Time (slots)	260.61	2896.71	6944.62
Channel Utilization	78.239%	76.813%	75.183%
System Throughput	105.549%	102.328%	99.124%

Table 5-3: Unicast maximum metric values for 16-node, two-area connected network.

5.2.2 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 7 and 8 are not the transmitting nodes.

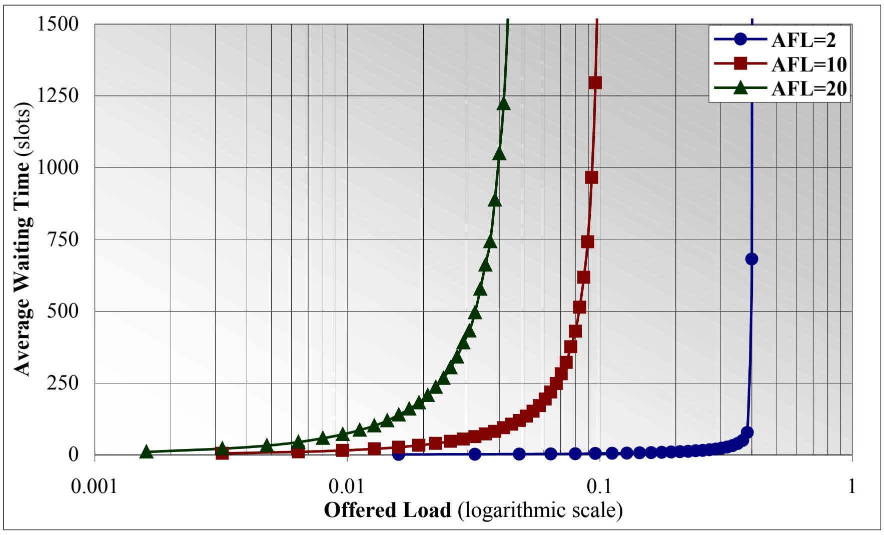

Figure 5-7: Broadcast average waiting time for 16-node, two-area network.

Once again average waiting time is decreased (see Figure 5-7) and the maximum supported offered load is increased (see Table 5-1 and Table 5-5) compared with that for the corresponding fully connected network case, due to the spatial channel reuse in the two-area network case.

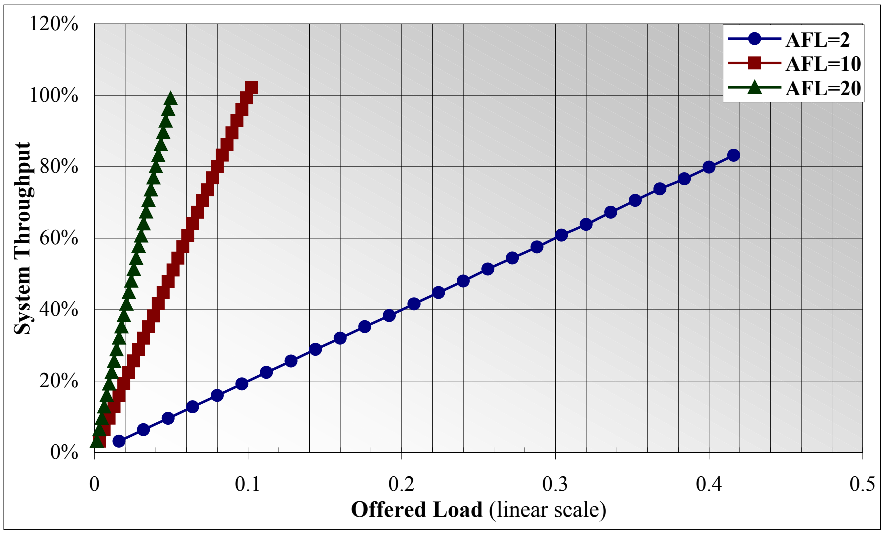

Figure 5-8: Broadcast system throughput for 16-node, two-area connected network.

	Fully connected network			Two-area network		
AFL	2	10	20	2	10	20
Offered Load	0.208	0.08	0.0416	0.208	0.08	0.0416
Average Message Delay (slots)	35.1	826.2	1839.4	26.9	575.6	1527.7
Average Waiting Time (slots)	18.1	681.2	1534.9	10.0	430.6	1223.0
Channel Utilization	41.6%	79.8%	83.0%	38.0%	66.3%	68.3%
System Throughput	41.6%	79.8%	83.0%	41.6%	80.0%	83.1%

Table 5-4: Broadcast comparison between fully connected and two-area network.

AFL	2	10	20
Offered Load	0.416	0.1024	0.0496
Average Message Delay (slots)	698.98	3725.66	7236.51
Average Waiting Time (slots)	681.97	3580.78	6931.73
Channel Utilization	65.842%	79.01%	77.563%
System Throughput	83.187%	102.123%	99.105%

Table 5-5: Broadcast maximum metric values for 16-node, two-area connected network

Like in the 8-node, two-area network, nodes 7 and 8 that connect the two areas are the reason for the difference between unicast transmissions and broadcast transmissions (see Table 5-3 and Table 5-5). 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 the same maximum offered load. Broadcast transmission in a particular slot from node 7 or 8 requires that the remaining fifteen 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 7 or 8 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 eight nodes in this case). Thus in broadcast transmissions, nodes 7 and 8 have almost twice as many competing nodes than in unicast transmissions.

5.3 SIXTEEN-AREA NETWORK TOPOLOGY

In this topology, nodes are connected as shown in Figure 5-9. Each node has three neighbors (maximum node degree is d=3) and defines its own transmission area. Thus the frame length, L, used for this network topology is $L=d^2+1=10$. This topology is selected to demonstrate the effect of smaller frame length compared to the previously examined 16-node topologies.

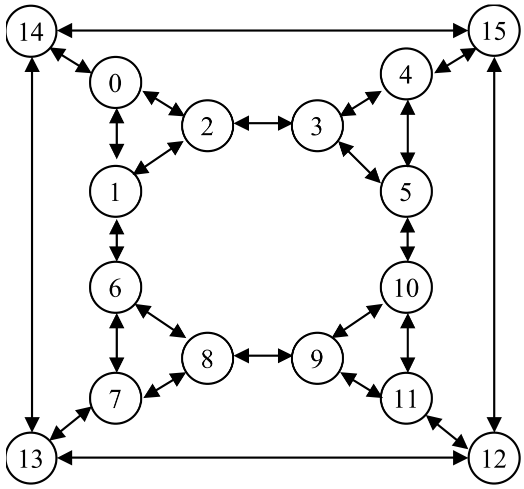

Figure 5-9: A 16-node, sixteen-area network topology

5.3.1 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 5-10 shows the average waiting time versus offered load. Average waiting time, for the same offered load, has been significantly reduced (see Table 5-6). This is caused not only by the higher spatial reuse but also by the smaller frame size compared to the previous network topologies. Remember that one of the parameters that affect the message waiting time is the service time of its predecessor message. In this case if for example a message has size of AFL=20 packets, it has service time of (AFL-1)*L+1 = (20-1)*10+1 = 191 slots, while in the previous 16-node networks, the same message had service time of (20-1)*16+1 = 305 slots. Thus service time of a 20 packet message has been reduced by 37.4%.

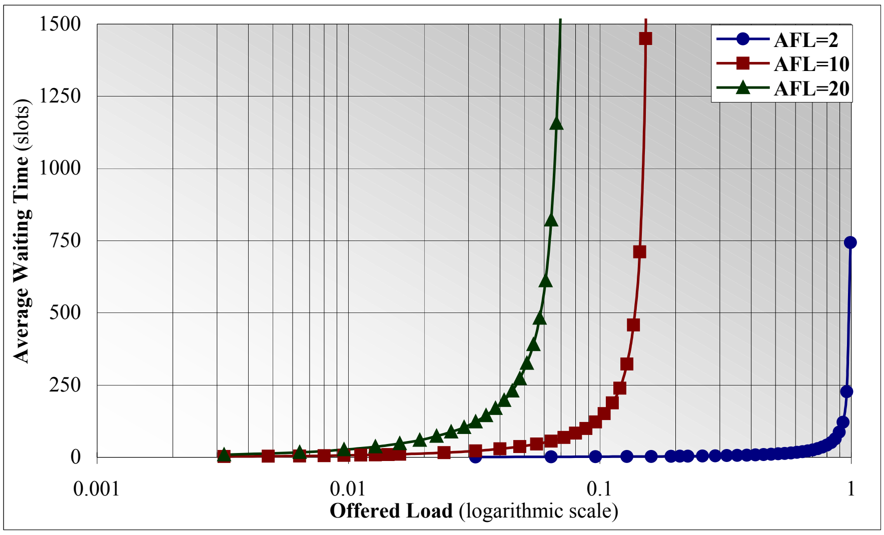

Figure 5-10: Unicast average waiting time for 16-node, sixteen-area network.

	Fully connected network			Two-area network			Sixteen-area network		
AFL	2	10	20	2	10	20	2	10	20
Offered Load	0.208	0.08	0.0416	0.208	0.08	0.0416	0.208	0.08	0.0416
Average Message Delay	35.1	826.2	1839.4	26.1	563.4	1553.5	14.8	174.8	388.0
Average Waiting Time	18.1	681.2	1534.9	9.1	418.9	1248.3	3.8	83.7	197.2
Channel Utilization	41.6%	79.8%	83.0%	37.5%	64.3%	66.4%	36.1%	60.5%	62.1%
System Throughput	41.6%	79.8%	83.0%	41.6%	79.7%	83.1%	41.6%	80.0%	83.2%

Table 5-6: Unicast comparison between fully connected, two-area and sixteen-area network.

Figure 5-11 shows the system throughput versus offered load. Higher channel slot utilization and higher throughput than previous area-networks is achieved due to higher spatial reuse of the communication channel.

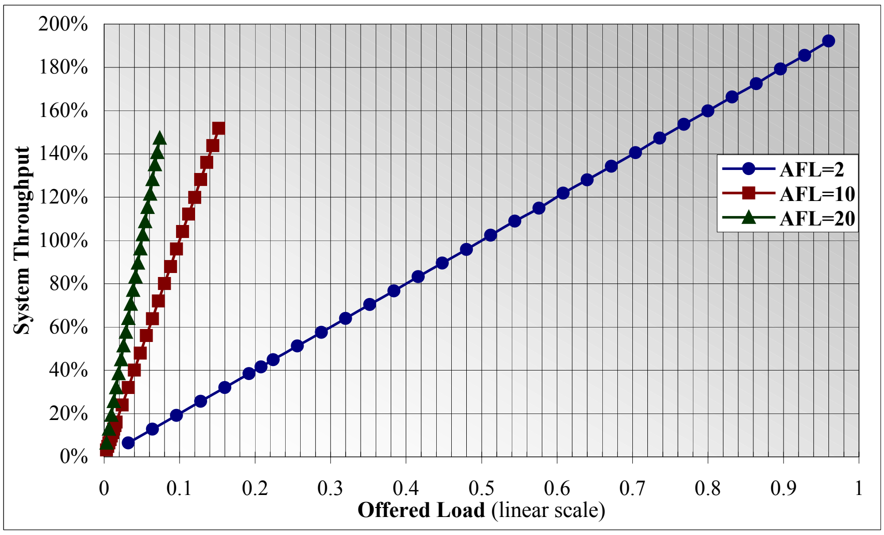

Figure 5-11: Unicast system throughput for 16-node, sixteen-area connected network.

Table 5-7 shows the maximum offered load and the maximum channel slot utilization and throughput supported by this network. As explained in section 4.2.1 , 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 60% for all AFL values. For AFL=2, this percentage reaches 100% which means that we would need twice as many slots to deliver all messages to their destination nodes, if spatial reuse was not allowed and only one data packet was transmitted per slot.

AFL	2	10	20
Offered Load	0.96	0.1520	0.0736
Average Message Delay (slots)	238.22	1540.60	3666.32
Average Waiting Time (slots)	227.21	1449.68	3475.29
Channel Utilization	94.537%	87.279%	86.144%
System Throughput	192.177%	151.813%	147.289%

Table 5-7: Unicast maximum metric values for 16-node, sixteen-area connected network.

5.3.2 Broadcast transmission requests

Spatial reuse of the communication channel is possible in the sixteen-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.

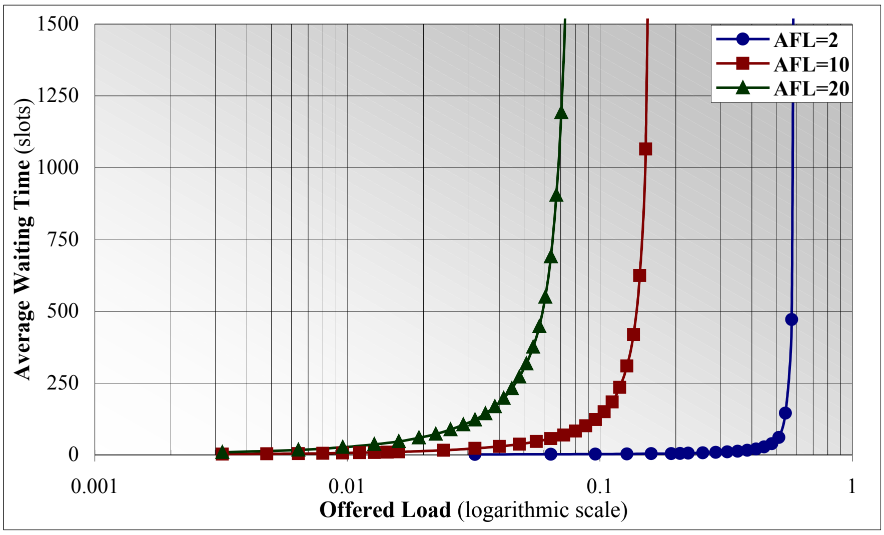

Figure 5-12: Broadcast average waiting time for 16-node, sixteen-area network.

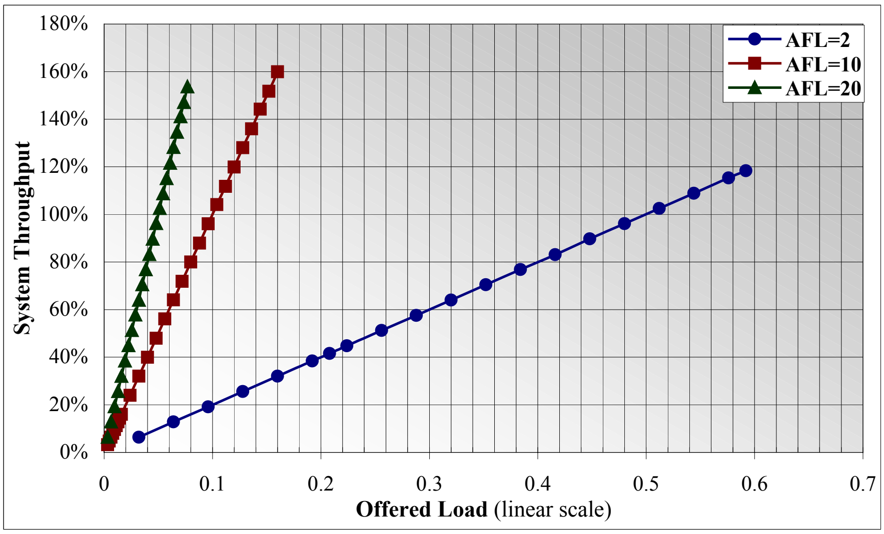

Figure 5-13: Broadcast system throughput for 16-node, sixteen-area connected network.

	Fully connected network			Two-area network			Sixteen-area network		
AFL	2	10	20	2	10	20	2	10	20
Offered Load	0.208	0.08	0.0416	0.208	0.08	0.0416	0.208	0.08	0.0416
Average Message Delay	35.1	826.2	1839.4	26.9	575.6	1527.7	15.7	173.8	388.0
Average Waiting Time	18.1	681.2	1534.9	10.0	430.6	1223.0	4.7	82.9	197.0
Channel Utilization	41.6%	79.8%	83.0%	38.0%	66.3%	68.3%	37.1%	62.9%	64.8%
System Throughput	41.6%	79.8%	83.0%	41.6%	80.0%	83.1%	41.6%	80.0%	83.1%

Table 5-8: Broadcast comparison between fully connected, two-area and sixteen-area network.

AFL	2	10	20
Offered Load	0.592	0.16	0.0784
Average Message Delay (slots)	481.96	3127.81	4699.44
Average Waiting Time (slots)	470.95	3036.92	4508.63
Channel Utilization	77.083%	92.296%	91.822%
System Throughput	118.359%	159.903%	156.387%

Table 5-9: Broadcast maximum metric values for 16-node, sixteen-area connected network

Lower message delays and waiting times (see Table 5-8) and higher channel utilization and system throughput (see Table 5-1, Table 5-5 and Table 5-9) are once again achieved, compared to previous 16-node network topologies, due to higher spatial reuse and smaller frame size. In addition, as in the 8-node eight-area network case, broadcast transmission performs poorer for small messages and better for large messages, compared to unicast transmissions, for reasons already explained in section 4.3.2

CHAPTER 6

BACKOFF ALGORITHM EVALUATION

In this chapter 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 4-1 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[19] 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.

6.1 SLOTTED ALOHA BACKOFF ALGORITM

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.

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

- 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 6-1 shows the average waiting time versus offered load.

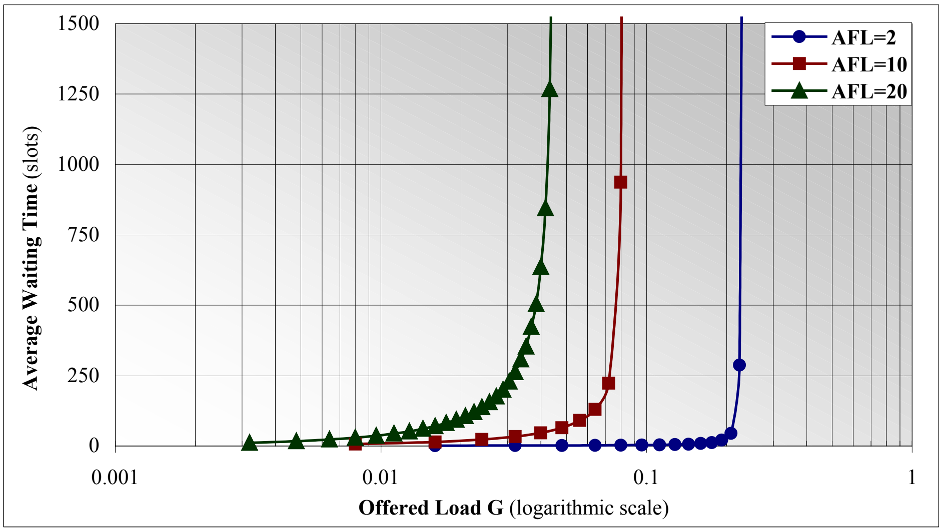

Figure 6-1: Slotted aloha average waiting time for 8-node, fully connected network.

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 6-2 shows the maximum offered load, average message delay and average waiting time values, for which the system is stable. Table 6-1 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 4.2.1 large 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 (data) are delay sensitive as in audio and video applications.

Figure 6-2 shows the channel utilization of the system versus offered load

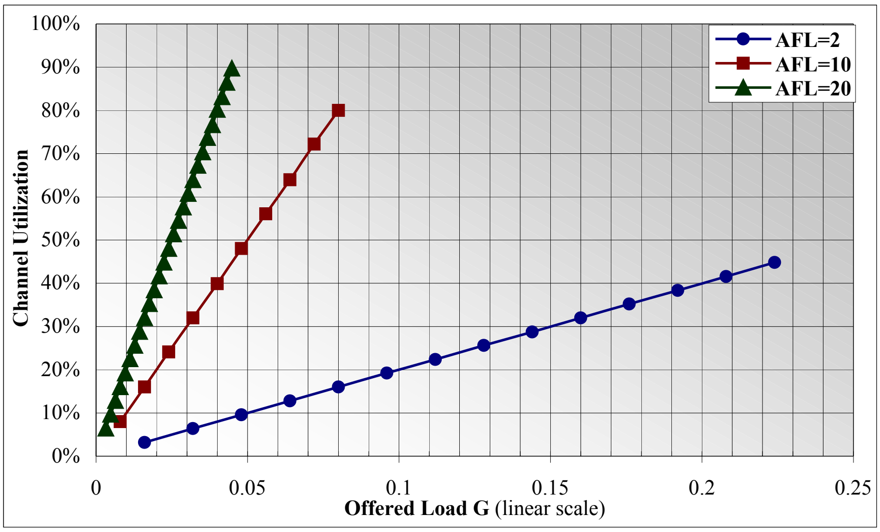

Figure 6-2: Slotted aloha channel utilization for 8-node, fully connected network.

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 6-1: Comparison between ABA and slotted aloha 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 6-2: Slotted aloha maximum metric values for 8-node, fully connected network

CHAPTER 7

CONCLUSIONS – FUTURE WORK

7.1 SUMMARY

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.

In the first part of this work, we gave the definition of an ad-hoc network, some examples of its applications together with its benefits and drawbacks. In the second part, 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 presented protocols which attempt to solve these problems. 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.

7.2 FUTURE WORK

Several design issues in order to further improve the performance of CATA, arise from this work. Some of them are:

- Node reservation scheduling: As we stated in section 4.2.1 , nodes reserve the first available slot in which they do not experience a collision in order to successfully complete their intended transmission. This sometimes results in slot underutilization that decreases the overall system performance. In addition the equation $Min\{d^2+1,N\}$gives the worst case number of slots in a frame that a network topology needs in order to provide collision free slots to all network nodes. A node reservation, rescheduling algorithm that takes into account the special topology characteristics of a network can decrease the number of slots in a frame thus decreasing message delays and increasing channel utilization. Real time dynamic frame resizing might prove a good idea that can improve the performance of CATA.
- In our simulation experiments we considered that data are not delay sensitive and data packets are not "dropped-out" from queues, after a certain time period, which is usually the case in real time applications such as those involving video and audio. In order for CATA to efficiently support the demands of real time applications (e.g. message delays, allowed packet drop-out percentage, data transfer rates, etc.), performance and design issues must be further examined. Some designing issues that can be taken into account are the creation of priority queues in order to reduce real time application message delays, and finding optimal slot sizes in order to increase utilization and decrease start-up delays.

Finally, a comparison between CATA with ABA backoff algorithm and IEEE802.11 in terms of the performance metrics presented in this thesis should be made. Although it is hard to tell because IEEE802.11 is not a slot reservation mechanism but rather a full channel reservation mechanism (explained in section 2.2.6), our intuition tells us that

IEEE802.11 should perform worse. First, its backoff algorithm is very similar to that of Slotted Aloha and second, from the specifications of the channel access and collision resolution mechanism, it is possible that some nodes monopolize the channel at the expense of other nodes. In addition, due to the full channel reservation mechanism, the nodes of the network are not treated fairly especially when messages are large, as some nodes will reserve the channel for long periods and will experience small message delays while other nodes will suffer very long message delays.

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.