IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 6, 2019, accepted August 31, 2019, date of publication September 18, 2019, date of current version October 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2942114

Self-Organizing TDMA: A Distributed Contention-Resolution MAC Protocol

MAHSA DERAKHSHANI[1], (Member, IEEE), YAHYA KHAN[2], DUC TUONG NGUYEN[2], SAEEDEH PARSAEEFARD[3], (Senior Member, IEEE), ATOOSA DALILI SHOAEI[2], AND THO LE-NGOC[2], (Fellow, IEEE)

[1]Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, Loughborough LE11 3TU, U.K.
[2]Department of Electrical and Computer Engineering, McGill University, Montreal, QC, Canada
[3]Iran Telecommunication Research Center (ITRC), Communication Technologies Department, Tehran 1439955471, Iran

Corresponding author: Mahsa Derakhshani (m.derakhshani@lboro.ac.uk)

This work was supported in part by the Huawei Technologies Canada, and in part by the Natural Sciences and Engineering Research Council of Canada.

ABSTRACT This paper presents a self-organizing time division multiple access (SO-TDMA) protocol for contention resolution aiming to support delay-sensitive applications. The proposed SO-TDMA follows a cognition cycle where each node independently observes the operation environment, learns about the network traffic load, and then makes decisions to adapt the protocol for smart coexistence. Channel access operation in SO-TDMA is similar to carrier-sense multiple-access (CSMA) in the beginning, but then quickly converges to TDMA with an *adaptive* pseudo-frame structure. This approach has the benefits of TDMA in a high-load traffic condition, and overcomes its disadvantages in low-load, heterogeneous traffic scenarios. Furthermore, it supports distributed and asynchronous channel-access operation. These are achieved by adapting the transmission-opportunity duration to the common idle/busy channel state information acquired by each node, without any explicit message passing among nodes. The process of adjusting the transmission duration is modeled as a congestion control problem to develop an additive-increase-multiplicative-decrease (AIMD) algorithm, which monotonically converges to fairness. Furthermore, the initial access phase of SO-TDMA is modeled as a Markov chain with one absorbing state and its required convergence time is studied accordingly. Performance of SO-TDMA in terms of effective capacity, system throughput, collision probability, delay-outage probability and fairness is investigated. Simulation results illustrate its effectiveness in performance improvement, approaching the ideal case that needs complete and precise information about the queue length and the channel conditions of all nodes.

INDEX TERMS Pseudo-TDMA, CSMA, MAC, effective capacity, Markov chain with one absorbing state, machine-type communications (MTC), Internet-of-Things.

I. INTRODUCTION

A. BACKGROUND AND MOTIVATION

The Internet of Things (IoT) is a promising technology paradigm that enables physical devices to connect to the internet and share their data without human intervention. The IoT is expected to provide the fundamentals of the highly automated society. However, there are still numerous challenges to be resolved to unlock the full potential of the IoT. One such challenge is how to enable and improve the efficiency of machine-type communication (MTC). MTC is different from human-type communication (HTC) in several aspects. First, MTC has many applications with diverse QoS requirements. For example, self-driving cars and smart healthcare applications are very delay-sensitive while smart cities applications such as smart cameras are less QoS demanding. Second, MTC devices have diverse traffic profiles, ranging from short and bursty to periodic traffic types. Finally, MTC networks typically contain a large number of devices, orders of magnitude above that in HTC networks. These three features of MTC lead to the demand for a robust and resilient medium access control (MAC) protocol which can support diverse QoS requirements at low signaling overhead.

The fundamental MAC mechanism of IEEE 802.11, called distributed coordination function (DCF), is a contention-based MAC scheme based on carrier-sense multiple-access

The associate editor coordinating the review of this manuscript and approving it for publication was Javed Iqbal.

with collision avoidance (CSMA/CA). Traditionally designed to handle only data traffic, DCF fails to provide guaranteed QoS for delay-sensitive applications. This problem mostly stems from inevitable collisions occurring in CSMA/CA [1]–[3]. Specifically, as the traffic load or the number of nodes increases, the number of collisions and hence the delay significantly increase due to the heavy contention among nodes using DCF. Thus, DCF is unsuitable to IoT networks containing large numbers of devices. To alleviate this issue, the IEEE 802.11ah standard [4] introduces the restricted access window (RAW) mechanism to reduce collisions in dense IoT network by grouping devices into different RAW groups and allowing only devices in the same group to compete with each other for transmission. This approach improves throughput, latency and energy efficiency as figured out by [5]–[8] but collisions might still happen within each group.

To support QoS provisioning, extensive research has ensued focusing on improving the MAC layer protocol since the inception of the IEEE 802.11 standard [9]–[12]. Several researches such as [13]–[16] focus on how to reduce collisions by combining the advantages of CSMA and TDMA. In [13] and [16], a node will use a constant predefined backoff value after a successful transmission. This enables the system to converge to a fixed transmission round and operate with no collision. However, both the transmission duration and the deterministic backoff value are non-adaptive, which might leave many time slot wasted. Zero-collision (ZC) [14] and learning-ZC [15] protocols allow dynamic duration of one transmission round, however, this might cause problems in QoS in delay-sensitive applications.

Apart from combining CSMA and TDMA, recent researches such as [17]–[19] concentrate on collision resolution by different methods. In [17], initially, devices contend to send access requests and those succeeding in this phase will be allowed to transmit in the next frame. This scheme alleviates collisions in the data transmission phase, however, some parts of a frame have to be reserved for request signals rather than data transmission. In [18], the contention window of each device is optimized to minimize collisions and in [19], the probability of transmission of each device is adjusted. These methods have low signaling overhead but collisions might still happen.

One potential alternative MAC protocol for collision resolution that has low signaling overhead and can converge to collision-free data transmission is pseudo-TDMA (PTDMA). [20]–[24]. In PTDMA, each active node starts its transmission by choosing random back-off times as in CSMA. But, after a node successfully transmits once and receives the relevant acknowledgement, it picks a fixed deterministic back-off value and switches to periodic transmission as in TDMA. After any collision, the node switches back to random back-off. Although periodic transmission provides reservation guarantees by having fixed slots and a frame-length by which the slot allocation is repeated, it has a drawback of being inflexible and cannot adapt to heterogeneous traffic. Thus, a MAC protocol that employs PTDMA but circumvents its inflexibility problem is necessary and will be the focus of this paper.

B. CONTRIBUTIONS

In order to identify and also quantify both potentials and shortcomings of PTDMA in QoS support, in this paper, we first investigate the *effective capacity* (EC) of PTDMA in comparison with CSMA. This study, for the first time to the best of our knowledge, evaluates the statistical QoS performance of PTDMA under a variety of traffic conditions. EC has been proposed in [25] as a QoS-aware metric that determines the maximum constant arrival rate that can be supported by a network, while satisfying a target statistical delay requirement. Our studies show that PTDMA can improve EC in *saturated* traffic scenarios compared with CSMA by reducing the collision probability.

However, when traffic is *unsaturated*, our studies reveal that CSMA could provide better EC performance. Such unsaturated traffic condition results in short-term random fluctuations in the number of active nodes (denoted by N_a), that have packets in their queues and are contending for the channel. This variation in N_a is not catered for in the PTDMA frame structure. In other words, PTDMA parameters such as the time-slot size and the frame-length are not adjusted based on N_a, while CSMA is flexible to the traffic demand and the number of active nodes owing to its opportunistic behavior. Consequently, in an unsaturated traffic scenario, PTDMA suffers from underutilization that may lead to high delays. The reason is that a portion of the channel might be left unoccupied even though there are nodes waiting to send a packet.

In order to overcome this problem, we propose a self-organizing TDMA (SO-TDMA) protocol. In this protocol, nodes can operate in a distributed and asynchronous manner by carrier sensing, but eventually in an efficient and opportunistic TDMA manner. More specifically, each node initiates its transmission through CSMA and then switches to periodic transmission as in PTDMA. But, different from PTDMA, the wireless channel frame structure in SO-TDMA is adaptable to the changing traffic and channel conditions.

In particular, a distributed learning-based MAC algorithm is developed, in which each node independently adapts its transmission length to the optimal values over time by learning the number of active nodes based on locally available information. This fully distributed SO-TDMA protocol eliminates the need for any central coordination that would suffer from scalability issues or any information exchange that would degrade throughput due to additional overhead. The process of transmission length adaptation of SO-TDMA will be analytically derived from the network congestion control problem, where each node independently adapts its own transmission rate to avoid any congestion in the bottleneck link due to the limited capacity.

The convergence behavior of proposed SO-TDMA is analytically studied for two phases of the proposed algorithm.

First, the required time to pass the initial CSMA phase (modeled as a Markov chain with one absorbing state) is studied. It is proved that the expected time for transition to a periodic transmission phase is only linearly increasing with the network size. Second, for the periodic transmission phase, it is proved that the proposed additive-increase-multiplicative-decrease (AIMD) time-slot adaptation algorithm for SO-TDMA converges to fairness.

Simulation results are obtained to evaluate the performance of the proposed SO-TDMA algorithm in terms of system throughput, fairness, collision probability, and effective capacity for QoS. They reveal that SO-TDMA can be a good MAC candidate for QoS provisioning in MTC for better EC as compared with CSMA and PTDMA in both saturated and unsaturated traffic scenarios.

The preliminary results of this work have been presented in [26]. As compared with [26], in this paper, we focus on theoretic analysis to obtain optimal adjustment rate of time slot sizes and prove the favorable convergence behaviors of our proposed protocol. More simulations under several traffic types are also carried out to verify the effectiveness of our protocol and its convergence under changes in the number of users.

C. STRUCTURE

The rest of this paper is organized as follows. Section II describes the system model for all the MAC protocols considered as well as a brief review of PTDMA and other related works. After details of the proposed SO-TDMA protocol given in Section III, Section IV studies the convergence properties of SO-TDMA. This is followed by the performance evaluation and benchmarking results in Section V. Finally Section VI presents concluding remarks.

II. SYSTEM MODEL

We consider a network comprised of a single AP and a set of nodes $\mathcal{N} = \{1, \cdots, N\}$ sharing a communication channel, where the average arrival bit rate of each node n is μ_n as shown in Fig. 1. We study two different types of arrival traffic in this paper, including constant bit rate (CBR) and Poisson arrival. Suppose Q_n denotes the queue length of node n, i.e., the number of backlogged packets in its queue. Accordingly, the number of active nodes (i.e., N_a) is defined as the number of nodes with $Q_n > 0$ whose instantaneous SNR is also above the minimum required threshold for transmission.

We assume a slotted transmission, where T denotes the length of the smallest unit of time which is a back-off time-slot. Subsequently, the length of a single transmission opportunity of node n is represented by $T_{\mathrm{s},n}$, which is defined in terms of the number of back-off time-slots. Furthermore, for node n, the required number of back-off time-slots to transmit one packet can be calculated as $S_{\mathrm{req},n} = P_\mathrm{s}/(R_n \times T)$, where P_s is the fixed packet size and R_n is the transmission data rate of node n.

A block fading channel model is assumed in which each node's channel gain remains constant for a block of time equal to the channel coherence time T_c, while it independently and asynchronously varies from one block to another. The transmission data rate of node n, R_n, is determined based on the instantaneous channel power gain, which is reported back to each node by the AP.

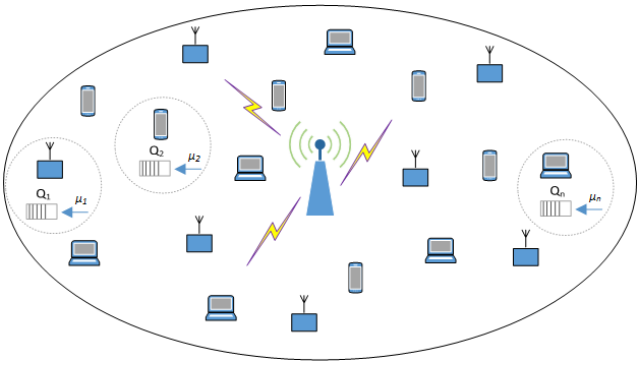


FIGURE 1. Network model.

Based on this information, a limited number of transmission modes, K, corresponding to certain transmission rates, is selected in order to guarantee a minimum packet error rate (PER) for each node. Let g_n be the channel power gain between the node n and the AP, the instantaneous SNR is given as

$$SNR_n = Pg_n/\sigma^2, \tag{1}$$

where P is the transmit power of the node and σ^2 represents the noise power. A transmission mode $k \in \{1, \ldots, K\}$ corresponding to the channel rate R_k requires a minimum SNR threshold η_k where $\eta_1 < \eta_2 < \cdots < \eta_K$. Therefore, the selected channel rate R_k for transmission of packets corresponds to the η_k such that $SNR_n \geq \eta_k$. If $SNR_n < \eta_1$ no transmission takes place.

The basic concept of PTDMA is illustrated in Fig. 2. PTDMA starts with CSMA and then switches to a *periodic transmission phase* with a frame-length equal to T_f. There are two important parameters related to PTDMA: the frame-length, T_f, and the transmission time-slot size, $T_{\mathrm{s},n}$. Both are vital to its performance as they directly impact the channel utilization, throughput, packet delay and fairness.

Since PTDMA has fixed parameters that define the frame structure, due to the dynamic nature of the channel and traffic arrival, it is often the case that some portion of the channel is left unused, adversely affecting channel utilization. This is due to two possible scenarios: 1) The dynamic queue length of nodes leads to empty queues in several of them, 2) The channel gain between a node and the AP falls below the minimum threshold, prohibiting the node from transmitting or rendering its signal too weak to be decoded at the AP.

Nodes that are not in the two scenarios mentioned above (have a non-empty queue and channel gain equal to or over the thresholds) are called *active nodes*. In order to achieve 100% channel utilization, we need $\sum_{n=1}^{N_a} T_{s,n} = T_f$, where N_a is the number of active nodes. Ideally, $T_{\mathrm{s},n}$ should instantaneously be updated to $T_{\mathrm{s},n} = T_f/N_\mathrm{a}$. In the rest of this paper,

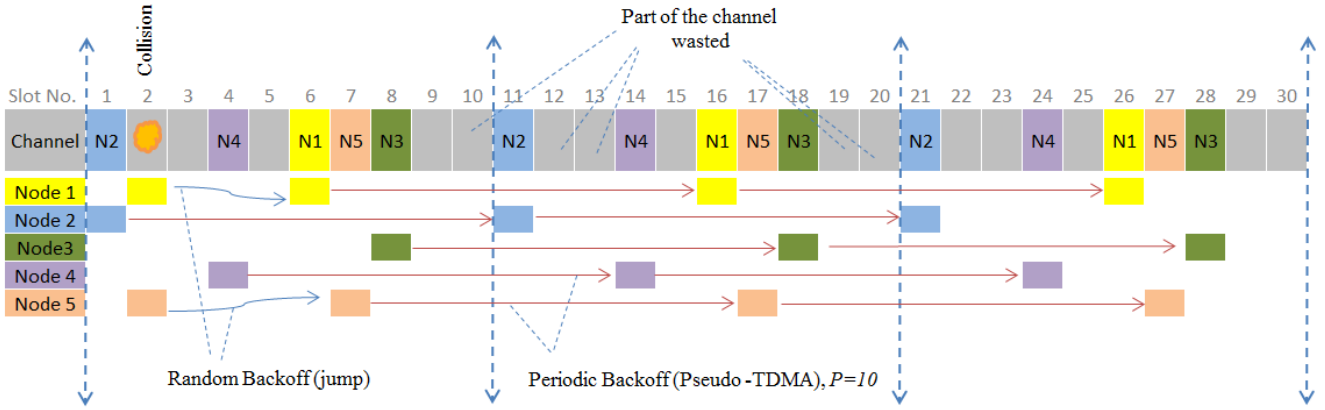


FIGURE 2. Illustration of PTDMA operation.

we refer to PTDMA with $T_{s,n} = T_f/N_a$ as "Ideal-PTDMA" and implement it to represent an upper bound for PTDMA performance. However, in reality, since the information of N_a is hard to be available to every node in the network, in PTDMA, $T_{s,n}$ is generally set to $T_{s,n} = T_f/N$, where N is the total number of nodes and its value is assumed to be known by every node.

An enhanced version of PTDMA, called period-controlled MAC (PCMAC), is proposed in [27], in which the frame-length is not fixed and adapts to the number of idle slots, while maintaining the relative position of each node in each frame-length. Since the convergence of the PTDMA is very sensitive to the frame-length, changing it can be problematic considering fluctuating heterogeneous traffic demand and random channel gains. Moreover, since T_f determines the delay between successive transmissions, its value can significantly impact QoS for delay sensitive applications and changing it could not be the best option.

Thus, in the following, we propose SO-TDMA, an alternative of PTDMA with adjustable time-slot sizes and fixed frame-length. With this approach, we can still have the adaptability with the network load and channel gains as in PCMAC, while securing reliable convergence and better EC performance. In the proposed SO-TDMA, the transmission length of each node is distributively optimized during the *periodic transmission phase* in order to maximize channel utilization and fairness. Through AIMD control, the value of transmission time-slot size $T_{s,n}$ converges from an initial value of T_0 to $\frac{T_f}{N_a}$ and aims to closely follow this value. Therefore, this protocol allows users to optimize the pseudo-frame structure according to the changing channel and traffic conditions, repeating the cycle of *observe*, *decide*, *act* and *learn* every pseudo-frame while approaching the performance of the Ideal-PTDMA. Since for the Ideal-PTDMA algorithm the instantaneous value of N_a is assumed to be available, its performance is optimal. In order to verify the benefits of the proposed SO-TDMA protocol, through numerical results, its performance is compared with CSMA, PTDMA, Ideal-PTDMA, and PCMAC as benchmarks.

III. PROPOSED SELF-ORGANIZING TDMA (SO-TDMA)

Here, we present in detail the SO-TDMA protocol that aims to improve the spectrum sharing efficiency in terms of channel utilization, fairness among nodes, and QoS in terms of packet delay. The key advantage of the proposed SO-TDMA is the ability to dynamically adapt the transmission time-slot of each node, $T_{s,n}(f_n)$, to changing traffic and channel conditions, where f_n represents the pseudo-frame index for node n. Thus, this protocol aims to cater as many nodes on the common channel as possible in their respective pseudo-frames, while minimizing the number of collisions as well as the wastage of resources. The access mechanism in SO-TDMA consists of two phases for each node: 1) *initial access phase* where each node tries to access the channel via the random back-off procedure of CSMA; 2) *periodic transmission phase* where each node independently attempts to shape the network pseudo-frame structure, while maximizing the channel utilization and the fair slot allocation among nodes.

A. INITIAL ACCESS PHASE

In SO-TDMA, the *initial access phase* for a node is the same as in CSMA, where the exponential random back-off procedure is executed. More specifically, a node with a packet to transmit should monitor the channel before transmission for a period of time called distributed inter frame space (DIFS). If the channel is sensed idle, the node will transmit. Otherwise, it continues monitoring until the channel is measured idle for a duration of DIFS and backs off for a random period of time, uniformly chosen from a range of $[0, w-1]$, where w is known as the contention window given in terms of back-off time-slots. Initially, w is set to the minimum contention window size $CW_{\min}$. In case of any failure attempt, the value of w is doubled but not exceeding the maximum contention window size $CW_{\max}$ [28].

The back-off counter is decremented and a node transmits when it reaches zero. While counting down, if the channel is sensed busy, the node will freeze its back-off counter and continue decrementing once the channel becomes idle again.

For each transmission opportunity, the node can transmit multiple back-to-back packets for a fixed period of time equal to $T_{s,n}$. Once the packet is received successfully, the AP waits for a period of time called short inter-frame space (SIFS) and then sends an acknowledgment (ACK) for the successful reception.

In the *initial access phase*, the transmission time-slot duration is fixed to $T_{s,n}(f_n) = T_0$ for all nodes. After a node grabs the channel for the first time using CSMA, a countdown timer of length T_f is initiated at the start of its transmission. In this phase, if the node could empty its queue during any transmission (i.e., $Q_n = 0$), the timer is reset to T_f again. In case the timer expires and the node still has some packets in its queue (i.e., $Q_n > 0$), the node would enter the *periodic transmission phase* (which is presented in Algorithm 1) until Q_n becomes 0 again.

B. PERIODIC TRANSMISSION PHASE

In the *periodic transmission phase*, all nodes periodically transmit with the same back-off value equal to the *pseudo-frame length* T_f. At the end of each node's respective periodic back-off, it senses the channel and starts transmission if the channel is sensed idle with the transmission time-slot size $T_{s,n}(f_n + 1)$. The value of $T_{s,n}(f_n + 1)$ is updated according to Algorithm 1, which will fully be described later. In case the channel is not idle (due to arrival of a newly joined node or expansion of time-slot sizes of previously accessed nodes), the node *does not revert* to the initial access phase but tries to find another back-off value to resume the *periodic transmission phase* by the following manner. First, the node continues sensing the channel until it becomes idle again. Then, the node carries out an additional random back-off procedure before returning to the periodic transmission phase. This additional random back-off is necessary to avoid collisions between two nodes with periodic back-off already expired and waiting to transmit as the channel becomes free. In the following section, the transmission time-slot adaptation algorithm is explained.

1) AIMD TIME-SLOT ADAPTATION ALGORITHM

Suppose that all users picked their locations in the frame through CSMA in the initial phase and started to transmit in the *periodic transmission phase*. The process of adjusting the transmission time-slot size can be modeled as a network congestion control problem, in which different nodes independently adapt their transmission rates to avoid any congestion in the bottleneck link due to the limited capacity [29]–[31].

Let define the *utilization rate* of node n as $r_n = \frac{T_{s,n}}{T_f}$ where $0 \le r_n \le 1$. The rate allocation can be modeled as an optimization problem to maximize the channel utilization and to achieve fairness as

$$\max_{\boldsymbol{r}\ge 0} \sum_{n=1}^{N} U_n(r_n), \quad \text{subject to,} \quad \sum_{n=1}^{N} r_n \le 1 \tag{2}$$

where $\boldsymbol{r} = [r_1, \ldots, r_N]$. $U_n(r_n)$ represents the utility earned by node n with allocated r_n. The linear flow constraint also states that the sum of time-slot sizes should be kept smaller than the frame size.

Assuming $U_n(r_n) = \log(r_n)$, in [30], it has been proven that a utilization rate control, based on a system of differential equations as

$$\frac{\partial}{\partial t} r_n(t) = \kappa\,(1 - r_n(t)\beta(t)) \tag{3}$$

where

$$\beta(t) = h\Big(\sum\nolimits_{n'=1}^{N} r_{n'}(t)\Big), \tag{4}$$

will converge to a unique stable point, which satisfies *proportional fairness*. In the system (3)-(4), $\kappa > 0$ is a constant and $h(x)$ is a non-negative, continuous, and increasing function of x. A feasible vector $\boldsymbol{r}^* = [r_1^*, \ldots, r_N^*]$ is *proportionally fair* if

$$\sum_{n=1}^{N} \frac{r_n - r_n^*}{r_n^*} \le 0 \quad \Longleftrightarrow \quad \frac{1}{N}\sum_{n=1}^{N} \frac{r_n}{r_n^*} \le 1 \tag{5}$$

for any other feasible vector $\boldsymbol{r}$, i.e., the aggregate of proportional changes is zero or negative.

Assuming $h(x) = [x - 1 + \epsilon]^+/\epsilon^2$ where $[x]^+ = \max(x, 0)$, it has been shown that (3) converges to the optimal solution of (2) when $\epsilon \to 0$. $\beta(t) = \Big[\sum_{n'=1}^{N} r_{n'}(t) - 1 + \epsilon\Big]^+/\epsilon^2$ is a pricing function to control the rate of change of $r_n(t)$ in a distributed manner. The pricing function is zero when the current total flow is small enough (i.e., $\sum_{n'=1}^{N} r_{n'}(t) \le 1 - \epsilon$) and the rate of node n will be increased with a constant rate of κ. When the system is congested, $\beta(t)$ is large (for small ϵ) and the rate of node n will be decreased proportional to its rate as $1 - r_n(t)\beta(t) < 0$. Supposing $\Delta t = T_f$, the utilization rate control can be written as

$$r_n(f_n + 1) \approx r_n(f_n) + \kappa T_f\Big(1 - r_n(f_n)\big[\sum\nolimits_{n'=1}^{N} r_{n'}(f_n) - 1 + \epsilon\big]^+/\epsilon^2\Big). \tag{6}$$

Consequently, (6) becomes

$$r_n(f_n + 1) \approx r_n(f_n) + \begin{cases} \kappa T_f & \text{if} \quad \sum_{n'=1}^{N} r_{n'}(f_n) \le 1 - \epsilon \\ \kappa T_f\Big(1 - r_n(f_n)\Big[\sum_{n'=1}^{N} r_{n'}(f_n) - 1 + \epsilon\Big]/\epsilon^2\Big) & \\ \qquad \text{otherwise.} & \end{cases} \tag{7}$$

Let define $R_I = \kappa T_f$ and $R_D = \kappa T_f/\epsilon$. Moreover, we define the *expected* idle fraction of the frame f_n as

$$y(f_n) = 1 - \sum\nolimits_{n'=1}^{N} r_{n'}(f_n). \tag{8}$$

According to these definitions, (7) becomes

$$r_n(f_n+1) \approx \begin{cases} r_n(f_n)+R_I & \text{if } y(f_n)\geq\epsilon \\ r_n(f_n)+R_I-R_D r_n(f_n)[1-y(f_n)/\epsilon] & \text{if } y(f_n)\leq\epsilon. \end{cases} \quad (9)$$

This utilization rate control is kind of an *additive-increase-multiplicative-decrease (AIMD)* algorithm, in which if the expected idle fraction of a current frame is larger than a threshold, ϵ, then the utilization rate of each node will be increased by adding a constant to its previous demand. However, if the current frame is mostly busy and the expected idle fraction is smaller than ϵ, then the utilization rate of each node will be reduced and the decrease is multiplicative proportional to the previous utilization rate. Here, ϵ can be interpreted as

$$\epsilon = I_{\text{th}}/T_f, \quad (10)$$

where I_{th} is defined as the target number of idle back-off units in each frame, in order to provide the non-active nodes a non-zero access chance to the channel.

To transform the utilization rate control in (9) into a time-slot adaptation, setting $W_I = R_I T_f$ and $W_D = R_D$, we have

$$T_{s,n}(f_n+1) \approx \begin{cases} T_{s,n}(f_n)+W_I & \text{if } J(f_n)\geq I_{\text{th}} \\ T_{s,n}(f_n)+W_I-W_D T_{s,n}(f_n)(1-\frac{J(f_n)}{I_{\text{th}}}) & \text{if } J(f_n)\leq I_{\text{th}} \end{cases} \quad (11)$$

where

$$J(f_n) = T_f - \sum_{n'=1}^{N} T_{s,n'}(f_n). \quad (12)$$

$J(f_n)$ gives an estimation of the network load, i.e., the lower J indicates the busier channel we would expect. However, in reality, it is hard to inform all nodes about the time slot size of all other nodes, so each node tries to measure the number of idle back-off units in its pseudo-frame as an estimation of $J(f_n)$. Nonetheless, this measurement is noisy due to collision, in other words, if we define $I_n(f_n)$ as the *measured* number of idle back-off units sensed by node n during the pseudo-frame f_n, we have

$$I_n(f_n) = J(f_n) + z_n(f_n), \quad (13)$$

where $z_n(f_n)$ represents the noise during the measurement by node n in pseudo-frame f_n. In addition to collisions that may cause an error in measuring the real number of idle time-slots, asynchronicity in measurements can cause a problem. In other words, since the start and end of the pseudo-frame for each node might be different, $I_n(f_n)$ measurements during the time period T_f after each transmission are asynchronous. However, assuming that all the nodes can perfectly sense the channel status, i.e., busy or idle, a moving average $\bar{I}_n(f_n)$ on the number of idle time-slots enables all active nodes to ultimately converge to a same value.

$$\bar{I}_n(f_n) = \alpha I_n(f_n) + (1-\alpha)\bar{I}_n(f_n-1), \quad 0<\alpha<1 \quad (14)$$

Algorithm 1 SO-TDMA: AIMD Time-Slot Adaptation at Active Node n

1: **Initialization:**
   Set $T_{\min}$, $T_{\max}$, and T_0
   Set $f_n = 0$, $T_{s,n}(0) = T_0$ and T_f
   Set I_{th}, $\bar{I}_n(0) = I_{\text{th}}$
2: **Periodic Transmission Phase**
3: **while** $Q_n(f_n) > 0$ **do**
4: Measure $I_n(f_n)$
5: Set $\bar{I}_n(f_n) = \alpha I_n(f_n) + (1-\alpha)\bar{I}_n(f_n-1)$
6: **if** $\bar{I}_n(f_n) > I_{\text{th}}$ **then**
7: $T_{s,n}(f_n+1) = T_{s,n}(f_n) + W_I$
8: **else if** $\bar{I}_n(f_n) < I_{\text{th}}$ **then**
9: $T_{s,n}(f_n+1) = T_{s,n}(f_n)(1 - W_D(1 - \frac{\bar{I}_n(t)}{I_{\text{th}}})) + W_I$
10: **end if**
11: $T_{s,n}(f_n+1) = \max(\min(T_{s,n}(f_n+1), T_{\max}), T_{\min})$
12: $f_n = f_n + 1$
13: **if** $Q_n(f_n) = 0$ **then**
14: Return to Initial access phase
15: $f_n = 0$, $T_{s,n}(0) = T_0$
16: **end if**
17: **end while**

Thus, we can update the control algorithm in (11) using a moving average of I_n as a measure for J, which can partly phase out the noise.

$$T_{s,n}(f_n+1) = \begin{cases} T_{s,n}(f_n)+W_I & \text{if } \bar{I}_n(f_n)\geq I_{\text{th}} \\ T_{s,n}(f_n)+W_I-W_D T_{s,n}(f_n)(1-\frac{\bar{I}_n(t)}{I_{\text{th}}}) & \text{if } \bar{I}_n(f_n)\leq I_{\text{th}} \end{cases} \quad (15)$$

Thus, using the time-slot adaption in (15), each node aims to ensure that there are at least I_{th} idle time-slots in all pseudo-frames. To reach this goal, each node adaptively adjusts its transmission time-slot size in the next frame, i.e., $T_{s,n}(f_n+1)$, based on the information from the last frame $T_{s,n}(f_n)$ and $\bar{I}_n(f_n)$. When the frame is underutilized, i.e., $\bar{I}_n(f_n) > I_{\text{th}}$, $T_{s,n}$ will be increased by adding a constant (W_I) to the previous demand. However, when the frame is over-utilized $\bar{I}_n(f_n) < I_{\text{th}}$, the time-slot size will be decreased proportional to the previous size. The details of the proposed time-slot adaptation algorithm are presented in Algorithm 1 (Lines 8-12).

In Algorithm 1, it should be noted that $T_{\min}$ and $T_{\max}$ are considered as the lower and upper bounds on the value of $T_{s,n}$, respectively. In other words, at any frame, $T_{s,n}$ needs to be kept in a range as $T_{\min} \leq T_{s,n} \leq T_{\max}$. Thus, after adjusting $T_{s,n}$ based on AIMD technique in Lines 8-12, such enforcement is implemented in Line 13 as

$$T_{s,n}(f_n+1) = \begin{cases} T_{\max} & \text{if } T_{s,n}(f_n+1) > T_{\max} \\ T_{\min} & \text{if } T_{s,n}(f_n+1) < T_{\min} \\ T_{s,n}(f_n+1) & \text{Otherwise.} \end{cases} \quad (16)$$

$T_{\min}$ can be set to an arbitrary value so that the payload is still a significant portion of a transmission as compared to the overheads. However, $T_{\min}$ directly limits the maximum number of active nodes in the fixed frame-length T_f to $T_f/T_{\min}$. Moreover, we need to have $T_{\max} \leq T_f - I_{\text{th}}$ in order to ensure at least I_{th} idle time-slots in any frame, even if there is a single active node.

In this algorithm, the moving average parameter α determines the amount of stability in the algorithm (a higher value gives more weightage to the latest measure of idle-slots as opposed to previous values). Decreasing α increases stability but also increases the convergence time, while increasing α can reduce the convergence time up to a certain point beyond which it becomes too unstable to converge. Furthermore, I_{th} (the target threshold for the number of idle time-slots in one frame), defined to give access probability to new incoming nodes, also faces a compromise. Increasing its value will make channel utilization poor as more back-off slots per frame are always kept idle but will reduce the initial channel access delay as the probability of finding idle slots for new nodes becomes higher. Similarly, decreasing it will have the opposite impact.

This iterative process in Algorithm 1 runs independently and in a distributed manner on node n until $Q_n(f_n) = 0$. In a nutshell, Algorithm 1 ensures how nodes converge to a common pseudo-frame structure without any information sharing, while achieving fairness at the pseudo-frame level (avoiding starvation of nodes) and high channel utilization.

IV. CONVERGENCE ANALYSIS

In this section, we study convergence properties of the proposed SO-TDMA, including the required time to pass the initial access phase and convergence to fairness in the periodic transmission phase using AIMD time-slot adaptation.

A. CONVERGENCE TIME: INITIAL ACCESS PHASE

In this subsection, we aim to study the required time for passing the initial phase and moving on to the periodic transmission phase. To this end, we first try to mathematically model the initial access phase of SO-TDMA. Then, using the developed model, we investigate the asymptotic behavior of convergence time with respect to the number of nodes in the network.

To present a tractable model, we make a few assumptions. First, we assume all nodes have always packets to transmit, which can represent the worst case scenario. Since all the nodes start with a time-slot equal to T_0, we assume there are $M = \frac{T_f}{T_0}$ different positions available in a frame to be picked by each node. We also assume nodes will choose their locations in a round-robin manner. More specifically, in each round, one node randomly picks (via CSMA) a location out of the M available ones. If the node chooses a location which has already been picked by others, a collision occurs. Thus, the two nodes who have been involved in this collision need to keep trying to find appropriate locations.

With these assumptions, the initial access phase of SO-TDMA can be modeled as a discrete-time Markov chain with one absorbing state. Let assume that the state of the network (denoted by S) is represented by the number of nodes, who have already accessed the channel successfully through CSMA and grabbed their locations in the frame. Thus, S belongs to a set $\mathcal{S} = \{0, \ldots, N\}$.

This procedure can be modeled as a discrete-time Markov chain since the probability of the network having N settled nodes in the future only depends on the current number of nodes in the network not the past. In other words, the initial access phase of SO-TDMA has the Markov property as

$$\begin{aligned}\Pr(S_{i+1} = N | S_0 = n_0, S_1 = n_1, \ldots, S_i = n_i) \\ = \Pr(S_{i+1} = N | S_i = n_i).\end{aligned} \tag{17}$$

This Markov chain is a birth and death chain, except for the last state that represents the case when all nodes settled down in the frame through CSMA. Thus, the state N is the absorbing state with $\Pr(S_{i+1} = N | S_i = N) = 1$.

Otherwise, if the network is at the state $S_i = s \in \{0, \ldots, N-1\}$, the chance that a node, whose turn is to pick, chooses an idle location is $1 - \frac{s}{M}$. Thus, we have

$$\Pr(S_{i+1} = s+1 | S_i = s) = 1 - \frac{s}{M}. \tag{18}$$

Otherwise, there is a chance of $\frac{s}{M}$ that the node chooses a location, which has been already picked by others. In this case, a collision occurs and the system state will go back to $s-1$. Thus, we have

$$\Pr(S_{i+1} = s-1 | S_i = s) = \frac{s}{M}. \tag{19}$$

In order to calculate the convergence time in the initial phase, we study the mean time to absorption in the underlying Markov chain. In other words, our objective is to find the expected number of visits in the transient states before being absorbed at $s = N$, when starting at $s = 0$.

Proposition 1: The expected time to converge to a periodic transmission phase is linearly increasing with N when $N < \sqrt{M}$.

Proof: The transition matrix of the proposed Markov chain can be represented as

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{I_1} \end{bmatrix}, \tag{20}$$

where the $N \times N$ *absorbing matrix* $\boldsymbol{A}$ represents the probabilities of transitions among the transient states, the $N \times 1$ matrix $\boldsymbol{B}$ represents the probabilities to reach the absorbing state from other transient states in one step, $\boldsymbol{0}$ is an $1 \times N$ zero matrix, and $\boldsymbol{I_1}$ is the 1×1 identity matrix. In this problem, the entries of $\boldsymbol{B}$ are 0, except for the last one $b_N = 1 - \frac{N-1}{M}$, which represents the transition probability from $s = N-1$ to the absorbing state $s = N$. In an absorbing Markov chain, it is well-known that the matrix of the expected number of visits to each state before absorption given the initial state is

$$\boldsymbol{L} = [l_{ij}]_{N \times N} = (\boldsymbol{I}_N - \boldsymbol{A})^{-1}, \tag{21}$$

where l_{ij} represents the expected number of times the chain is in state j, given that the chain started in state i [32]. Here we are looking for

$$\lambda_0 = \sum_{j=0}^{N-1} l_{1j}, \tag{22}$$

which is the expected number of visits in the transient states before reaching $s = N$ when starting at $s = 0$. In the proposed Markov chain, $\boldsymbol{I}_N - \boldsymbol{A}$ is a triangular matrix as

$$\boldsymbol{I}_N - \boldsymbol{A} = \begin{bmatrix} 1 & b_1 & 0 & 0 & \dots & 0 \\ c_1 & 1 & b_2 & 0 & \dots & 0 \\ 0 & c_2 & 1 & b_3 & \dots & 0 \\ & & \ddots & \ddots & \ddots & \\ 0 & 0 & \dots & c_{N-2} & 1 & b_{N-1} \\ 0 & 0 & \dots & 0 & c_{N-1} & 1 \end{bmatrix},$$

where $b_j = -1 + \frac{j-1}{M}$ and $c_j = -\frac{j}{M}$ for $j \in \{1, \dots, N-1\}$. The inversion of such tridiagonal matrix can be calculated numerically by

$$l_{1j} = (-1)^{1+j} b_1 b_2 \dots b_{j-1} \frac{\Phi_j + 1}{\Theta_N}, \tag{23}$$

where Θ_i and Φ_i satisfy the recurrence relation as

$$\Theta_j = \Theta_{j-1} - b_{j-1} c_{j-1} \Theta_{j-2}, \tag{24}$$

$$\Phi_j = \Phi_{j+1} - b_j c_j \Phi_{j+2}, \tag{25}$$

with initial conditions $\Theta_0 = \Theta_1 = \Phi_N = \Phi_{N+1} = 1$ [33].

To find a closed-form approximation for λ_0, we compute an upper bound. Considering that $b_j c_j > 0$, from (25), it is clear that $\Phi_j \leq 1$. Thus, (23) can be bounded as

$$l_{1j} \leq \frac{(-1)^{1+j}}{\Theta_N} b_1 b_2 \dots b_{j-1}. \tag{26}$$

Consequently,

$$\lambda_0 = \sum_{j=0}^{N-1} l_{ij} \leq \frac{1}{\Theta_N} \left[2 + \sum_{j=1}^{N-2} (1 - \frac{j}{M}) \right] \leq \frac{N}{\Theta_N}. \tag{27}$$

Since $0 < b_j c_j \leq -c_j$, (24) can be lower bounded by

$$\Theta_j \geq \Theta_{j-1} + c_{j-1} \Theta_{j-2} = \Theta_{j-1} - \frac{j-1}{M} \Theta_{j-2}. \tag{28}$$

Lemma 1: The lower bound on the Θ_N can be derived as

$$\Theta_N \geq 1 - \sum_{j=1}^{N-1} \frac{j}{M} = 1 - \frac{N(N-1)}{2M}. \tag{29}$$

Proof: See Appendix.

Thus, considering (27) and *Lemma 1*, an upper bound can be obtained for λ_0 as

$$\lambda_0 \leq \frac{N}{1 - \frac{N(N-1)}{2M}}. \tag{30}$$

From (30), the mean time to absorption in the presented Markov chain starting at $s = 0$ is linearly increasing with N when $N < \sqrt{M}$. □

Proposition 1: confirms that SO-TDMA passes the initial access phase in a time window that is linearly proportional to the network size and then settles in the periodic transmission phase.

B. CONVERGENCE TO FAIRNESS - PERIODIC TRANSMISSION PHASE

In this subsection, we prove that the proposed AIMD time-slot adaptation algorithm in the periodic transmission phase of SO-TDMA converges to fairness. In other words, following Algorithm 1, nodes can reach to a state, where each has an equal share of channel-access. The Jain's index [34] to evaluate fairness at pseudo-frame f_n can be calculated as

$$F(f_n) = \frac{1}{N} \frac{(\sum_{n=1}^{N} T_{\mathrm{s,n}}(f_n))^2}{\sum_{n=1}^{N} T_{\mathrm{s,n}}^2(f_n)}. \tag{31}$$

Proposition 2: The fairness index monotonically converges to one, i.e.,

$$\forall n \; F(f_n) < F(f_n + 1) \; \textit{and} \; \lim_{f_n \to +\infty} F(f_n) = 1. \tag{32}$$

Proof: In [29], it has been shown that

$$F(f_n + 1) = F(f_n) + (1 - F(f_n)) \Big(1 - \frac{\sum_{n=1}^{N} T_{\mathrm{s,n}}^2(f_n)}{\sum_{n=1}^{N} (C_n + T_{\mathrm{s,n}}(f_n))^2} \Big), \tag{33}$$

where, according to (11),

$$C_n = \begin{cases} W_I, & \text{if } J(f_n) \geq I_{\mathrm{th}} \\ W_I \times \frac{1}{1 - W_D(1 - J(f_n)/I_{\mathrm{th}})}, & \text{if } J(f_n) \leq I_{\mathrm{th}} \end{cases}.$$

Since for $W_D < 1$, C_n is always positive, the second term in (33) will be positive and $F(f_n) < F(f_n + 1)$. As a result of the strict increase of $F(f_n)$ over f_n and $0 \leq F \leq 1$, it can be concluded that $\lim_{f_n \to +\infty} F(f_n) = 1$. □

V. ILLUSTRATIVE RESULTS

In this section, first, the simulation setup is described. Subsequently, illustrative results are presented to evaluate performance of the proposed SO-TDMA in comparison with CSMA, PTDMA and PCMAC. Performance is investigated in terms of EC, delay-outage probability, system throughput, fairness, and collision probability.

A. SIMULATION SETUP AND ASSUMPTIONS

In this paper, we used MATLAB for simulations of all MAC protocols in the 20 MHz frequency band. We implemented the basic CSMA MAC functionality first (as explained in Section IV-A) and tested it rigorously to confirm its performance matched with the literature (e.g., [28]) and then implemented Algorithms 1 and 2 for the proposed SO-TDMA simulations. The PTDMA and Ideal-PTDMA (or upper bound) were also implemented as explained in Section III. For the Ideal-PTDMA, it is assumed each node has updated the knowledge of channel state and queue length information of all nodes (i.e., N_{a}). In particular, the value of $T_{\mathrm{s},n}$ is fixed

TABLE 1. Simulation parameters.

Simulation Parameter	Value
T_{m}	50 s
N	$2-10$
B	20 MHz
T	10 μs
I_{th}	$30 \times T$
T_f	$1000 \times T$
T_{min}	$40 \times T$
T_{max}	T_f-I_{th}=$970 \times T$
T_0	$100T$
T_{c}	10 ms
D_{max}	50 ms
Average SNR	20 dB
DIFS	$4 \times T$
SIFS	$1 \times T$
P_{s}	2400 bytes
$\mathrm{CW}_{\mathrm{min}}$	16
$\mathrm{CW}_{\mathrm{max}}$	1024
α	0.7
ε	0.001
δ	0.5
W_I	$5 \times T$
W_D	0.01

TABLE 2. Transmission rates R_{k} vs. SNR ranges used in the illustrative results.

R_{k} (Mbps)	SNR (dB) $[\eta_k, \eta_{k+1})$
6	[5,8)
9	[8,10)
12	[10,13)
18	[13,16)
24	[16,19)
36	[19,22)
48	[22,25)
54	[25,∞)

to $T_{\mathrm{s},n} = T_0$ for CSMA, $T_{\mathrm{s},n} = T_f/N$ for PTDMA, and $T_{\mathrm{s},n} = T_f/N_{\mathrm{a}}$ for Ideal-PTDMA. For SO-TDMA, the value of $T_{\mathrm{s},n}$ is optimized as in Algorithm 1.

The frame-length update in PCMAC has been implemented exactly as given in Algorithm 1 of [27]. All PCMAC simulation parameters in the implementation have been kept the same as described in Section IV of [27]. All common MAC simulation parameters like packet size, back-off time-slot size, traffic and channel model characteristics for consistency have been set to the same value assumed in other MAC protocols simulated in this paper (listed in Tables 1 and 2).

Furthermore, we assume a Rayleigh fading channel where the channel power gain has an exponential distribution and is independent and identically distributed for all nodes. For the packet arrival, two traffic models are examined in the provided numerical results by feeding nodes with (i) Poisson traffic and (ii) CBR traffic, where μ_n is the mean arrival bit rate for node n.

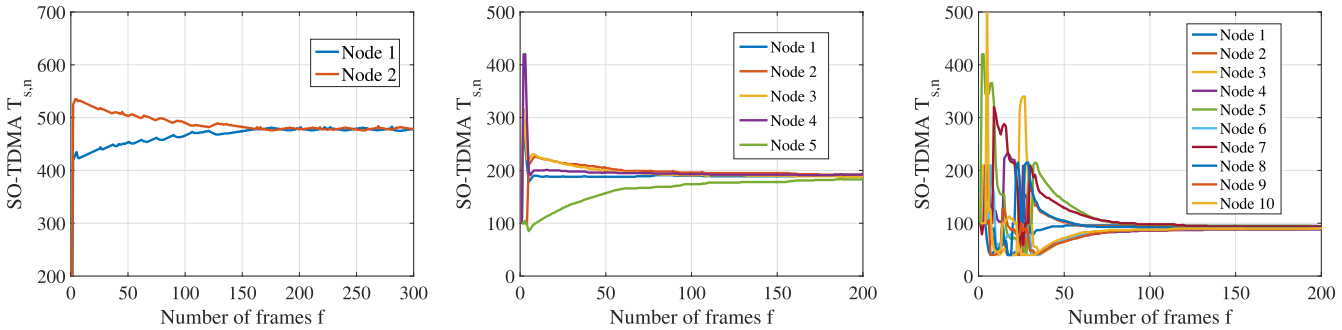
It should be noted that Poisson packet arrival is selected in most of the numerical results since it would represent the worse case compared with the constant bit rate (CBR) traffic according to Fig. 5. However, we present some results to evaluate the performance of SO-TDMA with CBR traffic model as well.

The parameter settings used for all MAC protocols are summarized in Tables 1 and 2. The values of T, CW_{min}, and CW_{max} chosen in Table 1 and the data rates listed in Table 2 are based on OFDM PHY specifications given in Sections 18.4.4 and 18.1.1 of [9], respectively. Moreover, the values of SIFS and DIFS are chosen as the multiples of the time slot size (i.e., T) for the sake of simplicity in implementation. The relations between T, SIFS and DIFS have been largely kept as in OFDM PHY specifications, although there are minimal differences in SIFS and DIFS from values in [9].

B. SO-TDMA CONVERGENCE

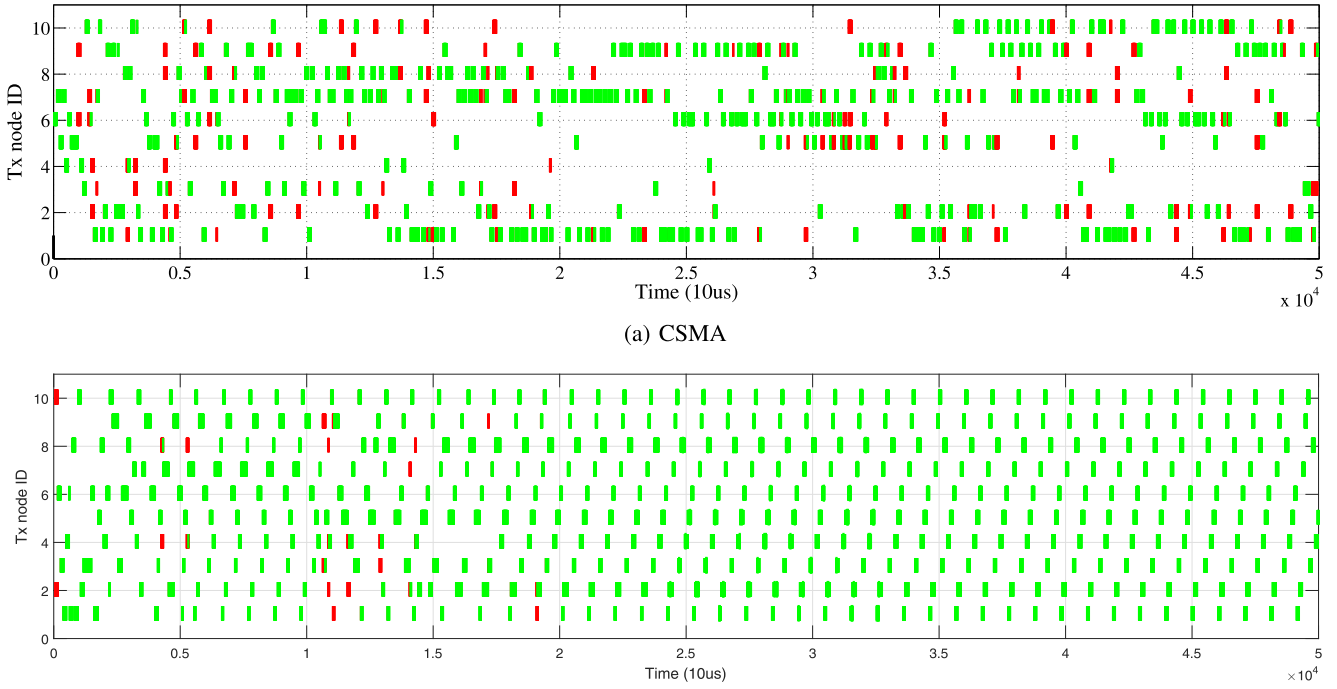
Fig. 3 illustrates the convergence process of transmission time-slot of different nodes using Algorithm 1 for different values of N_{a}, assuming $W_I = 5$ time-slots and $W_D = 0.05$. It is shown that $T_{\mathrm{s},n}$ in SO-TDMA protocol can converge to the target value (T_f/N_{a}) without knowledge of N_{a}. For clear illustration of the algorithm performance, we consider the saturated network scenario where all nodes have the non-empty queues. Therefore, the fluctuations in N_{a} due to traffic have been ignored in this figure. It can be seen that for different values of N_{a}, SO-TDMA converges to the ideal value in merely *a few pseudo-frames*. However, the convergence time increases with increasing N_{a}.

Fig. 4 shows an snapshot of the *resource allocation* procedure in CSMA and SO-TDMA along with successful transmissions (light green) and collisions (dark red) under saturated traffic conditions, assuming $W_I = 5$ time-slots and $W_D = 0.05$. Fig. 4(b) confirms how quickly SO-TDMA protocol approaches a network stable state (i.e., Algorithm 1 converges), while it eliminates the collision probability among nodes. However, Fig. 4(a) shows that there exists a non-zero collision probability in CSMA, resulting in poor network performance. It can also be observed that the fairness in CSMA is worse in the saturated condition as compared to SO-TDMA. For example, in Fig. 4(a), the nodes 1 and 7 hold the channel for much longer time as compared to nodes 3 and 4 that suffer from starvation. In contrast, in SO-TDMA, Fig. 4(b) shows that fairness is improved where all nodes transmit sequentially in the network stable state (achieved in 200 ms) and approach the same time-slot length as shown in Fig. 3. This approves how the proposed algorithm without additional complexity can result in more stable MTC networks in terms of less collisions and can lead to higher throughput. Now, we can look at the performance metrics and compare these protocols under different traffic scenarios.



(a) N_a=2 (b) N_a=5 (c) N_a=10

FIGURE 3. Convergence of T_s for $N_a = 2$, 5 and 10 vs. pseudo frames in SO-TDMA.



(a) CSMA

(b) SO-TDMA

FIGURE 4. Channel state vs. time ($N_a = 10$, dark red: collision, light green: successful transmission).

C. EFFECTIVE CAPACITY AND DELAY OUTAGE PROBABILITY

In order to evaluate the performance for QoS support, we empirically calculate EC of a network in order to compare the QoS provisioning capabilities of different MAC protocols. EC was first introduced in [25] as a QoS-aware metric that determines the maximum constant arrival rate that can be supported by a network, while satisfying a target statistical delay requirement. In particular, EC is defined subject to a maximum tolerable *delay-outage* probability (i.e., probability that packet delay exceeds a target delay bound (e.g., $D_{\max}$). Note that the average delay is not a key metric to assess the performance in terms of delay for emerging applications since some packets might be immediately served and some with a large delay with a certain mean average delay. This paper rather focuses on the delay outage probability, i.e, the probability that the experienced delay of each packet exceeds a certain threshold.

In [25], the delay-outage probability is approximated as

$$\Pr(D(t) \geq D_{\max}) \approx \gamma(\mu) e^{-\theta(\mu) D_{\max}}, \tag{34}$$

where $D(t)$ represents the total delay experienced by a packet, including the queuing delay and the negligible channel service delay (i.e., RF propagation/transmission) at time-slot t, γ denotes the probability of non-empty queue, and θ represents the delay-exponent. It should be noted that both γ and θ are functions of the constant traffic arrival rate μ.

The pair of $\{\gamma(\mu), \theta(\mu)\}$ characterizes the queue behavior of a time-varying service process being offered a constant arrival rate μ. In [25], [35], it is shown that these parameters can be estimated measuring the average delay sensed by a packet as

$$\gamma(\mu)/\theta(\mu) = \mathbb{E}[D(t)]. \tag{35}$$

In order to empirically measure EC, an estimation procedure is proposed in [25], [36]. In this method, $\gamma(\mu)$ and $\theta(\mu)$ can be estimated by taking average over N_s samples collected during an interval of length T_m from all nodes as

$$\hat{\gamma} = \frac{1}{N_s}\sum_{k=1}^{N_s} q_k, \tag{36}$$

$$\hat{d} = \frac{1}{N_s}\sum_{k=1}^{N_s} D_k. \tag{37}$$

where $q_k \in \{0, 1\}$ is the indicator of whether or not a packet is in queue and D_k is the total delay experienced by a packet at the k^{th} sampling epoch. Based on (35)-(37), the estimated delay-exponent becomes

$$\hat{\theta} = \hat{\gamma}/\hat{d}. \tag{38}$$

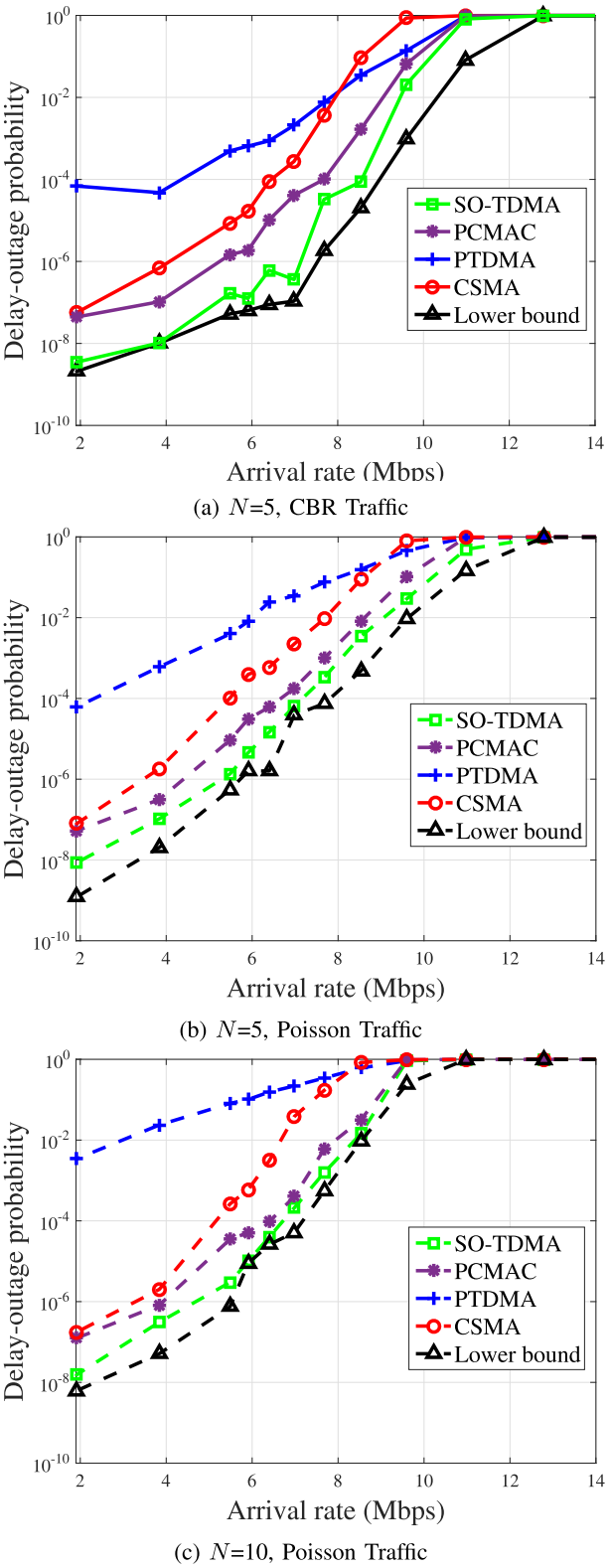
Consequently, employing $\hat{\gamma}$ and $\hat{\theta}$, the delay-outage probability in (34) can be approximated as

$$\Pr(D(t) \geq D_{\max}) \approx \hat{\gamma} e^{-\hat{\theta} D_{\max}} \tag{39}$$

Given a delay bound $D_{\max}$, based on (39), supplying a traffic source with a constant rate μ shall result in a certain delay-outage probability. Thus, via a bisection search method of different values of μ, we can empirically compute the EC, i.e., the maximum value of μ corresponding to the target delay-outage probability threshold ε.

Fig. 5 shows the delay-outage probability in different protocols, measured based on (39) versus μ_{sys}. Two traffic models are examined by feeding nodes with (i) CBR traffic and (ii) Poisson traffic. For the same load (mean traffic rate), Poisson traffic introduces higher delay-outage probability due to its variability/burstiness. For both traffic types, at high load, PTDMA outperforms CSMA since CSMA suffers from a large number of collisions, which are avoided in PTDMA. Fig. 5 also includes the case of Ideal-PTDMA (in which the time-slot size $T_{s,n}$ is adapted based on N_a) to provide the *lower bound* on the delay-outage probability. The results reveal a need for a *self-organizing* PTDMA protocol that is able to derive this information fast enough in a distributed manner to approach the *lower bound*.

Interestingly, the proposed SO-TDMA achieves a performance very close to the lower bound and outperforms CSMA, PTDMA and PCMAC. Thus, for a fixed acceptable threshold for delay-outage probability (as QoS requirements), much larger effective capacity can be supported in SO-TDMA than CSMA and PTDMA. For example, Fig. 5(c) shows that to maintain a delay-outage probability not exceeding 10^{-3}, SO-TDMA offers an effective capacity of 7.6 Mbps or 95% of the lower-bound of 8 Mbps, while CSMA achieves 75% and PTDMA has only 15% of the lower-bound of 8 Mbps. This result shows the effectiveness of SO-TDMA in supporting QoS of users in MTC networks without the need for

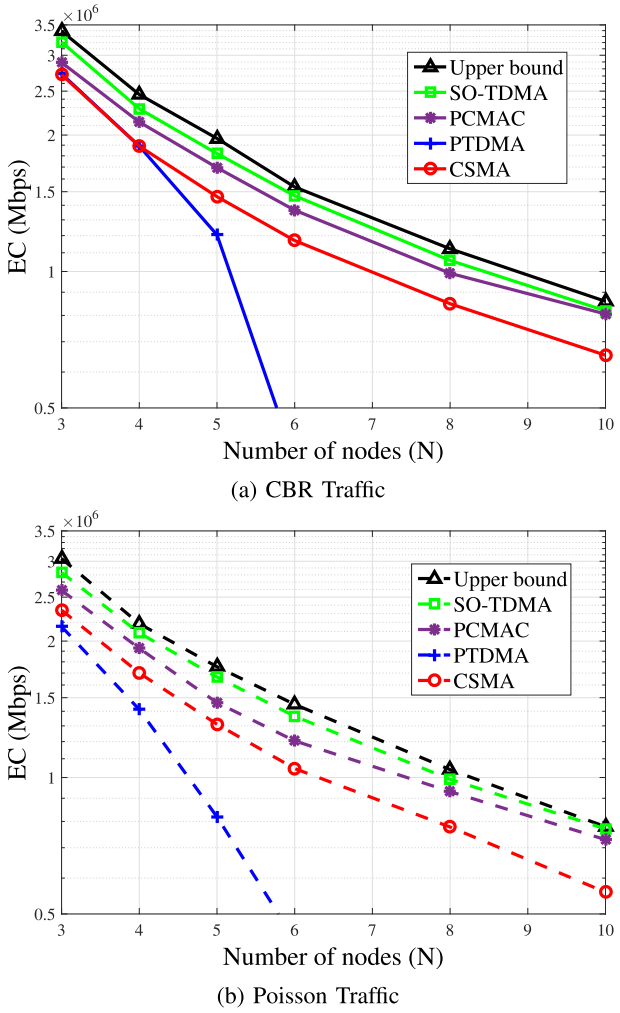


(a) N=5, CBR Traffic

(b) N=5, Poisson Traffic

(c) N=10, Poisson Traffic

FIGURE 5. Delay-outage probability vs. arrival rate μ_{sys} (Mbps).



(a) CBR Traffic

(b) Poisson Traffic

FIGURE 6. Effective capacity vs. number of nodes.

any additional message passing between nodes or a central entity.

The comparison in terms of EC (normalized per number of nodes) is shown in Fig. 6(a) and Fig. 6(b) plotted against the number of nodes in the network (N) for both CBR and Poisson traffic arrival, respectively. The EC of SO-TDMA protocol approaches the upper bound (i.e., Ideal-PTDMA) and is significantly better than CSMA. Hence, in terms of QoS, the SO-TDMA protocol shows better performance than CSMA, PTDMA and PCMAC for the different values of N. For example, it can be observed from Fig. 6(a) and Fig. 6(b) that the performance gain of SO-TDMA over CSMA is 15-40% in terms of effective capacity. In other words, 15-40% additional multimedia or delay sensitive services can be catered for with SO-TDMA MAC protocol as compared to CSMA with the same QoS requirements.

D. SYSTEM THROUGHPUT

In order to evaluate spectral efficiency provided by different MAC protocols, we study the maximum achievable system

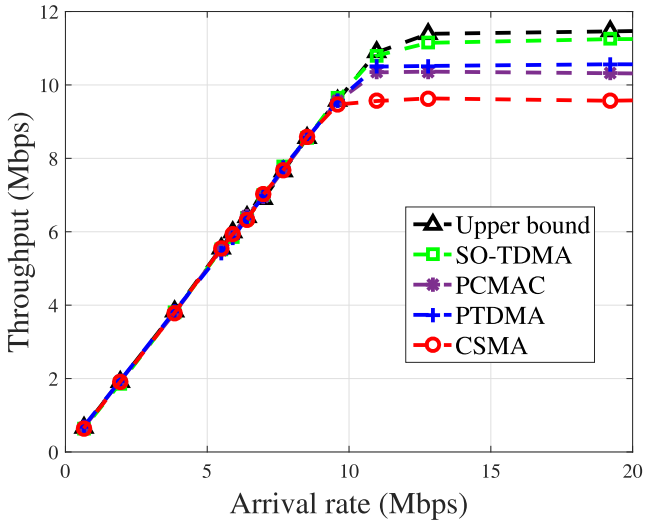


FIGURE 7. System throughput vs. arrival rate μ_{sys} (Mbps), $N = 5$.

throughput (denoted by S_{sys}) that is defined as

$$S_{\text{sys}} = \sum_{n=1}^{N} S_n, \tag{40}$$

where S_n is the up-link throughput for node n. The throughput of each node is defined as the rate of successful packet transmission. Assuming a fixed packet size (i.e., P_{s}), S_n is empirically calculated as $(K_{\text{suc},n} \times P_{\text{s}})/T_{\text{m}}$ where $K_{\text{suc},n}$ is the number of packets successfully transmitted by node n and T_{m} is the measurement period.

Fig. 7 plots S aggregated over all nodes in the network as μ_{sys} is varied between 0 to 20 Mbps. When the load (i.e., arrival rate) is low the system throughput S_{sys} in all the protocols increases linearly with μ_{sys} as expected. However, as the system load increases further in a higher range, the system throughput reaches saturation at 9.5 Mbps for CSMA, 10.3 for PCMAC, 10.5 Mbps for PTDMA, 11.25 Mbps for SO-TDMA, and 11.5 Mbps for upper-bound. The proposed SO-TDMA outperforms CSMA, PTDMA and PCMAC.

E. FAIRNESS

It is known that CSMA suffers from poor fairness performance in high traffic load conditions as studied in [37] and [38]. In case of PTDMA, the resource allocation of each node is equal thus achieving high fairness at pseudo-frame level. However, for SO-TDMA the percentage of channel occupied by an active node in the network is adaptive and optimized independently by each active node as opposed to PTDMA. Therefore, it is very important to look at its fairness performance as compared to other existing protocols.

To illustrate the short-term achievable fairness of the MAC protocols versus N, the Jain's index $J(S_1, S_2, \ldots, S_{\text{N}})$ is computed over non-overlapping 2-second periods and then averaged and plotted in Fig. 8. The results indicate that SO-TDMA can keep short-term fairness reasonably high with Jain's fairness index$>$ 0.9 despite its dynamic

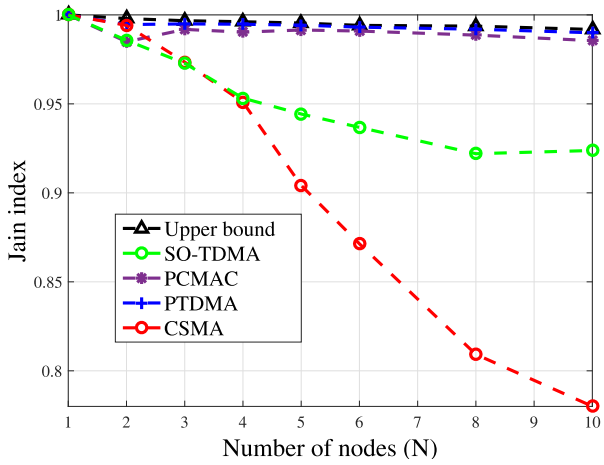


FIGURE 8. Short-term fairness vs. number of nodes.

behavior, confirming the monotonic convergence to fairness of SO-TDMA as stated in Proposition 2. In PTDMA and PCMAC, due to their fixed time-slot durations, every node will have a slot in the frame and apart from some occasional collisions, their fairness is better, close to perfect. On the other hand, SO-TDMA outperforms these protocols in other performance metrics: delay-outage probability (Fig. 5), effective capacity (Fig. 6), throughput (Fig. 7).

F. COLLISION PROBABILITY

One cause for a low system throughput can be a large number of collisions, which are inevitable in random access schemes. Therefore, the collision probability of a contention-based MAC protocol is an important measure to assess its performance. The network collision probability can empirically be calculated as

$$P_{\mathrm{c}} = \frac{1}{N}\sum\nolimits_{n=1}^{N} \frac{K_{\mathrm{col},n}}{K_{\mathrm{tot},n}}, \tag{41}$$

where $K_{\mathrm{col},n}$ is the number of collided packets of node n and $K_{\mathrm{tot},n}$ is the total number of packets transmitted by node n.

Collision probability in the different protocols is shown in Fig. 9 for $N = 5$ and $N = 10$. PTDMA has the lowest collision probability. SO-TDMA has significantly lower collision probability than CSMA. Compared to PCMAC, SO-TDMA has has lower collision probability for $N = 5$, and higher collision probability for $N = 10$ with high load (i.e., arrival rate>8Mbps). But, it should be noted that collision probability of SO-TDMA is still reasonably small for $N = 10$ (max 0.3 at worst case). This is mainly because PTDMA and PCMAC operate with a fixed time-slot length, while the time-slot length is dynamic in SO-TDMA.

Different from PTDMA with a fixed structure, SO-TDMA needs to learn the frame structure. Thus, increasing the number of nodes as well as arrival rates leads to larger convergence times, making it more difficult to approach stability. Thus, a larger number of collisions will happen.

PCMAC is more dynamic than PTDMA as the frame length is adaptive to the number of users. Compared

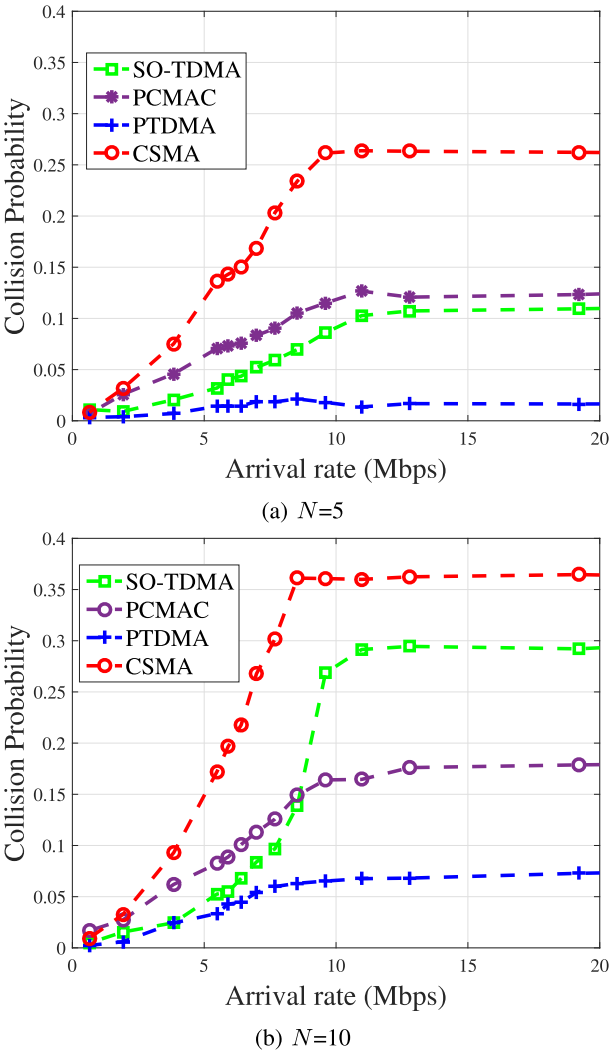


FIGURE 9. Collision probability vs. arrival rate μ_{sys} (Mbps).

to SO-TDMA, since PCMAC can stretch out the frame-length relative to the number of users, each frame will be less congested, and hence, fewer collisions will happen. However, effective capacity would suffer from stretching the frame-length due to the longer delays as shown in Fig. 6.

It should be noted that with a slightly greater risk of collisions compared to PTDMA and PCMAC, much better performance in terms of EC has been achieved in SO-TDMA.

G. IMPACT OF USER DYNAMICS

To study the impacts of user dynamics on SO-TDMA, Fig. 10 illustrates the convergence process of transmission time-slot of different nodes using SO-TDMA when more nodes join a network. This result shows the transient behavior of SO-TDMA and confirms that SO-TDMA can shortly converge to a stable state after a change in the number of users. Initially, the network has 8 active nodes, and then every 5 seconds, another node joins the network. The plots for

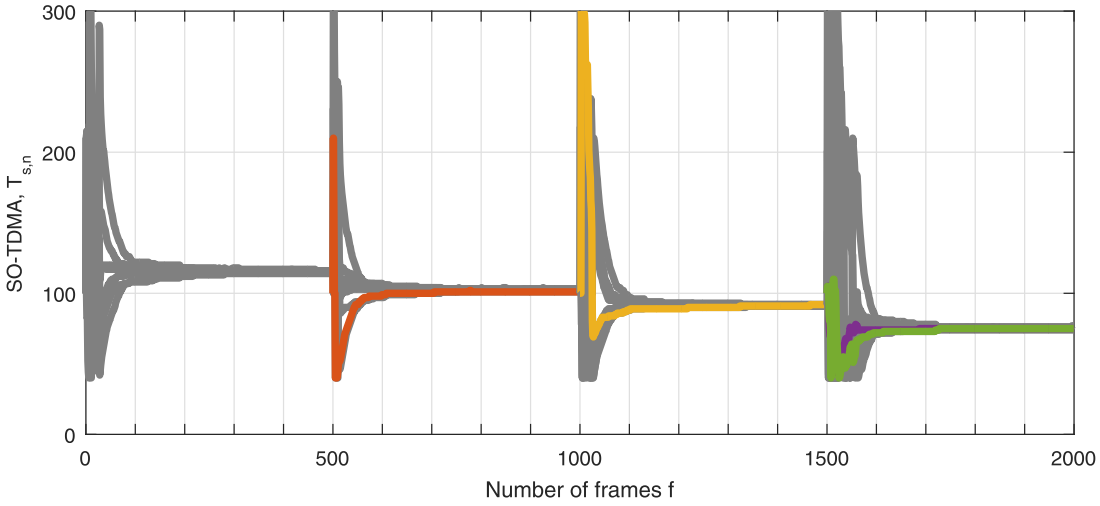


FIGURE 10. Convergence of T_S vs. pseudo frames in SO-TDMA when new nodes join a network of eight nodes every 5 seconds.

newly joined nodes are in colors different from grey. It can be observed that the convergence time of SO-TDMA is short (approximately 1 second). Even though a drastically mobile situation is considered with a very short period of change, the user dynamics could not cause any instability in the network. In the last period, to check how sensitive the stability is to the number of changes, two nodes simultaneously join the network. It is shown that SO-TDMA converges to the ideal value of $T_{s,n}$ in a few pseudo frames.

Overall, SO-TDMA outperforms CSMA, PTDMA and PCMAC in terms of EC, delay-outage probability, and system throughput. This is because of the improved and self-organized channel-access by all nodes while avoiding collisions. All the results of this section along with aforementioned discussions show how SO-TDMA can improve the spectrum utilization in MTC networks.

VI. CONCLUSION

To support QoS guarantee for dynamic channel access in MTC networks, we have proposed the distributed and self-organizing protocol SO-TDMA that utilizes both CSMA and TDMA concepts to form a dynamic, asynchronous pseudo-frame structure for transmission and access. An AIMD algorithm is proposed for time-slot adaptation process of SO-TDMA, which monotonically converges to fairness. It has been proved that the convergence time required for passing the initial phase in SO-TDMA and moving to the periodic transmission phase is only linearly increasing with respect to the number of nodes in the network. Simulation results illustrate the effectiveness of the proposed SO-TDMA in improving channel utilization and effective capacity for MTC networks. The proposed SO-TDMA can approach the performance in the ideal case which needs complete and precise information about the queue length and the channel conditions of all nodes.

APPENDIX
PROOF OF LEMMA 1

We prove Lemma 1 by induction. First, considering inequality in (28), for $N = 2$ and $N = 3$, we have

$$\Theta_2 \geq \Theta_1 - \frac{1}{M}\Theta_0 = \frac{M-1}{M}. \tag{42}$$

$$\Theta_3 \geq \Theta_2 - \frac{2}{M}\Theta_1 \geq \frac{M-3}{M}. \tag{43}$$

Considering that $\Theta_0 = \Theta_1 = 1$, the statement in (29) holds for $N = 2$ and $N = 3$. Let assume (29) holds for $N-2$ and $N-1$. Then, considering (28) for N, we have

$$\begin{aligned}\Theta_N &\geq \Theta_{N-1} - \frac{N-1}{M}\Theta_{N-2} \\ &\geq 1 - \frac{(N-1)(N-2)}{2M} - \frac{N-1}{M}\left(1 - \frac{(N-2)(N-3)}{2M}\right) \\ &\geq 1 - \frac{(N-1)(N-2)}{2M} - \frac{N-1}{M} = 1 - \frac{N(N-1)}{2M}. \quad \square\end{aligned}$$

REFERENCES

[1] S. Choi, K. Park, and C.-K. Kim, "On the performance characteristics of WLANs: Revisited," in *Proc. ACM Int. Conf. Meas. Modeling Comput. Syst. (SIGMETRICS)*, 2005, pp. 97–108.

[2] J. W. Robinson and T. S. Randhawa, "Saturation throughput analysis of IEEE 802.11e enhanced distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 5, pp. 917–928, Jun. 2004.

[3] D. He and C. Q. Shen, "Simulation study of IEEE 802.11e EDCF," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, vol. 1, Apr. 2003, pp. 685–689.

[4] *IEEE Standard for Information Technology—Telecommunications and Information Exchange Between Systems—Local and Metropolitan Area Networks—Specific Requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 2: Sub 1 GHz License Exempt Operation*, IEEE Standard 802.11ah-2016 (Amendment to IEEE Std 802.11-2016, as amended by IEEE Std 802.11ai-2016), May 2017, pp. 1–594.

[5] L. Tian, J. Famaey, and S. Latré, "Evaluation of the IEEE 802.11ah restricted access window mechanism for dense IoT networks," in *Proc. IEEE 17th Int. Symp. World Wireless, Mobile Multimedia Netw. (WoWMoM)*, Jun. 2016, pp. 1–9.

[6] B. Badihi, L. F. Del Carpio, P. Amin, A. Larmo, M. Lopez, and D. Denteneer, "Performance evaluation of IEEE 802.11ah actuators," in *Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, May 2016, pp. 1–5.

[7] A. Šljivo, D. Kerkhove, L. Tian, J. Famaey, A. Munteanu, I. Moerman, J. Hoebeke, and E. D. Poorter, "Performance evaluation of IEEE 802.11ah networks with high-throughput bidirectional traffic," *Sensors*, vol. 18, no. 2, p. 325, 2018.

[8] A. Šljivo, I. Moerman, E. De Poorter, and J. Hoebeke, "Evaluating the suitability of IEEE 802.11ah for low-latency time-critical control loops," *IEEE Internet Things J.*, to be published.

[9] *Wireless LAN Medium Access Control (MAC) Physical Layer (PHY) Specifications*, IEEE Standard 802.11, IEEE Computer Society, 2012.

[10] H. Zhu, M. Li, I. Chlamtac, and B. Prabhakaran, "A survey of quality of service in IEEE 802.11 networks," *IEEE Wireless Commun.*, vol. 11, no. 4, pp. 6–14, Aug. 2004.

[11] E. Charfi, L. Chaari, and L. Kamoun, "PHY/MAC Enhancements and QoS Mechanisms for Very High Throughput WLANs: A Survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 1714–1735, 4th Quart., 2013.

[12] M. Derakhshani and T. Le-Ngoc, "Cognitive MAC designs: Background," in *Cognitive MAC Designs for OSA Networks*. Cham, Switzerland: Springer, 2014, pp. 15–31.

[13] J. Barcelo, B. Bellalta, C. Cano, and M. Oliver, "Learning-BEB: Avoiding collisions in WLAN," in *Proc. Int. Fed. Inf. Process. Digit. Library, 14th Eunice Open Eur. Summer School*, Jan. 2008, pp. 1–132.

[14] J. Lee and J. C. Walrand, "Design and analysis of an asynchronous zero collision MAC protocol," 2008, *arXiv:0806.3542*. [Online]. Available: https://arxiv.org/abs/0806.3542

[15] M. Fang, D. Malone, K. R. Duffy, and D. J. Leith, "Decentralised learning MACs for collision-free access in WLANs," *Wireless Netw.*, vol. 19, pp. 83–98, Jan. 2010.

[16] Y.-W. Kuo and J.-H. Huang, "A CSMA-based MAC protocol for WLANs with automatic synchronization capability to provide hard quality of service guarantees," *Comput. Netw.*, vol. 127, pp. 31–42, Nov. 2017.

[17] A. Laya, C. Kalalas, F. Vazquez-Gallego, L. Alonso, and J. Alonso-Zarate, "Goodbye, ALOHA!" *IEEE Access*, vol. 4, pp. 2029–2044, 2016.

[18] Y. Maraden, W. Hardjawana, and B. Vucetic, "Contention resolution algorithm for industrial Internet-of-Things networks," in *Proc. IEEE 4th World Internet Things (WF-IoT)*, Feb. 2018, pp. 493–498.

[19] Y.-J. Chang, W. Jin, and S. Pettie, "Simple contention resolution via multiplicative weight updates," in *Proc. 2nd Symp. Simplicity Algorithms (SOSA)*, in OpenAccess Series in Informatics (OASIcs), vol. 69, J. T. Fineman and M. Mitzenmacher, Eds. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018.

[20] G. Jakllari, M. Neufeld, and R. Ramanathan, "A framework for frameless TDMA using slot chains," in *Proc. IEEE Int. Conf. Mobile Adhoc Sensor Syst. (MASS)*, Oct. 2012, pp. 56–64.

[21] X. Chen, Z. Chen, and Y. Wu, "Leveraging pseudo-TDMA to a controllable and bandwidth-efficient village area WiFi networks," in *Proc. Int. Conf. Wireless Commun., Netw. Mobile Comput.*, Oct. 2008, pp. 1–5.

[22] I. Tinnirello and P. Gallo, "Supporting a pseudo-TDMA access scheme in mesh wireless networks," in *Wireless Access Flexibility*, vol. 8072. Berlin, Germany: Springer, 2013, pp. 80–92.

[23] G. S. Paschos, I. Papapanagiotou, S. A. Kotsopoulos, and G. K. Karagiannidis, "A new MAC protocol with Pseudo-TDMA behavior for supporting quality of service in 802.11 wireless LANs," *EURASIP J. Wireless Commun. Netw.*, vol. 2006, no. 1, Dec. 2006, Art. no. 065836.

[24] Q. Wei, I. Aad, L. Scalia, J. Widmer, P. Hofmann, and L. Loyola, "E-MAC: An elastic MAC layer for IEEE 802.11 networks," *Wireless Commun. Mobile Comput.*, vol. 13, no. 4, pp. 393–409, Mar. 2013.

[25] D. Wu and R. Negi, "Effective capacity: A wireless link model for support of quality of service," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 630–643, Jul. 2003.

[26] Y. Khan, M. Derakhshani, S. Parsaeefard, and T. Le-Ngoc, "Self-organizing TDMA MAC protocol for effective capacity improvement in IEEE 802.11 WLANs," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2015, pp. 1–6.

[27] M. Lee, Y. Kim, and C.-H. Choi, "Period-controlled MAC for high performance in wireless networks," *IEEE/ACM Trans. Netw.*, vol. 19, no. 4, pp. 1237–1250, Aug. 2011.

[28] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535–547, Mar. 2000.

[29] D.-M. Chiu and R. Jain, "Analysis of the increase and decrease algorithms for congestion avoidance in computer networks," *Comput. Netw. ISDN Syst.*, vol. 17, no. 1, pp. 1–14, 1989.

[30] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *J. Oper. Res. Soc.*, vol. 49, pp. 237–252, Mar. 1998.

[31] D. Loguinov and H. Radha, "End-to-end rate-based congestion control: Convergence properties and scalability analysis," *IEEE/ACM Trans. Netw.*, vol. 11, no. 4, pp. 564–577, Aug. 2003.

[32] J. L. Katz and R. L. Burford, "Estimating mean time to absorption for a Markov chain with limited information," *Appl. Stochastic Models Data Anal.*, vol. 4, no. 4, pp. 217–230, 1988.

[33] Y. Huang and W. F. McColl, "Analytical inversion of general tridiagonal matrices," *J. Phys. A, Math. Gen.*, vol. 30, no. 22, p. 7919, 1997.

[34] T. Juhana and R. Hersyandika, "Experimental study of TCP unfairness in IEEE802.11-based ad hoc network," in *Proc. Int. Conf. Telecommun. Syst., Services, Appl. (TSSA)*, Oct. 2011, pp. 191–194.

[35] B. L. Mark and G. Ramamurthy, "Real-time estimation and dynamic renegotiation of UPC parameters for arbitrary traffic sources in ATM networks," *IEEE/ACM Trans. Netw.*, vol. 6, no. 6, pp. 811–827, Dec. 1998.

[36] A. Davy, B. Meskill, and J. Domingo-Pascual, "An empirical study of effective capacity throughputs in 802.11 wireless networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2012, pp. 1770–1775.

[37] Y. Jian, M. Zhang, and S. Chen, "Achieving MAC-layer fairness in CSMA/CA networks," *IEEE/ACM Trans. Netw.*, vol. 19, no. 5, pp. 1472–1484, Oct. 2011.

[38] T. Nandagopal, T.-E. Kim, X. Gao, and V. Bharghavan, "Achieving MAC layer fairness in wireless packet networks," in *Proc. Int. Conf. Mobile Comput. Netw.*, 2000, pp. 87–98.

MAHSA DERAKHSHANI (S'10–M'13) received the B.Sc. and M.Sc. degrees from the Sharif University of Technology, Tehran, Iran, in 2006 and 2008, respectively, and the Ph.D. degree from McGill University, Montréal, Canada, in 2013, all in electrical engineering. From 2013 to 2015, she was a Postdoctoral Research Fellow with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada, and a Research Assistant with the Department of Electrical and Computer Engineering, McGill University. From 2015 to 2016, she was an Honorary NSERC Postdoctoral Fellow with the Department of Electrical and Electronic Engineering, Imperial College London. She is currently a Lecturer (Assistant Professor) in digital communications with the Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University. Her current research interests include non-orthogonal multiple access (NOMA), ultra-reliable low-latency communications (URLLC), software-defined wireless networking, and the applications of optimization and machine learning for radio resource management. Dr. Derakhshani received the John Bonsall Porter Prize, the McGill Engineering Doctoral Award, and the Fonds de Recherche du Québec-Nature et Technologies (FRQNT) and Natural Sciences and Engineering Research Council of Canada (NSERC) Postdoctoral Fellowships. She currently serves as an Associate Editor of the IET *Signal Processing Journal*.

YAHYA KHAN received the M.Eng. degree in electrical engineering from McGill University, Montreal, QC, Canada, in 2015, under the supervision of Prof. T. Le-Ngoc. He is currently a Wireless Network Architect with SSi Micro Ltd., Ottawa, ON, Canada. His current research interests include MAC design and dynamic resource allocation for next-generation wireless networks.

DUC TUONG NGUYEN received the B.Eng. degree in electrical and electronics engineering from the University of Danang–University of Science and Technology, Danang, Vietnam, in 2017. Since 2019, he has been with the Department of Electrical and Computer Engineering, McGill University, Canada, where he is currently pursuing the M.Eng. degree. His current research interests include multiple access in machine-type communications, and the applications of optimization and machine learning into multiple access.

SAEEDEH PARSAEEFARD (S'09–M'14– SM'18) received the B.Sc. and M.Sc. degrees from the Amirkabir University of Technology (TehranPolytechnic), Tehran, Iran, in 2003 and 2006, respectively, and the Ph.D. degree in electrical and computer engineering from Tarbiat Modares University, Tehran, in 2012. She was a Postdoctoral Research Fellow with the Telecommunication and Signal Processing Laboratory, Department of Electrical and Computer Engineering, McGill University, Canada. From 2010 to 2011, she was a Visiting Ph.D. Student with the Department of Electrical Engineering, University of California at Los Angeles, Los Angeles, CA, USA. She is currently a Faculty Member with the Iran Telecommunication Research Center and a Visiting Faculty Member with the University of Toronto. Her current research interests include the applications of AI and ML in future cognitive network management, resource management in software-defined networking, the Internet of Things, and the fifth generation of wireless networks, and applications of robust optimization theory and game theory on the resource allocation and management in wireless networks. She received the IEEE Iran Section Women in Engineering (WIE) Award, in 2018.

ATOOSA DALILI SHOAEI received the B.Sc. degree in information technology engineering from the Isfahan University of Technology, Isfahan, Iran, in 2009, the M.Sc. degree in information technology engineering from the Amirkabir University of Technology, Tehran, Iran, in 2012, and the Ph.D. degree in electrical engineering from McGill University, in 2019, where he is currently a Postdoctoral Fellow. His current research interests include medium access control techniques, the Internet of Things (IoT), and wireless virtualization.

THO LE-NGOC (F'97) received the B.Eng. degree (Hons.) in electrical engineering and the M.Eng. degree from McGill University, Montréal, QC, Canada, in 1976 and 1978, respectively, and the Ph.D. degree in digital communications from the University of Ottawa, Canada, in 1983. From 1977 to 1982, he was a Research and Development Senior Engineer with Spar Aerospace Ltd., Sainte-Anne-de-Bellevue, Canada, where he was involved in the development and design of satellite communications systems. From 1982 to1985, he was the Engineering Manager with the Radio Group, Department of Development Engineering, SR Telecom Inc., Saint-Laurent, QC, where he developed the new point-to-multipoint DA-TDMA/TDM subscriber radio system SR500. From 1985 to 2000, he was a Professor with the Department of Electrical and Computer Engineering, Concordia University, Montréal, Canada. Since 2000, he has been with the Department of Electrical and Computer Engineering, McGill University. His current research interest includes broadband digital communications. He is a Fellow of the Engineering Institute of Canada, the Canadian Academy of Engineering, and the Royal Society of Canada. He was a recipient of the 2004 Canadian Award in telecommunications research and the IEEE Canada Fessenden Award, in 2005. He is a Canada Research Chair (Tier I) in broadband access communications.