MAC for Machine-Type Communications in Industrial IoT—Part I: Protocol Design and Analysis

Jie Gao, *Member, IEEE*, Weihua Zhuang, *Fellow, IEEE*, Mushu Li, *Student Member, IEEE*, Xuemin Shen, *Fellow, IEEE*, and Xu Li, *Member, IEEE*

***Abstract*—In this two-part paper, we propose a novel medium access control (MAC) protocol for machine-type communications in the Industrial Internet of Things. The considered use case features a limited geographical area and a massive number of devices with sporadic data traffic and different priority types. We target supporting the devices while satisfying their Quality-of-Service (QoS) requirements with a single access point and a single channel, which necessitates a customized design that can significantly improve the MAC performance. In Part I of this paper, we present the MAC protocol that comprises a new slot structure, corresponding channel access procedure, and mechanisms for supporting high device density and providing differentiated QoS. A key idea behind this protocol is sensing-based distributed coordination for significantly improving channel utilization. To characterize the proposed protocol, we analyze its delay performance based on the packet arrival rates of devices. The analytical results provide insights and lay the groundwork for the fine-grained scheduling with QoS guarantee as presented in Part II.**

***Index Terms*—Industrial Internet of Things (IIoT), machine-type communications, medium access control (MAC), protocol design.**

I. Introduction

NEXT-GENERATION wireless communications are envisioned to empower vertical industries [1]. Among potential enterprise use cases, industrial communication networks for applications such as factory automation and process control are particularly important as they play a crucial role in Industrial 4.0, the upcoming fourth industrial revolution [2]. An indispensable step toward Industrial 4.0 is the development and standardization of the Industrial Internet of Things (IIoT), which connects devices, such as sensors, actuators, and controllers, to facilitate data collection and analysis and to support applications such as factory automation [3].

As in the general Internet of Things (IoT), machine-type communication (MTC), which facilitates automated data communications among devices, is a primary enabler of the IIoT. According to 3GPP, features of MTC include low mobility, small data packets, and time-controlled access [4]. Meanwhile, compared with the general IoT, IIoT has some unique characteristics: first, connectivity in IIoT is usually structured, featuring centralized network management and devices at fixed locations [5]; second, IIoT scenarios generally involve densely deployed devices in a relatively limited area. For example, the IIoT application of process monitoring may involve a density of up to 10 000 devices per square kilometer [6]; and third, certain IIoT applications can be mission-critical and have extremely stringent Quality-of-Service (QoS) requirements. For example, the communication latency tolerance for machine tool motion control can be as low as 0.5 ms [7]. The last two characteristics pose a significant challenge for supporting MTC in IIoT. Specifically, within a limited geographical area such as in a factory, the communication network may need to support a massive number of devices and, simultaneously, satisfy exceptionally strict QoS requirements for some devices.

In recent years, standards such as LTE-M, NB-IoT, and IEEE 802.11ah have been developed for supporting MTC. However, these standards usually focus on the general IoT instead of the IIoT use cases. As a result, they may not meet the stringent QoS requirements in IIoT, e.g., millisecond-level or submillisecond-level delay. For example, LTE-M and NB-IoT, both targeting at low power and long-range communications, are more concerned with bandwidth usage and power consumption than latency. Specifically, the latency of LTE-M is no less than 10 ms, while the latency of NB-IoT can be up to 10 s [8]. The IEEE 802.11ah can support a less than 10-ms latency but only under a low-load condition, which limits its supported device density in practice [9].

To address challenges in supporting high device density and stringent QoS requirements, various solutions have been proposed for different layers in the network protocol stack. At the physical layer, exploiting the vast spectrum resource beyond 30 GHz or even in the Terahertz

Manuscript received August 12, 2020; revised November 16, 2020 and December 14, 2020; accepted January 6, 2021. Date of publication January 12, 2021; date of current version June 7, 2021. This work was supported in part by the Research Grants from Huawei Technologies Canada and in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada. *(Corresponding author: Jie Gao.)*

Jie Gao is with the Department of Electrical and Computer Engineering, Marquette University, Milwaukee, WI 53233 USA (e-mail: j.gao@marquette.edu).

Weihua Zhuang, Mushu Li, and Xuemin Shen are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: wzhuang@uwaterloo.ca; mushu.li@uwaterloo.ca; sshen@uwaterloo.ca).

Xu Li is with the Department of CRC, Huawei Technologies Canada Inc., Ottawa, ON K2K 3J1, Canada (e-mail: xu.lica@huawei.com).

Digital Object Identifier 10.1109/JIOT.2021.3051181

band potentially provides support for a high device density [10], [11]. In addition, nonorthogonal multiple access could be a solution [12], [13]. For example, compressed sensing-based multiuser detection has been investigated for massive MTC (mMTC) [14], [15]. However, the advances in the physical layer alone may not be sufficient for meeting the demand of MTC in IIoT for two reasons: first, the prevalence of IIoT relies on low-complexity and low-budget devices that may not support advanced physical-layer techniques and second, even if physical-layer solutions can be applied, there is no guarantee that they will meet the stringent QoS requirements of IIoT. Therefore, developing a reliable link-layer solution becomes appealing as it is less limited by hardware and can be implemented with physical-layer solutions.

There have been extensive research efforts in supporting MTC by link-layer design, despite few targeting specifically at IIoT. For cellular-based MTC, the random access (RA) procedure for the connection setup is a bottleneck that many medium access control (MAC) designs aim to address. When a large number of devices seek to set up connections in the RA procedure, the network can be congested [16]. MAC designs have been proposed with a focus on grouping and prioritizing devices, e.g., distributing devices into different groups after collisions [17], grouping devices based on their delay requirements [18], and prioritizing device transmissions using distributed binary sequences [19]. 3GPP release 15 includes the design of early data transmission (EDT), which replaces a standard four-step RA procedure with a two-step procedure [20]. While this change reduces delay [21], a grant-free access mechanism is still of great interest [22].[1] After the connection setup, delay in the data transmission phase can be reduced for all devices, via scalable transmission time intervals (TTIs) [24], [25] and, for high-priority (HP) devices in particular, via preempting scheduled low-priority (LP) transmissions [26].

For wireless local-area network (WLAN)-based MTC, many proposed MAC protocols focus on the improvement of 802.11ah. In 802.11ah, the mechanism for reducing transmission collision probability under high device density is the restricted access window (RAW), which divides devices into groups and uses time division in the channel access for different groups. Related MAC design efforts have been focused on the window size of RAW as well as the device grouping, e.g., optimizing the window size based on the number of devices [27], adapting window size based on the number of transmission attempts in each group [28], and using traffic-aware device grouping based on an estimation of channel usage in the groups [29].

Beyond the preceding two categories of research works, which build on and improve existing solutions, more ambitious approaches have been investigated. For example, machine learning-based device-level traffic arrival forecast is adopted in [30], which simplifies MAC into proactive channel scheduling. Another example is the reconfigurable MAC proposed in [31] and [32] that dynamically adjusts the partition between contention-free and contention-based sections based on network traffic for maximizing network throughput. Machine learning-based solutions for MAC are also discussed in [33].

The existing studies provide important insights on MAC for MTC, such as the importance of coordinating and prioritizing devices and the necessity of a flexible design. However, for IIoT and, in particular, for applications such as factory automation and process control, further research on customized MAC protocols is necessary. In this paper, we focus on coordination-based grant-free MAC as a link-layer solution to support such applications in IIoT. Specifically, we aim to achieve the following objectives: 1) to support a high device density in a limited geographical area, e.g., a manufacturing facility, with a single access point (AP) and a single channel; 2) to provide differentiated QoS to different types of devices, while satisfying stringent QoS requirements of HP devices, e.g., millisecond or submillisecond level delay; 3) to minimize messaging and control overhead; and 4) to accommodate devices with simple and low-cost hardware, i.e., without requiring advanced physical-layer techniques. Achieving all the objectives simultaneously requires a protocol that fully exploits the potential of the MAC design. We present our proposed MAC solution in two parts, with details of a new MAC protocol in Part I and a customized scheduling scheme to complement the protocol in Part II [34]. Through the two parts, the integration of delicate distributed coordination and fine-grained centralized scheduling composes the unique strength of our MAC solution.

The contribution of Part I is twofold. First, we propose a novel MAC protocol for applications in IIoT such as factory automation and process control. The protocol design comprises a new time slot structure, corresponding channel access procedure, and two mechanisms for providing differentiated QoS and for supporting ultradense networks, respectively. Featuring delicate distributed coordination, the protocol can significantly improve channel utilization efficiency without packet collision or, in the case of high device density, support a large number of devices at the cost of low packet collision probabilities. Second, we provide thorough performance analysis for the proposed MAC protocol based on limited data traffic information. Specifically, we characterize the delay performance for the proposed channel access strategy as well as the impact of the two mechanisms. Without assuming a specific traffic arrival model, we establish our analysis based only on the packet arrival rates at the devices. The analytical results provide insight for scheduling and are later demonstrated to be accurate by simulations in Part II.

The remainder of Part I is organized as follows. Section II describes the networking scenario under consideration. Section III presents the proposed MAC protocol. In Section IV, we provide performance analysis for the proposed MAC design. Section V concludes Part I. Proofs of the theorems are given in the Appendix. A list of main symbols is given in Table I.

[1] 3GPP release 16 also includes 5G support for IIoT through time sensitive communications (TSCs) [23]. However, the focus of TSC is time synchronization instead of access control.

TABLE I
LIST OF MAIN SYMBOLS

$d_{m,l}$	the index of the device assigned mini-slot m of slot l
D	the number of all devices
$\mathcal{D}$	the set of all devices
$\mathcal{D}^{\mathrm{H}}/\mathcal{D}^{\mathrm{R}}/\mathcal{D}^{\mathrm{L}}$	the set of all HP/RP/LP devices
$\mathcal{D}_{m,l}$	the set of devices assigned mini-slot m of slot l
$\mathcal{D}_l$	the set of devices assigned slot l
n_{m}	the number of mini-slots in a slot
n_{s}	the number of slots in a frame
$r^{\mathrm{H}}/r^{\mathrm{R}}/r^{\mathrm{L}}$	the number of slots in a HP/RP/LP assignment cycle
T_{m}	the length of a mini-slot
T_{s}	the length of a slot
T_{x}	the length of a packet transmission duration
$\delta^{\mathrm{H}}/\delta^{\mathrm{R}}/\delta^{\mathrm{L}}$	the maximum tolerable delay of HP/RP/LP devices
λ_i	the packet arrival rate of device i
$\lambda_{m,l}$	the packet arrival rate of device $d_{m,l}$
$\lambda'_{m,l}$	the effective packet arrival rate of device $d_{m,l}$
$\rho^{\mathrm{H}}/\rho^{\mathrm{R}}/\rho^{\mathrm{L}}$	the maximum tolerable packet collision probability of HP/RP/LP devices
$\tau_{m,l}$	the AD-F for the device assigned mini-slot m of slot l
$\tau^{\mathrm{b}}_{m,l}$	the AD-F for the device assigned mini-slot m of slot l in the case with buffer

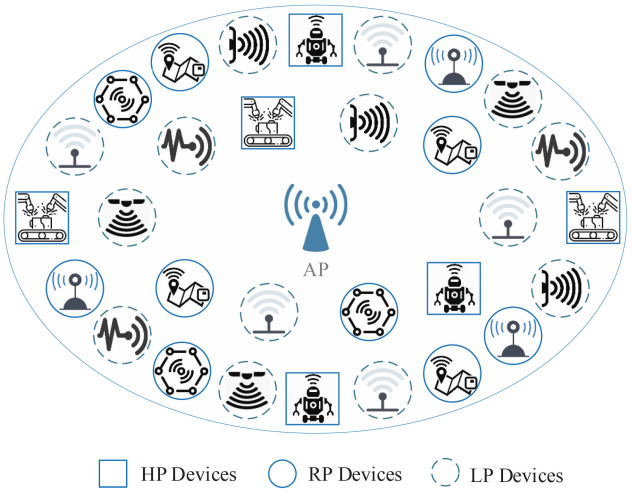


Fig. 1. Illustration of the networking scenario.

II. NETWORKING SCENARIO

Consider a fully connected network with one AP covering a limited geographical area, e.g., a manufacturing facility.[2] A large number of devices such as sensors, actors, and controllers are densely deployed in the area. The devices are categorized into three types, i.e., HP devices, regular-priority (RP) devices, and LP devices. An illustration of the considered scenario is given in Fig. 1.

The overall number of devices and the set of devices are denoted by D and $\mathcal{D}$, respectively. The number and set of HP, RP, and LP devices are denoted by D^{H} and $\mathcal{D}^{\mathrm{H}}$, D^{R} and $\mathcal{D}^{\mathrm{R}}$, and D^{L} and $\mathcal{D}^{\mathrm{L}}$, respectively. Without loss of generality, we assume that the devices are indexed such that the first till the D^{H}th devices are the HP devices, the next D^{R} devices are the RP devices, and the last D^{L} devices are the LP devices.

[2]The target area is assumed to be less than 1 km^2.

Communication Characteristics: The communication characteristics include the following.

1) Short data packets—the length of physical-layer packets is normally in the range between several bytes to several hundred bytes [35].
2) Uplink-dominated transmission—a significant portion of the data traffic is attributed to sensor readings or device status reports [36].

QoS Requirements: The considered QoS metrics are delay, from the instant of packet arrival to the instant of successful packet transmission, and packet transmission collision probability. Different types of devices have different QoS requirements. Specifically, the maximum tolerable delay and packet collision probability for HP, RP, and LP devices are denoted by δ^{H} and ρ^{H}, δ^{R} and ρ^{R}, and δ^{L} and ρ^{L}, respectively, where $\delta^{\mathrm{H}} < \delta^{\mathrm{R}} < \delta^{\mathrm{L}}$ and $\rho^{\mathrm{H}} < \rho^{\mathrm{R}} < \rho^{\mathrm{L}}$. The value of δ^{H} is assumed to be small such as on the millisecond level.

Device Packet Arrivals: For practicality, we do not assume a specific traffic model. However, we consider the following data packet arrival properties.

1) The packet arrival statistics at each device are constant during a relatively long period with respect to packet interarrival time. The packet arrival rate of device i in the considered time duration is denoted by λ_i.
2) The packet arrival rate is relatively low so that $1/\lambda_i$ is much larger than δ^{H} for any i. This is in accordance with the sporadic transmission characteristic of machine-type communications, where the packet interarrival time can range from tens of milliseconds to several minutes [8].
3) For tractability, we assume that the transmission time for data packets is identical and equal to T_{x}.

Given the networking scenario, we aim to develop an MAC solution with the following features.

1) Accommodating a large number of devices on a single channel with a single AP.
2) Satisfying the differentiated QoS requirements for each type of devices.
3) Keeping control overhead as low as possible.
4) Exploring the role of machine learning, specifically in device transmission scheduling.

In Part I, we focus on items 1)–3), while in Part II, our emphasis is placed on items 1), 2), and 4).

III. PROPOSED MAC PROTOCOL

The proposed MAC protocol is based on time-slotted channel access, which suits short packets. Tailored for the considered networking scenario, our protocol comprises the following elements.

1) Minislot-based carrier sensing (MsCS).
2) Synchronization carrier sensing (SyncCS).
3) Differentiated assignment cycles.
4) Superimposed minislot assignment (SMsA).

In the list, the first two elements target improving channel utilization efficiency through implicit distributed coordination, the third targets providing differentiated QoS for different device types, and the last targets increasing the number of supported devices.

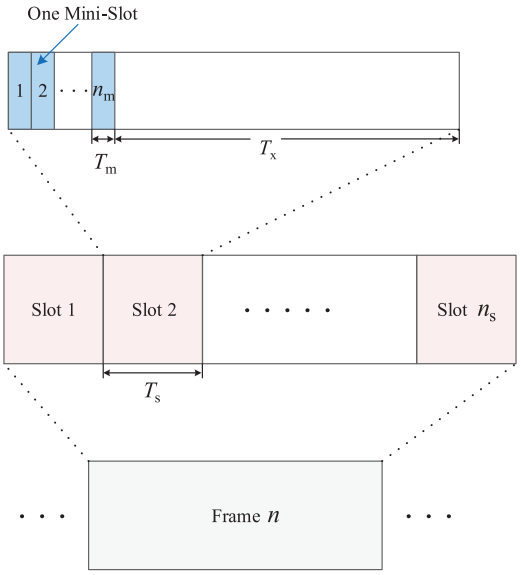


Fig. 2. Illustration of the frame, slot, and minislot structure.

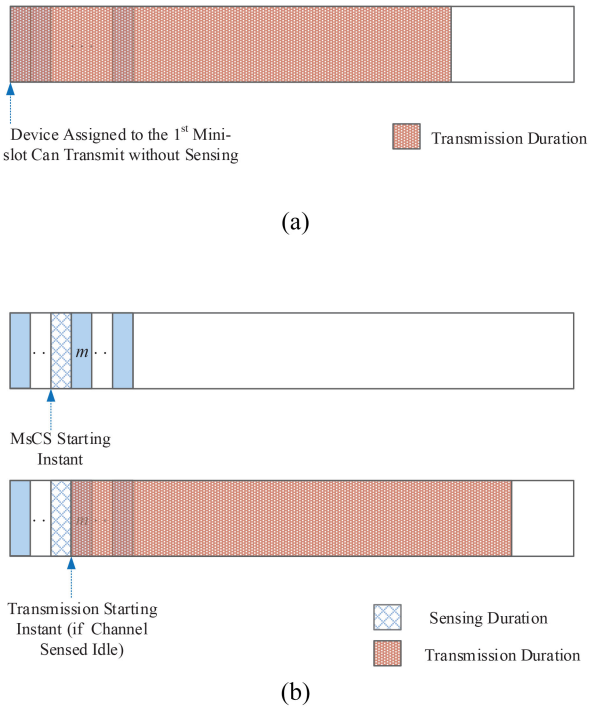


Fig. 3. Illustration of the MsCS. (a) Devices assigned to the first minislot of any slot starts transmission immediately when the slot begins, without sensing the channel. (b) Devices assigned to minislot $m(>1)$ must sense the channel during the $(m-1)$th minislot, and starts transmission at the beginning of the mth minislot if the channel is sensed to be idle.

A. Time Frame and Slot Structure

Time is partitioned into frames and each frame is partitioned into n_s slots, as shown in Fig. 2. A slot begins with n_m minislots, each of length T_m, followed by a duration of length T_x. Accordingly, the length of a slot, denoted by T_s, depends on the number of minislots and is equal to $n_m \times T_m + T_x$.

Given the high device density and sporadic transmission pattern, each slot is assigned to multiple devices, in order to achieve high channel utilization efficiency via reducing idle slots. Different devices associated with a slot are assigned different minislots of the slot. Different from the existing designs with minislots (e.g., [37], [38], or [39]), where each minislot is used for transmitting one or more packets, the minislots in our protocol are very short (e.g., less than 10 μs) and are used for channel sensing instead of sending reservation requests or data packets (as detailed in Section III-B). In the proposed protocol, the minimum time unit for transmitting a packet is a slot, and each slot accommodates at most one successful packet transmission. Clearly, without a proper coordination, transmission collision may happen when multiple devices are assigned to the same slot.

B. MsCS

The purpose of minislots is to enable channel sensing for collision-free distributed channel access. When the AP assigns a slot to a device, it also specifies a minislot for the device. Suppose that device i is assigned minislot m of slot l. Then, the following rules are used in the proposed protocol.

1) If device i has a packet to transmit and $m = 1$, it starts transmitting right away when slot l begins.
2) If device i has a packet to transmit and $m > 1$, it needs to sense the channel during minislot $m-1$ of slot l and starts transmitting from minislot m of slot l only if the channel is sensed idle; otherwise, it will skip this slot and wait for the next transmission opportunity.
3) If device i does not have a packet to send, it simply stays idle in the corresponding slot.

The first two cases are illustrated in Fig. 3.

With MsCS, different minislots correspond to different transmission priorities. Specifically, a minislot with a larger index corresponds to a lower transmission priority. Therefore, minislots with small indices can be used to accommodate HP devices. Through MsCS, before accessing the channel, a device makes sure that none of the devices with higher priority is using the channel. As a result, the devices can avoid packet collision while sharing the same slot. Note that the MsCS is fully distributed and does not require any control message exchange, given the assignment of slots and minislots to devices by the AP. The cost for avoiding collision is the overhead of using minislots for sensing. Specifically, the ratio of packet transmission duration over slot length is T_x/T_s.

For MsCS to work, the following conditions should be satisfied.

1) The minislot length T_m must be longer than the maximum propagation delay across the network coverage area.[3]
2) The overall length of all minislots, i.e., $n_m T_m$, should be less than the packet transmission duration

[3]A possible choice for minislot length is 9 μs, which follows from the distributed coordination function (DCF) slot time in IEEE 802.11ac.

T_x (for each slot to accommodate at most one transmission).[4]

3) The aggregated packet arrival rate of all devices assigned the same slot must be less than 1 per frame.

C. SyncCS

Even though MsCS improves channel utilization efficiency, as a result of multiple devices sharing each slot, none of the devices may have a packet to transmit in a slot. Increasing the number of minislots in each slot can reduce the slot idle probability. However, it may violate the delay requirements for devices assigned high-index minislots or the aforementioned condition on the overall length of minislots.

Alternatively, if idle slots can be identified and avoided, the channel utilization efficiency can be further improved, and so will the resulting QoS. To achieve this, the following rules of SyncCS are used in the proposed protocol.

1) All devices in $\mathcal{D}$ sense the channel in the last minislot, i.e., minislot n_m, of everyone slot. The exceptions are: a) any device that is transmitting and b) the device that is assigned minislot n_m, under the condition that it has a packet to transmit.[5]
2) If the last minislot is idle, the rest of the current slot is skipped and the next slot starts immediately after this last minislot.
3) If the last minislot is busy, the next slot starts after the current slot ends.

The above rules are illustrated in Fig. 4, and the rationale is explained as follows. Given the condition that $n_m T_m < T_x$ as mentioned in Section III-B, no device is or will be transmitting in a slot if the last minislot of that slot is idle. Therefore, upon sensing an idle last minislot, all devices know that the rest of the slot can be skipped and the next slot can start after this minislot. The SyncCS allows devices to synchronize slots even though the length of a slot is no longer fixed. With SyncCS, a busy slot has the full length of $n_m \times T_m + T_x$, while an idle slot has the reduced length of $n_m \times T_m$.

The SyncCS has two main differences from the MsCS.

1) In SyncCS, devices must perform sensing regardless of whether they have a packet to transmit or not (with exceptions as mentioned above).
2) In SyncCS, all devices, not just the devices assigned to the slot, need to sense the channel in each slot.

Similar to the MsCS, SyncCS is fully distributed and does not require any control message exchange. The cost for further improving channel utilization efficiency via SyncCS is the extra channel sensing. In addition, accurate time synchronization is required among all devices. Without SyncCS, a device can be in the sleep mode for most of the time in a frame and only wake up before its assigned minislot for MsCS if it has a packet to transmit. With SyncCS, each device needs to perform sensing in each slot and resynchronize once for each idle slot. In the IIoT scenario under consideration, it is possible that energy consumption of devices is less of a concern (e.g., as compared with sensors deployed in remote areas such as in forests); otherwise, the design element of SyncCS can be omitted in the proposed protocol.[6]

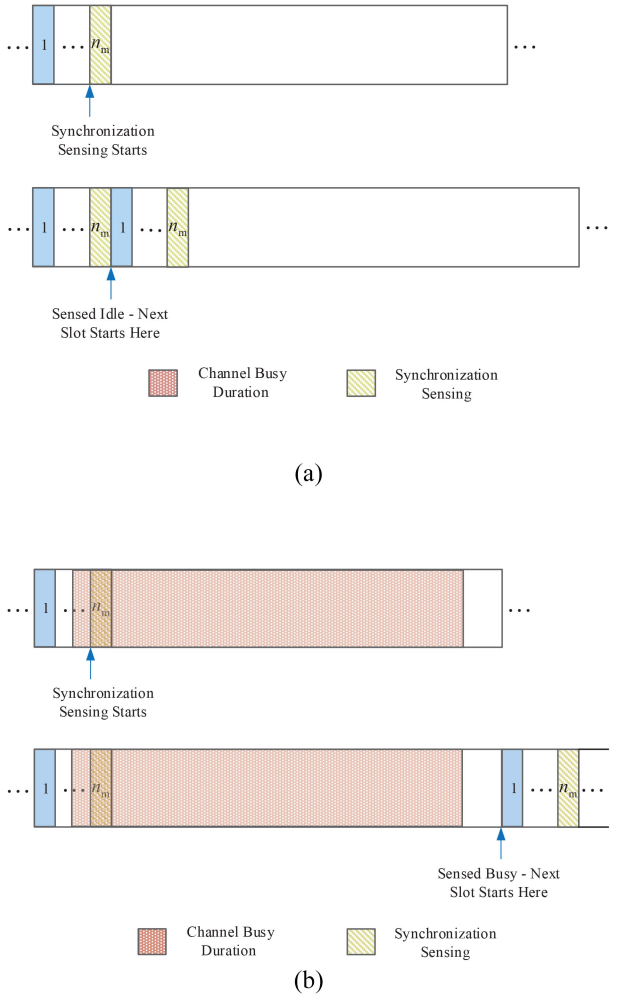


Fig. 4. Illustration of the syncCS. (a) When the last minislot of a slot is sensed idle, the remaining transmission duration of this slot is skipped, and the next slot starts right after the last minislot of this slot. (b) When the last minislot of a slot is sensed busy, the next slot starts after the entire duration of this slot.

D. Differentiated Assignment Cycles

Using the slot structure in Fig. 2, the delay for a device depends on the frame length if each device has at most one transmission opportunity in each frame. However, one transmission opportunity in each frame for every device does not provide sufficient flexibility to support differentiated QoS. Particularly, the maximum delay threshold of HP devices, i.e., δ^H, can be much smaller than that of RP/LP devices. To address this problem, we extend the frame in Fig. 2 to differentiated assignment cycles. Specifically, each HP, RP, and

[4]In an extreme case when a device assigned a low-index minislot transmits a very short packet, it is possible that a device assigned a high-index minislot senses channel idle and transmits a packet in the same slot. This extreme case is ignored in the protocol design and performance analysis.

[5]The device assigned minislot n_m knows whether the slot is idle or not from sensing the channel during minislot $n_m - 1$ as mandated by the MsCS.

[6]Alternatively, the AP may broadcast frame synchronization beacons. In such a case, when a device has a packet to send, it can wake up and synchronize to the next frame. It may remain awake and synchronized to each slot until the packet is transmitted.

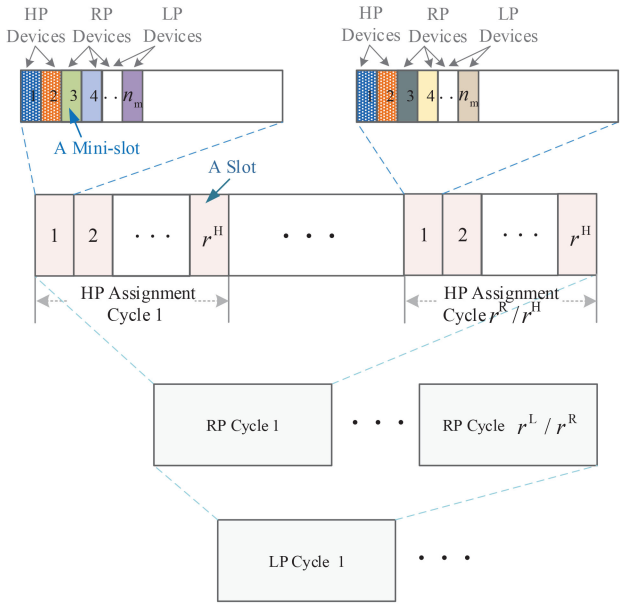


Fig. 5. Illustration of differentiated assignment cycles.

LP assignment cycle consists of r^H, r^R, and r^L slots, respectively, where $r^H < r^R < r^L$. Each HP, RP, or LP device is assigned one minislot of one slot in each HP, RP, or LP assignment cycle, respectively. Thus, an HP/RP/LP cycle serves as a frame for the HP/RP/LP devices, respectively. In the case when all devices have the same priority, the HP, RP, and LP cycles become identical and reduce to a standard frame. The differentiated assignment cycles are illustrated in Fig. 5, in which different color patterns in the minislots represent different assigned devices. In the illustration, r^L is a multiple of r^R, and r^R is a multiple of r^H.[7] The HP devices assigned to the same slot in any different HP assignment cycles are identical, as shown by the two illustrated slots on the top of Fig. 5, while the RP or LP devices assigned to the two slots are different.

With differentiated assignment cycles, it becomes possible to achieve the stringent delay requirement of HP devices, by setting r^H small, and at the same time support a large number of devices, by using a large r^R and/or r^L. Note that similar idea of differentiated cycles can be found in existing work such as [40], where two different cycle lengths are used for real-time and nonreal-time traffic, respectively. With a different slot structure and three different cycle lengths, we adopt the same essential idea here. This is because, for scheduling-based channel access, achieving lower delay translates to more frequently scheduled transmission opportunities. This naturally leads to differentiated cycles for different device or traffic types.

E. SMsA

The proposed MAC protocol aims to support a high device density. The MsCS and SyncCS contribute to the solution by improving channel utilization efficiency, along with differentiated assignment cycles with a large r^R and/or r^L. In addition, if devices can share a minislot, beyond only sharing a slot, the capacity of the network in terms of the number of supported devices can be significantly improved, at the cost of nonzero packet transmission collision probabilities.

The final element in our proposed protocol, i.e., SMsA, allows the assignment of one minislot to multiple devices, provided that packet transmissions associated with such an assignment can be properly scheduled as not to violate the QoS requirements of the devices. For simplicity in presentation, we limit the SMsA to devices of the same type, i.e., an HP device can share a minislot only with other HP devices. With SMsA, a minislot in Fig. 5 may no longer be assigned to a device exclusively.

Transmission collision may happen among devices sharing a minislot, and the collision probability depends on the following factors.

1) The device packet arrival rates.
2) The number of minislots and the minislot assignment.
3) The HP, RP, and LP assignment cycle lengths.

While the device packet arrival rates are not controllable, the collision probability may be reduced by properly determining the last two factors (as presented in Part II).

We do not consider collision resolution in this work. However, a design element for collision detection can be added in our proposed MAC protocol. The following is an example. If two or more devices assigned the same minislot simultaneously start sending packets to the AP, the AP will detect the collision. As soon as the AP detects the collision, it will start broadcasting a collision beacon that fills the rest of the current slot. On the device side, the sending devices will switch to the sensing mode to check for a collision beacon after transmitting their packets. If a beacon is sensed, the device knows that a packet collision happened during its transmission and may decide to retransmit the packet.

F. Downlink Control

The AP broadcasts the minislot and slot assignment to devices via downlink control messages. Based on the assumption of stationary traffic statistics in a relatively long duration,[8] the assignment does not need to be updated frequently. The AP may either broadcast the entire assignment in one downlink control message or breakdown the assignment information into multiple messages.

Consider an example of 10 minislots per slot (i.e., $n_m = 10$) and 200 slots per LP assignment cycle (i.e., $r^L = 200$). In such a case, 2 bytes is more than sufficient to represent the slot and minislot assignment for each device. For 1000 devices, the assignment message payload size is no more than 2 kB. For a slot length of 200 μs, an LP assignment cycle is about 40 ms in length. Even if the traffic statistics change as frequently as once in every 5 minutes, the 2-kB downlink assignment message is needed just once in every 7500 LP assignment cycles or, equivalently, 1.5×10^6 slots.

As downlink control messages are infrequent in comparison with the dominating uplink messages, we neglect the impact of downlink control messages while analyzing the performance of the proposed protocol.

[7]While r^L does not have to be a multiple of r^R or r^H in theory, the overall device assignment cycle is the lowest common multiple of r^H, r^R, and r^L. Limiting the lowest common multiple to be r^R itself can reduce the complexity of the device assignment by the AP.

[8]The stationary duration, if denoted by T_{st}, should satisfy $T_{st} \gg r^L T_s$.

To summarize, the core of the MAC protocol is how to coordinate transmissions from devices, while prioritizing and device grouping are two important aspects of coordination. There are various approaches for prioritizing, such as using both contention-based and contention-free access in an MAC protocol [41]. Similarly, there are many grouping approaches, such as limiting contention to devices generating packets around the same instant [42]. In our proposed MAC protocol, the utilization of minislots is inherently capable of both prioritizing and grouping. Meanwhile, the differentiated assignment cycles further strengthen the design's capability in prioritizing, while the SMsA further strengthens the capability in grouping.

IV. Performance Analysis

In this section, we present a performance analysis of the proposed MAC protocol, focusing on the MsCS, SyncCS, and SMsA. Note that the proposed MAC protocol works under the following conditions.

1) The expected number of packet arrivals summarized over all devices sharing a slot is less than 1 per frame.[9]
2) The average packet arrival interval of any device is larger than the maximum tolerable packet delay of that device.

In practice, some devices can have a high packet arrival rate that violates the above conditions. In such a case, more than one slot can be assigned to such a device in the corresponding assignment cycle so that the expected number of packet arrivals of the device per scheduled slot is less than 1. In the subsequent analysis, we simply assume that the number of packet arrivals for any device is less than 1 per its assignment cycle.

Without assuming a specific traffic model, we focus on the first-order statistic. The expected number of packet arrivals at device i in a frame is given by $\lambda_i T_\mathrm{f}$, where T_f denotes the length of a frame. Denote the set of all HP, RP, and LP devices assigned to slot l by $\mathcal{D}_l$. Denote the delay of device i, averaged over packet transmissions while the traffic is stationary, by τ_i. The aforementioned two conditions correspond to the following equations:

$$\sum_{i\in\mathcal{D}_l} \lambda_i T_\mathrm{f} \le 1 \quad \forall l \tag{1a}$$

$$\frac{1}{\lambda_i} \ge \tau_i \quad \forall i. \tag{1b}$$

A. Delay Performance With No Buffer

We investigate the impact of minislots in the case without SMsA, given the slot assignment and device packet transmission probabilities (estimated from the packet arrival rates). Starting from a simplified scenario, the analysis here is based on the following assumptions.

1) The condition in (1a) is satisfied.
2) A packet not in transmission is dropped when a new packet is generated. The scenario where devices have buffers is analyzed in Section IV-B.
3) All devices are of the same type and priority. Consequently, the three assignment cycles reduce to a unified frame with n_s slots.
4) The SyncCS is not adopted. The analysis of SyncCS is given in Section IV-D.

We focus on the delay analysis since the collision probability is 0 without SMsA. Let τ_0 denote the *base delay*, defined as the time duration from the packet arrival instant till the first assigned minislot. Under the aforementioned assumptions, the average base delay is equal to $n_\mathrm{s}T_\mathrm{s}/2$ for all devices, as each device has one assigned minislot in each frame. The *overall delay* is the base delay plus the access delay (AD), i.e., the duration from the first assigned minislot since the packet arrival till the end of the packet transmission. Since the average base delay is a constant here, we focus on finding the average AD.

Denote the device assigned the mth minislot of the lth slot by $d_{m,l}$. Denote by $\tau_{m,l}$ the average AD counted in frames (AD-F), i.e., the number of logical frames since device $d_{m,l}$'s packet arrival till device $d_{m,l}$'s packet transmission.[10] Different from a physical frame, a logical frame counted in the AD-F for device $d_{m,l}$ is the duration from slot l of one physical frame to slot l of the next physical frame. Therefore, a logical frame has the same length as a physical frame, but different starting and ending points for different devices. Accordingly, the arrival and transmission of a packet can happen within one logical frame, and the resulting AD-F is 1 in such a case.[11] Note that AD-F $\tau_{m,l}$ corresponds to a duration slightly longer than the AD defined in the preceding paragraph. This is because the AD ends when a packet transmission is completed, while the AD-F counts the entire frame in the delay, including the duration after device $d_{m,l}$'s packet transmission. Accordingly, the AD of device $d_{m,l}$ can be obtained from the AD-F by calculating $(\tau_{m,l}-1)\times T_\mathrm{f} + T_\mathrm{x}$, where the frame length T_f is equal to $n_\mathrm{s}T_\mathrm{s}$.

Since any device assigned the first minislot of any slot can transmit right away without sensing when the slot begins, we have

$$\tau_{1,l} = 1 \quad \forall l. \tag{2}$$

For devices assigned the subsequent minislots, the AD-F can be found using the following result.

Theorem 1: For any integer m such that $1 \le m \le n_\mathrm{m}-1$, the following relation between the AD-F of device $d_{m+1,l}$ and device $d_{m,l}$ holds:

$$\begin{aligned}\tau_{m+1,l} = {} & \frac{1}{1-\gamma_{m,l}-T_\mathrm{f}\lambda'_{m,l}} \\ & \times \left(-\frac{\big(1-\gamma_{m,l}\big)T_\mathrm{f}\lambda'_{m,l}}{2}\tau_{m,l}^2 \right. \\ & \left. + \big(1-\gamma_{m,l}+T_\mathrm{f}\lambda'_{m,l}\big)\tau_{m,l} - \frac{T_\mathrm{f}\lambda'_{m,l}\big(1+\gamma_{m,l}\big)}{2}\right)\end{aligned} \tag{3}$$

[9]This condition applies to the case without differentiated assignment cycles. In the case with differentiated assignment cycles, the condition is different.

[10]When packet retransmission is considered, the definition of AD-F should be changed to "the number of logical frames since packet arrival till successful packet transmission." Meanwhile, the "packet arrival rate" in our analysis should be replaced by "packet transmission rate" as a packet may need retransmission(s).

[11]In the rest of this article, we do not distinguish physical and logical frames and refer to both as "frame" since they are equal in length.

where

$$\lambda'_{m,l} = \frac{\lambda_{m,l}}{\left(1 + T_{\mathrm{f}}\lambda_{m,l}\left(\tau_{m,l} - 1/2\right)\right)} \tag{4}$$

represents the effective packet arrival rate of device $d_{m,l}$ excluding dropped packets due to packet replacement (as there is no buffer), and

$$\gamma_{m,l} = T_{\mathrm{f}} \sum_{r=1}^{m} \lambda'_{r,l} \tag{5}$$

represents the expected overall number of packet arrivals in a frame for devices $d_{1,l}$ till $d_{m,l}$ (excluding replaced packets).

Using the fact that $\tau_{1,l}$ is equal to 1 for any l, (3) can be used to obtain the AD-F for devices assigned to all subsequent minislots in a slot recursively.

B. Delay Performance With Buffer

Now, consider the case when each device has a buffer. Recall that different minislots correspond to different transmission priorities. In the proposed protocol, any proper slot and minislot assignment ensures that the expected number of packets in the buffer of device $d_{m,l}$ is less than 1, for any $m < n_{\mathrm{m}}$ and any l. The reason is that if the expected number of buffered packets at $d_{m,l}$ is larger than or equal to 1, devices assigned minislots $m+1, \ldots, n_{\mathrm{m}}$ of slot l have almost no opportunity to transmit. As a result, we neglect the case when there are more than one packet in a buffer and use the following approximation. Specifically, at any instant, a device is in one of the three states.

1) No packet.
2) One packet, transmitting or waiting for channel access.
3) Two packets, one transmitting or waiting for channel access and the other arriving and going into the buffer.

Accordingly, for any given device, there is either no packet or one packet transmitting or waiting for channel access when a new packet arrives.

Denote by $\tau^{\mathrm{b}}_{m,l}$ the average AD-F of device $d_{m,l}$ in the case with buffer, the following result is in order.

Theorem 2: In the case with buffers, for any integer m such that $1 \le m \le n_{\mathrm{m}} - 1$, the relation between the AD-F of device $d_{m+1,l}$ and device $d_{m,l}$ is given by

$$\tau^{\mathrm{b}}_{m+1,l} = \frac{1-\gamma^{\mathrm{b}}_{m,l}}{1-\gamma^{\mathrm{b}}_{m+1,l}} \left(\frac{1}{1-\gamma^{\mathrm{b}}_{m,l} - T_{\mathrm{f}}\lambda_{m,l}} \times \left(-\frac{\left(1-\gamma^{\mathrm{b}}_{m,l}\right)T_{\mathrm{f}}\lambda_{m,l}}{2} \cdot \left(\tau^{\mathrm{b}}_{m,l}\right)^2 + \left(1-\gamma^{\mathrm{b}}_{m,l} + T_{\mathrm{f}}\lambda_{m,l}\right)\tau^{\mathrm{b}}_{m,l} - \frac{T_{\mathrm{f}}\lambda_{m,l}\left(1+\gamma^{\mathrm{b}}_{m,l}\right)}{2} \right) - 1 \right) + 1 \tag{6}$$

where

$$\gamma^{\mathrm{b}}_{m,l} = T_{\mathrm{f}} \sum_{r=1}^{m} \lambda_{r,l} \tag{7}$$

represents the expected overall number of packet arrivals in a frame for devices $d_{1,l}$ till $d_{m,l}$.

C. Slot Idle Probability

A slot is idle if none of its associated devices transmits. Under stationary packet arrival statistics, the expected slot idle probability of MsCS can be obtained. In the cases with and without buffer, the slot idle probability is approximately given by

$$\eta^{\mathrm{b}}_{l} = 1 - \sum_{m=1}^{n_{\mathrm{m}}} \lambda_{m,l} T_{\mathrm{f}} \tag{8a}$$

$$\eta_{l} = 1 - \sum_{m=1}^{n_{\mathrm{m}}} \lambda'_{m,l} T_{\mathrm{f}} \tag{8b}$$

respectively, where $\lambda'_{m,l}$ is given in (4). Note that the right-hand side of either of the two equations above is nonnegative when condition (1a) is satisfied, i.e., when the slot is not overloaded. The above approximation of the slot idle probability also assumes a negligible packet collision probability, which means that the expected number of transmitted packets and the expected number of packet arrivals (that cause no packet replacement) are equal in any slot.

Define the throughput of slot l as the expected number of packets transmitted in slot l. The slot throughput equals $1-\eta^{\mathrm{b}}_{l}$ and $1-\eta_{l}$ for the cases with and without buffers.

D. Impact of SyncCS

As SyncCS results in two possible lengths for each slot, i.e., the full and the reduced lengths, the frame length becomes a random variable. Denote the expected frame length with SyncCS in the case with and without buffer by $T^{\mathrm{e,b}}_{\mathrm{f}}$ and $T^{\mathrm{e}}_{\mathrm{f}}$, respectively. Denote n'_{s} as the number of busy slots out of the n_{s} slots in a frame. In the case without buffer, it follows that:

$$T^{\mathrm{e}}_{\mathrm{f}} = n_{\mathrm{s}} n_{\mathrm{m}} T_{\mathrm{m}} + n'_{\mathrm{s}} T_{\mathrm{x}}. \tag{9}$$

Since there is no collision

$$T^{\mathrm{e}}_{\mathrm{f}} \sum_{l} \sum_{m} \lambda'_{m,l} = n'_{\mathrm{s}} \tag{10}$$

because the expected number of packet transmissions should equal the expected number of arriving packets (that are not replaced) in a frame duration. From (4), (9), and (10) (with T_{f} replaced by $T^{\mathrm{e}}_{\mathrm{f}}$), n'_{s} and $T^{\mathrm{e}}_{\mathrm{f}}$ can be solved.

In the case with buffers, we have

$$T^{\mathrm{e,b}}_{\mathrm{f}} = n_{\mathrm{s}} n_{\mathrm{m}} T_{\mathrm{m}} + n'_{\mathrm{s}} T_{\mathrm{x}} \tag{11a}$$

$$T^{\mathrm{e,b}}_{\mathrm{f}} \sum_{l} \sum_{m} \lambda_{m,l} = n'_{\mathrm{s}} \tag{11b}$$

which gives

$$T^{\mathrm{e,b}}_{\mathrm{f}} = \frac{n_{\mathrm{s}} n_{\mathrm{m}} T_{\mathrm{m}}}{1 - \sum_{l}\sum_{m} \lambda_{m,l} T_{\mathrm{x}}}. \tag{12}$$

Substituting T_{f} in (3) and (6) with $T^{\mathrm{e}}_{\mathrm{f}}$ and $T^{\mathrm{e,b}}_{\mathrm{f}}$, respectively, gives the AD-F of the proposed design with MsCS and SyncCS. In the case without buffer, $T^{\mathrm{e}}_{\mathrm{f}}$ depends on $\tau_{m,l}$ through (4), which renders a complicated relation.

E. Impact of SMsA

The AD-F in Sections IV-A and IV-B is obtained when each minislot is assigned to a device exclusively. With SMsA, we have the following questions.

1) What is the relation among the AD-F of different devices assigned the same minislot?
2) How does the SMsA impact the relation in the AD-F between devices assigned adjacent minislots?

Denote the set of all devices assigned minislot m of slot l by $\mathcal{D}_{m,l}$. The following theorem answers the first question.

Theorem 3: In the case without buffer, all devices in $\mathcal{D}_{m,l}$ have the same AD-F, regardless of the difference in their individual packet arrival rates. In the case with buffer, assuming negligible packet collision probability and

$$\lambda_i \ll \sum_{r=1}^{m} \sum_{j \in \mathcal{D}_{r,l}} \lambda_j \quad \forall i \in \mathcal{D}_{m,l} \tag{13}$$

the differences among the AD-Fs of devices in $\mathcal{D}_{m,l}$ are negligible.

For the second question, similar to (3) and (6), the relation between the AD-Fs of devices in adjacent minislots in the case of SMsA can be characterized. For brevity, the characterization is given in the proof of Theorem 3 in Appendix C. It is worth mentioning that the packet collision probability has an impact on the AD-F even if devices do not detect collisions or retransmit. Given the aggregated packet arrival rate of devices sharing a minislot, a higher collision probability implies less channel busy duration for transmitting the same amount of packets. Consequently, the average packet waiting time and the AD-F reduce as the collision probability increases. However, if the collision probability is low, such an impact can be negligible.

With the AD-F, we can estimate the packet collision probability. Consider the case with buffer as an example and assume that the condition in (13) is satisfied. Based on Theorem 3, all the devices in $\mathcal{D}_{m,l}$ have the same AD-F, denoted by $\tau_{m,l}^{\mathrm{s,b}}$. Then, any device in $\mathcal{D}_{m,l}$ with a packet to send is expected to have one transmission opportunity in every $\tau_{m,l}^{\mathrm{s,b}}$ frames. The expected number of packet arrivals at device $i \in \mathcal{D}_{m,l}$ between any two consecutive transmission opportunities, which must be less than 1, can be estimated by $\tau_{m,l}^{\mathrm{s,b}} T_{\mathrm{f}} \lambda_i$. With the MsCS, all devices in $\mathcal{D}_{m,l}$, which have packets to send, share the same transmission opportunities. Therefore, the probability that device i's packet encounters a collision is approximately given by

$$q_i^{\mathrm{c,b}} = 1 - \prod_{j \in \mathcal{D}_{m,l} \setminus \{i\}} \left(1 - \tau_{m,l}^{\mathrm{s,b}} T_{\mathrm{f}} \lambda_i\right). \tag{14}$$

Note that knowing only the average packet arrival rates, the above approximation may be limited in accuracy. An accurate determination of the collision probability requires the traffic arrival model of all devices, which can be difficult to obtain in practice. We demonstrate through numerical results that our approximation can be a useful tool for device assignment in the case with SMsA in Part II.

V. Conclusion

In Part I of this paper, we tailor an MAC protocol for MTC in IIoT. To increase channel utilization efficiency, we propose MsCS and SyncCS, both of which feature distributed coordination. To prioritize devices and guarantee the QoS requirement of HP devices, we adopt differentiated assignment cycles for different types of devices. To further increase the supported number of devices, we develop the idea of SMsA, which can multiply the network capacity with a delay-collision tradeoff. Thanks to the strategies, the overall protocol has the potential to simultaneously achieve the targets of improving channel usage, minimizing messaging overhead, satisfying stringent QoS constraints, and providing differentiated performance. Meanwhile, customized and effective packet transmission scheduling that complements the protocol is necessary for achieving the full potential of the proposed protocol and is studied in Part II, which also presents numerical results to evaluate performance of the proposed MAC protocol.

Appendix A
Proof of Theorem 1

We first prove the effective packet arrival rate in (4). A packet is subject to replacement in the duration from its arrival till the beginning of its transmission. Given the average AD-F $\tau_{m,l}$, the average of the aforementioned duration is $\tau_{m,l} - 1 + 1/2$ frames, where -1 excludes the frame of transmission while $1/2$ adds half frame due to the average base delay. Given packet arrival rate $\lambda'_{m,l}$ and neglecting the correlation between packet arrivals, the probability that a new packet arrives in this duration can be estimated by $(\tau_{m,l} - 1/2)/(1/(n_{\mathrm{s}} T_{\mathrm{s}} \lambda'_{m,l}))$, where $1/(n_{\mathrm{s}} T_{\mathrm{s}} \lambda'_{m,l})$ is the average number of frames per packet arrival. Therefore, the following equation regarding packet arrival and replacement holds:

$$\lambda'_{m,l} = \lambda_{m,l}\left(1 - n_{\mathrm{s}} T_{\mathrm{s}} \lambda'_{m,l}\left(\tau_{m,l} - 1/2\right)\right) \tag{15}$$

which gives (4).

Suppose that device $d_{m,l}$ has a packet ready to transmit at the beginning of slot l. We have the following observations.

1) Device $d_{m+1,l}$ has the same AD-F as device $d_{m,l}$ if: a) device $d_{m,l}$ is removed from minislot m and b) the packet arrival probability of $d_{m+1,l}$ is the same as the packet arrival probability of device $d_{m,l}$.
2) Consider independent packet arrivals at devices $d_{m,l}$ and $d_{m+1,l}$. Even if the packet arrival rates of devices $d_{m,l}$ and $d_{m+1,l}$ are different, the following two probabilities are the same, given any realizations of the packet arrival processes of devices $d_{1,l}$ to $d_{m-1,l}$ and for any arbitrary $x \geq 0$. The first is the probability that the AD-F of a packet of device $d_{m,l}$ is x, while the second is the probability that the AD-F of a packet of device $d_{m+1,l}$ is x if device $d_{m,l}$ is removed from the slot.
3) If a packet of device $d_{m+1,l}$ arrives when a packet of device $d_{m,l}$ is waiting for channel access, then the packet of device $d_{m+1,l}$ needs to wait till at least one frame after the packet of device $d_{m,l}$ is sent. During the waiting in such a case, it is possible that there are

new packet arrivals at devices $d_{0,l}, \ldots, d_{m,l}$, which can trigger further waiting for the packet of device $d_{m+1,l}$.

Accordingly, for devices assigned to minislot m ($1 \le m \le n_{\mathrm{m}} - 1$), we can obtain the following relation between $\tau_{m+1,l}$ and $\tau_{m,l}$:

$$\begin{aligned}\tau_{m+1,l} = {} & \left(1 - \alpha_{m,l} - n_{\mathrm{s}} T_{\mathrm{s}} \lambda'_{m,l} - \beta_{m,l}\right)\tau_{m,l} \\ & + \alpha_{m,l}\left(\left(\frac{\tau_{m,l}-1}{2} + 1\right)\frac{1}{1-\gamma_{m,l}} + 1\right) \\ & + n_{\mathrm{s}} T_{\mathrm{s}} \lambda'_{m,l}\left(1 + \frac{1}{1-\gamma_{m,l}}\right) \\ & + \beta_{m,l}\left(\tau_{m,l} + \frac{1}{1-\gamma_{m,l}}\right) \end{aligned} \tag{16}$$

where

$$\alpha_{m,l} = \frac{\tau_{m,l}-1}{1/\left(n_{\mathrm{s}} T_{\mathrm{s}} \lambda'_{m,l}\right)} = \left(\tau_{m,l}-1\right) n_{\mathrm{s}} T_{\mathrm{s}} \lambda'_{m,l} \tag{17}$$

represents the probability that a packet of device $d_{m+1,l}$ arrives while device $d_{m,l}$ has a packet waiting for channel access but not transmitting. The probability is approximated by the ratio of the average number of frames for device $d_{m+1,l}$'s packet to wait to the average number of frames between two packet arrivals of $d_{m+1,l}$. The parameter $\beta_{m,l}$ in (16) is given by

$$\beta_{m,l} = \frac{\tau_{m+1,l}-1}{1/\left(n_{\mathrm{s}} T_{\mathrm{s}} \lambda'_{m,l}\right)} \tag{18}$$

and it represents the probability that a packet of device $d_{m+1,l}$ arrives when device $d_{m,l}$ has no packet yet, but device $d_{m,l}$ will have a packet arrival and transmit that packet before device $d_{m+1,l}$ transmits its packet.

The four terms on the right-hand side of (16) correspond to the following four cases.

1) With probability $1 - \alpha_{m,l} - n_{\mathrm{s}} T_{\mathrm{s}} \lambda_{m,l} - \beta_{m,l}$, device $d_{m,l}$ has no packet waiting for channel access between the arrival and transmission of a packet of device $d_{m+1,l}$. In such a case, the expected AD-F of this packet of device $d_{m+1,l}$ is the same as the expected AD-F of a packet of device $d_{m,l}$.
2) With probability $\alpha_{m,l}$, a packet of device $d_{m+1,l}$ arrives when a packet of device $d_{m,l}$ is waiting for channel access (but not transmitting). In such a case, the AD-F for the packet of $d_{m+1,l}$ is equal to the number of frames device $d_{m,l}$ needs to wait for from now on, added by one more frame plus the average number of packet arrivals at the devices from $d_{0,l}$ till $d_{m,l}$.
3) With probability $n_{\mathrm{s}} T_{\mathrm{s}} \lambda'_{m,l}$, a packet of device $d_{m+1,l}$ arrives when device $d_{m,l}$ is transmitting. In such a case, there was no packet waiting for channel access at devices assigned to precedent minislots when the packet transmission started. Therefore, the expected AD-F is equal to the expected number of packet arrivals of devices $d_{0,l}$ to $d_{m,l}$ in this frame plus the expected number of packet arrivals during the transmission of these packets plus one frame of transmission time.
4) With probability $\beta_{m,l}$, a packet of device $d_{m+1,l}$ arrives when $d_{m,l}$ has no packet waiting or transmitting, and then a packet of device $d_{m,l}$ arrives while the packet of device $d_{m+1,l}$ is waiting for channel access. In such a case, compared to the expected AD-F of a new packet at device $d_{m+1,l}$ with no packet at $d_{m,l}$, which is equal to $\tau_{m,l}$ based on the argument in the first case above, the waiting time of $d_{m+1,l}$ is increased by 1 (for device $d_{m,l}$ to transmit its packet) plus the average number of packet arrivals at devices $d_{0,l}$ to $d_{m,l}$.

Substituting (17) and (18) into (16), $\tau_{m,l}$ can be found as in (3).

Appendix B
Proof of Theorem 2

Consider the conditional probability that one packet is waiting for channel access given that a new packet arrives at device $d_{m,l}$ in the scenario with buffers. When the packet arrival is independent on the packet transmission, the above probability is the same as the probability that one packet is transmitting or waiting for channel access at device $d_{m,l}$. Denote this probability by $p^{\mathrm{b}}_{m,l}$. Then, for minislot 1, the AD-F, i.e., $\tau^{\mathrm{b}}_{1,l}$, is 1 with probability $1 - p^{\mathrm{b}}_{1,l}$. With probability $p^{\mathrm{b}}_{1,l}$, a new packet of $d_{1,l}$ arrives in the frame in which the existing packet of $d_{1,l}$ is or will be transmitting. The average AD-F $\tau^{\mathrm{b}}_{1,l}$ in the latter case is 1.5 frames, with 1 transmission frame and an average of 0.5 waiting frame.

The overall average AD-F is given by

$$\tau^{\mathrm{b}}_{1,l} = \left(1 - p^{\mathrm{b}}_{1,l}\right) + 1.5 p^{\mathrm{b}}_{1,l} = 1 + 0.5 p^{\mathrm{b}}_{1,l}. \tag{19}$$

Under the aforementioned approximation of at most 1 packet in buffer, the probability $p^{\mathrm{b}}_{m,l}$ can be estimated as

$$p^{\mathrm{b}}_{1,l} = \left(\tau^{\mathrm{b}}_{1,l} - 0.5\right)\lambda_{1,l} T_{\mathrm{f}} \tag{20}$$

where -0.5 corresponds to deducting the frame of transmission and adding the 0.5 frame of base delay. From the above two equations, the average AD-F can be derived as

$$\tau^{\mathrm{b}}_{1,l} = 1 + \frac{\lambda_{1,l} T_{\mathrm{f}}}{2\left(2 - \lambda_{1,l} T_{\mathrm{f}}\right)}. \tag{21}$$

For subsequent minislots, the average AD-F when a packet arrives with no existing packet at the same device, denoted by $\hat{\tau}^{\mathrm{b}}_{m,l}$, can be found using the same approach as in the case with no buffer. Meanwhile, with probability $p^{\mathrm{b}}_{m,l}$, a packet arrives when another packet is waiting for channel access, which will experience the average AD-F given by

$$\tilde{\tau}^{\mathrm{b}}_{m,l} = \hat{\tau}^{\mathrm{b}}_{m,l} + \frac{1}{1 - \gamma^{\mathrm{b}}_{m,l}}. \tag{22}$$

Thus, the overall average AD-F is

$$\begin{aligned}\tau^{\mathrm{b}}_{m,l} &= p^{\mathrm{b}}_{m,l}\left(\hat{\tau}^{\mathrm{b}}_{m,l} + \frac{1}{1-\gamma^{\mathrm{b}}_{m,l}}\right) + \left(1 - p^{\mathrm{b}}_{m,l}\right)\hat{\tau}^{\mathrm{b}}_{m,l} \\ &= \hat{\tau}^{\mathrm{b}}_{m,l} + p^{\mathrm{b}}_{m,l}\frac{1}{1-\gamma^{\mathrm{b}}_{m,l}}. \end{aligned} \tag{23}$$

The probability $p^{\mathrm{b}}_{m,l}$ can be found as

$$p^{\mathrm{b}}_{m,l} = \frac{\tau^{\mathrm{b}}_{m,l} - 1}{1/\left(\lambda_{m,l} T_{\mathrm{f}}\right)}. \tag{24}$$

Similarly, from (23) and (24), it can be found that

$$\tau^{\mathrm{b}}_{m,l} = \frac{\hat{\tau}^{\mathrm{b}}_{m,l} - 1}{1 - \lambda_{m,l} T_{\mathrm{f}} \frac{1}{1-\gamma^{\mathrm{b}}_{m,l}}} + 1 = \frac{1 - \gamma^{\mathrm{b}}_{m,l}}{1 - \gamma^{\mathrm{b}}_{m+1,l}} \left(\hat{\tau}^{\mathrm{b}}_{m,l} - 1 \right) + 1. \tag{25}$$

Using (3), but with the effective packet arrival rate $\lambda'_{m,l}$ replaced by $\lambda_{m,l}$ (since there is no packet replacement in the case with buffer), it follows that:

$$\hat{\tau}^{\mathrm{b}}_{m+1,l} = \frac{1}{1 - \gamma^{\mathrm{b}}_{m,l} - T_{\mathrm{f}} \lambda_{m,l}} \left(-\frac{1}{2} \left(1 - \gamma^{\mathrm{b}}_{m,l} \right) T_{\mathrm{f}} \lambda_{m,l} \tau^2_{m,l} + \left(1 - \gamma^{\mathrm{b}}_{m,l} + T_{\mathrm{f}} \lambda_{m,l} \right) \tau_{m,l} - \frac{1}{2} T_{\mathrm{f}} \lambda_{m,l} \left(1 + \gamma^{\mathrm{b}}_{m,l} \right) \right). \tag{26}$$

Using (25) and (26), the average AD-F for minislot $m \geq 1$ in (6) can be obtained.

APPENDIX C
PROOF OF THEOREM 3

A. Case With No Buffer

Denote by q^{c}_i the conditional packet collision probability of device i given that device i is transmitting. Note that as each minislot is assigned to multiple devices, we cannot use $d_{m,l}$ to represent the device assigned to minislot m of slot l. Because of a nonzero collision probability, the expected number of transmitting packets given that device i is transmitting, denoted by n^{c}_i, can be larger than 1.

Let $\Lambda'_{m,l}$ denote the aggregated/cumulative effective packet arrival rate of all devices assigned to minislot m of slot l, given by

$$\Lambda'_{m,l} = \sum_{i \in \mathcal{D}_{m,l}} \lambda'_i \left(\left(1 - q^{\mathrm{c}}_i \right) \cdot 1 + q^{\mathrm{c}}_i \cdot \frac{n^{\mathrm{c}}_i - 1}{n^{\mathrm{c}}_i} \right) = \sum_{i \in \mathcal{D}_{m,l}} \lambda'_i \left(1 - \frac{q^{\mathrm{c}}_i}{n^{\mathrm{c}}_i} \right) \tag{27}$$

where λ'_i is the effective arrival rate in (4). The aggregated/cumulative effective arrival rate $\Lambda'_{m,l}$ is defined from the perspective of channel usage and, thus, colliding packets are treated as one packet because a collision occupies the channel for one packet transmission duration. The impact of packet replacement and packet collision is considered in (27). Note that (27) is an approximation based only on the average packet arrival rates for the case of low collision probability and uncorrelated packet arrival processes at the devices. While q^{c}_i and n^{c}_i are conditional (on a packet transmission of device i), λ'_i implicitly indicates the packet transmission rate of device i (as each packet taken into account by the effective arrival rate will be transmitted). As a result, $\Lambda'_{m,l}$ is unconditional.

Denote the average cumulative effective packet arrival rate for all devices assigned to the first m minislots by

$$\Gamma_{m,l} = T_{\mathrm{f}} \sum_{r=1}^{m} \Lambda'_{m,l}. \tag{28}$$

Given $\Lambda'_{m,l}$ $\forall m, l$, the average AD-F in the case of SMsA can be found by extending the result in (3) as follows:

$$\tau^{\mathrm{s}}_{m+1,l} = \frac{1}{1 - \Gamma_{m,l} - T_{\mathrm{f}} \Lambda'_{m,l}} \left(-\frac{\left(1 - \Gamma_{m,l} \right) T_{\mathrm{f}} \Lambda'_{m,l}}{2} \left(\tau^{\mathrm{s}}_{m,l} \right)^2 + \left(1 - \Gamma_{m,l} + T_{\mathrm{f}} \Lambda'_{m,l} \right) \tau^{\mathrm{s}}_{m,l} - \frac{1}{2} T_{\mathrm{f}} \Lambda'_{m,l} \left(1 + \Gamma_{m,l} \right) \right). \tag{29}$$

With $\tau^{\mathrm{s}}_{m,l}$, the packet collision probability for device i in $\mathcal{D}_{m,l}$ can be approximated by

$$q^{\mathrm{c}}_i = 1 - \prod_{j \in \mathcal{D}_{m,l} \setminus \{i\}} \left(1 - \tau^{\mathrm{s}}_{m,l} T_{\mathrm{f}} \lambda_i \right). \tag{30}$$

Meanwhile, n^{c}_i can be estimated by

$$n^{\mathrm{c}}_i = 1 + \sum_{j \in \mathcal{D}_{m,l} \setminus \{i\}} \tau^{\mathrm{s}}_{m,l} T_{\mathrm{f}} \lambda_j \tag{31}$$

where constant 1 corresponds to the given fact that a packet of device i is involved in the collision.

Using (27)–(31) and the fact that $\tau^{\mathrm{s}}_{1,l} = 1$ $\forall l$, the average AD-F and packet collision probability for all devices assigned to all minislots can be derived. The result in (29) suggests that although the packet arrival rates for devices assigned to one minislot can be different, the average AD-Fs for the devices are the same. This is not unexpected if we compare (3) and (29). From the comparison, it can be seen that the impact of the individual packet arrival rate on the average AD-F is replaced by the impact of the aggregated packet arrival rate of the minislot in the case of SMsA. Thus, while individual devices may have different packet arrival rates, their average AD-Fs become identical as all devices assigned to the same minislot share the same aggregated packet arrival rate.

B. Case With Buffers

Denote by $q^{\mathrm{c,b}}_i$ the conditional packet collision probability of device i given that device i is transmitting. Because of a nonzero collision probability, the expected number of transmitting packets given that device i is transmitting, denoted by $n^{\mathrm{c,b}}_i$, can be larger than 1.

Let $\Lambda^{\mathrm{b}}_{m,l}$ denote the aggregated effective packet arrival rate of all devices assigned to minislot m of slot l, given by

$$\Lambda^{\mathrm{b}}_{m,l} = \sum_{i \in \mathcal{D}_{m,l}} \lambda_i \left(1 - \frac{q^{\mathrm{c,b}}_i}{n^{\mathrm{c,b}}_i} \right). \tag{32}$$

Denote the cumulative effective packet arrival rate for all devices assigned to the first m minislots by

$$\Gamma^{\mathrm{b}}_{m,l} = T_{\mathrm{f}} \sum_{r=1}^{m} \Lambda^{\mathrm{b}}_{m,l}. \tag{33}$$

Following the analysis leading to the results in (3), (6), and (29), the average AD-F for device i assigned to the mth

minislot of slot l can be obtained:

$$\tau^{\mathrm{s,b}}_{i,m+1,l} = \frac{1-\Gamma^{\mathrm{b}}_{m,l}}{1-\Gamma^{\mathrm{b}}_{m,l}-T_{\mathrm{f}}\lambda_i} \times \left(\frac{1}{1-\Gamma^{\mathrm{b}}_{m,l}-T_{\mathrm{f}}\Lambda^{\mathrm{b}}_{m,l}} \times \left(-\frac{1}{2}\left(1-\Gamma^{\mathrm{b}}_{m,l}\right)\cdot T_{\mathrm{f}}\Lambda^{\mathrm{b}}_{m,l}\left(\bar{\tau}^{\mathrm{s,b}}_{m,l}\right)^2 + \left(1-\Gamma^{\mathrm{b}}_{m,l}+T_{\mathrm{f}}\Lambda^{\mathrm{b}}_{m,l}\right)\bar{\tau}^{\mathrm{s,b}}_{m,l} - \frac{T_{\mathrm{f}}\Lambda^{\mathrm{b}}_{m,l}\left(1+\Gamma^{\mathrm{b}}_{m,l}\right)}{2} \right) - 1 \right) + 1 \quad (34)$$

where

$$\bar{\tau}^{\mathrm{s,b}}_{m,l} = \frac{1}{\left|\mathcal{D}_{m,l}\right|}\sum_{i\in\mathcal{D}_{m,l}} \tau^{\mathrm{s,b}}_{m,l} \quad (35)$$

represents the average AD-F of all devices assigned the mth minislot of slot l. From (34), it can be seen that, different from the case in (29), where the average AD-F is identical for all devices assigned to the same minislot, the average AD-F can be different for different devices here due to $-T_{\mathrm{f}}\lambda_i$ in the denominator of the first term. However, when the collision probability is low and the condition in (13) is satisfied, the difference by $T_{\mathrm{f}}\lambda_i$ is negligible in comparison with $\Gamma^{\mathrm{b}}_{m,l}$.

With $\tau^{\mathrm{s,b}}_{m,l}$, $q^{\mathrm{c,b}}_i$ and $n^{\mathrm{c,b}}_i$ can be estimated similarly as in the case without buffer.

REFERENCES

[1] X. Shen *et al.*, "AI-assisted network-slicing based next-generation wireless networks," *IEEE Open J. Veh. Technol.*, vol. 1, pp. 45–66, 2020.

[2] S. Vitturi, C. Zunino, and T. Sauter, "Industrial communication systems and their future challenges: Next-generation Ethernet, IIoT, and 5G," *Proc. IEEE*, vol. 107, no. 6, pp. 944–961, Jun. 2019.

[3] G. Aceto, V. Persico, and A. Pescapé, "A survey on information and communication technologies for industry 4.0: State-of-the-art, taxonomies, perspectives, and challenges," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3467–3501, 4th Quart., 2019.

[4] *Service Requirements for Machine-Type Communications (MTC), Version 16.0.0, Release 16*, 3GPP Standard TS 22.368, Jul. 2020.

[5] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, and M. Gidlund, "Industrial Internet of Things: Challenges, opportunities, and directions," *IEEE Trans. Ind. Informat.*, vol. 14, no. 11, pp. 4724–4734, Nov. 2018.

[6] G. Brown, "Ultra-reliable low-latency 5G for industrial automation," San Diego, CA, USA, Qualcomm, White Paper, 2018.

[7] "5G for connected industries and automation," Frankfurt, Germany, 5G Alliance Connected Ind. Autom., White Paper, Feb. 2019.

[8] S. K. Sharma and X. Wang, "Toward massive machine type communications in ultra-dense cellular IoT networks: Current issues and machine learning-assisted solutions," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 1, pp. 426–471, 1st Quart. 2020.

[9] M. Z. Ali, J. Mišić, and V. B. Mišić, "Performance evaluation of heterogeneous IoT nodes with differentiated QoS in IEEE 802.11ah RAW mechanism," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3905–3918, Apr. 2019.

[10] H. Sarieddeen, N. Saeed, T. Y. Al-Naffouri, and M.-S. Alouini, "Next generation terahertz communications: A rendezvous of sensing, imaging, and localization," 2019. [Online]. Available: arXiv:1909.10462.

[11] X. Shen, "Device-to-device communication in 5G cellular networks," *IEEE Netw.*, vol. 29, no. 2, pp. 2–3, Mar./Apr. 2015.

[12] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.

[13] Z. Shi, W. Gao, S. Zhang, J. Liu, and N. Kato, "Machine learning-enabled cooperative spectrum sensing for non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, pp. 5692–5702, Sep. 2020.

[14] Y. Du *et al.*, "Efficient multi-user detection for uplink grant-free NOMA: Prior-information aided adaptive compressive sensing perspective," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2812–2828, Dec. 2017.

[15] G. Ma, B. Ai, F. Wang, and Z. Zhong, "Joint design of coded tandem spreading multiple access and coded slotted ALOHA for massive machine-type communications," *IEEE Trans. Ind. Informat.*, vol. 14, no. 9, pp. 4064–4071, Sep. 2018.

[16] H. G. Moussa and W. Zhuang, "RACH performance analysis for large-scale cellular IoT applications," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3364–3372, Apr. 2019.

[17] A.-T. H. Bui, C. T. Nguyen, T. C. Thang, and A. T. Pham, "A comprehensive distributed queue-based random access framework for mMTC in LTE/LTE-A networks with mixed-type traffic," *IEEE Trans. Veh. Technol.*, vol. 68, no. 12, pp. 12107–12120, Dec. 2019

[18] R. Abbas, M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "Random access for M2M communications with QoS guarantees," *IEEE Trans. Commun.*, vol. 65, no. 7, pp. 2889–2903, Jul. 2017.

[19] M. Vilgelm, S. R. Liñares, and W. Kellerer, "Dynamic binary countdown for massive IoT random access in dense 5G networks," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6896–6908, Aug. 2019.

[20] *Technical Specification Group Services and System Aspects; Release 15 Description; Summary of Rel-15 Work Items (Release 15), Version 15.0.0*, 3GPP Standars TR 21.915, Sep. 2019.

[21] J. Thota and A. Aijaz, "On performance evaluation of random access enhancements for 5G uRLLC," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Marrakesh, Morocco, Apr. 2019, pp. 1–7.

[22] Y. Liu, Y. Deng, M. Elkashlan, A. Nallanathan, and G. K. Karagiannidis, "Analyzing grant-free access for URLLC service," *IEEE J. Sel. Areas Commun.*, early access, Aug. 24, 2020, doi: 10.1109/JSAC.2020.3018822.

[23] *Technical Specification Group Services and System Aspects; Release 16 Description; Summary of Rel-16 Work Items (Release 16), Version 0.6.0*, 3GPP Standard TR 21.916, Sep. 2020.

[24] G. Pocovi, B. Soret, K. I. Pedersen, and P. Mogensen, "MAC layer enhancements for ultra-reliable low-latency communications in cellular networks," in *Proc. IEEE ICC Workshops*, Paris, France, May 2017, pp. 1005–1010.

[25] L. You, Q. Liao, N. Pappas, and D. Yuan, "Resource optimization with flexible numerology and frame structure for heterogeneous services," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2579–2582, Dec. 2018.

[26] *Study on NR Industrial Internet of Things (IoT), V16.0.0, Release 16*, 3GPP Standard TR 38.825, Mar. 2019.

[27] C. W. Park, D. Hwang, and T.-J. Lee, "Enhancement of IEEE 802.11ah MAC for M2M communications," *IEEE Commun. Lett.*, vol. 18, no. 7, pp. 1151–1154, Jul. 2014.

[28] Y. Kim, G. Hwang, J. Um, S. Yoo, H. Jung, and S. Park, "Throughput performance optimization of super dense wireless networks with the renewal access protocol," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3440–3452, May 2016.

[29] T.-C. Chang, C.-H. Lin, K. C.-J. Lin, and W.-T. Chen, "Traffic-aware sensor grouping for IEEE 802.11ah networks: Regression based analysis and design," *IEEE Trans. Mobile Comput.*, vol. 18, no. 3, pp. 674–687, Mar. 2019.

[30] V. Rodoplu, M. Nakip, D. T. Eliiyi, and C. Güzeliş, "A multiscale algorithm for joint forecasting–scheduling to solve the massive access problem of IoT," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8572–8589, Sep. 2020.

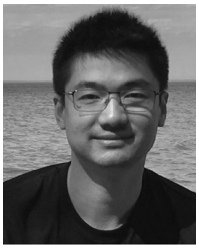
[31] Q. Ye and W. Zhuang, "Distributed and adaptive medium access control for Internet-of-Things-enabled mobile networks," *IEEE Internet Things J.*, vol. 4, no. 2, pp. 446–460, Apr. 2017.

[32] A. D. Shoaei, M. Derakhshani, and T. Le-Ngoc, "Reconfigurable and traffic-aware MAC design for virtualized wireless networks via reinforcement learning," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5490–5505, Aug. 2019.

[33] N. Kato, B. Mao, F. Tang, Y. Kawamoto, and J. Liu, "Ten challenges in advancing machine learning technologies toward 6G," *IEEE Wireless Commun.*, vol. 27, no. 3, pp. 96–103, Jun. 2020.

[34] J. Gao, M. Li, W. Zhuang, X. Shen, and X. Li, "MAC for machine type communications in industrial IoT—Part II: Scheduling and numerical results," *IEEE Internet Things J.*, early access, Dec. 18, 2020, doi: 10.1109/JIOT.2020.3045831.

[35] D. M. Kim, R. B. Sorensen, K. Mahmood, O. N. Osterbo, A. Zanella, and P. Popovski, "Data aggregation and packet bundling of uplink small packets for monitoring applications in LTE," *IEEE Netw.*, vol. 31, no. 6, pp. 32–38, Nov./Dec. 2017.
[36] C. Bockelmann *et al.*, "Massive machine-type communications in 5G: Physical and MAC-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, Sep. 2016.
[37] L. Alonso, R. Agusti, and O. Sallent, "A near-optimum MAC protocol based on the distributed queueing random access protocol (DQRAP) for a CDMA mobile communication system," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 9, pp. 1701–1718, Sep. 2000.
[38] P. Wang and W. Zhuang, "A collision-free MAC scheme for multimedia wireless mesh backbone," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3577–3589, Jul. 2009.
[39] H. S. Dhillon, H. Huang, H. Viswanathan, and R. A. Valenzuela, "Fundamentals of throughput maximization with random arrivals for M2M communications," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 4094–4109, Nov. 2014.
[40] K. R. Malekshan, W. Zhuang, and Y. Lostanlen, "An energy efficient MAC protocol for fully connected wireless ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 10, pp. 5729–5740, Oct. 2014.
[41] K. R. Malekshan, W. Zhuang, and Y. Lostanlen, "Coordination-based medium access control with space-reservation for wireless ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1617–1628, Feb. 2016.
[42] J. Gao, M. Li, L. Zhao, and X. Shen, "Contention intensity based distributed coordination for V2V safety message broadcast," *IEEE Trans. Veh. Tech.*, vol. 67, no. 12, pp. 12288–12301, Dec. 2018.



Jie Gao (Member, IEEE) received the M.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Alberta, Edmonton, AB, Canada, in 2009 and 2014, respectively.

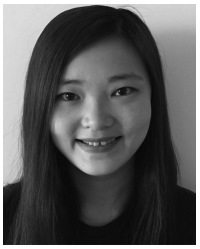
He joined the Department of Electrical and Computer Engineering, Marquette University, Milwaukee, WI, USA, as an Assistant Professor in August 2020. He was a Research Associate with the University of Waterloo, Waterloo, ON, Canada, from 2019 to 2020 and a Postdoctoral Fellow with Ryerson University, Toronto, ON, Canada, from 2017 to 2019. His research interests include machine learning for communications and networking, Internet of Things (IoT) and industrial IoT solutions, and cloud and edge computing.

Dr. Gao was a recipient of the Natural Science and Engineering Research Council of Canada Postdoctoral Fellowship and the Ontario Centres of Excellence TalentEdge Fellowship. He is an Editor for IEEE ACCESS.

Weihua Zhuang (Fellow, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from Dalian Maritime University, Dalian, China, and the Ph.D. degree in electrical engineering from the University of New Brunswick, Fredericton, NB, Canada.

She has been with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, since 1993, where she is a Professor and a Tier I Canada Research Chair of Wireless Communication Networks.

Prof. Zhuang was a recipient of the 2017 Technical Recognition Award from the IEEE Communications Society Ad Hoc and Sensor Networks Technical Committee, and a co-recipient of several Best Paper Awards from IEEE conferences. She was the Editor-in-Chief of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY from 2007 to 2013, the Technical Program Chair/Co-Chair of IEEE VTC 2017/2016 Fall, and the Technical Program Symposia Chair of IEEE Globecom 2011. She is an Elected Member of the Board of Governors and Vice President—Publications of the IEEE Vehicular Technology Society. She was an IEEE Communications Society Distinguished Lecturer from 2008 to 2011. She is a Fellow of the Royal Society of Canada, the Canadian Academy of Engineering, and the Engineering Institute of Canada.



Mushu Li (Student Member, IEEE) received the B.Eng. degree from the University of Ontario Institute of Technology, Oshawa, ON, Canada, in 2015, and the M.A.Sc. degree from Ryerson University, Toronto, ON, Canada, in 2017. She is currently pursuing the Ph.D. degree in electrical engineering with the University of Waterloo, Waterloo, ON, Canada.

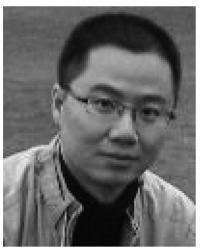
Her research interests include the system optimization in VANETs and machine learning in wireless networks.

Ms. Li was a recipient of NSERC CGS scholarship in 2018 and OGS in 2015 and 2016, respectively.

Xuemin (Sherman) Shen (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990.

He is currently a University Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research focuses on network resource management, wireless network security, Internet of Things, 5G and beyond, and vehicular *ad hoc* and sensor networks.

Dr. Shen received the R. A. Fessenden Award in 2019 from IEEE, Canada, the Award of Merit from the Federation of Chinese Canadian Professionals (Ontario) in 2019, the James Evans Avant Garde Award in 2018 from the IEEE Vehicular Technology Society, the Joseph LoCicero Award in 2015 and Education Award in 2017 from the IEEE Communications Society, the Technical Recognition Award from Wireless Communications Technical Committee in 2019, and the AHSN Technical Committee 2013. He has also received the Excellent Graduate Supervision Award in 2006 from the University of Waterloo and the Premier's Research Excellence Award in 2003 from the Province of Ontario, Canada. He served as the Technical Program Committee Chair/Co-Chair for IEEE Globecom'16, IEEE Infocom'14, IEEE VTC'10 Fall, and IEEE Globecom'07, and the Chair for the IEEE Communications Society Technical Committee on Wireless Communications. He was the Elected IEEE Communications Society Vice President for Technical and Educational Activities, a Vice President for Publications, the Member-at-Large on the Board of Governors, a Chair of the Distinguished Lecturer Selection Committee, and a member of IEEE ComSoc Fellow Selection Committee. He was the Editor-in-Chief of the IEEE INTERNET OF THINGS JOURNAL, IEEE NETWORK, and *IET Communications*. He is a registered Professional Engineer of Ontario, Canada, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, a Chinese Academy of Engineering Foreign Member, and a Distinguished Lecturer of the IEEE Vehicular Technology Society and Communications Society.



Xu Li (Member, IEEE) received the B.Sc. degree in computer science from Jilin University, Changchun, China, in 1998, the M.Sc. degree from the University of Ottawa, Ottawa, ON, Canada, in 2005, and the Ph.D. degree from Carleton University, Ottawa, in 2008.

He worked as a Research Scientist (with tenure) with Inria, Le Chesnay, France. He is a Senior Principal Researcher with Huawei Technologies Canada, Ottawa. He contributed extensively to the development of 3GPP 5G standards through over 90 standard proposals. He has published over 100 refereed scientific papers and is holding over 30 issued U.S. patents. His current research interests are focused in 5G.

Dr. Li is/was on the editorial boards of *IEEE Communications Magazine*, the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, the *Transactions on Emerging Telecommunications Technologies* (Wiley), and a number of other international archive journals.