Minimizing Average Network Delay for Ultra-Wideband Wireless Networks

Christopher S. Taggart, Yannis Viniotis, and Mihail L. Sichitiu
North Carolina State University
Department of Electrical and Computer Engineering
Raleigh, NC 27695, USA
E-mail: christopher.taggart@gd-ais.com; candice@ncsu.edu; mlsichit@ncsu.edu

***Abstract*— Ultra-wideband wireless (UWB) may provide the physical layer for high capacity personal area networks in the future. Certain characteristics of the wireless links possible using this technology give rise to properties not seen with other wireless technologies. Two such properties are long synchronization times for link establishment and the ability to change individual link capacities by choosing PN codes of different lengths. This paper formulates a novel routing problem in UWB; in particular we investigate the impact of UWB physical layer characteristics on average network transit delay for a UWB network with a ring topology. The paper derives an expression of average network transit delay as a function of the capacity between each pair of nodes in the network. Each node is assumed to have one input and one output link resulting in a ring network topology, which guarantees access to every node. An aggregate network capacity bound for all links is assumed. A Lagrangian function is defined to obtain the optimum link capacities in order to minimize the average network delays and the resulting formula for delay is explained.**

I. INTRODUCTION

Since the Federal Communication Commission (FCC) issued a Report and Order [1] allocating 7,500 MHz of spectrum between 3.1 and 10.6 GHz for unlicensed use by ultra-wideband wireless devices, the landscape of UWB communications research has greatly changed. The primary drivers of the changes were the stringent spectral masks imposed for indoor and outdoor hand-held UWB communications devices.

Much previous research had focused on UWB as an enabling technology for covert, high capacity, multi-user communication systems [2]. The new FCC power and spectral density limitations pushed UWB squarely indoors, cutting its potential range to not more than about 10 meters at high data rates [3]. Previous UWB research was focused on multi-user systems with higher transmit power and reasonably long ranges. These proposed systems used very short ($\sim$1 ns) duration radio frequency pulses to transmit data bits. Coding techniques to channelize data included use of pseudo-random (PN) [4], orthogonal [5], or chaotic [6] codes. These systems achieved the wide bandwidth characteristic of UWB by the use of the extremely short pulses.

More recent UWB schemes include more traditional ways to use the 7,500 MHz of unlicensed spectrum defined by the FCC. Multiband schemes defining several channels of $\geq$ 500 MHz have been proposed. In addition, techniques such as frequency hopping, spectral keying, and orthogonal frequency division multiplexing have been discussed [7]. The power limits imposed by the FCC order restrict the range of practical high capacity UWB systems to approximately 10 meters [8]. As with most wireless data communications systems, lower data rates are possible at longer ranges. UWB is now viewed as a strong contender of providing high data rate wireless connectivity for personal area networks in the home or office environment, for example, replacing cables carrying high definition television (HDTV) signals or allowing high speed wireless access to printers [3]. The IEEE 802.15 Wireless Personal Area Network (WPAN) working group has two committees looking at UWB as the physical layer for both high data rate and low power, low data rate applications [9].

Intel recently demonstrated 220 Mbps communication over a distance of about one meter using a multiband approach [10]. Intel has also joined with a number of other high powered communications technology companies to form the Multiband OFDM Alliance to promote orthogonal frequency division multiplexing (OFDM) as the best alternative for the 802.15.3a UWB PHY and MAC specification [3].

Prior to the new FCC limits, the communication systems using UWB were developed for the U.S. Government. Because of its extremely low power spectral density, UWB transmissions are very difficult to detect - ideal for covert communications. A UWB ad hoc network of eight nodes was even built and tested at Fort Campbell, KY. Called DRACO, this system was capable of sending unencrypted data at 1.544 Mbps at a range in excess of 1 km [11].

An additional characteristic of UWB signals is the ability to easily penetrate walls and inherent immunity to multipath interference due to the wide bandwidth of the transmitted signals. These characteristics make UWB ideal for indoor use. Combined with UWB's ability to handle high capacity data rates, these properties make UWB ideal for self-organizing personal area networks in homes or offices.

A self-organizing personal area network consists of nodes that must not only generate and receive node-specific traffic, but may be required to forward traffic from other nodes. Some of the nodes in a personal area network may be portable, implying possible mobility induced changes to the established network topology. Channelization of the available

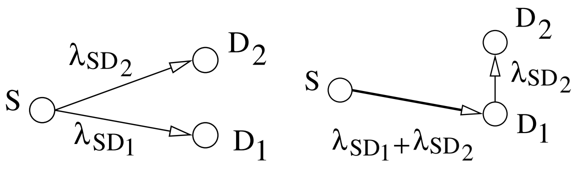
0-7803-8966-2/05/$20.00 (C) 2005 IEEE

UWB capacity and the limited range of the UWB nodes create a challenging routing environment.

The rest of this paper is organized as follows. Section II presents our assumptions about the characteristics of UWB and their influence on networks based on the technology. Section III develops the salient characteristics for the network to be investigated. Section IV derives the formula for minimizing average network delay. Finally, Section V concludes and describes possible future work building on this topic area.

II. UWB CHARACTERISTICS

Due to the very short pulse lengths used in UWB communications, the synchronization time between the receiver and transmitter can be long - on the order of milliseconds [12], [13]. Finding ways to shorten this synchronization time is an ongoing research area [14], [15], [16].



(a) (b)

Fig. 1. Two options to transfer information from a source S to two different destinations D_1 and D_2

Due to the relatively long acquisition times, it might be impractical for a source to frequently change between two destinations. Figure 1 depicts two different strategies for networking a source S and two destinations D_1 and D_2. In Fig. 1(a), the traditional approach is depicted with the source sending packets to both destinations, alternating transmissions between them. In Fig. 1(b), the source directs all traffic to the first destination which, in turn, forwards the traffic to the second destination. The strategy in Fig. 1(a) is commonly employed in wireless networks. However, in UWB, the source has to synchronize with each destination before it can start the data transmission, a potentially very expensive (in terms of efficiency) proposition. Therefore, it is very likely that in UWB the scheme depicted in Fig. 1(b) will be far more efficient than the scheme in Fig. 1(a).

If M transceivers are used for each node, M distinct links can be maintained and used without the overhead of resynchronization; however, it is likely that such a node will be significantly more complex and potentially expensive. In this paper, we focus our attention on the simplest case where only one transceiver is available for each node; and, hence, each network node can send to a single destination and receive from a single source.

Proposed UWB systems transmit tens to hundreds of very short monopulses per information bit [4]. By changing the number of monopulses per bit, the transmission rate can be varied. For example, for 1ns monopulses, using 100 pulses per bit results in a data rate of 10 Mbps, while using 10,000 pulses per bit results in 100 kbps. This property of UWB (also existent in CDMA networks) allows the allocation of different data rates to different nodes in a UWB network.

Whether a UWB system uses frequency division multiple access (FDMA) or code division multiple access (CDMA) to provide separate data channels, the aggregate capacity of nodes within the wireless range of one another has an upper bound, typically much lower than the Shannon limit [8]. The 7,500 MHz of bandwidth allocated by the FCC limits the number of multiband UWB channels. Likewise, if PN codes are used to create separate channels, longer codes create more channels; but, each channel then has a lower bit rate. This feature is not unique to UWB wireless networks.

We assume that all nodes of the network are effectively collocated (a realistic assumption given the relatively reduced range allowed by FCC); and, hence, all nodes interfere with each other. Therefore, while we have the flexibility to allocate different data rates to different nodes, the total capacity of all of the links in the network is limited (we will denote the capacity limit by B). We assume perfect multiple user access through code division multiple access (CDMA), so that each node may transmit and receive simultaneously with no interference.

III. NETWORK DESCRIPTION

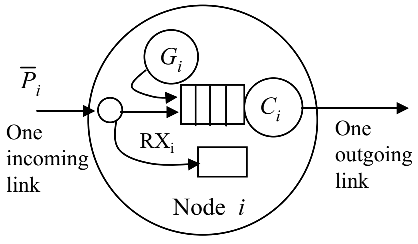


Fig. 2. Block diagram of a network node

Define a network of R stationary wireless nodes, each of which generates its own traffic and routes traffic received from other nodes. Each node i has one incoming wireless link and one outgoing wireless link. Multiple input, multiple output nodes are certainly possible; however, for simple personal area networks where link redundancy is not a major concern, a single input single output assumption is reasonable and cost effective.

Each node generates packets exponentially distributed in time with an average packet generation rate of G_i packets per second. Each node has a single packet queue of infinite size and serves packets such that the inter-service intervals are exponentially distributed with an average service rate of c_i packets per second, which is synonymous with the capacity of the outgoing link from node i. Note that this assumption implies that packet sizes are exponentially distributed.

Arriving packets with destination i are received at an average rate of RX_i . Finally, the average traffic on the incoming link to node i is given by $\overline{P}_i$. Arriving packets destined for other nodes are immediately queued for forwarding. A block diagram of a node from this network is depicted in Fig. 2. We define the traffic matrix

$$\mathbf{T} = \begin{bmatrix} 0 & \lambda_{1,2} & \lambda_{1,3} & \dots & \lambda_{1,R} \\ \lambda_{2,1} & 0 & \lambda_{2,3} & \dots & \lambda_{2,R} \\ \lambda_{3,1} & \lambda_{3,2} & 0 & \dots & \lambda_{3,R} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_{R,1} & \lambda_{R,2} & \lambda_{R,3} & \dots & 0 \end{bmatrix} \tag{1}$$

where each $\lambda_{i,j}$ represents average offered load in packets or bits per second between node i and node j in the network. Let c_i represent the capacity of the outgoing link for node i. Let B represent an upper bound on the aggregate capacity of the network, such that for all connections of capacity c_i between nodes (as discussed in Section II),

$$\sum_{i=1}^{R} c_i \leq B. \tag{2}$$

In general, this upper bound B depends on channel characteristics and multi-user interference, which are beyond the scope of this paper. Our goal is to find a capacity matrix $\mathbf{c}$ that will minimize the average network transit delay for the traffic specified in $\mathbf{T}$ where c_i is the capacity of the outgoing link from node i.

$$\mathbf{c} = \left[\begin{array}{ccccc} c_1 & c_2 & c_3 & \dots & c_R \end{array} \right]. \tag{3}$$

We will show that the nodes in the network must form a ring. Because the traffic matrix requires traffic to be delivered to every node i in the network, no two nodes may connect with their outgoing links to the same node. As shown in Fig. 3, if they did so, one of the nodes would be unable to receive its traffic. No matter how the links are arranged, if two nodes connect their outgoing links to the same node, at least one node cannot receive traffic. The only topology that can be constructed so that all traffic in the traffic matrix $\mathbf{T}$ can be delivered is a ring. The choice of exactly how the nodes should form the ring for optimal performance will not be investigated in this paper but is a topic of future research. Examining Fig. 2, it is clear that the capacity of the outgoing link must be greater than the arrival rate into the service queue, (or the queue will be unstable):

$$c_i \geq G_i + \overline{P}_i - RX_i. \tag{4}$$

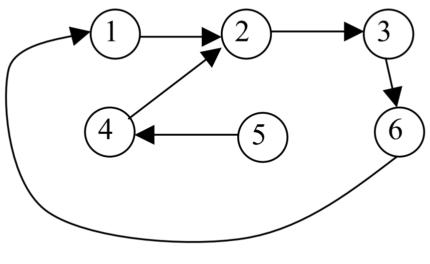


Fig. 3. If the topology is not a ring, some of the nodes cannot receive any traffic.

We can easily write equations for G_i and RX_i in terms of the traffic matrix $\mathbf{T}$:

$$G_i = \sum_{k=1}^{R} \lambda_{i,k}, \tag{5}$$

and

$$RX_i = \sum_{m=1}^{R} \lambda_{m,i}. \tag{6}$$

Figure 4 shows R nodes arranged in a ring. We will determine a value for $\overline{P}_i$ (the average number of packets per second on the incoming link to node i based on the traffic matrix $\mathbf{T}$. In a ring, the average flow $\overline{P}_i$ is the offered load from node $i-1$ to node i. To elaborate, $\overline{P}_i$ is the sum of all traffic from node $i-1$ to any other node plus the sum of all traffic into node i, making sure not to count the flow from node $i-1$ to node i twice. Added to this is the sum of all traffic from nodes $2, 1, \dots, i+2$, to nodes $i+1, \dots, 2$ that must traverse across the link between nodes $i-1$ and i, to reach the destination node. We are only concerned with the traffic to and from these nodes because we have already accounted for all traffic to and from nodes $i-1$ and i in the previous step. Mathematically, this translates to the following equation:

$$\overline{P}_i = \sum_{h=1}^{R} \lambda_{i-1,h} + \sum_{k=1}^{R} \lambda_{k,i} + \sum_{r=i-2}^{i+2} \sum_{s=r-1}^{i+1} \lambda_{r,s}. \tag{7}$$

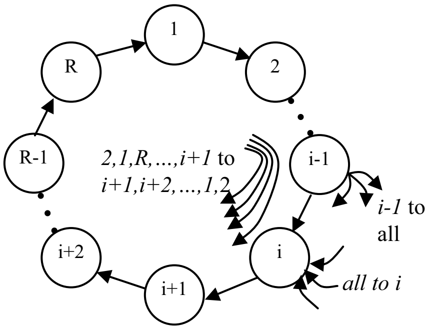


Fig. 4. Traffic on each link is the sum of offered loads that must traverse that link

We can, thus, calculate $\overline{P}_i$ simply by knowing the $\lambda_{i,j}$ values in the traffic matrix $\mathbf{T}$. We will now use the equations from this section to derive an expression for the capacity matrix $\mathbf{c}$ that minimizes average network transit delay.

IV. MINIMIZING NETWORK DELAY

We define the average network transit delay as follows:

$$T_D(\mathbf{c}) = \sum_{i=1}^{R} w_i T_i(\mathbf{c}). \tag{8}$$

In this expression, R is the total number of nodes in the network, w_i is a weighting factor (it can be selected to be different for each node to give nodes preferential treatment), $T_i(\mathbf{c})$ is the average delay in seconds on link i, j caused by transmission, processing, queuing, media access, and propagation delays, and $\mathbf{c}$ is the capacity matrix defined in (3).

The delay on link i, j is the sum of the transmission, queuing, processing, media access, and propagation delays from node i to node j. That is,

$$\begin{aligned} T_i(\mathbf{c}) &= T_{transmission\ i,j} + T_{queuing\ i,j} + T_{processing\ i,j} \\ &\quad + T_{propagation\ i,j} + T_{media\ access\ i,j}. \end{aligned} \tag{9}$$

We assume that processing and propagation delays are small compared to the others and can be neglected. Note that the long synchronization times typically associated with UWB transmissions do not have an impact on the considered problem as the ring topology is fixed after network initialization. We also assume that media access time of the media access control (MAC) protocol is negligible (a realistic assumption given the CDMA character of UWB transmissions: the nodes do not have to wait for other nodes to complete their transmissions before starting their owns). Thus, the delay at node i is

$$T_i(\mathbf{c}) = T_{transmission\ i,j} + T_{queuing\ i,j}. \tag{10}$$

Finally, the queuing and transmission delays can be determined assuming, for simplicity, $M/M/1$ queues. We recognize that assumptions to justify the use of the $M/M/1$ queuing formulas may not apply here, but use them to achieve a closed-form solution to a minimization problem, rather than to calculate some explicit value for delay. In addition, packet framing issues are not addressed here in order to achieve a closed-form solution for delay. As usual in delay studies, frame error rates are assumed to be zero (and, hence, frame error recovery techniques are not taken into account). No packet fragmentation is assumed to occur.

Under these conditions, we can write the delay equation using the $M/M/1$ system delay formula [17]:

$$T_i(\mathbf{c}) = \frac{\frac{1}{c_i}}{1 - \frac{AR_i}{c_i}} \tag{11}$$

where AR_i is the arrival rate of packets into the service queue at node i and c_i is the service rate at node i (equal to the capacity c_i of the outgoing link). The arrival rate of packets into the queue is the sum of packets arriving via the incoming links and those generated within node i, itself, less those packets that are received by node i. Here "received" means taken out of the service queue and placed in a processing queue within the node. If we let RX_i be the rate at which packets destined for node i are received, and $\overline{P}_i$ be the average flow rate of packets on the incoming link to node i, we can write that the arrival rate to the node i queue AR_i is given by:

$$AR_i = \overline{P}_i + G_i - RX_i. \tag{12}$$

Using these identities, we can write the following equation for the delay at node i:

$$T_i(\mathbf{c}) = \frac{\frac{1}{c_i}}{1 - \frac{\overline{P}_i + G_i - RX_i}{c_i}} \tag{13}$$

where c_i is the outgoing link capacity for node i and RX_i is the rate at which packets whose destination is node i are received by the node. After simplifications, we obtain:

$$T_i(\mathbf{c}) = \frac{1}{c_i + RX_i - \overline{P}_i - G_i}. \tag{14}$$

So, we can now write the equation for average network transit delay as:

$$\begin{aligned} T_D(c_i) &= \sum_{i=1}^{R} w_i T_i \\ &= \sum_{i=1}^{R} \frac{w_i}{c_i + RX_i - \overline{P}_i - G_i}. \end{aligned} \tag{15}$$

Now, we state the optimization problem we are trying to solve, namely, to minimize the average network transit delay T_D subject to the constraint (2). We introduce the Lagrange multiplier β and the Lagrangian function L that we will use to determine the minimum values for c_i:

$$\begin{aligned} L &= \sum_{i=1}^{R} (w_i T_i(\mathbf{c}) + \beta c_i) \\ &= \sum_{i=1}^{R} \left(\frac{w_i}{c_i + RX_i - \overline{P}_i - G_i} + \beta c_i \right). \end{aligned} \tag{16}$$

Taking the partial derivatives with respect to c_i and setting the result equal to zero yields:

$$\begin{aligned} \frac{\partial L}{\partial c_i} &= \beta - \frac{w_i}{\left(c_i + RX_i - \overline{P}_i - G_i\right)^2} \\ &= 0. \end{aligned} \tag{17}$$

Solving for c_i, we obtain the following:

$$c_i = \sqrt{\frac{w_i}{\beta}} - RX_i + \overline{P}_i + G_i. \tag{18}$$

The total capacity constraint (2) becomes:

$$\begin{aligned} \sum_{i=1}^{R} c_i &= \sum_{i=1}^{R} \left(\sqrt{\frac{w_i}{\beta}} - RX_i + \overline{P}_i + G_i \right) \\ &= B. \end{aligned} \tag{19}$$

We solve this equation for $\sqrt{\frac{1}{\beta}}$ and substitute in (18), while considering the queue arrival rate (12). We obtain:

$$c_i = \left(\frac{\sqrt{w_i}}{\sum_{i=1}^{R}\sqrt{w_i}}\right)(B - \sum_{i=1}^{R}\overline{P}_i) + AR_i. \tag{20}$$

This result in (20) states that to minimize average network transit delays, any excess capacity must be distributed among the network links in proportion to the square roots of the weights assigned to each link:

$$c_i = \underbrace{\left(\frac{\sqrt{w_i}}{\sum_{i=1}^{R}\sqrt{w_i}}\right)}_{weighting\ factor}\underbrace{(B - \sum_{i=1}^{R}\overline{P}_i)}_{excess\ capacity} + \underbrace{AR_i}_{arrival\ rate}. \tag{21}$$

V. CONCLUSIONS

This paper developed a method to calculate optimal network link capacities for UWB nodes arranged in a ring. We obtained the result by constructing a network model, developing an equation specifying average network transit delay, and applying queuing theory delay equations for $M/M/1$ queues. We, then, used the Lagrangian function to find link capacities to minimize the delay.

While not specifically applied to mobile nodes, this method can be used to specify link capacities required in an ad hoc UWB network with a ring topology to minimize average network transit delay.

This work can be extended by examining the optimum order of nodes in the ring topology to minimize delay based on required network traffic. In addition, the effect of adding additional incoming and outgoing links to the UWB nodes can be studied to determine topologies and capacities to minimize delay. Finally, the effect of expanding the UWB network so that some nodes are not in range of other nodes can be investigated to determine methods for minimizing the network transit delay in extended UWB networks.

REFERENCES

[1] "First Report And Order Before the Federal Communications Commission In the matter of Revision of Part 15 of the Commission's Rules Regarding Ultra-Wideband Transmission Systems," Federal Communications Commission Washington, D.C., *ET Docket 98-153,* Adopted: February 14, 2002, Released: April 22, 2002.

[2] S.S. Kolenchery, J.K. Townsend, J.A. Freebersyser, "A Novel Impulse Radio Network for Tactical Military Wireless Communications," *Proceedings of IEEE MILCOM '98*, November 1998.

[3] S. Stroh, "Ultra-Wideband: Multimedia Unplugged," *IEEE Spectrum,* pp. 23-27, September 2003.

[4] M.Z. Win, R.A. Scholtz, "Impulse Radio: How It Works," *IEEE Communications Letters*, Vol. 2, No. 1, pp. 10-12, January 1998.

[5] K. Eshima, Y. Hase, S. Oomori, F. Takahashi, R. Kohno, "M-ary UWB System Using Walsh Codes," *Proceedings IEEE Conference on Ultra Wideband Systems and Technologies,* pp. 37-40, May 2002.

[6] G.M. Maggio, D. Laney, F. Lehman, L. Larson, "A Multi-Access Scheme for UWB Radio Using Pseudo-Chaotic Time Hopping," *Proceedings IEEE Conference on Ultra Wideband Systems and Technologies,* pp. 225-229, May 2002.

[7] G.R. Aiello and G.D. Rogerson "Ultra-Wideband Wireless Systems," *IEEE Microwave Magazine,* June 2003.

[8] S. Roy, J.R. Foerster, V.S. Somayazulu, D.G. Leeper, "Ultrawideband Radio Design: The Promise of High-Speed, Short-Range Wireless Connectivity," *The Proceedings of the IEEE*, Vol. 92, pp. 295-311, March 2004.

[9] M. Welborn, T. Miller, J. Lynch, J. McCorkle, "Multi-User Perspectives in UWB Communications Networks," *Proceedings IEEE Conference on Ultra Wideband Systems and Technologies,* pp. 271-275, May 2002.

[10] M. Williams, "Intel takes Ultra Wideband to 220 Mbps," *InfoWorld,* April 11, 2003.

[11] R. Fontana, A. Ameti, E. Richley, L. Beard, D. Guy, "Recent Advances in Ultra Wideband Communications Systems," *Proceedings IEEE Conference on Ultra Wideband Systems and Technologies,* pp. 129-133, May 2002.

[12] R.A. Scholtz, "Multiple Access with Time-Hopping Impulse Modulation," *Proceedings of the Military Communications Conference (MILCOM '93),* Vol. 2, Boston, MA, pp. 447-450, October 1993.

[13] S.S. Kolenchery, J.K. Townsend, J.A. Freebersyser, G. Bilbro, "Performance of Local Power Control in Peer-to-Peer Impulse Radio Networks with Bursty Traffic," *Proceedings of IEEE Globecom'97,* pp. 910-916, Phoenix, November 1997.

[14] R. Fleming, C. Kushner, G. Roberts, U. Nandiwada, "Rapid Acquisition for Ultra-Wideband Localizers," *Proceedings IEEE Conference on Ultra Wideband Systems and Technologies,* pp. 245-249, May 2002.

[15] Y. Ma, F. Chin, B. Kannan, "Acquisition Performance of An Ultra Wide-band communications System Over A Multiple-Access Fading Channel," *Proceedings IEEE Conference on Ultra Wideband Systems and Technologies,* pp. 99-103, May 2002.

[16] E. Homier, R. Scholtz, "Rapid Acquisition of Ultra-Wideband Signals in the Dense Multipath Channel," *Proceedings IEEE Conference on Ultra Wideband Systems and Technologies,* pp. 105-109, May 2002.

[17] D.P. Bertsekas, R. Gallagher, *Data Networks*, Prentice-Hall, Second Ed, 1992.