International Journal of Power Control Signal and Computation(IJPCSC)
Vol3. No1. Jan-Mar 2012 ISSN: 0976-268X
www.ijcns.com

Minimum Bandwidth Reservations for Periodic Streams in Wireless Real-Time Systems

[1]Shanmuga priya a,[2]Sumathi.p
[1]charmingsp88@gmail.com, [2]sumathip84@mail.com

Abstract

Reservation-based (as opposed to contention-based) channel access in WLANs provides predictable and deterministic transmission and is therefore able to provide timeliness guarantees for wireless and embedded real-time applications. Also, reservation-based channel access is energy-efficient since a wireless adaptor is powered on only during its exclusive channel access times. While scheduling for Quality of Service at the central authority (e.g., base station) has received extensive attention, the problem of determining the actual resource requirements of an individual node in a wireless real-time system has been largely ignored. This work aims at finding the minimum channel bandwidth reservation that meets the real-time constraints of all periodic streams of a given node. Keeping the bandwidth reservation of a node to a minimum leads to reduced energy and resource requirements and leaves more bandwidth for future reservations by other nodes. To obtain a solution to the minimum bandwidth reservation problem, we transform it to a generic uniprocessor task schedulability problem, which is then addressed using a generic algorithm. This algorithm works for a subclass of priority-driven packet scheduling policies, including three common ones: fixed-priority, EDF, and FIFO.

Moreover, we then specialize the generic algorithm to these three policies according to their specific characteristics. Their computation complexities and bandwidth reservation efficiencies are evaluated and guidelines for choosing scheduling policies and stream parameters are presented.

Index Terms—Bandwidth reservation, schedulability test, earliest deadline first, fixed-priority, first-in-first-out, medium access control, real time.

1. INTRODUCTION

WIRELESS embedded real-time systems are becoming prevalent with the continuous increase in streaming applications such as video/audio communications, industrial automation, networked and embedded control systems, and wireless sensor and actuator networks. This has called for research efforts to enhance the support of timeliness and Quality of Service (QoS) in wirelessly networked embedded environments. Wireless networks are inherently broadcast and media-shared. Contentionbased media accesses such as CSMA are nondeterministic and thus incapable of providing predictable QoS support to periodic communications often found in wireless real-time systems.

Moreover, multiple nodes are active simultaneously and continuously sense and contend for the shared media, leading to excessive energy consumption. Recently, reservation-based channel access protocols that explicitly allow wireless devices to negotiate channel access intervals have been receiving increasing attention. Such access mechanisms allow for contention-free and exclusive accesses, providing deterministic bounds on the delays experienced by the traffic streams and conserving energy (since wireless adaptors having no channel access can be temporarily powered down). Therefore, such access mechanisms are ideally suited for providing real-time services in wireless environments. For example, in ad hoc networks, coordinated sleep mechanisms have been designed [1] to allow wireless devices to coordinate medium access with their neighbors and to reduce their energy requirements. Similarly, in infrastructure-based systems, wireless end devices can coordinate medium access with base stations (BSs) using protocols such as the IEEE 802.11e [2]. Reservation-based channel management requires each node to negotiate its desired channel access duration for a given period based on its traffic constraints. However, the computation of such requirements has largely been ignored which has often resulted in poor real-time support, overprovisioning of valuable resources, and poor scalability.

The goal of this work is to develop a strategy for the computation of the required channel access reservations for a given packet scheduling policy, such that 1) the real-time constraints of each node's traffic are satisfied and 2) resource reservations are minimized.

To solve the minimum bandwidth reservation problem at a given node, we treat the complement of the periodic bandwidth reservation as a special periodic stream (the periodic sleep stream), i.e., the bandwidth reservation per channel access period is equal to the complement of the execution time (sleep time) of the sleep stream with period equal to the channel access period. We add the sleep stream to the original stream set to form an extended stream set.

Accordingly, the scheduling policy for the extended stream Set is extended from the original scheduling policy for the original stream set such that 1) the sleep stream always has the highest priority and is noninterruptible and 2) the priority relationship among the original stream set is unchanged. As a consequence, we transform the minimum bandwidth reservation problem to the maximum sleep time problem. In other words, there exists a schedule for the original stream set with a given scheduling policy if and only if there exists a schedule for the extended stream set using the extended scheduling policy. Therefore, minimizing the bandwidth reservation is equivalent to maximizing the sleep time of the sleep stream.

2. RELATED WORK

Scheduling and schedulability analysis have been extensively studied in previous work, particularly for processing resources. In networking environments, reservation-based mechanisms are becoming highly prominent in supporting latency-critical and energy-aware traffic. In this section, we discuss existing protocol standards and techniques related to resource and channel access reservations.

A well-known wireless standard that offers channel access reservations is the IEEE 802.11e protocol [2]. The IEEE 802.11e standard proposes a Hybrid Coordination. Function (HCF) that provides both contention-based and contention-free channel accesses through two modes: the Enhanced Distributed Channel Access (EDCA) and the HCF Controlled Channel Access (HCCA) [2]. With HCCA, the Hybrid Coordinator (HC), which usually resides at the base station, continuously polls every node. TXOPs (Transmit Opportunities) are assigned by the HC to a node at a regular interval and for a specified duration, which are determined based on the node's traffic specification. The research results reported in this work can be applied to the HCCA mode to help each node reserve the smallest amount of bandwidth necessary to meet all packet deadlines.

RI-EDF [4] is a table-driven, slotted reservation protocol based on earliest-deadline first (EDF). It uses the periodic nature of traffic in a fully connected network to deduce a shared packet transmission schedule. Packets are transmitted during their allocated slots, thus avoiding contention.

Traffic from the same node is interspersed into discrete slots. In contrast, our reservation model reserves a continuous time interval for every individual node and nodes only wake up during their allocated intervals.

An earlier work studying the same problem [5] relies on several strong assumptions. It assumes that 1) the channel access period is smaller than every stream period, 2) datagram deadlines must be equal to or less than their respective periods, and 3) datagrams must be transmitted consecutively and without interruption. Also, the work implicity assumes that the underlying scheduling policy is FIFO. Although it has a linear complexity, it overreserves bandwidth in general cases. This paper removes these restrictions, takes into account the impact of different scheduling policies on the computation of the required bandwidth, and presents algorithms for several scheduling policies to efficiently compute the minimum bandwidth reservation, although at the cost of higher complexity compared to [5]. However, even with increased complexity, the presented algorithms are practical considering that the computational capacity of wireless end devices continuously increases and that a wireless end device usually has only a limited number of concurrent real-time streams.

In this work, we transform the minimum bandwidth reservation problem to the maximum sleep time problem, which is then solved by computing the schedulable execution time of the sleep stream for a subclass of scheduling policies, including fixed-priority, EDF, and FIFO.

A similar approach has been taken in [6] to solve the minimum EDF-feasible deadline problem of a given task, given its period and execution time. There exist a lot of efforts on exact schedulability tests for various scheduling policies, e.g., fixed-priority [7], [8], [9], EDF [10], [11], [12], [13], and FIFO [14]. Our generic algorithm for the bandwidth reservation problem is based on the time-demand analysis techniques provided by these earlier research results, but applied to a new problem.

Our work is also closely related to previous work on resource partition/composition models, which usually focus on processor resources in real-time systems. Such models include the static resource partition model [3], the bounded delay resource partition model [3], [15], [16], the periodic resource model [17], [18], and the explicit deadline periodic resource model [19], [20]. These prior efforts differ from each other mainly in the chosen scheduling model. The resource partition/reservation model used in this paper corresponds to the single time slot periodic partition (STSPP) model (a special case of the static resource partition model) introduced in [3].

The schedulability test scheme for the STSPP model in [3] can also be used to solve the problem, by iteratively executing the schedulability test algorithm proposed in [3] (similar to a binary search). Therefore, this approach would be very costly. It is particularly inefficient for the fixedpriority policy since the change of the resource supply and the change of job response times are nonlinear and irregular. Our algorithm augments the traditional timedemand analysis to compute the finished/unfinished portion before a job's deadline, which avoids computing fixed-point equations iteratively. As a result, our solution to the reverse problem (the minimum resource requirement problem) has the same complexity as the original problem (the exact schedulability test problem).

Besides resource partition/composition models, hierarchical schedulers (e.g., [21], [22]) can also provide temporal isolation among applications on a uniprocessor. This property prevents a misbehaving task from interfering with other tasks in another application, i.e., only the tasks within the same application as the misbehaving one could be affected. In hierarchical scheduling, each application is composed of a set of correlating entities (e.g., tasks or streams), where applications are scheduled by a global scheduler and each application schedules its tasks using its local scheduler. The approach proposed in this paper can be applied to hierarchical schedulers, i.e., it can be used to determine the minimum resource requirements of an application, given this application's local scheduling policy (fixed-priority, EDF, or FIFO).

3. BANDWIDTH RESERVATION MODEL

This section presents our network access model, traffic model, and the problem statement.

A. Network Access Model

We briefly discuss the concept of reservation-based channel access model (which corresponds to the single time slot periodic partition model introduced in [3]) since it forms the basis for the problem we intend to solve. Such a mechanism uses resource reservations to ensure contention-free accesses. This is achieved through a central authority at a BS that regulates the channel accesses of individual nodes.

Here, the BS takes control of the channel and starts polling each of the nodes in a predetermined order (e.g., roundrobin). Upon reception of a polling frame, a node gains access to the channel. The HCCA mode defined in the IEEE 802.11e standard [2] is an example of a protocol which adopts the reservation-based channel access approach to enhance the QoS support for real-time applications in wireless environments.

Borrowing the terminology from the IEEE 802.11e standard, in a reservation-based channel access mechanism, each node is provided a Service Period (SP), during which the node has exclusive access to the wireless medium.

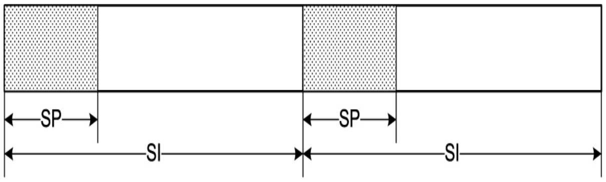


Figure.1.A wireless device's bandwidth profile (SP; SI).

Shaded intervals (SPs) are the exclusive access periods for a node, which are repeated every SI time units. Polling frames issued by the BS specify the start time and maximum duration of the SP allotted to a node. At the end of an SP for a node, the BS begins polling the next node in its schedule. The period of recurrence of the SPs is referred to as the Service Interval (SI), which is usually specified by the BS in advance and equal to a multiple of the beacon interval of the BS shared by all client nodes. We call the pair B ¼ ðSP; SIÞ the bandwidth profile of a node. The SP parameter at each node must be negotiated with the BS based on the requirements of the node's expected realtime traffic. This reservation-based channel access model is very practical and valuable in both wireless local area networks (WLANs) and wireless sensor networks (WSNs): There are three advantages of using STSPP in these network areas: 1) it saves energy since nodes only need to wake up to communicate within their respective reserved time intervals; 2) it leads to better latency predictability and possibly higher throughput since wireless contention is avoided a priori; 3) it greatly decreases the runtime complexity of resource partition scheduling due to its simple partition structure.

B. Traffic Model

We consider a set of wireless nodes with applications on each node generating one or more periodic real-time streams. Nodes connect wirelessly to a common BS to access an external network. We denote the set of periodic streams generated by a node as S ¼ fS1; . . . ; Sng. Each stream Si periodically generates a certain number (worst case or average case) of bytes (called a datagram) for a given period pi for transmission. The datagram generated at the beginning of the jth period of Si for transmission is denoted as Ji;j. Wireless channel conditions are time-varying and error-prone.

The worst-case estimation (denoted as ei) of the transmission time of a datagram of Si is needed, which has been the focus of many prior efforts (e.g., in [5] and [24]). Each datagram of Si has a relative transmission completion deadline Di. The release time and deadline of Ji;j are denoted as ri;j and di;j ¼ ri;j þ Di, respectively. Our framework requires no specific relationship between stream periods and datagram deadlines, i.e., Di can be less than, equal to, or greater than pi. Due to the similarity between the concept of tasks in the literature and the concept of streams in this paper, we will use stream and task interchangeably in this paper. Similarly, the terms datagram and job are also used interchangeably.

Datagrams are often fragmented at the network and/or link layer, depending on the datagram size, network parameters (e.g., the maximum transfer unit or MTU), and the scheduling policy. Therefore, a datagram can be also YI ET AL.: MINIMUM BANDWIDTH RESERVATIONS FOR PERIODIC STREAMS IN WIRELESS REAL-TIME SYSTEMS 481 Fig. 1. A wireless device's bandwidth profile (SP; SI). Shaded intervals (SPs) are the exclusive access periods for a node, which are repeated every SI time units. treated as a logical conglomeration of a series of physical packets. The maximum size of packets cannot be greater than the MTU. When a packet is in flight, no other packet can interrupt it. Therefore, the worst-case non preemption portion of a datagram is equal to the MTU. Since all streams at the same node share the same MTU value, we set the nonpreemption portion of every stream to be the MTU value and denote it as Although packet interruption is not allowed, interleaved transmissions of packets of a datagram with other packets of other datagrams are allowed. For example, when packet pkt1;1;1 from datagram J1;1 of stream S1 finishes its transmission and a more urgent datagram J2;1 from stream S2 arrives, the scheduler may need to transmit packets from J2;1 before other packets from J1;1 (e.g., pkt1;1;2). As another example, after packet pkt1;1;1 finishes its transmission, the node's allocated time interval for network access is used up and the node is forced to sleep.

During the sleep duration, more urgent packets may arrive at the network queue. When the network resource is available again (after the sleep duration), the scheduler may need to transmit these newly arrived urgent packets before packets from J1;1. These interleaved transmissions happen frequently since applications treat the network as a dedicated resource and issue packets regardless of the network reservation.

C. Problem Definition and Objectives

Each client node requests its desired bandwidth reservation from the BS and the normalized bandwidth (i.e., the ratio of SP to the given SI) of a node should be minimum as long as all real-time streams meet their deadlines. Minimizing the reserved bandwidth ensures that a node has the maximum amount of sleep time, thereby minimizing the energy consumption of its wireless network card. From the perspective of the BS, the normalized bandwidth of each node should be minimum as well so that the BS's throughput is maximized and the maximum amount of bandwidth is available for potential future reservation requests of other nodes. Therefore, the problem and objective can be stated as:

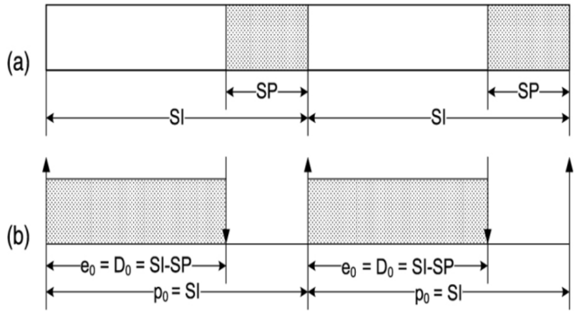


Figure.2. Transformation of the bandwidth profile and the sleep stream:
(a) A given bandwidth profile (SP; SI). (b) The corresponding sleep stream S0 with D0 ¼ e0 ¼ SI-SP and p0 ¼ SI. It always has the highest priority and it is noninterruptible.

Problem1(MinimumBandwidth Reservation (MBR)).

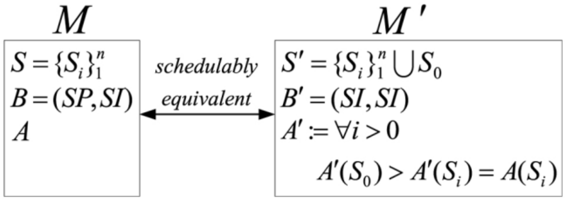
Given a node's set S ¼ fSi gn 1 of periodic streams, a fixed service interval SI, and the node's scheduling policy A, determine the minimum SP such that all streams in S meet their deadlines.

Normally, the SI value given to a node is a multiple of the beacon interval of the BS [2]. The beacon interval is determined by the application scenarios for which the WLAN is deployed. Once bandwidth reservations are allocated to nodes by the BS, a change of the SI value will necessitate changes of all allocated bandwidth reservations and trigger renegotiations between the BS and all of its nodes.

A solution/algorithm to the problem can run at each client node if the SI value is given from the BS to the client node, or run at the BS if the client node communicates all of its stream parameters to the BS. The actual implementation choice depends on communication and computation capacities of nodes and the BS.

4. MODEL TRANSFORMATIONS

To approach the MBR problem, we first transform the scheduling model M ¼ ðS ¼ fSign 1;B ¼ ðSP; SIÞ;AÞ of the MBR problem to another scheduling modelM0 ¼ ðS0;B0;A0Þ,



where the stream set S0 extends S by adding a sleep stream; the scheduling policy A0 extends A by assigning the sleep stream the highest priority; and B0 represents the dedicated Resource allocation for S0. We show that the two scheduling models are schedulable equivalent, i.e., M is schedulable if and only if M0 is schedulable. As a corollary, we show that the MBR problem in M is a dual of the maximum execution time problem of the sleep stream in M0.

5. MINIMUM BANDWIDTH RESERVATIONS

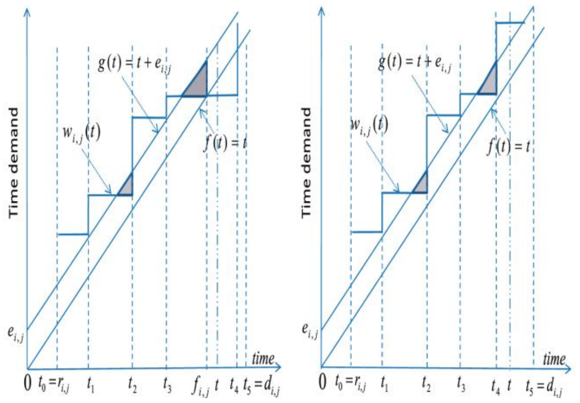
A priority-driven scheduling policy (a set of priority rules) can be considered as a time-varying function AðJ i;k; J j;l; tÞ for any two jobs J i;k and J j;l, taking on values _1, 0, and 1.

Function AðJ i;k; J j;l; tÞ ¼ 1 (-1 and 0, respectively) if and only if J i;k has a higher (lower and equal, respectively) policy A is priority-driven if and only if 1) a backlogged processor is not idle (i.e., AðJ i;k; J _1;lÞ ¼ 1 for every job J i;l, where i 6¼ _1) and 2) the processor invariantly executes the highest priority backlogged job J i;k at any time t (i.e., AðJ i;k; J j;l; tÞ ¼ 1 for every J j;l 6¼ Ji;k) In systems that use dynamic job-level scheduling, a datagram (a job) may consist of multiple packets, whose priorities may be different from each other or they may change over time. This, while adding flexibility, would significantly add to the complexity of the network scheduler.

Instead, this work assumes that the network scheduler is job-level static, i.e., all packets from the same datagram have the same priority, which is assigned at the release time of the datagram. We would like to point out that many prioritydriven scheduling policies (especially widely used ones such as RM, DM, EDF, FIFO, LILO, and roundrobin) satisfy these constraints. Policies such as least slack time (LST) do not fall into this category. In the following sections, we restrict our focus to the class of priority-driven policies that satisfy these constraints.

A. Generic Framework

In this section, we develop a generic algorithmic framework (Algorithm 1) to solve the MET problem, based on an augmented time-demand analysis. In the following description, we distinguish between the finished portion and the unfinished portion of execution before a given deadline. This concept allows us to conservatively reduce the sleep time of the sleep stream to approach its minimum value, assuming that the reduced sleep time (equal to the unfinished portion) is solely utilized by the job missing its deadline. To compute the finished/unfinished portions of a job, we define a generic time-demand function at every job release event point. The available time at a job release event point for lower priority jobs is equal to the dedicated time supply minus the generic time demand. As a result, the finished portion of a job before its deadline is equal to, provided it has not finished, the maximum available time over all job release event points between its release time and its deadline.



6. RESULTS AND PERFORMANCE

This section describes a set of simulation experiments that have been conducted to evaluate the over reservation ratio of bandwidth and the computational overhead of our algorithm under typical application scenarios using four different scheduling policies (EDF, FIFO, RM, and DM).

These results can provide the system designer and administrator with guidance on which policy to choose in what workload scenarios, also indicating how much bandwidth would be wasted. The periodic streams are generated with random parameters within given ranges and distributions. Moreover, we experimentally verify the correctness of our algorithm using various test cases.

7. CONCLUSIONS AND FUTURE WORK

The benefits of reservation-based channel accesses are twofold: 1) they provide contention-free access within allocated/reserved channel access intervals to meet timing constraints predictably and 2) they allow a wireless radio to be powered down when the channel is not needed. Careless resource allocations may lead to poor support for real-time traffic or overprovisioning of scarce network resources. This paper solves the minimum bandwidth reservation problem to allow all streams to meet their timing constraints.

To obtain a solution to the minimum bandwidth reservation problem, we transform it to a generic uniprocessor task schedulability problem, which is then addressed using a generic algorithm based on time-demand analysis. The generic minimum bandwidth reservation algorithm works for a subclass of priority-driven packet scheduling policies, including three common ones: fixed-priority (e.g., RM and DM), EDF, and FIFO. Refinements of the generic solution to these three types of policies are presented and discussed as well. The simulation results show that the generic algorithm is correct and practical in terms of computation complexity.

The proposed bandwidth reservation scheme leads to minimal amounts of bandwidth waste if appropriate scheduling policies and stream parameters are selected for a given stream set. However, it also leads to potentially large energy savings, while being simple to implement and deploy. In our future work, we will address 1) how the base station chooses the optimal SI value to minimize the bandwidth/energy consumption of the entire wireless local area network, 2) how different client nodes use an SI which is a multiple of the beacon interval of the base station, and
3) how reservations (SPs) of nodes with different SIs can be composed efficiently to form a superframe.

REFERENCES

[1] L. Wang and Y. Xiao, "A Survey of Energy-Efficient Scheduling Mechanisms in Sensor Networks," Mobile Networks and Applications, vol. 11, no. 5, pp. 723-740, 2006.
[2] IEEE 802.11e Standard, IEEE 802.11 WG, Part II: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QoS), Nov. 2005.
[3] A.K. Mok, X.A. Feng, and D. Chen, "Resource Partition for Realtime Systems," Proc. Seventh IEEE Real-Time Technology and Applications Symp., pp. 75-84, 2001.
[4] T.L. Crenshaw, S. Hoke, A. Tirumala, and M. Caccamo, "Robust Implicit EDF: A Wireless MAC Protocol for Collaborative Realtime Systems," ACM Trans. Embedded Computing Systems, vol. 6, no. 4, 2007.
[5] D. Rajan, C. Poellabauer, X.S. Hu, L. Zhang, and K. tten, "Wireless Channel Access Reservation for Embedded Real-Time Systems," Proc. Seventh ACM Int'l Conf. Embedded Software, pp. 129- 138, 2008.