Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Reliable and Efficient Access for Alarm-initiated and Regular M2M Traffic in IEEE 802.11ah Systems

Germán Corrales Madueño, Čedomir Stefanović, Petar Popovski

Abstract—**IEEE 802.11ah is a novel WiFi-based protocol, aiming to provide an access solution for the machine-to-machine (M2M) communications. In this paper, we propose an adaptive access mechanism that can be seamlessly incorporated into IEEE 802.11ah protocol operation and that supports all potential M2M reporting regimes, which are periodic, on-demand and alarm reporting. The proposed access method is based a periodically re-occurring pool of time slots, whose size is proactively determined on the basis of the reporting activity in the cell. We show that it is possible to both efficiently and reliably resolve all reporting stations in the cell, within the limits of the allowed deadlines. As a side result, we also provide a rationale for modeling the inter-arrival time in alarm events by using the Beta distribution, a model that is considered in the 3GPP standardization.**

I. INTRODUCTION

IEEE 802.11ah is a novel wireless local area network protocol, offering low-power, long-range and scalable operation. The operating frequency of IEEE 802.11ah is below 1 GHz, allowing a single access point (AP) to provide service to a area of up to 1 km. Additionally, IEEE 802.11ah is designed to support very large number of stations (up to 8100) in a single cell, in comparison to the other standards from the IEEE 802.11 family. As such, IEEE 802.11ah aims to establish itself as an access solution with a low infrastructure cost for capillary M2M access.

At the time of writing, the IEEE 802.11ah task group is still working on the completion of the protocol definition, where the second draft was initially expected by mid-2014 [1]. Nevertheless, the foreseen EEE 802.11ah access mechanism has already been topic of several recent studies. The number of devices supported for different uplink-downlink traffic ratios with different data rates was investigated in [2]. The performance of the grouping strategy, labeled as restricted access window (RAW), was studied in [3]. It was shown that limiting the number of contending stations through use of RAW could produce significant throughput improvements. In

[4], the authors propose to estimate the number of contending devices in order to determine the adequate length of the RAW.

The main figure of interest in the above studies is the throughput, where the analysis is based on a full-buffer approach, i.e., it is assumed that the stations always have data to transmit. However, this assumption does not correspond to the nature of typical reporting from M2M devices (terminals), where the terminals only sporadically report data, triggered by the reporting period or certain external events.

In contrast to previous works, in this paper we consider the reporting regimes that are characteristic for M2M, which differ significantly from the approach that assumes full buffer. The first regime is the *asynchronous periodic* reporting, characterized with reporting intervals of several minutes [5]. The second regime is *synchronous alarm* reporting, where potentially thousands of stations are triggered almost simultaneously. The showcase example of an M2M application in which terminals report in both regimes is smart metering. Under normal conditions, smart meters asynchronously and periodically report the energy consumption to a remote server. However, in special occasions, such as in a power outage, thousands of devices almost simultaneously try to report the failure before the battery dies [6].

The topic of this paper is the design of an access method tailored for IEEE 802.11ah protocol, whose main aim is the provision a reliable and efficient service, both for asynchronous and synchronous M2M reporting. The proposed solution consists of the periodically reoccurring pool of resources, i.e., time slots, which is proactively dimensioned based on the outcomes of the contention among the reporting terminals and operates such that the dominant reporting regime, i.e., asynchronous vs. synchronous, is matched. Moreover, in contrast to previous works, we consider both the efficiency and reliability of the access mechanism, as the latter performance parameter plays an important role in M2M communications, particularly in alarm reporting scenarios. We show that it is possible to detect all alarms reports while efficiently using the system resources, i.e., wireless link time. The proposed scheme draws inspiration from the work presented in [7], where the main topic is a static solution for provision of a reliable communication to a set of M2M terminals. However, the scenario is limited only to the case in which the set of M2M terminals report asynchronously. Differently from [7], here we thoroughly modify and extend the solution from [7] towards a dynamic version that is adapted to the reporting demands in a cell. Specifically, the proposal in this paper provides an efficient reliable support of
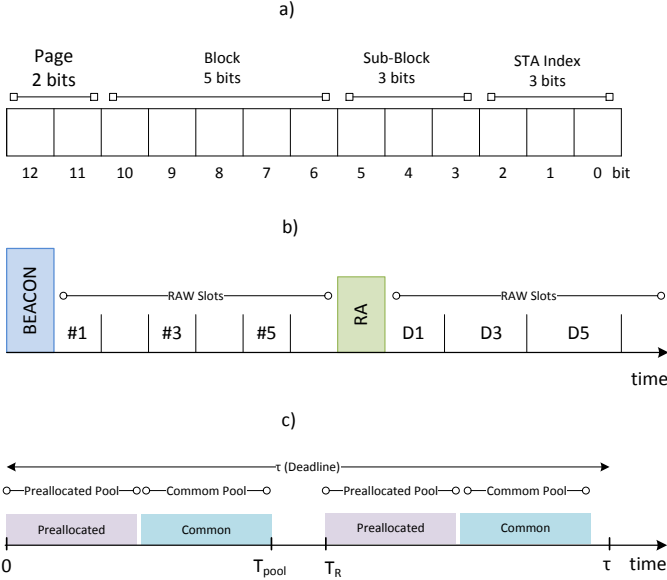
Fig. 1. a) AID format. b) Beacon frame announcing a RAW with 6 slots followed by a Resource Allocation (RA) frame that allocates data slots for stations 1, 3 and 5. c) Illustration of T_R, T_{pool} and the deadline τ.

synchronously reported alarms, which is a very challenging traffic class.

Another important contribution of our work is related to the Beta distribution, used to model alarms that are reported synchronously. The use of the Beta distribution has been considered in standardization [8], however, without rationale on the choice of the distribution and its parameters. This paper closes this knowledge gap and provides a systematic treatment of the Beta distribution in relation to the M2M traffic.

The rest of the paper is organized as follows. Section II provides an overview of the 802.11ah access mechanism. Section III presents the system model. Section IV is devoted the modeling of alarm reporting. Section V introduces the proposed access method. Section VI is the central part of the paper, presenting analysis of the proposed scheme. The evaluation of the scheme's performance is done in Section VII. Finally, Section VIII concludes the paper.

II. BASICS OF IEEE 802.11AH ACCESS

The IEEE 802.11ah access mechanism operates as follows. Every station (i.e., terminal) connected to the AP receives an unique identifier, denoted as an association identifier (AID). The AID is a 13-bit word that follows a four-level hierarchical structure as depicted in Fig. 1a). The first two bits are used to organize the stations in four pages. The next 5 bits split a page into 32 blocks. The following three bits are used to divide a block into 8 sub-blocks. Finally, the last 3 bits are used to determine the station's index within the sub-block, i.e., there could be a maximum of 8 stations per sub-block. Obviously, this hierarchical structure allows for a straightforward grouping of stations.

The AID can be used to determine which stations are allowed to access the medium. Specifically, IEEE 802.11ah introduces the concept of restricted access window (RAW),

during which only certain stations are allowed to contend based on their AIDs. This limitation the number of contending stations in RAW enhances the access efficiency, seen in the number of stations whose transmissions do not experience collisions. RAW structure is depicted in Fig. 1b), where it can be seen that RAW is subdivided in slots. The information concerning the number of slots in the RAW, slot duration and assignment is indicated in the beacon frame that is periodically sent by the AP to the stations. Every slot can be assigned to a single or multiple stations, according to the number of stations and number of the available slots [9]. The active stations contend only in the assigned slots. Henceforth, we assume that the contention is based on slotted ALOHA [10], where the contention outcomes can be an idle, singleton or collision slot, and that the cross-boundary slot option [1] is not used, implying that a station is allowed to contend only within the boundaries of the allocated slot. The amount of information that can be sent within a slot depends on the slot duration and the modulation used; the available modulation schemes range from BPSK to 256-QAM and the corresponding data rates from 150 kbps to 4000 kbps [1].

Two main RAW operation modes are generic RAW and triggered RAW. In case of the generic RAW mode, the stations contend with their data packets. In the triggered RAW mode, the access is reservation based and there are actually two RAWs, the initial RAW corresponding to the reservation phase and the second RAW corresponding to the data phase. In the initial RAW, the stations contend with short poll frames[1], trying to reserve (i.e., obtain from the AP) a dedicated RAW slot for the pending downlink/uplink data transmissions. Based on the received polls, the AP allocates a second RAW consisting of dedicated slots, and notifies the corresponding stations via resource allocation frame (RA) about the RAW slot assignment, see Fig. 1b). The advantage of triggered over the generic RAW mode is that idle or collided slots are less costly due to their shorter length, as they just have to fit a poll frame instead of a data packet.

For a better understanding, we provide an example of the triggered RAW mode in Fig. 1b). The AP provides an initial RAW composed of 6 RAW slots intended for 6 stations, i.e., a slot per station. It is assumed that only stations #1, #3 and #5 have data to transmit and thus send their polls, whereas #2, #4 and #6 are idle and do not transmit. Immediately after the initial RAW, the AP allocates another RAW containing a slot dedicated to each station, denoted as D1, D3 and D5 in Fig. 1b), and informs the stations about the allocation via RA frame.

We conclude this section by noting that proposed access method is tailored for the triggered RAW mode of operation. In particular, the proposed method is intended as the efficient and reliable solution for the reservation phase of the triggered RAW mode.

III. SYSTEM MODEL

We consider a single IEEE 802.11ah cell. We assume that the cell is circular, with radius r, and that the AP is residing

---
[1]At the time of writing, the structure of the poll frame has not been fully standardized.

in its center. There are N stations in the cell, whose distances from the AP are uniformly randomly distributed. We consider the scenario where the locations of the stations are fixed. This is a widely used assumption for M2M deployment based on scenarios with smart meters and sensors; this is also why one of the main changes to the cellular standards [5], in order to make them suitable for M2M communications, is to remove the mobility-related features in the protocol.

All stations perform periodic, on-demand and alarm reporting. *Periodic reporting* refers to the automatic reporting performed by stations, which is modeled by a Poisson arrival process on a station basis with rate $\lambda_{\mathrm{p}} = 1/T_{\mathrm{RI}}$, where T_{RI} denotes the duration of the reporting interval (RI) [7]. In the smart-metering context, periodic reporting is the automatic consumption reporting performed by the meters. *On-demand reporting* is the reporting that is demanded by the application server and can be also modeled as a Poisson arrival process with rate λ_{d}; it could be expected that $\lambda_{\mathrm{p}} > \lambda_{\mathrm{d}}$ [11]. An example for on-demand reporting is a delivery of consumption reports triggered by the customer. Both the periodic and on-demand reporting are asynchronous; henceforth refer to their combination as *regular* reporting, as it corresponds to a regular, i.e., standard reporting in the context of M2M applications, when there are no alarm events. Finally, *alarm reporting* corresponds to traffic generated by an event in which affected devices are activated almost simultaneously. Modeling of alarm reporting is significantly more involved than in the previous two cases; we devote the following section to the derivation of the appropriate model.

A parameter of paramount importance is the maximum allowed delay, i.e., the deadline from the report generation until its delivery. A report whose delay exceeds the maximum allowed value is considered as outdated and dropped by the station. Typically, it is assumed that regular reporting is more delay tolerant, i.e., a report in the current RI could be delivered in the beginning of the following RI [5]. Therefore, the maximum allowed delay of the periodic reporting is:

$$\tau_{\mathrm{p}} = T_{\mathrm{RI}}. \tag{1}$$

On the other hand, on-demand and, particularly, alarm reporting have stricter deadlines [11]:

$$\tau_{\mathrm{a}} \ll \tau_{\mathrm{d}} \leq \tau_{\mathrm{p}}, \tag{2}$$

where τ_{a} and τ_{d} denote alarm and on-demand reporting deadlines, respectively.

IV. REPORTING OF ALARMS

The typical assumption is that in case of an alarm event a large number of stations will be affected and their reporting will become correlated in time. In [8], 3GPP proposed a model for highly correlated traffic arrivals, where the interval-arrival time follows a Beta distribution:

$$p(t) = \frac{t^{\alpha-1}(T-t)^{\beta-1}}{T^{\alpha+\beta-1}\mathrm{B}(\alpha,\beta)} \text{ for } 0 \leq t \leq T, \tag{3}$$

where $\alpha > 0$ and $\beta > 0$ are shape parameters, $\mathrm{B}(\alpha,\beta)$ is the Beta function [12] and T is the activation period, i.e., the
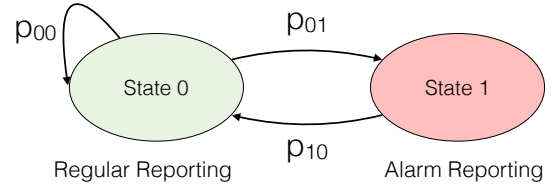

Fig. 2. State diagram of a station, with two states that correspond to regular reporting state (state 0), comprising periodic and on-demand reporting, and alarm reporting state (state 1).

time period from the activation of the first station until the last station is activated. The number of stations triggered during an interval T_I is given by:

$$N_A = N \int_0^{T_I} p(t)dt, \tag{4}$$

where N is the total number of stations.

The values of the parameters α, β and T suggested in [8] are $\alpha = 3$, $\beta = 4$ and $T = 10\,\mathrm{s}$. However, reference [8] neither elaborates on the reasons to use Beta distribution, nor provides a rationale for the choice of the values of the parameters. Our present work closes this knowledge gap as follows: modeling the reporting behavior of the stations, providing a justification for the use of Beta distribution for the characterization of the alarm reporting regime, providing insights into the choice of the parameters α, β and T, as well as their relation to the values suggested by 3GPP.

A. The Proposed Model of Alarm Reporting

We assume that every station can be in two reporting states, see Fig. 2. The first, regular reporting state, denoted as state 0, comprises periodic and on-demand reporting, and the second state, denoted as state 1, corresponds to alarm reporting. We assume that the state transitions occur in discrete time, where time step is denoted as Δt.[2] The arrival rates for each state during Δt are:

$$\lambda_0 = (\lambda_{\mathrm{p}} + \lambda_{\mathrm{d}})\Delta t, \tag{5}$$

where λ_{p} and λ_{d} are arrival rates of the periodic and on-demand reporting, which are adjusted to the time step Δt in (5). For analytical tractability, we also assume Poisson arrivals in state 1 [13], with the arrival rate during Δt:

$$\lambda_1 = 1. \tag{6}$$

i.e., it is assumed that a single report is generated on average.

In order to model the correlated behavior of the stations, we assume a coupled Markov modulated Poisson process, as proposed in [13]. The main idea is that a background process Θ determines the transition probability matrices $\mathbf{P}_n(t)$ for all stations n, $1 \leq n \leq N$:

$$\begin{aligned} \mathbf{P}_n(t) &= \begin{pmatrix} p_{n,00}(t) & p_{n,01}(t) \\ p_{n,10}(t) & p_{n,11}(t) \end{pmatrix} \\ &= (1 - \theta_n(t)) \cdot \mathbf{P}_0 + \theta_n(t) \cdot \mathbf{P}_1, \end{aligned} \tag{7}$$

[2]To appropriately model the alarm reporting with respect to the proposed access method, it is sufficient that $\Delta t \leq T_{\mathrm{R}}$, where T_{R} is the pool reoccurring period, see Section V.
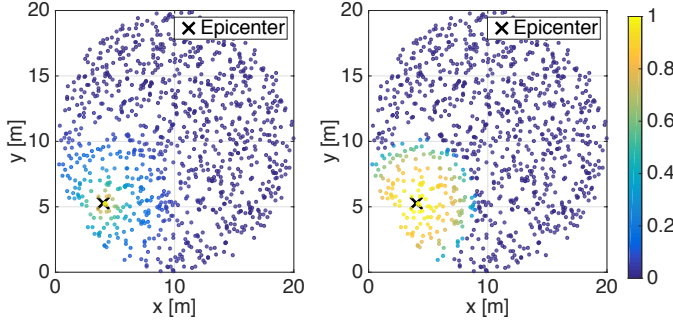
Fig. 3. Values of spatial correlation factor in a cell with $N = 1000$ stations and radius of $r = 10\,\mathrm{m}$. a) Exponentially-decaying spatial-correlation factor $\Psi^{\mathrm{Exp}}(d)$, $a = 1.2$. b) Square-root spatial-correlation factor $\Psi^{\mathrm{Sq}}(d)$, $d_{\max} = 4\,\mathrm{m}$.

where the samples of the background process $\theta_n(t)$ are in the range $[0, 1]$, and where $\mathbf{P}_0$ and $\mathbf{P}_1$ are the transition probabilities matrices in the regular state (state 0) and alarm reporting state (state 1), respectively:

$$\mathbf{P}_0 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \qquad \mathbf{P}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

In other words, in the regular reporting state a station never reports alarm, while in the alarm reporting state, a station reports a single alarm and goes back to the regular reporting regime.

Using the fact that the sum of independent Poisson random variables is a Poisson random variable [14], the aggregated arrival rate λ of station n during the interval of duration $T_I = M\Delta t$ is:

$$\lambda_n(T_I) = \sum_{m=0}^{M-1} \lambda_0 \cdot \pi_{n,0}(m\Delta t) + \lambda_1 \cdot \pi_{n,1}(m\Delta t), \quad (8)$$

where $\pi_n(t) = (\pi_{n,0}(t) \;\; \pi_{n,1}(t))$ can be obtained from the balance equation $\pi_n(t) = \pi_n(t) \cdot \mathbf{P}_n(t)$.

Obviously, the samples of background process $\theta_n(t)$ determine the features of the larm reporting on the cell level, see (7). In [13], $\theta_n(t)$ is modeled using Beta distribution with the parameters given in [8]. In contrast to this approach, we characterize $\theta_n(t)$ using a physical model of alarm propagation. Specifically, we assume that $\theta_n(t)$ can be expressed as:

$$\theta_n(t) = \Psi(d_n) \cdot \delta\left(t_{\mathrm{a}} - \frac{d_n}{v}\right), \; n = 1, \ldots, N, \quad (9)$$

where $\Psi(d_n)$ is the spatial correlation factor, $\delta\left(t - \frac{d_n}{v}\right)$ is a delta pulse that describes the propagation of the event that triggered alarm reporting, t_{a} is the moment of the occurrence of the alarm event, d_n is the stations distance from the epicenter of the alarm event and v is the propagation speed of the alarm event.

The spatial correlation factor $\Psi(d)$ expresses probability that a station is affected by an alarm event that took place at distance d_n. We consider three different models for $\Psi(d)$. The first model corresponds to the situation where all stations in the cell are affected by the event:
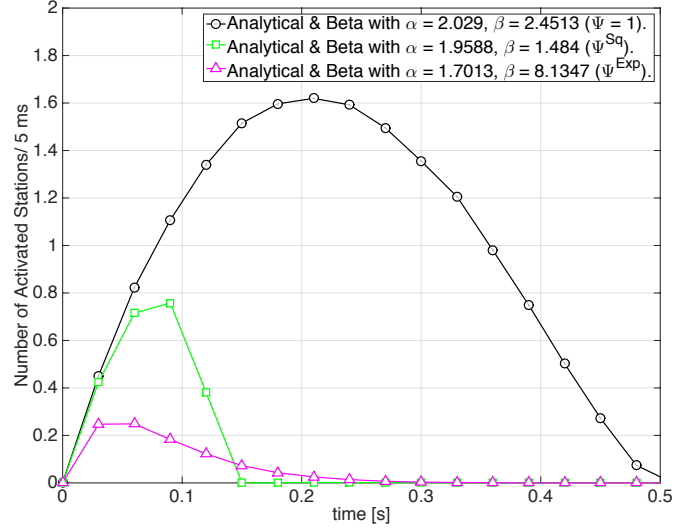
$$\Psi(d) = 1. \quad (10)$$

Fig. 4. Evolution of the number of alarm reports over time for three spatial correlation factors: $\Psi(d) = 1$, $\Psi^{\mathrm{Sq}}(d)$ with $d_{\max} = 500\,\mathrm{m}$, and Ψ^{Exp} with $a = 0.005$. Scenario corresponds to a cell with $r = 1000\,\mathrm{m}$, where the event propagates at $v = 4\,\mathrm{km/s}$.

In the second model, we consider a correlation exponentially decaying with the distance, given as:

$$\Psi^{\mathrm{Exp}}(d) = \begin{cases} 0, & d < 0, \\ e^{-a \cdot d}, & d \geq 0, \end{cases} \quad (11)$$

where d is the distance of the station to the epicenter of the alarm event and a is the decay constant[3]. Finally, as the third model we consider a square root function:

$$\Psi^{\mathrm{Sq}}(d) = \begin{cases} \sqrt{d_{\max}^2 - d^2}, & 0 \leq d \leq d_{\max}, \\ 0, & \text{else}, \end{cases} \quad (12)$$

where $d_{\max}$ is the maximum distance with respect to the epicenter that the event can reach.

The examples of the exponentially-decaying and the square-root spatial correlation factor are shown in Fig. 3. For the sake of comparison, the decay constant a is chosen such that $\Psi^{\mathrm{Sq}}(d) < 0.01$ for $d \geq d_{\max}$, where $d_{\max}$ is the parameter that determines the maximum spreading of the alarm for $\Psi^{\mathrm{Exp}}(d)$. Obviously, the values of spatial correlation are higher when the spatial correlation of alarm reporting is modeled by (12).

B. Illustrative Examples

We provide a brief illustration of the alarm reporting patterns in a cell for the different correlation factors. The example scenario corresponds to a power outage in the grid due to an earthquake, where the primary wave propagates at an average speed $v = 4$ km/s [8], the cell radius is set to $r = 1000\,\mathrm{m}$ and the number of stations is $N = 1000$.

Fig. 4 shows how the number of alarm reports (i.e., the number of the alarm reporting stations) evolves over intervals of $5\,\mathrm{ms}$, for the three correlation factors: $\Psi(d) = 1$, $\Psi^{\mathrm{Sq}}(d)$ with $d_{\max} = 500\,\mathrm{m}$ and $\Psi^{\mathrm{Exp}}(d)$ with $a = 0.005$. Obviously, for $\Psi(d) = 1$ all stations are triggered and the activation

---

[3]We assume that a is given in m^{-1}.

period is the longest. The total number of triggered stations for $\Psi^{\mathrm{Sq}}(d)$ is larger than for $\Psi^{\exp}(d)$; however, the activation period is longer for $\Psi^{\mathrm{Exp}}(d)$. The depicted results were also validated in a simulation setup that implements alarm reporting in a cell with the corresponding number of stations and using the models of the background process (9).

Interestingly, it can be shown that all activation patterns showed in Fig. 4 can be described by a Beta distribution with the appropriate choice of the shape parameters and activation periods, see (3). The corresponding values are indicated in the figure's legend. Obviously, these values significantly differ from the ones proposed by 3GPP [8]. Specifically, the activation period T is one order magnitude shorter than the value suggested in [8]. Also, it could be shown that the number of activated devices over time for the model proposed in [8] is negligible in comparison to the values depicted in Fig. 4. We note that the similar conclusions can be drawn for other values of spatial correlation factor.

The observations above justify the use of Beta distribution to model synchronous reporting by a massive number of stations in a cell. However, the parameters of the distribution have to be carefully tuned to the properties of the phenomenon that describes the spreading of the alarm event, and there is no one-size-fits-all solution.

In Section VII we demonstrate the performance of the proposed allocation mechanism in the case of an alarm whose spatial correlation factor is modeled by $\Psi^{\mathrm{Sq}}(d)$ and note that similar studies based on different spatial correlation factors could be done in an analogous manner.

V. THE PROPOSED ACCESS METHOD

The main feature of the proposed access method is that it consists of a periodically reoccurring pool of reservation slots (RSs). We assume that RS is a RAW slot dimensioned to fit a poll frame that consists of a station's AID and a single bit that indicates that the station has a pending uplink report, which is required by the IEEE 802.11ah protocol. The pool is reoccurring with a fixed period T_{R}. The pool duration T_{pool} is adaptive and determined by the reporting demands of the stations, such that all stations that have pending report and are requesting a dedicated slot are identified.[4] The bound on joint size of T_{R} and T_{pool} is posed by the deadline τ that corresponds to the minimum allowed delay for the identification of the reporting station:

$$\tau > T_{\mathrm{R}} + \max\{T_{\mathrm{pool}}\}, \tag{13}$$

where in the assumed reporting model $\tau = \tau_{\mathrm{a}}$, see (2), and where $T_{\mathrm{pool}} \leq T_{\mathrm{R}}$, see Fig. 1c).

The pool is divided into two parts: the preallocated and the common pool, as depicted in Fig. 5a). The preallocated pool consists of fixed number of RSs, each one dedicated to Ω stations. A station that experience packet arrivals attempts access in the next dedicated RSs when it occurs, i.e., the access is gated. In a straightforward approach, one could design preallocated pool such that there is a RS dedicated to every

station, which corresponds to the case when $\Omega = 1$. Although inherently reliable, this solution is very inefficient. Typically the stations are only periodically reporting with $T_{\mathrm{RI}} >> T_{\mathrm{R}}$, which implies that a station does not have a report to transmit in every pool and, therefore, most of the RSs will be idle (i.e., wasted).[5] The key idea is to preallocate the same RS to a group of $\Omega > 1$ stations, such that the probability of a RS being idle is kept at the tolerable level. The group is defined by the stations AIDs, the allocation of the respective RS is announced in the beacon frame, and the total size of the preallocated pool is $\lceil N/\Omega \rceil$ RSs. The access in the preallocated pool is contention based, where in the worst case there could be Ω stations contending per RS. However, as shown in Section VI, dimensioning of Ω is also done such that, under regular reporting, there is a low probability that two or more stations from the same group will be active in the same RS and cause a collision.

The purpose of the common pool is to foster resolution of collisions that occur in the preallocated pool. The actual resolution mechanism, as well as the size of the common pool depend on the number of the collisions observed in the preallocated pool k_C. In particular, k_C is compared to a predefined threshold Δ_C to decide whether to resolve collisions using a *contention-based* or a *contention-free* approach. If $k_C < \Delta_C$, the AP assumes that the stations perform only regular reporting. For every group of stations that collided in a RS of the preallocated pool, a new frame with $L_1 < \Omega$ RSs is allocated in the common pool, in which the stations from the group contend again by transmitting in a random RSs, see Fig. 5b). In other words, every collision RS from the preallocated pool expands into frame with L_1 slots in the common pool, where the collided stations contend using frame slotted ALOHA approach [15]. If there are collisions in this frame, a second frame with L_2 slots is allocated for their resolution, where $L_2 \leq L_1$, in which collided stations contend again using frame slotted ALOHA approach, see Fig. 5c).[6] Ultimately, if there are still collisions in the second frame, the AP provides a contention-free frame with Ω RSs, i.e., a RS dedicated to each, potentially unresolved station from the original group, as shown in Fig. 5c). On the other hand, if $k_C \geq \Delta_C$, the AP immediately opts for a contention-free strategy, assuming that the collisions are due to an ongoing alarm reporting that affects many stations. In that case, for every collided RS in the preallocated pool, Ω RSs are allocated in the common pool, see Fig. 5d), and each station transmits in its dedicated RS. Finally, once all the active stations have been identified, the AP assigns a data slot for each station, and the operation of the access mechanism continues as in the standard triggered RAW mode.

Ω, Δ_C, L_1 and L_2 are design parameters. In the next section we analyze their optimal dimensioning such that the proposed

---

[4]We do not consider the potential impact of the noise-induced channel errors in this paper, as our focus is on the operation of the access protocol.

[5]Specifically, we assume that there is a maximum of a single arrival per station during T_{R}, and that in regular mode, most of the stations do not experience packet arrivals during T_{R}. This is justified when (1), (2), (13) and the alarm reporting model from Section IV-A are taken into account.

[6]When the collision resolution is frame slotted Aloha-based, a higher efficiency can achieved using two frames with (optimized) lengths of L_1 and L_2 slots, respectively, than providing a single frame with $L = L_1 + L_2$ slots [16].
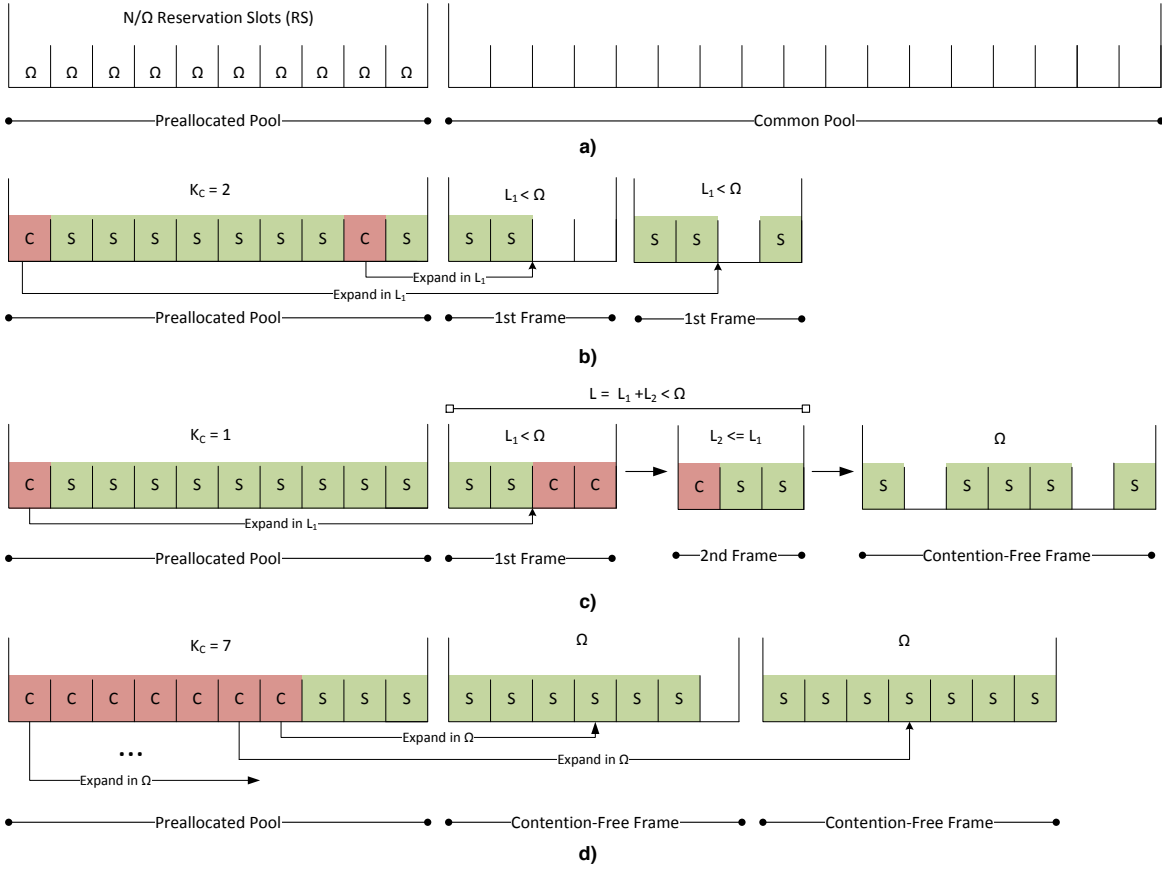
Fig. 5. a) Structure of the pool of resources: the preallocated and the common pool. b) Preallocated pool working under regular regime, where RSs in collision split in L_1 extra RS. c) Preallocated pool working under regular regime, where a collision in the common pool is resolved using frames with L_1, L_2 and, finally, Ω RSs. d) Preallocated pool under alarm regime where every collision RS expands into frame with Ω RSs.

access method is able to detect if there is on-going alarm reporting in the cell, identify all active stations and use the resources efficiently.

VI. ANALYSIS

As already elaborated, the proposed access method tunes its operation to the reporting activity in the cell, i.e., whether there is only the regular or a combination of the regular and alarm reporting in the cell. Denote by H_0 the hypothesis that there is only regular reporting in the cell, and denote by H_1 the alternative hypothesis, i.e., that an alarm event has taken place, affecting a significant number of stations in the cell. The threshold Δ_c is used to determine what is the nature of the reporting activity in the cell, by comparing it to the number of the collisions observed in the periodic pool:

$$H_0 : k_c < \Delta_c, \qquad (14)$$
$$H_1 : k_c \geq \Delta_c. \qquad (15)$$

Let C_{ij} denote the cost of deciding H_i when H_j is true, $i, j = 0, 1$. The costs C_{ij}, $i, j = 0, 1$, represent the number of RSs used in the respective modes of operation of the

access mechanism, given the actual reporting activity. The total expected cost is:

$$E[C] = \sum_{i=0}^{1} \sum_{j=0}^{1} E[C_{ij}] \cdot P(H_i|H_j) \cdot P(H_j), \qquad (16)$$

where $P(H_j)$, $j = 0, 1$, denotes the a priori probability of the hypothesis, which can be calculated using, e.g., the knowledge of the history of alarm events:

$$P(H_1) = \frac{\text{no. of alarms/day}}{\text{no. of pools/day}}, \qquad (17)$$
$$P(H_0) = 1 - P(H_1). \qquad (18)$$

The probability of correctly deciding that there is only regular reporting activity in the cell is:

$$P(H_0|H_0) = P(k_C < \Delta_C|H_0)$$
$$= \sum_{k=0}^{\Delta_c-1} \binom{\lceil N/\Omega \rceil}{k} \cdot P_{C_{H0}}^k \cdot (1 - P_{C_{H0}})^{\lceil N/\Omega \rceil - k}, \qquad (19)$$

where $\lceil N/\Omega \rceil$ is the number of RSs in the preallocated pool and $P_{C_{H0}}$ denotes the probability of a collision in a RS in the regular reporting regime. We assume in (19) that the contention outcomes over RSs independent identically distributed (i.i.d.) random variables; this assumption was verified

by the evaluation that simulates an IEEE 802.11ah cell and the proposed reporting setup, see Section VII. The probability of false alarm detection $P(H_1|H_0)$ is simply:

$$P(H_1|H_0) = 1 - P(H_0|H_0). \qquad (20)$$

The probability of detecting the alarm is:

$$P(H_1|H_1) = P(k_C \geq \Delta_C|H_1)$$
$$= \sum_{k=\Delta_c}^{\lceil N/\Omega \rceil} \binom{\lceil N/\Omega \rceil}{k} \cdot P_{C_{H1}}^k \cdot (1 - P_{C_{H1}})^{\lceil N/\Omega \rceil - k}, \quad (21)$$

where $P_{C_{H1}}$ denotes the probability of collision in a RS in the alarm regime. Again, we assume that the contention outcomes over RSs are (i.i.d.) random variables, which was verified by the evaluation. Finally, the probability of alarm miss $P(H_0|H_1)$ is:

$$P(H_0|H_1) = 1 - P(H_1|H_1). \qquad (22)$$

A. Derivation of Collision Probabilities

The probabilities of collision $P_{C_{H0}}$ and $P_{C_{H1}}$ are the probabilities that there are at least two stations that are active in the same RS:

$$P_{C_{H0}} = 1 - (1 - p_{A,0})^\Omega - \Omega \cdot p_{A,0} \cdot (1 - p_{A,0})^{\Omega-1}, \quad (23)$$
$$P_{C_{H1}} = 1 - (1 - p_{A,1})^\Omega - \Omega \cdot p_{A,1} \cdot (1 - p_{A,1})^{\Omega-1}. \quad (24)$$

where $p_{A,0}$ and $p_{A,1}$ indicate the probabilities of a station reporting when in state 0/state 1, respectively:

$$p_{A,0} = 1 - e^{-\lambda_0 \cdot T_R}, \qquad (25)$$
$$P_{A,1} = 1 - e^{-\lambda(T_R)}, \qquad (26)$$

see Section IV-A.

B. Derivation of Costs

Consider first the cost C_{00}, i.e., the number of RSs used in the contention-based operation of the access method when there is only regular reporting in the cell. C_{00} is computed as:

$$C_{00} = \lceil N/\Omega \rceil + \sum_{i=1}^{k_{C_{00}}} S_i, \qquad (27)$$

where the first term on the right is the size of the preallocated poll in RSs, while the second term corresponds to the size of the common pool. Specifically, $k_{C_{00}}$ is the number of RSs in collision in the preallocated pool, while S_i denotes number of RSs needed to resolve i-th collision from the preallocated pool. $k_{C_{00}}$ and S_i, $i = 1, \ldots, k_{C_{00}}$, are independent random variables, while the latter are also identically distributed. Using Wald's equation [17], the expected value $E[C_{00}]$ can be computed as:

$$E[C_{00}] = \lceil N/\Omega \rceil + E[k_{C_{00}}] E[S]. \qquad (28)$$

We proceed by computation of $E[S]$, while derivation of $E[k_{C_{00}}]$ is presented in Section VI-C.

Recall from Section V that every RS in collision from the preallocated pool expands first into a frame with L_1 RSs, then in a second frame with L_2 RSs if the collision is not resolved

in the first frame, and finally, in a contention-free frame with Ω RSs if the collision is not resolved in the second frame. It is easy to verify that:

$$E[S] = L_1 + L_2(1 - R_1) + \Omega(1 - (R_1 + R_2)), \qquad (29)$$

where R_1 is the probability of resolving the collision in the first frame and R_2 is the probability of resolving the collision in the second frame.

We proceed by deriving R_1 and R_2. Denote by $R(h|m, L)$ the conditional probability that h users are resolved when m users contend in the frame of length L, assuming frame slotted ALOHA-based contention. It could be shown that [18]:

$$R(h|m, L) = \frac{\binom{L}{h}}{L^m} G(L - h, m - h) \prod_{i=0}^{h-1} (m - i), \qquad (30)$$

where:

$$G(u, v) = u^v + \sum_{t=1}^{v} (-1)^t \prod_{j}^{t-1} [(v-j)(u-j)](u-t)^{v-t} \frac{1}{t!}. \quad (31)$$

R_1 can be computed as:

$$R_1 = \sum_{m=0}^{\Omega} R(m|m, L_1) P_A(m), \qquad (32)$$

where $P_A(m)$ denotes the probability that m stations are contending in the first frame. $P_A(m)$ can be computed as:

$$P_A(m) = \begin{cases} \frac{\binom{\Omega}{k} p_{A,0}^k (1 - p_{A,0})^{\Omega-k}}{F(2, \Omega, p_{A,0})}, & 2 \leq m \leq \Omega, \\ 0, & \text{else}, \end{cases} \qquad (33)$$

where:

$$F(l, v, u) = \sum_{t=l}^{v} \binom{v}{t} u^t (1 - u)^{v-t}. \qquad (34)$$

In other words, (33) expresses the fact that there has to be at least two colliding stations if there was a collision. Analogously as in (32), R_2 can be computed as:

$$R_2 = \sum_{m=0}^{\Omega} \sum_{h=2}^{m} R(h|h, L_2) R(m - h|m, L_1) P_A(m). \quad (35)$$

$E[C_{10}]$, i.e., the expected cost when operating in the contention-free mode when there are no alarms is simply:

$$E[C_{10}] = \lceil N/\Omega \rceil + E[k_{C_{10}}] \Omega, \qquad (36)$$

where $E[k_{C_{10}}]$ is the expected number of collision RSs in the preallocated pool, and each of these expands into frame with Ω RSs.

Further, $E[C_{01}]$, i.e., the cost of operating in the contention-based mode when there are alarms, is:

$$E[C_{01}] = \lceil N/\Omega \rceil + E[k_{C_{01}}] (L_1 + L_2 + \Omega), \qquad (37)$$

where we assumed that the probability of resolving collision RSs in the preallocated pool using the first and the second frame in the common pool is negligible, such that every RS in collision is finally resolved using a contention-free frame of Ω RSs.

Finally, $E[C_{11}]$, i.e., the expected cost of operating in contention-free mode when there is an alarm is:

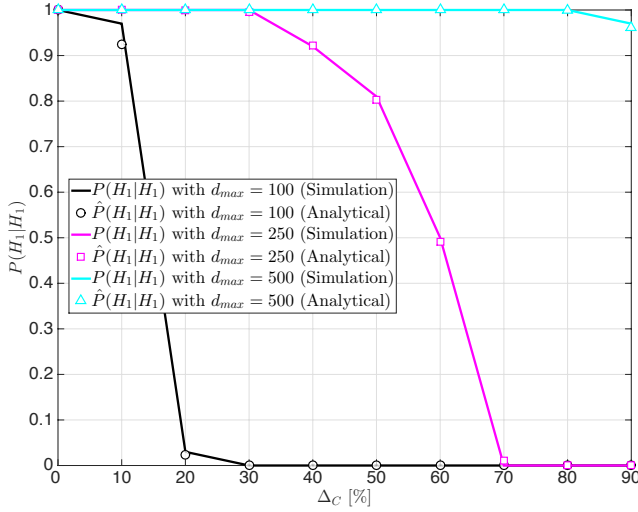$$E[C_{11}] = \lceil N/\Omega \rceil + E[k_{C_{11}}] \Omega. \qquad (38)$$

Fig. 6. $P(H_1|H_1)$ in a 1000-m radius cell with $N = 8000$ stations and $\Omega = 40$. The alarm threshold Δ_C is given as the percentage of RSs in collision in the preallocated pool.

C. Expected Number of Collision RSs

We proceed by deriving the expected number of collision RSs, required for the computation of the expected costs. It can be shown that:

$$E[k_{C_{00}}] = \frac{\sum_{k=0}^{\Delta_C - 1} k \cdot \binom{\lceil \frac{N}{\Omega} \rceil}{k} \cdot P_{C_{H0}}^k \cdot (1 - P_{C_{H0}})^{\lceil \frac{N}{\Omega} \rceil - k}}{F(0, \Delta_C - 1, P_{C_{H0}})},$$
(39)

$$E[k_{C_{10}}] = \frac{\sum_{k=\Delta_C - 1}^{\lceil \frac{N}{\Omega} \rceil} k \cdot \binom{\lceil \frac{N}{\Omega} \rceil}{k} \cdot P_{C_{H0}}^k \cdot (1 - P_{C_{H0}})^{\lceil \frac{N}{\Omega} \rceil - k}}{F(\Delta_C, \lceil \frac{N}{\Omega} \rceil, P_{C_{H0}})},$$
(40)

$$E[k_{C_{01}}] = \frac{\sum_{k=0}^{\Delta_C - 1} k \cdot \binom{\lceil \frac{N}{\Omega} \rceil}{k} \cdot P_{C_{H1}}^k \cdot (1 - P_{C_{H1}})^{\lceil \frac{N}{\Omega} \rceil - k}}{F(0, \Delta_C - 1, P_{C_{H1}})},$$
(41)

$$E[k_{C_{11}}] = \frac{\sum_{k=\Delta_C - 1}^{\lceil \frac{N}{\Omega} \rceil} k \cdot \binom{\lceil \frac{N}{\Omega} \rceil}{k} \cdot P_{C_{H1}}^k \cdot (1 - P_{C_{H1}})^{\lceil \frac{N}{\Omega} \rceil - k}}{F(\Delta_C, \lceil \frac{N}{\Omega} \rceil, P_{C_{H1}})},$$
(42)

where Δ_C is the detection threshold, $P_{C_{H0}}$, $P_{C_{H1}}$ and $F(l, u, v)$ are given in (23), (24) and (34), respectively, and where we assumed that the contention outcomes over RSs are i.i.d. random variables.

VII. EVALUATION

In this section we evaluate the proposed solution in terms of the expected costs $E[C]$ and the probability of alarm detection $P(H_1|H_1)$. The investigated scenario corresponds to a heavily loaded case with 8000 smart meters in a 1000-m radius cell, close to the maximum capacity of a IEEE 802.11ah cell (i.e., 8100 stations). The traffic model of smart meters comprises periodic reporting, on-demand reporting and alarm reporting. For the periodic reporting we use the reporting interval $T_{\mathrm{RI}} = 5\,\mathrm{min}$, thus the corresponding arrival rate is $\lambda_{\mathrm{p}} = 300^{-1}$ report/s, which is a typical value [5]. For the on-demand reporting, RI is set to 15 min, i.e., $\lambda_{\mathrm{d}} =$

1500^{-1} report/s, which is a rather demanding configuration in comparison with typical values [11]. For the alarm reporting we assume a scenario where the alarm propagates at an average speed of $v = 4$ km/s [8], which corresponds to a power outage or power fluctuation in the grid due to, e.g., an earthquake, while the correlation between stations with respect to the distance is given by the function $\Psi^{Sq}(d)$, see (12). According to a relevant description of smart grid traffic messages from [11], we assume that a maximum tolerable delay for alarm is $\tau = 5\,\mathrm{s}$. We adopt a rather conservative approach and illustrate the performance when the pool reoccurring period is set to $T_{\mathrm{R}} = 2.5\,\mathrm{s}$, which allows for the maximum pool duration of $\max T_{\mathrm{pool}} = 2.5\,\mathrm{s}$, see (13). RS duration is set to $200\,\mu\mathrm{s}$, which, at the lowest modulation scheme, is enough to fit the transmission of the 13 bits AID and an uplink data flag bit, and leaves room for the transmission of the additional 23 bits that may be devoted to other information, e.g., a CRC trailer, etc. Finally, we note that all the analytical results presented in this section were verified using simulations of a IEEE 802.11ah cell with the same communication setup.

Fig. 6 illustrates $P(H_1|H_1)$ as function of the alarm detection threshold Δ_C, for varying maximum propagation distance from the epicenter $d_{\max}$. Obviously, Δ_C has to be selected according to d_{max}, as d_{max} affects the number of alarm reporting stations. For example, if we aim to detect an alarm event where $d_{\max} = 250$ m, Δ_C should be set below 30% of RSs in collision, and when $d_{\max} = 500$ m, the alarm is always detected if Δ_C is set below 80% of RSs in collision.

Next, we investigate the efficiency of the proposed mechanism as function of the alarm threshold Δ_C and RS degree Ω. We assume that the frames sizes used for collision resolution in the common pool are $L_1 = 0.6\,\Omega$ and $L_2 = 0.4\,\Omega$; these values were obtained through an exhaustive search optimization that bests fit the arrival rates of the stations and statistics of the number of collision RSs in the preallocated pool in the regular reporting regime. Further, we assume the probability of alarm occurrence is $P(H1) = 5 \cdot 10^{-3}$. Fig. 7 illustrates the expected cost $E[C]$ per pool as function of Ω and Δ_C. Obviously, for low values of Ω the expected cost $E[C]$ is high; this is due to the fact the for low Ω the number of RSs in the preallocated pool is overdimensioned with respect to the reporting activity in the cell and many of these RSs are idle. The worst case is for $\Omega = 1$, when every station has a reserved RS per pool. Increasing Ω sharply reduces the cost, until the breakpoint at roughly around $\Omega = 100$ after which too many collisions start to occur, requiring an increased number of resources in the common pool to resolve them. The alarm threshold Δ_C also plays an important role in the efficiency of the system. If Δ_C is set too low, probability of false alarm detection increases, incurring the waste of resources in the common pool, as every collision RS in the preallocated is automatically expanded in a frame of Ω RSs. On the other hand, with a high Δ_C the probability of alarm miss increases, which also leads to the waste of resources in the common pool. Specifically, in this case there is a high chance that a collision RS from the preallocated pool has to be expanded in frames with L_1, L_2 and Ω RSs. A closer inspection of the results depicted in Fig. 7 reveals that the optimal configuration is $\Omega = 40$ and
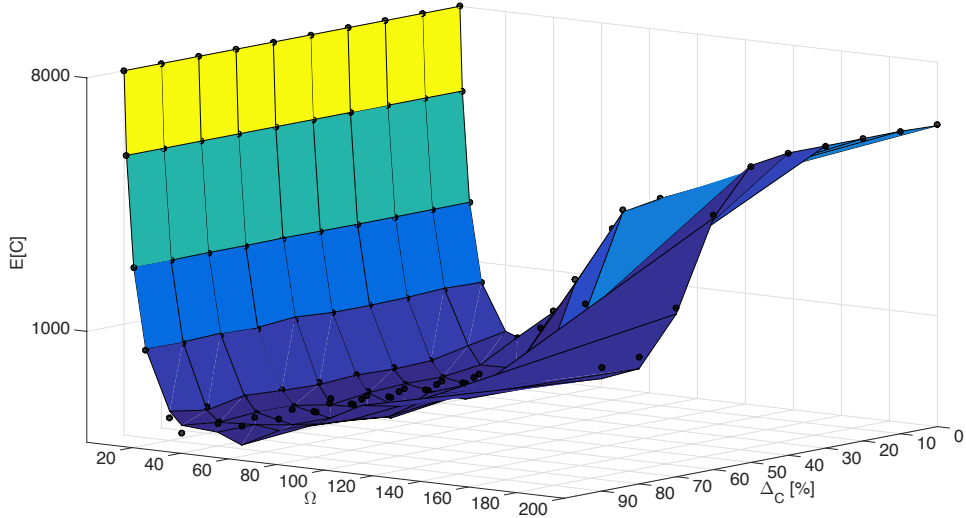
Fig. 7. Expected cost $E[C]$ with respect to the alarm threshold Δ_C as percentage of total RSs in collision and the slot degree Ω, where $L_1 = 0.6\,\Omega$, $L_2 = 0.4\,\Omega$ and the alarm probability $P(H_1) = 5 \cdot 10^{-3}$; the dots indicate analytical results and the solid lines simulation results.

$\Delta_C = 50\%$ of total RSs in collision, where $E[C] \approx 400$ RSs.

A more detailed view of the configuration corresponding to $\Delta_C = 50\%$ of total RSs in collision is shown in Fig. 8. For the sake of comparison we have also included the expected cost of the naive approach in which each collision RS in preallocated pool is always expanded into Ω additional RSs in the common pool (i.e., the access mechanism in the common pool operates only in the contention-free mode). Obviously, the proposed scheme is less sensitive in the variations in Ω and exhibits a superior performance, justifying the use of the contention-based collision resolution in the common pool. Specifically, when the corresponding minimum expected costs are compared, the naive approach requires in average roughly 2 times more RSs than the proposed scheme.

We conclude by commenting on the reliability/delay/efficiency tradeoff that the proposed access method offers. Taking into account the assumed regular reporting interval of 5 min and the period of the pool reoccurring of $T_R = 2.5$ s in the considered scenario, there are 120 pools during the regular reporting interval. When the optimal configuration parameters $\Omega = 40$ and $\Delta_C = 50\%$ are taken into account, the expected cost per pool is $E[C] \approx 400$ RSs, and the pool duration is just $T_{pool} = 80$ ms. Also, on average there are 6 RS are spend per station during the interval of 5 min, i.e., in all 120 pools, while at the same time, all the reporting stations are always resolved within the reporting deadline of $\tau = 5$ s. For the simple polling mechanism that ensures that the same deadline is met, which is for $\Omega = 1$, the cost per pool is 8000 RSs and the pool duration is fixed to $T_{pool} = 1.6$ s, and there are 120 RSs per station during the regular reporting interval. The proposed access method could be further optimized, i.e., T_{RI} could be prolonged and the number of pools reduced, by taking into account the statistics of the pool duration conditioned T_{RI} and dimensioning T_{RI} such that (13) is satisfied. However, we leave such a study for further work. Finally, we note that, if the duration of RS should be increased in order to accommodate longer messages (which could include, e.g., short message
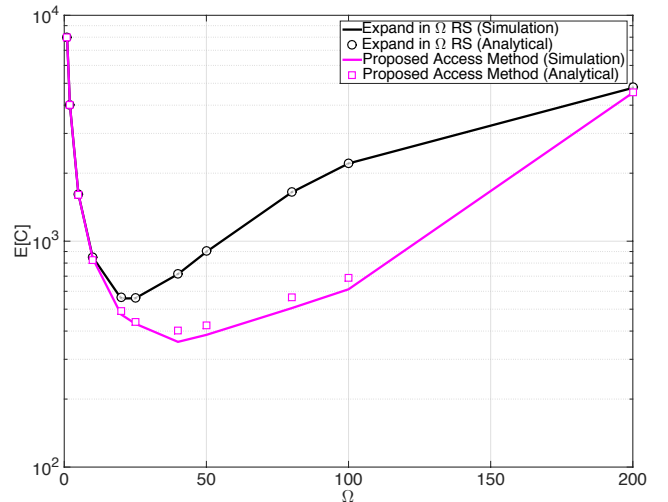

Fig. 8. Expected cost $E[C]$ vs. Ω, for $\Delta_C = 50\%$ of RSs in collision and $P(H_1) = 5 \cdot 10^{-3}$.

identification or other information elements, besides AID and the uplink flag), this reduction in number of required RSs becomes even more important, as it directly influences the pool duration and hence pool reoccurring interval to achieve the reporting deadlines.

VIII. DISCUSSION AND CONCLUSIONS

In this paper we proposed an access mechanism tailored for M2M reporting applications in IEEE 802.11ah networking scenario. Its main features are tuning of the operation to the reporting activity in the call by proactively dimensioning the number of used resources, i.e., slots, while resolving all reporting stations within the given reporting deadline. The proposed solution can be seamlessly incorporated in the triggered RAW operation of IEEE 802.11ah access protocol.

As a side result, we investigated the modeling of alarm reporting in a IEEE 802.11ah cell, using spatio-temporal

models of a alarm propagation and provided justifications for the use of the Beta distribution to model inter-arrival times of alarm reporting. To the best of our knowledge, such a study does not exist in the previous literature.

Finally, we note that the we assumed a static, i.e., fixed placement of the M2M terminals in the cell. However, this assumption is not restrictive and the presented results could be applied to the cases when the terminals are mobile. In particular, the terminal mobility within the cell does not affect the regular reporting model, and, for the alarm reporting model, the spreading of alarms is typically several orders of magnitude faster than the potential velocity of M2M terminals.

REFERENCES

[1] Qualcomm, *White Paper 802.11ah*, 2015 (accessed in July 15, 2015). [Online]. Available: https://www.qualcomm.com/invention/research/projects/wi-fi-evolution/80211ah

[2] T. Adame, A. Bel, B. Bellalta, J. Barcelo, J. Gonzalez, and M. Oliver, "Capacity analysis of IEEE 802.11 ah WLANs for M2M communications," in *Multiple Access Communications*, ser. Lecture Notes in Computer Science. Springer, 2013, vol. 8310, pp. 139–155.

[3] L. Zheng, L. Cai, J. Pan, and M. Ni, "Performance analysis of grouping strategy for dense ieee 802.11 networks," in *Proc. IEEE GLOBECOM 2013*, Atlanta, GA, USA, Dec. 2013.

[4] C. W. Park, D. Hwang, and T.-J. Lee, "Enhancement of ieee 802.11ah mac for m2m communications," *IEEE. Commun. Letters*, vol. 18, no. 7, pp. 1151–1154, Jul. 2014.

[5] G. Corrales Madueño, C. Stefanovic, and P. Popovski, "Reengineering GSM/GPRS Towards a Dedicated Network for Massive Smart Metering," in *Proc. IEEE SmartGridComm 2014*, Venice, Italy, Nov. 2014.

[6] IEEE 802.16p, "IEEE 802.16p Machine to Machine (M2M) Evaluation Methodology Document (EMD)," IEEE 802.16 Broadband Wireless Access Working Group (802.16p), EMD 11/0005, 2011.

[7] G. Madueño, C. Stefanovic, and P. Popovski, "Reliable Reporting for Massive M2M Communications With Periodic Resource Pooling," *IEEE Wireless Commun. Letters*, vol. 3, no. 4, pp. 429–432, Aug. 2014.

[8] 3GPP, "Study on RAN Improvements for Machine-type Communications," 3rd Generation Partnership Project (3GPP), TR 37.868 V11.0, Aug. 2010.

[9] Task Group ah (TGah), "Uplink channel access general procedure 0831r00," 2012.

[10] L. G. Roberts, "Aloha packet system with and without slots and capture," *SIGCOMM Comput. Commun. Rev.*, vol. 5, no. 2, pp. 28–42, Apr. 1975.

[11] Open Smart Grid SG-Network task force, "Smart grid networks system requirements specification, release version 5," 2010.

[12] G. Andrews, R. Askey, and R. Roy, *Special Functions*. Cambridge University Press, 2000.

[13] M. Laner, P. Svoboda, N. Nikaein, and M. Rupp, "Traffic models for machine type communications," in *Proc. ISWCS 2013*, Ilmenau, Germany, Aug. 2013.

[14] L. Asimow and M. Maxwell, *Probability and Statistics with Applications: A Problem Solving Text*, ser. ACTEX academic series. ACTEX Publications, 2010.

[15] H. Okada, Y. Igarashi, and Y. Nakanishi, "Analysis and application of framed ALOHA channel in satellite packet switching networks - FADRA method," *Electronics and Communications in Japan*, vol. 60, pp. 60–72, Aug. 1977.

[16] G. Corrales Madueño, N. K. Pratas, C. Stefanovic, and P. Popovski, "Massive M2M Access with Reliability Guarantees in LTE Systems," in *Proc. IEEE ICC 2015*, London, UK, Jun. 2015.

[17] S. M. Ross, *Introduction to probability and statistics for engineers and scientists*. Academic Press, 2004.

[18] H. Vogt, "Efficient Object Identification with Passive RFID Tags," in *Pervasive Computing*, ser. Lecture Notes on Computer Science. Springer, 2002, vol. 2414, pp. 98–113.