Distributed and Adaptive Medium Access Control for Internet-of-Things-Enabled Mobile Networks

Qiang Ye, *Student Member, IEEE*, and Weihua Zhuang, *Fellow, IEEE*

***Abstract*—In this paper, we propose a distributed and adaptive hybrid medium access control (DAH-MAC) scheme for a single-hop Internet-of-Things (IoT)-enabled mobile ad hoc network (MANET) supporting voice and data services. A hybrid superframe structure is designed to accommodate packet transmissions from a varying number of mobile nodes generating either delay-sensitive voice traffic or best-effort data traffic. Within each superframe, voice nodes with packets to transmit access the channel in a contention-free period using distributed time division multiple access (TDMA), while data nodes contend for channel access in a contention period using truncated carrier sense multiple access with collision avoidance (T-CSMA/CA). In the contention-free period, by adaptively allocating time slots according to instantaneous voice traffic load, the MAC exploits voice traffic multiplexing to increase the voice capacity. In the contention period, a throughput optimization framework is proposed for the DAH-MAC, which maximizes the aggregate data throughput by adjusting the optimal contention window size according to voice and data traffic load variations. Numerical results show that the proposed MAC scheme outperforms existing QoS-aware MAC schemes for voice and data traffic in the presence of heterogeneous traffic load dynamics.**

***Index Terms*—IoT, MANETs, medium access control, TDMA, T-CSMA/CA, voice capacity, data throughput.**

I. Introduction

INTERNET-OF-THINGS (IoT)-enabled mobile networks have been envisioned as an important revolution for future generation wireless networks, which facilitate a growing number of smart heterogeneous objects (e.g., smart sensors, home appliances, monitoring devices, etc.) being connected via suitable wireless technologies to the Internet, in order to achieve ubiquitous information access and seamless communication interaction [1]. A mobile ad hoc network (MANET) is one typical example of IoT-enabled mobile networks, allowing a group of smart mobile devices (e.g., smartphones, smart sensor nodes, etc.) interconnected in a peer-to-peer manner and communicating without relying on any network infrastructure. MANETs are widely deployed for pervasive IoT-oriented applications, for example, smart home/office networking [1], emergency communications and crisis response in disaster areas [2]–[4], and tactical networks for the purpose of command interactions [5]. Device-to-Device (D2D) communication networks have been recently popularized as one of the typical realizations for IoT-enabled MANETs, which rely on the ad hoc networking of spatially-distributed smart devices for information relaying and sharing in some postdisaster areas where the communication infrastructures are temporarily destroyed [3] [4]. In [3], the information dissemination in disaster-stricken areas is realized through collaborative routing based on intermittent group-forming and group-disbanding among both pedestrians and vehicles equipped with smart devices. Further, D2D communications are also employed in the next generation wireless cellular networks to fulfill the ever-increasing communication demands by sharing the spectrum resources with cellular users [6] [7]. Some IoT-based wireless networks are designed to support a large number of power-constrained sensor nodes or autonomous devices generating low data rate traffic. Therefore, energy-efficient medium access control (MAC) is required to coordinate packet transmissions among wireless nodes to prolong the whole network lifetime. Such studies include the carrier sense multiple access with collision avoidance (CSMA/CA)-based IEEE 802.15.4 ZigBee supporting energy efficient and low data rate communications in wireless sensor networks (WSNs) [8], a data gathering protocol in a IoT-based WSN with time division multiple access (TDMA) employed for intra-cluster data transmissions [9], the IEEE 802.11ah operating at lower frequency bands to cover an increasing number of machine-to-machine (M2M) devices for an outdoor M2M network environment [10]. However, these MAC protocols, developed to support low data rate applications on power-limited devices, cannot guarantee the differentiated quality-of-service (QoS) requirements for heterogeneous services [2].

For a typical IoT-enabled MANET with power-rechargeable mobile nodes [11] (i.e., smartphones, laptops) generating a high volume of heterogeneous traffic, supporting heterogeneous services with differentiated QoS guarantee becomes an important but challenging task. The network is expected to not only provide as high as possible throughput for best-effort data traffic, but also ensure a bounded packet loss rate for delay-sensitive voice communications or even multimedia streaming. Therefore, QoS-aware MAC is required to coordinate the channel access for heterogeneous traffic in a differentiated way to satisfy QoS requirements of all individual users [12]. However, the characteristics of MANETs pose technical challenges in the QoS-aware MAC design: 1) Since MANETs do not depend on any central control, distributed MAC is required to coordinate the transmissions of neighboring nodes based on their local information exchanges; 2) Nodes are mobile, making the heterogeneous network traffic load change with time. The traffic load variations can lead to QoS performance

This work was supported by a research grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Qiang Ye and Weihua Zhuang are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, N2L 3G1. E-mail: {q6ye, wzhuang}@uwaterloo.ca.

degradation. Thus, MAC is expected to be context-aware, which adapts to the changing network traffic load to achieve consistently satisfactory service performance.

In this paper, we propose a distributed and adaptive hybrid MAC scheme (DAH-MAC), in which distributed TDMA is employed for voice packet transmissions to guarantee a voice packet loss rate bound and truncated CSMA/CA (T-CSMA/CA) is used for data nodes to access the channel. Most of the existing contention-based MAC schemes evaluate the average access delay for voice traffic in a saturation condition or with a constant arrival rate [13], which is not the case in reality. We use a more accurate *on/off* model [14] for voice traffic generation, and exploit voice traffic multiplexing to improve the voice capacity [14]. Since voice service is realtime, a packet not transmitted after a delay bound should be dropped at the source, and the voice packet delay has to be evaluated in a stochastic manner [15] for calculating the packet loss probability. In this way, the delay requirement for voice traffic can be satisfied probabilistically by guaranteeing the voice packet loss rate below a given bound. The contributions of this work are three-folded:

1) To guarantee the voice packet loss rate bound, we present a distributed and traffic-adaptive TDMA time slot allocation scheme to allocate one time slot for each active voice node according to its transmission buffer state. Also, we establish an analytical model so that the MAC scheme can determine the voice capacity region by adjusting a MAC parameter, i.e., the maximum time fraction allocation requirement for voice traffic in each superframe, which facilitates voice session admission control for QoS guarantee. By exploiting the voice traffic multiplexing, the resource utilization for voice traffic is improved significantly;
2) The T-CSMA/CA based contention scheme is employed for data traffic access. We establish an analytical model of data saturation throughput for the DAH-MAC. The saturation throughput is a function of the number of voice and data nodes as well as the packet transmission probability of each data node;
3) For the saturation throughput of the DAH-MAC, we derive an approximate closed-form expression of the optimal data packet transmission probability as a function of the heterogeneous network traffic load. Further, we obtain a closed-form expression of the optimal contention window size, which establishes a mathematical relationship between the MAC layer parameter and the heterogeneous network traffic load. Based on the analysis, the maximum best-effort data saturation throughput can be achieved by adjusting the contention window size according to variations of the number of voice and data nodes.

We organize the reminder of the paper as follows. Section II reviews existing MAC schemes for MANETs supporting heterogeneous services. The system model is described in Section III, and the DAH-MAC scheme is presented in Section IV. Section V provides performance analysis in terms of the average number of scheduled voice time slots and aggregate data throughput within a superframe. Numerical results are given in Section VI to evaluate the performance of the proposed DAH-MAC scheme. Finally, conclusions are drawn in Section VII. Important parameters and variables are summarized in Table I.

II. Related Work

In literature, contention-based MAC schemes with service differentiation are commonly used for supporting heterogeneous traffic [16]–[19]. The enhanced distributed channel access (EDCA), standardized in IEEE 802.11e [20], is one typical example, in which delay-sensitive realtime traffic is granted smaller arbitration interframe space (AIFS) and contention window size to access the channel with a higher probability than non-realtime traffic [16] [17]. It is demonstrated in [21] that the contention window size differentiation among realtime and non-realtime traffic is superior over the AIFS differentiation in achieving a smaller access delay for the realtime service in a traffic saturation condition. To grant voice traffic deterministic channel access priority for further improving the delay performance, busy-tone based contention protocols are proposed [14] [22], in which each voice node broadcasts a busy-tone signal, instead of decrementing a backoff counter, after an idle AIFS duration to prevent the contention intervention from data nodes. Even if the contention separation is achieved between voice traffic and data traffic in busy-tone based protocols, contention collisions still exist and accumulate among voice (data) nodes themselves after the voice (data) traffic load becomes relatively high, making the delay (throughput) performance degraded to an unacceptable level.

By avoiding contention collisions, distributed TDMA schemes [23] [24] allocate time slots to each node in a distributed way for exclusive use. They are more effective than contention-based MAC schemes in guaranteeing the delay of realtime traffic especially in a relative high traffic load condition, where channel time is accumulated and wasted for packet collisions resolution in contention-based schemes. To maximize resource utilization, the distributed TDMA time slot allocation should be adaptive to the instantaneous voice traffic load [23]. In [25], a TDMA-based distributed packet reservation multiple access (D-PRMA) protocol is proposed, in which voice nodes are granted a higher probability than data nodes to contend for the channel based on slotted-Aloha. Once a contention is successful, the same time slot in each subsequent frame is reserved for the successful voice node until the slot is detected idle. Each D-PRMA frame consists of a fixed number of time slots to support transmissions from a certain number of voice and data nodes, which is not flexible when the number of nodes varies over a wide range. In addition, TDMA-based schemes can be underperformed in supporting best-effort data traffic. Since the data traffic generation is bursty, some of the time slots are wasted when the traffic load is relatively low.

To guarantee the voice delay bound and achieve high resource utilization with multiplexing for best-effort data, hybrid MAC schemes are better options which combine a TDMA period for voice transmissions and a contention period for

TABLE I: Important Parameters and Variables

Parameter & Variable	Definition
$\frac{1}{\alpha}$ / $\frac{1}{\beta}$	Average duration in *on/off* state of a voice node
λ/λ_d	Average packet arrival rate at each voice/data node
τ	Data packet transmission probability
τ_{opt}	Optimal data packet transmission probability
φ	Maximum time fraction for voice traffic in each superframe
B	Average voice burst size
CW_{opt}	Optimal contention window size
M	Maximum number of voice packets generated in a superframe
N_a	Number of active voice nodes in each superframe
N_d/N_v	Number of data/voice nodes
$\overline{N_s}$	Average number of voice bursts scheduled for transmission
N_{sm}	Maximum number of voice bursts scheduled for transmission
N_{vm}	Number of minislots in each control period (voice capacity)
p	Data packet collision probability
P_L	Voice packet loss rate bound
p_v	Probability of a generic time slot inside the vulnerable period
S_d	Normalized saturation data throughput for T-CSMA/CA
T_c	Duration of a conflict period
$\overline{T_{cfp}}$	Average duration of contention-free period
T_{cfpm}	Maximum duration of contention-free period
$\overline{T_{cp}}$	Average duration of contention period
T_{ctrl}	Control period duration
T_m	Minislot duration
T_{pd}/T_{pv}	Data/voice packet duration
T_s	Complete data packet transmission time
T_{SF}	Superframe duration
$\overline{T_v}$	Average duration of vulnerable period
$\overline{T_{vd}}$	Average virtual transmission time
y_m	Maximum number of transmitted voice packets in each superframe

data transmissions using CSMA/CA-based mechanisms within a superframe [26] [27]. In [14], a hybrid MAC scheme is developed for wireless local area networks (WLANs), in which voice nodes in talk spurts are polled by an access point (AP) to transmit packets in a contention-free period, whereas the remaining idle voice nodes, once having packets to transmit, contend with data nodes to access the channel according to the busy-tone contention protocol in a contention period. Since contention-based MAC schemes experience throughput degradation with an increase of network traffic load, some existing methods adapt the contention window size to node density [28] or node relative velocity [29] to achieve consistently high network throughput. However, within a hybrid MAC superframe structure, how to achieve a consistently maximal data throughput over heterogeneous traffic load variations and how to adaptively allocate time slots based on instantaneous voice traffic load in a distributed way to maximize voice traffic multiplexing gain still remain unsolved.

III. System Model

Consider a single-channel fully connected MANET [25] [30] [31], where each node can receive packet transmissions from any other node. The fully connected network scenario can be found in various MANET applications, including office networking in a building or in a university library where users are restricted to move in certain geographical areas [31], users within close proximity are networked with ad hoc mode in a conference site [12], M2M communications in a residential network for a typical IoT-based smart home application where home appliances are normally within the communication range of each other [2]. The channel is assumed error-free, and packet collisions occur when more than one node simultaneously initiate packet transmission attempts. Without any network infrastructure or centralized controller, nodes exchange local information with each other and make their transmission decisions in a distributed manner. The network has two types of nodes, voice nodes and data nodes, generating delay-sensitive voice traffic and best-effort data traffic, respectively. Each node is identified by its MAC address and a unique node identifier (ID) that can be randomly selected and included in each transmitted packet [24]. We use N_v and N_d to denote the total numbers of voice and data nodes in the network coverage area, respectively. Nodes are mobile with a relatively low speed, making N_v and N_d change with time.

For delay-sensitive voice traffic, each packet should be successfully transmitted within a bounded delay to achieve an acceptable voice communications quality; otherwise, the packet will be dropped. Therefore, as a main QoS metric for voice traffic, packet loss rate should be carefully controlled under a given threshold, denoted by P_L (e.g., 10^{-2}). The generic *on/off* characteristic of voice traffic allows traffic multiplexing in transmission. Each voice source node is represented by an *on/off* model, which is a two-state Markov process with the *on* and *off* states being the talk spurt and silent periods, respectively. Both periods are independent and exponentially distributed with respective mean $\frac{1}{\alpha}$ and $\frac{1}{\beta}$. During a talk spurt, voice packets are generated at a constant rate, λ packet/s. As for best-effort data traffic, data nodes are expected to exploit limited wireless resources to achieve as high as possible aggregate throughput. It is assumed that each data node always has packets to transmit. Nodes in the network are synchronized in time, which can be achieved such as by using the 1PPS signal with a global positioning system (GPS) receiver [24] [25].

In the network, time is partitioned into superframes of constant duration, denoted by T_{SF}, which is set to have the same duration as the delay bound of voice traffic. Each superframe is further divided into three periods: control period (CTP), contention-free period (CFP) and contention period (CP), the durations of which are denoted by T_{ctrl}, T_{cfp} and T_{cp} respectively, as shown in Fig. 1. The control period consists

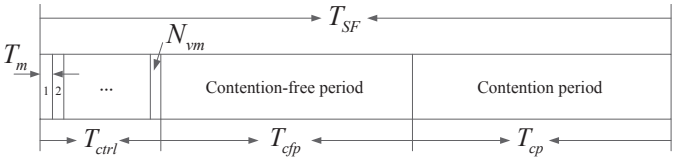


Fig. 1: Superframe structure.

of N_{vm} fixed-duration (T_m) minislots, each with a unique minislot sequence number. It is to support a varying number of voice nodes in the network. Each voice node selects a unique minislot and broadcasts local information in its selected minislot, for distributed TDMA time slot allocation in the following contention-free period [32]. In the context of higher service priority to voice traffic, to avoid a complete deprivation of data service, there is a maximum fraction of time, φ (< 1), for voice traffic in each superframe. The value of φ is assumed known to all nodes when the network operation starts, and can be broadcast by the existing nodes in each control period. The voice capacity is the maximum number of voice nodes allowed in the network, denoted by N_{vm} (same as the number of minislots in each CTP), under the QoS constraint, which depends on φ. The period following the control period is the CFP, which is further divided into multiple equal-duration TDMA time slots, each slot having a unique sequence number. Each voice node with packets to transmit (referred to as active voice node) occupies one time slot to transmit a number of voice packets, called a *voice burst*[1]. Thus, the number of TDMA slots in the CFP is determined by the number of voice burst transmissions scheduled for the superframe, denoted by N_s ($N_s \leq N_v$).

The last period CP is dedicated to best-effort data nodes for transmission according to T-CSMA/CA. Data packet transmissions are based on CSMA/CA and are periodically interrupted by the presence of CTP and CFP.

IV. The DAH-MAC Scheme

In the following, we illustrate how voice nodes access their TDMA slots in each CFP without a central controller. For data nodes accessing the channel using T-CSMA/CA, we highlight differences between the T-CSMA/CA within the proposed hybrid superframe structure and the traditional CSMA/CA.

A. Accessing Minislots

In the distributed MAC, each voice node needs to exchange information with neighboring voice nodes by broadcasting control packets in the minislots in the control period of each superframe. When the network operation starts, the number of minislots (voice capacity N_{vm}) in the control period should be determined in a distributed way under the constraint that voice packet loss rate is bounded by P_L and the summation of T_{ctrl} and T_{cfp} does not exceed $\varphi \cdot T_{SF}$ in each superframe. After N_{vm} is determined, each voice node randomly chooses one minislot in the CTP of a superframe, and broadcasts a control packet in its selected minislot [32]. Each node broadcasts its control packet in the same occupied minislot of each subsequent superframe[2], until it is powered off or departs from the network. A control packet, shown in Fig. 2, includes five fields: a header, a set of IDs of the node's neighbors including the node itself, the node's occupied minislot sequence number (MSN, chosen from 1 to N_{vm}), buffer occupancy indication bit (BIB), the node's scheduled TDMA slot sequence number (SSN, a number within 0 to N_s) in the previous superframe.

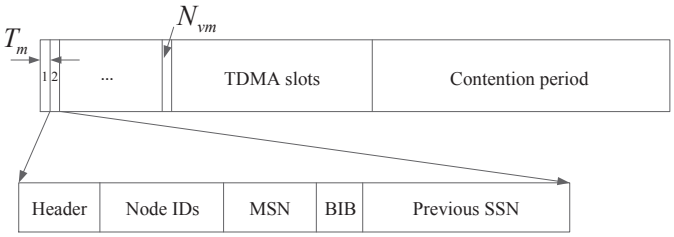


Fig. 2: Format of control packet broadcast in each minislot.

Accessing a minislot from a tagged voice node is considered successful if the control packets received from other nodes in subsequent minislots contain the tagged node's ID [34]. Otherwise, an *access collision* happens due to simultaneous control packet transmissions in the same minislot by more than one node. All nodes involved in the collision wait until the next superframe to re-access one of the vacant minislots. The minislot accessing process is completed when all existing nodes successfully acquire their respective minislots. When

[1]A voice burst is the packets generated by an active voice node within one superframe that can be transmitted over a time slot.

[2]To ensure fair minislot access, voice nodes re-select minislots after using the previous ones for a predefined number of successive superframes [33].

a new node is powered on or entering the network coverage area, it first synchronizes in time with the start of a superframe, determines the number of minislots (based on φ), N_{vm}, and listens to all control packets in the CTP. Then, it randomly selects an unoccupied minislot and broadcasts a control packet in the minislot in the next superframe. If all N_{vm} minislots are occupied, which means the whole network reaches its voice capacity, the node defers its channel access and keeps sensing the CTPs of subsequent superframes until some existing minislots are released due to node departures. After the minislot accessing is successful, the node keeps using the same minislot of subsequent superframes to broadcast its control packet.

B. Adaptive TDMA Time Slot Allocation

For efficient resource utilization, time slot allocation to voice nodes should adapt to traffic load variations. Taking account of the voice traffic *on/off* characteristic, only active nodes should be allocated one time slot each, in a superframe. We divide active voice nodes into two categories: Type I and Type II nodes. Type I nodes in the current superframe were not allocated a time slot in the previous superframe[3], and are named "current-activated" nodes; Type II nodes remain active in both previous and current superframes, and are called "already-activated" nodes.

For each Type I node, voice traffic transits from the *off* state to the *on* state during previous superframe and generates voice packets before the node broadcasts a control packet in current superframe. Because of the randomness of state transition time from *off* to *on* in the previous superframe, packet transmissions from Type I nodes should have higher priority to be scheduled as early as possible according to their minislot accessing sequence, as long as Type II nodes can transmit within the delay bound, in order to minimize the possibility of Type I packet loss due to delay bound violation. Each Type II node has a time slot in the previous superframe and remains active in the current superframe. It should transmit packets no later than in the same time slot in the current superframe to meet the delay bound requirement. As an example, Fig. 3 illustrates how TDMA time slots are allocated in one superframe, with $N_v = 9$ and $N_{vm} = 10$, i.e., how to obtain current SSN based on the information in control packets. Each node has a unique minislot. Three types of important information in a broadcast control packet are shown: 1) MSN, the unique sequence number of a specific minislot; 2) BIB, which is 1 if the node has packets to transmit and 0 otherwise; 3) previous SSN, showing the sequence number of a TDMA time slot allocated to a voice node in the previous superframe, with SSN = 0 if the node was not allocated a TDMA time slot. Each entry in the left part of Fig. 3 discloses the information broadcast by the nodes in their minislots. The information broadcast by Type I and Type II active nodes is distinguished by the dashed-line and solid-line bounding rectangles, respectively. It can be seen that Nodes 8 and 5 are Type I with BIB = 1 and with previous SSN = 0, whereas Nodes 7, 1 and 4 are Type II with BIB = 1 and previous SSN $\neq$ 0.

[3]In most cases, the reason of not having a time slot is that the voice node has no packets to transmit. In some occasions when the instantaneous voice traffic load becomes heavy, an active node may not be able to get a time slot, resulting in packet dropping at the transmitter.

Node ID	MSN	BIB	Previous SSN		Current SSN
Node 7	1	1	3		3
Node 8	2	1	0		1
Node 5	3	1	0		4
Node 3	4	0	1		0
	5				
Node 1	6	1	2		2
Node 2	7	0	5		0
Node 6	8	0	4		0
Node 4	9	1	6		5
Node 9	10	0	0		0

Fig. 3: An example of TDMA time slot allocation.

Based on the information provided by the control packets, the right part of Fig. 3 shows the time slot allocation in the current superframe. Node 8 accessing the first minislot among all the Type I nodes will transmit in the first time slot (with current SSN = 1) in the CFP. Node 5 is allocated a time slot after Nodes 1 and 7 because the latter two nodes are Type II nodes and should transmit packets no later than in their previously allocated time slots (with previous SSN = 2 and 3 respectively). Packet transmissions in current superframe from Node 4 (a Type II node with the largest previous SSN) are scheduled in a time slot with the index less than its previously allocated time slot.

C. T-CSMA/CA based Contention Access

In DAH-MAC, best-effort data nodes access the channel within a CP of each superframe according to the T-CSMA/CA contention protocol, in which data nodes attempt packet transmissions according to CSMA/CA with exponential backoff [35] and the transmissions are periodically interrupted by the presence of a CTP and a CFP. Thus, the performance of T-CSMA/CA is different from the traditional CSMA/CA contention protocol without interruptions. First, the packet waiting time for transmission is increased by the interrupted periods; Second, before each source node initiates a packet transmission attempt at the end of its backoff counter decrementing process, it is required to check whether the remaining time in the CP is enough to support at least one packet transmission. To have an acceptable transmission attempt, the remaining time should be not less than the summation of a data packet duration (T_{pd}, including acknowledgment) and a gurad time (T_{gt}). This summation is called *conflict period*. If the remaining time in current CP is not long enough, a *virtual conflict* occurs with the imminent CTP of next superframe. A *hold-on* strategy can be used to resolve the conflict [26] [27], in which the packet attempts are suspended until the start of next CP. Other nodes that are not involved in the conflict

can still decrement their backoff counters within the conflict period until the end of current CP. When the next CP arrives, the transmission process resumes and the suspended packets are transmitted immediately after the channel is sensed idle for a distributed interframe space (DIFS).

By referring to some methods in [26], we give a detailed illustration inside the CP of each superframe, shown in Fig. 4, to highlight the differences between the T-CSMA/CA and the traditional CSMA/CA protocol. Suppose that the CP starts at

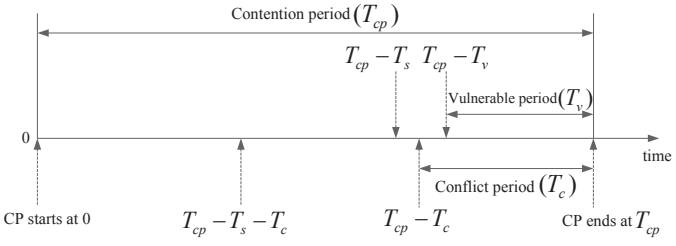


Fig. 4: An illustration of the CP.

time instant 0 and ends at T_{cp}. If a packet transmission attempt is initiated within the interval $[0, T_{cp}-T_s]$, the packet can be transmitted according to the CSMA/CA, either successfully or in collision, with a complete transmission duration T_s (T_{pd} plus a DIFS interval). The time instant $T_{cp}-T_c$ denotes the last time instant at which a packet transmission attempt can be initiated, and the conflict period is the following interval with duration T_c (T_{pd} plus T_{gt}), which is smaller than T_s. Thus, if a packet transmission starts in the interval $[T_{cp}-T_s, T_{cp}-T_c]$, the packet transmission time is on average $\frac{T_c+T_s}{2}$, assuming the transmission initiation instant is uniformly distributed within the interval. On the other hand, if the last packet transmission within the CP starts in the interval $[T_{cp}-T_s-T_c, T_{cp}-T_c]$, the transmission finishing point, denoted by $T_{cp}-T_v$, lies in the conflict period, where T_v is called *vulnerable period* [26] indicating the residual idle interval between the last transmission finishing point and the end of the CP; If no transmissions initiate during $[T_{cp}-T_s-T_c, T_{cp}-T_c]$, the starting point of the vulnerable period is $T_{cp}-T_c$, and the vulnerable period is the same as the conflict period. Thus, it can be seen that the vulnerable period is always not longer than the conflict period. The time interval $[0, T_{cp}-T_v]$ before the vulnerable period is called *non-vulnerable period*.

V. Performance Analysis

In this section, firstly, for a given maximum fraction of time (φ) for voice traffic in each superframe, the voice capacity, N_{vm}, under the packet loss rate bound is derived, which can facilitate voice session admission control. Secondly, for a specific N_v, the average number of voice burst transmissions scheduled in each superframe, $\overline{N_s}$, is derived, with which the average time duration of each CFP and CP, denoted by $\overline{T_{cfp}}$ and $\overline{T_{cp}}$, can be determined. Then, the aggregate throughput of the DAH-MAC for N_d data nodes is evaluated for each superframe, and maximized by adjusting the contention window size to the optimal value according to variations of N_v and N_d.

A. Voice capacity

As mentioned in Section IV, when nodes come into the network coverage area, they distributedly calculate the number of minislots, N_{vm}, in each CTP, indicating the maximum number of voice nodes supported in the network. Thus, within the voice capacity region, the following inequality needs to be satisfied to guarantee that the time duration for voice traffic not exceed the maximum fraction (φ) of each superframe time,

$$T_{ctrl}+T_{cfpm}=N_{vm}T_m+N_{sm}\lceil B\rceil T_{pv}\le \varphi T_{SF} \quad (1)$$

where T_{cfpm} denotes the maximum value of T_{cfp}, T_m is the duration of each minislot, N_{sm} is the maximum value of N_s, indicating the maximum number of scheduled voice burst transmissions in a CFP to maintain the packet loss rate bound, B is the average size (number of voice packets) of a voice burst, $\lceil\rceil$ is the ceiling function, T_{pv} is the voice packet duration including header, and $\lceil B\rceil T_{pv}$ indicates the duration of each TDMA time slot (the duration of one time slot should allow an integer number of packet transmissions).

To determine N_{sm}, we first estimate that, with N_v voice nodes, how many generated voice packets in a superframe are required to be transmitted in the CFP to guarantee the packet loss rate bounded by P_L. Let X_i denote the number of packets generated by voice node i ($i=1,2,...,N_v$) within a superframe, and y_m denote the maximum number of transmitted voice packets in the CFP to guarantee P_L. Since the length of a superframe is to be the same as the voice packet delay bound, lost packets are estimated as those generated but not transmitted within one superframe. Thus, y_m can be calculated by solving the following equation

$$\frac{E\left[X-y_m|X>y_m\right]}{E\left[X\right]}=P_L \quad (2)$$

where $X=\sum_{i=1}^{N_v}X_i$.

Since $\{X_i, i=1,2,...,N_v\}$ are independent and identically distributed (i.i.d) random variables, X can be approximated as a Gaussian random variable when N_v becomes relatively large (based on central limit theorem) [14], with mean $E[X]$ and variance $D[X]$ being $N_vE[X_i]$ and $N_vD[X_i]$ respectively. Thus, we estimate the distribution of X as a normal distribution, with which (2) is approximated as

$$\frac{\int_{y_m}^{N_v\cdot M}\frac{x-y_m}{\sqrt{2\pi N_vD[X_i]}}\cdot e^{-\frac{(x-N_vE[X_i])^2}{2N_vD[X_i]}}dx}{N_vE[X_i]}=P_L \quad (3)$$

where $M=\lambda\cdot T_{SF}$ denotes the maximum number of packets generated by a voice source node within one superframe.

In (3), to derive $E[X_i]$ and $D[X_i]$, we calculate the distribution of X_i, which is the probability of generating k packets by voice node i within a superframe ready for transmission in the CFP, denoted by $P(k)$[4]. According to the *on/off* source model, the probability of a voice node staying at *on* (*off*) state, denoted by P_{on} (P_{off}), at any time instant, is $\frac{\beta}{\alpha+\beta}$ $\left(\frac{\alpha}{\alpha+\beta}\right)$.

[4]Since $\{X_i, i=1,2,...,N_v\}$ have identical probability distribution, we simply drop the voice node index i.

Let T_{on} (T_{off}) denote the time duration a voice node stays at *on* (*off*) state. We have

$$\begin{aligned} P(k) &= P_{on} \cdot P\left\{\frac{k-1}{\lambda} < T_{on} \le \frac{k}{\lambda}\right\} \\ &\quad + P_{off} \cdot P\left\{T_{SF} - \frac{k}{\lambda} < T_{off} \le T_{SF} - \frac{k-1}{\lambda}\right\} \\ &= \frac{\beta}{\alpha+\beta}\left[e^{-\frac{\alpha(k-1)}{\lambda}} - e^{-\frac{\alpha k}{\lambda}}\right] \\ &\quad + \frac{\alpha}{\alpha+\beta}\left[e^{-\beta\left(T_{SF}-\frac{k}{\lambda}\right)} - e^{-\beta\left(T_{SF}-\frac{k-1}{\lambda}\right)}\right] \\ &\qquad\qquad (1 \le k \le M-1) \end{aligned} \tag{4}$$

$$\begin{aligned} P(M) &= P_{on} \cdot P\left\{T_{on} > \frac{M-1}{\lambda}\right\} + P_{off} \cdot P\left\{T_{off} \le \frac{1}{\lambda}\right\} \\ &= \frac{\beta}{\alpha+\beta} e^{-\frac{\alpha(M-1)}{\lambda}} + \frac{\alpha}{\alpha+\beta}\left(1 - e^{-\frac{\beta}{\lambda}}\right) \end{aligned} \tag{5}$$

and

$$P(0) = 1 - \sum_{k=1}^{M} P(k). \tag{6}$$

With the probability distribution of X_i, the average voice burst size B can be obtained. Based on y_m and B, the maximum number of scheduled voice bursts, N_{sm}, in each CFP is derived. Then, with a specific φ, Algorithm 1 can be used to determine the maximum number of voice nodes (voice capacity), N_{vm}, that can be supported in the network.

Algorithm 1: Voice capacity

Input : The maximum fraction of time, φ, for voice traffic in each superframe.

Output: Voice capacity N_{vm}, the CTP duration T_{ctrl}.

```
Initialization: N_v ← 1;
do
    y_m ← solving (3);
    N_sm ← y_m / B;
    if N_v T_m + N_sm ⌈B⌉ T_pv ≤ φ T_SF then
        N_v ← N_v + 1;
    else
        N_vm ← N_v − 1;
        T_ctrl ← N_vm T_m;
        break;
    end
while N_v > 0;
return N_vm and T_ctrl.
```

B. Average number of scheduled voice bursts in a CFP

The actual number of generated voice bursts is likely less than N_{sm} and varies depending on the buffer occupancy states broadcast at the beginning of each superframe. In the following, for a specific N_v, we calculate the average number of scheduled voice bursts, $\overline{N_s}$, with which $\overline{T_{cfp}}$ and $\overline{T_{cp}}$ can be obtained.

We first determine the probability distribution of the number of active voice nodes, denoted by N_a, which broadcast control packets with BIB = 1 in their respective minislots of each superframe. Due to the voice source *on/off* characteristic, N_a is composed of two portions: 1) the number of nodes with a nonempty buffer staying at the *on* state, denoted by N_a^{on}; and 2) the number of nodes with a nonempty buffer staying at the *off* state, denoted by N_a^{off}.

Since packets periodically arrives at the transmission buffer every $\frac{1}{\lambda}$ second for each voice node at the *on* state, the probability of a voice node being active in its occupied minislot conditioned on that the node is at the *on* state can be derived by calculating a *posterior probability* as

$$P_{a|on} = P\left\{\text{on at } \left(t_i - \frac{1}{\lambda}\right) \middle| \text{ on at } t_i\right\} = e^{-\frac{\alpha}{\lambda}} \tag{7}$$

where t_i is the time instant that voice node i broadcasts in its selected minislot in the current superframe. Eq. (7) indicates that the time duration of voice node i staying at the *on* state should last for at least the duration of $\frac{1}{\lambda}$ before it broadcasts at t_i to ensure a nonempty transmission buffer.

Similarly, the probability of voice node i being active at t_i, conditioned on that the node stays at the *off* state, is calculated as

$$\begin{aligned} P_{a|off} &= P\left\{\text{on at } (t_i - T), \text{on at } \left(t_i - T + \frac{1}{\lambda}\right) \middle| \text{ off at } t_i\right\} \\ &= \frac{\beta}{\alpha}\left(e^{-\frac{\alpha}{\lambda}} - e^{-\alpha T}\right) \end{aligned} \tag{8}$$

where $T = T_{SF} - T_{cfpm}$. Thus, $t_i - T$ is a calculation for the time instant of the end of the CFP in previous superframe[5]. Eq. (8) indicates that voice node i should stay at the *on* state for at least the time interval of $\frac{1}{\lambda}$ after the end of previous CFP to ensure a nonempty transmission buffer before time instant t_i. Then, the probability distribution of N_a, denoted by $P_{N_a}(k)$, can be derived as

$$\begin{aligned} P_{N_a}(k) &= P\{N_a^{on} + N_a^{off} = k\} \\ &= \sum_{i=0}^{N_v} \sum_{j \in \mathbf{A}} P\{N_a^{off} = k - N_a^{on} \,|\, N_a^{on} = j, N^{on} = i\} \\ &\quad \cdot P\{N_a^{on} = j \,|\, N^{on} = i\} \cdot P\{N^{on} = i\} \\ &= \sum_{i=0}^{N_v} \sum_{j \in \mathbf{A}} P(i,j,k) \qquad\qquad (0 \le k \le N_v) \end{aligned} \tag{9}$$

where N^{on} denotes the number of nodes in the *on* state, set $\mathbf{A}$ denotes the value range of j depending on k and i, and

$$\begin{aligned} P(i,j,k) = &\binom{N_v - i}{k - j} P_{a|off}^{k-j} \left(1 - P_{a|off}\right)^{N_v - i - k + j} \\ &\cdot \binom{i}{j} P_{a|on}^{j} \left(1 - P_{a|on}\right)^{i-j} \cdot \binom{N_v}{i} P_{on}^{i} P_{off}^{N_v - i}. \end{aligned}$$

Then, the complete expression of $P_{N_a}(k)$ is obtained by delimiting j in (9) considering the following three cases:

[5]For calculation simplicity, we assume that the duration spent in current CTP before t_i is T_{ctrl}, and the duration of previous CFP is T_{cfpm}.

(i) $N_v - k > k$,

$$P_{N_a}(k) = \sum_{i=0}^{k}\sum_{j=0}^{i} P(i,j,k) + \sum_{i=k+1}^{N_v-k}\sum_{j=0}^{k} P(i,j,k) + \sum_{i=N_v-k+1}^{N_v}\sum_{j=i-N_v+k}^{k} P(i,j,k); \tag{10}$$

(ii) $N_v - k < k$,

$$P_{N_a}(k) = \sum_{i=0}^{N_v-k}\sum_{j=0}^{i} P(i,j,k) + \sum_{i=N_v-k+1}^{k}\sum_{j=i-N_v+k}^{i} P(i,j,k) + \sum_{i=k+1}^{N_v}\sum_{j=i-N_v+k}^{k} P(i,j,k); \tag{11}$$

(iii) $N_v - k = k$,

$$P_{N_a}(k) = \sum_{i=0}^{k}\sum_{j=0}^{i} P(i,j,k) + \sum_{i=k+1}^{N_v}\sum_{j=i-N_v+k}^{k} P(i,j,k). \tag{12}$$

Thus, with $P_{N_a}(k)$, the probability mass function (pmf) of the number of scheduled voice bursts, N_s, is given by

$$P_{N_s}(k) = \begin{cases} P_{N_a}(k) & (0 \le k \le N_{sm}-1) \\ \sum_{r=N_{sm}}^{N_v} P_{N_a}(r) & (k = N_{sm}). \end{cases} \tag{13}$$

Finally, $\overline{N_s}$, $\overline{T_{cfp}}$, and $\overline{T_{cp}}$ can be obtained accordingly, based on $P_{N_s}(k)$.

C. A data throughput optimization framework for the DAH-MAC

In a CP, data nodes access the channel according to the T-CSMA/CA contention protocol. Since we evaluate the average aggregate throughput for data nodes in each superframe, $\overline{T_{cp}}$ is used to denote the average duration of a CP[6]. For the T-CSMA/CA, nodes are restricted to transmit packets within each CP. Before any transmission attempts, nodes are required to ensure that the remaining time in current CP is long enough for at least one packet transmission; otherwise, all transmission attempts are suspended until the next CP starts.

According to the illustration inside the CP of each superframe in Fig. 4, the average value of T_v can be calculated by (14), assuming the transmission starting point is uniformly distributed inside $\left[\overline{T_{cp}} - T_s - T_c, \overline{T_{cp}} - T_c\right]$.

$$\overline{T_v} = (1-\tau)^{N_d T_s}\cdot T_c + \left[1-(1-\tau)^{N_d T_s}\right]\cdot\frac{T_c}{2} = \frac{\left[1+(1-\tau)^{N_d T_s}\right]T_c}{2} \tag{14}$$

where τ is the packet transmission probability of a data node with a nonempty transmission buffer at any backoff slot.

A generic time slot in the non-vulnerable period of the CP can be divided into three categories: 1) an idle backoff slot; 2) a complete transmission slot with the duration T_s (a successful transmission duration is assumed the same as a collision duration [36]); and 3) a restricted transmission slot with the duration $\frac{T_c+T_s}{2}$. Thus, the average duration of a generic time slot in the non-vulnerable period is calculated as

$$\sigma = (1-\tau)^{N_d} + \left[1-(1-\tau)^{N_d}\right]T_a \tag{15}$$

where $T_a = \left(\frac{T_s-T_c}{\overline{T_{cp}}-T_c}\cdot\frac{T_s+T_c}{2} + \frac{\overline{T_{cp}}-T_s}{\overline{T_{cp}}-T_c}\cdot T_s\right)$ denotes the average duration of a transmission slot in the non-vulnerable period.

Then, the probability that a generic slot is inside the vulnerable period is given by

$$p_v = \frac{\overline{T_v}}{\frac{\overline{T_{cp}}-\overline{T_v}}{\sigma} + \overline{T_v}}. \tag{16}$$

Thus, the duration of a generic slot including the vulnerable period is derived as

$$\sigma_d = p_v + (1-p_v)\sigma = p_v + (1-p_v)\left[(1-\tau)^{N_d} + \left[1-(1-\tau)^{N_d}\right]T_a\right]. \tag{17}$$

Since packet transmissions or collisions cannot happen in T_v, the packet collision probability for each node at any backoff slot in a traffic saturation case is expressed as

$$p = 1-(1-p_v)(1-\tau)^{N_d-1}. \tag{18}$$

In (18), the transmission probability τ can be approximated, based on renewal reward theory, as a ratio of the average reward received during a renewal cycle over the average length of the renewal cycle [26] [36]. That is,

$$\tau = \frac{E[A]}{E[A]+E[W]} = \frac{\sum_{j=0}^{R_l-1} p^j}{\sum_{j=0}^{R_l-1} p^j + \sum_{j=0}^{R_l-1}\left(\frac{CW_j}{2}\cdot p^j\right)} \tag{19}$$

where $E[A]$ and $E[W]$ denote the average number of transmission attempts and backoff slots experienced, respectively, before a successful packet transmission; R_l is the retransmission limit; and $CW_j = 2^j CW (j = 0, 1, ..., M_b)$ is the contention window size in backoff stage j (CW is the minimum contention window size and M_b is the maximum backoff stage).

Therefore, the aggregate data saturation throughput[7] for the DAH-MAC is expressed as

$$S_d = \frac{N_d T_{pd}\tau(1-p)}{\sigma_d}\cdot\frac{\overline{T_{cp}}}{T_{SF}} = \frac{N_d T_{pd}\tau(1-p_v)(1-\tau)^{N_d-1}}{p_v + (1-p_v)\left[(1-\tau)^{N_d} + \left[1-(1-\tau)^{N_d}\right]T_a\right]}\cdot\frac{\overline{T_{cp}}}{T_{SF}} \tag{20}$$

where T_{pd} is the data packet duration, and $\overline{T_{cp}}$ is a function of N_v.

[6]All time intervals in a CP are normalized to the unit of an idle backoff time slot duration as commonly seen in CSMA/CA based systems.

[7]The throughput in this paper is normalized by the channel capacity.

From (20), we can see that when N_v and N_d are given and other system parameters are set, e.g., according to IEEE 802.11b standard [37], the saturation throughput S_d is a function of τ, and can be evaluated by solving (18) for τ numerically. We rewrite (20) as

$$S_d = \frac{T_{pd} \cdot \frac{\overline{T_{cp}}}{\overline{T_{SF}}}}{\overline{T_{vd}}} \tag{21}$$

where $\overline{T_{vd}} = \frac{p_v+(1-p_v)\left[(1-\tau)^{N_d}+\left[1-(1-\tau)^{N_d}\right]T_a\right]}{N_d\tau(1-p_v)(1-\tau)^{N_d-1}}$. $\overline{T_{vd}}$ is called *average virtual transmission time* [38], which indicates the total average time experienced (including backoff waiting time, collision time and packet transmission time) in each CP to successfully transmit one packet.

Therefore, with different values of N_d, we evaluate the relationship between $\overline{T_{vd}}$ and τ, shown in Fig. 5. It can be seen

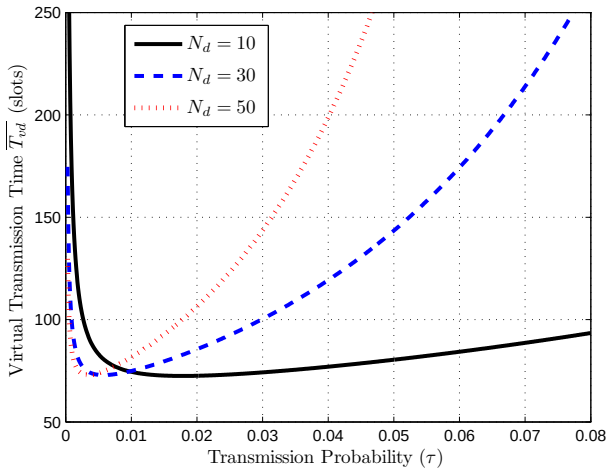


Fig. 5: The evaluation of $\overline{T_{vd}}$ in a function of τ ($\varphi = 0.5$, $N_v = 20$).

that there exists an optimal transmission probability τ_{opt} that achieves a minimum of $\overline{T_{vd}}$. The reason of existing such an optimal transmission probability can be explained as follows: When $\tau < \tau_{opt}$, an increasing amount of channel time remains idle before an transmission initiates, which consistently enlarges $\overline{T_{vd}}$ even if transmission collisions rarely happen in this scenario; However, when τ continues to increase beyond τ_{opt}, due to more frequent transmission attempts, the number of packet collisions rises, which consume an increasing fraction of channel time before packets are successfully transmitted. Therefore, the existence of τ_{opt} can be regarded as a compromise of the preceding two effects and achieves a minimum virtual transmission time and a maximum throughput. Since overheads consumed in one transmission collision are much greater than in an idle backoff slot, the optimal transmission probability is obtained as a relatively small value, as shown in Fig. 5, to lower the collision probability at the expense of consuming more idle slots. Therefore, our objective is to first derive τ_{opt} as a function of N_v and N_d. Then, by substituting τ_{opt} into (14)-(19), a closed-form mathematical relationship can be established between the optimal value of contention window size CW, denoted by CW_{opt}, and the heterogeneous network traffic load.

To do so, the expression of $\overline{T_{vd}}$ in (21) can be further derived as the summation of the following three terms:

$$\overline{T_{vd}} = \frac{p_v}{N_d(1-p_v)\tau(1-\tau)^{N_d-1}} + \frac{1-\tau}{N_d\tau} + \frac{\left[1-(1-\tau)^{N_d}\right]T_a}{N_d\tau(1-\tau)^{N_d-1}}. \tag{22}$$

From (22), it is computational complex to obtain the first order derivative function of $\overline{T_{vd}}$ with respect to τ. The complexity mainly results from p_v which is a complex function of τ in (16). Thus, to make the derivation of $\overline{T_{vd}}$ tractable, an approximation of p_v can be obtained by simplifying $\overline{T_v}$, considering the following two cases:

1) For $\tau \geq 0.005$, since all time durations in a CP are normalized to the unit of an idle backoff slot duration, we have $T_s \gg 1$ (according to the IEEE 802.11b specification) and $N_d T_s \gg 1$. Thus,

$$\overline{T_v} = \frac{\left[1+(1-\tau)^{N_d T_s}\right]T_c}{2} \approx \frac{T_c}{2} \quad (\tau \geq 0.005); \tag{23}$$

2) For $\tau < 0.005$, the average duration of a generic time slot in the non-vulnerable period, σ, approaches 1. Moreover, since $\overline{T_v} \in \left[\frac{T_c}{2}, T_c\right]$, we have $\overline{T_v} \ll \overline{T_{cp}}$. Thus, Eq. (16) can be approximated as

$$p_v = \frac{\sigma\overline{T_v}}{\overline{T_{cp}} + (\sigma-1)\overline{T_v}} \approx \frac{\sigma\overline{T_v}}{\overline{T_{cp}}} \approx \frac{\sigma T_c}{2\overline{T_{cp}}} \quad (\tau < 0.005). \tag{24}$$

As a result, by using $\frac{T_c}{2}$ (the lower bound of $\overline{T_v}$) to approximate $\overline{T_v}$, the approximation of p_v is

$$\widetilde{p_v} = \frac{\frac{T_c}{2}}{\frac{\overline{T_{cp}}-\frac{T_c}{2}}{\sigma} + \frac{T_c}{2}}. \tag{25}$$

Therefore, by substituting (25) into (22) and after some algebraic manipulation, the approximation of $\overline{T_{vd}}$ is obtained as

$$\widetilde{\overline{T_{vd}}} = \frac{\overline{T_{cp}}}{\overline{T_{cp}} - \frac{T_c}{2}} \cdot \frac{T_a - (T_a-1)(1-\tau)^{N_d}}{N_d\tau(1-\tau)^{N_d-1}}. \tag{26}$$

Then, by taking the first order derivative of $\widetilde{\overline{T_{vd}}}$ with respect to τ and letting the derivative function equal to 0, we solve for an approximation of τ_{opt} (under the condition of $\tau \ll 1$) as a closed-form function of N_v and N_d, given by

$$\widetilde{\tau_{opt}} = \frac{\sqrt{1+\frac{2(T_a-1)(N_d-1)}{N_d}}-1}{(T_a-1)(N_d-1)}. \tag{27}$$

Fig. 6 shows the accuracy of the approximation by plotting τ_{opt} and $\widetilde{\tau_{opt}}$ over a wide range of N_d. Note that although the optimal transmission probability, τ_{opt}, for each data node in a backoff slot decreases to a relatively small value with the increase of N_d, the probability of a successful packet transmission in a backoff slot is much higher than τ_{opt} when N_d becomes large, to achieve the maximized throughput.

Then, by substituting $\widetilde{\tau_{opt}}$ and $\widetilde{p_v}$ into (15)-(19), we derive an approximate expression for the optimal contention window

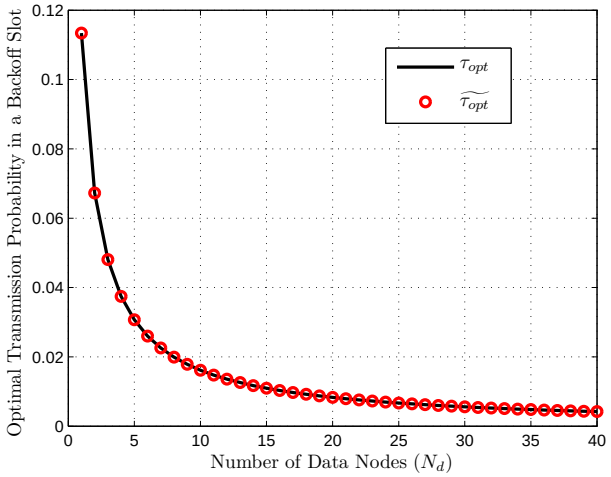


Fig. 6: Optimal transmission probability in each backoff slot for data nodes ($N_v = 20$, $\varphi = 0.5$).

CW_{opt}, as a closed-form function of N_v and N_d, given by

$$\widetilde{CW_{opt}} = \frac{\left(1-\widetilde{\tau_{opt}}\right)\left(1-\widetilde{p}^{R_l}\right)}{\widetilde{\tau_{opt}}\left(1-\widetilde{p}\right)\left(\sum_{j=0}^{M_b} 2^{j-1}\widetilde{p}^{j} + \sum_{j=M_b+1}^{R_l-1} 2^{M_b-1}\widetilde{p}^{j}\right)} \tag{28}$$

where $\widetilde{p} = 1-(1-\widetilde{p_v})(1-\widetilde{\tau_{opt}})^{N_d-1}$.

The proposed analytical framework not only provides an effective way to evaluate the performance of the DAH-MAC in supporting both voice and data traffic, but also provides some insights in MAC design in practical engineering for performance improvement: First, the voice capacity, N_{vm}, and the maximum number of voice bursts, N_{sm}, scheduled for each superframe are derived based on the analytical model and used as a reference for engineers in the protocol design to guarantee the voice delay bound in presence of the voice traffic load dynamics; Second, with the closed-form mathematical relationship provided in (28), data nodes operating T-CSMA/CA in each CP can adaptively adjust the minimum contention window size CW to the optimal value $\widetilde{CW_{opt}}$, based on the updated heterogeneous network traffic load information acquired in each superframe, to achieve consistently maximum aggregate data throughput.

VI. Numerical Results

In this section, simulation results are provided to validate the accuracy of the analytical results. All simulations are carried out using OMNeT++ [39] [40] [41]. Nodes are interconnected and each source node randomly selects one of the rest nodes as its destination node. We run each simulation for 10000 superframe intervals to generate one simulation point. The main simulation settings for different MAC schemes with voice and data traffic are listed in Table II. For a voice source, the GSM 6.10 codec is chosen for encoding the voice stream, with which voice packet payload size is 33 bytes and packets interarrival interval is 20 ms when the voice source node is at the *on* state [14]. Packet arrivals for each best-effort data node follow a Poisson process with the average arrival rate λ_d. We set λ_d to 500 packet/s to ensure each data transmission buffer is always saturated.

We first study the voice capacity, determined by Algorithm 1 in Section V, with a variation of the maximum fraction of time (φ) for voice traffic in each superframe. Then, with a specific φ, the maximum number of voice burst transmissions supported in a CFP to guarantee the packet loss rate bound and the average number of scheduled voice bursts are both evaluated. For performance metrics, the voice packet loss rate and aggregate data throughput are considered. We also compare the performance of the proposed DAH-MAC scheme with two well-known MAC protocols.

A. Voice capacity

The voice capacity in the network, with a variation of φ, for the DAH-MAC is plotted in Fig. 7. The analytical results are obtained according to Algorithm 1 in Section V-A. It can be seen that the analytical results closely match the simulation results especially when the voice capacity region is relatively large, since using the central limit theorem to approximate the distribution of X in (2) becomes more accurate when N_v gets larger.

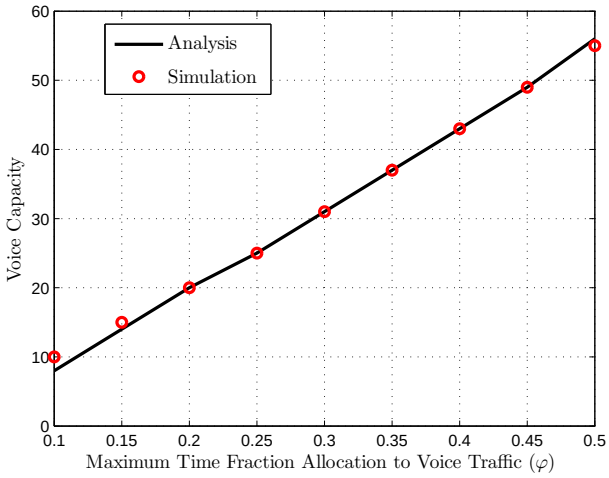


Fig. 7: Voice capacity region with different φ.

B. Number of scheduled voice bursts (time slots) in a CFP

We also evaluate the number of voice burst transmissions (time slots) in a CFP with different N_v in the voice capacity region. Fig. 8 shows the average number of scheduled voice bursts ($\overline{N_s}$) in each superframe. It can be seen that the analytical and simulation results closely match, which verifies the accuracy of our analysis. In Fig. 8, we also plot the maximum supported voice bursts (N_{sm}) in each CFP. Due to the randomness of voice packet arrivals, the instantaneous voice traffic load fluctuates on a per-superframe basis. The gap between N_{sm} and $\overline{N_s}$ indicates the number of time slots allocated to voice bursts in each CFP is commonly below the maximum allowable value. Therefore, by adapting to the realtime voice traffic load in each superframe, the proposed distributed TDMA time slot allocation achieves a high resource utilization.

TABLE II: Simulation parameter settings [14] [37]

Parameters \ MAC schemes	DAH-MAC	Busy-tone contention protocol [14]	D-PRMA [25]
Channel capacity	11Mbps	11Mbps	11Mbps
Backoff slot time	20 μs	20 μs	n.a.
Minimum contention window size (voice/data)	n.a.	8/32	n.a.
Maximum contention window size (voice/data)	n.a.	16/1024	n.a.
Backoff Stage limit (voice/data)	n.a./5	1/5	n.a.
Retransmission limit (voice/data)	n.a./7	2/7	n.a.
PLCP & Preamble	192 μs	192 μs	192 μs
MAC header	24.7 μs	24.7 μs	24.7 μs
RTP/UDP/IP headers (voice)	$\frac{4\cdot 8}{11}$ μs	$\frac{4\cdot 8}{11}$ μs	$\frac{4\cdot 8}{11}$ μs
Packet payload length (voice/data)	$\frac{33\cdot 8}{11}$ / $\frac{8184}{11}$ μs	$\frac{33\cdot 8}{11}$ / $\frac{8184}{11}$ μs	$\frac{33\cdot 8}{11}$ / $\frac{8184}{11}$ μs
AIFS/DIFS (voice/data)	n.a./50 μs	30/50 μs	n.a.
Minislot duration	0.25 ms	n.a.	0.41 ms
Time slot duration	1.22 ms	n.a.	1.64 ms
Transmission time (voice/data)	0.244/1.18 ms	0.244/1.18 ms	0.244/1.18 ms
Gurad time (T_{gt})	20 μs	n.a.	n.a.
Average *on/off* time $\left(\frac{1}{\alpha} \ / \ \frac{1}{\beta}\right)$	352/650 ms	352/650 ms	352/650 ms
Minislot contention probability (voice/data)	n.a.	n.a.	0.6/0.2
Transmission queue length	10000 packets	10000 packets	10000 packets
Superframe time (delay bound)	100 ms	100 ms	100 ms

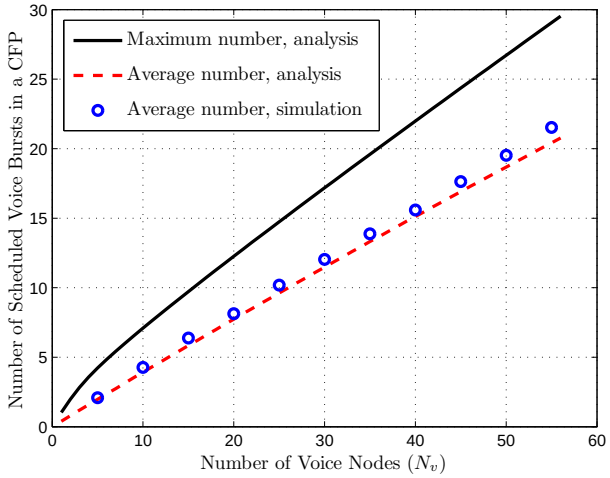


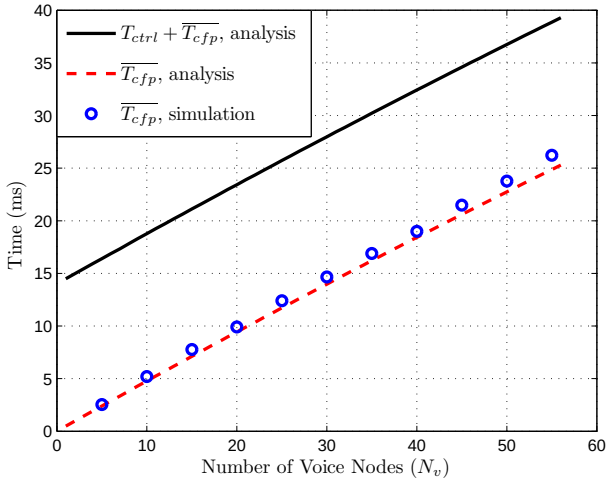
Fig. 8: Number of scheduled time slots for voice traffic ($\varphi = 0.5$).

Fig. 9 shows the average time allocated to voice and data traffic in each superframe with a specific φ. In Fig. 9(a), we can see that the average time of each CFP ($\overline{T_{cfp}}$) for voice burst transmissions increases with N_v, with a fixed duration of each CTP (T_{ctrl}) for accommodating a certain number of voice nodes, which is determined according to the voice capacity with $\varphi = 0.5$; Fig. 9(b) shows the average percentage of time allocated to voice and data traffic in each superframe. It can be seen that within the capacity region, the average time allocated to voice traffic is always bounded by φT_{SF} under the packet loss rate constraint, and the residual average superframe time are occupied by data traffic.

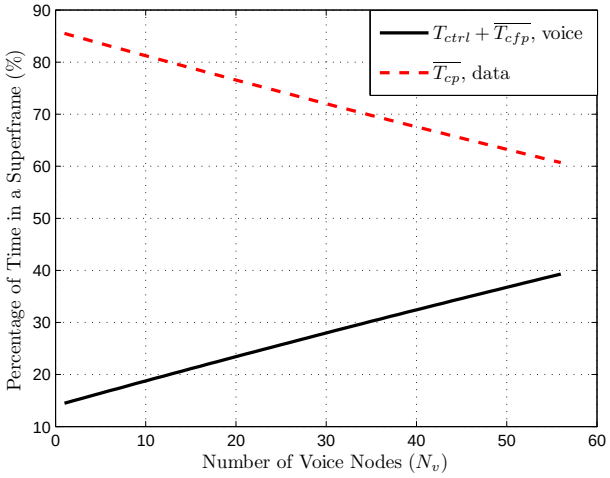
C. Voice packet loss rate

Packet loss rate for voice traffic in a CFP is evaluated with different φ in Fig. 10. It is observed that the simulation results are close to the analytical results. Although some performance fluctuations appear when N_v is relatively small due to the central limit theorem approximation and the rounding-off effect in deriving N_{sm} (set as a simulation parameter), the packet loss rate is always below the analytical bound within the voice capacity region, which verifies the effectiveness of our proposed MAC in supporting voice service. If the number of minislots in the control period of each superframe is set beyond the voice capacity N_{vm}, the packet loss rate rises dramatically as shown in Fig. 10. Therefore, Algorithm 1 in Section V-A is employed in the DAH-MAC to calculate N_{vm} with different requirement of φ, which controls the number of voice nodes N_v within the capacity region to guarantee a bounded packet loss rate.

Fig. 11 displays a comparison of voice packet loss rates between the proposed DAH-MAC and two well-known MAC protocols: D-PRMA protocol [25] and busy-tone contention protocol [14], with a variation of N_v. The latter two MAC protocols are both effective in supporting voice packet transmissions. We can see that the D-PRMA can guarantee a bounded packet loss rate when N_v is relatively small. However, the packet loss rate increases dramatically since contention collisions rise when an increasing number of voice nodes start to contend for the transmission opportunity in each available time slot. Thus, the voice capacity region for D-PRMA is limited. Different from the D-PRMA, the busy-tone contention protocol grants a deterministic channel access priority for voice traffic. Thus, it can be seen that the voice packet loss



(a)



(b)

Fig. 9: Average time allocation in a superframe ($\varphi = 0.5$). (a) Durations of CTP and CFP for voice traffic. (b) Percentage of time for voice and data traffic.

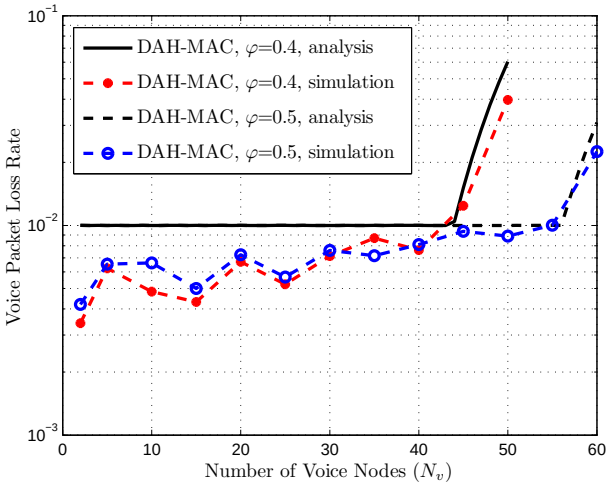


Fig. 10: Voice packet loss rate in a CFP with different φ.

rate is guaranteed over a wide range of N_v. Nevertheless, due to the contention nature, a consistent increase of voice packet collisions lead to accumulated channel access time,

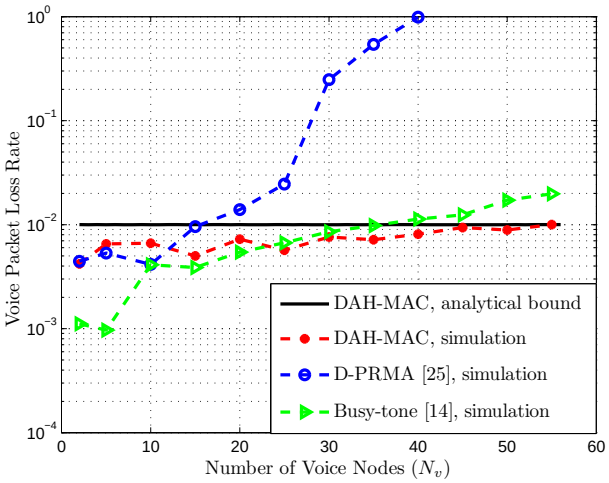


Fig. 11: A comparison of voice packet loss rates ($N_d = 10$, $\varphi = 0.5$).

and the packet loss rate eventually exceeds the bound after around 35 voice nodes are admitted. In the DAH-MAC, the proposed distributed TDMA can admit more voice sessions by setting a higher value of φ, at the expense of local information exchanges in each enlarged control period. As can be seen in Fig. 11, the voice capacity region of the proposed MAC can be larger than the other two MAC protocols with a bounded packet loss rate.

D. Aggregate best-effort data throughput

We first evaluate the average channel utilization of T-CSMA/CA in each CP, defined as the ratio of average time used for successful data packet transmissions in a CP to average duration of the CP. It can be seen in Fig. 12 that the T-CSMA/CA achieves consistently high channel utilization with variations of data traffic load, since the proposed throughput analytical framework maximizes the T-CSMA/CA channel utilization within the DAH-MAC superframe structure.

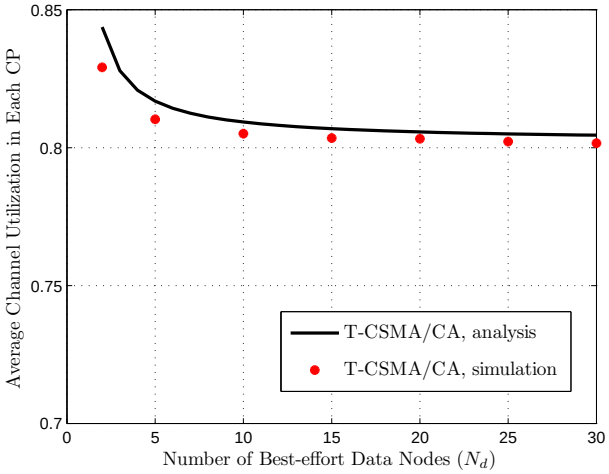
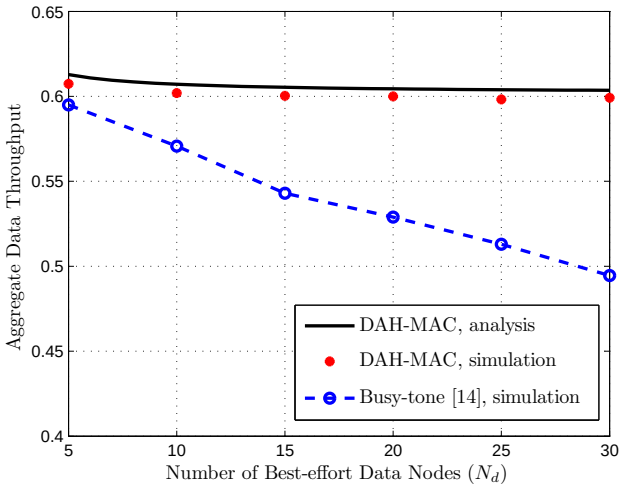


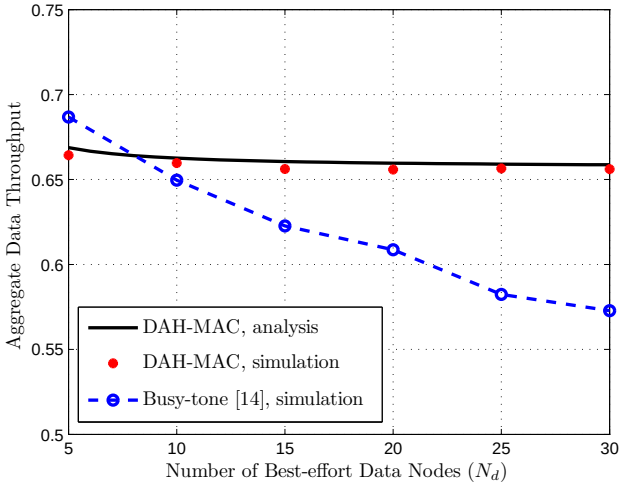
Fig. 12: Channel utilization for data traffic in each CP ($N_v = 20$, $\varphi = 0.5$).

Then, we make a comparison of aggregate data throughput between the proposed DAH-MAC and the busy-tone contention protocol with a variation of N_d and under different

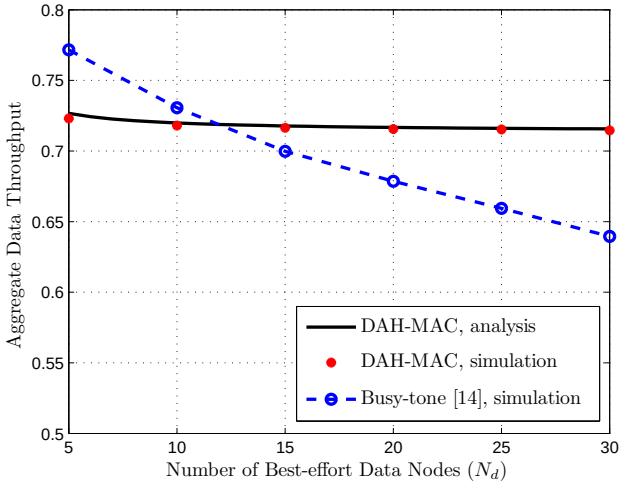
voice traffic load conditions. To ensure a fair comparison, we set φ as 0.33 for DAH-MAC to achieve the same voice capacity ($N_{vm} = 35$) with the busy-tone contention protocol. First, when $N_v = 35$ representing a high voice traffic load condition, it can seen from Fig. 13(a) that the DAH-MAC can achieve a consistently higher data throughput than the busy-tone contention protocol. This is because the distributed TDMA can achieve better resource utilization in high voice traffic load conditions, and the aggregate data throughput is maximized over a wide range of N_d; for the busy-tone contention protocol, with a high voice traffic load, an increasing fraction of channel time is consumed for voice collisions resolution, and a large number of voice nodes also limit the channel access opportunity for data traffic, since voice traffic has absolute priority of accessing the channel. The data throughput comparison is also conducted when the voice traffic load is moderate. It can be seen in Fig. 13(b) that throughputs of the DAH-MAC and the busy-tone contention protocol experience a similar trend with the increase of N_d as in Fig. 13(a), except for the cases where the busy-tone contention protocol achieves a higher throughput when N_d becomes relatively small. The advantage of busy-tone contention becomes more notable when the voice traffic load gets lower, as can be seen in Fig. 13(c) where $N_v = 5$. Therefore, the results from Fig. 13(b) - 13(c) demonstrate some effectiveness of busy-tone based contention when N_v and N_d are relatively small, since in a low heterogeneous network traffic load condition, contention collisions among voice or data nodes are largely reduced, thus improving the channel utilization of the contention-based MAC protocol. Overall, over a wide range of N_d, our proposed MAC can achieve a consistently higher throughput especially in a relatively high voice traffic load condition.



(a)



(b)



(c)

Fig. 13: A comparison of the DAH-MAC maximum data throughput ($\varphi = 0.33$) with the busy-tone contention protocol. (a) $N_v = 35$. (b) $N_v = 20$. (c) $N_v = 5$.

We further conduct the throughput comparison between the DAH-MAC and the D-PRMA. It is shown in Fig. 14 that the DAH-MAC can achieve both a larger voice capacity region ($N_{vm} = 35$ with $\varphi = 0.33$) and a higher data throughput than the D-PRMA. Since the D-PRMA is designed to support the QoS of voice traffic, the throughput for best-effort data traffic is suppressed because data nodes contend the channel with a lower probability and they can only transmit once in a slot upon the successful contention. Also, the D-PRMA uses slotted-Aloha based mechanism to contend for the transmission opportunity, which is inferior to the CSMA/CA

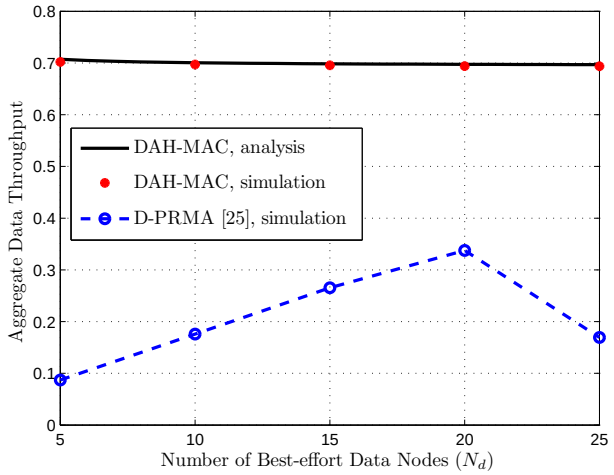


Fig. 14: A comparison of the DAH-MAC maximum data throughput ($\varphi = 0.33$, $N_v = 10$) with the D-PRMA protocol.

based mechanism in terms of collisions resolution. We can see that the data throughput starts to decrease when N_d becomes large due to accumulated contention collisions.

VII. Conclusion

In this paper, we propose a distributed and traffic-adaptive hybrid MAC scheme for a single-hop IoT-enabled MANET. A hybrid MAC superframe structure is devised, in which voice nodes are allocated a time slot in a distributed way by adapting to their instantaneous transmission buffer states and data nodes contend to access the channel in a contention period of each superframe according to the T-CSMA/CA. The proposed hybrid MAC exploits the voice traffic multiplexing to improve the resource utilization while guaranteeing a voice packet loss rate bound, and reduces the congestion level for data nodes by the contention separation between voice and data traffic. A data throughput analytical and optimization framework is developed for the hybrid MAC, in which a closed-form mathematical relationship is established between the MAC layer parameter (i.e., the optimal contention window size) and the number of voice and data nodes in the network. With this framework, the maximum aggregate data throughput can be achieved and be adaptive to variations of the heterogeneous network traffic load. Based on a comparison with two well-known MAC protocols, simulation results demonstrate the effectiveness of our proposed MAC in supporting voice and data services in the presence of heterogeneous traffic load dynamics.

References

[1] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," *Future Gener. Comput. Syst*, vol. 29, no. 7, pp. 1645–1660, 2013.

[2] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surv. Tutor.*, vol. 17, no. 4, pp. 2347–2376, 2015.

[3] H. Nishiyama, T. Ngo, S. Oiyama, and N. Kato, "Relay by smart device: Innovative communications for efficient information sharing among vehicles and pedestrians," *IEEE Veh. Technol. Mag.*, vol. 10, no. 4, pp. 54–62, 2015.

[4] J. Liu and N. Kato, "A markovian analysis for explicit probabilistic stopping-based information propagation in postdisaster ad hoc mobile networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 81–90, 2016.

[5] J.-R. Cha, K.-C. Go, J.-H. Kim, and W.-C. Park, "TDMA-based multi-hop resource reservation protocol for real-time applications in tactical mobile adhoc network," in *Proc. IEEE MILCOM'10*, 2010, pp. 1936–1941.

[6] C. Xu, L. Song, Z. Han, Q. Zhao, X. Wang, X. Cheng, and B. Jiao, "Efficiency resource allocation for device-to-device underlay communication systems: A reverse iterative combinatorial auction based approach," *IEEE J. Select. Areas Commun.*, vol. 31, no. 9, pp. 348–358, 2013.

[7] J. Qiao, X. Shen, J. W. Mark, Q. Shen, Y. He, and L. Lei, "Enabling device-to-device communications in millimeter-wave 5G cellular networks," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 209–215, 2015.

[8] M. Palattella, N. Accettura, X. Vilajosana, T. Watteyne, L. Grieco, G. Boggia, and M. Dohler, "Standardized protocol stack for the Internet of (important) Things," *IEEE Commun. Surv. Tutor.*, vol. 15, no. 3, pp. 1389–1406, 2013.

[9] M. Dong, K. Ota, A. Liu, and M. Guo, "Joint optimization of lifetime and transport delay under reliability constraint wireless sensor networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, no. 1, pp. 225–236, 2016.

[10] M. Park, "IEEE 802.11ah: Sub-1-GHz license-exempt operation for the Internet of Things," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 145–151, 2015.

[11] S. Tozlu, M. Senel, W. Mao, and A. Keshavarzian, "Wi-Fi enabled sensors for Internet of Things: A practical approach," *IEEE Commun. Mag.*, vol. 50, no. 6, pp. 134–143, 2012.

[12] M. Natkaniec, K. Kosek-Szott, S. Szott, and G. Bianchi, "A survey of medium access mechanisms for providing QoS in ad-hoc networks," *IEEE Commun. Surv. Tutor.*, vol. 15, no. 2, pp. 592–620, 2013.

[13] L. Cai, X. Shen, J. W. Mark, L. Cai, and Y. Xiao, "Voice capacity analysis of WLAN with unbalanced traffic," *IEEE Trans. Veh. Technol.*, vol. 55, no. 3, pp. 752–761, 2006.

[14] P. Wang, H. Jiang, and W. Zhuang, "Capacity improvement and analysis for voice/data traffic over WLANs," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1530–1541, 2007.

[15] A. Abdrabou and W. Zhuang, "Stochastic delay guarantees and statistical call admission control for IEEE 802.11 single-hop ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 3972–3981, 2008.

[16] D. Xu, T. Sakurai, H. Vu, and T. Sakurai, "An access delay model for IEEE 802.11e EDCA," *IEEE Trans. Mobile Comput.*, vol. 8, no. 2, pp. 261–275, 2009.

[17] P. Serrano, A. Banchs, P. Patras, and A. Azcorra, "Optimal configuration of 802.11e EDCA for real-time and data traffic," *IEEE Trans. Veh. Technol.*, vol. 59, no. 5, pp. 2511–2528, 2010.

[18] Y. Xiao and Y. Pan, "Differentiation, QoS guarantee, and optimization for real-time traffic over one-hop ad hoc networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 16, no. 6, pp. 538–549, 2005.

[19] A. Veres, A. Campbell, M. Barry, and L.-H. Sun, "Supporting service differentiation in wireless packet networks using distributed control," *IEEE J. Select. Areas Commun.*, vol. 19, no. 10, pp. 2081–2093, 2001.

[20] "IEEE Standard for Information technology–Local and metropolitan area networks–Specific requirements–Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications - Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements," *IEEE Std 802.11e-2005 (Amendment to IEEE Std 802.11, 1999 Edition (Reaff 2003)*, pp. 1–212, 2005.

[21] Y. Xiao, "Enhanced DCF of IEEE 802.11e to support QoS," in *Proc. IEEE WCNC '03*, vol. 2, 2003, pp. 1291–1296.

[22] J. Sobrinho and A. Krishnakumar, "Quality-of-service in ad hoc carrier sense multiple access wireless networks," *IEEE J. Select. Areas Commun.*, vol. 17, no. 8, pp. 1353–1368, 1999.

[23] 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, 2009.

[24] H. Omar, W. Zhuang, and L. Li, "VeMAC: A TDMA-based MAC protocol for reliable broadcast in VANETs," *IEEE Trans. Mobile Comput.*, vol. 12, no. 9, pp. 1724–1736, 2013.

[25] S. Jiang, J. Rao, D. He, X. Ling, and C. C. Ko, "A simple distributed PRMA for MANETs," *IEEE Trans. Veh. Technol.*, vol. 51, no. 2, pp. 293–305, 2002.

[26] R. Zhang, L. Cai, and J. Pan, "Performance analysis of reservation and contention-based hybrid MAC for wireless networks," in *Proc. IEEE ICC'10*, 2010, pp. 1–5.

[27] ——, "Performance study of hybrid MAC using soft reservation for wireless networks," in *Proc. IEEE ICC'11*, 2011, pp. 1–5.

[28] M. Wang, Q. Shen, R. Zhang, H. Liang, and X. Shen, "Vehicle-density-based adaptive MAC for high throughput in drive-thru networks," *IEEE Internet Things J.*, vol. 1, no. 6, pp. 533–543, 2014.

[29] W. Alasmary and W. Zhuang, "Mobility impact in IEEE 802.11p infrastructureless vehicular networks," *Ad Hoc Netw.*, vol. 10, no. 2, pp. 222–230, 2012.

[30] H. Jiang, P. Wang, and W. Zhuang, "A distributed channel access scheme with guaranteed priority and enhanced fairness," *IEEE Trans. Wireless Commun.*, vol. 6, no. 6, pp. 2114–2125, 2007.

[31] L. Lei, S. Cai, C. Luo, W. Cai, and J. Zhou, "A dynamic TDMA-based MAC protocol with QoS guarantees for fully connected ad hoc networks," *Kluwer J. Telecommun. Syst.*, pp. 1–11, 2014.

[32] N. Wilson, R. Ganesh, K. Joseph, and D. Raychaudhuri, "Packet CDMA versus dynamic TDMA for multiple access in an integrated voice/data PCN," *IEEE J. Select. Areas Commun.*, vol. 11, no. 6, pp. 870–884, 1993.

[33] A. Aijaz and A. Aghvami, "Cognitive machine-to-machine communications for Internet-of-Things: A protocol stack perspective," *IEEE Internet Things J.*, vol. 2, no. 2, pp. 103–112, 2015.

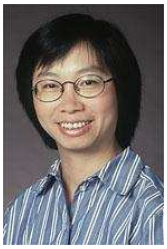
[34] H. Omar, W. Zhuang, A. Abdrabou, and L. Li, "Performance evaluation of VeMAC supporting safety applications in vehicular networks," *IEEE Trans. Emerging Topics in Comput.*, vol. 1, no. 1, pp. 69–83, 2013.

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

[36] X. Ling, K.-H. Liu, Y. Cheng, X. Shen, and J. W. Mark, "A novel performance model for distributed prioritized MAC protocols," in *Proc. IEEE GLOBECOM'07*, 2007, pp. 4692–4696.
[37] "Supplement to 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: Higher-Speed Physical Layer Extension in the 2.4 GHz Band," *IEEE Std 802.11b-1999*, pp. i–90, 2000.
[38] F. Cali, M. Conti, and E. Gregori, "Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit," *IEEE/ACM Trans. on Netw.*, vol. 8, no. 6, pp. 785–799, 2000.
[39] http://www.omnetpp.org/omnetpp.
[40] Q. Ye, W. Zhuang, L. Li, and P. Vigneron, "Traffic load adaptive medium access control for fully-connected mobile ad hoc networks," *IEEE Trans. Veh. Technol.*, 2016, to be published, DOI: 10.1109/TVT.2016.2516910.
[41] J. Ren, Y. Zhang, K. Zhang, A. Liu, J. Chen, and X. Shen, "Lifetime and energy hole evolution analysis in data-gathering wireless sensor networks," *IEEE Trans. Ind. Informat.*, vol. 12, no. 2, pp. 788–800, 2016.

Qiang Ye (S'16) received the B.Sc. and M.Sc. degrees from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2009 and 2012, respectively. He is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada.

His research interests include medium access control and performance optimization in mobile ad hoc networks and Internet of Things.



Weihua Zhuang (M'93-SM'01-F'08) received the B.Sc. and M.Sc. degrees from Dalian Maritime University, Dalian, China, and the Ph.D. degree from the University of New Brunswick, Fredericton, NS, Canada, all in electrical engineering.

Since 1993, she has been with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, where she is currently a Professor and a Tier I Canada Research Chair in wireless communication networks. Her current research interests include resource allocation and quality-of-service provisioning in wireless networks, and smart grid.

Ms. Zhuang is an elected member of the Board of Governors and VP Mobile Radio of the IEEE Vehicular Technology Society. From 2008 to 2011, she was an IEEE Communications Society Distinguished Lecturer. In 2011, she served as the Technical Program Symposia Chair of the IEEE Global Communications Conference. From 2007 to 2013, she served as the Editor-in-Chief for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. She coreceived several Best Paper Awards from IEEE conferences. In 2001, she received the Premiers Research Excellence Award from the Ontario Government and since 2005 has received the Outstanding Performance Award four times from the University of Waterloo. She is a Fellow of the Canadian Academy of Engineering and the Engineering Institute of Canada.