The Case for Using Traffic Forecasting in Schedule-Based Channel Access

Vladislav Petkov and Katia Obraczka
School of Engineering
University of California, Santa Cruz
Email: {vladi,katia}@soe.ucsc.edu

***Abstract*—In this paper, we explore the idea of using traffic forecasting to improve the delay performance of a schedule-based medium access control protocol. Schedule-based channel access has been shown to utilize network and energy resources efficiently but is often hindered by the extra delay that scheduling introduces. We explore the use of traffic forecasting to anticipate transmission schedules instead of establishing them reactively, thereby reducing scheduling delays. We show the potential performance benefits traffic forecasting can bring to schedule-based medium access in the context of an existing MAC protocol called DYNAMMA [14]. Preliminary results using a machine-learning based traffic forecasting technique are also presented.**

I. Introduction

As a result of the wide availability and variety of wireless devices and the ubiquity of wireless communication infrastructure, a number of new applications and services have emerged – notably the "Smart Environments", which include the "Digital Home". One of the challenges imposed by Smart Environments in general, and the Digital Home in particular, is their high throughput and stringent quality-of-service (QoS) requirements.

As new physical layer technology such as highly integrated, small foot-print, single-chip radios combined with advanced coding and modulation techniques are able to support very high data rates, the fundamental limits challenging development and deployment of Digital Home applications have been shifting from the physical (PHY) layer to the medium access control (MAC) layer. Consequently, it is imperative to design efficient MAC techniques that will "expose" the underlying PHY's high data rates to the applications while meeting their QoS requirements such as low delay and delay-jitter.

Random access (a.k.a., contention-based) channel access methods, such as IEEE 802.11's DCF, are known to have their performance deteriorate sharply as traffic load and network density increase. This degradation in performance is due to collisions that reduce channel efficiency and increase per-packet energy cost.

Schedule-based solutions to the medium access problem – such as those presented in [4], [3], [14] – eliminate collisions, improving channel efficiency substantially. Furthermore, using time slots to arbitrate medium access makes it possible to do more effective sleep-scheduling [14] and thus considerably improve energy efficiency. Dynamic schedule-based medium access approaches schedule data transmissions in response to application traffic. While these solutions maximize channel utilization by allocating channel time only to nodes that need it, they may exhibit longer data delivery delays. This extra delay is due to the fact that before data can be sent, traffic information needs to propagate so that nodes can schedule accordingly.

In this paper we are interested in answering the following question: "If we use predictions of the behaviors of flows in the network, can we decrease the delay in schedule-based medium access control?" The main idea is to use traffic forecasting to **anticipate** transmission schedules instead of establishing them reactively, i.e., as traffic arrives at the MAC layer. Although not all applications generate forecastable traffic, we contend that many applications do, in particular most Digital Home services that have stringent QoS demands. Examples of predictable network traffic include Voice-over-IP (VoIP) applications such as Skype, iChat, and Google talk. Video streaming applications have lower QoS demands but also contain many predictable patterns. All of these applications are becoming increasingly commonplace in the home networks of today. As we move towards "smart" homes and offices, we argue that even more media-carrying, forecastable data streams will flow over the underlying networks.

The remainder of this paper is organized as follows. We start by presenting an overview of related work in Section II. Then, in Section III, we present some preliminary results showing the potential performance benefits traffic forecasting can bring to schedule-based channel access. We then describe a machine-learning based traffic forecasting technique using the expert framework ([8], [9]) in Section IV. We conclude with a discussion of open research issues and the direction of our future work.

II. Related Work

A. Schedule Based MAC Protocols

Contention-based medium access methods suffer from collisions that cause performance deterioration with increased load and also contribute significantly to the energy consumption of the radios. Schedule-based medium access methods on the other hand eliminate collisions and provide much higher delivery-ratio performance. Through structured use of the channel, they can also eliminate overhearing and idle-listening, which gives them very good energy saving properties.

The *TRAMA* (TRaffic Adaptive Medium Access) [13] protocol, is a schedule-based protocol designed to be bandwidth- and energy-efficient. TRAMA achieves energy efficiency by eliminating collisions and allowing nodes to sleep when they are not the intended receivers of the current transmission. While TRAMA achieves considerable energy savings (when compared to contention-based MAC protocols such as S-MAC), it incurs high delay.

DYNAMMA (DYNAmic Multi-channel Medium Access) [14], like TRAMA, gives attention to energy consumption and provides more application independence than FLAMA [12], a scheduled-access protocol designed specifically for sensor network applications. The DYNAMMA framework boasts the ability to adapt to application traffic patterns, provide collision-free multi-channel operation, while achieving reduced signaling overhead. DYNAMA's performance in terms of medium utilization and energy consumption is superior to that of TRAMA and contention-based protocols. However, while the average delay packets incur in DYNAMA is less than that of TRAMA, it is still significantly larger than that of IEEE 802.11 (which we use here as a baseline for contention-based MACs). Our goal is to explore whether traffic forecasting can be used to reduce this delay commonly found in schedule-based MAC protocols. In Section III we use a modified version of DYNAMMA, which we call DYNAMMA-PRED, to show the potential improvement possible if traffic forecasting is used.

B. Traffic Forecasting

Dynamic schedule-based protocols achieve collision-freedom by exchanging some form of scheduling information in advance of actual data transmission. Transmission slots are allocated based on this information. Allocating more slots than the traffic needs wastes channel time, while allocating less slots than needed will increase the delivery delay. We propose to use a traffic forecaster at each traffic source that constantly adapts its forecast of how many slots and with what spacing will be required to best serve each data flow. The forecasts are disseminated to the two-hop neighborhood and used to schedule the medium access. A survey of the literature shows that traffic forecasting has not been used at the MAC layer in the way we are proposing.

In [1], the authors present a dynamic bandwidth allocation strategy that, combined with the use of the Renegotiated Constant Bit Rate (RCBR) service model can accommodate variable bit rate (VBR) video while reducing queue sizes and increasing network utilization. The method used is adaptive linear prediction minimizing mean square error. With the RCBR service model, a large frequency of bandwidth renegotiations corresponds to higher network utilization but also larger signaling overhead; the authors present some methods to control this tradeoff. In our forecast-enabled, schedule-based MAC we encounter a similar set of problems in terms of: allocating enough time slots to a flow to satisfy its bandwidth needs while trying to dampen fluctuations in the forecasts to reduce dissemination overhead.

In the work described in [2] the authors investigate several bandwidth allocation policies that can be used at the adaptation layer between ATM and IP networks. More specifically, the paper compares the performance of static algorithms (where bandwidth allocation does not change), periodic algorithms (where the bandwidth allocation is changed periodically) and adaptive algorithms (where the bandwidth allocation is changed as necessary within some restrictions) is compared in the paper.

Koutsakis et al. have shown how traffic models can be used in call admission schemes to provide better quality of service when delivering video conferencing, MP3 downloads, and MPEG-4 streaming in cellular wireless networks [10],[5].

Traffic prediction or forecasting has also been used by Liu et al. in their work [11] to dynamically set the duration of a transmission opportunity (TXOP) in IEEE 802.11e in order to improve the quality of service given to variable bit-rate video.

C. Quality of Service

One of the benefits of a schedule-based MAC is the ease with which Quality of Service (QoS) guarantees can be made. IEEE 802.11 on the other hand struggles in this regard. There exists a large body of work concerned with adding QoS to IEEE 802.11. One notable example is the work described in [15] focusing on bandwidth management in ad-hoc networks. The solution presented there is a dynamic bandwidth management scheme that provides an application level solution to the problem of providing QoS in an ad-hoc network. Under the centralized control of a bandwidth manager entity, each node adapts its rate factoring in both the requirements of its applications and the available bandwidth. The whole framework runs on top of IEEE 802.11. Although the proposed dynamic bandwidth management is an effective method of providing QoS, it is complex and depends on a central management entity, which makes the system less flexible and prone to a central point of failure. The traffic forecasting, schedule-based MAC that we aim to develop as a follow up to this work will not only be able to deliver QoS guarantees without a central management authority, but it will be able to infer the QoS needs of applications and meet them without the need for any cross-layer awareness, greatly reducing the overall system complexity.

III. Benefits of Traffic Forecasting

In this section we quantify the impact of traffic forecasting on the performance of scheduled-based MACs. We do this by assuming that traffic can be forecast with 100% accuracy and evaluating the performance under this assumption. Assuming an optimal forecaster will give us an upper bound on the performance of forecast-based medium access.

Evaluation Methodology

We chose to use a simulation-based evaluation methodology to determine an upper bound on the performance of schedule-based medium access given perfect traffic forecasting. Specifically, the DYNAMMA protocol presented in [14] is adapted to use "oracle" traffic forecasting and its performance is compared to DYNAMMA's unmodified version.

DYNAMMA is a schedule based protocol. Each node that participates in the network synchronizes to its neighbors and broadcasts a beacon during its assigned beaconing slot at the beginning of each superframe (a construct commonly seen in schedule-based MAC protocols that defines a repeating pattern of time slots). When a node has traffic for one of its neighbors, it means it has an outgoing flow to that neighbor. Beacon packets are used to inform a node's neighbors of its own flows as well as those of it's neighbors. Using this method two-hop flow knowledge is established.

At the start of each data slot of the the superframe one or more flows are pseudorandomly selected to use that slot. Flows requiring more bandwidth are favored by the pseudorandom generator. Multiple flows can use the same slot only if their simultaneous transmissions will not cause a collision.

DYNAMMA-PRED uses the same framework as DYNAMMA with the addition of an "oracle" forecaster, which provides global queue state knowledge to the whole network. Unlike DYNAMMA, in DYNAMMA-PRED only flows that are known to have at least one packet to send are eligible to contend for a given slot. Although this method is not achievable in practice because it relies on all nodes knowing instantly when a packet is ready to be sent at any of the nodes in their 2-hop neighborhood, the result provides a baseline of what is achievable with traffic forecasting.

Observations

To compare the performance of the two protocols we examine queuing delay, delivery ratio (total packets received divided by total packets sent), and percentage of time spent with radios in sleep mode using the Qualnet network simulator.

The scenario depicted in Figure 1 represents an infrastructure-less home network. An ad-hoc deployment will be essential to alleviating the burden of set-up and installation of smart devices on the consumer. The simulation scenario consists of 16 nodes arranged in a grid formation. They are spaced apart such that the diagonal spacing between two nodes is within radio range (25 meters) and creates a neighborhood pattern. The radio transmission rate is 53.3Mbps using a WiMedia physical layer model as defined in the ECMA-368 ultra wide-band physical and MAC layer standard [7]. Each node is a traffic source and randomly generates packets for random neighbors.

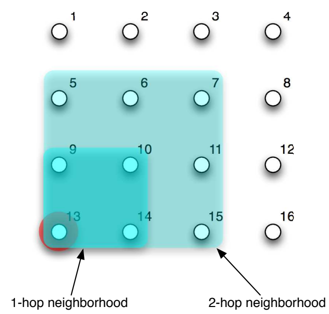

Fig. 1. Simulation scenario: each node randomly generates traffic for its 1-hop neighbors.

The effect of the "oracle" forecaster on queueing delay is of particular interest. In Figure 2 we observe that the delay performance of DYNAMMA-PRED is better than that of DYNAMMA for inter-arrival times of $5ms$ and more. At high loads, when bandwidth demand exceeds availability and MAC layer queues grow, DYNAMMA-PRED has comparable delay performance to DYNAMMA. DYNAMMA-PRED reaches a new delay floor of around $0.5ms$. To put this into perspective, the width of a slot is $638.125\mu s$ so the average delay is less than one slot length. DYNAMMA-PRED will try to schedule a flow for transmission during the slot after the packet arrives. The worst case delay in a lightly loaded system would be one slot length and the best case delay would approach zero, so on average the delay should be close to half a slot duration. The low delay of 802.11 at high load is misleading – the significant number of dropped packets are not accounted for in the MAC-layer delay measurement.

DYNAMMA is a multi-channel MAC protocol and we include results for 1–, 2– and 3– channel operation for both the original DYNAMMA and augmented DYNAMMA-PRED.

Figures 3 and 4 confirm that DYNAMMA-PRED retains the high delivery ratio and energy efficiency of DYNAMMA. It is not traffic forecasting but rather just the schedule-based nature of the medium access that provides good performance in these two aspects.

It is worth noting that the delivery ratio of schedule-based protocols drops only when the MAC-level queues fill up. Contention based protocols, on the other hand, begin to experience collisions even before the queues fill and so their delivery ratio steadily degrades with increased traffic load.

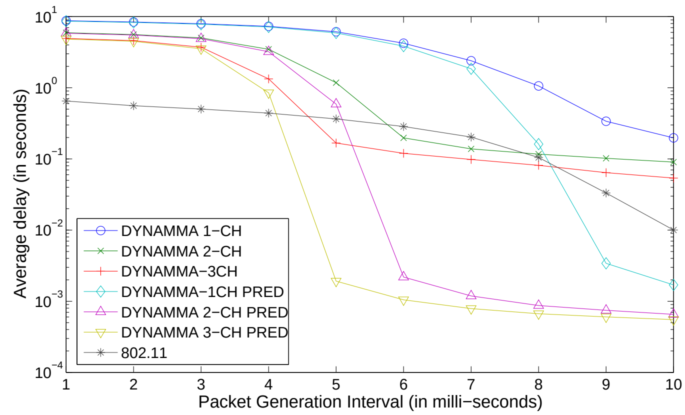

Fig. 2. Average queueing delay of IEEE 802.11, DYNAMMA, and DYNAMMA-PRED.

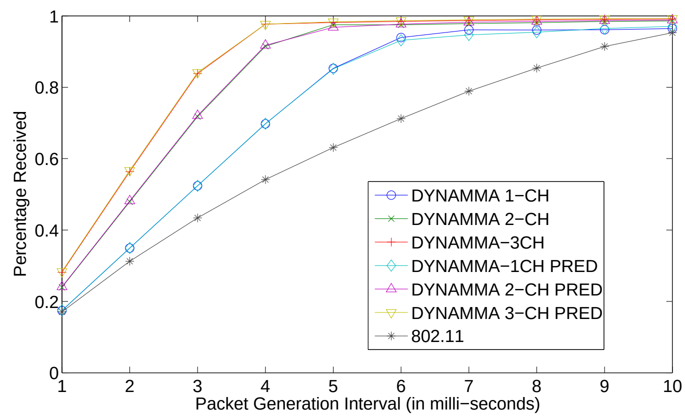

Fig. 3. Average delivery ratio of IEEE 802.11, DYNAMMA, and DYNAMMA-PRED.

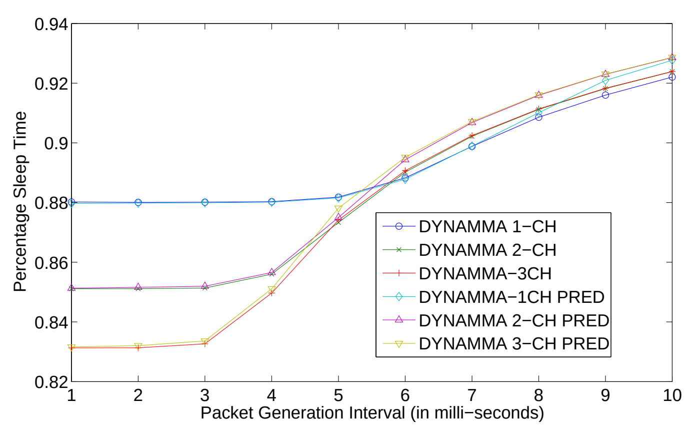

Fig. 4. Percentage of time spent sleeping in DYNAMMA, and DYNAMMA-PRED.

IV. FORECASTING TRAFFIC

The positive effect of the oracle forecaster can be seen in the results of DYNAMMA-PRED. Though "oracle" forecasting is not realizable, past traffic trends and patterns can be harnessed to anticipate future traffic behaviors.

There are several properties of a flow's traffic that may need to be forecast such as amount of throughput it will need and for what duration, the arrival time of its next packet, the size of the next packet or the average packet size. For the purpose of slot scheduling, we formulate the traffic forecasting problem in terms of the expected arrival time of the flow's next packet. In other words, the forecasting approach described in this section will focus on predicting a flow's expected packet inter-arrival time in order to schedule future transmission slots for that flow. The proposed traffic forecasting algorithm, which is based on the experts framework [9], selects the best inter-slot duration that a node should use to service a flow with minimal delay and minimal slot wastage.

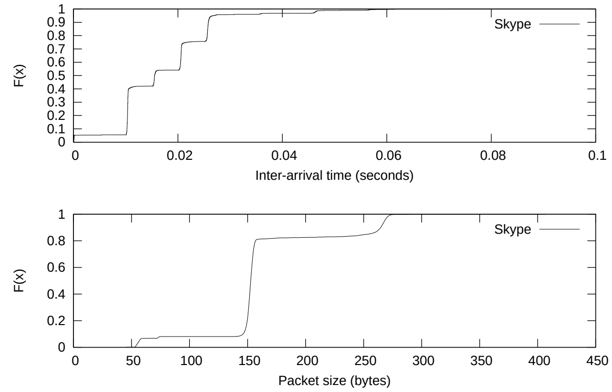

Fig. 5. Cumulative distribution function plots showing packet inter-arrival time and packet size. There are around 5 discrete inter-arrival times and 3 discrete packet sizes.

Expert Framework Forecaster

The experts framework [9] is an on-line machine learning algorithm in which the output at each prediction trial depends on the individual outputs of a set of experts. All experts are given identical input data which they use to predict an output. The goal in the experts framework is to have the output of the overall algorithm track that of the best performing expert. An expert can be a complex algorithm that processes the input data to predict the output or it can be a simple function or even a constant that always gives the same prediction. Each expert has a weight that is multiplicatively updated at each trial of the algorithm. An expert's weight is reduced proportionally to how poorly it predicts. The output of the overall algorithm at each trial is the weighted average of the experts' outputs. The variable-share expert algorithm is used here to help experts that begin to forecast accurately quickly regain their weight.

The forecasting algorithm we designed operates on the assumption that packet inter-arrival times are persistent over time. This assumption is based on empirical observations of traffic generated by applications that fall in the category described above, i.e., high date rate, real-time. Figure 5 shows cumulative distribution function (CDF) plots of the packet inter-arrival time and packet size properties of a Skype audio trace. The trace contains one direction of the Skype conversation and was recorded using the Wireshark [6] traffic monitoring tool at the source. The CDF plots reveal that only a few discrete values of inter-arrival time and packet size dominate the distributions. As home networks evolve, we expect to see many applications that, like Skype, are rich with application-dependent patterns that can be exploited by a traffic forecaster such as the one we outline here.

With MAC layer traffic forecasting, the time-scale of interest is rather small. It is determined by the size of a superframe in the protocol. Given that at least two superframe durations are required to propagate traffic information to a 2-hop neighborhood, the time scale of interest is around two superframe durations, e.g. in DYNAMMA it would be $320ms$.

The forecasting algorithm is designed around a schedule-based MAC's time structure. More specifically, the notions of superframe and time slot have been built into its operation. In a real implementation, the forecaster will run in real time and can be implemented in hardware for speed and efficiency. Each forecast trial happens at the beginning of a superframe. At each trial, the algorithm evaluates the forecast of each expert and comes up with a new value for the best inter-slot duration. The experts' forecasts are evaluated using a loss function that takes into account both how many slots that forecast would waste and what the average per-packet delay would be. The forecast of each expert is a constant value in this algorithm and ranges from 1 to the number of slots per superframe. For example, a forecast of 1 corresponds to requesting every slot for the flow.

Forecasting Results

Our forecaster was run on the trace depicted in Figure 5. Figure 6 shows that the forecaster accurately predicts the slot period that will lead to best slot usage and lowest average delay. In this case, the best slot period corresponds to the inter-arrival time indicated by the the 0.01s dark line in the top plot of Figure 5.

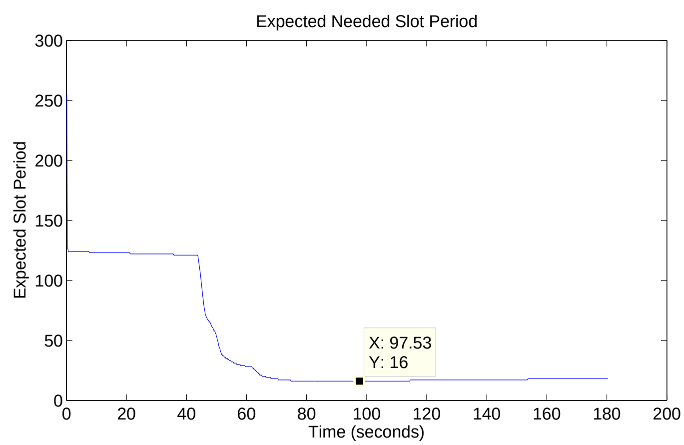

Fig. 6. Forecaster output: prediction of slot period required to service a flow that produces lowest delay and wasted slots. Given that the slot duration was set at $638.125\mu s$, a slot period of 16 corresponds to $10.210ms$ and this is exactly the inter-arrival time indicated by the lowest dark cluster in the top plot of Figure 5.

The average delay incurred by a packet is depicted in Figure 7. When the forecaster converges on the correct slot period, the average delay is around half the packet inter-arrival time. The forecaster constantly tries to reduce this delay while at the same time keeping the number of slots wasted low. Some amount of slot wastage must be traded in order to reduce the delay, and this can be tuned in the loss function that governs the operation of the experts algorithm. Figure 8 shows the slot usage of the algorithm to service the Skype audio flow. To reduce wasted slots, a backup transmitter can be elected along with the primary transmitter for each slot. The backup transmitter can use the slot if it does not detect the primary transmission within some timeout from the start of the slot.

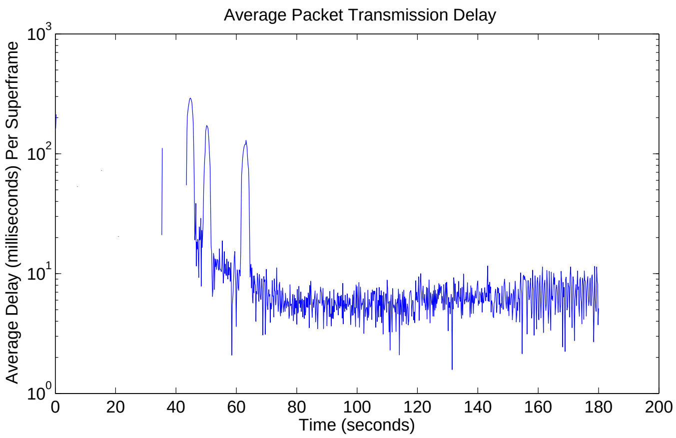

Fig. 7. Average delay incurred by the experts forecasting algorithm.

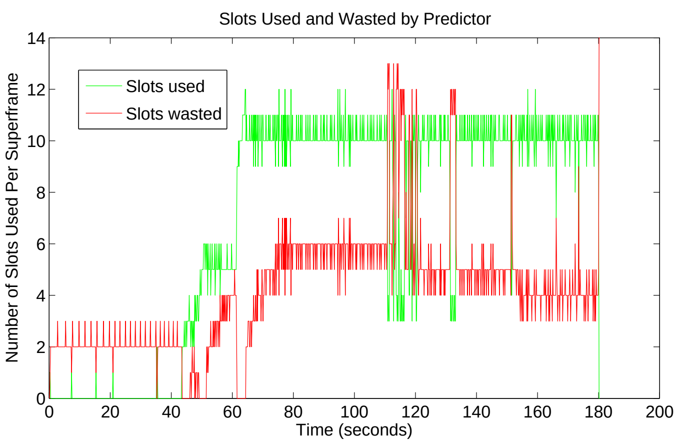

Fig. 8. Slots used and wasted by the forecaster over time.

V. Challenges and Future Work

The forecaster presented in Section IV works very effectively at forecasting the best slot period for a flow; it is therefore a promising approach to forecasting flow traffic at the MAC layer in order to improve schedule-based performance.

Thus far we have only investigated forecasting of packet arrival times, but there are other forecastable properties of the flows. For instance, we can see from Figure 5 that there is a trend in packet size as well. If packet size forecasting is used, time slot size can be reduced (it is currently sized for the maximum transmission unit) and multiple contiguous time slots can be given to flows to send an entire packet instead of single time slots. This way we can minimize unused channel time even further, while addressing stringent QoS requirements such as delay and delay jitter.

We have shown forecasting working in a simplified scenario: a single node running a single application generating a single flow. In realistic scenarios each node may be host to flows from multiple network applications and in a multi-hop-capable network each node must forward traffic on behalf of other nodes. An example of the targeted applications are real-time multimedia applications whose traffic exhibits strong pattern and persistence. Furthermore, in the applications we are interested in, single-hop activity is more common making multi-hop aggregation less of a concern.

In traditional schedule-based medium access protocols, traffic information needs to propagate around these 2-hop neighborhoods in order for nodes to be able to set up schedules for traffic to be sent. Traffic forecast information needs to propagate around 2-hop neighborhoods in a similar fashion. The length of time required for this propagation to complete is implementation dependent, but in typical superframe-structured protocols it will take 2 superframes. The signaling done during the first superframe propagates the forecast to the 1-hop neighbors, and the signaling done during the second superframe propagates it to all the 2-hop neighbors. This non-negligible delay must be considered because if the traffic pattern does not persist for longer than this period, the forecast will be out of date by the time it is propagated.

References

[1] A. Adas. Using adaptive linear prediction to support real-time vbr video under rcbr network service model. *Networking, IEEE/ACM Transactions on*, 6(5):635–644, Oct 1998.

[2] Y. Afek, M. Cohen, E. Haalman, and Y. Mansour. Dynamic bandwidth allocation policies. *INFOCOM '96. Fifteenth Annual Joint Conference of the IEEE Computer Societies. Networking the Next Generation. Proceedings IEEE*, 2:880–887 vol.2, Mar 1996.

[3] L. Bao and J. Garcia-Luna-Aceves. Hybrid channel access scheduling in ad hoc networks. *Network Protocols, 2002. Proceedings. 10th IEEE International Conference on*, pages 46–57, Nov. 2002.

[4] L. Bao and J. J. Garcia-Luna-Aceves. A new approach to channel access scheduling for ad hoc networks. In *MobiCom '01: Proceedings of the 7th annual international conference on Mobile computing and networking*, pages 210–221, New York, NY, USA, 2001. ACM.

[5] S. Chatziperis, P. Koutsakis, and M. Paterakis. On achieving accurate call admission control for videoconference traffic transmission over wireless cellular networks. pages 3221 –3226, mar. 2007.

[6] G. Combs. Wireshark network protocol analyzer. http://www.wireshark.org/.

[7] ECMA International. High rate ultra wideband PHY and MAC standard. Technical report, ECMA International Standard, www.ecma-international.org, Dec 2005.

[8] Y. Freund, D. Haussler, D. Helmbold, and R. Schapire. How to use expert advice. *Journal of the ACM*, Jan 1997.

[9] M. Herbster and M. Warmuth. Tracking the Best Expert. *Machine Learning*, 32(2):151–178, 1998.

[10] P. Koutsakis, M. Paterakis, and S. Psychis. Call admission control and traffic policing mechanisms for the wireless transmission of layered videoconference traffic from mpeg-4 and h.263 video coders. volume 5, pages 2155 – 2159 vol.5, sep. 2002.

[11] H. Liu and Y. Zhao. Adaptive edca algorithm using video prediction for multimedia ieee 802.11e wlan. pages 10 –10, jul. 2006.

[12] V. Rajendran, J. Garcia-Luna-Aceves, and K. Obraczka. Energy-efficient, application-aware medium access for sensor networks. *IEEE Conference on Mobile Ad-hoc and Sensor Systems (MASS)*, Jan 2005.

[13] V. Rajendran, K. Obraczka, and J. Garcia-Luna-Aceves. Energy-efficient collision-free medium access control for wireless sensor networks. *SenSys '03: Proceedings of the 1st international conference on Embedded networked sensor systems*, Nov 2003.

[14] V. Rajendran, K. Obraczka, and J. Garcia-Luna-Aceves. Dynamma: A dynamic multi-channel medium access framework for wireless ad hoc networks. *Mobile Adhoc and Sensor Systems*, Jan 2007.

[15] S. Shah, K. Chen, and K. Nahrstedt. Dynamic bandwidth management for single-hop ad hoc wireless networks. *Pervasive Computing and Communications, 2003. (PerCom 2003). Proceedings of the First IEEE International Conference on*, pages 195–203, March 2003.