Optimality of Periodwise Static Priority Policies in Real-time Communications

I-Hong Hou, Anh Truong, Santanu Chakraborty and P. R. Kumar

***Abstract*— We consider the problem of real-time communication with delay constraints. In earlier work it has been shown that a certain weighted-debt policy is feasibility-optimal in the sense that if any scheduling policy can satisfy the throughput-with-deadline requirements of all the clients, then the weighted-debt policy can do so. This raises the interesting question: Why is it that a periodwise static priority policy can satisfy any set of requirements that the more general class of history dependent policies can? We answer this by showing that the set of feasible timely-throughput vectors is a polymatroid. We do so by establishing a submodularity property of the complement of the unavoidable idle time function. This shows that a periodwise static priority policy, where the priority order is revised at the beginning of each period, but never in the middle of a period, can attain any feasible timely-throughput vector.**

We next go on to investigate a more general problem where the packet arrivals and channel conditions can vary over periods, and establish the existence of an optimal periodwise static priority policy.

Keywords: Real-time communication, Polymatroids, Submodularity, Periodwise Static Priority Policies, Throughput, Deadlines, Timely throughput.

I. INTRODUCTION

Wireless networks are becoming an integral part of the global communication infrastructure. There is increasing demand for wireless data services for multimedia applications, video streaming, VoIP, real-time monitoring, networked control, etc. Thus, providing temporal Quality of Service (QoS) over unreliable wireless channels has become an important service for wireless networks to support. In previous work [2], a model and framework have been developed for jointly addressing the QoS criteria requirements of supporting required delivery ratios, given channel reliabilities, subject to the constraint that the delay of each delivered packet is less than one time period.

This material is based upon work partially supported by USARO under Contract Nos. W911NF-08-1-0238 and W-911-NF- 0710287, NSF under Contracts CNS-1035378, CNS-0905397, CNS-1035340, and CCF-0939370, and AFOSR under Contract FA9550-09-0121.

I-Hong Hou is with CSL and the Department of Computer Science, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA ihou2@illinois.edu

Anh Truong is with CSL and the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA truong3@illinois.edu

Santanu Chakraborty is with the Illinois Center for Transportation, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA santanu@illinois.edu

P. R. Kumar is with Computer Engineering and Systems Group, Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, USA prk@tamu.edu

In the model of [2], time is slotted, and slots are broken into groups of τ slots each, called periods. One packet arrives for each of a number of clients at the beginning of each period. This packet is dropped if it is not delivered by the end of the period. In each slot, the access-point (in the case of downlink, or some one client in the case of uplink) can attempt to transmit one packet. The transmission may however be successful or not, depending on the channel unreliability between the access-point and the particular client. The timely-throughput for a client is defined as the long-term rate at which packets are successfully delivered to (or from, in the case of uplink) the client. Given a period for the whole system, a timely-throughput requirement for each client, and a channel reliability for each client, the fundamental question addressed in [2] is whether it is possible to satisfy a given set of such requirements. It was shown there that a certain weighted-debt policy is feasibility optimal in the sense it can meet the QoS requirements of any system for which there is a history dependent policy that can do so.

This weighted-debt policy orders the clients at the beginning of each period, and persistently transmits the packets of a client until it is successful, before moving on to the packet of the next lower client in the ordering. Of course a period may run out before all clients are served. At the commencement of the next period, the policy again orders the clients, possibly in a different order, and serves them in that order during the period, etc. It should be noted that this weighted-debt policy never revises the priority order of clients in the middle of a period. In particular the order is not changed within a period due to the random events that may occur in that period (e.g., a certain packet is unsuccessfully attempted a certain number of times). However, the order is revised at the beginning of the next period. We call such a policy a *periodwise static priority policy*. The intriguing question that arises is why such a restrictive policy can meet the QoS requirements of any set of clients whenever there exists a more general history-dependent policy that can meet their requirements. Why is it that one does not need to consider a more general and powerful history dependent policy?

In this paper, we directly show why it is that a *a periodwise static priority policy* can meet the QoS requirements whenever there is a more general history dependent policy that can do so. Having done so, we move on to consider a more general model for supporting QoS. We allow random arrivals of packets to clients at the beginning of each period, which relaxes the earlier requirement that one packet arrives for each client at the beginning of the period. Packet arrivals to clients can even be correlated, as for example may happen in a multimedia situation where there are stereo cameras. We also allow for the channel for each client to vary from period to period. Such a model was considered in [3], and a certain joint debt-channel policy was shown to be feasibility-optimal in the class of all periodwise static priority policies. Here we show the new result that the joint debt-channel is actually optimal in the general class of all history-dependent policies.

Our main contributions in this paper are therefore twofold. First, we prove why consideration of only *periodwise static priority polices* is sufficient to support any real-time traffic. We do this by showing that the feasibility conditions satisfy a certain submodularity property. From this it follows that the convex hull of occupation measures achieved by history dependent policies is simply the convex combination of those achievable by periodwise static priority policies. So, only randomized periodwise static priority policies are the extreme points of the convex hull of occupation measures of the history dependent policies, since time sharing of randomized priority policies can cover all points in that set. The second contribution is the extension of the above result to systems where the packet arrivals at the beginning of a period are random and possibly correlated, and where channel conditions change from period to period. For this more general model we prove that a earlier joint debt-channel policy [3] is feasibility optimal in the general class of history dependent policies. This generalizes the previous result where optimality is only established in the class of periodwise static priority policies.

The rest of the paper is organized as follows: Section II summarizes some related work. Section III describes the system model. Section IV establishes that there exists a periodwise static priority policy that is feasibility optimal. We then use this result to extend the optimality of such a policy for the case of various packet arrival patterns and time-varying channels in Section V. Finally, Section VI concludes this paper.

II. Related work

In recent years, scheduling for QoS constraints on unreliable wireless networks has been increasingly gaining research interest. Tassiulas and Ephremides [7] have proposed a max-weight scheduling policy and proved that it is throughput optimal. Neely [9] has further evaluated this policy and shown that the policy achieves a constant average delay. Shakkottai and Stolyar [8] have evaluated various scheduling policies to support a mixture of real-time and non-real-time traffic. In Hou, Borkar, and Kumar [2], a model and framework for admission control and scheduling under QoS constraints have been proposed for the case where all clients generate traffic periodically with the same period. A periodwise static priority policy has been proposed and proved to be feasibility optimal in that it fulfills the requirements of all feasible systems. However, a thorough understanding has been lacking on why a periodwise static priority policy can be feasibility optimal. The model studied in [2] is further extended in [4] to accommodate different traffic arrival patterns generated by different flows, and in [3] to time-varying channel reliabilities. In this work, a certain periodwise static priority policy has been proved to be optimal in the class of all such policies. However, whether it is optimal in the class of all history dependent policies was left unresolved.

III. Model for QoS

In this section we describe the model for supporting time-based QoS described in [2], for the class of problems where the delay allowed for each delivered packet is no longer than one period of τ slots. Consider a wireless access point serving N clients numbered $1, 2, \ldots, N$. Time is slotted and packets arrive at the access point at time slots $0, \tau, 2\tau...$. The quantity τ is the *period*. At each such time slot $n\tau$, there is simultaneous arrival of N packets, one for each client. Each packet has to be delivered to its respective client before the end of the period, or else it is dropped. Thus, in this model, a delay constraint of τ time slots is enforced on all successfully delivered packets.

The access point may make one packet transmission in each slot. However, the channels to the clients are unreliable. When the access point transmits a packet to a client (or, in the case of uplink, when the client i transmits a packet to the access point), it is successful with probability p_i, and unsuccessful with probability $(1 - p_i)$.

Each client requires a *timely throughput* of q_i packets per period. This is the long-term average of the number of packets per period that need to be delivered to client i. We call this the timely throughput because packets are counted as delivered only if they are delivered within the deadline τ. The entire problem is therefore characterized by the following quantities:

1) The period τ.
2) The unreliabilities of the channels to the N clients $\{p_1, p_2, \ldots, p_N\}$.
3) The timely throughputs required by the clients $\{q_1, q_2, \ldots, q_N\}$.

The question of interest is this: Can the access point satisfy the timely throughput requirements of these N clients?

In [2] it is shown that if there is *any* history dependent scheduling policy that can satisfy all the clients, then the following *weighted-debt policy* is feasibility optimal: At the beginning of period $n\tau$, the access point computes the *weighted debt* owed to client i as

$$\frac{1}{p_i}\left(q_i - \frac{\text{\# packets delivered to client } i \text{ in } [0,(n-1)\tau]}{n}\right).$$

This is essentially the shortfall in the fraction of jobs that the server needed to have delivered for client i in order to have met its demand, weighted by the factor $\frac{1}{p_i}$. (If the debt is negative, the client is ahead of its requirement). The clients are then served in the order of their weighted debt, with higher weighted debt clients getting higher priority. It should note that this is a periodwise static priority policy that orders the clients at the commencement of the period, but never revises this order within the period in response to events that may have unfolded within the current period.

IV. Optimality of periodwise static priority policy

In this section we directly establish the existence of a *periodwise static priority policy* which satisfies a set of clients that can be satisfied by some history dependent policy. We start with the feasibility region for the vectors $(q_1, q_2, ..., q_N)$ that has been characterized in [2] as follows. Note that the number of transmissions to serve client i is a geometrically distributed random variable, with parameter p_i. Let $I(S)$ denote the expected fraction of unavoidable idle time that is incurred by the access point, if it can serve only the clients in the set $S \subseteq \{1, 2, \dots, N\}$. Thus,

$$I(S) := E\{\frac{1}{\tau}\left(\tau - \sum_{i \in S} \gamma_i\right)^+\},$$

where $\gamma_i \sim \text{Geom}(p_i)$. We also define $f(S) = 1 - I(S)$, which is the expected fraction of time that the access point spends transmitting packets, if it is non-idling, by which we mean that it never idles as long as there are undelivered packets, and it has only the set S of clients to serve. Then a vector $(q_1, q_2, ..., q_N)$ is feasible if and only if it lies in the polytope:

$$\mathcal{P} = \left\{(q'_1, q'_2, ..., q'_N) \middle| \sum_{i \in S} x_i \le f(S), \forall S \subseteq \{1, 2, \dots, N\}, \right.$$
$$\left. \text{where } x_i = \frac{q'_i}{\tau p_i}, \text{and } q'_i \ge 0, \forall i, \right\}$$

We begin our analysis by showing that the above set is a polymatroid. This is shown by proving that the function $f(S)$ is submodular (see Theorem 1).

We note that the polytope $\mathcal{P}$ is a polymatroid if and only if $f(S)$ satisfies the following properties (see Yao [1]).

(a) $f(\phi) = 0$.
(b) f is non-decreasing.
(c) f is submodular, i.e., if $E, F \subseteq S$, then $f(E) + f(F) \ge f(E \cup F) + f(E \cap F)$.

The first property is obvious for f. The second property follows from the fact that the idle time function $I(\cdot)$ is a non-increasing function. That is, if there were a larger set of clients, then any non-idling policy would necessarily idle less. Thus, the only condition that remains to be checked is the submodularity of f.

Definition 1. *For two disjoint set of clients A and B, let $f(A|B)$ denote the expected fraction of time the access point spends transmitting packets for the clients in A, given that the access point always schedules transmissions for the clients in A right after it delivers all packets for the clients in B, and only clients in the set $A \cup B$ are available. Note that, obviously, $f(A|B) = f(A \cup B) - f(B)$.*

Lemma 1. *For subsets A, B, and C, if $B \subseteq C$ and $A \cap C = \phi$, then $f(A|B) \ge f(A|C)$.*

Proof. By definition, we have

$$\begin{aligned}
&f(A|B) \\
=&f(A \cup B) - f(B) \\
=&I(B) - I(A \cup B) \\
=&E\{\frac{1}{\tau}\left(\tau - \sum_{i \in B} \gamma_i\right)^+\} - E\{\frac{1}{\tau}\left(\tau - \sum_{i \in A \cup B} \gamma_i\right)^+\} \\
=&E\{\frac{1}{\tau}\min[\left(\tau - \sum_{i \in B} \gamma_i\right)^+, \sum_{i \in A} \gamma_i]\}.
\end{aligned}$$

Similarly,

$$f(A|C) = E\{\frac{1}{\tau}\min[\left(\tau - \sum_{i \in C} \gamma_i\right)^+, \sum_{i \in A} \gamma_i]\}.$$

Since $B \subseteq C$, $\sum_{i \in C} \gamma_i \ge \sum_{i \in B} \gamma_i$, and therefore $f(A|B) \ge f(A|C)$. ☐

Theorem 1. *For any subsets of clients E and F, $f(E) + f(F) \ge f(E \cup F) + f(E \cap F)$.*

Proof. We have

$$f(E) = f(E \setminus F|E \cap F) + f(E \cap F),$$

and

$$f(E \cup F) = f(E \setminus F|F) + f(F).$$

Therefore,

$$\begin{aligned}
&[f(E) + f(F)] - [f(E \cup F) + f(E \cap F)] \\
=&[f(E \setminus F|E \cap F) + f(E \cap F) + f(F)] \\
&- [f(E \setminus F|F) + f(F) + f(E \cap F)] \\
=&f(E \setminus F|E \cap F) - f(E \setminus F|F) \ge 0.
\end{aligned}$$

The last inequality holds by Lemma 1 and the fact that $(E \cap F) \subseteq F$. ☐

Next, note that a history dependent policy is a policy which uses the information of the clients from past to the current time, to decide which packet is to be served in any slot. It keeps calculating during the period and can change the order of clients within a period even if those have made prior unsuccessful attempts and would like to retransmit. On the contrary, a *periodwise static priority policy* is a static priority policy (e.g., debt-first policy) which maintains the order of the clients throughout one period regardless of the outcome of the tranmission of packets in that period. A randomized periodwise static priority policy is a periodwise static priority policy which, at the begining of each period, randomly chooses a particular order for all clients, and maintains that order for the whole period. If a history dependent policy, or a periodwise static priority policy is feasibility optimal, it must satisfy the above constraints. Theorem 2 proves the existence of a randomized periodwise static priority policy.

Theorem 2. *If there exists a history dependent policy which is feasibility optimal, then there exists a randomized periodwise static priority policy, which is also feasibility optimal.*

Proof. Let $\mathscr{C}$ be the convex set satisfying the constraints describing the polytope $\mathcal{P}$. Consider an ordering, which could be any permutation over the N clients, $\pi = [\pi_1, \pi_2, ..., \pi_N]$. If the randomized periodwise static priority policy (here we only consider non-idling policies) picks this order, its expected busy times spent on various subsets of clients are described by the following equations:
$x_{\pi_1} = f(\{\pi_1\})$,
$x_{\pi_2} = f(\{\pi_1, \pi_2\}) - f(\{\pi_1\})$,
...
$x_{\pi_N} = f(\{\pi_1, \pi_2, ..., \pi_N\}) - f(\{\pi_1, \pi_2, ..., \pi_{N-1}\})$.
Because the number of time slots given for clients is bounded by τ, $\mathscr{C}$ is a closed and bounded convex set. It follows that if every extreme point of a polytope is some static priority policy, then the class of periodwise static priority policies covers all extreme points of $\mathscr{C}$, and thus can realize any history dependent policy by time sharing. We therefore only need to show that every extreme point of $\mathscr{C}$ is a static priority policy x_π. It is indeed shown in [1] that, for a polymatroid, every extreme point corresponds to a static priority policy, proving the result. □

V. Extensions for Time-Varying Channels and Variable-Bit-Rate Traffic

We now consider a more general model for arrivals and channels considered in [3]. Such a model supports systems with time-varying channels and variable-bit-rate traffic. In [3], there is proposed a scheduling policy, called a *joint debt-channel policy*, which is a particular periodwise static priority policy, and it is proved that such a policy is feasibility optimal among all periodwise static priority policies. In this section, we extend this and show that the joint debt-channel policy is actually feasibility optimal among the class of all history dependent policies.

We first describe the extended model. It is assumed that the channel reliability between the access point and a client can vary over time. In particular, when the access point schedules a transmission for client i at any time for period k, then the transmission is successful with probability $p_k(t)$. The channel reliability, $p_k(t)$, is assumed to remain the same within the period k, and only changes from period to period.

The other feature allowed in this more general model is that clients may generate packets according to a more general traffic pattern. At the beginning of a period, only a subset of clients may generate packets.

It is assumed that both the channel reliabilities for clients and packet generations in each period evolve as an irreducible (without loss of generality) Markov chain with finite number of states. The state of the system at the beginning of a period is fully described by the channel reliabilities and packet generations in this period. We denote the set of all possible states by C, and suppose that, in steady state, a particular state $c \in C$ occurs with probability r_c.

Similar to the previous model, each client i requires its long-term timely throughput to be at least q_i. As earlier, we define $I_c(S)$ as the expected fraction of forced idle time when the access point only serves clients in S when the state of the system is c, and define $f_c(S) := 1 - I_c(S)$. The same argument as in Theorem 1 shows that $f_c(S)$ is also submodular for each c.

As in [3], the problem of providing long-term timely throughput to each client can be solved by decoupling the system states. Let us suppose that we aim at providing a timely throughput $q_{c,i}$ to client i under state c. The choices of $[q_{c,i}]$ are made so that the long-term timely throughput of each client averaged over all states is at least q_i. Clearly, the feasibility conditions for each state c need to be satisfied. Thus we formulate the decoupled problem as one of solving the following optimization problem:

$$\max \sum_{c,i} r_c q_{c,i}$$
$$\text{s.t.} \sum_{c} r_c q_{c,i} \geq q_i, \forall i,$$
$$\sum_{i \in S} \frac{q_{c,i}}{\tau p_c} \leq f_c(S), \forall c \in C \text{ and } S \subseteq \{1, 2, \dots, N\}.$$

A similar argument as in Theorem 2 shows that if the vector of timely throughput requirements of clients, $[q_i]$, can be achieved by any policy, then there exists a randomized periodwise priority policy that achieves the same timely throughput requirements. Noting that Theorem 4 in [3] has shown that the joint debt-channel policy is feasibility optimal among all periodwiwe static priority policies, we obtain the following theorem.

Theorem 3. *The joint debt-channel policy proposed in [3] is feasibility optimal among all history dependent policies.*

VI. Conclusion

In this paper we have studied the problem of providing delay based QoS in unreliable wireless networks. We have proved the existence of a periodwise static priority policy which is feasibility optimal for the class of problems where packets have to delivered before the end of the packet in which they arrive. Our analysis has proceeded by establishing the key property that the function representing the unavoidable idle time incurred by the access point in each period is supermodular. We have further investigated a more general problem with variable-bit-rate packet arrivals and time-varying channels and established the optimality of a certain joint debt-channel policy over the class of all history dependent policies.

References

[1] D. D. Yao, *Dynamic Scheduling via Polymatroid Optimization*, Performance Evaluation of Complex Systems: Techniques and Tools, Performance 2002, Tutorial Lectures, Springer-Verlag, London, UK, 2002.

[2] I-Hong Hou, V. Borkar and P. R. Kumar, *A Theory of QoS for Wireless*, Proceedings of Infocom 2009, April 19-25, 2009, Rio de Janeiro, Brazil.

[3] I-H. Hou and P.R. Kumar. Scheduling heterogeneous real-time traffic over fading wireless channels. In *Proc. of IEEE INFOCOM*, 2010.

[4] I-Hong Hou, and P. R. Kumar, *Admission Control and Scheduling for QoS Guarantees for Variable- Bit-Rate Applications on Wireless Channels*. Proceedings of MobiHoc 2009, pp. 175-184, New Orleans, May 28-May 21, 2009

[5] J. Edmonds, *Submodular Functions, Matroids, and Certain Polyhedra*, Combinatorial optimization - Eureka, you shrink!, Springer-Verlag New York, Inc., New York, NY, USA 2003.

[6] V. Borkar, *Convex analytic methods in Markov decision processes*, Borkar, V. 2001. E. Feinberg, A. Shwartz, eds. Handbook of Markov Decision Processes: Methods and Applications. Kluwer, Boston, MA

[7] L. Tassiulas and A. Ephremides, *Dynamic server allocation to parallel queues with randomly varying connectivity*. IEEE Trans. on Information Theory, 39, March 1993

[8] S. Shakkottai and A. L. Stolyar, *Scheduling algorithms for a mixture of real-time and non-real-time data in HDR*. In Proc. of 17th International Teletraffic Congress (ITC-17), 2001

[9] M. J. Neely, *Delay analysis for max weight opportunistic scheduling in wireless systems*. In Proc. of Allerton Conf., 2008