Fast Node Cardinality Estimation and Cognitive MAC Protocol Design for Heterogeneous M2M Networks

Sachin Kadam, Chaitanya S. Raut and Gaurav S. Kasbekar

***Abstract*—Machine-to-Machine (M2M) networks are an emerging technology with applications in numerous areas including smart grids, smart cities, vehicular telematics, healthcare, security and public safety. In this paper, we design a medium access control (MAC) protocol that supports multi-channel operation for a heterogeneous M2M network, with three types of M2M devices (*e.g.*, those that send emergency, periodic and normal type data), operating as a secondary network using Cognitive Radio technology. Also, we design an estimation protocol for rapidly obtaining separate estimates of the number of active nodes of each traffic type, and use these estimates to find the optimal contention probabilities to be used in the Cognitive MAC protocol. We compute a closed form expression for the expected number of time slots required by our estimation protocol to execute as well as a simple upper bound on it, which shows that the expected number of time slots required by our protocol to obtain the above estimates is small. Also, we mathematically analyze the performance of the Cognitive MAC protocol and obtain expressions for the expected number of successful contentions and the expected amount of energy consumed per frame. Finally we evaluate the performance, in terms of average throughput and average delay, of our MAC protocol using simulations.**

I. Introduction

Machine-to-Machine (M2M) communications is an emerging technology, in which data generation, processing and transmission is done with minimal human intervention [1]. M2M networks have applications in numerous areas including smart grids, smart cities, vehicular telematics, healthcare, industrial automation, security and public safety [1]–[3]. It is challenging to design medium access control (MAC) protocols for M2M networks due to their unique characteristics such as limited access to energy sources (most M2M devices are battery operated), need to provide network access to a very large number of devices, the fact that the Quality of Service (QoS) requirements of M2M devices differ from those of Human-to-Human (H2H) communications and are also different for different M2M devices [1] etc [1], [3], [4].

Several wireless technologies such as Bluetooth, Wi-Fi, ZigBee and cellular networks including LTE-Advanced and 802.16 are potential candidates for enabling M2M communications; however, these technologies have some shortcomings [2] [5]. Cognitive Radio technology is a promising alternative to the above wireless technologies for enabling M2M communications [2]. Cognitive Radio Networks (CRNs) have emerged as a promising solution to alleviate the artificial spectrum scarcity (wherein most of the usable radio spectrum is allocated, but underutilized) caused by the traditional spectrum regulation policy of assigning *exclusive* licenses to users to operate their networks in different geographical regions [6]. In CRNs, there are two types of spectrum users– primary users (PUs), which have prioritized access to channels, and secondary users (SUs) that detect and use spectrum holes, *i.e.*, chunks of spectrum that are currently not in use by the PUs [6]. Operating an M2M network as a secondary network using Cognitive Radio technology has the advantage that a large amount of spectrum, which is allocated to other users, but underutilized, becomes available for M2M communications [5]. However, this requires the design of efficient *Cognitive MAC protocols* in order to provide channel access to an extremely large number of M2M devices, while satisfying the unique service requirements of M2M applications described in the first paragraph of this section, as well as ensuring avoidance of interference to PUs. The design of a Cognitive MAC protocol *that supports multi-channel operation* involves addressing additional challenges [7] including achieving coordination among nodes [3], overcoming the multi-channel hidden terminal problem [8], and balancing the traffic load of the secondary (M2M) nodes over the free channels in real-time [9]. In this paper, *we design a Cognitive MAC protocol for M2M networks that overcomes the above challenges.*

Now, consider an M2M network in which a large number of M2M devices intermittently transmit some information (*e.g.*, smart meter readings, information collected by sensors) to a base station (BS). In any given time frame, the BS is unaware of the number of *active* nodes, *i.e.*, those that need to transmit some data to the BS in the current frame. There is a need to rapidly estimate the number of active nodes since this estimate can be used to determine the optimal values of various parameters of the MAC protocol such as contention probabilities and the amounts of time to be used for contention and for data transmission in the current frame. For example, recall that for the Slotted ALOHA protocol, the optimal contention probability is the reciprocal of the number of active nodes [10]. Also, in [11]–[13], cardinality estimation

---

The authors are with the Department of Electrical Engineering, Indian Institute of Technology (IIT) Bombay, Mumbai 400076, India. Their email addresses are {sachink, csraut, gskasbekar}@ee.iitb.ac.in.

[1] For example, some M2M nodes need to transmit data (*e.g.*, smart meter readings) periodically, some need it for sending emergency or alarm messages (*e.g.*, in healthcare and security applications), some need it for transmission of normal data traffic and some need it for reliable transmission of data packets (*e.g.*, in remote payment gateway systems) [1], [3], [4].

[2] Specifically, Wi-Fi has high power consumption, due to which it is not suitable for battery operated M2M devices, and Bluetooth has high latency when the number of devices is large, as is the case in M2M networks [5]. ZigBee operates on unlicensed bands and is prone to interference from Wi-Fi networks and other equipment (*e.g.*, microwave ovens) that use those bands [2], [5]. Due to the high demand for H2H communication services such as voice, video, emails etc, only a limited amount of radio spectrum is available with cellular operators to support M2M communications [2].

[3] Note that for two nodes to be able to exchange data, both must have their wireless transceiver tuned to a common channel at a time.

is performed and using the estimates obtained, the contention probabilities that maximize the throughput of their respective MAC protocols for M2M networks are determined.

In a *heterogeneous* M2M network, *i.e.*, one in which different types of nodes are present (*e.g.*, those that send emergency, periodic and normal type data), we need to obtain *separate* estimates of the number of active nodes of each traffic type. In prior work, several protocols have been designed [11]–[13] to estimate the number of active nodes of a *homogeneous* M2M network (see Section II). However, to the best of our knowledge, so far *no estimation protocol has been designed for obtaining separate estimates of the number of active nodes of each traffic type in a heterogeneous M2M network*. Note that executing a cardinality estimation protocol for a homogeneous M2M network multiple times to do this is inefficient.

In this paper, we consider an M2M network with three types of nodes, which we refer to as Type 1, Type 2 and Type 3 nodes; *e.g.*, these may be emergency, periodic and normal data type nodes. We design an estimation protocol to rapidly obtain separate estimates of the number of active nodes of each traffic type (see Section III). We compute a closed form expression for the expected number of time slots required by our estimation protocol to execute as well as a simple upper bound on it, which shows that the expected number of time slots required by our protocol to obtain the above estimates is small. Next, we use our estimation protocol as part of a Cognitive MAC protocol that we design for heterogeneous M2M networks (see Section IV). In the proposed MAC protocol, time is divided into frames of equal duration, with each frame containing an estimation window, a contention window (CW) and a data transmission window (DTW). Whenever a node succeeds in contention on a given channel during the CW, the BS reserves the requested number of time slots for data transmission by that node in the DTW. Slotted ALOHA [10] is used for contention in the CW, and the contention probability used by each node is the reciprocal of the estimated number of contending nodes on the channel; *thus, the estimates obtained using our estimation protocol are used for optimizing the contention probabilities*. We mathematically analyze the performance of the proposed MAC protocol and obtain expressions for the expected number of successful contentions and the expected amount of energy consumed per frame (see Section V). Finally, using simulations, we evaluate the performance, in terms of average throughput and average delay, of our MAC protocol and compare it with that of a hypothetical "ideal protocol", which is assumed to know the exact number of active nodes at any time (see Section VI).

II. Related Work

A scheme to estimate the number of active nodes in M2M networks is proposed in [12]. In the proposed method every device selects a slot uniformly at random from a set of slots and transmits a Power Save-poll message. The access point (AP) estimates the number of nodes by using the number of empty slots with the maximum likelihood (ML) estimation method. In [11] an iterative method is proposed for estimation of nodes and is obtained using drift analysis on the access results of consecutive empty and collision slots of the past slots. In [13] estimation of nodes is carried out in two phases. (i) Coarse phase: In this phase every node sends a busy tone with probability $(1/2)^i$ in slot i. This process continues till all nodes stop sending busy tones in some slot j. Average length of this phase is $\log_2 N$ where N is number of active nodes. (ii) Refine phase: In this phase each node sends a busy tone with probability used by node to send last busy tone in the coarse phase. Average length of this phase depends on accuracy requirement. In this estimation scheme, it uses only one control channel for node estimation due to which all other channels remain unused during estimation phase. In the proposed estimation scheme, we use all the available channels for node estimation due to which the efficiency of the scheme is improved.

Tag cardinality estimation methods are extensively discussed in literature of Radio-frequency identification (RFID) systems. Most of the tag cardinality estimation schemes aim to achieve the following accuracy requirement.

$$Pr\{|\hat{n}-n| \leq \varepsilon n\} \geq 1-\delta, \tag{1}$$

where n, $\hat{n}$ are actual and estimated number of tags respectively, $|\varepsilon|$ is confidence interval and δ is error probability. The scheme proposed in [14] tags select a slot uniformly at random and transmit a packet in the selected slot. The number of empty slots and collision slots are counted by the reader (server) and these values are used to obtain the zero estimator (ZE) and collision estimator (CE) respectively. More accurate among ZE and CE is chosen as the estimator of tags and it is called as unified probabilistic estimator (UPE). Limitations of this scheme are all tags must be readable in any given iteration and approximate number of tags need to inform the server. These limitations are addressed in [15], which uses only empty slots for the cardinality estimation process and estimator is called as enhanced zero based estimator (EZB). In both the described estimation algorithms require large number of slots. Energy efficient cardinality estimation process is proposed in [16]. At every polling server sends a request packet with contention probability p along with frame size f. Any user can poll in one of the slots with contention probability p. Polling stops when the accuracy requirement described in (1) gets fulfilled. Three different cardinality estimation algorithms; Maximum Likelihood Estimation Algorithm (MLEA), Average Sum Estimation Algorithm (ASEA) and Enhanced MLEA are proposed in [16]. In MLEA optimised contention probability $p_i = (1.594/\hat{n}_{i-1})$ and in ASEA $p_i = (1/\hat{n}_{i-1})$, where $\hat{n}_i$ is estimated number of tags in i^{th} iteration, are used. Basic cardinality estimation algorithm presented in [17] require $\mathcal{O}(\log n)$ slots. Optimized version of the same algorithm require $\mathcal{O}(\log\log n)$ slots which also satisfy accuracy requirement of (1). In [18] RFID tag cardinality estimation scheme is based on new distinct element counting method described in [19]. Every tag has a counter corresponding to a random number. After every time slot counter decreases by one and tag transmits when counter becomes zero. Here server need not scan the entire estimation window. Server informs frame length f to every node. It awaits response from every node. If k be number of waiting slots then $k+\log_2 f$ are required number of slots to achieve the accuracy described in (1). Reader observes the position of empty (0) and non empty (1) slots. First Non-Empty slots Based (FNEB) estimator is used to estimate the number of tags. 7 times faster estimation scheme than UPE and EZB is proposed in [20] and it is called Average Run based Tag (ART) cardinality estimation method which uses slotted ALOHA protocol. Reader sends a packet to all tags indicating frame size. Every tag picks a slot randomly to respond. Empty (0) and non empty (1) responses are collected by reader, cardinality estimation is done based on the average run size of 1s in the obtained binary sequence.

However, all the node cardinality estimation schemes studied in the above papers are for a *homogeneous* network, wherein all nodes are alike. In contrast, in this paper, we propose an estimation scheme for a *heterogeneous* network with three different types of nodes, which efficiently computes separate estimates of the number of active nodes of each type.

Extensive surveys related to MAC protocol design for M2M networks are provided in [3], [21]. In [22], a hybrid MAC protocol that uses contention-based channel access (CSMA/CA) when the network load is low and reservation-based access when the load is high is proposed. In [23], a hybrid MAC protocol, in which each time frame consists of a contention period followed by a transmission period, is proposed. The devices that successfully contend in the contention period are assigned a time slot for data transmission in the transmission period. Extension of protocol proposed in [23] is extended to heterogeneous M2M networks in [24], where service requirements of different devices are different along with different priorities. In [8], the MAC protocol is proposed for multichannel ad hoc networks and this is modified in [13] to suite for M2M networks. Time is divided into frames and they are further divided into 3 phases namely estimation phase, contention phase and data transmission phase. Number of active users are estimated in estimation phase. In the contention phase, all active users tune to a common control channel and contend for channel access with contention probabilities which are obtained with the help of number of estimated nodes. The nodes which are successful in contention transmit their data packets in data transmission phase simultaneously on different channels. In the protocol proposed in [25], time is divided into slots, and in each slot, nodes contend with contention probability (which is the statistical estimate of present traffic load), using a Request to Send (RTS) packet and it is responded with Clear to Send (CTS) packet, followed by transmission of a data packet. In [12], the 802.11ah MAC protocol is modified for M2M communications as follows: first estimation of the number of active M2M devices is done and this estimation is used to adapt the length of the Restricted Access Window, in which only M2M devices are allowed to contend. In [11], a modified version of the Slotted-ALOHA scheme is presented, in which results of the previous slots are considered to estimate the transmission attempt probability of the current slot which results in maximising the throughput. In [26], an overload control mechanism is presented for M2M communication in LTE-Advanced networks, in which based on the traffic load on the random access channel (RACH) base station adjusts the number of RACH resources.

However, to the best of our knowledge, our Cognitive MAC protocol is the first to employ separate estimates of the numbers of active nodes of different types for selecting the optimal contention probabilities in a heterogeneous M2M network.

III. Fast Node Cardinality Estimation Scheme

In this section, we present our fast node cardinality estimation scheme for heterogeneous M2M networks. The estimation problem is defined in Section III-A. In Section III-B, the Lottery Frame (LoF) based protocol [27], [28], which is a cardinality estimation scheme for homogeneous networks, and which we extend to estimate cardinalities in heterogeneous M2M networks, is briefly described. Our proposed fast estimation scheme is described in Section III-C. A closed form expression for the expected number of time slots required by our estimation protocol to execute is computed in Section III-D and a simple upper bound on it is established in Section III-E.

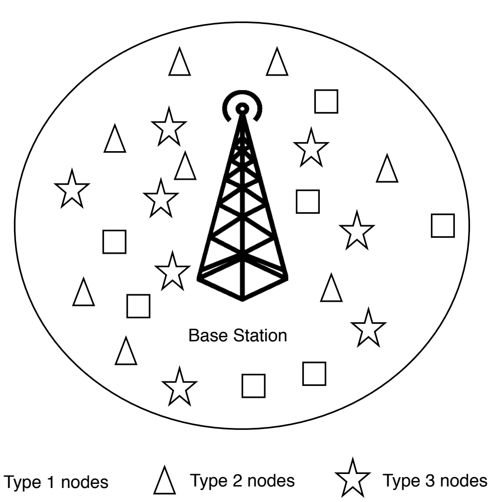

Figure 1. The figure shows a base station and three types of nodes within its range (the area inside the circle).

A. The Estimation Problem

Consider a heterogenous M2M network consisting of a base station (BS) and three different types, say Type 1, Type 2 and Type 3, of M2M devices (nodes) in its range as shown in Figure 1. We denote the sets of nodes of Type 1, Type 2 and Type 3 as $\mathscr{N}_1$, $\mathscr{N}_2$ and $\mathscr{N}_3$ respectively; let [4] $|\mathscr{N}_b| = n_b$, $b \in \{1,2,3\}$. Time is divided into frames of equal duration, and in each frame only a subset of the nodes of each type are *active*, *i.e.*, have data to send to the BS. Also, each frame is divided into time slots of equal durations. Let n_b be the number of active nodes of Type b, $b \in \{1,2,3\}$, in a given frame. Our objective is to design an estimation protocol to estimate the values of n_1, n_2 and n_3 rapidly, *i.e.*, using a small number of time slots.

B. Review of the LoF Based Protocol

The LoF based estimation protocol was designed in [27] and uses the probabilistic bitmap counting technique proposed in [28] for tag cardinality estimation in RFID systems. The LoF based estimation protocol is designed for a homogeneous network. Our proposed protocol extends the LoF based protocol to a heterogeneous network with three types of nodes. So we provide a brief review of the LoF based protocol in this subsection.

Every tag (equivalent to a node in M2M networks) has a unique binary identification (ID) number that is l bits in length. The hash value h of any tag is defined as the 'position

---

[4] $|A|$ denotes the cardinality of set A.

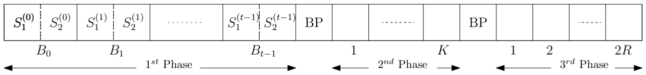

Figure 2. The figure shows the structure of the Estimation Window. “BP” denotes a Broadcast Packet.

of the least significant zero bit' in its ID. For example, $h(01001001)=1$ and $h(00101111)=4$, where $h(I)$ denotes the hash value corresponding to ID I. So if h is the hash value of a random tag, then assuming that each of the l bits of the corresponding ID independently equal 0 or 1 with probability 1/2 each, $P(h=i)=1/2^{(i+1)}$, $i=0,1,2,\dots l-1$ and [5] $P(h=l)=1/2^l$.

Now, time is divided into slots of equal duration. During the estimation process, each active tag with hash value h transmits a packet in the h^{th} time slot, for $h=0,1,2,\dots,l$. A corresponding bitmap (BM) of 0s and 1s is generated by the RFID system reader (equivalent to the BS in M2M networks) based on the slot results; the h^{th} bit of the BM is 0 if the h^{th} time slot is empty (*i.e.*, one in which no node transmits) and 1 if the h^{th} time slot is non-empty (*i.e.*, one in which one or more nodes transmit). Let $\rho=\min\{h|BM(h)=0\}$; then the estimated value of n (the actual number of active tags) is $\hat{n}=1.2897\times 2^{\rho}$ [27]. It is also proved in [27] that the LoF based protocol executes within $\log_2 n_{all}$ slots, where n_{all} is the total number of all possible binary IDs.

Note that if the LoF based protocol is executed thrice to separately estimate n_1, n_2 and n_3 in the network model for M2M networks described in Section III-A, then $\log_2(n_{1,all}n_{2,all}n_{3,all})$ slots are required, where $n_{b,all}$ is the total number of all possible binary IDs of the b^{th} type of nodes. To reduce the number of slots, we propose a fast estimation scheme, which is described in the following subsection.

C. Proposed Estimation Scheme

The estimation process is carried out in three phases, which we describe in Sections III-C1, III-C2 and III-C3. We refer to the set of slots used during the estimation process as the Estimation Window (EW). The structure of a typical EW is shown in Figure 2. At the end of the estimation process, separate estimates, say $\hat{n}_1,\hat{n}_2$ and $\hat{n}_3$, of the number of active nodes of the three types, n_1,n_2 and n_3 (see Section III-A), are obtained. For each $b\in\{1,2,3\}$, the estimate $\hat{n}_b$ *is equal to (and hence, as accurate as)* the estimate of n_b that would have been obtained if the LoF protocol [27], [28] were used for the estimation. However, note that the total number of time slots used in our estimation scheme is much smaller than the number of time slots that would have been required if the LoF protocol were separately executed thrice to estimate n_1,n_2 and n_3.

At a high level, our estimation scheme operates as follows. Let $t=\log_2(\max(n_{1,all},n_{2,all},n_{3,all}))$. Also, for $b\in\{1,2,3\}$ and $i\in\{0,1,\dots,t-1\}$, let $B_p(b,i)$ be 1 (respectively, 0) if the i^{th} slot would have been non-empty (respectively, empty) if the LoF protocol were used to estimate the number of active nodes of Type b. From Section III-B, it is clear that if the bit patterns $B_p(b,i)$, $b\in\{1,2,3\}$, $i\in\{0,1,\dots,t-1\}$, are known, then the LoF estimates $\hat{n}_1,\hat{n}_2$ and $\hat{n}_3$, of n_1,n_2 and n_3 respectively, can be deduced. In our estimation scheme, the bit patterns $B_p(b,i)$, $b\in\{1,2,3\}$, for most values of i are found in the first phase; ambiguity about the rest remains, which is resolved in the second and third phases.

1) First Phase: In the first phase, $2t$ slots are used. Every two consecutive slots constitute a block (see Figure 2); let B_i denote the i^{th} block. In block B_i, $i\in\{0,1,\dots,t-1\}$, active nodes from $\mathscr{N}_1$ whose hash value is i send a packet containing the symbol α in both the slots of B_i. Also, in block B_i, active nodes from $\mathscr{N}_2$ (respectively, $\mathscr{N}_3$) whose hash value is i send a packet containing the symbol β only in the first slot (respectively, only in the second slot) of B_i. So every slot has four possible outcomes, which are as follows: (i) Empty (E) if no node transmits in the slot, (ii) Collision (C) if two or more nodes transmit, (iii) α if exactly one node of Type 1 transmits, (iv) β if exactly one node of Type 2 or Type 3 transmits. The possible outcomes in a block are shown in the first two columns of Table I [6]. Note that for $b\in\{1,2,3\}$, $i\in\{0,1,\dots,t-1\}$, $B_p(b,i)$ equals 1 (respectively, 0) if and only if atleast one node (respectively, no node) of Type b transmits in block B_i. The bit patterns $B_p(b,i)$, $b\in\{1,2,3\}$ corresponding to each possible block outcome are shown in the last three columns of Table I. For example if Slot1 results in C and Slot2 results in α, then it implies that exactly one node from $\mathscr{N}_1$, at least one node from $\mathscr{N}_2$ and none from $\mathscr{N}_3$ have transmitted. Similarly if both the slots result in β, then it implies that exactly one node each from $\mathscr{N}_2$ and $\mathscr{N}_3$, and none from $\mathscr{N}_1$ have transmitted. The outcome (C, C) in which collisions occur in both the slots may be due to transmissions by atleast two nodes from $\mathscr{N}_1$, or by one node from $\mathscr{N}_1$ and at least one node each from $\mathscr{N}_2$ and $\mathscr{N}_3$, or by atleast two nodes each from $\mathscr{N}_2$ and $\mathscr{N}_3$. Due to the above ambiguity, if the outcome (C, C) occurs in the i^{th} block, then the second phase is used to find the bit patterns $B_p(1,i)$, $B_p(2,i)$ and $B_p(3,i)$. Let C_I be the set of block numbers i in which the ambiguous outcome (C, C) has occurred. A broadcast packet (BP) is sent by the BS after the first phase (see Figure 2), which contains a list of the block numbers in set C_I.

Outcome in Block i		Bit patterns		
Slot1	Slot2	$B_p(1,i)$	$B_p(2,i)$	$B_p(3,i)$
E	E	0	0	0
E	C	0	0	1
E	β	0	0	1
C	E	0	1	0
C	C	$\star$	$\star$	$\star$
C	α	1	1	0
C	β	0	1	1
α	C	1	0	1
α	α	1	0	0
β	E	0	1	0
β	C	0	1	1
β	β	0	1	1

Table I. E, C and $\star$ denote “Empty”, “Collision” and “ambiguous result” respectively.

2) Second Phase: In the second phase, only the nodes from $\mathscr{N}_1$ whose hash value belongs to the set C_I participate. Specifically, for each $j=1,2,\dots,|C_I|$, in the j^{th} slot of the second phase, the nodes from $\mathscr{N}_1$ whose hash value equals the block number of the j^{th} block whose outcome was (C, C)

---

[5] If all the bits of the ID are 1, then its hash value is defined to be l.

[6] Note that the block results (E, α), (α, E), (α, β) and (β, α) cannot occur under the above protocol.

in the first phase transmit. Nodes from $\mathscr{N}_2$ and $\mathscr{N}_3$ do not transmit in the second phase.

Now, consider the slot in the second phase corresponding to the i^{th} block in the first phase, where $i \in C_I$. If the slot result is empty, then it follows that $B_p(1,i) = 0$, $B_p(2,i) = 1$ and $B_p(3,i) = 1$; also, if the slot result is one packet transmission, then $B_p(1,i) = 1$, $B_p(2,i) = 1$ and $B_p(3,i) = 1$. If the slot result is C, then $B_p(1,i) = 1$; however, ambiguity about the values of $B_p(2,i)$ and $B_p(3,i)$ still remains and it is resolved in the third phase. Let $C_{II} \subseteq C_I$ be the set of block numbers i for which a collision occurred in the second phase. A BP is sent by the BS after the second phase (see Figure 2), which contains a list of the block numbers in set C_{II}.

3) Third Phase: In this phase, only those active nodes from $\mathscr{N}_2$ and $\mathscr{N}_3$ participate, whose corresponding blocks in the first phase and corresponding slots in the second phase resulted in collisions. That is, the active nodes from $\mathscr{N}_2$ and $\mathscr{N}_3$ whose hash value belongs to C_{II} participate. The odd (respectively, even) numbered slots of the third phase are used by nodes from $\mathscr{N}_2$ (respectively, $\mathscr{N}_3$). Specifically, for each $j = 1, 2, \ldots, |C_{II}|$, in slot $2j-1$ (respectively, $2j$) of the third phase, the active nodes from $\mathscr{N}_2$ (respectively, $\mathscr{N}_3$) whose hash value equals the first phase block number, say i, of the j^{th} element of C_{II} transmit. If slot $2j-1$ is empty, then $B_p(2,i) = 0$, else $B_p(2,i) = 1$. Similarly, if slot $2j$ is empty, then $B_p(3,i) = 0$, else $B_p(3,i) = 1$. Also, since $B_p(1,i) = 1$, the above ambiguity is resolved in the third phase.

D. Determination of Expected Number of Time Slots Required by Estimation Protocol to Execute

Recall that $2t$ slots are required in the first phase. Let K (respectively, $2R$) be the number of slots required in the second phase (respectively, third phase).

1) Determination of $E[K]$: Note that $0 \le K \le t$. Let $S_1^{(i)}$ (respectively, $S_2^{(i)}$) represent the result of the first (respectively, second) slot of B_i. Also, let $I_{\mathscr{F}}$ denote the indicator random variable corresponding to event $\mathscr{F}$, *i.e.*,

$$I_{\mathscr{F}} = \begin{cases} 1, & \text{if } \mathscr{F} \text{ occurs,} \\ 0, & \text{else.} \end{cases}$$

Clearly, $K = \sum_{i=0}^{t-1} I_{\{S_1^{(i)}=C, S_2^{(i)}=C\}}$. So:

$$E[K] = \sum_{i=0}^{t-1} P(S_1^{(i)} = C, S_2^{(i)} = C). \qquad (2)$$

The conditions under which collisions occur in both the slots of B_i are as follows: 1) At least two nodes from $\mathscr{N}_1$ transmit in B_i, 2) Exactly one node from $\mathscr{N}_1$ and at least one node each from $\mathscr{N}_2$ and $\mathscr{N}_3$ transmit in B_i, 3) At least two nodes each from $\mathscr{N}_2$, $\mathscr{N}_3$ and none from $\mathscr{N}_1$ transmit in B_i. Let $Q_1(i)$, $Q_2(i)$ and $Q_3(i)$ denote the probabilities of the events in 1), 2) and 3) respectively. Then:

$$P(S_1^{(i)} = C, S_2^{(i)} = C) = Q_1(i) + Q_2(i) + Q_3(i). \qquad (3)$$

It is easy to show that $Q_1(i) = 1 - u(n_1) - v(n_1)$, $Q_2(i) = v(n_1)(1-u(n_2))(1-u(n_3))$ and $Q_3(i) = u(n_1)(1-u(n_2) - v(n_2))(1-u(n_3)-v(n_3))$, where $u(n) = (1-(1/2)^{i+1})^n$ and $v(n) = n(1/2)^{i+1}(1-(1/2)^{i+1})^{n-1}$. By (2) and (3), the expected number of slots required in the second phase is:

$$E[K] = \sum_{i=0}^{t-1} \{Q_1(i) + Q_2(i) + Q_3(i)\}. \qquad (4)$$

2) Determination of $E[R]$: Note that $0 \le R \le K$. It is easy to show that:

$$E[R] = \sum_{i=0}^{t-1} Q_1(i). \qquad (5)$$

The expected total number of slots required by the estimation protocol to execute is $2t + 2 + E[K] + 2E[R]$ (see Figure 2), where $E[K]$ and $E[R]$ are given by (4) and (5) respectively.

E. Upper Bound on Expected Number of Time Slots Required by Estimation Protocol to Execute

Although the expressions derived in Section III-D are exact, they are complicated. So to gain insight, in this subsection, we provide simple upper bounds on $E[K]$ and $E[R]$ and use them to obtain an upper bound on the expected total number of slots required by the estimation protocol to execute. Let $n_m = \max(n_1, n_2, n_3)$, $\lceil x \rceil =$ the smallest integer greater than or equal to x, and $l_y = \lceil (\log_2 y) \rceil$.

Theorem 3.1: $E[K] \le l_{n_m} - 1 + \frac{2n_1^2}{3n_m^2}\left[1 + \frac{2n_2^2 n_3^2}{5n_1^2 n_m^2} + \frac{12 n_2 n_3}{7 n_1 n_m}\right]$.

Theorem 3.2: $E[R] \le l_{n_m} - 1 + \frac{2n_1^2}{3n_m^2}$.

The proofs of Theorems 3.1 and 3.2 are provided Appendix. As an example, consider the case where $n_1 = n_2 = n_3 = n$ (say) and $n_{1,all} = n_{2,all} = n_{3,all}$. Then the expected total number of slots required by the estimation protocol to execute is bounded by $2t + 2 + E[K] + 2E[R] \le 2t + 2 + l_n + 1.076 + 2(l_n - 0.333) = 2t + 3l_n + 2.41$. Hence, *the number of time slots saved compared with the case where the LoF protocol is executed thrice to separately estimate n_1, n_2 and n_3 is at least* $3t - (2t + 3l_n + 2.41) = t - 3l_n - 2.41$.

IV. COGNITIVE MAC PROTOCOL FOR MULTI-CHANNEL M2M NETWORKS

A. Overview

For concreteness, we henceforth assume that Type 1, Type 2 and Type 3 nodes are emergency, periodic and normal data nodes respectively. Time is divided into frames of equal durations. Let M_T be the total number of channels and q_i be the probability that a primary user (PU) is present on channel $i \in \{1, \ldots, M_T\}$ in any given frame [7]. Also, in a frame, suppose there are M_f free channels, say $\{a_1, a_2, \ldots, a_{M_f}\}$; then out of these, we use the M_{f_e} channels with the lowest values of q_i for emergency data traffic, the M_{f_p} channels with the next lowest values of q_i for periodic data traffic and the rest for normal data traffic, for some M_{f_e}, M_{f_p}. We assume that all M2M devices are in the range of the base station (BS) (see Figure 1). In each frame, only the BS senses the M_T channels

---

[7] We assume that the probabilities q_i are known to the base station; for example, they can be estimated using past observations of PU occupancies on different channels.

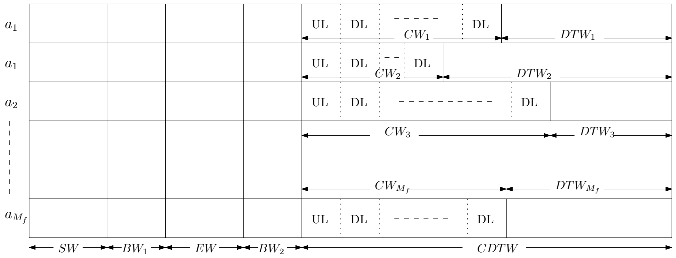

Figure 3. The figure shows the structure of a frame. Only the free channels are shown.

to check for the presence of PUs [8]. Figure 3 shows the structure of a frame. The BS senses every channel in the sensing window (SW) to check for the presence of PUs. In the first broadcast window (BW_1), all the active nodes receive the list of channels that are free in the current frame from the BS (see Section IV-B). The fast node cardinality estimation scheme described in Section III is executed in the estimation window (EW) to estimate the number of active nodes of each type (see Section IV-C). In the second broadcast window (BW_2), the list of channels assigned to each type of node and the optimal contention probabilities (which are computed using the estimates obtained in the EW) are broadcast by the BS (see Section IV-D). In the Contention and Data Transmission Window (CDTW), active nodes contend on the channels assigned to them using Slotted ALOHA [10]; upon each successful contention, the BS reserves the requested number of slots for data transmission by the node in the DTW (see Section IV-E). Periodic nodes require channels for periodically transmitting data. In particular, when a periodic node r with T_r data packets contends successfully, the BS reserves one slot each in T_r successive frames for data transmissions by node r. Node r does not participate again in the contention process in these T_r frames.

B. First Broadcast Window (BW_1)

The BS and every node store the list of all channels, sorted in increasing order of q_i. In BW_1, the BS repeatedly broadcasts a packet on the first free channel (say m_f) of the above list; this packet contains the list of channels that are free in the current frame. Each active node tunes to channels in increasing order of q_i, listening for one time slot on each channel, until it tunes to channel m_f and receives the list broadcast by the BS.

C. Estimation Window (EW)

Recall that the fast node cardinality estimation scheme described in Section III requires $2t+2+|C_I|+2|C_{II}|=R_s$ (say) slots to execute. R_s slots are reserved [9] in the EW for the estimation process. In the EW, *all the M_f free channels in the frame are utilized* for the estimation. The scheme used for numbering the reserved R_s slots in the multi-channel environment is shown in Figure 4.

a_1	1	M_f+1	.	.	.	.	.
a_2	2	M_f+2	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	R_s-1
.	.	.	.	.	.	.	R_s
.	.	.	.	.	.	.	
.	.	.	.	.	.	.	
a_{M_f}	M_f	$2M_f$	.	.	.	.	

Figure 4. The figure shows the scheme used for numbering the reserved R_s slots in the EW. The first slot of channel a_1 is numbered 1, the first slot of channel a_2 is numbered 2, …, the first slot of channel a_{M_f} is numbered M_f, the second slot of channel a_1 is numbered M_f+1 and so on.

D. Second Broadcast Window (BW_2)

After the EW, the BS knows the estimated numbers of active nodes with emergency ($\hat{n}_e$), periodic ($\hat{n}_p$) and normal ($\hat{n}_n$) data packets. Based on the values of $\hat{n}_e$, $\hat{n}_p$ and $\hat{n}_n$, out of the M_f free channels, M_{f_e}, M_{f_p} and M_{f_n} channels are assigned to emergency, periodic and normal data nodes respectively, where $M_f = M_{f_e} + M_{f_p} + M_{f_n}$; the BS broadcasts the lists of channels assigned to each type of node in BW_2. For instance, let w_e, w_p and w_n be *weights* (positive real numbers) associated with the emergency, periodic and normal data classes respectively. Then M_{f_e},M_{f_p} and M_{f_n} may be selected to be approximately $\frac{\hat{n}_e w_e M_f}{(\hat{n}_e w_e + \hat{n}_p w_p + \hat{n}_n w_n)}$, $\frac{\hat{n}_p w_p M_f}{(\hat{n}_e w_e + \hat{n}_p w_p + \hat{n}_n w_n)}$ and $\frac{\hat{n}_n w_n M_f}{(\hat{n}_e w_e + \hat{n}_p w_p + \hat{n}_n w_n)}$ respectively. We use $w_e \geq w_p \geq w_n$ to ensure that emergency (respectively, periodic) data is provided a higher priority than periodic (respectively, normal) data. To balance the load across the assigned channels, each emergency (respectively, periodic, normal) node selects one channel from the M_{f_e} (respectively, M_{f_p}, M_{f_n}) channels at random and tunes to it in the CDTW. Now, recall that if n nodes contend using Slotted ALOHA, then the value of the contention probability p that maximizes the throughput is $p = 1/n$ [10]. So the BS sets the probabilities of contention for emergency, periodic and normal data nodes to $\hat{p}_e = \min(M_{f_e}/\hat{n}_e, 1)$, $\hat{p}_p = \min(M_{f_p}/\hat{n}_p, 1)$ and $\hat{p}_n = \min(M_{f_n}/\hat{n}_n, 1)$ respectively and broadcasts the values of $\hat{p}_e$, $\hat{p}_p$ and $\hat{p}_n$ in BW_2. Finally, there may be some periodic data nodes with time slots in the DTW of the current frame reserved in past frames; a packet containing a list of such reserved slots is also broadcast by the BS in BW_2.

E. Contention and Data Transmission Window (CDTW)

After BW_2, all nodes switch to their respective selected channels for contention and data transmission. Every channel in this window is divided into a Contention Window (CW) and a Data Transmission Window (DTW) of variable lengths (see Figure 3). Each active node from $\mathcal{N}_1$ contends using Slotted ALOHA [10] with contention probability $\hat{p}_e$ in the first slot of the CW on its channel, which is an uplink (UL) slot, and listens to the channel for an acknowledgment (ACK) packet from the BS in the second slot, which is a downlink (DL) slot. If a node successfully contends in the UL slot, then it is allotted the requested number of slots in the DTW by the BS, starting from the rightmost available slot of the current frame; the BS includes the list of allotted slots in the ACK packet that it broadcasts in the following DL slot. The node then switches to idle (sleep) state to save energy and wakes up to transmit in its allotted slots in the DTW. Now, since the

---

[8] Since M2M devices are low-cost and battery-operated devices, our protocol does not require them to have sensing capability.

[9] Note that although the value of R_s is not known in advance, after the first (respectively, second) phase of the estimation scheme, the BS can find the value of $|C_I|$ (respectively, $|C_{II}|$) (see Sections III-C1 and III-C2). So the information required to reserve R_s slots is available with the network.

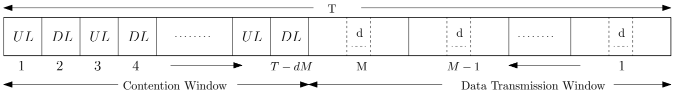

Figure 5. The figure shows the CDTW used in the analysis in Section V-A.

number of contending nodes has reduced by 1, the BS modifies $\hat{p}_e$ to $\min(1/((\hat{n}_e/M_{f_e})-1),1)$ and broadcasts this value in the *DL* slot. In case of an unsuccessful contention (collision or empty slot), the BS does not send any ACK. This process continues until the CW and DTW on that channel are separated by a single slot; then, the BS transmits a broadcast packet informing the remaining contending nodes to switch to idle state (to save energy) for the rest of the frame. However, if three successive UL slots are empty, then it is taken by the BS to be an indication that with a high probability all the active nodes on the channel have already successfully contended; in this case, the BS can allot the remaining free slots of the channel to active nodes of other channels. A similar procedure is followed by active nodes from $\mathscr{N}_2$ and $\mathscr{N}_3$ on their selected channels with parameter sets $(\hat{p}_p,\hat{n}_p)$ and $(\hat{p}_n,\hat{n}_n)$ respectively.

V. Performance Analysis

In this section, we obtain closed-form expressions for the expected number of successful contentions and the expected amount of energy consumed per frame under the Cognitive MAC protocol described in Section IV.

A. Expected Number of Successful Contentions

Here, we focus on only one channel and hence only nodes of a single type contend on it. Assume that n nodes of this type are active on the channel at the start of a given frame and let $\hat{n}$ be the estimated value of n. Let the length of the CDTW of the frame be T slots.

Let M be the number of successful contentions in the given frame. Recall that contentions occur only in UL slots. For tractability, we assume that upon every successful contention, the BS reserves a constant number, say d, of slots for the successful node from the last available slot in that frame as shown in Figure 5. If no successful contentions take place in the frame, then $M=0$ and if all contentions are successful, then $M=T/(2+d)$ since $2M+dM=T$. So $0 \le M \le T/(2+d)$. For each x, let $T_x = 0.5(T-xd)$. By definition:

$$E(M) = \sum_{m=0}^{T/(2+d)} mP(M=m) \tag{6}$$

Let S_x be the event that a successful contention occurs when x nodes contend. Let $P(S_x)=r_x$ and p_x be the contention probability used when x nodes contend. In the proposed protocol, $p_{n-j}=\min(1/(\hat{n}-j),1)$, $j=0,1,2\ldots$ (see Section IV-E). We now find the distribution of M. There are $M=m$ successful contentions if and only if for some integers $k_1,k_2,\ldots,k_m$, the first (k_1-1) contention attempts are unsuccessful with n contending nodes and the k_1^{th} attempt is successful, $(k_1+1)^{th}$ to $(k_2-1)^{th}$ attempts are unsuccessful with $n-1$ contending nodes and k_2^{th} attempt is successful, $\ldots$, and $(k_m+1)^{th}$ to T_m^{th} attempts are unsuccessful with $n-m$ contending nodes. So:

$$\begin{aligned}P(M=m) = &\sum_{k_1=1}^{T_m-m+1}\sum_{k_2=k_1+1}^{T_m-m+2}\cdots\sum_{k_{j+1}=k_j+1}^{T_m-m+j+1}\cdots\sum_{k_m=k_{m-1}+1}^{T_m} \\ &(1-r_n)^{k_1-1}r_n(1-r_{n-1})^{k_2-k_1-1}r_{n-1}\ldots \\ &(1-r_{n-j})^{k_{j+1}-k_j-1}r_{n-j}\ldots(1-r_{n-m})^{T_m-k_m}.\end{aligned} \tag{7}$$

Note that $r_{n-j}=(n-j)p_{n-j}(1-p_{n-j})^{(n-j-1)}$. $E(M)$ can be obtained from (6) and (7).

B. Expected Amount of Energy Consumed per Frame

Let $\gamma_I,\gamma_T,\gamma_R$ be the energy spent by a node per slot in the idle state, transmission state and reception state respectively. Let us classify the slots in the given frame into uplink slots, downlink slots and data transmission slots; let the total energy spent by all the active nodes in them be $\mathscr{E}_{UL},\mathscr{E}_{DL}$ and $\mathscr{E}_{DT}$ respectively. So the total expected amount of energy spent per frame is $E(\mathscr{E}_{UL})+E(\mathscr{E}_{DL})+E(\mathscr{E}_{DT})$. We compute $E(\mathscr{E}_{UL})$, $E(\mathscr{E}_{DL})$ and $E(\mathscr{E}_{DT})$ in Sections V-B1, V-B2 and V-B3 respectively.

1) Energy Spent in UL Slots: Note that there are a total of T_M uplink slots in the frame. In each of these slots, some of the active nodes are in transmission state and the rest are in idle state. So, $\mathscr{E}_{UL}=\sum_{i=1}^{T_M}(L_i\gamma_T+(n-L_i)\gamma_I)$, where L_i is the number of nodes that transmit in UL slot i, which depends on N_i (the number of contending nodes in UL slot i) and p_{N_i}. Taking expectations and conditioning on the values taken by M:

$$\begin{aligned}E(\mathscr{E}_{UL}) = &\sum_{m=0}^{T/(2+d)}\Bigg(\sum_{i=1}^{T_m}\Bigg(E(L_i/M=m)\gamma_T \\ &+\Big(n-E(L_i/M=m)\Big)\gamma_I\Bigg)\Bigg)P(M=m).\end{aligned} \tag{8}$$

So $E(L_i/M=m)$ can be calculated as,

$$\begin{aligned}E(L_i/M=m) = &\sum_{j=0}^{m}\Big(\sum_{l_i=0}^{n-j} l_i P(L_i=l_i/M=m,N_i=n-j)\Big) \\ &P(N_i=n-j/M=m)\end{aligned} \tag{9}$$

Distribution of N_i follows from the famous Gambler's ruin problem [29], i.e.,

$$\begin{aligned}P(N_i=n-j/M=m) = &P(N_{i-1}=n-j/M=m) \\ &\Big(1-P_i(S_{n-j}/M=m)\Big)+P(N_{i-1}=n-j+1/M=m) \\ &\Big(P_i(S_{n-j+1}/M=m)\Big)\end{aligned} \tag{10}$$

where $P_i(S_x/M=m)=$ Probability of success with x nodes given $M=m$ in slot i and $j=0,1,\ldots i-1$. Now,

$$P_i(S_{n-j}/M=m) = \frac{P_i(M=m/S_{n-j})P_i(S_{n-j})}{P(M=m)}, \tag{11}$$

$$P_i(M=m/S_{n-j}) = \sum_{k_1=1}^{T_m-m+1}\sum_{k_2=k_1+1}^{T_m-m+2}\cdots\sum_{k_{j+1}=k_j+1}^{T_m-m+j+1}\cdots\sum_{k_m=k_{m-1}+1}^{T_m}$$
$$(1-r_n)^{k_1-1}r_n(1-r_{n-1})^{k_2-k_1-1}r_{n-1}\ldots$$
$$(1-r_{n-j})^{k_{j+1}-k_j-1}\ldots 1\ldots(1-r_{n-m})^{T_m-k_m}. \tag{12}$$

Where $P_i(M=m/S_{n-j})=0$ for $m \le j$. Now distribution of L_i is given by, $P(L_i=l)=\binom{N_i}{l}p_{N_i}^l(1-p_{N_i})^{N_i-l}$. Let $L_{i,n-j}=$ The number of nodes that transmit in UL slot i given $(n-j)$ nodes, $j=0,1,2,\ldots,m$. So conditional distribution is given by,

$$P(L_i=l_i/M=m,N_i=n-j)=$$
$$\Big(1-P(L_{i,n-j}=l_i)\Big)^{T_m-m}\Big(P(L_{i,n-j}=l_i)\Big)^m \tag{13}$$

By using (7), (8), (9), (10), (11), (12), (13), we get $E(\mathscr{E}_{UL})$.

2) Energy Spent in DL Slots: In these slots, contending nodes are in reception state and the rest are in idle state. So, $\mathscr{E}_{DL}=\sum_{i=1}^{T_M}(N_i\gamma_R+(n-N_i)\gamma_I)$. Taking expectations and conditioning on the values taken by M, we get:

$$E(\mathscr{E}_{DL})=\sum_{m=0}^{T/(2+d)}\left(\sum_{i=1}^{T_m}\Big(E(N_i/M=m)\gamma_R\right.$$
$$\left.+\Big(n-E(N_i/M=m)\Big)\gamma_I\Big)\right)P(M=m). \tag{14}$$

By definition of conditional expectation, $E(N_i/M=m)=\sum_{j=0}^{m}(n-j)P(N_i=n-j/M=m)$.

$E(\mathscr{E}_{DL})$ can then be found using (7), (10) and (14).

3) Energy Spent in DT Slots: Since only one node is transmitting in these slots, all other nodes are in the idle state. So, $\mathscr{E}_{DT}=dM\Big(\gamma_T+(n-1)\gamma_I\Big)$ and

$$E(\mathscr{E}_{DT})=dE(M)\Big(\gamma_T+(n-1)\gamma_I\Big). \tag{15}$$

Using (6) and (15), we can calculate $E(\mathscr{E}_{DT})$.

C. Expectation of the Efficiency with the Proposed Estimation Scheme

Suppose there are n active nodes of a class in a given frame and let $\hat{n}$ be the estimated value of n obtained using the scheme described in Section III-C. Also, suppose m free channels are allocated to the class in the CDTW in the frame. Consider one of these channels and $n' \le n$ nodes select this channel. Recall that on this channel, each of the n' nodes contends using slotted ALOHA with contention probability $m/\hat{n}$.

From [28], we know that $\hat{n}=c\times 2^\rho$, where $c=1.2897$ and ρ is defined in Section III-B as the position of right most zero in BITMAP. From [28], we get the distribution of ρ as,

$$P(\rho=k)=P(\rho\ge k)-P(\rho\ge k+1), \tag{16}$$

where $P(\rho\ge k)=\sum_{j=0}^{2^k}(-1)^{\nu(j)}\Big(1-\frac{j}{2^k}\Big)^n$ and $\nu(j)$ indicates the number of one bits in the binary representation of j. We now find the expected value of efficiency, say η_A, which is defined to be the probability of successful contention in the first UL slot of the frame. Note that η_A is a measure of accuracy of the proposed estimation scheme. So, $\eta_A=\frac{mn'}{\hat{n}}(1-m/\hat{n})^{n'-1}$.

$$E(\eta_A)=mn'\times E\left(\frac{1}{c\times 2^\rho}\Big(1-\frac{m}{c\times 2^\rho}\Big)^{n'-1}\right),$$
$$=mn'\times\sum_r\left(\frac{1}{c\times 2^\rho}\Big(1-\frac{m}{c\times 2^\rho}\Big)^{n'-1}\right)P(\rho=r). \tag{17}$$

By using (16) and (17), we can find $E(\eta_A)$.

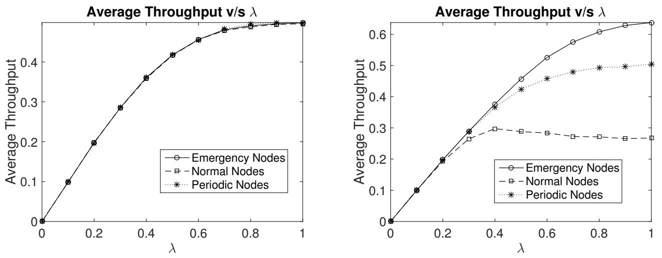

Figure 6. The following parameters are used in these plots: $M_T=30$, $N=50$, $k_e=k_p=k_n=1$. In the left plot we use $w_e=w_p=w_n=1$ whereas in the right plot we use $w_e=3$, $w_p=2$, $w_n=1$.

VI. Simulations

In this section, we evaluate the performance of the proposed Cognitive MAC protocol, in terms of average throughput and average delay, via simulations. Also, we compare the performance of the proposed protocol with a hypothetical "ideal protocol" to find out how accurate the proposed estimation scheme is. The ideal protocol is similar to the proposed protocol, with the difference being that it is assumed to know the *exact* number of active nodes at any time [10].

Let $M_T, M_f, \hat{n}_e, \hat{n}_p, \hat{n}_n, w_e, w_p, w_n$ and q_i be as defined in Section IV. At the beginning of each frame, data packets arrive at random at each node; the number of packets that arrive at a node belonging to the emergency (respectively, periodic, normal) data class is a Poisson random variable with mean λ_e (respectively, λ_p, λ_n). Also, each frame is divided into 50 slots and transmission of a data packet takes 1 slot. An emergency (respectively, normal) node which has successfully contended for access during the CW can reserve at most k_e (respectively, k_n) consecutive slots in the DTW. A periodic node can reserve one slot per frame for at most k_p consecutive frames. The limits k_e, k_n and k_p are imposed to ensure short-term fairness in the transmission opportunities that different nodes get. We consider a balanced load condition wherein there are an equal number, say N, of nodes of each type in the network and $\lambda_e=\lambda_p=\lambda_n=\lambda$ (say).

Figure 6 shows the average throughput per node versus λ for the three classes with different parameter values. It can be seen that initially the average throughput for any given class equals the arrival rate λ, but after a particular value of λ, the average throughput saturates, which is the unstable

---

[10] Note that the ideal protocol is not practically implementable and is considered only for comparison with the proposed protocol.

region of operation. The left plot in Figure 6 shows that when $w_e = w_p = w_n = 1$ and $k_e = k_p = k_n = 1$, the average throughput curves of all three classes roughly coincide; this is because they are treated alike by the protocol. In contrast, the right plot in Figure 6 shows that in the unstable region, emergency (respectively, periodic) nodes achieve a higher average throughput than periodic (respectively, normal) nodes when $w_e = 3$, $w_p = 2$, $w_n = 1$; this is because a higher weight results in more channels being allocated to a class. Figure 7 shows the average throughput under the proposed protocol and the ideal protocol versus λ for the emergency and normal classes. These plots show that *the performance of the proposed protocol is close to that of the ideal protocol for both classes*: in particular, in the unstable region of operation, on average, the proposed protocol achieves 87.5% (respectively, 68%) of the average throughput under the ideal protocol for the emergency (respectively, normal) class.

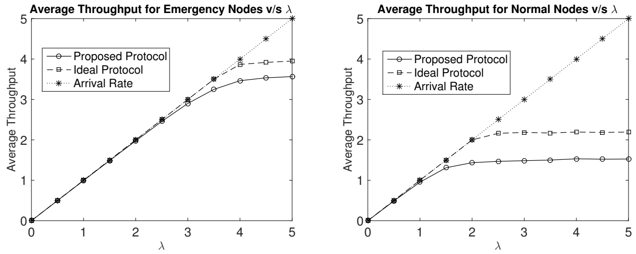

Figure 7. The following parameters are used in these plots: $M_T = 30$, $N = 50$, $w_e = 3$, $w_p = 2$, $w_n = 1$, $k_e = k_p = k_n = 5$.

Figure 8 shows the average packet delay versus λ for the emergency and normal classes with different parameter values. The left plot of Figure 8 shows that when $w_e = w_p = w_n = 1$ and $k_e = k_p = k_n = 5$, the average delay curves of the two classes roughly coincide; on the other hand, when the weights $w_e = 3$, $w_p = 2$, $w_n = 1$ are used (see the right plot of Figure 8), the average delay for emergency nodes is much lower than that of normal nodes, which is because more channels are allocated to emergency nodes.

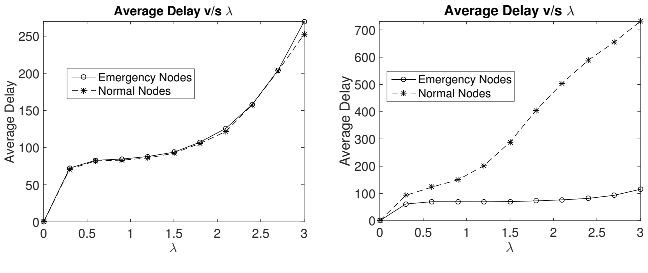

Figure 8. The following parameters are used in these plots: $M_T = 30$, $N = 50$, $k_e = k_p = k_n = 5$. In the left figure we use $w_e = w_p = w_n = 1$ whereas in the right figure we use $w_e = 3$, $w_p = 2$, $w_n = 1$.

VII. Conclusions and Future Work

We designed a Cognitive MAC protocol for a heterogeneous M2M network with three types of nodes. Our MAC protocol incorporates a fast cardinality estimation protocol to rapidly estimate the number of active nodes of each type; these estimates are used to find the optimal contention probabilities to be used in the MAC protocol. We mathematically analyzed the number of time slots required by our estimation protocol to execute as well as the performance of the Cognitive MAC protocol. Also, we evaluated the performance, in terms of average throughput and average delay, of our MAC protocol using simulations. In this paper we have considered a heterogeneous network with three types of nodes; a direction for future research is to extend our results to the case where there are an arbitrary number of node types.

References

[1] G. Wu, S. Talwar, K. Johnsson, N. Himayat, and K. D. Johnson, "M2M: From Mobile to Embedded Internet," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 36–43, 2011.

[2] Y. Zhang, R. Yu, M. Nekovee, Y. Liu, S. Xie, and S. Gjessing, "Cognitive Machine-to-Machine Communications: Visions and Potentials for the Smart Grid," *IEEE Network*, vol. 26, no. 3, pp. 6–13, 2012.

[3] A. Rajandekar and B. Sikdar, "A Survey of MAC Layer Issues and Protocols for Machine-to-Machine Communications," *IEEE Internet of Things Journal*, vol. 2, no. 2, pp. 175–186, 2015.

[4] 3GPP, "Service Requirements for Machine-Type Communications," 2014.

[5] A. Aijaz and A.-H. Aghvami, "PRMA-based Cognitive Machine-to-Machine Communications in Smart Grid Networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 8, pp. 3608–3623, 2015.

[6] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks*, vol. 50, no. 13, pp. 2127–2159, 2006.

[7] A. De Domenico, E. C. Strinati, and M.-G. Di Benedetto, "A Survey on MAC Strategies for Cognitive Radio Networks," *IEEE Communications Surveys and Tutorials*, vol. 14, no. 1, pp. 21–44, 2012.

[8] J. So and N. H. Vaidya, "Multi-Channel MAC for Ad Hoc Networks: Handling Multi-Channel Hidden Terminals using a Single Transceiver," in *Proc. of ACM MobiHoc*, pp. 222–233, 2004.

[9] C. Cordeiro and K. Challapali, "C-MAC: A Cognitive MAC Protocol for Multi-Channel Wireless Networks," in *Proc. of DySPAN*, 2007.

[10] D. P. Bertsekas and R. G. Gallager, *Data Networks*. PHI, 2nd ed., 1992.

[11] H. Wu, C. Zhu, R. J. La, X. Liu, and Y. Zhang, "FASA: Accelerated S-ALOHA Using Access History for Event-Driven M2M Communications," *IEEE/ACM Transactions on Networking*, vol. 21, no. 6, pp. 1904–1917, 2013.

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

[13] C.-Y. Hsu, C.-H. Yen, and C.-T. Chou, "An Adaptive Multichannel Protocol for Large-Scale Machine-to-Machine (M2M) Networks," in *Proc. of IWCMC*, pp. 1223–1228, 2013.

[14] M. Kodialam and T. Nandagopal, "Fast and Reliable Estimation Schemes in RFID Systems," in *Proc. of ACM MobiCom*, 2006.

[15] M. Kodialam, T. Nandagopal, and W. C. Lau, "Anonymous Tracking using RFID Tags," in *Proc. of IEEE INFOCOM*, pp. 1217–1225, 2007.

[16] T. Li, S. Wu, S. Chen, and M. Yang, "Energy Efficient Algorithms for the RFID Estimation Problem," in *Proc. of IEEE INFOCOM*, 2010.

[17] Y. Zheng and M. Li, "PET: Probabilistic Estimating Tree for Large-Scale RFID Estimation," *IEEE Transactions on Mobile Computing*, vol. 11, no. 11, pp. 1763–1774, 2012.

[18] H. Han, B. Sheng, C. C. Tan, Q. Li, W. Mao, and S. Lu, "Counting RFID Tags Efficiently and Anonymously," in *Proc. of IEEE INFOCOM*, 2010.

[19] Z. Bar-Yossef, T. Jayram, R. Kumar, D. Sivakumar, and L. Trevisan, "Counting distinct elements in a data stream," in *International Workshop on Randomization and Approximation Techniques in Computer Science*, pp. 1–10, Springer, 2002.

[20] M. Shahzad and A. X. Liu, "Every Bit Counts: Fast and Scalable RFID Estimation," in *Proc. of ACM MobiCom*, pp. 365–376, 2012.

[21] M. T. Islam, M. T. Abd-elhamid, and S. Akl, "A Survey of Access Management Techniques in Machine Type Communications," *IEEE Communications Magazine*, vol. 52, no. 4, pp. 74–81, 2014.

[22] F. Azquez-Gallego, J. Alonso-Zárate, I. Balboteo, and L. Alonso, "DPCF-M: A Medium Access Control Protocol for Dense M2M Area Networks with Dynamic Gateways," in *Proc. of IEEE SPAWC*, 2013.

[23] Y. Liu, C. Yuen, J. Chen, and X. Cao, "A Scalable Hybrid MAC Protocol for Massive M2M Networks," in *Proc. of IEEE WCNC*, 2013.

[24] 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 of Things Journal*, vol. 1, no. 1, pp. 99–111, 2014.

[25] G. Wang, X. Zhong, S. Mei, and J. Wang, "An Adaptive Medium Access Control Mechanism for Cellular based Machine-to-Machine (M2M) Communication," in *Proc. of IEEE ICWITS*, pp. 1–4, 2010.

[26] A. Lo, Y. W. Law, M. Jacobsson, and M. Kucharzak, "Enhanced LTE-Advanced Random-Access Mechanism for Massive Machine-to-Machine Communications," in *Proc. of WWRF Meeting*, pp. 1–5, 2011.

[27] C. Qian, H. Ngan, Y. Liu, and L. M. Ni, "Cardinality Estimation for Large-Scale RFID Systems," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 9, pp. 1441–1454, 2011.

[28] P. Flajolet and G. N. Martin, "Probabilistic Counting Algorithms for Data Base Applications," *Journal of Computer and System Sciences*, vol. 31, no. 2, pp. 182–209, 1985.

[29] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 1. Wiley, January 1968.

APPENDIX

Proof for Theorem 3.1

Proof: The expression for the expected number of slots required in the second phase is, $E(K) = \sum_{i=0}^{t-1} P(C_i)$, where C_i is the event that collision occurs in both the slots of i^{th} block. So $C_i = A_i \cup B_i \cup D_i$.
A_i : Event that at least two Type 1 nodes transmit in block i ;
B_i : Event that at least two Type 2 and Type 3 nodes transmit in block i ;
D_i : Event that exactly one Type 1 node and one Type 2, one Type 3 nodes transmit in block i.

So, upper bounds for $P(A_i)$, $P(B_i)$ and $P(D_i)$ are computed to be,

$$^{11}P(A_i) \leq \binom{n_1}{2} p_i^2 = \frac{n_1^2}{2} p_i^2; \tag{18}$$

$$^{12}P(B_i) \leq \binom{n_2}{2} p_i^2 \binom{n_3}{2} p_i^2 = \frac{n_2^2 n_3^2}{4} p_i^4; \tag{19}$$

$$^{13}P(D_i) \leq (n_1 p_i)(n_2 p_i)(n_3 p_i) = n_1 n_2 n_3 p_i^3; \tag{20}$$

where $p_i = (1/2)^{i+1}$. As we know, $l_{n_m} = \lceil (\log_2 n_m) \rceil$. Using union bound, $P(C_i) \leq P(A_i) + P(B_i) + P(D_i) \leq (n_1^2/2)p_i^2 + (n_2^2 n_3^2/4)p_i^4 + n_1 n_2 n_3 p_i^3$. This implies,

$$E(K) = \sum_{i=0}^{t-1} P(C_i) = \sum_{i=0}^{l_{n_m}-2} P(C_i) + \sum_{i=l_{n_m}-1}^{t-1} P(C_i). \tag{21}$$

Let us consider the first summation of the Equation (21),

$$\sum_{i=0}^{l_{n_m}-2} P(C_i) \leq \sum_{i=0}^{l_{n_m}-2} 1 = l_{n_m} - 1. \tag{22}$$

Now consider the second summation of the Equation (21), where t can be written as $t = l_{n_m} + r$,

$$\begin{aligned}\sum_{i=l_{n_m}-1}^{l_{n_m}+r-1} P(C_i) &\leq \sum_{i=l_{n_m}-1}^{l_{n_m}+r-1} \left(\frac{n_1^2}{2} p_i^2 + \frac{n_2^2 n_3^2}{4} p_i^4 + n_1 n_2 n_3 p_i^3 \right) \\ &= \frac{n_1^2}{2} \sum_{i=l_{n_m}-1}^{l_{n_m}+r-1} \left(\frac{1}{4}\right)^{i+1} + \frac{n_2^2 n_3^2}{4} \sum_{i=l_{n_m}-1}^{l_{n_m}+r-1} \left(\frac{1}{16}\right)^{i+1} \\ &\quad + n_1 n_2 n_3 \sum_{i=l_{n_m}-1}^{l_{n_m}+r-1} \left(\frac{1}{8}\right)^{i+1}\end{aligned} \tag{23}$$

Let us consider individual upper bounds for each quantity in Equation (23). We get,

$$\begin{aligned}\frac{n_1^2}{2} \sum_{i=l_{n_m}-1}^{l_{n_m}+r-1} \left(\frac{1}{4}\right)^{i+1} &\leq \frac{n_1^2}{2} \left[\frac{1}{n_m^2} + \frac{1}{4n_m^2} + \cdots + \frac{1}{4^{r-1} n_m^2} \right]^{14} \\ &\leq \frac{2}{3} \frac{n_1^2}{n_m^2}.\end{aligned} \tag{24}$$

$$\begin{aligned}\frac{n_2^2 n_3^2}{4} \sum_{i=l_{n_m}-1}^{l_{n_m}+r-1} \left(\frac{1}{16}\right)^{i+1} &\leq \frac{n_2^2 n_3^2}{4} \left[\frac{1}{n_m^4} + \frac{1}{16 n_m^4} + \cdots + \frac{1}{16^{r-1} n_m^4} \right]^{15} \\ &\leq \frac{4}{15} \frac{n_2^2 n_3^2}{n_m^4}.\end{aligned}$$

$$\begin{aligned}n_1 n_2 n_3 \sum_{i=l_{n_m}-1}^{l_{n_m}+r-1} \left(\frac{1}{8}\right)^{i+1} &\leq n_1 n_2 n_3 \left[\frac{1}{n_m^3} + \frac{1}{8 n_m^3} + \cdots + \frac{1}{8^{r-1} n_m^3} \right]^{16} \\ &\leq \frac{8}{7} \frac{n_1 n_2 n_3}{n_m^3}.\end{aligned}$$

Therefore,

$$E(K) \leq l_{n_m} - 1 + \frac{2n_1^2}{3n_m^2} \left[1 + \frac{2 n_2^2 n_3^2}{5 n_1^2 n_m^2} + \frac{12 n_2 n_3}{7 n_1 n_m} \right]. \tag{25}$$

■

Proof for Theorem 3.2

Proof: The expression for the expected number of slots required in the third phase is, $E(R) \leq \sum_{i=0}^{t-1} P(A_i)$, where A_i is the event that at least two Type 1 nodes transmit in both the slots of block i in the first phase. As we know, $l_{n_m} = \lceil (\log_2 n_m) \rceil$. From the inequalities shown in (18) and (24), we get $E(R) \leq l_{n_m} - 1 + \frac{2n_1^2}{3n_m^2}$. ■

---

[11] Let $E_j = j^{th}$ node transmission event, for $j = 1, 2, \ldots n$ and for some n. So $P(A_i) = P(\cup_{1 \leq j_1 < j_2 \leq n_1} (E_{j_1} \cap E_{j_2})) \leq \binom{n_1}{2} P(E_{j_1} \cap E_{j_2})$.

[12] $P(B_i) = P(\cup_{1 \leq j_1 < j_2 \leq n_2} (E_{j_1} \cap E_{j_2})) \quad P(\cup_{1 \leq k_1 < k_2 \leq n_3} (E_{k_1} \cap E_{k_2})) \leq \binom{n_2}{2} P(E_{j_1} \cap E_{j_2}) \ \binom{n_3}{2} P(E_{k_1} \cap E_{k_2})$.

[13] $P(D_i) \leq n_1 P(E_{j_1}) n_2 P(E_{j_2}) n_3 P(E_{j_3})$, for $j_1 = 1, 2, \ldots n_1; j_2 = 1, 2, \ldots n_2; j_3 = 1, 2, \ldots n_3$.

[14] $\left(\frac{1}{4}\right)^{l_{n_m}+k} \leq \left(\frac{1}{4}\right)^{\log_2 n_m} \left(\frac{1}{4}\right)^k = \left(\frac{1}{4^k n_m^2}\right)$ for $k = 0, 1, 2, \ldots, r-1$

[15] $\left(\frac{1}{16}\right)^{l_{n_m}+k} \leq \left(\frac{1}{16}\right)^{\log_2 n_m} \left(\frac{1}{16}\right)^k = \left(\frac{1}{16^k n_m^4}\right)$ for $k = 0, 1, 2, \ldots, r-1$

[16] $\left(\frac{1}{8}\right)^{l_{n_m}+k} \leq \left(\frac{1}{8}\right)^{\log_2 n_m} \left(\frac{1}{8}\right)^k = \left(\frac{1}{8^k n_m^3}\right)$ for $k = 0, 1, 2, \ldots, r-1$