Token-Based Adaptive MAC for a Two-Hop Internet-of-Things Enabled MANET

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

***Abstract*—In this paper, a distributed token-based adaptive medium access control (TA-MAC) scheme is proposed for a two-hop IoT-enabled mobile ad hoc network (MANET). In the TA-MAC, nodes are partitioned into different one-hop node groups, and a time division multiple access (TDMA)-based superframe structure is proposed to allocate different TDMA time durations to different node groups to overcome the hidden terminal problem. A probabilistic token passing scheme is devised to distributedly allocate time slots to nodes in each group for packet transmissions, forming different token rings. The distributed time slot allocation is adaptive to variations of the number of nodes in each token ring due to node movement. To optimize the MAC design, performance analytical models are presented in closed-form functions of both the MAC parameters and the network traffic load. Then, an average end-to-end delay minimization framework is established to derive the optimal MAC parameters under a certain network load condition. Analytical and simulation results demonstrate that, by adapting the MAC parameters to the varying network condition, the TA-MAC achieves consistently minimal average end-to-end delay, bounded delay for local transmissions, and high aggregate throughput. Further, the performance comparison with other MAC schemes shows the scalability of the proposed MAC in an IoT-based two-hop environment with an increasing number of nodes.**

***Index Terms*—IoT, multi-hop MANET, token passing, TDMA, adaptive MAC, scalability, end-to-end delay, aggregate throughput.**

I. Introduction

THE Internet-of-Things (IoT) has great potentials to be one of the most promising network infrastructures towards the next generation wireless network evolution. The IoT framework will interconnect a growing number of heterogeneous objects, i.e., smartphones, sensors and actuators, autonomous devices, via suitable wireless technologies for ubiquitous Internet access and pervasive information sharing [1]–[3]. Within this framework, various IoT-oriented intelligent applications can be realized, e.g., disaster monitoring and response, intelligent control for smart homing, and industrial automation. To support an increasing node number and user demands, an IoT-enabled mobile ad hoc network (MANET) emerges as an important means to provide seamless Internet access for end users[1]. A MANET consists of a group of self-organized nodes, interconnected for communication in a peer-to-peer manner, without any centralized control. Due to low cost and simplified implementation, MANETs are widely deployed for applications such as smart home networking [1], emergency communications and prompt response in postdisaster areas [4]–[6].

For an IoT-enabled MANET, to maintain consistently satisfactory performance in presence of network traffic load variations due to node mobility, an efficient medium access control (MAC) protocol is imperative to coordinate packet transmissions of each node in a distributed way and to adapt to the network traffic load variations [7]. However, the distinctive characteristics of IoT pose new technical challenges on MAC for MANETs: 1) The IoT infrastructure should accommodate an increasing number of users. For example, in disaster-affected areas without conventional communication infrastructures, an increasing number of smart devices from victims can be connected via ad hoc networking to support an abrupt rise of data traffic and communication demands after the catastrophe. Therefore, the MAC protocol should be scalable to the number of nodes to achieve high network throughput and low transmission delay, especially under high network load conditions [8]; 2) The increased number of nodes can enlarge the network coverage area, making the communication distance between a pair of end users beyond the one-hop transmission (communication) range; 3) For a multi-hop network, some nodes staying in the transmission ranges of both source and destination nodes (that are far apart) may relay traffic for the end nodes. Thus, the compound traffic arrival rate (superposition of the external traffic arrival rate and the relay traffic arrival rate) at each relay node can become high, resulting in a large overall delay for relay transmissions and thus for end-to-end transmissions. Therefore, how to maintain a consistently minimized end-to-end packet delay in a multi-hop environment with an increased number of nodes is critical for MAC.

In literature, contention-based carrier sense multiple access with collision avoidance (CSMA/CA) using request-to-send/clear-to-send (RTS/CTS) handshaking schemes, e.g., IEEE 802.11 MAC [8], [9], has been demonstrated not scalable in a high network load condition in a multi-hop environment, due to increased transmission collisions caused by the hidden terminal problem [10] and/or the receiver blocking problem [11], which become worse with an increasing number of nodes. The above problems can be solved by a dual-channel busy-tone based MAC solution [12] at the price of increased protocol complexity and additional circuitry [11]. Time division multiple access (TDMA) protocols [13], [14] perform better for multi-hop transmissions, achieving high channel utilization by eliminating unintentional packet col-

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.

Copyright (c) 2012 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

[1]An end user can also mean an end device in this paper.

lisions due to the hidden terminal problem. In [13], a joint TDMA-based MAC and routing protocol is proposed for packet transmissions in a multi-hop vehicular ad hoc network (VANET), in which every vehicle can acquire a time slot not occupied by any of its two-hop neighbors upon listening to the neighboring information exchange within each frame. Dynamic TDMA time slot assignment (DTSA) is presented in [15] to support a varying number of users in a multi-hop MANET, where the frame length is doubled each time when no time slots are available for newly arriving nodes in current frame. Recently, hybrid MAC protocols, combining C-SMA/CA with TDMA, are re-visited for a multi-hop environment to achieve a performance tradeoff between the two MAC approaches, which can be effective in a low load condition, for example, the unused transmission time slot contention [16] and the CSMA/CA-based time slot scheduling [17]. However, the network scalability is still throttled due to contention collision accumulation in a high load condition. Token-based MAC protocols, as a subset of contention-free protocols, have also gained many research interests for MANETs, due to its quality-of-service (QoS) provisioning capability [18]–[20] and the flexibility in supporting network topology changes [21]. A multi-channel token ring-based MAC protocol is proposed in [22] for supporting both safety and non-safety packet transmissions in a multi-hop VANET, where inter-ring communications are based on the CSMA/CA and token passing is employed for intra-ring data communications. In [23], a dual-channel token-based MAC protocol is proposed for multi-hop MANETs, which have a control channel for token passing and channel reservation, and a data channel for data transmissions. The performance analysis is carried out for a single-hop scenario.

The end-to-end packet delay is an important performance metric to reflect the effectiveness of a MAC protocol in a multi-hop environment. However, most of the existing TDMA and token-based protocols [22], [23] allocate time slots and schedule the token passing without considering the end-to-end delay satisfaction due to the intractability of analytical modeling for the end-to-end delay and its optimization in a multi-hop network. Thus, the end-to-end packet delay can increase to an unacceptable level with an increasing node number, if transmission opportunities are not adaptively allocated. Therefore, adapting the allocation of TDMA time slots or the scheduling of token rotation cycles to the network traffic load variations is of paramount importance to ensure the protocol scalability, with a low end-to-end delay and a high aggregate network throughput. In this paper, we consider a two-hop network as the first step towards a more general multi-hop environment, and propose a token-based adaptive MAC (TA-MAC) scheme. In the TA-MAC, both the number of token rotation cycles and the superframe duration are optimized and adapted to the instantaneous network traffic load, to achieve a consistently minimal average end-to-end packet delay. Our contributions are three-folded:

1) First, to eliminate the hidden terminal problem, a distributed TDMA-based superframe structure is considered for the TA-MAC, in which different one-hop node groups are allocated different TDMA durations. Inspired by [19] [20], each node group forms a token ring by adopting a probabilistic token passing scheme to distributedly allocate time slots to the group members for packet transmissions. Each token ring maintains and updates its node members in a distributed way, and the transmission time slot allocation is adaptive to the instantaneous number of nodes in the network;
2) Second, to determine the MAC parameters for performance optimization, we evaluate the average delay for end-to-end packet transmissions, the average delay for local packet transmissions, and the aggregate network throughput for the TA-MAC in closed-form functions of the MAC protocol parameters and the network traffic load;
3) Third, with a predefined superframe length, an optimization framework is established for minimizing the average end-to-end delay under the constraints of guaranteeing the bounded average delay for local transmissions and maintaining stable transmission queues of each node. The original non-convex minimization problem is then decoupled into a convex subproblem and a biconvex subproblem, which can be solved sequentially to obtain the minimized number of token rotation cycles for each token ring. Then, a distributed computation algorithm is proposed to determine the optimal superframe length and the associated optimal numbers of token rotation cycles for each token ring, with which the minimal average end-to-end delay can be achieved.

The remainder of the paper is organized as follows. The system model is described in Section II. In Section III, the TA-MAC scheme is presented to support packet transmissions from a varying number of mobile nodes in the two-hop MANET. Section IV provides the performance analysis of the proposed MAC scheme, where the average end-to-end delay is derived as a closed-form function of a set of MAC parameters. An end-to-end delay optimization framework is established in Section V to obtain the MAC parameters, with which the minimal average end-to-end delay can be achieved. Extensive analytical and simulation results are presented in Section VI to show the scalability of the proposed MAC scheme in supporting two-hop packet transmissions in a high traffic load condition. Finally, we draw conclusions in Section VII. Main parameters and variables are listed in Table I.

II. System Model

For a multi-hop MANET, the communication distance between a pair of source-destination (S-D) nodes can be larger than the one-hop transmission range. Therefore, some intermediate relay (R) nodes, residing within both transmission ranges of the S-D nodes that are not reachable to each other directly, not only transmit data packets from their own application layers, but may also relay packets between the S-D node pair. Fig. 1(a) illustrates a general MANET, where each dashed circle represents a fully-connected network (a one-hop cluster), i.e., nodes in the network are within the one-hop communication range of each others. Some nodes, denoted by black dots, staying in the overlapping areas of different one-hop clusters can act as relays to forward packets for other

TABLE I: Main parameters and variables

Parameter & Variable	Definition
D_{ac} (D_{bc})	Average delay for packet transmissions from area A (B) to C
D_{ca} (D_{cb})	Average delay for relay packet transmissions from area C to A (B)
D_a (D_b)	Average delay for local packet transmissions within area A (B)
D_{ab} (D_{ba})	Average end-to-end delay for the transmission direction from area A (B) to B (A)
D^*	Minimized average end-to-end delay given a certain superframe length
D^{opt}	Minimal average end-to-end delay
D_{th}	A delay bound for local packet transmissions within area A and area B
k_j	Number of token rotation cycles scheduled for token ring R_j ($j = ac, bc, a, b$)
k_j^*	Optimal number of scheduled token rotation cycles given a certain superframe length
k_j^{opt}	Optimal number of scheduled token rotation cycles
L_j	Number of nodes in token ring R_j ($j = ac, bc, a, b$)
M	Total number of time slots in each superframe
M^{opt}	Optimal total number of time slots for each superframe
N	Total number of nodes in the two-hop network
N_a, N_b, N_c	Number of nodes in network areas A, B, C
R_{ac}, R_{bc}, R_a (R_b)	Four token rings among different node groups
T_{ac}, T_{bc}, T_{ab}	Three TDMA time durations in each superframe
T_f	Superframe length
T_f^{opt}	Optimal superframe length
T_s	Transmission time slot duration
T_1/T_2	Idle duration before data (token) packet / REQUEST packet transmissions
$W_{s,j}$	Packet service time for a node in token ring R_j ($j = ac, bc, a, b$)
λ	External traffic arrival rate at each node
λ_{ac} (λ_{bc})	External traffic arrival rate heading to area C from nodes in area A (B)
λ_{ca} (λ_{cb})	Compound traffic arrival rate at a relay node for a destination in area A (B)
λ_a (λ_b)	External traffic arrival rate at nodes in area A (B) for local transmissions
μ_j	Packet service rate for nodes in R_j ($j = ac, bc, a, b$)
$\mathcal{L}(j)$	Probabilistic token passing list for token ring R_j ($j = ac, bc, a, b$)
$\mathcal{N}(x)$	One-hop neighbor node IDs of node x

S-D node pairs (denoted by white dots) beyond the direct communication range. Therefore, the basic communication unit in a multi-hop MANET is a two-hop network, and some nodes can be R nodes in addition to S-D nodes. With mobility, nodes can leave one two-hop network and become members of another one. In this paper, we consider a basic two-hop network model as the first step towards a general multi-hop environment, shown in Fig. 1(b). There are three logical areas (A, B and C); Nodes enter or depart from the network coverage region, or move around in the three areas. For a given packet transmission direction, such as from left to right, a node can be an S (D) node or an R node depending on whether its location is in area A (B) or C.

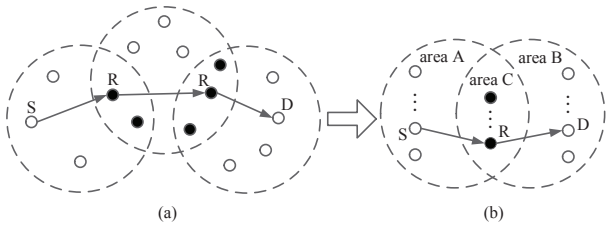


Fig. 1: (a) A general multi-hop MANET. (b) A simplified two-hop network.

Let N denote the total number of nodes in the network which can slowly vary with time due to node mobility, and N_a, N_b and N_c denote the numbers of nodes in areas A, B, and C, respectively. There is a single type of data traffic in the network. The compound traffic arrivals for each R node are different from those of an S (D) node, which consists of not only the traffic arrivals generated from its own application layer but also the relay traffic coming from nodes in both area A and area B [13]. Packet arrivals at each node are split into two traffic streams according to different transmission directions: an arriving packet at each S node in area A (B) is transmitted either to a local D node in the same area or to an end D node two-hop away in area B (A). Similarly, each R node transmits its self-generated packets to a D node in either area A or area B, and also relays packets from area A (B) to a D node in area B (A).

There is a single radio channel in the network, without transmission errors. Nodes access the channel in a distributed manner. We assume that each node is equipped with a global positioning system (GPS) receiver, and the time synchronization among nodes in the network can be achieved by using the 1PPS signal provided by any GPS receiver [24]. Transmission failures can happen due to packet collisions, i.e., more than one transmission attempts are initiated simultaneously by different nodes. Each node has an exclusive node identifier (ID) that can be selected at random and included in each transmitted packet [24]. For a tagged node x, we denote the set of node IDs of all one-hop neighbors of node x (including x itself) as $\mathcal{N}(x)$. In the network, time is partitioned into a sequence

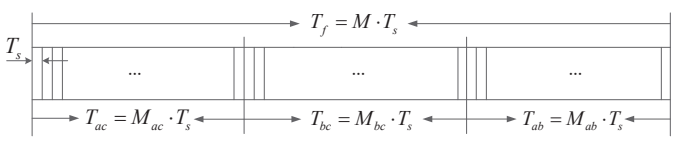


Fig. 2: Superframe structure.

of superframes, and the length of each superframe, denoted by T_f, is determined based on the numbers of nodes in the network areas. As shown in Fig. 2, we partition each superframe into durations, T_{ac}, T_{bc} and T_{ab}, which consist of, M_{ac}, M_{bc} and M_{ab}, numbers of time slots of equal duration T_s, respectively. Therefore, the duration of each superframe, T_f, is equal to $M \cdot T_s$, where M is the total number of time slots within a superframe. Each time slot can accommodate one data packet transmission, and nodes transmit packets at the start of each time slot. To resolve the hidden terminal problem [14], [25] for the two-hop network, durations T_{ac} and T_{bc} are reserved for communications between nodes in areas A and C and between nodes in areas B and C, respectively, where the transmitting and receiving nodes of a communication pair are in different areas, as shown in Fig. 3 (a)-(b); The last duration, T_{ab}, is reserved for simultaneous communications among nodes in area A and among nodes in area B, where both the transmitting and receiving nodes are in the same area, as shown in Fig. 3(c). With this spatial reservation of transmission time slots for the four node groups forming four one-hop subnetworks, packet collisions caused by hidden terminals can be completely eliminated.

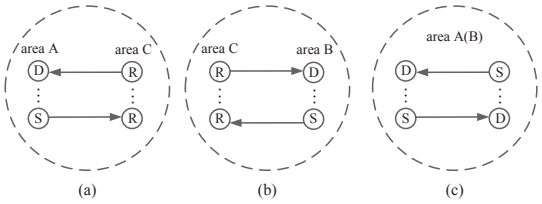


Fig. 3: Packet transmissions in four one-hop subnetworks during (a) T_{ac}. (b) T_{bc}. (c) T_{ab}.

III. The TA-MAC Scheme

A. Probabilistic Token Passing within Each Node Group

In the TA-MAC, there are four tokens circulated separately among nodes in each group for packet transmissions, forming four token rings, R_{ac}, R_{bc} and R_a (R_b). For each token ring, when a node holds a token, it is assigned a time slot with duration T_s for transmission of either a data packet[2] or a token packet [22], [23], and the current token holder decides which node is the next token holder. We define *one token rotation cycle* as the time duration a token has visited all node members once in token ring R_j ($j = ac, bc, a, b$), which equals $L_j \cdot T_s$, where L_j is the number of nodes in token ring R_j. We also define *probabilistic token passing list*, $\mathcal{L}(j)$, as the set of node IDs of all node members in token ring R_j. Each node in R_j records a sequence of node IDs that the token has already visited for current token rotation cycle, and the current token holder selects the next token holder with equal probability from those nodes that have not been visited, to achieve fairness in channel access among all the nodes.

At the beginning of a superframe, a token starts to circulate among nodes in areas A and C during T_{ac}, forming token ring R_{ac}. Once a designated node in area A (C) gets a token, it first waits for the channel to be idle for the duration of T_1 [20], and then piggybacks the token on the head-of-line (HOL) packet (if any) waiting in its queue and transmits the packet to its destination node in area C (A). Note that the destination node (or the next-hop relay node) and the next token holder are not necessarily the same node. If the token holder does not have packets in its queue, it simply passes the token to the next token holder after T_1. When the current token rotation cycle finishes, a new token rotation cycle of token ring R_{ac} starts, conforming to the same token passing rule until the end of T_{ac}. Once the duration T_{ac} elapses, the current token circulation for R_{ac} ceases, and another token starts to circulate among nodes in areas C and B during T_{bc}, forming token ring R_{bc}, which proceeds in the same way as in T_{ac}. The token rings R_a and R_b are formed when two tokens are circulated among nodes in area A and among nodes in area B, respectively, during T_{ab}. These two token rings operate simultaneously and independently forming two disjoint one-hop subnetworks in the areas. Therefore, the durations, T_{ac}, T_{bc} and T_{ab}, can be denoted by k_{ac}, k_{bc} and k_a (k_b) token rotation cycles for token rings, R_{ac}, R_{bc} and R_a (R_b), respectively, as shown in Fig. 4, indicating the number of times a token is held by each node in each token ring for packet transmissions. Note that the duration T_{ab} can be denoted by k_a token rotation cycles of R_a, or k_b token rotation cycles of R_b (not depicted in Fig. 4 for clarity). It is possible that the numbers of time slots, M_{ac}, M_{bc} and M_{ab}, in each duration are not an integer multiple of the numbers of nodes, L_{ac}, L_{bc} and L_a (L_b), in respective token rings, making the numbers of token rotation cycles, k_j ($j = ac, bc, a, b$), a non-integer. In this case, the number of time slots in the last token rotation cycle, denoted by $M_j - (\lceil k_j \rceil - 1)L_j$ ($\lceil \cdot \rceil$ is the ceiling function), is less than L_j. Since nodes are granted a random time slot in each token rotation cycle based on the probabilistic token passing, each node in token ring R_j is statistically guaranteed to hold the token for k_j times in each superframe. To ensure fair channel access among nodes, all the nodes in each token ring at least hold the token once in each duration (i.e., $k_j \geq 1$). Both k_j and M are MAC parameters that affect the performance of the TA-MAC scheme (to be discussed in Subsection III-E).

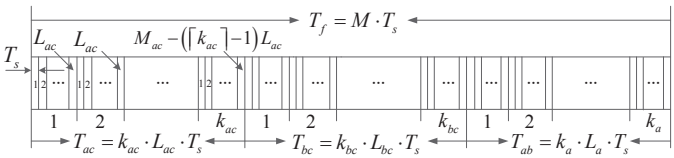


Fig. 4: Token rotation cycles within T_{ac}, T_{bc} and T_{ab}.

Any node, x, in the network, transmits two types of (data/token) packets: Type I and Type II packets. A Type I packet contains a header, a set of IDs of the one-hop neighbors of node x, $\mathcal{N}(x)$, including the probabilistic token passing list, $\mathcal{L}(j)$, for the current token ring R_j, and a payload for either a data packet or a token packet; A Type II packet is composed of a header and a payload. Each node in token ring R_j ($j = ac, bc, a, b$) transmits exactly one Type I packet in the first token rotation cycle to exchange local information with its two-hop neighbors for detecting (updating) the node

[2]To smooth the delay jitter, we assume that each backlogged node transmits one data packet each time that the node holds the token.

location and for distributedly calculating the durations, T_{ac}, T_{bc} and T_{ab}, in each superframe. If more than one token rotation cycles are scheduled for token ring R_j, Type II packets are transmitted in other token rotation cycles.

B. Nodes Joining/Leaving the Network

A node needs to join corresponding token rings for packet transmissions when entering the network. To do so, it first specifies its location in the two-hop network. Suppose a new node, x, is powered on, and synchronizes in time with its one-hop neighbors. Then, it listens to packet transmissions on the channel for one superframe duration, from which it obtains $\mathcal{N}(x)$. Then, the node determines that

1) it is an S (D) node in area A or B, if $\exists\ ID_y \in \mathcal{N}(x) \backslash ID_x$, such that $\mathcal{N}(x) \subset \mathcal{N}(y)$, where ID_x and ID_y denote the IDs of node x and node y;
2) it is an R node in area C, if $\forall\ ID_y \in \mathcal{N}(x) \backslash ID_x$, we have $\mathcal{N}(y) \subseteq \mathcal{N}(x)$, and $\exists\ ID_z \in \mathcal{N}(x) \backslash ID_x$, such that $\mathcal{N}(z) \subset \mathcal{N}(x)$.

Furthermore, if node x is an S-D node and can only receive packets from R nodes in area C during T_{ac} (T_{bc}), it is located in area B (A).

After determining its location, node x broadcasts a REQUEST packet, which is a Type III packet with a higher priority than Types I and II data/token packets, after waiting for the channel to be idle for a duration of T_2 ($< T_1$), to join corresponding token rings. Each REQUEST packet contains a header, the transmitting node ID, and other two important information fields: JOINING and LEAVING, indicating the current network area that the node stays in and the previous area that it departed from. If the node is newly powered on, the LEAVING field is left blank. For instance, when node x is powered on in area A, it broadcasts a REQUEST packet within T_{ac}, after sensing an idle channel for T_2, to join the token rings R_{ac} and R_a, respectively. Upon receiving the REQUEST packet, each one-hop neighbor y of x adds ID_x in the set $\mathcal{N}(y)$ (ID_x is added in the probabilistic token passing list in $\mathcal{N}(y)$). Consequently, if subsequent packet transmissions from any node z in R_{ac} (R_a) have $ID_x \in \mathcal{N}(z)$ and $ID_x \in \mathcal{L}(ac)$ ($\mathcal{L}(a)$), the admissions to corresponding token rings are successful.

On the other hand, when node x is expected to drain its power, it sends a REQUEST packet within T_{ac} before leaving area A, with the LEAVING field specifying area A (the JOINING field is left blank). Then, each one-hop neighbor y of x removes ID_x from $\mathcal{N}(y)$. If node x is the current token holder in R_{ac} (R_a), the token is also passed to the next token holder after the REQUEST packet is transmitted before node departure.

C. Existing Nodes Moving Across Network Areas

When an existing node, x, moves across network areas, its location change can be detected[3]:

1) When moving from area A to C, node x detects its location change within T_{bc} after receiving packets from nodes in area B; Then, it broadcasts a REQUEST packet, with JOINING and LEAVING fields specifying area C and area A, to join R_{bc} and leave R_a. Its ID, ID_x, is added in $\mathcal{L}(bc)$ by each node in R_{bc} and removed from $\mathcal{L}(a)$ by each node in R_a;
2) When moving from area C to A, node x detects its location change within T_{bc} if no packet transmission activity can be detected from nodes in area B. Similarly, a REQUEST packet is broadcast from node x after the location change detection, and ID_x is then added in $\mathcal{L}(a)$ and removed from $\mathcal{L}(bc)$, respectively. For token ring R_{bc}, if the current token holder is a node from area C, it removes ID_x from $\mathcal{L}(bc)$ directly upon receiving the REQUEST packet. Any node, y, in area B also removes ID_x from $\mathcal{N}(y)$ when receiving the updated $\mathcal{L}(bc)$ from the token holder. If the current token holder is a node from area B, it removes ID_x from $\mathcal{L}(bc)$ when it selects node x as the next token holder and no transmission activity is detected within a retransmission timeout (see details in Subsection III-D).

Access collisions happen when two or more nodes, either newly arriving nodes or existing nodes, within the same one-hop transmission range broadcast REQUEST packets at the same time, which can be detected by the nodes involved when their node IDs are not updated in corresponding token passing lists $\mathcal{L}(i)$ received from subsequent packet transmissions. Some random backoff based collision resolution schemes can be used for the nodes involved before re-broadcasting a REQUEST packet [20]. When nodes move at a low speed (e.g., a walking speed), access collisions are unlikely to happen.

D. Lost Token Recovery

Occasionally, existing nodes are not aware of a node departure in the following three situations[4]: 1) The REQUEST packet broadcast by a node being powered off is in collision, and new REQUEST packet cannot be re-initiated due to the power depletion; 2) The current token holder cannot correctly receive the broadcast REQUEST packet from a moving node since the communication range exceeds the one-hop distance; 3) The next token holder departs from the network due to node movement.

When one of the preceding situations happens, the token is lost, which can be detected by the previous token holder as there is no packet transmission from the current token holder. Then, the previous token holder enters into a token recovery process, in which it regenerates and passes a new token to the same current token holder for a maximum of N_{re} times [19], [22]. If there is still no transmission activity discovered after N_{re} is reached, a retransmission timeout is triggered and the previous token holder resends the token to a new node, with the old one removed from the probabilistic token passing list.

E. Important MAC Parameters

The average packet service time for each node is the duration from the instant that a packet arrives at the head

[3]Because of the geographical symmetry of areas A and B, we only consider S (D) nodes moving between areas A and C. Similar results can be obtained when nodes move between areas B and C.

[4]Since nodes move at a low speed, we assume that a token holder can pass the token to the next token holder before moving to a new network area.

of a node queue to the instant it is successfully transmitted, averaged over all transmitted packets from the node. The average end-to-end delay in the two-hop network, denoted by D_{ab} (D_{ba}), for the transmission direction from area A (B) to B (A) is the summation of i) the delay from the time a packet arrives at an S node in area A (B) to the time it is received by an R node, averaged over all transmitted packets for the same transmission direction from area A (B) to C, denoted by D_{ac} (D_{bc}), and ii) the delay from the moment a packet reaches a selected R node to the moment it is received by its D node in area B (A), averaged over all transmitted packets for the same transmission direction from area C to B (A), denoted by D_{cb} (D_{ca}).

To achieve minimal average end-to-end packet delay, each node in token ring R_j ($j = ac, bc, a, b$) should determine the number of token rotation cycles k_j ($j = ac, bc, a, b$) scheduled in T_{ac}, T_{bc} and T_{ab} of each superframe, which also indicates the number of time slots allocated to each node in token ring R_j. A small k_j gives better time slot utilization in a token rotation cycle, i.e., the percentage of nonempty time slots in a token rotation cycle, due to the increased node queue utilization ratios, but can result in a longer transmission queue length for each node. Thus, with an increase of k_j, the packet delay is expected to be reduced due to more resource reservation for high service capability. However, excessive resource reservation for one token ring lowers the delay for single-hop packet transmissions within the token ring, but reduces the channel resources for other token rings, thus increasing the packet delay for other transmission hops. Moreover, an excessive k_j prolongs the length of each superframe, which may cause the increase of average packet service time for each node. Therefore, to minimize the average end-to-end delay, first, time slot allocation for one individual token ring should be balanced with the others; Second, for the purpose of using a minimum total amount of channel resources to achieve the lowest average end-to-end delay, the total number of time slots for each superframe, M, should also be optimized and adaptive to the varying numbers of nodes in each network area. We define k_j^{opt}, M^{opt}, and D^{opt} as the optimal number of token rotation cycles scheduled for token ring R_j ($j = ac, bc, a, b$), the optimal total number of time slots for each superframe, and the minimal average end-to-end delay, respectively. Therefore, we aim at finding the set of MAC parameters k_j^{opt} and M^{opt}, with a varying network load, to achieve D^{opt} under the constraints that each node queue is stable and the average delay for local packet transmissions is bounded. With an increasing number of nodes in each network area, M^{opt} is expected to consistently increase and then remain stable in high network load conditions to maintain a low packet service time, thus making k_j^{opt} decrease continuously. The set of optimal MAC parameters, k_j^{opt} and M^{opt}, are distributedly calculated based on the current traffic load conditions for all the three network areas, and are also dynamically updated upon variations of the numbers of nodes in each area. According to $\left[k_j^{opt}, M^{opt}\right]$, nodes can also determine the optimal durations for T_{ac}, T_{bc} and T_{ab} in each superframe.

IV. Performance Analysis

In this section, we develop performance analytical models for the TA-MAC scheme in closed-form functions of k_j and M for each superframe.

A. Compound Packet Arrival Rate

Traffic arrivals at each node are modeled as a Poisson process with an arrival rate λ packet/slot [20]. Each node in area A (B) transmits a packet either to a local destination node in the same area during T_{ab}, or to a relay node in area C during T_{ac} (T_{bc}). Thus, the traffic arrivals at each node in area A (B) are split into two streams with the average arrival rates denoted by λ_a (λ_b) and λ_{ac} (λ_{bc}) for each transmission direction. For analysis simplicity, we assume that every packet generated from an S node's own application layer is transmitted equally likely for both directions. Thus, λ_a (λ_b) and λ_{ac} (λ_{bc}) are equal to $\frac{\lambda}{2}$ packet/slot. On the other hand, traffic arrivals at each relay node consist of two portions: 1) traffic generated at the node's own application layer and destined for a node either in area A or in area B equally as assumed for an S node, with the average arrival rate $\frac{\lambda}{2}$ packet/slot for both transmission directions; 2) the relay traffic received from nodes in area A (B) during T_{ac} (T_{bc}), which will be forwarded to a destination node chosen in area B (A) during T_{bc} (T_{ac}).

To analyze the average end-to-end delay from a source node in area A (B) to a destination node in area B (A) as well as the aggregate throughput for the two-hop network, a network of queues should be established. However, the exact relay traffic arrival process at each node in area C, consisting of the superposition of the departure processes from the nodes in area A (B), is difficult to be precisely modeled [26]. Therefore, inspired by [13], we approximate the relay traffic arrival process at a relay node as a single Poisson process with rate parameter, λ_{ar} (λ_{br}), being the sum of the traffic arrival rates heading to the common relay node from nodes in area A (B), as shown in Fig. 5. Under the assumption that each source

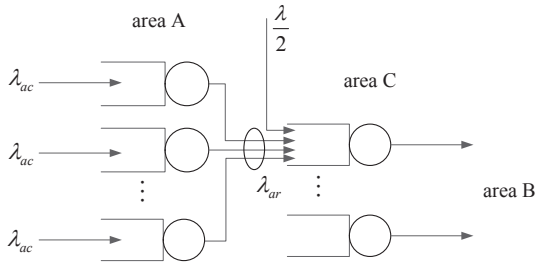


Fig. 5: Poisson approximation for relay traffic arrival rate for transmission direction from area C to B.

node selects its relay from N_c nodes with equal probability, the compound traffic arrival rates at each relay node with the combination of the external traffic and the relay traffic for the transmission directions from area C to B and from area C to A, denoted by λ_{cb} and λ_{ca}, respectively, are approximately given by

$$\lambda_{cb} \approx \frac{\lambda}{2} + \lambda_{ar} = \frac{\lambda}{2} + \frac{\lambda N_a}{2N_c} \qquad (C \to B), \quad (1)$$

$$\lambda_{ca} \approx \frac{\lambda}{2} + \lambda_{br} = \frac{\lambda}{2} + \frac{\lambda N_b}{2N_c} \qquad (C \to A). \quad (2)$$

As an extension to *Kleinrock independence approximation* [26], this Poisson traffic approximation on each relay node is effective in analytically modeling the two-hop network as a network of M/G/1 queues for evaluating the average end-to-end delay. To justify the accuracy of this approximation, the analytical results are further compared with the simulation results in Subsection VI-B. The approximation becomes more accurate when the network traffic load increases [26] (i.e., a large number of nodes in each network area with increased queue utilization ratios for each node), under which the TA-MAC scheme can also achieve high channel utilization.

B. Average Packet Service Time

We calculate average packet service time for head-of-line (HOL) packet transmissions. Packet service time (in the unit of time slot), $W_{s,j}$, for a node in token ring R_j ($j = ac, bc, a, b$), is the duration from the instant that a packet arrives at the head of the node queue to the instant it is successfully transmitted. We use index q ranging from 1 to L_j to indicate the end instant of each time slot in a token rotation cycle of R_j. Take nodes in token ring R_{ac} as an example, in which there are k_{ac} token rotation cycles scheduled within each T_{ac}. HOL packets from a tagged node x in area A and area C can appear at the end of each allocated time slot in any of k_{ac} token rotation cycles along each T_{ac}. For analysis tractability, we neglect the possibility that an HOL packet arrives within the duration of a time slot, and the possibility that a packet arrives at a node in R_{ac} and finds the node queue empty during T_{bc} and T_{ab} [27], which is more unlikely for a higher traffic load condition. We make the same assumption for HOL packet arrivals at nodes in all token rings. Define a random variable H_j, which equals 0 if an HOL packet from a tagged node in R_j appears during the k_jth token rotation cycle, and equals 1 otherwise. As each node has a randomly selected transmission slot in the corresponding token passing sequence of each token rotation cycle, the service times for consecutive packet transmissions are independent and identically distributed (i.i.d) random variables [13], which is a necessary condition for the M/G/1 queue modeling for each node. Thus, the HOL packet is transmitted in a time slot randomly chosen from the next token rotation cycle following the packet arriving instant. Therefore, we derive the average packet service time, denoted by $E[W_{s,j}]$, by considering the following two cases:
(1) When $H_j = 1$, the probability mass function (pmf) of $W_{s,j}$ conditioned on the arriving time instant Q of the HOL packet is expressed as

$$\begin{aligned} &P\{W_{s,j} = i | Q = q, H_j = 1\} \\ &= \frac{1}{L_j} \quad (L_j - q + 1 \le i \le 2L_j - q,\ 1 \le q \le L_j), \end{aligned} \quad (3)$$

where $L_{ac} = N_a + N_c$, $L_{bc} = N_b + N_c$, $L_a = N_a$, $L_b = N_b$, and the expectation of $W_{s,j}$ conditioned on $H_j = 1$ can be derived as

$$\begin{aligned} E[W_{s,j} | H_j = 1] &= \sum_{i=L_j-q+1}^{2L_j-q} \sum_{q=1}^{L_j} i P\{W_{s,j} = i | Q = q, H_j = 1\} \\ &\quad \cdot P\{Q = q | H_j = 1\} \\ &= L_j; \end{aligned} \quad (4)$$

(2) When $H_j = 0$, we similarly have the conditional pmf and the conditional expectation of $W_{s,j}$ as

$$\begin{aligned} &P\{W_{s,j} = i | Q = q, H_j = 0\} \\ &= \frac{1}{L_j} \quad (U_j - q + 1 \le i \le U_j - q + L_j,\ 1 \le q \le L_j) \end{aligned} \quad (5)$$

where $U_j = M - (k_j - 1)L_j$, and

$$\begin{aligned} E[W_{s,j} | H_j = 0] &= \sum_{i=U_j-q+1}^{U_j-q+L_j} \sum_{q=1}^{L_j} i P\{W_{s,j} = i | Q = q, H_j = 0\} \\ &\quad \cdot P\{Q = q | H_j = 0\} \\ &= U_j. \end{aligned} \quad (6)$$

Therefore, the average service time for each HOL packet from nodes in R_j can be derived as

$$\begin{aligned} E[W_{s,j}] &= E[W_{s,j} | H_j = 1] P\{H_j = 1\} \\ &\quad + E[W_{s,j} | H_j = 0] P\{H_j = 0\} \\ &= L_j \cdot \left(1 - \frac{1}{k_j}\right) + U_j \cdot \frac{1}{k_j} = \frac{M}{k_j}, \quad j = ac, bc, a, b. \end{aligned} \quad (7)$$

Define the average packet service rate, μ_j, as the number of packets transmitted per slot time from a node in R_j. Thus,

$$\mu_j = \frac{1}{E[W_{s,j}]} = \frac{k_j}{M}, \quad j = ac, bc, a, b. \quad (8)$$

C. Aggregate Network Throughput

The aggregate network throughput is defined as the ratio of average number of transmitted packets over total number of time slots in each superframe, which is also the aggregate time slot utilization for each superframe, given by

$$\begin{aligned} S = \frac{1}{M} \Bigg(& k_{ac} \frac{N_a \lambda_{ac} + N_c \lambda_{ca}}{\mu_{ac}} + k_{bc} \frac{N_b \lambda_{bc} + N_c \lambda_{cb}}{\mu_{bc}} \\ & + \sum_{j \in \{a,b\}} k_j \frac{N_j \lambda_j}{\mu_j} \Bigg). \end{aligned} \quad (9)$$

From (7) and (9), the aggregate network throughput, S, is a function of the numbers of nodes in each network area and the traffic arrival rates at the nodes. Actually, given certain numbers of nodes in the network areas, the variations of k_j and M affect the average packet service rate μ_j for each node in R_j and the time slot utilization for each token rotation cycle, resulting in a different packet delay. But the aggregate channel utilization for an entire superframe remains unchanged with variations of k_j and M. Therefore, higher aggregate network throughput is expected to be achieved with more nodes in all individual network areas.

D. Average End-to-End Delay

As defined in Subsection III-E, the average end-to-end delay in the two-hop network, D_{ab} (D_{ba}), for the transmission direction from area A (B) to B (A) is the summation of average packet delays D_{ac} (D_{bc}) and D_{cb} (D_{ca}). The average packet delay is composed of the average queueing delay, i.e., the average duration the packet stays in the transmission queue after its arrival, and the average service time.

We derive the second moment of the packet service time $W_{s,j}$ for each node in token ring R_j ($j = ac, bc, a, b$) as

$$\begin{aligned} E[W_{s,j}^2] = E[W_{s,j}^2|H_j = 1]P\{H_j = 1\} \\ + E[W_{s,j}^2|H_j = 0]P\{H_j = 0\}. \end{aligned} \tag{10}$$

Then, based on *P-K formula* [26], the average end-to-end delay, D_{ab} (D_{ba}), for either transmission direction, and the average delay for local transmissions within area A (B), denoted by D_a (D_b), can be derived as

$$\begin{aligned} D_{ab} &= D_{ac} + D_{cb} \\ &= \sum_{(n,j)\in\{(ac,ac),(cb,bc)\}} \left(E[W_{s,j}] + \frac{\lambda_n E[W_{s,j}^2]}{2\left(1 - \lambda_n E[W_{s,j}]\right)} \right) \\ &= \sum_{(n,j)\in\{(ac,ac),(cb,bc)\}} \left(\frac{\varepsilon_j}{k_j} + \frac{\lambda_n[\alpha_j k_j^2 + \beta_j k_j + \gamma_j]}{2(k_j - \lambda_n \varepsilon_j)} \right), \end{aligned} \tag{11}$$

$$\begin{aligned} D_{ba} &= D_{bc} + D_{ca} \\ &= \sum_{(n,j)\in\{(bc,bc),(ca,ac)\}} \left(\frac{\varepsilon_j}{k_j} + \frac{\lambda_n[\alpha_j k_j^2 + \beta_j k_j + \gamma_j]}{2(k_j - \lambda_n \varepsilon_j)} \right), \end{aligned} \tag{12}$$

and

$$D_j = \frac{\varepsilon_j}{k_j} + \frac{\lambda_j[\alpha_j k_j^2 + \beta_j k_j + \gamma_j]}{2(k_j - \lambda_j \varepsilon_j)} \quad (j = a, b) \tag{13}$$

where $\alpha_j = L_j^2$, $\beta_j = -\frac{5L_j^2 + 12ML_j + 1}{6}$, $\gamma_j = M^2 + 2ML_j$, and $\varepsilon_j = M$.

From (11) to (13), it is observed that, with certain numbers of nodes in each network area, both the average end-to-end delay and average delay for local transmissions are functions of k_j and M. As stated in Subsection III-E, with a certain M value, an increased k_j for one token ring reduces the one-hop average packet delay among its node members, and also the time resources for other token rings, which can increase the average delay for other transmission hops. Thus, the numbers of token rotation cycles scheduled for each token ring should be balanced to minimize the average end-to-end delay. At the same time, the total number of time slots, M, for each superframe should be properly chosen to further improve the delay performance. Therefore, our design objective is to determine k_j and M to achieve the minimal average end-to-end delay.

V. Optimal MAC Parameters

In this section, we propose an optimization framework to find the set of optimal MAC parameters, $\left[k_j^{opt}, M^{opt}\right]$, for each superframe, with which the minimal average end-to-end delay, D^{opt}, is achieved, and a bounded average delay, D_{th}, for local transmissions is guaranteed.

A. Average End-to-End Delay Minimization

An average end-to-end delay minimization problem is formulated as (P1), to derive the set of optimal numbers of token rotation cycles, $\mathbf{k}^* = [k_{ac}^*, k_{bc}^*, k_a^*, k_b^*]$, for each token ring, for a given M.

$$(\mathbf{P1}): \min_{\mathbf{k}=[k_{ac},k_{bc},k_a,k_b]} \{\max\{D_{ab}(k_{ac}, k_{bc}), D_{ba}(k_{ac}, k_{bc})\}\}$$

$$\text{s.t.} \begin{cases} k_{ac}L_{ac} + k_{bc}L_{bc} + k_a L_a = M & \text{(14a)} \\ k_a L_a = k_b L_b & \text{(14b)} \\ \rho_n = \dfrac{\lambda_n}{\mu_j} < 1 \quad (n,j) \in \{(ca,ac),(cb,bc)\} & \text{(14c)} \\ \rho_j = \dfrac{\lambda_j}{\mu_j} < 1 \quad (j = a, b) & \text{(14d)} \\ D_j(k_j) \le D_{th} \quad (j = a, b) & \text{(14e)} \\ k_j \ge 1 \quad (j = ac, bc, a, b). & \text{(14f)} \end{cases}$$

In (P1), the objective is to minimize the average end-to-end delay for both transmission directions from area A to B and from area B to A, by finding the set of optimal numbers of token rotation cycles for each token ring in each superframe. Constraints (14a) - (14b) indicate the total number of time slots for each superframe is M, and the time slots allocated to nodes in area A and area B are balanced based on the numbers of nodes in both areas. Constraint (14c) guarantees a stable relay node queue in token rings R_{ac} and R_{bc}, where $\rho_n = \frac{\lambda_n}{\mu_j}$ denotes queue utilization ratio for relay nodes in area C[5]. Similarly, ρ_j in constraint (14d) denotes queue utilization ratio for nodes in token rings R_a and R_b. Constraint (14e) states that the average delays for local packet transmissions within both area A and area B are bounded by threshold D_{th}. Constraint (14f) guarantees that nodes in each token ring hold the token at least once in each superframe to ensure fair channel access.

The main difficulty to solve (P1) is the non-convexity of the objective function in terms of $\mathbf{k}$. Therefore, to make the problem tractable, we decouple (P1) into a convex subproblem and a biconvex subproblem [28], [29] with two separate sets of decision variables. By solving these two subproblems sequentially, the original problem can be solved. First, we introduce an important proposition and its corollary.

Proposition 1. *In (P1), the one-hop average delays, D_{ac} (D_{ca}), D_{bc} (D_{cb}), and D_a (D_b), for packet transmissions from area A (C) to area C (A), from area B (C) to area C (B), and within the local area A (B), are all strictly convex functions in terms of k_{ac}, k_{bc}, and k_a (k_b), respectively.*

The proof of Proposition 1 is given in Appendix A.

Corollary 1. *In (P1), the one-hop average delays, D_{ac} (D_{ca}), D_{bc} (D_{cb}), and D_a (D_b), are strictly decreasing functions of k_{ac}, k_{bc}, and k_a (k_b), respectively.*

[5]Since traffic arrival rates at each relay node are greater than that at each source node, queue stability of source nodes in R_{ac} and R_{bc} is also guaranteed by constraint (14c).

The proof of Corollary 1 is given in Appendix B.

According to Corollary 1, each one-hop average delay is a decreasing function of k_j for corresponding token ring. Thus, to minimize the average end-to-end delay, the maximum amount of time slots are expected to be allocated to nodes in token rings R_{ac} and R_{bc} among all feasible solutions for (P1). In other words, the minimum number of time slots should be reserved for the token rings R_a and R_b under constraints (14b) - (14f). Therefore, to tackle (P1) efficiently, we first solve the following subproblem (SP1) to obtain the set of optimal numbers of token rotation cycles, $\mathbf{k_1^*} = [k_a^*, k_b^*]$, scheduled for token rings R_a and R_b, respectively.

$$(\mathbf{SP1}): \min_{\mathbf{k_1}=[k_a,k_b]} k_a L_a$$

$$\text{s.t.} \begin{cases} k_a L_a = k_b L_b & (15a) \\ \rho_j < 1 \quad (j=a,b) & (15b) \\ D_j(k_j) \le D_{th} \quad (j=a,b) & (15c) \\ k_j \ge 1 \quad (j=a,b). & (15d) \end{cases}$$

We further simplify (SP1) as (SP1a) with a single decision variable k_a, by substituting constraint (15a) into constraints (15b) - (15d).

$$(\mathbf{SP1a}): \min_{k_a} k_a L_a$$

$$\text{s.t.} \begin{cases} \rho_a < 1 & (16a) \\ \rho_b \left(\frac{k_a L_a}{L_b} \right) < 1 & (16b) \\ D_a(k_a) \le D_{th} & (16c) \\ D_b \left(\frac{k_a L_a}{L_b} \right) \le D_{th} & (16d) \\ \frac{\max\left\{1, \frac{L_b}{L_a}\right\}}{k_a} \le 1. & (16e) \end{cases}$$

In (SP1a), the objective function and the left-hand sides of all inequality constraints (16a) - (16c) and (16e) are convex functions of k_a. For constraint (16d), the function $D_b\left(\frac{k_a L_a}{L_b}\right)$ is a composite function of k_a, where $D_b(\cdot)$ is a convex and decreasing function of k_b and $k_b = \frac{k_a L_a}{L_b}$ is a linear function of k_a. Therefore, according to the *scalar composition rules* [30], $D_b\left(\frac{k_a L_a}{L_b}\right)$ is also a convex function of k_a. Hence, (SP1a) is proved to be a convex optimization problem, which can be efficiently solved to get the optimal solution $k_a^\dagger$. Since the total number of time slots, M_{ab}, reserved for R_a or R_b is required to be integer, we obtain the optimal numbers of token rotation cycles for each of the token rings as

$$k_a^* = \frac{\lceil k_a^\dagger L_a \rceil}{L_a} \tag{17}$$

and

$$k_b^* = \frac{\lceil k_a^\dagger L_a \rceil}{L_b}. \tag{18}$$

Note that k_a^* and k_b^* are guaranteed the global optimal in the feasible set of (SP1a), since all the inequality constraint functions of (SP1a) are decreasing functions of the decision variable k_a.

By substituting the optimal set of values $[k_a^*, k_b^*]$ into (P1), the original optimization problem is reduced to the second subproblem (SP2).

$$(\mathbf{SP2}): \min_{\mathbf{k_2}=[k_{ac},k_{bc}]} \{\max\{D_{ab}(k_{ac}, k_{bc}), D_{ba}(k_{ac}, k_{bc})\}\}$$

$$\text{s.t.} \begin{cases} k_{ac} L_{ac} + k_{bc} L_{bc} = M^* & (19a) \\ \rho_n < 1 \quad (n = ca, cb) & (19b) \\ k_j \ge 1 \quad (j = ac, bc) & (19c) \end{cases}$$

where $M^* = M - k_a^* L_a$.

Theorem 1. *In (SP2), the two-dimensional decision vector $\mathbf{k_2}$ represents a biconvex set, if $\mathbf{k_2}$ is a convex set with respect to k_{bc} (k_{ac}) for any given k_{ac} (k_{bc}) from the feasible solutions.*

Proof: Rewrite the set of constraints of (SP2) in a standard form for a given k_{ac} (k_{bc}) in $\mathbf{k_2}$, the equality constraint is an affine function of k_{bc} (k_{ac}), and both inequality constraints are convex functions of k_{bc} (k_{ac}). Therefore, the set of feasible solutions of k_{bc} (k_{ac}) satisfying all the constraints form a convex set [30].

Theorem 2. *In (SP2), the objective function defined on the biconvex set $\mathbf{k_2}$ represents a biconvex function, if the objective function is a convex function in terms of k_{bc} (k_{ac}) for any given k_{ac} (k_{bc}) from the feasible solutions.*

Proof: Given k_{ac} (k_{bc}) in $\mathbf{k_2}$, both $D_{ab}(\cdot)$ and $D_{ba}(\cdot)$ functions are a linear combination of a convex function in terms of k_{bc} (k_{ac}) and a constant, which are also convex. Moreover, the *max* function, $\max\{x, y\}$, proved to be convex on $\mathbf{R}^2$ in [30], is also nondecreasing in each of its two arguments. Therefore, according to the *vector composition rules* [30], the objective function $\max\{D_{ab}, D_{ba}\}$ is a convex function with respect to k_{bc} (k_{ac}).

Based on Theorem 1 and Theorem 2, (SP2) is a biconvex optimization problem since we have a biconvex objective function minimized over a biconvex set, which often has multiple local optima and is difficult to determine the global optimal solution [31]. Therefore, to solve (SP2) efficiently, we further simplify (SP2) into a single variable optimization problem (SP2a), by substituting equality constraint (19a) into the objective function and other constraints.

$$(\mathbf{SP2a}): \min_{k_{ac}} \{\max\{D_{ab}(k_{ac}, h(k_{ac})), D_{ba}(k_{ac}, h(k_{ac}))\}\}$$

$$\text{s.t.} \begin{cases} \rho_{ca} < 1 & (20a) \\ \rho_{cb}(h(k_{ac})) < 1 & (20b) \\ 1 \le k_{ac} \le \frac{M^* - L_{bc}}{L_{ac}} & (20c) \end{cases}$$

where $h(k_{ac}) = k_{bc} = \frac{M^* - k_{ac} L_{ac}}{L_{bc}}$.

Proposition 2. *(SP2a) is a convex optimization problem, with respect to the decision variable k_{ac}.*

The proof of Proposition 2 is given in Appendix C.

Based on Proposition 2, the convex optimization problem (SP2a) can also be efficiently solved for the optimal solutions $k_{ac}^\dagger$ and $k_{bc}^\dagger$. Similar to (17) and (18), we further obtain the optimal numbers of token rotation cycles, k_{ac}^* and k_{bc}^*, for

R_{ac} and R_{bc} respectively, by rounding $k_{ac}^{\dagger}L_{ac}$ to the integer number (within the feasible region) that achieves the minimum value of the objective function in (SP2a).

Based on the optimal numbers of token rotation cycles, $\mathbf{k}^* = [k_{ac}^*, k_{bc}^*, k_a^*, k_b^*]$, in each superframe under a predefined M, the average end-to-end packet delay is minimized, denoted as D^*.

B. Optimal Total Number of Time Slots for Each Superframe

As discussed in Subsection III-E, D^* decreases with an increase of M, due to more time slots reserved for each token ring. If M is set too large, excessive resource reservation prolongs the superframe length, causing an increase of packet delay. Therefore, we aim at determining the optimal total number of time slots, M^{opt}, for each superframe, given the optimal number of token rotation cycles, k_j^{opt}, for token ring R_j ($j = ac, bc, a, b$), to minimize average end-to-end delay. To this end, we propose an optimal superframe length calculation algorithm, Algorithm 1, for each node to distributedly determine and update the set of optimal MAC parameters, $\left[k_j^{opt}, M^{opt}\right]$. The procedure of the algorithm is summarized as follows:

Algorithm 1: The optimal superframe length calculation algorithm

Input : L_{ac}, L_{bc}, L_a, L_b, λ, D_{th}.
Output: k_j^{opt}, D^{opt}, M^{opt}, and T_f^{opt}.

Initialization: $M \leftarrow L_{ac} + L_{bc} + \max\{L_a, L_b\}$, $D^{opt} \leftarrow +\infty$, $M_s \leftarrow 0$, set the maximum iteration limit;
do
  $[k_j^*, D^*] \leftarrow$ solving (SP1a) and (SP2a);
  if *No feasible solutions are found* **then**
    if *The maximum iteration limit is reached* **then**
      break;
    end
    $M \leftarrow M + 1$;
  else
    $M_s \leftarrow M$;
    break;
  end
while *(P1) is not feasible*;
if $M_s > 0$ **then**
  while *The maximum iteration limit is not reached* **do**
    if $D^* < D^{opt}$ **then**
      $k_j^{opt} \leftarrow k_j^*$;
      $D^{opt} \leftarrow D^*$;
      $M^{opt} \leftarrow M$;
    end
    $M \leftarrow M + 1$;
    $[k_j^*, D^*] \leftarrow$ solving (SP1a) and (SP2a);
  end
  $T_f^{opt} \leftarrow M^{opt} \cdot T_s$;
  return k_j^{opt}, M^{opt}, T_f^{opt}, and D^{opt}.
end

Step 1. The minimum value for M is set to satisfy constraint (14f) in (P1). Both average end-to-end delay and M_s (the minimum value of M to make (P1) feasible) are initialized, and the maximum iteration limit is set to a large number;
Step 2. The sequential subproblems (SP1a) and (SP2a) are repeatedly solved by increasing M with the increment of one time slot in each iteration until a set of feasible solutions, $[k_j^*, D^*]$, for (P1) are found at $M = M_s$ (If no feasible M_s is found at the maximum iteration limit under current network load condition, the newly arriving node will not be admitted so that the numbers of nodes in each network area can be controlled within the network capacity);
Step 3. Starting from a feasible M_s, we iteratively search for M^{opt} and k_j^{opt} to achieve the minimal average end-to-end delay D^{opt}, by continuously increasing M and solving (SP1a) and (SP2a) at each updated M until the maximum iteration limit is reached.

VI. Numerical Results

In this section, analytical and simulation results are presented to demonstrate the accuracy of performance analysis. All the simulations are carried out using OMNeT++ [32], [33]. In the simulation, nodes are scattered over a 150m $\times$ 150m square region, forming a two-hop network with three network areas similar to that shown in Fig. 1(b), where nodes within the transmission range (set to 50m) of all other nodes can relay traffic from the S-D node pairs that are not reachable to each other directly. For each transmitted packet, the source node randomly selects a next-hop node or a destination node, according to the packet's destination area, among a specific group of nodes. External traffic arrivals for each node are modeled as a Poisson process with the rate of 0.01 packet/slot (10 packet/s). The delay bound for local packet transmissions within area A and area B is set as 400 ms. Each simulation point is generated by running the simulation for 10000 superframes. The main simulation parameters are summarized in Table II.

By solving the sequential subproblems (SP1a) and (SP2a) using Algorithm 1, we determine the optimal total number of time slots, M^{opt}, for each superframe, and the optimal number of token rotation cycles, k_j^{opt}, scheduled for token ring R_j ($j = ac, bc, a, b$), upon which the minimal average end-to-end delay D^{opt}, the bounded average delays for local transmissions, D_a and D_b, in area A and area B, can be achieved by the TA-MAC scheme. Then, we analyze D^{opt}, D_a and D_b, and the aggregate network throughput with respect to variations of the network traffic load. Lastly, the TA-MAC scheme is compared with a hybrid MAC scheme and a dynamic TDMA scheme in terms of delay and throughput over a wide range of network traffic load.

A. Optimal MAC Parameters

Fig. 6 shows the optimal number of token rotation cycles, k_j^*, scheduled for token ring R_j ($j = ac, bc, a, b$) versus M. We can see that k_j^* increases with M, and more token rotation cycles are scheduled for token rings R_{ac} and R_{bc} to minimize the average end-to-end delay. The set of optimal MAC parameters, $\left[M^{opt}, k_j^{opt}\right]$, is also shown, based on Algorithm 1, which achieves the minimal average end-to-end delay.

TABLE II: Simulation parameters

Parameters \ MAC schemes	TA-MAC	LA-MAC [16]	DTSA [15] [16]
Channel capacity	11Mbps	11Mbps	11Mbps
Time slot duration	1 ms	1 ms	1 ms
Idle duration (T_1/T_2)	20/10 μs	n.a.	n.a.
Node ID	7 bit	7 bit	7 bit
Data/Token packet duration	0.98/0.28 ms	0.98 ms/n.a.	0.98 ms/n.a.
REQUEST packet duration	0.22 ms	n.a.	n.a.
Packet arrival rate (λ)	0.01 packet/slot	0.01 packet/slot	0.01 packet/slot
Transmission queue length	10000 packets	10000 packets	10000 packets
Delay bound for local transmissions (D_{th})	400 ms	n.a.	n.a.

Fig. 7 demonstrates the minimal average end-to-end delay, D^*, and the average delays for local packet transmissions, D_a and D_b, versus M. It can be seen that D^* decreases with M when M is small, which indicates that more token rotation cycles are required to achieve a smaller end-to-end delay; When M becomes large, D^* starts to increase since excessive time slot reservation for each superframe increases packet service time. Therefore, the optimal total number of time slots for each superframe, M^{opt}, can be determined based on Algorithm 1 to achieve the minimal average end-to-end delay, D^{opt}. We can also see that the average delays for local transmissions, D_a and D_b, are below threshold $D_{th} = 400$ ms. Nodes in token rings R_a and R_b are always guaranteed the minimum amount of time slots to maintain bounded average delays for local packet transmissions.

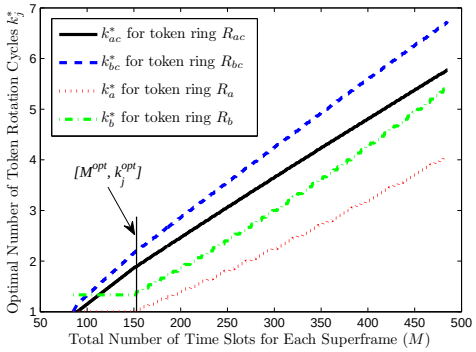


Fig. 6: Optimal number of token rotation cycles k_j^* for token ring R_j ($j = ac, bc, a, b$) versus M ($N_a = 20$, $N_b = 15$, $N_c = 15$).

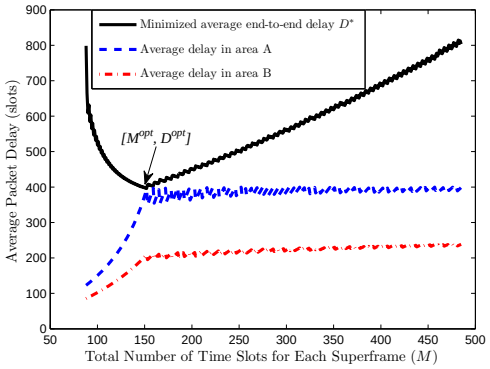


Fig. 7: Average packet delay versus M ($N_a = 20$, $N_b = 15$, $N_c = 15$).

In Fig. 8, we evaluate M^{opt} and k_j^{opt} as the total number of nodes, N, in the network increases, with the same numbers of

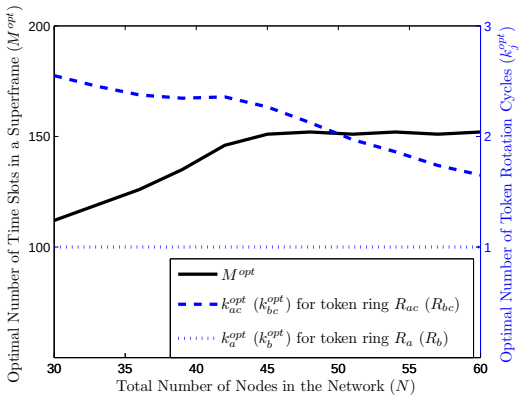


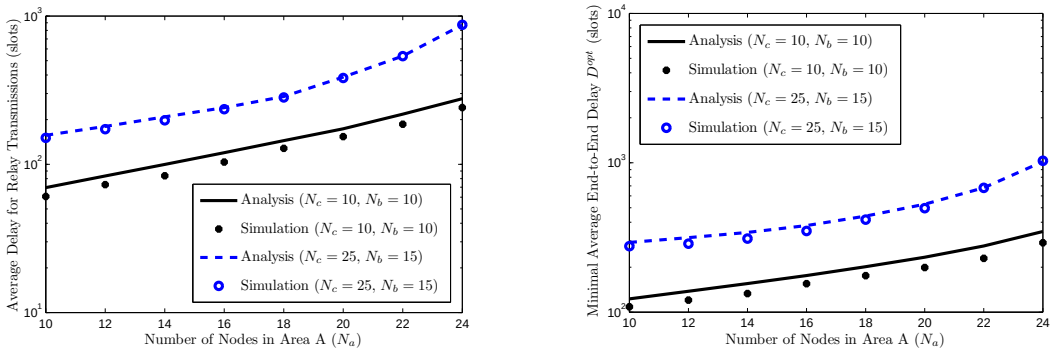
Fig. 8: The optimal total number of time slots, M^{opt}, for each superframe and the optimal number of token rotation cycles, k_j^{opt}, for token ring R_j with respect to the total number of nodes, N ($N_a = N_b = N_c$).

nodes in each area ($N_a = N_b = N_c$). It can be seen that M^{opt} increases consistently with N and remains at a steady value when the network load becomes high, which indicates that the optimal superframe length for the TA-MAC is adaptive to the network traffic load variations and is stable at high traffic load conditions. Within each superframe, the optimal numbers of token rotation cycles, k_{ac}^{opt} and k_{bc}^{opt}, keep decreasing with an increase of N, and k_a^{opt} and k_b^{opt} maintain at the minimum value in order to provide the maximum amount of resources for nodes in R_{ac} and R_{bc} to achieve the minimal average end-to-end delay.

B. Performance Metrics for the TA-MAC

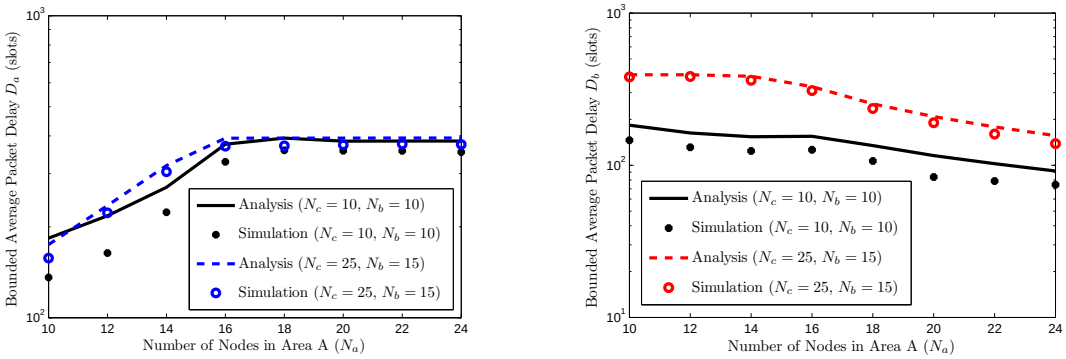
We evaluate the average delay for relay packet transmissions and the average end-to-end delay for both low and high network traffic load conditions, as the number of nodes, N_a, in area A varies. In Fig. 9a, it is shown that the average delay for relay transmissions increases consistently with N_a, and the simulation results match the analytical results more closely when the network traffic load becomes high ($N_c = 25, N_b = 15$), which verifies the effectiveness of the Poisson compound traffic arrival rate approximation on each relay node used in the analysis. Basically, the approximation becomes more accurate with an increasing number of nodes and node queue utilization ratios. Similar trends are observed in Fig. 9b. As expected, the minimal average end-to-end delay increases with the number of nodes. The simulation results match well with the analytical results.

Fig. 10a and 10b demonstrate average delays for local transmissions in areas A and B, D_a and D_b, which are



(a) (b)

Fig. 9: An evaluation of average delay for relay transmissions and average end-to-end delay under different network load conditions.



(a) (b)

Fig. 10: An evaluation of average delay for local transmissions in different network load conditions.

guaranteed to be bounded under threshold $D_{th} = 400$ ms with a varying network load. The TA-MAC reserves the minimum amount of time slots for both areas A and B to achieve the minimal average end-to-end delay for pairs of end users. Similarly, the analytical results match the simulation results.

Fig. 11 shows the aggregate network throughput for the TA-MAC versus the number of nodes. The throughput continuously increases with N_a, and the simulation results match well with the analytical results. A higher network throughput is achieved in the higher network load condition ($N_c = 25, N_b = 15$), with more packets transmitted in each superframe.

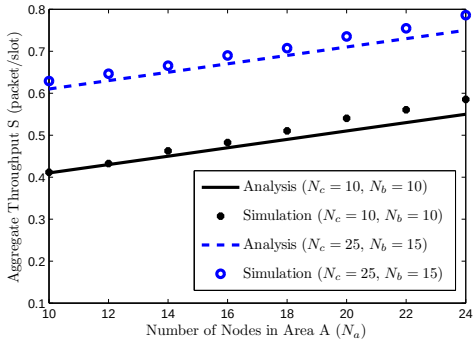


Fig. 11: Aggregate network throughput under different network load conditions

C. Performance Comparison

We compare the performance of the TA-MAC scheme with that of two existing MAC schemes proposed for multi-hop MANETs: a load adaptive MAC (LA-MAC) scheme [16] and a dynamic TDMA time slot assignment (DTSA) [15], [16]. LA-MAC is a hybrid MAC scheme, in which each node is allocated one time slot for exclusive use based on a dynamic assignment. If current slot owner has no packets to transmit, all its two-hop neighbors can contend for the transmission opportunity in its designated time slot, based on a mechanism similar to the CSMA/CA. If a node fails transmission attempts for a consecutive number of times, referred to as the state switch threshold, the node switches to a high contention state and broadcasts a notification message, allowing its one-hop neighbors to contend in its designated time slot. This state switch is to reduce packet collisions caused by hidden nodes in a high traffic load condition. Additional simulation parameters for the LA-MAC are set as follows: Each backoff slot duration is 20 μs; Both minimum and maximum contention window sizes (in the unit of backoff slot time) for slot owners and non slot owners are 1 and 8 backoff slots, and 9 and 16 backoff slots, respectively; The state switch threshold and packet retransmission limit are 5 and 7, respectively, and the high contention state duration is 100 superframes. The DTSA is a dynamic TDMA scheme, where each node in a two-hop network is allocated one exclusive time slot within a time frame. The first slot in each frame is reserved for new nodes broadcasting request messages to join the network. If the current frame has no available time slot for newly arriving nodes, the whole frame length is doubled to generate new available time slots. Thus, the scheme guarantees each node occupying two time slots at most in every time frame.

Fig. 12 shows the comparison of average end-to-end delay versus the network load, with the same numbers of nodes in

each area, between the TA-MAC scheme and the other two schemes. The LA-MAC achieves a smaller end-to-end delay in a low traffic load condition since nodes can contend to exploit the transmission opportunities in those time slots not used by the slot owners, making the MAC scheme behaving like CSMA/CA. However, the increase of node numbers in areas A, C and B results in a reduced number of empty slots and accumulated contention collisions for the LA-MAC, making it similar to the DTSA. Since the node time slot allocation in DTSA is not optimized, the traffic of relay nodes becomes quickly saturated, making the average end-to-end delay for both LA-MAC and DTSA increase dramatically to a large value in high network load conditions, whereas the TA-MAC achieves consistently minimal average end-to-end delay within a wide range of network traffic load. Its advantage becomes more obvious with a high number of nodes in the network by maintaining the end-to-end delay at a low level. The aggregate

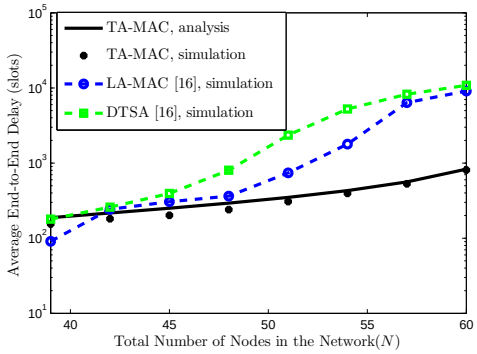


Fig. 12: Average end-to-end packet delay comparison between the TA-MAC scheme and other MAC schemes

network throughput comparison is plotted in Fig. 13. The throughput of all three schemes consistently increases with the traffic load. In a low network load condition, all the schemes achieve similar channel utilization. However, when the network load increases, the proposed TA-MAC scheme achieves consistently higher throughput than the other two schemes, by optimizing the scheduling of token rotation cycles for each token ring and controlling the queue of each node in an unsaturated condition, whereas the throughputs for both LA-MAC and DTSA start to saturate from a moderate network load condition.

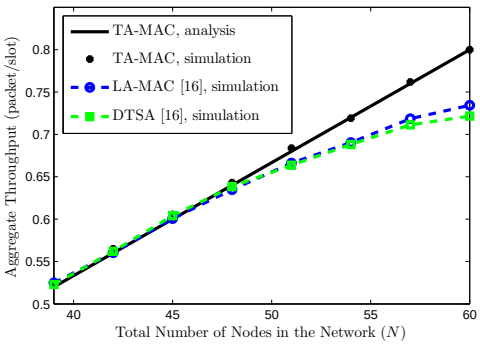


Fig. 13: Aggregate throughput comparison between the TA-MAC scheme and other MAC schemes

VII. Conclusion

In this paper, we propose a distributed token-based adaptive MAC scheme for a two-hop IoT-enabled MANET. To eliminate the hidden terminal problem, a TDMA-based superframe structure is proposed to accommodate packet transmissions from different groups of nodes in separate TDMA durations. In each individual node group, a token is circulated probabilistically among the node members for distributed time slot allocation which is adaptive to variations of the numbers of nodes in each network area. To optimize the design parameters of the TA-MAC scheme, performance analytical models are developed in closed-form functions of the MAC parameters (i.e., the numbers of token rotation cycles scheduled for each token ring and the superframe length) and the network traffic load. Then, an optimization framework is proposed to minimize the average end-to-end delay by acquiring the set of optimal MAC parameters for each superframe. Analytical and simulation results demonstrate that the TA-MAC scheme achieves consistently minimal average delay for end-to-end transmissions, bounded delays for local transmissions and high aggregate throughput with variations of the number of nodes in the network. Based on a comparison with other two MAC schemes, the TA-MAC demonstrates much better scalability for the IoT-based two-hop environment in presence of network load dynamics, especially in a high traffic load condition.

Appendix A
Proof of Proposition 1

For brevity, we only provide the proof for D_a. The proofs for other average delay functions can be carried out in a similar way. From (13), D_a is the combination of average queueing delay, D_q, and average service delay D_t. That is,

$$\begin{aligned} D_a = D_q + D_t &= \frac{\lambda_a}{2} \cdot \frac{x_1 k_a^2 + x_2 k_a + x_3}{k_a + x_4} + \frac{\varepsilon_a}{k_a} \\ &= \frac{\lambda_a}{2} f_1(k_a) + f_2(k_a) \end{aligned} \tag{21}$$

where x_1, x_2, x_3 and x_4 equal to the corresponding values in (13) ($x_1 = \alpha_a$, $x_2 = \beta_a$, $x_3 = \gamma_a$, $x_4 = -\lambda_a \varepsilon_a$).

In (21), $f_2(k_a)$ is a strictly convex function of k_a, since $f_2''(k_a) > 0$, $\forall k_a \geq 1$. On the other hand, the second-order derivative of $f_1(k_a)$ can be derived as

$$f_1''(k_a) = \frac{2x_1 x_4^2 - 2x_2 x_4 + 2x_3}{(k_a + x_4)^3} = \frac{g_1(x_4)}{(k_a + x_4)^3}. \tag{22}$$

Theoretically, x_1, x_2, and x_3 are fixed with a certain number of nodes, L_a, in area A, and k_a can take values from the interval $\left[1, \frac{M - L_{ac} - L_{bc}}{L_a}\right]$, due to constraints (14a) and (14f). Thus, by conforming to constraint (14d), we have $x_4 \in \left(-\frac{M - L_{ac} - L_{bc}}{L_a}, 0\right)$ with the variation of λ_a, to guarantee $(k_a + x_4)^3 > 0$ in (22). The numerator of (22) can be regarded as a quadratic function of x_4, denoted by $g_1(x_4)$. We define $\widetilde{g_1(x_4)}$ as an extension of $g_1(x_4)$ with $\mathbf{dom}\, \widetilde{g_1} \in (-\infty, \infty)$. Since $x_1 > 0$ and $x_2 < 0$, $\widetilde{g_1(x_4)}$ represents a parabola, opening upward with the horizontal axis coordinate of its vertex $x_4^* = -\frac{5L_a^2 + 12ML_a + 1}{12L_a^2}$. Because $x_4^* < -\frac{M - L_{ac} - L_{bc}}{L_a}$, it

is concluded that $g_1(x_4)$ with $\mathbf{dom}\, g_1 \in \left(-\frac{M-L_{ac}-L_{bc}}{L_a}, 0\right)$ is a strictly increasing function of x_4. Furthermore, since $g_1\left(-\frac{M^- -L_{ac}-L_{bc}}{L_a}\right) > 0$, it is proved that $g_1(x_4) > 0$, $\forall x_4 \in \mathbf{dom}\, g_1$, and thus we have $f_1''(k_a) > 0$, $\forall k_a \in \left[1, \frac{M-L_{ac}-L_{bc}}{L_a}\right]$. Hence, D_a is a linear combination of two strictly convex functions of k_a, which is also known to be strictly convex [30].

Appendix B
Proof of Corollary 1

According to Proposition 1, D_a is a convex function of k_a. Thus, we have $D_a''(k_a) > 0$, indicating that $D_a'(k_a)$ is a strictly increasing function which is derived as

$$\begin{aligned} D_a'(k_a) &= \frac{\lambda_a}{2} \cdot \frac{x_1 k_a^2 + 2x_1 x_4 k_a + x_2 x_4 - x_3}{(k_a + x_4)^2} - \frac{\varepsilon_a}{k_a^2} \\ &= \frac{\lambda_a}{2} f_3(k_a) + f_4(k_a), \quad k_a \in \left[1, \frac{M - L_{ac} - L_{bc}}{L_a}\right]. \end{aligned} \tag{23}$$

Therefore, the maximum value of $D_a'(k_a)$ is obtained when $k_a = \frac{M-L_{ac}-L_{bc}}{L_a}$. That is,

$$\begin{aligned} &D_a'\left(\frac{M - L_{ac} - L_{bc}}{L_a}\right) \\ &= \frac{\lambda_a}{2} f_3\left(\frac{M - L_{ac} - L_{bc}}{L_a}\right) + f_4\left(\frac{M - L_{ac} - L_{bc}}{L_a}\right). \end{aligned} \tag{24}$$

In (24), $f_3\left(\frac{M-L_{ac}-L_{bc}}{L_a}\right)$ is a function of x_4, with $x_4 \in \left(-\frac{M-L_{ac}-L_{bc}}{L_a}, 0\right)$, which is expressed as

$$\begin{aligned} &f_3\left(\frac{M - L_{ac} - L_{bc}}{L_a}\right) \\ &= \frac{\left[\frac{2x_1(M-L_{ac}-L_{bc})}{L_a} + x_2\right] x_4 + x_1\left(\frac{M-L_{ac}-L_{bc}}{L_a}\right)^2 - x_3}{\left(\frac{M-L_{ac}-L_{bc}}{L_a} + x_4\right)^2} \\ &= \frac{g_2(x_4)}{\left(\frac{M-L_{ac}-L_{bc}}{L_a} + x_4\right)^2}. \end{aligned} \tag{25}$$

Since $\frac{2x_1(M-L_{ac}-L_{bc})}{L_a} + x_2 < 0$, the linear function $g_2(x_4)$ is a strictly decreasing function with its maximum value being $g_2\left(-\frac{M^- -L_{ac}-L_{bc}}{L_a}\right) < 0$. Thus, we have $f_3\left(\frac{M-L_{ac}-L_{bc}}{L_a}\right) < 0$, and $D_a'\left(\frac{M-L_{ac}-L_{bc}}{L_a}\right) < 0$, $\forall x_4 \in \left(-\frac{M-L_{ac}-L_{bc}}{L_a}, 0\right)$. Hence, it is proved that $D_a'(k_a) < 0$, $\forall k_a \in \left[1, \frac{M-L_{ac}-L_{bc}}{L_a}\right]$. Similar proofs for the same property of other one-hop average delay functions can be made, which are omitted here.

Appendix C
Proof of Proposition 2

If we rewrite (SP2a) in a standard form, it is easy to verify that all the inequality constraint functions are convex with respect to k_{ac}. In the objective function, $D_{ab}(\cdot)$ is a summation of $D_{ac}(\cdot)$ and $D_{cb}(\cdot)$ according to (11), which is further expressed as

$$D_{ab}(k_{ac}, h(k_{ac})) = D_{ac}(k_{ac}) + D_{cb}(h(k_{ac})). \tag{26}$$

Based on Proposition 1, we know that $D_{ac}(\cdot)$ is a convex function of k_{ac} and $D_{cb}(\cdot)$ is convex function of $h(k_{ac})$. It is also found that $h(k_{ac})$ is a linear function of k_{ac}. Thus, according to the scalar composition rules, $D_{cb}(h(k_{ac}))$ is also a convex function of k_{ac}. Hence, $D_{ab}(\cdot)$, a linear combination of two convex functions, is also convex with respect to k_{ac}. The same property also holds for $D_{ba}(\cdot)$ with a similar proof. Moreover, as stated before, the two dimensional max function, $\max\{x, y\}$, is convex on $\mathbf{R}^2$ and nondecreasing in each of its two arguments. Therefore, according to the vector composition rules, the objective function of (SP2a) is a convex function with respect to the decision variable k_{ac}, which ends the proof.

References

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

[2] 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.

[3] N. Zhang, N. Cheng, A. T. Gamage, K. Zhang, J. W. Mark, and X. Shen, "Cloud assisted HetNets toward 5G wireless networks," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 59–65, 2015.

[4] 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.

[5] A. Aijaz and A. H. 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.

[6] 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.

[7] 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.*, vol. 65, no. 11, pp. 9358–9371, 2016.

[8] A. Rajandekar and B. Sikdar, "A survey of MAC layer issues and protocols for machine-to-machine communications," *IEEE Internet Things J.*, vol. 2, no. 2, pp. 175–186, 2015.

[9] "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.

[10] R. Jurdak, C. V. Lopes, and P. Baldi, "A survey, classification and comparative analysis of medium access control protocols for ad hoc networks," *IEEE Commun. Surv. Tutor.*, vol. 6, no. 1, pp. 2–16, 2004.

[11] K. Ghaboosi, M. Latva-aho, Y. Xiao, and Q. Zhang, "eMAC-A medium-access control protocol for the next-generation ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 58, no. 8, pp. 4476–4490, 2009.

[12] H. Zhai, J. Wang, and Y. Fang, "DUCHA: A new dual-channel MAC protocol for multihop ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 11, pp. 3224–3233, 2006.

[13] H. A. Omar, W. Zhuang, and L. Li, "Gateway placement and packet routing for multihop in-vehicle Internet access," *IEEE Trans. Emerging Topics in Comput.*, vol. 3, no. 3, pp. 335–351, 2015.

[14] I. Rhee, A. Warrier, J. Min, and L. Xu, "DRAND: Distributed randomized TDMA scheduling for wireless ad hoc networks," *IEEE Trans. Mobile Comput.*, vol. 8, no. 10, pp. 1384–1396, 2009.

[15] A. Kanzaki, T. Uemukai, T. Hara, and S. Nishio, "Dynamic TDMA slot assignment in ad hoc networks," in *Proc. IEEE AINA'03.*, 2003, pp. 330–335.

[16] W. Hu, H. Yousefi'zadeh, and X. Li, "Load adaptive MAC: A hybrid MAC protocol for MIMO SDR MANETs," *IEEE Trans. Wireless Commun.*, vol. 10, no. 11, pp. 3924–3933, 2011.
[17] Y. Liu, C. Yuen, X. Cao, N. U. Hassan, and J. Chen, "Design of a scalable hybrid MAC protocol for heterogeneous M2M networks," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 99–111, 2014.
[18] P. Koutsakis, "Token- and self-policing-based scheduling for multimedia traffic transmission over WLANs," *IEEE Trans. Veh. Technol.*, vol. 60, no. 9, pp. 4520–4527, 2011.
[19] H. Liang and W. Zhuang, "DFMAC: DTN-friendly medium access control for wireless local area networks supporting voice/data services," *ACM Mobile Netw. Appl. (MONET)*, vol. 16, no. 5, pp. 531–543, 2011.
[20] P. Wang and W. Zhuang, "A token-based scheduling scheme for WLANs supporting voice/data traffic and its performance analysis," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1708–1718, 2008.
[21] I. S. Liu, F. Takawira, and H. J. Xu, "A hybrid token-CDMA MAC protocol for wireless ad hoc networks," *IEEE Trans. Mobile Comput.*, vol. 7, no. 5, pp. 557–569, 2008.
[22] Y. Bi, K. H. Liu, L. X. Cai, X. Shen, and H. Zhao, "A multi-channel token ring protocol for QoS provisioning in inter-vehicle communications," *IEEE Trans. Wireless Commun.*, vol. 8, no. 11, pp. 5621–5631, 2009.
[23] P. Teymoori, N. Yazdani, and A. Khonsari, "DT-MAC: An efficient and scalable medium access control protocol for wireless networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 3, pp. 1268–1278, 2013.
[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] 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.
[26] D. P. Bertsekas, R. G. Gallager, and P. Humblet, *Data networks*. Englewood Cliffs, NJ, USA: Prentice-hall, 1987, vol. 2.
[27] 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.
[28] J. Ren, Y. Zhang, N. Zhang, D. Zhang, and X. Shen, "Dynamic channel access to improve energy efficiency in cognitive radio sensor networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3143–3156, 2016.
[29] D. Zhang, Z. Chen, J. Ren, N. Zhang, M. Awad, H. Zhou, and X. Shen, "Energy harvesting-aided spectrum sensing and data transmission in heterogeneous cognitive radio sensor network," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 831–843, 2017.
[30] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
[31] J. Gorski, F. Pfeuffer, and K. Klamroth, "Biconvex sets and optimization with biconvex functions: a survey and extensions," *Math. Method. Oper. Res.*, vol. 66, no. 3, pp. 373–407, 2007.
[32] "OMNeT++ 5.0," [Online]. Available: http://www.omnetpp.org/omnetpp.
[33] Q. Ye and W. Zhuang, "Distributed and adaptive medium access control for Internet-of-Things-enabled mobile networks," *IEEE Internet Things J.*, to appear. DOI: 10.1109/JIOT.2016.2566659.

Qiang Ye (S'16-M'17) received the B.S. degree in network engineering, the M.S. degree in communication and information system from Nanjing University of Posts and Telecommunications, Nanjing, China, and the Ph.D. degree in electrical and computer engineering from University of Waterloo, Waterloo, ON, Canada, in 2009, 2012, and 2016, respectively. Since December 2016, He has been a postdoctoral fellow with the Department of Electrical and Computer Engineering, University of Waterloo. His research interests include medium access control and performance optimization in mobile ad hoc networks and Internet of Things, resource virtualization for 5G networks, and software-defined networking.

Weihua Zhuang (M'93-SM'01-F'08) has been with the Department of Electrical and Computer Engineering, University of Waterloo, Canada, since 1993, where she is a Professor and a Tier I Canada Research Chair in Wireless Communication Networks. Her current research focuses on resource allocation and QoS provisioning in wireless networks, and on smart grid. She is a co-recipient of several best paper awards from IEEE conferences. Dr. Zhuang was the Editor-in-Chief of IEEE Transactions on Vehicular Technology (2007-2013), and the Technical Program Chair/Co-Chair of the IEEE VTC Fall 2017/2016. She is a Fellow of the IEEE, a Fellow of the Canadian Academy of Engineering, a Fellow of the Engineering Institute of Canada, and an elected member in the Board of Governors and VP Publications of the IEEE Vehicular Technology Society.