Enhanced Markov Chain Model and Throughput Analysis of the Slotted CSMA/CA for IEEE 802.15.4 Under Unsaturated Traffic Conditions

Chang Yong Jung, *Student Member, IEEE*,
Ho Young Hwang, *Member, IEEE*,
Dan Keun Sung, *Senior Member, IEEE*, and
Gang Uk Hwang, *Member, IEEE*

***Abstract*—In this paper, we propose an analytical Markov chain model of the slotted carrier-sense multiple-access/collision-avoidance (CSMA/CA) protocol for IEEE 802.15.4 under unsaturated traffic conditions. Our proposed Markov chain model reflects the characteristics of the IEEE 802.15.4 medium-access control (MAC) protocol, such as a superframe structure, acknowledgements, and retransmissions with and without limit. We evaluate the throughput performance of the slotted CSMA/CA and verify the analytical model using simulation results.**

***Index Terms*—Carrier sense multiple access/collision avoidance (CSMA/CA), IEEE 802.15.4, medium-access control (MAC), throughput.**

I. Introduction

As wireless sensor networks (WSNs) have been widely deployed in our lives, connecting sensor nodes in a simple and efficient manner with low power and low cost has become an important issue. The IEEE 802.15.4 standard for low-rate wireless personal area networks (LR-WPANs) has been introduced to achieve these requirements [1].

IEEE 802.15.4 networks can operate in either a beacon-enabled mode or a nonbeacon-enabled mode. In the nonbeacon-enabled mode, nodes in a personal area network (PAN) communicate with each other according to an unslotted carrier-sense multiple-access/collision-avoidance (CSMA/CA) protocol. On the other hand, in the beacon-enabled mode, nodes communicate with each other according to a slotted CSMA/CA protocol based on a superframe structure. Each superframe consists of an active period and an inactive period. The active period consists of a beacon period, a contention access period (CAP), and a contention-free period (CFP). During the inactive period, the coordinator and the nodes shall not interact with each other and may enter a low-power mode. To investigate the slotted CSMA/CA protocol of IEEE 802.15.4, we consider the beacon-enabled mode in this paper.

During the CAP in the beacon-enabled mode, IEEE 802.15.4 has adopted a slotted CSMA/CA protocol, which is different from that of the IEEE 802.11 wireless local area network (WLAN) [2], [3]. In the case of the WLAN, the backoff count value decreases by one only if the channel is idle; otherwise, it is frozen. On the other hand, in the case of IEEE 802.15.4, the backoff count value decreases by one, regardless of whether the channel is idle or busy. Thus, we need to make a new analytical model for the CSMA/CA algorithm that is used in IEEE 802.15.4.

There have been a few papers regarding this technical issue. Park *et al.* [4], Tao *et al.* [5], Pollin *et al.* [6], and Lee *et al.* [7] proposed analytical Markov chain models for the slotted CSMA/CA of the IEEE 802.15.4 medium-access control (MAC) protocol. They assumed that the duration of a CAP is infinite without considering the superframe structure under saturated traffic conditions. Since most applications in wireless sensor networks (WSNs) are expected to operate under unsaturated traffic conditions, and they use superframes, including inactive periods, to reduce power consumption, we need an analytical model considering unsaturated traffic conditions and the superframe structure. From this point of view, Ramachandran *et al.* [8] proposed an analytical model of the slotted CSMA/CA under unsaturated traffic conditions; however, they did not consider the inactive period in the superframe structure. Mišić and Mišić [9] also proposed a Markov chain model to analyze the slotted CSMA/CA algorithm under unsaturated traffic conditions considering the superframe structure, acknowledgements, and retransmission schemes of the IEEE 802.15.4 MAC protocol. However, they assumed that there was no retransmission limit, and they did not verify their analytical model. Park *et al.* [4] showed that the model of Mišić and Mišić did not match the simulation results under saturated traffic conditions.

In this paper, we propose an enhanced Markov chain model to observe the throughput performance of the slotted CSMA/CA algorithm under unsaturated traffic conditions, and the proposed Markov chain model reflects the characteristics of the IEEE 802.15.4 MAC protocol, such as a superframe structure, acknowledgements, and retransmission schemes with and without a retry limit. Through simulations, we verify that the analytic results from our model are well matched with the simulation results. In addition, we show that our model provides more accurate results than the model of Mišić and Mišić [9].

The rest of this paper is organized as follows. In Section II, we briefly describe the IEEE 802.15.4 MAC protocol. In Section III, we propose an analytical model of the slotted CSMA/CA for the IEEE 802.15.4 MAC protocol and evaluate the throughput performance of the slotted CSMA/CA. Finally, we present conclusions in Section IV.

II. Overview of the IEEE 802.15.4 MAC Protocol

At the start of each superframe, the PAN coordinator transmits a beacon frame that carries system parameters, such as a beacon order (BO) that determines the length of a beacon interval ($BI = 16 * 60 * 2^{BO}$ symbols) and a superframe order (SO) that determines the length of a superframe duration ($SD = 16 * 60 * 2^{SO}$ symbols). In the CAP period, each node communicates with the PAN coordinator and other nodes using the slotted CSMA/CA. The duration of one slot is $aUnitBackoffPeriod$ (default value $= 20$ symbols). When a node has a new data frame to transmit, it initializes relevant parameters such as a backoff exponent (BE) and the number of backoffs or backoff stages (NB), which are set to $macMinBE$ (default value $= 3$) and 0, respectively. In addition, it uniformly selects a backoff counter value from a window $[0, 2^{BE} - 1]$. The backoff counter value is decremented by one for each time slot, regardless of the channel state, and whenever the backoff counter value is zero, the node performs carrier sensing that requires two clear channel assessments (CCAs) at the physical (PHY) layer before a transmission. If the channel is assessed

Manuscript received July 24, 2007; revised January 17, 2008 and March 20, 2008. First published April 18, 2008; current version published January 16, 2009. This work has been supported in part by a grant "Next Generation PC Project" from the Institute of Information Technology Assessment (IITA) and in part by the Korea Research Foundation Grant funded by the Korean Government (KRF-2008-357-D00181). The review of this paper was coordinated by Prof. J. Li.

C. Y. Jung and D. K. Sung are with the School of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea (e-mail: cyjung@cnr.kaist.ac.kr; dksung@ee.kaist.ac.kr).

H. Y. Hwang is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: hyhwang@bbcr.uwaterloo.ca).

G. U. Hwang is with the Department of Mathematical Sciences, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea (e-mail: guhwang@kaist.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2008.923669

to be idle at the two consecutive CCAs, then it transmits the data frame. If the channel is assessed to be busy, it increases the values of BE and NB by one and delays the transmission for a random number of time slots that are uniformly chosen from $[0, 2^{BE} - 1]$, where BE is no more than $aMaxBE$ (default value $= 5$). The above procedure is continued until the successful transmission; however, if the NB value is greater than $macMaxCSMABackoffs$ (default value $= 4$), then the CSMA/CA algorithm shall be terminated with a channel access failure. If either a channel access failure occurs or a frame transmission failure occurs due to a collision, the node retries the aforementioned procedure for retransmissions up to $aMaxFrameRetries$ (default value $= 3$) times. Since the transmission of an Ack frame shall commence at a slot boundary, the duration t_{ack} from the reception of the last symbol of the data frame to the transmission of the first symbol of its Ack frame is between $aTurnaroundTime$ (default value $= 12$ symbols) and $aTurnaroundTime + aUnitBackoffPeriod$ (32 symbols), that is, one time slot can be included in the duration t_{ack} at most. After transmitting a data frame, the node waits for its Ack frame during the duration that is specified by the parameter $macAckWaitDuration$ (default value $= 54$ symbols).

III. Analysis of the Slotted CSMA/CA in the IEEE 802.15.4 MAC Protocol

A. Markov Chain Models

In this paper, we consider a single-hop wireless network consisting of a PAN coordinator and n sensor nodes. We assume that all nodes are within the transmission range of each other and time-synchronized by the PAN coordinator's beacon. We also assume that there are no transmission errors and no channel sensing errors. We consider a star topology and an uplink data transmission scenario so that transmitted frames can be lost only due to collisions. Note that WSNs might have a different topology other than a star topology; however, considering a general topology is beyond the scope of this paper.

We assume that the data frame arrives at each node according to a Poisson process with rate λ and that each node can store a single data frame. Thus, when a node has a data frame to transmit, it cannot accept any more new data frames from upper layers. We further assume that all n nodes are homogeneous, and accordingly, the performance of all nodes is identical. For this reason, we tag an arbitrary node and call it the tagged node.

For the analysis, we construct a discrete-time Markov chain, which models the operation of the CSMA/CA algorithm in the tagged node and captures the key characteristics of the IEEE 802.15.4 MAC protocol such as a superframe structure, acknowledgements, and retransmission schemes with and without a retry limit. For convenience, when we construct the Markov chain, we consider only the time epochs where the states of the Markov chain defined below are changed. The state transition diagram of the Markov chain is given in Fig. 1. The states of the Markov chain at time t are classified into three types. The first type is of the form $\{s(t), b(t), w(t), r(t)\}$. Here, $s(t) \in [0, m]$ represents the value of NB at time t, where $m = macMaxCSMABackoff$. $b(t)$ represents the value of the backoff counter at time t. When $s(t) = i$, $b(t)$ is in $[0, W_i - 1]$, where $W_0 = 2^{macMinBE}$, and $W_i = W_0 2^{min(i, aMaxBE - macMinBE)}$, $1 \leq i \leq m$. $w(t) \in \{1, 2\}$ represents the remaining number of CCAs to be done for a transmission at time t. Thus, if $b(t) = 0$ and $w(t) = 2$, then the tagged node performs the first CCA at time t. Similarly, if $b(t) = 0$ and $w(t) = 1$, the tagged node performs the second CCA at time t. P_I and P_{II} in Fig. 1 denote the probabilities that the channel is idle when performing the first CCA and the second CCA, respectively. Finally, $r(t) \in [0, R]$ represents the value of the retransmission plane at time t, which is shown by a rectangular box in Fig. 1, where $R = aMaxFrameRetries$. Note that our model can simply be applied to the retransmission scheme without limit if R tends to ∞.

The second type is of the form $\{T^{nd}_{s(t)}, r(t)\}$ and $\{T^{d}_{s(t)}, r(t)\}$. These states represent the nondeferred and deferred transmissions in the $s(t)$th backoff stage on the $r(t)$th retransmission plane at time t. Since we consider the superframe structure with a finite CAP and an inactive period, note that the tagged node needs to defer a transmission until the start of the next CAP when the transmission cannot be completed within the current CAP. P^{nd}_c and P^{d}_c in Fig. 1 denote the collision probabilities for nondeferred transmissions and deferred transmissions, respectively.

The third type is of the form $\{D_{s(t)}, r(t)\}$. These states represent the waiting state in the $s(t)$th backoff stage on the $r(t)$th retransmission plane at time t in order for the tagged node to defer the transmission until the next CAP due to the lack of the remaining slots in the current CAP. P_d is the probability that a deferred transmission occurs.

In Fig. 1, we have, in fact, one more state $\{-1, -1, -1, -1\}$ that represents the state where the tagged node has no data frame to transmit. The transition probability α is the probability that a new data frame occurs between state transition times at the tagged node.

Now, to complete the construction of the Markov chain in Fig. 1, we need to determine the probabilities $P_I, P_{II}, P^{nd}_c, P^{d}_c, P_d$, and α. As shown below, we can directly determine the probabilities P_d and α, but it seems to be difficult to directly determine P_I, P_{II}, P^{nd}_c, and P^{d}_c. To solve this problem, assuming that P_I, P_{II}, P^{nd}_c, and P^{d}_c are independent of the backoff stages and retransmission planes, we obtain expressions of P_I, P_{II}, P^{nd}_c, and P^{d}_c in terms of the steady-state probabilities of the Markov chain and numerically solve them. The details are summarized in the following sections.

B. Probabilities P_d and α

To compute the probability P_d, note that the total number of time slots that are needed for a single transmission is $2 + N^{\text{slot}}_{\text{data}} + N^{\text{slot}}_{t_{\text{ack}}} + N^{\text{slot}}_{\text{ack}}$. Here, two slots are included due to the number of time slots for performing two CCAs, $N^{\text{slot}}_{\text{data}}$ is the number of time slots for the transmission of a data frame, $N^{\text{slot}}_{t_{\text{ack}}}$ is the number of time slots for the period t_{ack}, and $N^{\text{slot}}_{\text{ack}}$ is the number of time slots for the transmission of an Ack frame. Since the transmission is deferred due to the lack of the remaining slot times in a CAP, P_d is approximated by

$$P_d \approx \frac{N^{\text{slot}}_{\text{def}}}{N^{\text{slot}}_{\text{CAP}}} = \frac{2 + N^{\text{slot}}_{\text{data}} + N^{\text{slot}}_{t_{\text{ack}}} + N^{\text{slot}}_{\text{ack}}}{N^{\text{slot}}_{\text{CAP}}}$$

where $N^{\text{slot}}_{\text{CAP}}$ is the total number of time slots in a CAP.

Next, α is the probability that a new data frame occurs between state transition times at the tagged node. Since the data frame arrives according to a Poisson process with rate λ, α is expressed as

$$\alpha = \frac{N^{\text{slot}}_{\text{CAP}} - 1}{N^{\text{slot}}_{\text{CAP}}} \int_0^{T_{\text{slot}}} \lambda e^{-\lambda \tau} d\tau + \frac{1}{N^{\text{slot}}_{\text{CAP}}} \int_0^{T_o} \lambda e^{-\lambda \tau} d\tau \qquad (1)$$

where $T_o = T_{\text{slot}} + T_{\text{CFP}} + T_{\text{Inactive}} + T_{\text{BEP}}$. Here, T_{slot}, T_{CFP}, T_{Inactive}, and T_{BEP} are the durations of one time slot, a CFP, an inactive period, and a beacon period, respectively. Note that $(N^{\text{slot}}_{\text{CAP}} - 1)/N^{\text{slot}}_{\text{CAP}}$ is the probability that an arbitrary time slot is not the last time slot of a CAP.

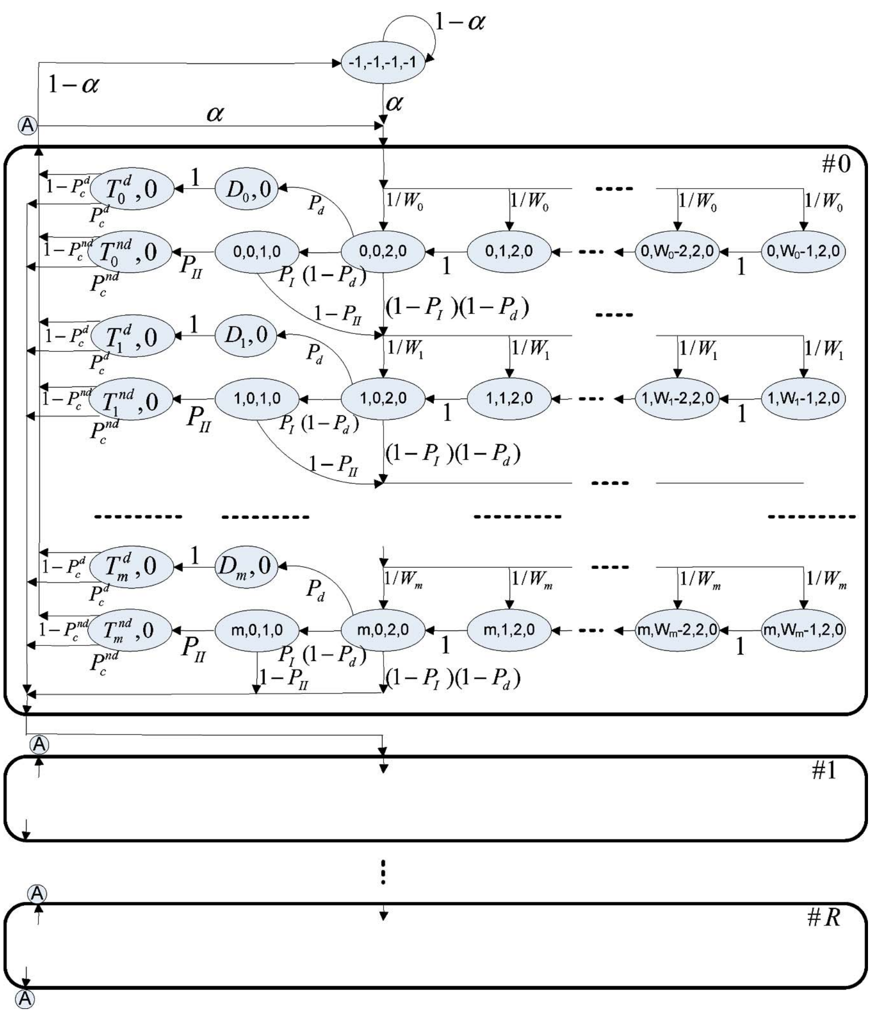

Fig. 1. Markov chain model.

C. Probabilities P_c^{nd} and P_c^d

Let $b_{i,j,k,l} = \lim_{t\to\infty} P\{\, s(t) = i, b(t) = j, w(t) = k, r(t) = l\}$, $b_{T_i^{nd},l} = \lim_{t\to\infty} P\{T_{s(t)^{nd}} = T_i^{nd}, r(t) = l\}$, $b_{T_i^d,l} = \lim_{t\to\infty} P\{T_{s(t)}^d = T_i^d, r(t) = l\}$, and $b_{D_i,l} = \lim_{t\to\infty} P\{D_{s(t)} = D_i, r(t) = l\}$, where $i \in [0, m]$, $j \in [0, W_i - 1]$, $k \in [1, 2]$, and $l \in [0, R]$, be the stationary probabilities of the Markov chain. We assume that the probabilities P_I, P_{II}, P_c^{nd}, and P_c^d are independent of the backoff stages and the retransmission planes. Then, some transition probabilities in Fig. 1 can be expressed as

$$\begin{aligned}
b_{T_i^d,l} &= P_d b_{i,0,2,l}, \quad i \in [0,m],\ l \in [0,R] \\
b_{T_i^{nd},l} &= C_1 b_{i,0,2,l} = P_I P_{II}(1-P_d) b_{i,0,2,l} \\
&\quad i \in [0,m], l \in [0,R] \\
b_{i+1,0,2,l} &= C_2 b_{i,0,2,l} = (1-P_I P_{II})(1-P_d) b_{i,0,2,l} \\
&\quad i \in [0,m-1],\ l \in [0,R] \\
b_{0,0,2,l+1} &= C_3 b_{0,0,2,l} \\
&= \left[\left(P_c^{nd} C_1 + P_c^d P_d\right) \sum_{i=0}^{m} C_2^i + C_2^{m+1} \right] b_{0,0,2,l} \\
&\quad l \in [0, R-1]
\end{aligned}$$

where C_1 represents the transition probability from the state where the tagged node performs the first CCA to the state where the tagged node performs a nondeferred transmission, and C_2 and C_3 represent the transition probabilities between the backoff stages and the retransmission planes, respectively.

Let τ_{nd} and τ_d be the stationary probabilities that the tagged node carries out nondeferred and deferred transmissions, respectively. Then, they satisfy

$$\tau_{nd} = \sum_{l=0}^{R} \sum_{i=0}^{m} b_{T_I^{nd},l}$$

$$\tau_d = \sum_{l=0}^{R} \sum_{i=0}^{m} b_{T_i^d,l}.$$

Let τ be the stationary probability that the tagged node transmits a data frame. Then, it follows that

$$\tau = \tau_{nd} + \tau_d.$$

The collision probabilities P_c^{nd} and P_c^d for nondeferred and deferred transmissions of the tagged node, respectively, are expressed as

$$P_c^{nd} = 1 - (1 - \frac{\tau_{nd}}{C_1})^{(n-1)} \quad (2)$$

$$P_c^d = 1 - (1 - P_{\mathrm{CAP}}^{d\_tx})^{(n-1)} \quad (3)$$

where τ_{nd}/C_1 in (2) is the conditional probability that the tagged node performs the first CCA, given that a nondeferred transmission occurs. $P_{\mathrm{CAP}}^{d\_tx}$ in (3) is the probability that the tagged node defers the transmission in a CAP. To compute $P_{\mathrm{CAP}}^{d\_tx}$, note that $\tau_d N_{\mathrm{CAP}}^{\mathrm{slot}}/N_{\mathrm{state}}^{\mathrm{slot}}$ represents the average number of deferred transmissions in a CAP at the tagged node, where $N_{\mathrm{state}}^{\mathrm{slot}}$ is the average number of time slots staying in an arbitrary state of the Markov chain in Fig. 1. $N_{\mathrm{state}}^{\mathrm{slot}}$ is given by

$$N_{\mathrm{state}}^{\mathrm{slot}} = (1-\tau)\cdot 1 + \tau(1-P_c)\left(N_{\mathrm{data}}^{\mathrm{slot}} + N_{t_{\mathrm{ack}}}^{\mathrm{slot}} + N_{\mathrm{ack}}^{\mathrm{slot}}\right) + \tau P_c\left(N_{\mathrm{data}}^{\mathrm{slot}} + N_{\mathrm{timeout}}^{\mathrm{slot}}\right)$$

where $N_{\mathrm{timeout}}^{\mathrm{slot}}$ is the number of time slots until the Ack timer expires due to no Ack. Since there is at most one deferred transmission in a CAP for the tagged node, we have

$$P_{\mathrm{CAP}}^{d\_tx} = \frac{\tau_d N_{\mathrm{CAP}}^{\mathrm{slot}}}{N_{\mathrm{state}}^{\mathrm{slot}}}. \tag{4}$$

From (2) and (3), the overall collision probability P_c for the tagged node is expressed as

$$P_c = P_c^{nd}\frac{\tau_{nd}}{\tau_d+\tau_{nd}} + P_c^{d}\frac{\tau_d}{\tau_d+\tau_{nd}}.$$

D. Probabilities P_I and P_{II}

To obtain the probabilities P_I and P_{II} that the channel is idle at the first and second CCAs, respectively, we need to compute the average number of busy and idle time slots in a CAP from the viewpoint of the network or the channel. The average number of busy slots in a CAP is

$$N_{\mathrm{CAP}}^{b\_\mathrm{slot}} = \left[N_{\mathrm{net}}^{s\_tx}\left(N_{\mathrm{data}}^{\mathrm{slot}} + N_{\mathrm{ack}}^{\mathrm{slot}}\right) + N_{\mathrm{net}}^{c\_tx}N_{\mathrm{data}}^{\mathrm{slot}}\right] \tag{5}$$

where $N_{\mathrm{net}}^{s\_tx}$ and $N_{\mathrm{net}}^{c\_tx}$ are the average numbers of successful transmissions and collided transmissions in the network except for the tagged node in a CAP, respectively, and can be expressed as

$$N_{\mathrm{net}}^{s\_tx} = (n-1)\left[N_{\mathrm{node}}^{nd\_tx}P_{c^*|n-2,0}^{nd} + N_{\mathrm{node}}^{d\_tx}P_{c^*|n-2,0}^{d}\right] \tag{6}$$

$$N_{\mathrm{net}}^{c\_tx} = (n-1) \times \left[N_{\mathrm{node}}^{nd\_tx}\sum_{i=1}^{n-2}\frac{P_{c^*|n-2,i}^{nd}}{i+1} + N_{\mathrm{node}}^{d\_tx}\sum_{i=1}^{n-2}\frac{P_{c^*|n-2,i}^{d}}{i+1}\right] \tag{7}$$

where $N_{\mathrm{node}}^{nd\_tx}$ is the average number of nondeferred transmissions for a nontagged node in a CAP and is given by $N_{\mathrm{node}}^{nd\_tx} = \tau_{nd}N_{\mathrm{CAP}}^{\mathrm{slot}}/N_{\mathrm{state}}^{\mathrm{slot}}$, and $N_{\mathrm{node}}^{d\_tx}$ is the average number of deferred transmissions for a nontagged node in a CAP and is given by $N_{\mathrm{node}}^{d\_tx} = \tau_d N_{\mathrm{CAP}}^{\mathrm{slot}}/N_{\mathrm{state}}^{\mathrm{slot}}$, which is the same as $P_{\mathrm{CAP}}^{d\_tx}$ in (4).

In addition, $P_{c^*|n-2,i}^{nd}$ is the conditional collision probability that, under the given condition that one node among $n-1$ nontagged nodes carries out a nondeferred transmission, i nodes among $n-2$ other nontagged nodes carry out nondeferred transmissions and $i+1$ nodes collide accordingly and is expressed as

$$P_{c^*|n-2,i}^{nd} = \binom{n-2}{i}\left(\frac{\tau_{nd}}{C_1}\right)^i\left(1-\frac{\tau_{nd}}{C_1}\right)^{n-2-i}.$$

In the same manner, $P_{c^*|n-2,i}^{d}$ for deferred transmissions is expressed as

$$P_{c^*|n-2,i}^{d} = \binom{n-2}{i}{P_{\mathrm{CAP}}^{d\_tx}}^i\left(1-P_{\mathrm{CAP}}^{d\_tx}\right)^{n-2-i}.$$

Since $P_{c^*|n-2,0}^{nd}$ in (6) is the probability that a nondeferred transmission of one node among $n-1$ nontagged nodes is successful, $(n-1)N_{\mathrm{node}}^{nd\_tx}P_{c^*|n-2,0}^{nd}$ in (6) is the average number of successful nondeferred transmissions for the network except the tagged node in a CAP.

Regarding (7), we consider the case that $i+1$ nontagged nodes collide. From the viewpoint of the network or the channel, it is considered as a single collided transmission. Therefore, we can say that each of the $i+1$ nontagged nodes that is involved in the collision contributes the amount of the $1/(i+1)$ portion to the single collided transmission. $(n-1)N_{\mathrm{node}}^{nd\_tx}\sum_{i=1}^{n-2}(P_{c^*|n-2,i}^{nd}/(i+1))$ in (7) is the average number of nondeferred collided transmissions for the network except the tagged node in a CAP.

In the same way, $(n-1)N_{\mathrm{node}}^{d\_tx}P_{c^*|n-2,0}^{d}$ in (6) and $(n-1)N_{\mathrm{node}}^{d\_tx}\sum_{i=1}^{n-2}(P_{c^*|n-2,i}^{d}/(i+1))$ in (7) are the average numbers of successful deferred transmissions and collided deferred transmissions, respectively, for the network except the tagged node in a CAP.

Note that the first and second CCAs for nondeferred transmissions cannot be performed at any time slots of a CAP but can be performed in a limited range of a CAP. The length of the limited range is $N_{\mathrm{CAP}}^{\mathrm{slot}} - N_{\mathrm{def}}^{\mathrm{slot}}$. $N_{\mathrm{CAP}}^{b\_\mathrm{slot}}$ in (5) can include the number of busy slots in the range where the tagged node does not perform the first CCA due to deferred transmissions. Since the number of those busy slots is very small compared with that of slots that the tagged node can perform the first CCA, i.e., $(N_{\mathrm{CAP}}^{\mathrm{slot}} - N_{\mathrm{def}}^{\mathrm{slot}})$, we assume that this effect can be negligible. Therefore, the probability P_I is approximated as

$$P_I \approx \frac{\left(N_{\mathrm{CAP}}^{\mathrm{slot}} - N_{\mathrm{def}}^{\mathrm{slot}}\right) - N_{\mathrm{CAP}}^{b\_\mathrm{slot}}}{N_{\mathrm{CAP}}^{\mathrm{slot}} - N_{\mathrm{def}}^{\mathrm{slot}}}. \tag{8}$$

Next, we observe that the second CCA can occur when the channel is idle at the first CCA. Note that $N_{\mathrm{CAP}}^{b|i\_\mathrm{slot}}$ is the average number of slots where the channel is idle, but it is busy at the next slot. Since those slots can just occur before data or Ack frame transmissions, $N_{\mathrm{CAP}}^{b|i\_\mathrm{slot}}$ is expressed as

$$N_{\mathrm{CAP}}^{b|i\_\mathrm{slot}} = \left[N_{\mathrm{net}}^{s\_tx}\left(1 + N_{t_{\mathrm{ack}}}^{\mathrm{slot}}\right) + N_{\mathrm{net}}^{c\_tx}\cdot 1\right].$$

Thus, the probability P_{II} is approximated as

$$P_{II} \approx \frac{1}{P_I}\frac{\left(N_{\mathrm{CAP}}^{\mathrm{slot}} - N_{\mathrm{def}}^{\mathrm{slot}}\right) - N_{\mathrm{CAP}}^{b\_\mathrm{slot}} - N_{\mathrm{CAP}}^{b|i\_\mathrm{slot}}}{N_{\mathrm{CAP}}^{\mathrm{slot}} - N_{\mathrm{def}}^{\mathrm{slot}}}. \tag{9}$$

Finally, each parameter of P_I, P_{II}, P_c^{nd}, and P_c^{d} can be numerically solved from (2), (3), (8), (9), and the normalization condition of the Markov chain, as mentioned before.

E. Throughput Analysis

Let S be the system throughput. Then, S is computed as

$$S = \frac{nD_R\tau(1-P_c)E[T_{\mathrm{Payload}}]}{(1-\tau)E\left[T_{\mathrm{slot}}^*\right] + \tau(1-P_c)E[T_{\mathrm{suc}}] + \tau P_c E[T_{\mathrm{col}}]}$$

where D_R is the data rate (in bits per second), $E[T_{\mathrm{Payload}}]$ is the average duration of a data frame payload, and $E[T_{\mathrm{suc}}]$ and $E[T_{\mathrm{col}}]$ are the average durations of a successful transmission and a collided transmission, respectively, and are expressed as

$$E[T_{\mathrm{suc}}] = T_{\mathrm{slot}}N_{\mathrm{suc}}^{\mathrm{slot}} = T_{\mathrm{slot}}\left(N_{\mathrm{data}}^{\mathrm{slot}} + N_{t_{\mathrm{ack}}}^{\mathrm{slot}} + N_{\mathrm{ack}}^{\mathrm{slot}}\right)$$

$$E[T_{\mathrm{col}}] = T_{\mathrm{slot}}N_{\mathrm{col}}^{\mathrm{slot}} = T_{\mathrm{slot}}\left(N_{\mathrm{data}}^{\mathrm{slot}} + N_{\mathrm{timeout}}^{\mathrm{slot}}\right).$$

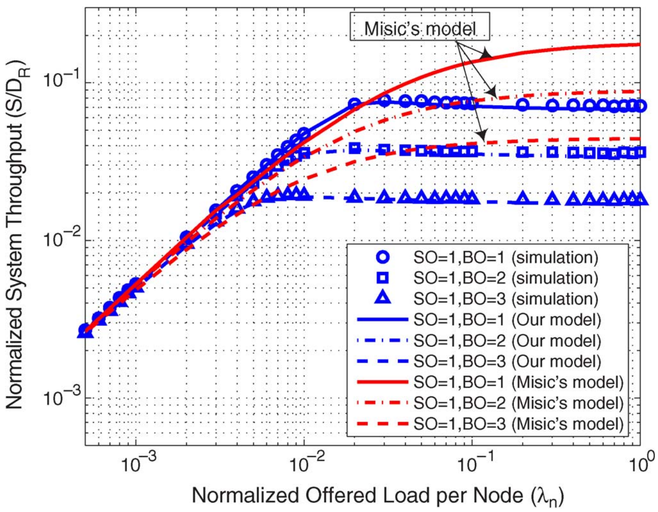

Fig. 2. Normalized system throughput for varying SO and BO values and retransmissions without limit.

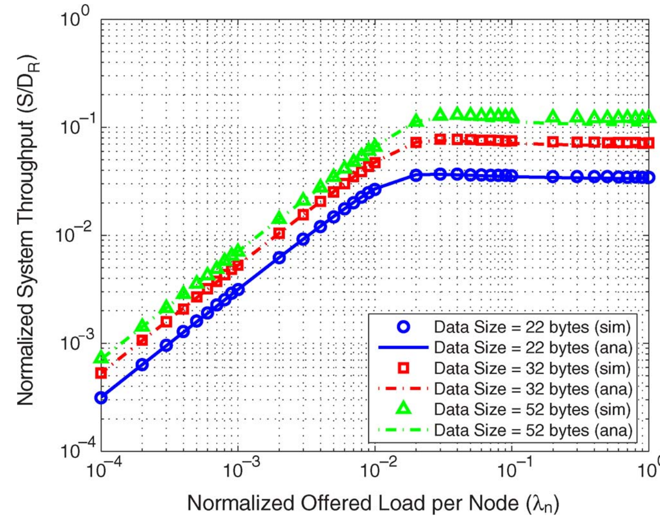

Fig. 3. Normalized system throughput for varying data sizes and retransmissions with limit.

$E[T^*_{\text{slot}}]$ is the average duration of a nontransmission slot. Note that the "transmission slots" include slots of data and Ack frame transmissions and t_{ack} for successful transmissions, as well as slots of data frame transmissions and waiting until the Ack timer expires for collided transmissions. To compute $E[T^*_{\text{slot}}]$, we first consider the average number of transmission slots of the tagged node in a CAP, which is denoted by $N^{tx\_\text{slot}}_{\text{node}}$. Since the average number of transmissions of the tagged node in a CAP is given by

$$N^{tx}_{\text{node}} = N^{nd\_tx}_{\text{node}} + N^{d\_tx}_{\text{node}} = \frac{\tau N^{\text{slot}}_{\text{CAP}}}{N^{\text{slot}}_{\text{state}}}$$

it follows that

$$N^{tx\_\text{slot}}_{\text{node}} = N^{tx}_{\text{node}} \left[(1 - P_c)N^{\text{slot}}_{\text{suc}} + P_c N^{\text{slot}}_{\text{col}}\right].$$

Observing that, among the nontransmission slots, the last slot in a CAP is followed by a CFP period, an inactive period, and a beacon period, $E[T^*_{\text{slot}}]$ is given by

$$E\left[T^*_{\text{slot}}\right] = \frac{\left(N^{\text{slot}}_{\text{CAP}} - N^{tx\_\text{slot}}_{\text{node}} - 1\right) T_{\text{slot}} + 1 \cdot T_o}{N^{\text{slot}}_{\text{CAP}} - N^{tx\_\text{slot}}_{\text{node}}}$$

where T_o is given in (1).

F. Performance Evaluation

To evaluate the throughput performance of the IEEE 802.15.4 MAC protocol, we consider a 2.4-GHz PHY layer and $D_R = 250$ kbps. A data frame consists of PHY and MAC headers and a MAC payload. Let the size of data frames L_{data} be 32 bytes, the size of PHY and MAC headers of data frames be 15 bytes, and the size of Ack frames L_{ack} be 11 bytes. The number of nodes in a network is 10. Each period is set as $T_{\text{BEP}} = 1 * 60 * 2^{SO}$ symbols, $T_{\text{CAP}} = 10 * 60 * 2^{SO}$ symbols, $T_{\text{CFP}} = 5 * 60 * 2^{SO}$ symbols, and $T_{\text{Inactive}} = 16 * 60 * 2^{BO-SO}$ symbols, respectively.

Fig. 2 shows the normalized system throughput S/D_R when we vary the normalized offered load per node λ_n $(= \lambda L_{\text{data}}/D_R)$ and the values of SO and BO. In Fig. 2, we consider the case of retransmissions without limit to compare our results with those of Mišić and Mišić. Note that we can obtain the results for retransmissions without limit by letting the maximum number of retransmissions R go to ∞. As shown in Fig. 2, the model of Mišić and Mišić fails to match the simulation results as the normalized offered load λ_n increases. There are some reasons for the mismatch of the model of Mišić and Mišić. First, they did not separately obtain the collision probabilities for nondeferred and deferred transmissions. Second, when calculating the average number of busy slots in a CAP to obtain the probability that the channel is idle at the first CCA, they did not consider the fact that there are no acknowledgements for collided transmissions, and they did not properly obtain the average number of transmissions in a CAP in the viewpoint of the network. Third, when obtaining the probability that the channel is idle at the second CCA, the condition that the channel is idle at the first CCA was not considered in the model of Mišić and Mišić. These factors affect the inaccurate results of the model of Mišić and Mišić. On the other hand, our model is well matched with the simulation results. For the normalized offered load λ_n smaller than 10^{-2}, since most of the transmissions are successful due to low traffic, the model of Mišić and Mišić seems to be well matched with the simulation results. As the value of BO increases with a fixed SO value, the normalized system throughput decreases for the value of λ_n larger than 10^{-2}. This is because the proportion of a CAP is getting smaller due to a longer inactive period under these traffic conditions. On the other hand, for the value of λ_n that is smaller than 10^{-3}, there is almost no difference in the system throughput, regardless of the duration of a beacon interval BI due to low traffic.

Fig. 3 shows the normalized system throughput S/D_R when we vary the normalized offered load per node λ_n and the data size. In this case, we consider retransmissions with limit $(R = 3)$. Let the values of SO and BO be set to 1. As the data size increases, the normalized system throughput increases. Since most of the transmissions are successful in a low traffic load such as λ_n smaller than 10^{-2}, large data frame sizes yield high throughput performance. In a high traffic load such as λ_n larger than 10^{-2}, if a node transmits a large data frame, it occupies the channel for a long time. This fact causes a low transmission probability and a low collision probability, which results in high throughput performance.

Fig. 4 shows the normalized system throughput S/D_R when we vary the number of nodes n and the value of the retransmission limit R. In this case, the normalized offered load per node λ_n is set to 0.03, the values of SO and BO are set to 1, and the maximum value of backoff stages m is set to 0. As the value of R increases, the normalized system throughput slightly increases for a low traffic load

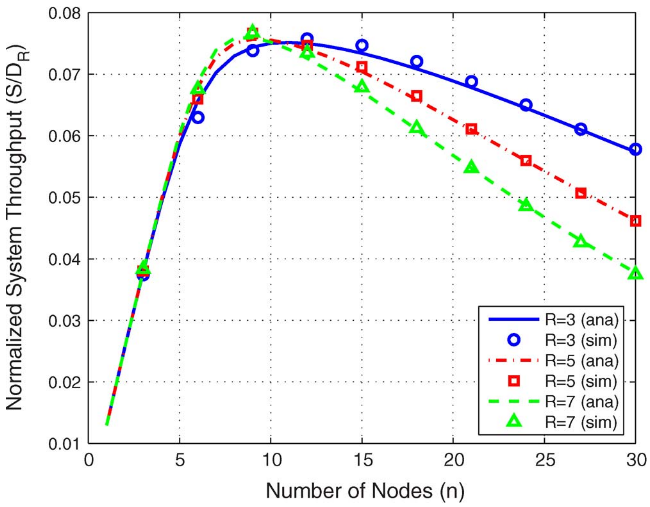

Fig. 4. Normalized system throughput for varying the value of the retransmission limit R.

such as n smaller than 6. Retransmissions decrease the probability of frame drops. However, for a high traffic load such as n larger than 12, the system throughput decreases as the value of R increases. A smaller value of R causes earlier frame drops. Since these frame drops reduce the number of contending nodes in the network, the collision probability decreases. This yields higher system throughput.

IV. Conclusion

In this paper, we have proposed an analytical Markov chain model of the slotted CSMA/CA in the IEEE 802.15.4 MAC protocol considering a superframe structure, acknowledgements, and retransmissions with and without limit under unsaturated traffic conditions. With the proposed model, we have evaluated the throughput performance of the slotted CSMA/CA. We have validated our proposed analytical model by simulation.

References

[1] *Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low Rate Wireless Personal Area Networks (LR-WPANs)* , IEEE Std. 802.15.4 Specification, Oct. 1, 2003.

[2] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications,* IEEE Std. 802.11 Specification, Jun. 26, 1997.

[3] G. Bianchi, "IEEE 802.11 saturation throughput analysis," *IEEE Commun. Lett.*, vol. 2, no. 12, pp. 318–320, Dec. 1998.

[4] T. R. Park, T. H. Kim, J. Y. Choi, S. Choi, and W. H. Kwon, "Throughput and energy consumption analysis of IEEE 802.15.4 slotted CSMA/CA," *Electron. Lett.*, vol. 41, no. 18, pp. 1017–1019, Sep. 2005.

[5] Z. Tao, S. Panwar, D. Gu, and J. Zhang, "Performance analysis and a proposed improvement for the IEEE 802.15.4 contention access period," in *Proc. IEEE WCNC*, Las Vegas, NV, Apr. 3–6, 2006, pp. 1811–1818.

[6] S. Pollin, M. Ergen, S. C. Ergen, B. Bougard, L. V. Perre, F. Catthoor, I. Moerman, A. Bahai, and P. Varaiya, "Performance analysis of slotted carrier sense IEEE 802.15.4 medium access layer," in *Proc. IEEE GLOBECOM*, San Francisco, CA, Nov. 27–Dec. 1, 2006, pp. 1–6.

[7] T. Lee, H. R. Lee, and M. Y. Chung, "MAC throughput limit analysis of slotted CSMA/CA in IEEE 802.15.4 WPAN," *IEEE Commun. Lett.*, vol. 10, no. 7, pp. 561–563, Jul. 2006.

[8] I. Ramachandran, A. K. Das, and S. Roy, "Analysis of Contention Access Period of IEEE 802.15.4," Univ. Washington, Seattle, WA, UWEE Tech. Rep. UWEETR-2006-0003, Feb. 2006.

[9] J. Mišić and V. B. Mišić, "Access delay for nodes with finite buffers in IEEE 802.15.4 beacon enabled PAN with uplink transmissions," *Comput. Commun.*, vol. 28, no. 10, pp. 1152–1166, Jun. 2005.

Efficient Sum Rate Maximization and Resource Allocation in Block-Diagonalized Space-Division Multiplexing

Boon Chin Lim, Witold A. Krzymień, *Senior Member, IEEE*, and Christian Schlegel, *Senior Member, IEEE*

***Abstract*—For space-division multiplexing (SDM) via block diagonalization on multiuser multiple-input multiple-output (MIMO) wireless downlink, it is shown that receive antenna selection (RAS) is necessary for maximizing the achievable sum rate. This is true even when all receive antennas are equipped with radio frequency (RF) chains and RAS reduces the upper bound on the broadcast sum capacity, and when the orthogonalized channels use optimal processing. Similarly, spatial-mode selection (SMS) is necessary for sum rate maximization when receive-weight matrices are used for spatial-mode allocation. RAS/SMS may release transmission resources that can fully be utilized via additional user scheduling to yield further sum rate gains. Optimal user selection for sum rate maximization is subsumed within an exhaustive RAS/SMS process for multiantenna terminals, and both selection processes become identical for single-antenna terminals. RAS/SMS thus helps reduce the performance gap from the optimal sum capacity even for small user pool sizes. A block antenna/mode selection approach is introduced to help overcome the drawbacks of existing algorithms. Since RAS/SMS involves antenna/mode ranking, a systematic method for resource allocation with sum rate loss minimization is inherently provided. This way, a streamlined process that combines user selection, RAS/SMS, and resource allocation is developed for sum rate maximization of block-diagonalized SDM.**

***Index Terms*—Block diagonalization, MIMO spatial multiplexing, multiuser MIMO downlink, receive-antenna selection, spatial-mode selection, sum rate maximization.**

I. Introduction

For a wireless base station (BS) equipped with M antennas where coordination is feasible among the transmit chains but not among the K mobile user terminals, simultaneous downlink transmissions to multiple users are possible when channel state information is available at the transmitter. The optimal approach for sum rate maximization is dirty paper coding (DPC) [1], which is very complex, and beamforming is a reduced-complexity alternative. For a system with N receive antennas per user, the optimal beamforming sum rate scales as for DPC, i.e., with $M \log\log KN$ [2], when K is very large and scheduling is applied. However, optimal beamforming involves signal-to-interference-plus-noise ratio balancing and is also quite complex. A suboptimal beamforming method is zero-forcing beamforming (ZFBF), which enforces zero interference among spatial layers (streams). For multiantenna terminals, it is better to impose orthogonality between users only (cancel only multiuser interference), because antennas located at the same terminal can effectively cooperate. This is commonly referred to as block diagonalization (BD) [3], [4]. Although inferior when compared with optimal beamforming, the

Manuscript received June 11, 2007; revised December 6, 2007, February 15, 2008, and March 20, 2008. First published May 7, 2008; current version published January 16, 2009. This work was supported in part by DSO National Laboratories (Singapore), by TRLabs, by the Rohit Sharma Professorship, and by the Natural Sciences and Engineering Research Council of Canada. The review of this paper was coordinated by Dr. C. Yuen.

B. C. Lim and C. Schlegel are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2R3, Canada (e-mail: bclim@ece.ualberta.ca; schlegel@ece.ualberta.ca).

W. A. Krzymień is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2R3, Canada, and also with TRLabs, Edmonton, AB T6G 2W3, Canada (e-mail: wak@ece.ualberta.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2008.924973