

On an Efficient Random Access Scheme for Capillary Machine Type Communication

Kazi Ashrafuzzaman and Abraham O. Fapojuwo

Department of Electrical and Computer Engineering, University of Calgary, Calgary, AB, Canada

email: {kazi.ashrafuzzaman, fapojuwo}@ucalgary.ca

Abstract—In capillary machine type communications, random multi-access lends itself as a natural choice at the link layer due to the largely intermittent, event-driven traffic involved. A major challenge to that end arises, however, from the anticipated large scale of the network. Considering a carrier sense multiple access based random access protocol where infrastructural support is available in the form of a coordinator node, this paper presents a framework to help sustain efficient protocol performance when faced with temporal variation in packet traffic. The coordinator dynamically determines a timely contention parameter by means of an online algorithm based on certain analytic characterizations of the protocol performance, and the end-nodes get opportunistically notified of it. Throughput, fairness and delay-variation performance produced by the proposed scheme turns out to be consistently close to an optimal protocol in identical settings, in contrast with that of the binary exponential backoff based solutions.

Index Terms—Carrier sense multiple access; Random Access; Protocol Optimization;

1. Introduction

Aside from the more conspicuous high-rate consumer data networking, a vital component of the evolving infrastructure for the future networked systems is concerned with *capillary communication* among densely located nodes that are largely reduced-function and low-power in nature. Within the larger Internet of Things (IoT) paradigm [1], this category of networked communication constitutes the underlying enabling technology upon which smart environments are to be constructed [2], [3]. Intelligent pervasive systems and applications making such development happen rely on largely self-organized computing and communication tasks that overlap in regard to data acquisition, sensing and control activities (e.g., within home and building area, intelligent energy management systems, ambient intelligence and occupancy detection/prediction, security systems, heating, ventilation, and air conditioning control integrated with renewable generation etc.). Accordingly, general purpose networking primitives are called for, rather than the application-specific approaches often adopted in sensor networking in the past.

A crucial attribute in the range of applications involved is the prevalence of event-driven communication traffic. This

leads to the requirement that the link layer medium access coordination solution be in the form of a random access protocol, which comes with the kind of distributed operational simplicity that can take care of the intermittent node activities of the event-driven traffic. However, it has long been noted in protocol performance analysis that the random access mechanisms do not perform well when the scale grows [4], as is the case in the present setting. Crucially, though, associated IoT infrastructure is poised to have some full-function devices to act as gateways or coordinators in connecting the reduced-function end-nodes with the upper level nodes of the network hierarchy orchestrating the applications and services.

If organized to operate at its maximum equilibrium efficiency, a *carrier sense multiple access* (CSMA) protocol can still be feasibly adequate in terms of the deliverable performance in most practical settings. The goal of this work is to achieve such efficiency by means of a collision avoidance (CA) scheme that is realized by appropriately leveraging the coordinator function. Specifically, for a familiar slotted CSMA variety typically considered in networked communication among power-constrained nodes, an online adaptive algorithm is presented. The algorithm has as its basis an analytic characterization that equips the local network coordinator to dynamically facilitate contention throttling using the recent channel activity trace as input. Subsequently, the contextually proper contention parameter is determined that is opportunistically broadcast to achieve sustained efficiency of the access scheme in the local area neighborhood. The intermittently active end-nodes can continue to act in distributed fashion apart from the occasional message passing from the coordinator involved in getting notified of this contention parameter. Against traffic scenarios varied in terms of elasticity and scale of network contention, the proposed system attains high throughput in an agile manner, as opposed to the usual alternative, the *binary exponential backoff* (BEB), which fails to perform with high channel utilization while maintaining fairness and/or low delay variation. Besides, the solution retains the desirable feature of scalability in the sense that when larger networks are confronted, the CSMA channel access efficiency is not compromised substantially.

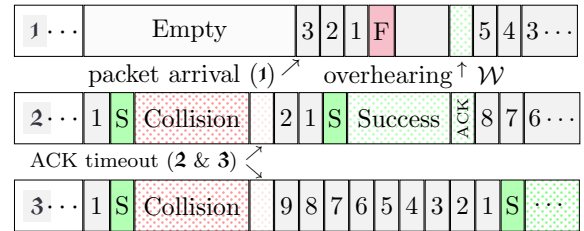
Paper organization. The rest of the paper is organized as follows: §2 next positions the present work in context with a brief review of the pertinent literature. §3 describes

the system model considered and the assumptions made before presenting, in § 4, the analytic performance model and the associated efficiency-sustaining algorithm. § 5 specifies the simulation setup and the evaluation scenarios, followed by a discussion of the numerical results in § 6. Concluding statements are made in § 7.

Beginning with the classical works on random access protocols, the mechanisms to enforce channel access efficiency have been intensively studied (see [4] for a summary). Pseudo-Bayesian stabilization [5] and related approaches [6] depend on knowing the arrival process within precise bounds, leaving the algorithmic case of dynamic arrival open. In recent era, motivated by the increasing scale of the networks and the diversity in expectation on the related networking primitives, researchers have proposed solutions to cope with large networks while addressing the demands on the service primitives. In a cellular-based machine-to-machine framework, [7] has presented a hybrid MAC protocol for communication involving access networks at massive scale. However, the paper is based on the assumption that the distribution of the packet arrival process in the nodes is known to the base station. An attempt to get the CSMA to minimize collisions is in [8], but it is rather specific to application types where spatial locality induces redundancy and suppression of related contention is desirable (*e.g.*, sensor network for fire-detection). Approaches to solutions involving low-power personal area and/or sensor networking have been proposed in [9], [10], [11], [12], which address standard specific beacon enabled CSMA/CA operation. This paper complements these works in that the proposed mechanism is amenable to be incorporated within them to yield robust estimation of more general arrival processes. And it contrasts with them in seeking to bring the core CSMA operation within an optimizing and adaptive framework instead of fine-tuning the parameters of a given specification, facilitating the handling of dynamic arrival processes.

the analytic treatment in this paper. In this backdrop, the contribution of this paper is in devising an efficient collision avoidance mechanism for the slotted CSMA system suitable for infrastructure-based local area access control considering intermittent, dynamic node activity and packet arrival. An analytic framework provides the underpinnings in the estimation within the adaptive procedure yielding near optimal aggregate throughput without disproportionately losing in fairness or jitter performance.

From within an IoT infrastructure where many protocols at different layers and multiple media at different link levels are potentially involved, of interest in this work is the machine type communication in an access network comprised of mostly reduced function communication nodes. Specifically, keeping to the medium access control layer's operational scope, a single-hop local area network in star topology is considered; a total of N end-nodes are associated with a resourceful coordinator which acts as the gateway between these nodes and the applications running in the wider networked systems. As the focus of this work is on the MAC layer, no PHY-layer specific requirement or facilitation is made except that carrier sensing is assumed to be available and that collision detection is infeasible. To focus mainly on the effectiveness of the access mechanism, neither packet capture in the event of collisions nor channel-error induced packet losses have been incorporated.



3.1. CSMA Procedure

Time is viewed as succession of specified unit length called slots. Compared to the likes of IEEE 802.11

CSMA/CA, such slots are longer for low-data rate systems, essentially longer than the inter-frame spacing between a packet and the corresponding ACK packet, thus protecting the ACK from potential collision with subsequent data packets from other nodes. Additionally, capillary systems have shorter packets, and are energy-conserving by design with contending nodes necessarily getting to know of the channel activity states only through explicit carrier sensing.

In the considered scheme, a slot is assumed to be able to accommodate an ACK packet, which is presumably short (this assumption is not a requirement for the functioning of the proposed framework, however). Also in the event of packet collision, the colliding nodes get to learn of the failure in delivering the packet as timeout for receiving the ACK occurs, which spans the single slot following the failed transmission. A crucial element in the process is that the ACK packet piggybacks a number: a backoff window size that is appropriate to the present CSMA channel load. Apart from the node engaged in the ongoing successful transmission, the ACK packet is also overheard and the window size information is therefore opportunistically extracted by the nodes doing channel sensing (the idea of resorting to piggybacking to convey similarly critical information has previously been employed in, *e.g.*, [17]).

As illustrated in Fig. 1, node 1 has its buffer empty for a while and once there is arrival of packet(s), it initially draws a random backoff value from a default window (*e.g.*, $[0, 2^3]$). As the channel sensing following this backoff is reported in failure (*i.e.* channel sensed busy), node 1 freezes its procedure until the ongoing transmission finishes and overhears the associated ACK that piggybacks the appropriate backoff window size $\mathcal{W}$. In the following cycle of attempt, it draws a backoff value from $[0, \mathcal{W}]$. Earlier in the activity trace, both nodes 2 and 3 incidentally finished backing off and conducted successful channel sensing (*i.e.*, channel sensed idle) at the same slot leading to concurrent transmissions that resulted in a collision on the channel. Both nodes 1 and 2 waited the following slot in ACK timeout (when other nodes could potentially conduct carrier sensing). Nodes that are possibly conducting carrier sensing during a colliding transmission would also see out, following the sensing failure, the remaining slots in transmission and would subsequently start the next backoff drawing from the last known window size. In the later steps, before node 3 succeeds, node 2 manages to transmit its packet successfully, during which node 1 happened to have found the channel to be busy.

3.2. Traffic model

Each node becomes and remains active contender as long as it has packets in its buffer handed down from the upper layer and it remains inactive otherwise. Number of estimated contending/active nodes at a point in time is denoted by n ($n \leq N$).

The evaluation carried out in this work is intended to be within the stable regime of the assumed CSMA/CA protocol. Hence buffer-sizing and packet drop due to potential buffer

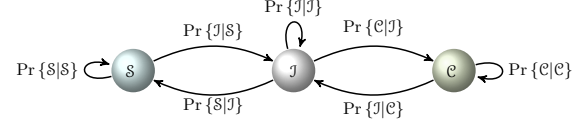


Figure 2: State transition diagram of the DTMC for the CSMA channel.

overflow are outside the scope of this study. Accordingly, Bernoulli trials are performed both at the end of a successful transmission by the contending node concerned, with probability p of having further packet(s) to contend with, and by the empty nodes at every slot, with probability q for each to encounter a non-empty packet-buffer.

Though this traffic model is homogeneous across the nodes, the variation in burstiness it incorporates essentially captures the aggregate effects on the channel that could emerge from heterogeneous load.

4. Adaptive Algorithm with Analytic Framework

This section articulates the main contribution of this paper – an analytic procedure for enforcing a contextually appropriate collision avoidance algorithm within the CSMA protocol specified in § II-A. First, a probabilistic model of the protocol performance is presented considering a saturated network of n active/contending nodes. While the number of active nodes at a given point in time is random with intermittent event driven traffic, the model can be utilized to map from the channel events observed to estimated number of contending nodes within the larger unsaturated network. An adaptive algorithm then seeks to throttle the contention to a level that is in the vicinity of the optimal operating point. These ideas drive the following parts in this section where expressions for the associated quantities are deduced before plugging them into the adaptive algorithm to estimate n and, therefore, $\mathcal{W}$.

4.1. Analytic Performance Model at Saturation

Let β be the rate at which each of the n contending nodes in a network are making channel sensing attempts. The channel is to be found in either of the three states at a given slot, namely, *idle*, *successful transmission* or *colliding transmission*, denoted by $\mathcal{J}$, $\mathcal{S}$ and $\mathcal{C}$, respectively. Fig. 2 captures the evolution of the stochastic process on the channel in a discrete-time Markov chain (DTMC). From the state $\mathcal{J}$ of the channel at a particular slot, a transition into the state $\mathcal{S}$ in the next slot occurs if there were a single node conducting channel sensing for possible packet transmission. On the other hand, the channel remains in the state $\mathcal{J}$ in the next slot if no node makes any sensing attempt. The associated

probabilities for the two transitions¹ are $n\beta(1-\beta)^{n-1}$ and $(1-\beta)^n$, respectively. Complementing the above two events is the transition into the state $\mathcal{C}$ if there were more than one node concurrently sensing the channel instead, occurring with probability (w.p.) $1 - (1-\beta)^n - n\beta(1-\beta)^{n-1}$.

The channel remains in the states $\mathcal{S}$ and $\mathcal{C}$ based on the packet length, which is considered to be L slots in the model. Observed uniformly, a busy slot in successful transmission can be any of the $L+1$ slots (including 1 slot for ACK). Hence there is a $1/(L+1)$ probability that the transmission ends in the observed slot with the next slot becoming idle, or the channel remains busy with probability $L/(L+1)$. Likewise, in the case of a colliding transmission, the probabilistic counterparts of transition are $1/L$ and $(L-1)/L$, respectively (as the ACK timeout slot can be a sensing slot for a potentially contending node).

The transition probabilities for the DTMC are thus established as follows

$$\Pr\{\mathcal{S}|\mathcal{J}\} = n\beta(1-\beta)^{n-1} \quad (1a)$$

$$\Pr\{\mathcal{J}|\mathcal{J}\} = (1-\beta)^n \quad (1b)$$

$$\Pr\{\mathcal{C}|\mathcal{J}\} = 1 - (1-\beta)^n - n\beta(1-\beta)^{n-1} \quad (1c)$$

$$\Pr\{\mathcal{J}|\mathcal{S}\} = \frac{1}{L+1} \quad (1d)$$

$$\Pr\{\mathcal{S}|\mathcal{S}\} = \frac{L}{L+1} \quad (1e)$$

$$\Pr\{\mathcal{J}|\mathcal{C}\} = \frac{1}{L} \quad (1f)$$

$$\Pr\{\mathcal{C}|\mathcal{C}\} = \frac{L-1}{L} \quad (1g)$$

Now, considering that n is relatively large while β is small ($\ll 1$), the remaining analysis can be simplified by approximating

$$(1-\beta)^{n-1} \approx (1-\beta)^n \approx e^{-n\beta} = e^{-g} \quad (2)$$

where g represents the aggregate attempt rate on the CSMA channel such that

$$g = n\beta \quad (3)$$

Accordingly, the stationary state probabilities of the DTMC $\{\pi_i, \pi_s, \pi_c\}$ can be derived from the balance equations

$$\pi_i(\Pr\{\mathcal{S}|\mathcal{J}\} + \Pr\{\mathcal{C}|\mathcal{J}\}) = \pi_s \Pr\{\mathcal{J}|\mathcal{S}\} + \pi_c \Pr\{\mathcal{J}|\mathcal{C}\} \quad (4a)$$

$$\pi_s \Pr\{\mathcal{J}|\mathcal{S}\} = \pi_i \Pr\{\mathcal{S}|\mathcal{J}\} \quad (4b)$$

$$\pi_c \Pr\{\mathcal{J}|\mathcal{C}\} = \pi_i \Pr\{\mathcal{C}|\mathcal{J}\} \quad (4c)$$

1. For more exact expressions and associated representation, idle slots would have to be covered through two different states due to a slight variation in the number of nodes potentially involved in carrier sensing activity: while n nodes would be involved probabilistically in sensing in a slot that is preceded by an idle slot or a successful transmission, the specific slot that follows a collision would likely involve $n-2$ nodes or fewer as the 2 or more colliding nodes are timing out in expectation of ACK in that slot. The simplification involved in the condensation of the idle states (as well as the approximation in (2) later) in the present formulation is analytically justified thanks to the asymptotic convergence associated, and is also a fairly reasonable approximation numerically with the larger values of n that are of interest in the capillary setting.

along with the normalization relation

$$\pi_i + \pi_s + \pi_c = 1 \quad (5)$$

as follows

$$\pi_s = \frac{g(1+L)}{e^g(L+1) - L + g} \quad (6)$$

$$\pi_c = \frac{(e^g - g - 1)L}{e^g(L+1) - L + g} \quad (7)$$

$$\pi_i = \frac{e^g}{e^g(L+1) - L + g} \quad (8)$$

Apart from β , the other key modeling parameter is the probability that a given sensing attempt by an arbitrary node will fail to find the CSMA channel idle. Let ϕ denote this sensing failure probability.

Given that an arbitrary sensing attempt on the part of a given node will fail whenever the channel is in transmission (*i.e.* not idle), ϕ can be expressed as

$$\phi = \pi_s + \pi_c = 1 - \pi_i = \frac{L(e^g - 1) + g}{e^g(L+1) - L + g} \quad (9)$$

4.1.1. Attempt rate by an arbitrary active node. The state of a contending node is modeled as a *renewal reward process* where the number of channel sensing is counted as rewards. A single cycle of the process associated with a reward of 1 is comprised of

- i) uniformly backing off by a random number of slots, with the expectation of this period being $\mathcal{W}/2$ slots given that it is drawn from a single window of size $\mathcal{W}$;
- ii) attempt to sense the channel idle using 1 slot; and
- iii) either to wait for the ongoing transmission to finish (including overhearing the value of $\mathcal{W}$ in the ACK) in the event of busy channel (w.p. ϕ) or else to transmit its packet (w.p. $1 - \phi$); the expected length in this part is therefore $\phi \frac{L}{2} + (1 - \phi)(L + 1)$.

An arbitrary node's sensing attempt rate β can now be expressed as the following *renewal rate* of this process

$$\beta = \frac{1}{\frac{\mathcal{W}}{2} + 1 + \phi \cdot \frac{L}{2} + (1 - \phi)(L + 1)} \quad (10)$$

The modeling problem addressed above has similarity with the ones studied in [18], [19], and the outcome is also fundamentally consistent with classical CSMA analyses [20].

4.2. Contention Estimation within Unsaturated Network and Optimization

The performance model presented in § III-A can be employed to discern into the realtime network contention load exerted through the intermittent node activity within the larger unsaturated network. The throttling of the effective contention then can be done according to the attempt rate deduced for optimal operating point. The method for estimation of the present contention is also assisted in part by the knowledge of the said optimal point.

The detail steps to that end are produced below.

4.2.1. Estimating Contention Level from Channel Event Trace. From the total of N nodes associated with the coordinator, the number of active/contending nodes, n , has to be periodically estimated and correspondingly, appropriate $\mathcal{W}$ is to be determined.

Let $\mathcal{O}$ be the recent most observation window from the immediate past slot-wise trace of the channel states, and k_i , k_c and k_s represent the proportion of the idle, colliding and success slots, respectively, within $\mathcal{O}$. In implementations, $|\mathcal{O}|$, the length of $\mathcal{O}$, can be chosen to have a balance between weighing the recently observed states in contention and keeping to the trend based on the longer record. The proportions k_i , k_s and k_c represent the short-term sampled measures of the steady-state quantities π_i , π_s and π_c , respectively. These short-term observations are used to make corresponding estimates of g , denoted $g_s(k_s)$, $g_c(k_c)$, $g_i(k_i)$, from the inverse relations of (6), (7) and (8), in that order.

Inverse functions $g_c(k_c)$ and $g_i(k_i)$, respectively, can be deduced as follows

$$g_c(k_c) = \frac{L(k_c - 1)}{L + k_c} - \mathbb{W}_{-1} \left(\frac{k_c(1 + L) - L}{L + k_c} e^{\frac{k_c - 1}{L + k_c}} \right) \quad (11)$$

$$g_i(k_i) = L - \mathbb{W}_0((1 + L - 1/k_i)e^L) \quad (12)$$

where $\mathbb{W}_0(\cdot)$ and $\mathbb{W}_{-1}(\cdot)$ are the principal and the secondary real branches, respectively, of the Lambert-W function [21].

On the other hand, the inverse relation of π_s from (6) is not a unique function. Rather, in correspondence with the bi-stability property of the CSMA system [15], there are two separate inverse functions partitioned in range across the point of convergence g^* (deduced later in the sequel in (17)). Denoting by $g_s^-(k_s)$ the function for the lower range of g ($g \leq g^*$), and by $g_s^+(k_s)$ the higher range ($g \geq g^*$), the following inverse mappings can be inferred from (6)

$$g_s^-(k_s) = -\frac{Lk_s}{L - k_s + 1} - \mathbb{W}_0 \left(-\frac{k_s(1 + L)}{L - k_s + 1} e^{-\frac{Lk_s}{L - k_s + 1}} \right) \quad (13a)$$

$$g_s^+(k_s) = -\frac{Lk_s}{L - k_s + 1} - \mathbb{W}_{-1} \left(-\frac{k_s(1 + L)}{L - k_s + 1} e^{-\frac{Lk_s}{L - k_s + 1}} \right) \quad (13b)$$

When g^* is applied to (8), the proportion of the idle slots, k_i^* , can be known for the occasion of the throughput optimal operation. If k_i turns out to be larger than k_i^* , the channel is underutilized and $g_s^-(k_s)$, given by (13a), is meant to be representing $g_s(k_s)$ correctly, otherwise $g_s^+(k_s)$ from (13b) has to be called for. Thus the estimate $g_s(k_s)$ can be unambiguously decided. With $g_s(k_s)$, $g_c(k_c)$ and $g_i(k_i)$ calculated, corresponding estimates of n_s , n_c and n_i can be made as $g_{(\cdot)}(k_{(\cdot)})/\beta$ according to the equation (3), once the running value of β is obtained. Recall that employing (10), the running value of β for the traced duration can be estimated using the previous estimate of $\mathcal{W}$, and the value $\phi = k_s + k_c$ observed in the immediate past $\mathcal{O}$.

The overall estimate of the contending nodes $\tilde{n}$ can now be made by a relation akin to multiple linear regression from the endogenous variables n_i , n_c and n_s using k_i , k_c and k_s as the corresponding coefficients

$$\tilde{n} = n_i k_i + n_s k_s + n_c k_c \quad (14)$$

And lastly, the updated $\mathcal{W}$ can be calculated as

$$\mathcal{W} = 2 \left[\frac{\tilde{n}}{g^*} - 1 - \phi^* \frac{L}{2} - (1 - \phi^*)(L + 1) \right] \quad (15)$$

where

$$\phi^* = 1 - k_i^* = \frac{(e^{g^*} - 1)}{e^{g^*}(1 + 1/L) - 1} \quad (16)$$

is the sensing failure rate analytically known for the throughput-optimal operations.

4.2.2. Condition for Throughput Optimality. The aggregate attempt rate that maximizes the stable throughput can be found using the first order conditions on equation (6). Taking the derivative from (6) with respect to g and setting it to zero leads to the optimal attempt rate g^* given by the following

$$g^* = 1 + \mathbb{W}_0 \left(-\frac{L}{e(1 + L)} \right). \quad (17)$$

4.3. Computing $\mathcal{W}$ for Collision Avoidance

Armed with the analytic characterization presented in this section so far, the collision avoidance procedure is based on periodically repeated computation by the coordinator in accord with the following algorithm:

loop

Step 1: yield k_i, k_c, k_s from updated $\mathcal{O}$ out of the channel event trace with $|\mathcal{O}| = (\frac{1}{3}N + \frac{2}{3}\tilde{n})L$
 $\phi = k_s + k_c$; update β per (10) using $\mathcal{W}$ from previous iteration.

Step 2: $g_c \leftarrow g_c(k_c)$, $g_i \leftarrow g_i(k_i)$, $g_s^- \leftarrow g_s^-(k_s)$, $g_s^+ \leftarrow g_s^+(k_s)$ per (11), (12) and (13)

if $k_i \leq k_i^*$ **then**

$g_s \leftarrow g_s^+$

else

$g_s \leftarrow g_s^-$

end if

Step 3: $n_c \leftarrow g_c/\beta$, $n_i \leftarrow g_i/\beta$, $n_s \leftarrow g_s/\beta$

update $\tilde{n}$ per (14)

update $\mathcal{W}$ per (15) using ϕ^* from (16) and $\tilde{n}$

end loop

Recall that the coordinator is invariably active, remaining aware of the channel activity trace, and is responsible to run the above algorithm. The coordinator here updates $\mathcal{W}$ based on its recent trace history $\mathcal{O}$ of length $|\mathcal{O}| = (\frac{1}{3}N + \frac{2}{3}\tilde{n})L$ slots²

The estimation algorithm is executed every L slots to update $\mathcal{W}$, thereby accounting for every possible transmission event on the CSMA channel in the estimation and the associated update procedure (other choices of this length are possible and, the performance of the algorithm turns out to

2. Alternative choices of the trace history length would also be possible; this particular length seeks to attain a weighted balance between exploitation of presently estimated information ($\frac{2}{3}\tilde{n}$) and emphasizing longer term observation in expectation of steady behavior regarding the activity pattern among the nodes ($\frac{1}{3}N$).

be only slightly sensitive to moderate changes in this update interval).

Complexity and scalability. The above algorithm has a constant-time complexity $O(1)$ with the number of nodes as input size while it is linear time with a complexity $O(|\mathcal{O}|)$ in regard to the length of the trace. The algorithm therefore has maximum scalability as adding more nodes does not bring extra computational costs. Also the k_i, k_c, k_s values can be updated in an online manner by the coordinator that remains continuously active, rendering the temporally linear complexity of the algorithm to be efficient enough to incur no additional overhead.

5. Simulation

The efficacy of the proposed mechanism has been evaluated, in comparison with the BEB and an optimal distributed protocol, by means of a packet level discrete-event simulator developed in C++. The candidate access schemes evaluated and the configurations considered in the simulation are described in the following.

5.1. Candidate access schemes

The proposed collision avoidance mechanism (labeled “proposed” in the plots) is equipped with the adaptive estimation procedure articulated in the previous section. The BEB based mechanism (“BEB”) is considered with an exponent range³ of [3, 12] that is wide enough to cope with larger scales of the network addressed. The hypothetical access scheme operates with backoff window sizes that are calculated on the basis of the exact knowledge of the number of active nodes at each point in time and thus represents an optimal distributed CSMA protocol (“OPT”). Apart from the backoff window size determination, the rest of the protocol behavior remains the same for all three access schemes.

5.2. Configurations

For a broader outline of pertinent use cases with machine type communication scenarios, consider a net-zero Carbon home or building equipped with applications and services to create smart spaces; in such context, changes at different hours in user activity and mobility as well as in environmental indicators can impart varying burstiness in communication of sensing data and control packets pertaining to tasks such as occupancy detection, ambient intelligence, appliance control and situational awareness. Accordingly, to include the variation of burstiness in numerical study, two traffic scenarios (cf. § II-B) are considered:

- scenario *i* with $(p, q) = (0.1, 0.001)$ has largely steady node activity, and

3. With each sensing failure in a transmission attempt, backoff value is uniformly drawn from a window size that is twice the size of the previous one, with the initial size and the upper bound being 2^3 and 2^{12} , respectively. Following a sensing success, the subsequent backoff starts with the initial window size.

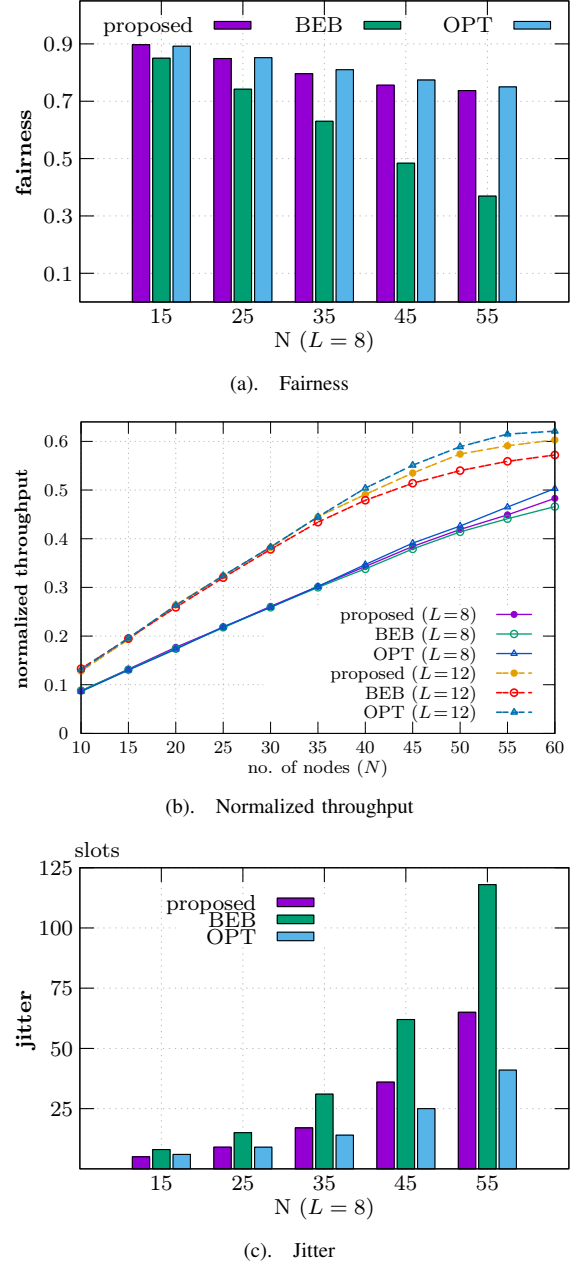
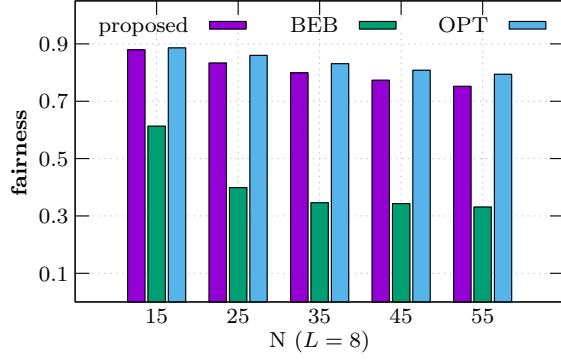


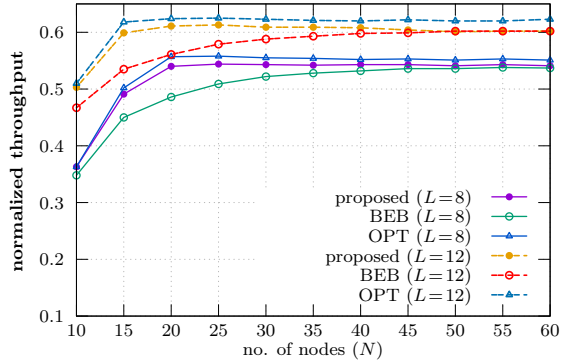
Figure 3: Fairness, throughput and jitter performance: scenario *i*. with $(p, q) = (0.1, 0.001)$.

- scenario *ii* with $(p, q) = (0.01, 0.005)$ captures relatively more fluctuation in node activity;

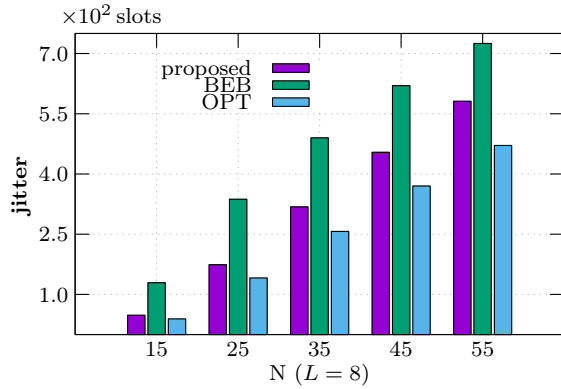
That is, active nodes are more likely to remain active in scenario *i* as opposed to scenario *ii* which sees idle nodes to be more likely to become temporarily active. Note that packet arrival is modulated by a semi-Markov process, pertaining to the active state of the node specifically where the interval between successive packet arrival events is de-



(a). Fairness



(b). Normalized throughput



(c). Jitter

Figure 4: Fairness, throughput and jitter performance: scenario *ii*. with $(p, q) = (0.01, 0.005)$.

pendent on packet service time⁴. In addition to the aggregate throughput indicating CSMA channel efficiency, *jitter* and *fairness* are calculated for each case as averages from 100 iterations of simulations each spanning 10K slots. Jitter is quantified as the standard deviation of the packet service times, taken as an average weighted against the number of packets transmitted by each node in the network. Standard

4. Packet service time is defined here as the time between the point (*i.e.* the slot) at which a packet arrives at the node's MAC buffer and the point in time (in slot) when its successful transmission finishes.

deviation of the mean packet service times experienced by the N nodes comprising the network is made use of in calculating the measure of fairness, according to [22], in granting channel access opportunities to the nodes. Quantitative evaluations are made for $N = 10$ to 60 to illustrate how the performance of each of the access schemes fare. Two different packet sizes, $L = 8$ and 12, are considered to account for packet size sensitivity to throughput and, for brevity, fairness and jitter results are presented for $L=8$ case only.

6. Discussion

Interpretation of the results obtained from the simulation experiments specified in the previous section yields a range of performance characteristics, as presented below.

Figures 3 and 4 show the results from the traffic scenarios *i* and *ii*, respectively. For all three candidate access schemes evaluated, as channel access contention deepens with larger N , jitter rises monotonically while aggregate throughput grows till reaching around the achievable peaks. Correspondingly, fairness continues to get reduced though the rate is considerably higher in the case of the BEB, unlike the other two.

With respect to aggregate throughput, there is very little to separate the three access schemes. The BEB remains within 7.4% ($L=8$) and 9.1% ($L=12$) in scenario *i*, and within 12.8% and 13.4% in scenario *ii* of corresponding OPT performances. The respective numbers for the proposed mechanism are 4% ($L=8$) and 3.9% ($L=12$) with scenario *i*, and 3% and 3.4% ($L=12$) with scenario *ii*. Thus the near optimality of both in regard to the throughput performance is clearly evident.

Performance of the proposed scheme sustains with significantly higher fairness and lower jitter than the BEB in both traffic scenarios. The proposed scheme achieves fairness index values within 3.4% (scenario *i*) and 6.2% (scenario *ii*) of the OPT whereas the BEB sees gaps as much as 52% (scenario *i*) and 58% (scenario *ii*) with the OPT. Likewise, the proposed mechanism keeps the jitter within 62% (scenario *i*) and 24% (scenario *ii*) of OPT as opposed to 194% and 231%, respectively, in the case of the BEB. The reason for the high jitter and the low fairness attained by the BEB in general is ingrained in the way the BEB enforces its collision avoidance procedure. Specifically, successful and newer nodes are prone to being rewarded and nodes unsuccessfully contending for long periods are prone to get punished in delay measures. Thus the range of dispersion in granting the packet transmission opportunities among the nodes and in time for a given node becomes exponentially wider, leading to the significant degradation in fairness and jitter performance.

For smaller size of networks and/or with fairly low arrival rates⁵ as in scenario *i*, the extent of BEB's performance

5. Note that the semi-Markov modulation of the arrival process is more sensitive to increment in q than in p , as the interval between packet arrival events in the idle state of a node is independent of packet service times determined by present contention, unlike the case in the busy state.

loss in fairness and jitter is relatively moderate compared to what it incurs in contrary settings. In other words, the BEB shows a strong trade-off between a) throughput, and b) fairness among the nodes, and in time (*i.e.* jitter). This observation is consistent with the theoretical underpinnings reported in [23] concerning distributed CSMA operation without message passing.

The proposed scheme, on the other hand, resorts to minimal message passing to convey the backoff window size via opportunistic piggybacking and ensures that the nodes contend with (almost) equal opportunities and the resulting collective contention probability remains in close agreement with the achievable rate corresponding to CSMA channel capacity. Consequently, retention of high aggregate throughput is not compromised in achieving low variation in packet service times for individual nodes nor is there a trading away in equitable distribution of channel usage among the nodes.

Consistent with equation (6), there is higher throughput for larger L when other factors remain unchanged (there is a corresponding increase in jitter and marginal reduction in fairness observed numerically, understandably, due to the added contribution of each single event outcome in sensing and/or transmission attempts to jitter, the delay-variation measure). Traffic scenario *ii* involves more intermittent node activity than *i* and produces higher levels of jitter accordingly for all three access schemes. Based on the semi-Markov modulation of the packet arrival process, contention level is more sensitive to changes in q than p and hence scenario *i* reaches traffic saturation with larger number of nodes than *ii*, as reflected in the throughput curves (Fig. 3(b) and 4(b)). The BEB performs reasonably well in all measures only when there is little contention as in the lower regime of N in scenario *i*; it suffers regarding fairness whenever there is higher contention even though it achieves near-optimal throughput in aggregate. The marginal decrease in fairness and proportional increase in jitter with the proposed scheme occurs in the same pattern as with the OPT.

7. Conclusions

This paper has presented a framework for efficient random access by enforcing a collision avoidance mechanism that is devised on the basis of an adaptive online algorithm. Unlike the BEB, the mechanism does not seek to exploit certain operating conditions (e.g., traffic characteristics and scale of the network) at the expense of performance in other operating conditions or achieving one metric at an attractive level while incurring undesirable results in others. Remaining largely immune to the conditions, the solution provided ensures that the CSMA channel capacity is retained in all cases and is distributed rather equitably among the nodes as the vital network resource. This is particularly useful in the context of the highly intermittent or bursty contention in large scale networks supporting capillary machine type communication. The evolving IoT systems can provide the

minimal infrastructure support required for deploying⁶ the proposed scheme so as to extract the benefit of realizing timely and equitably efficient local area packet communication.

References

- [1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols and applications," *IEEE Communications Surveys Tutorials*, vol. PP, no. 99, pp. 1–1, 2015.
- [2] N. Accettura, M. Palattella, M. Dohler, L. Grieco, and G. Boggia, "Standardized power-efficient & internet-enabled communication stack for capillary M2M networks," in *IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, April 2012, pp. 226–231.
- [3] V. B. Mišić, J. Mišić, X. Lin, and D. Nerandzic, "Capillary machine-to-machine communications: the road ahead," in *Ad-hoc, Mobile, and Wireless Networks*. Springer, 2012, pp. 413–423.
- [4] D. P. Bertsekas and R. G. Gallager, *Data networks*, 2nd ed. Prentice-Hall Inc., 1992.
- [5] R. L. Rivest, "Network control by bayesian broadcast," *Information Theory, IEEE Transactions on*, vol. 33, no. 3, pp. 323–328, 1987.
- [6] F. P. Kelly, "Stochastic models of computer communication systems," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 379–395, 1985.
- [7] Y. Liu, C. Yuen, X. Cao, N. Hassan, and J. Chen, "Design of a Scalable Hybrid MAC Protocol for Heterogeneous M2M Networks," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 99–111, Feb 2014.
- [8] Y. Tay, K. Jamieson, and H. Balakrishnan, "Collision-minimizing CSMA and its applications to wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1048–1057, 2004.
- [9] P. Park, C. Fischione, and K. H. Johansson, "Adaptive IEEE 802.15.4 protocol for energy efficient, reliable and timely communications," in *Proceedings of the 9th ACM/IEEE international conference on information processing in sensor networks*. ACM, 2010, pp. 327–338.
- [10] N. Abdeddaim, F. Theoleyre, M. Heusse, and A. Duda, "Adaptive IEEE 802.15.4 mac for throughput and energy optimization," in *Distributed Computing in Sensor Systems (DCOSS), 2013 IEEE International Conference on*. IEEE, 2013, pp. 223–230.
- [11] H.-W. Tseng, A.-C. Pang, J. Chen, and C.-F. Kuo, "An adaptive contention control strategy for IEEE 802.15.4-based wireless sensor networks," *Vehicular Technology, IEEE Transactions on*, vol. 58, no. 9, pp. 5164–5173, 2009.
- [12] M. D. Francesco, G. Anastasi, M. Conti, S. K. Das, and V. Neri, "Reliability and energy-efficiency in IEEE 802.15.4/ZigBee sensor networks: an adaptive and cross-layer approach," *Selected Areas in Communications, IEEE Journal on*, vol. 29, no. 8, pp. 1508–1524, 2011.

6. From the implementation perspective, functional customizations involved are of moderate complexity. In particular, the coordinator's link-layer driver would have to be programmed appropriately, while the channel access behavior of the end-nodes would have to be modified in small but non-trivial ways (enabling extraction of piggybacked information of $\mathcal{W}$, overhearing ACK by sensing nodes). Still, end-nodes operating with traditional BEB *can coexist* within the proposed access scheme as long as the CSMA slot boundary is maintained; although, in such scenarios, these end-nodes would likely suffer from higher jitter and possibly reduce fairness in the network.

- [13] G. Bianchi and I. Tinnirello, "Kalman filter estimation of the number of competing terminals in an IEEE 802.11 network," in *Proc. of the Annual IEEE Joint Conference of the Computer and Communications Societies (INFOCOM)*, 2003, pp. 844–852.
- [14] H. Wu, C. Zhu, R. J. La, X. Liu, and Y. Zhang, "FASA: Accelerated S-ALOHA using access history for event-driven M2M communications," *Networking, IEEE/ACM Transactions on*, vol. 21, no. 6, pp. 1904–1917, 2013.
- [15] A. Kumar, E. Altman, D. Miorandi, and M. Goyal, "New insights from a fixed point analysis of single cell IEEE 802.11 WLANs," in *Proc. of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, 2005, pp. 1550–1561.
- [16] P. K. Wong, D. Yin, and T. T. Lee, "Analysis of non-persistent CSMA protocols with exponential backoff scheduling," *IEEE Transactions on Communications*, vol. 59, no. 8, pp. 2206–2214, 2011.
- [17] W.-B. Pöttner, S. Schildt, D. Meyer, and L. Wolf, "Piggy-Backing link quality measurements to IEEE 802.15.4 acknowledgements," in *Proc. of the 8th IEEE International Conference on Mobile Adhoc and Sensor Systems (MASS)*, 2011, pp. 807–812.
- [18] K. Ashrafuzzaman and K. S. Kwak, "On the performance analysis of the contention access period of IEEE 802.15.4 MAC," *IEEE Communications Letters*, vol. 15, no. 9, pp. 986–988, 2011.
- [19] F. Shu and T. Sakurai, "A new analytical model for the IEEE 802.15.4 CSMA-CA protocol," *Computer networks*, vol. 55, no. 11, pp. 2576–2591, 2011.
- [20] L. Kleinrock and F. A. Tobagi, "Packet switching in radio channels: Part I – carrier sense multiple-access modes and their throughput-delay characteristics," *IEEE Transactions on Communications*, vol. 23, no. 12, pp. 1400–1416, 1975.
- [21] R. M. Corless, G. H. Gonnet, D. E. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W function," *Advances in Computational mathematics*, vol. 5, no. 1, pp. 329–359, 1996.
- [22] R. Jain, D.-M. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," Digital Equipment Corporation, Technical Report DEC-TR-301, 1984.
- [23] L. Jiaping, Y. Yung, A. Proutiere, C. Mung, and H. Poor, "Convergence and tradeoff of utility-optimal CSMA," in *Broadband Communications, Networks, and Systems (BROADNETS), Sixth International Conference on*, Sept 2009, pp. 1–8.