DEMAND ASSIGNED CHANNEL ALLOCATION APPLIED TO FULL DUPLEX UNDERWATER ACOUSTIC NETWORKING

John Gibson, Geoffrey G. Xie, and Andy Kaminski

Naval Postgraduate School, Department of Computer Science
Monterey, CA, U.S.A.
jhgibson@nps.navy.mil

ABSTRACT

Acoustic communications provide a viable means for underwater networking. However, extreme propagation delays, limited bandwidth, and half duplex communications, with its inherent use of delay inducing collision avoidance media access and "stop-and-wait" flow control, severely limit the throughput and power of such networks. While full duplex communications eliminate access coordination and enable more effective flow control, they impact transmission time and may lead to wasted channel capacity. The authors hold that, by combining demand assigned multiple access techniques with bandwidth on demand allocations, the cost of full duplex, in terms of latency, can be significantly reduced. This paper makes two contributions. First, it formally evaluates the limitations imposed on delay-constrained networks by adherence to stop-and-wait methods. Second, it demonstrates by simulation the potential to reduce message latency by inverse multiplexing, using the aforementioned techniques, for delay-challenged networks incorporating full duplex communications. These levels of latency improvement, observed through simulation, provide a lower bound for performance in delay challenged networks that employ bandwidth-on-demand and sliding-window techniques under similar traffic load parameters.

1. Introduction

The growth in remote maritime sensing and the employment of autonomous underwater vehicles has spurred the development and implementation of wireless underwater networks. Unlike traditional wireless networks, underwater networks use acoustic signals to carry data rather than electromagnetic signals. One measure of the utility of a network is its power. As defined by Jain [Stallings], the power of a network is the ratio of its achieved throughput to the total traffic delay. This metric provides a means for comparing the relative performance of various designs and implementations. Unfortunately, the underwater acoustic environment induces severe limitations on the achievable power of underwater acoustic networks (UANs). Two of the key limiting factors are the severe propagation delays, associated with the nominal 1500 meters per second propagation rate of the acoustic signal in sea water, and constrained bandwidth, due to extreme attenuation of frequencies above 50 Kilohertz over any appreciable distance. Further, the omni-directional nature of the medium leads to the same hidden terminal, exposed-station, and near-far problems associated with traditional (radio-based) wireless communications. To address these latter issues, UANs typically implement a form of collision avoidance access control to coordinate the transmissions of multiple users. Such techniques employ the exchange of "handshake" messages which grant the use of the medium to a single user while informing other

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JUN 2005	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Demand Assigned Channel Allocation Applied to Full Duplex Underwater Acoustic Networking	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Postgraduate School Department of Computer Science Monterey, CA 93943**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
Recent Advances in Marine Science & Technology, Journal of PACON International, pp. 81-95, June 2005, The original document contains color images.

14. ABSTRACT

Acoustic communications provide a viable means for underwater networking. However, extreme propagation delays, limited bandwidth, and half duplex communications, with its inherent use of delay inducing collision avoidance media access and stop-and-wait flow control, severely limit the throughput and power of such networks. While full duplex communications eliminate access coordination and enable more effective flow control, they impact transmission time and may lead to wasted channel capacity. The authors hold that, by combining demand assigned multiple access techniques with bandwidth on demand allocations, the cost of full duplex, in terms of latency, can be significantly reduced. This paper makes two contributions. First, it formally evaluates the limitations imposed on delay-constrained networks by adherence to stop-and-wait methods. Second, it demonstrates by simulation the potential to reduce message latency by inverse multiplexing, using the aforementioned techniques, for delay-challenged networks incorporating full duplex communications. These levels of latency improvement, observed through simulation, provide a lower bound for performance in delay challenged networks that employ bandwidth-on-demand and sliding-window techniques under similar traffic load parameters.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**SAR**	**19**	

users within range of the intended recipient of the pending transmission. This coordination is necessary to limit retransmissions due to frame collisions (overlapping data receptions) at the destination. While the exchange of small coordination messages does not induce severe delay penalties in a radio wireless environment, the five-orders of magnitude slower propagation rate of the acoustic medium induces significant delay costs, especially for traffic relayed through several intermediary nodes before reaching the intended destination [Xie]. As long as all hosts share a common channel, the wireless nature of the network limits access controls to collision avoidance types of methods and stop-and-wait flow control mechanisms.

We observed that if the available bandwidth is channelized and nodes are assigned unique channels within their respective two-hop neighborhoods then the need for collision avoiding coordination prior to message transmission is eliminated as each node is effectively operating over point-to-point links with its neighbors [Xie; Gibson et al]. The resulting full duplex communications also allow for more efficient flow control mechanisms, such as sliding-window based methods. Fundamental to this observation is the assumption that nodes are channel agile. That is, they can discriminate between multiple channels and receive signals simultaneously on any channel over which they are not transmitting and they can transmit over any channel on which they are not receiving signals. Channel agility is a necessary condition for full duplex communications. As we demonstrated, achieving such agility for underwater nodes by equipping each node with two transducers, one for transmission and the other for reception, and using a rudimentary CDMA scheme [Bektas], is not unreasonable. Additionally, it is assumed that the channel space is bidirectional and reflexive, such that if a node receives signals from another node then the latter can also receive signals from the former.

While the fundamental goal of implementing full-duplex connections for UANs and other propagation delay constrained networks is to mitigate the compounding of propagation delays inherent in establishing half-duplex based sessions, the resulting division of the available bandwidth into sufficient channels for full-duplex modality impacts the available data rate for applications using the network. The allocation of two unidirectional channels for each connection may result in very low bandwidth efficiency, unless the traffic from each node is both regular and constant. However, in most data communications exchanges, the data is irregular and bursty, resulting in periods where allocated bandwidth is unused. In bandwidth constrained environments, under utilized allocations among a collection of neighboring nodes may inhibit near-real-time transmission of information submitted by network applications for delivery, even though collectively the unused capacity of the individual links could have supported the requirement. Further, where bandwidth is severely limited, reducing the effective capacity of a link, by dividing the total capacity into discrete channels results in increased transmission delays, which may counter any gains made in reducing the compounding of propagation delays by eliminating transmission coordination. Tanenbaum asserts that, barring collision impacts resulting in retransmissions, if multiple hosts share a common medium and present bursty, non constant traffic loads, dividing the capacity into equal shares for each host will result in increased delays, due to the wasted capacity when hosts are idle [Tanenbaum]. Attention must be given, then, to recovering the bandwidth which is unallocated within a given node's neighborhood or which, though allocated to one of the nodes within the neighborhood, is not required by the node to which it is assigned and may be offered to any neighbor requiring increased capacity. This

unused capacity may be allocated on demand to any node within the neighborhood which has traffic queued requiring higher capacity throughput relative to that node's baseline allocation.

Proven techniques in satellite communications capacity management may hold keys to address the bandwidth inefficiency side effect of the channelization necessary for full-duplex communications. Specific among them are Demand Assigned Multiple Access (DAMA) controls and Bandwidth-on-Demand (BoD) techniques. DAMA allocates channels to users when the users request an allocation. These channels are typically fixed in size. Alternately, BoD provides users a variable sized allocation depending on the request of the individual user. By allocating multiple channels to a user on demand, these channels may be used to inverse multiplex two or more message frames, thus providing a coarse version of BoD. It is this coarse BoD implemented via DAMA that this paper suggests for use in full duplex acoustic networks and refers to as DAMA with regard to the analysis and simulation. In consideration of this potential, this paper provides two key contributions. First, it formally establishes, by way of utilization analysis, the inherent performance limitations imposed on delay-constrained networks by adherence to half-duplex communications, specifically with respect to stop-and-wait flow control mechanisms. This limitation is especially sensitive to frame size. Without implementing full duplex communications reliable data transfer mechanisms are limited to variants of the stop-and-wait protocol. However, implementation of full duplex links to improve message latency by eliminating media access delays must consider the potential performance penalty imposed by the required capacity channelization and whether that penalty can effectively be reduced. This leads to the second contribution, as the paper reports simulation results that demonstrate the utility of BoD enabled inverse multiplexing techniques to mitigate that induced performance penalty. The simulation modeled two very different channel allocation schemes: a simple *first-come-first-served* (FCFS) method, and a *fair distribution* (FD) method which allocated resources to sources proportional to the remaining size of the message being sent. Interestingly, the FCFS method consistently resulted in better average latency performance than the fair allocation method.

The rest of the paper is organized as follows. Section 2 considers the implementation of channel capacity conservation schemes, such as DAMA and BoD, employed in satellite systems and cellular telephone networks. It also contrasts the motivation for the implementation of such schemes with that for their proposed implementation in the underwater acoustic environment. Section 3 provides an analysis of the impact of propagation delays on a traditional half-duplex flow control scheme, the stop-and-wait protocol. Specifically, it addresses the relationship between propagation delay, frame size, transmission rate, and resource utilization providing insight into the usefulness of increasing the number of dynamic channels with respect to the expected message size. Section 4 provides an overview of our recent experimentation, through simulation, with the utility of DAMA applied to propagation delay constrained network environments employing a stop-and-wait protocol. Finally, Section 5 provides conclusions drawn from the initial experimental results and suggests areas for further study, analysis, and direction.

2. Related Work

Both satellite and cellular systems are concerned with capacity conservation for one principal reason: resource use efficiency. From the service provider's perspective this translates to servicing more customers with less capacity, while from the customer's perspective this means

only having to pay for the capacity actually used. Prior to the implementation of DAMA technologies to military satellite systems, capacity was allocated to various customers; either dedicated point-to-point links or shared "nets," where multiple users accessed the allocated channel in a "party-line" type of connection. In either case, the number of available channels was quickly outpaced by the demand for support. Further, the offered load was often far less than the capacity of the channel which had been dedicated for that load. For example, the typical use of dedicated channels for the 22nd Signal Brigade was 15 percent in 1999 [Stratman]. Implementing DAMA enables otherwise idle link time to be allocated to users on demand. The following algorithm provides insight into how allocations are made:

User requests channel allocation from controller over order-wire
If (channel available) then
    Controller returns allocation to requester
    Requester sends information over allocated channel
    Requester returns channel to controller
else controller advises requester of non availability of service

Allocation may be either for a predetermined, fixed capacity or may be tailored to the user's traffic characteristics, as described by Barda [Barda]. Using Bandwidth on Demand (BoD), an allocation based on traffic characteristics provides resources according to whether or not the traffic is predictable with respect to resources required. The former include database transactions and voice connections, both providing a predicable fixed size packet. However, transaction communications are better suited for contention-based (shared) media due to their burst nature and short frame size, while voice communications are better served by guaranteed resource services due to their constant bit rate requirements. Such resource guarantees may be provided by a DAMA controlled service, where once the allocated resources dedicated to the voice connection are no longer needed, they can be returned to a resource allocation pool to be used for later customer requests. Traffic whose patterns are less predicable, such as IP hosted traffic, which includes web surfing, electronic mail or various other computer applications, are best served by an allocation scheme which may provide resources specific to the traffic being sent and recovered when the transfer is complete so that they may be allocated to later requirements [Barda]. According to Barda's calculations, traffic patterns which require a constant data rate, such as voice communications, when serviced by DAMA constructs require as little as 20% of the resources necessary for the same level of service from a contention-based network. For traffic which varies appreciably in both size and frequency, Barda suggests BoD is the superior allocation scheme where his results imply a resource requirement of only 8% of that required for contention networks. Moreover, use of fixed size allocations such as those in DAMA, while still reducing the resources required over Aloha style protocols by 80%, require more than twice the amount of bandwidth as BoD protocols. It should be noted that much of the IP hosted traffic may not have the same near-real-time delay constraints that voice communications hold, thus they may be served by a reduced transmission rate resulting in longer delivery delays without adversely impacting their utility. These results are significant when considering the allocation of severely limited resources such as those available for underwater acoustic networks.

Requests for allocation, either in DAMA or BoD managed networks, are made over a control channel. This channel may be either contention-based or contention-less, where the channel is

divided into discrete sub-channels and each host assigned a sub-channel. Barring contention, a request for satellite channel access and an assigned allocation in response to the request may be exchanged in approximately 250 ms, the round trip propagation delay. This assumes the transmit time of the request and response are negligible, as well as any queuing delays encountered. The same is not the case for acoustic networks, where the round trip delay over a 1 km link is just over 1300 ms, more than five times as long as that experienced by geosynchronous satellite links.

While the incentive for using DAMA or BoD techniques for satellite service is to allow a larger user population to access a limited number of channels than would be possible using dedicated channel allocations, the motivation with respect to underwater acoustic networking, as proposed, is to provide a means of mitigating the cost, as measured in achievable throughput, associated with dividing the available capacity into discrete channels in order to support full duplex communications. The impetus for establishing full duplex communications is to eliminate the need for medium access controls. As noted, the channelization necessary to support full duplex communications may result in wasted capacity if supported traffic is bursty and the channels created are dedicated to specific nodes. To minimize this impact, we propose assigning a dedicated channel to each network node, sized to the modal message size and pooling other channels to be assigned on demand should nodes have larger messages to send. Thus, the use of DAMA for UANs would be to allocate a larger pool of channels to a smaller user set thereby enabling a coarse BoD service. A fundamental trade-off to be considered is whether to establish a large pool of small channels or a smaller pool of large channels, recognizing that in the realm of underwater acoustic channels, large is a relative term. This paper provides simulation results that help to quantify this trade-off.

An additional benefit of full duplex communications is the ability to implement the more efficient sliding-window flow control methods. Flow control methods are essential for providing reliable connectivity throughout the network, as well as to improve the efficiency of data transfer over propagation delay constrained links. Since the stop-and-wait protocol represents a worst-case performance baseline for sliding-window protocols, as a stop-and-wait protocol is essentially a sliding-window of size one, it provides a reasonable point of departure for assessing the benefit to be gained by implementing DAMA-like allocations for acoustic communications. The following section explores the impact of propagation delay on stop-and-wait protocols.

3. Performance of Stop-and-Wait Protocols over Delay Constrained Networks

In order to provide a baseline by which to evaluate the benefit of sliding-window protocols for delay and bandwidth constrained networks it is useful to understand the performance of stop-and-wait protocols over the same networks. It is also worth noting that by default, half duplex connections can only support stop-and-wait protocols as sliding-window protocols with a window size greater than one must have a full duplex link to fully function. Without a full duplex link the sliding-window functionality simply serves to segment a large frame into smaller fragments allowing improved error recovery; however, the source must relinquish the link in order for the destination to acknowledge traffic receipt, thereby following the stop-and-wait paradigm. Following is an analysis of the relationship between the number of channels available to a transmitting node, the number of frames remaining to be sent, the frame size, and the number of stop-and-wait cycles necessary to complete message delivery.

Suppose n child nodes compete for transmission to a single parent node. Let the total transmission capacity be r bits/second and assume a channelization scheme, such as FDMA or CDMA, is used to divide the total available capacity into M channels, where $M > n$. The surplus channels are held in reserve, i.e. pooled, for on-demand allocation. To manage the analysis complexity, assume the child nodes send data frames of a fixed size (F bits) and that the size of the acknowledgement (ACK) messages is fixed at B bits. The ACK message may also serve a dual role by incorporating the allocation control of pooled channel capacity. In this way, the ACK notifies a child node of additional channels allocated to it based on the child's reported traffic load [Gibson et al].

Consider a child node sending a message of size $k * F$ bits, where k is a positive integer. Assuming no retransmission is required, the total delay, from the start of transmission to acknowledgment delivery, is $(C+1) * T$, where T is the time necessary to transmit $(F + B)$ bits plus the round-trip propagation delay and C is the number of transmission cycles required for frames other than the first one. Thus,

$$T = \frac{(F+B)}{(r/M)} + 2\frac{d}{v} = T_x + 2T_p \tag{1}$$

where d is the distance, in meters, between the two nodes, v is the propagation velocity of the medium, in meters per second, T_x is the transmission time for the combined frame and ACK bits, and T_p is the one-way propagation delay.

For radio-borne networks, we have $T_x >> 2T_p$; therefore, $delay \approx (C+1) * T_x$. However, for UANs, we have that $T_x << 2T_p$ for most (r, d) values and small frame sizes. Therefore, $delay \approx (C+1) * 2T_p$. Increasing the frame size, to compensate for propagation delays, may result in increased retransmission loads as the likelihood of frames arriving with uncorrectable errors increases with the frame size.

To assess the steady state performance, assume the same message arrival pattern for each child node and an average message size of $k * F$. Clearly, the average inter-message arrival time at a child node must be less than the delay for the system to be stable; otherwise, the delay would increase without bound due to queuing delays at the source. On average, a node's spare channel request comes when e spare channels are already allocated to other nodes. Therefore, assuming no uncorrectable bit errors requiring frame retransmissions, the average number of additional transmission cycles required for a message is:

$$C = \left\lceil \frac{(k-1)}{(M-n-e+1)} \right\rceil \tag{2}$$

It is clear, then, that to minimize the total delay the number of transmission cycles must be reduced, assuming that the propagation delay remains constant. Such a reduction may be accomplished either by increasing the frame size or transmitting frames in parallel, thus implementing a form of inverse multiplexing. Increasing the frame size increases the duration of

the transmission time for a given transmission rate. It is also clear from Equation 2 that increasing the number of dynamic channels established such that the total capacity available to a node exceeds the number of frames to be transmitted does not further reduce the transmission cycles required. Rather such an action would simply increase the transmission time for each cycle, as noted, thereby adversely impacting the overall latency. Thus, the size of the dynamic channel pool should be adapted to the expected size of multiple frame messages. Note that if the message arrival distribution is known, the exact value of e may be derived based on the conditional probabilities:

Prob{0 spare channels allocated | new request},
Prob{1 spare channel allocated | new request},
etc.

When transmission delays dominate the network, as in the case with radio-based networks above, the total delay is minimized when the number of additional transmission cycles required is zero. Since T_x is proportional to M, M should be as small as possible, subject to the constraint $(M - n - e) \geq k - 1$, where n is the minimum number of channels required for full duplex operation. This result agrees with the observation by Tanenbaum that dividing the total capacity into smaller portions to support simultaneous access increases total delay, given intermittent, non constant traffic loads. Where access controls are not an issue, this result further supports a shared medium vice full duplex, especially in bandwidth constrained environments.

When propagation delays dominate the network, as is the case with acoustic-based networks, the total delay doesn't change appreciably if the number of remaining frames to be sent is small compared to the number of available channels [$(k-1)/(M-n-e+1) < 1$]. This situation occurs if the average message size is very small compared to the frame size, negating the need for multiple frames; or the number of available channels greatly exceeds the number of frames to be sent. Performance improves, that is the total delay decreases, as the number of channels allocated approaches the required transmission cycles and drops off once it exceeds the number of frames to be sent. In other words, the performance does not improve appreciably as the number of spare channels exceeds a certain threshold and that threshold depends on the average message size. Early simulation results, discussed in Section 4, indicate the performance is asymptotic.

Unless frame size is very large, the latter case is applicable to the acoustic environment. In shared access systems, increasing the frame size to reduce the dominating impact of propagation delays on message delivery results in extreme access delays, as nodes may have to wait for other nodes to complete their transmission prior to gaining access. In this way, access delays are further exacerbated as traffic loads increase. The impact of frame size on performance can also be appreciated when considering the flow control mechanism used to ensure message delivery and prevent the source from overloading the destination's capacity to receive data. For half-duplex systems, the only flow control mechanisms available are derived from the basic stop-and-wait protocol, which requires the source to wait for an acknowledgment for the last frame transmitted before sending the next frame. Thus, only one unacknowledged frame may be outstanding. As this is the norm for acoustic networks, this will form the baseline for assessing the delivery delay mitigation of a network which implements a DAMA capacity recovery

mechanism. Thus, an understanding of the impact of the relationship between frame size and propagation delay on the stop-and-wait protocol is useful.

Define the ratio, $a = T_p/T_x$, to be the length of the medium measured in frames. $T_x = F/r$, where F is the frame size and r the transmission rate. $T_p = d/v$, where d is the distance between communicating hosts and v is the signal propagation rate. Rewriting the formula for a, we have

$$a = d(r/vF). \tag{3}$$

Note that the ratio, a, is unit-less, as expected. Given a particular d, a is proportional to $(r/(vF))$. It is illustrative to compare the respective values of a for radio and acoustic wireless networks, using their typical data rates as constrained by available bandwidth and system noise.

Assume the frame sizes for the two types of networks are F_r and F_a bits, respectively. For radio-based networks, given r_r = 1000 to 10000 kbps, v_r = 300000 km/sec, then

$$a_r = dr_r/(v_r F_r) \rightarrow d/300F_r \le a_r \le d/30F_r \ . \tag{4}$$

For acoustic underwater networks, given r_a = 100 to 1000 bits/sec, v_a = 1500 m/sec, then

$$a_a = dr_a/(v_a F_a) \rightarrow d/15F_a \le a_a \le 2d/3F_a \ . \tag{5}$$

For the two types of networks to have the same value for a, $F_a = 20 * F_r$. That is, the frame size for acoustic networks must be twenty times that of radio networks in order for the lengths of the respective mediums, in frames, to be equal, given the differences in transmission rates and propagation rates as noted by the subscripts. Moreover, the larger the disparity in transmission rates, the greater the scaling factor for the frame sizes.

The utilization of the network capacity for a stop-and-wait protocol can be derived based on the same ratio, as determined by the following equations [Stallings]:

$$U = 1/(1+2a), \text{ given error free reception, and} \tag{6}$$
$$U = (1-p)/(1+2a) \tag{7}$$

where p is the probability a frame contains an unrecoverable error.

Let us assume that a link utilization of 20% is acceptable, which is about 10% larger than that of an Aloha-based packet radio network, a contention-based architecture without carrier-sense, collision-detection, or collision avoidance mechanisms. Then the ratio of propagation time to transmission time, a (from above), must be 2. The frame, then, must contain 1 bit for each factor of 3000 in the distance-rate product, km-bps, as defined in [Kilfoyle]. Thus, for a 1 km acoustic link with a data transmission rate of 300 bps, the minimum frame size must be 100 bits, as shown below.

$$F \geq {}^{rd}\!/_{av} \rightarrow F \geq \frac{1km * 300bps}{(2 * 1.5km/s)} \rightarrow F \geq 100bits \quad (8)$$

While for a 5 km link operating at 3000 bps the required frame size would be at least 5000 bits (Equation 9)

$$F \geq {}^{rd}\!/_{av} \rightarrow F \geq \frac{5km * 3000bps}{(2 * 1.5km/s)} \rightarrow F \geq 5000bits \quad (9)$$

The cost of achieving this utilization is a frame transmission time of 0.66 seconds and 1.66 seconds, respectively. Longer frames will yield a higher utilization rate, while greater distances or higher data rates will require longer frames to sustain the utilization rate – in either case, requiring a longer frame transmission time. It should be noted that this discussion assumes the link is only accessed by the two communicating parties. Increasing the number of communicating devices on the link introduces the likelihood that frames collide, thus reducing the achieved utilization by requiring colliding frames be retransmitted.

Contrast this condition with the contention-based Aloha packet radio network. In this case, the best utilization possible is approximately 18% [Tannenbaum]. As his argument also addresses the impact of multiple users on a single satellite channel, the impact of propagation delays can be ignored – if a station does not receive an acknowledgement for its transmitted message it simply retransmits after a random wait period, tacitly implied to be greater than the transmission time of the frame and its resulting acknowledgement and the round trip propagation delay. Thus, with no active access control mechanism, we can expect ALOHA-based acoustic networks to outperform stop-and-wait based networks where the frame length is very small and where the offered load is no more than one frame per two-frame transmit periods, that is, the time it takes to send two frames [Tanenbaum]. As the load increases active access controls will mitigate the impact of collisions resulting in greater achievable utilization rates but at the cost of access control overhead (increased delay due to additional propagation delay components and media-busy delays).

As demonstrated above, the difficulty with stop-and-wait flow control protocols is that *the frame size must be proportional to the rate-distance product*. That is, as the product of the link transmission rate and the link distance increases, so must the frame size in order to achieve the desired utilization. When multiple users share the medium care must be taken to minimize the impact of collisions. While the Aloha protocol accounts for collisions in a radio-based network where propagations delays are considered negligible, the increased frame size required for equivalent utilization rates for acoustic networks requires a collision avoidance scheme to minimize lost capacity due to frame collisions. Such schemes increase the effective propagation delay by requiring the exchange of access control messages prior to data transmission. This increase in effective propagation delay for data messages requires the frame size of the data frames be larger to achieve the desired utilization. Since the frame size is constrained by the quality of the transmission medium, the effective throughput of stop-and-wait is not optimal in a UAN after the rate-distance product exceeds a certain threshold, especially as multiple users are added to the system. It thus appears impractical to increase F to offset the negative effect of

large propagation delays on link utilization. Further, controlling message latency may be more important for some applications, such as remote device control or event monitoring, where the size of the data exchanged may be small, than overall bandwidth utilization. In such cases it becomes imperative to minimize or eliminate access delays such as those imposed by shared media systems. Therefore, in the presence of large traffic loads or where message latency is the dominating concern the alternative is to abandon half duplex links with their inherent stop-and-wait performance characteristic and implement a full-duplex mode of communication for which a more efficient sliding-window link transfer protocol can be used. This, however, carries with it the costs of implementing the full duplex capability. Having formally established the relationship between required transmission cycles, message size, and channel pool size, the next section describes the use of DAMA to implement a bandwidth-on-demand capability to mitigate the cost of capacity channelization. In order to provide a reasonable base of comparison, the stop-and-wait protocol is used as a worst case sliding-window implementation and the net impact on message latency is modeled by way of simulation.

4. Experimental Evaluation of DAMA Applied to Full-Duplex UANs

Our investigation into the utility of DAMA for mitigating the costs of channelization to support full duplex communications in UANs assessed the expected change in message delivery times when the message is sent across a single hop network, as shown in Figure 1. Messages exchanged between outlying nodes are relayed through the control node, similar to a primary/secondary multi-point wired network. The topology used for the network simulation, as depicted in Figure 1, contains five nodes, each with a dedicated transmission channel. All traffic flows to or from the control node, as noted. While the topology is quite rudimentary, it might be a small portion or cluster of a hierarchical network where one or more of the nodes interface with other local clusters. The control node is equidistant from each of the other four nodes and is responsible for allocating additional channels to those nodes based on requests contained in their frame message headers. As full duplex links are assumed, no access control is necessary. A base case, with no dynamic channels, was used to establish a baseline. This result was compared to delivery times for the same traffic patterns where each node could request additional channels to inverse-multiplex several frames. Each source has exclusive use of its dedicated channel as the channel is not considered part of the allocation pool. We chose a message generation algorithm that used message size to affect the offered load. Three levels of offered load were targeted; light, moderate, and heavy. Message arrivals followed a simple uniform distribution as the offered load levels were more of interest than specific traffic arrivals. While other traffic distributions may more closely mimic network traffic, the purpose here was to generate traffic which stressed the use of dynamic channels. Further, the traffic for each simulation run was generated a priori so the same pattern could be used for each level of dynamic channel availability. The same message generation algorithm is used by the four outlying nodes. The simulation determines the average number of

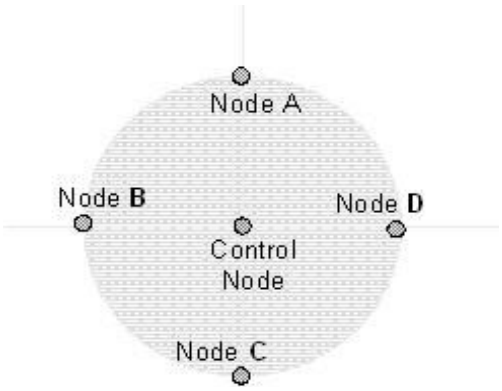

Figure 1. Simulation Topology

transmission cycles required to send a set of messages using dedicated channels only, as well as sending the same message set using both dedicated and dynamically allocated channels. It is important to note that the simulation did not address frame retransmission due to errors or loss. As such, the simulation represents an idealized model of the transmission environment.

Only messages longer than two frames are eligible for dynamic channels, because an allocation is not available any sooner than the cycle after a message arrives and one frame will be sent over the dedicated channel during every cycle, thus two frames will be sent before an allocation can be used. The simulation considered the impact of two allocation schemes, *first-come-first-served* (FCFS) and *fair distribution* (FD). With FCFS, once an allocation was given to a host, based on the indicated number of frames to be sent, the allocation remained in effect until all fames were transmitted. Requests for increased allocations from other hosts were serviced with any remaining resources or deferred until one of the active hosts completed use of its allotted resources if no dynamic channels were available. With fair distribution, resources for all sources were reallocated each transmission cycle based on remaining frames to be sent by each source. Thus, no source is able to prevent other hosts from gaining additional resources; however, the amount of dynamic capacity each source receives is proportional to the number of frames remaining in its current message.

In order to manage the complexity of the model, the simulation determined the allocations of dynamic channels based directly on the generated message arrivals and lengths rather than using an event-based simulation requiring additional intricacy to track the transmission cycle in which specific frames, and thus allocation request and statuses, are exchanged. While the latter would be preferred for evaluating the performance of an implementation of the allocation scheme, the simplified model is sufficient to determine the suitability of the concept, which was the goal of this effort. The conceptual structure of the allocation model may be thought of as a vector for the first-come-first-served allocation scheme sorted by arrival time regardless of source node, and an array for the fair distribution scheme, indexed by source node and message arrival time.

A key decision in developing the simulation was whether to keep the message frame size constant or the message transmission time constant. If the frame size is constant then the transmission time per frame must be increased proportional to the decrease in transmission capacity. If the frame transmission time is to be held constant then the number of frames per message must increase to account for the longer total transmission time. As the purpose of the simulation was to model the change in time periods necessary to send a message using dynamic channels, the frame size was kept constant and the change in message latency was evaluated in terms of time periods to deliver the message and the duration of the time periods as impacted by the transmission time. Also, as a stop-and-wait protocol is assumed, only one frame is sent per time period per available channel. The frame time for the base case, that is no dynamic channels, is assumed to be equal to the round trip propagation delay and the round trip propagation time is normalized to one. This results in a maximum capacity utilization of 50 percent. Traffic generated for the simulation is measured in frames rather than bytes, normalizing the message size to the topology. Thus, increasing the distance between the nodes requires the frame length to be increased accordingly.

As the number of dynamic channels created increases the frame transmission time is increased relative to the base case. For messages sent where no channels were established for dynamic allocation, the duration of the time period, T_{period}, is the sum of the transmission time of the data and the acknowledgement, T_x, and twice the one-way propagation delay, T_p. Thus,

$$T_{period} = T_x + 2T_p \tag{11}$$

If the transmission rate is adjusted for changes in the individual channel capacity due to increasing the number of channels formed to support dynamic allocations, then

$$T_{period} = (\frac{f+d}{f})(T_x) + 2T_p, \tag{12}$$

where f and d are the number of fixed and dynamic channels, respectively. Note that the reason for this tight coupling is the inability to increase the bandwidth available to the system in order to increase the number of channels. For example, if five dynamic channels are established each channel will have half the transmission rate of one of the base case channels, doubling the frame transmission time. While the frame latency increases in this example by 66 percent (from three time units to five), the channel utilization increases by 33 percent (from 50 percent to 66 percent). Thus, a six frame message requiring 11.5 time units in the base case would also require 11.5 time units in the example case, where the time units are normalized to the one way propagation delay. This is the break even point with the example case, assuming only one of the five dynamic channels is made available. That is, messages smaller than six frames may suffer from the creation of dynamic channels, however, messages longer than six frames would always benefit. Since creation of fewer dynamic channels results in a smaller increase in the frame transmission time, the break even point for a smaller dynamic channel pool will be less than six frames, while for larger pools it will be more than six frames. It can be seen then that larger pools favor messages with a larger frame count, while smaller pools are more forgiving of shorter messages.

The following pseudo-code highlights the simulation design. Pseudo-code in the appendix suggests an implementation for an event-driven simulation where the control node generates allocations upon receipt of message traffic.

```
// Traffic generation
1    for each host
2      for each message
3         Generate message start time  // start times concurrent with start of a transmission cycle
4         Generate message length      // length measured in frames
5         Insert message in FCFS vector based on arrival time
6         Insert message in FD array based on source and arrival time
7         Set cycle counter for this message = 0  // maintain separate cycle count for each message

// First-Come-First-Served Allocation Strategy
1    Set dynamic channel counter for each node
2    CurrentCycle = 0
3    dc = DynamicChannelPoolSize
4    while (CurrentCycle < simulation duration) do{
5      for (each message in the FCFS vector in order) do {
```

```
6         if (arrival time this message > CurrentCycle)
7           continue
8         if (frames remaining this message > 0)
9           increment cycle count this message
10          decrement frames remaining this node      // account for frame sent on fixed channel
11          if (frames remaining this message ≥ 1 AND arrival time < CurrentCycle)
12            if (frames remaining this message > dc)
13              dynamic channels assigned this source = dc
14              dc = 0
15              reduce frames remaining this message by dc
16            else
17              dc = dc - frames remaining this message
18              dynamic channels assigned this source = frames remaining this message
19              frames remaining this message = 0
20        else          // completed message: return any dynamic channels still allocated
21          dc = dc + dynamic channels allocated this source
        }
22      increment CycleCount
    }
// end FCFS processing

// Fair Distribution Allocation Strategy
1   CurrentCycle = 0
2   while (CurrentCycle < simulation duration) do (
3     dc = dynamic channel pool size   // allocations only valid for next cycle
4     demand = 0  // demand determined for next cycle
5     for (each message in FD array)  // determine total demand do {
6       if (arrival time == CurrentCycle)
7         increment cycle count this message // initial cycle (fixed channel)
8         decrement frames remaining this message  // account for initial frame
9       if (arrival time ≤ CurrentCycle AND frames remaining this message > 0)
10        increment cycle count this message  // next cycle
11        decrement frames remaining this message  // account for source's fixed channel
12        demand = demand + frames remaining this message
      }
13    for (each message in FD array)  // allocate each share of dynamic pool do {
14      if (arrival time ≤ CurrentCycle AND frames remaining this message > 0)
15        this message share = frames remaining this message/demand)
16        allocation this message = floor(this message share * dc)
17        reduce frames remaining this message by allocation this message
      }
18    increment CurrentCycle
    }
// end FD processing
```

The first section generates the message traffic to be used by the simulation run and stores it in two data structures, the first to be used to determine the FCFS allocation and the second to determine the fair distribution allocation. The next two segments, which determine the number of cycles each message requires, based on channel allocations, must be executed for each value of dynamic pool size. A separate cycle counter must be maintained for each message in each of the allocation strategies to assess the total number of transmission cycles required to send the message in each case. Theses counts are used to compare the average message latency between the various sizes of the dynamic pool and allocation strategies, with the cycle duration adjusted to account for the respective changes in frame transmission times.

Lines 6-8 of the FCFS strategy ensures messages which have not yet "arrived" are not allocated any dynamic channels nor is their frame count reduced. Only those messages that have "arrived" are considered after line 8. The source node's fixed channel capacity is accounted for in line 10. Lines 12-19 allocate channels to the message if there are sufficient frames left beyond the one that can be sent on the fixed channel during the next cycle. Since the messages are handled in order by arrival time, the oldest message gets allocated as many dynamic channels as are available, up to the remainder of frames to be sent. Any remaining channels are allocated to the next message until no channels are available. Since the allocations are recalculated each cycle, as messages complete any allocated channels are returned to the pool, at line 21, and are available for other messages, in turn.

The fair distribution strategy recovers all allocated channels with each cycle and determines the allocation for the next cycle. Thus, lines 6-8 must account for the initial frame sent by the source. Lines 9-11 account for a frame sent during the next cycle over the respective source's dedicated channel, while line 12 determines the total remaining frames to be sent, which forms the basis for each source's dynamic allocation for the next cycle. These allocations are determined in lines 15-16, and line 17 reduces the remaining frame counts for those sources, accordingly. Only those sources with frames remaining beyond that which can be sent over the dedicated channel are considered for an allocation from the dynamic pool. Each allocation is determined by the ratio of frames a source has waiting to the total number of frames waiting for all sources.

Size of Channel Pool	Ratio U_d to U_f	Reduction in Latency	
		Light (load)	Mod/Heavy
1	1.3	14 - 18%	11 – 12%
5	2.5	29 - 42%	39 – 47%
10	4.1	36 – 52%	53 – 62%
15	5.6	39 – 57%	57 – 66%

Table 1. Simulation Results

The simulation was run 30 times, with each run consisting of the three load levels, the base case, and each of the various dynamic channel pool sizes. Three different message sets were randomly generated for each run, each providing a specific traffic load: light, moderate, and heavy. Within each run, the same three sets were used for each size dynamic pool. The results were averaged across the runs, for each category. Table 1 provides a synopsis of the results. Under moderate and heavy loads, where frames requiring transmission accounted for 70% or more of the transmission periods available, the difference between the two allocation strategies was less

pronounced than under light loads. However, in all cases, the simpler FCFS appeared to outperform the more complicated fair distribution scheme. More rigorous simulation is necessary to determine which provides the optimum performance for a given traffic load and with flow control window sizes greater than one. It should also be noted that light traffic loads produced a decreasing scale of return compared to moderate and heavy loads as the number of channels allocated to the dynamic pool increased, which agrees with the analysis in Section 3 regarding transmission cycle reductions.

Figure 2 depicts the change in message latency due to the incorporation of DAMA under the various traffic loads. The graph shows the effect of increasing the dynamic channel pool on message latency. Note that the top line in each graph corresponds to the FCFS allocation. The generated traffic patterns resulted in a 28% capacity usage for light loads, 75% for moderate loads, and 92% for heavy loads. Larger messages are favored in the fair distribution strategy resulting in longer latency for shorter messages as compared to the FCFS scheme. This is reflected by the more pronounced improvement in latency for the FCFS scheme for the light load. *The difference in performance between the two schemes highlights the necessity of choosing an allocation method appropriate to the characteristics and purpose of the traffic being supported*. The most dramatic improvement occurs when the ratio of dynamic channels to fixed channels is less than one. It was observed that as the number of dynamic channels increased the occurrence of idle dynamic channels also increased. This agrees with the analysis of the transmission cycles in Section 3, in that, as the number of channels available exceeds the number of remaining frames to be sent the number of cycles required remains at one; however, the transmission duration of the cycle increases with the number of channels provided due to the fixed system bandwidth. Thus, there exists an optimum number of dynamic channels for a given traffic pattern and, therefore, the channelization scheme must be adaptable to the expected traffic load. For the configuration and message set used for this simulation, the results suggest that setting the dynamic pool size to five channels provides the most significant performance improvement, while allocating another

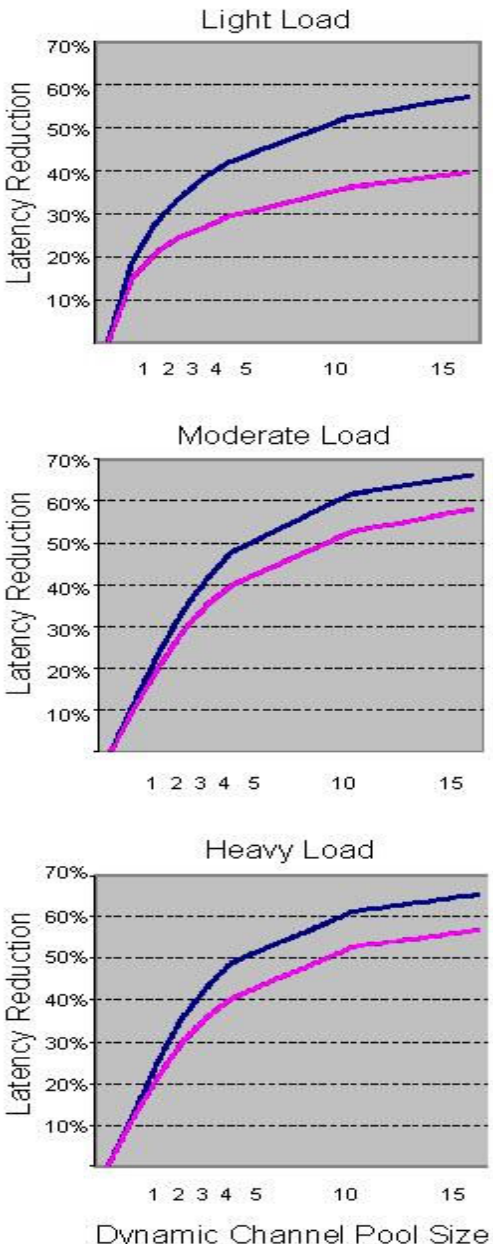

Figure 2. Latency reduction as a function of the dynamic channel pool size. Top lines show performance changes with FCFS scheduling, while the bottom lines reflect changes due to FD scheme.

five channels only resulted in an additional latency reduction of ten percent. Even less significant improvement was gained by allocating more than ten channels to the pool.

One of the principal benefits of implementing full duplex communications is the ability to employ sliding-window protocols, which seek to eliminate the need to pause data transmission while waiting for the acknowledgment of the last frame. As suggested in the previous section, stop-and-wait protocols provide a worst case for sliding-window performance in that they represent a sliding-window of size one, that is, at most one transmitted frame may be pending an acknowledgment. Sliding-window protocols allow multiple frames to be transmitted and pending acknowledgment, with the goal of acknowledgments arriving before the maximum number of frames pending acknowledgment are transmitted, thus allowing data transmission to continue without pauses. Thus, the improvement of sliding-windows protocols over stop-and-wait protocols is a factor of the window size. In effect, larger window sizes emulate larger frame sizes without requiring the pause for acknowledgement between frames.

5. Conclusions

The simulation results highlight the potential of dynamic channel allocation schemes to mitigate the costs of employing full duplex channels in the underwater environment. Further, as they provide a base case, emulating a stop-and-wait protocol, they suggest increased performance gains may be achievable when a sliding-window protocol is employed, thereby maximizing the benefit of full duplex communications. Current trends in underwater acoustic network traffic, to include imagery from periscopes or underwater camera, are for transmission of data messages nearing 4000 bytes [Rice2]. Depending on the arrival rate of such messages, the use of dynamic allocation schemes offer a means of supporting such traffic without effectively isolating other users from the network by precluding their access.

Given that channelization of the total capacity, which would otherwise be available to network users over contention-based access control methods, into multiple channels to support full duplex communications results in wasted capacity when network hosts present sporadic, non constant rate traffic, such capacity loss may be mitigated using dynamic channel allocation pools, sized to the expected traffic load. The simulation results suggest that the size of the pool is related both to the expected traffic load and the number of hosts using the network. Our initial studies indicate that providing a pool of channels approximately twice the number of nodes supported results in a reasonable message delivery latency reduction, while adding more channels to the pool offers diminishing returns. These results warrant further investigation to quantify the potential of such techniques over sliding-window based networks, a crucial and defining benefit of full duplex channels. Additional simulation and analysis is planned to investigate and quantify the benefits to be gained by employing both bandwidth-on-demand, based on DAMA techniques, and sliding-window based flow control to delay challenged networks such as those using underwater acoustic communications.

The selection of the allocation scheme impacts the degree of improvement. While the FCFS scheme showed better performance improvement, in terms of average message latency, than did the FD Scheme, the difference was less marked under heavier loads. The effect of the FD scheme is to implement a form of dynamic priority on larger messages, such that as the message

transmission progresses its relative priority to other messages decreases. This change in relative priorities decreases the effective transmission rate provided to the message as it nears completion, thus increasing the messages transmission delay over that that would be experienced if a static priority were enforced. Since the FD scheme offers no improvement over the simpler FCFS scheme, we recommend implementation of the FD scheme be avoided.

References and selected readings

Barda, A., Z. Ner, 2002, Resource Allocation Techniques for Satellite IP Systems. In *International Broadcasting Convention 2002 Conference*, Sep 2002,

Bektas, K., G. Xie, J. Gibson, 2004, Evaluating the Feasibility of Establishing Full-Duplex Underwater Acoustic Channels. In *Proceedings of the Third Annual Mediterranean Ad Hoc Networking Workshop*, June 2004.

Detlef J. Schmidt, D.J., 1989, DAMA – a New Method of Handling Packets. In *Proceedings of the 8th Computer Networking Conference*, Oct 1989

Gibson, J., A. Larraza, J. Rice, K. Smith, and G. Xie, "On the Impacts and Benefits of Implementing Full-Duplex Communications Links in an Underwater Acoustic Network," *Proceedings of the 5th Mine Symposium*, April 2002

Kilfoyle, D., and A. B. Baggeroer, 2000, The State of the Art in Underwater Acoustic Telemetry. In *IEEE Journal of Oceanic Engineering*, Volume 25, Number 1, January 2000.

Rice, J. A., C. L. Fletcher, R. K. Creber, J. E. Hardiman, and K. F. Scussel, 2001, Networked Undersea Acoustic Communications Involving a Submerged Submarine, Deployable Autonomous Distributed Sensors, and a Radio Gateway Buoy Linked to an Ashore Command Center. In *Proceedings UDT HAWAII 2001 Undersea Defence Technology*, Waikiki, HI, Oct, 2001

Rice, J., Personal Communicaton, 25 May 2004

Sozer, Ethem, M. Stojanovic, and J. Proakis, 2000, Undersea Acoustic Networks. In *IEEE Journal of Oceanic Engineering*, Volume 25, Number 1, January 2000

Stallings, 2004, Data and Computer Communications (7th Edition), New Jersey: Pearson Prentice Hall

Stratman, G., Demand Assigned Multiple Access. In U.S. Army V Corps News Archive, Feb 15, 1999, www.vcorps.army.mil/www/news/1999/february_15_dama.htm

Tannenbaum, A., 1988, Computer Networks, New Jersey: Prentice Hall

Xie, G. and J. Gibson, 2001, A Network Layer Protocol for UANs to Address Propagation Delay Induced Performance Limitations. In *Proceedings of the MTS/IEEE Oceans 2001 Conference*, Honolulu, HI, November, 2001

__, 2000, Information on DAMA. In http://www.qpcomm.com/dama.html, Quantum Prime Communications

__, 2000, DAMA VSAT 10000 Network. In http://www.stmi.com/products_services/dama-10000.asp, STM Networks

Appendix: Pseudo-code for network implementation

State variables on a source node:

NF is the number of frames waiting to be sent
NC is the number of available channels: includes both the dedicated and dynamic
SN is the allocation request sequence number (transmission cycle number)

State variable on a control node:

DCP is the pool of dynamic channels currently available for allocation

```
sendMessage(message msg) {
1   NF = number of frames in msg;
2   NC = 1; SN = 0;
3   while (NF > 0)  do {
4       framesRemaining = max(0, NF - NC);
5       loopCount = min(NF, NC);    // number of frames to be sent this cycle
6       pendingACKS = loopCount;
7       for (loopCount) do {
8          Insert SN and framesRemaining into next frame header;
9          Select available channel;
10         Send next frame on selected channel;
        }
11      SN++;
12      wait (arrival of ACK frame or timeout) {
13         if (arrival of ACK frame: ack)
14         then   processACK(ack);
15                if (--pendingACKS == 0)
16                   break;
17         else   re-queue all unacknowledged frames;
18                break;
        }
19      return;
}

processACK(frame ack) {
1   if (ack is first for SN)
2   then   Extract allocated channel list from ack
3          NC = size of allocation list +1;
4   NF = NF-1;
5   return;
}

processFrame(frame f) {
// Hybrid allocation
1   Extract framesRemaining from header of frame f;
2   if (f comes from a dynamic channel)
3   then   Reclaim channel into DCP;
4   if ((framesRemaining > 1) and (first frame for sequence SN))
5   then   allocationSize = min(framesRemaining-1, size of DCP);
6          Remove allocationSize channels from DCP and insert them into ACK;
7   Send ACK;
8   return;
}
```