Random Access for Machine-Type Communication based on Bloom Filtering

Pratas, Nuno; Stefanovic, Cedomir; Madueño, Germán Corrales; Popovski, Petar

Published in:
Global Communications Conference (GLOBECOM), 2016 IEEE

DOI (link to publication from Publisher):
10.1109/GLOCOM.2016.7842195

Publication date:
2016

Document Version
Accepted author manuscript, peer reviewed version

Random Access for Machine-Type Communication based on Bloom Filtering

Nuno K. Pratas, Čedomir Stefanović, Germán Corrales Madueño, Petar Popovski

Department of Electronic Systems, Aalborg University, Denmark
Email: {nup,cs,gco,petarp}@es.aau.dk

***Abstract*—We present a random access method inspired on Bloom filters that is suited for Machine-Type Communications (MTC). Each accessing device sends a *signature* during the contention process. A signature is constructed using the Bloom filtering method and contains information on the device identity and the connection establishment cause. We instantiate the proposed method over the current LTE-A access protocol. However, the method is applicable to a more general class of random access protocols that use preambles or other reservation sequences, as expected to be the case in 5G systems. We show that our method utilizes the system resources more efficiently and achieves similar or lower latency of connection establishment in case of synchronous arrivals, compared to the variant of the LTE-A access protocol that is optimized for MTC traffic. A dividend of the proposed method is that allows the base station (BS) to acquire the device identity and the connection establishment cause already in the initial phase of the connection establishment, thereby enabling their differentiated treatment by the BS.**

I. Introduction

Machine-type communications (MTC) are typically characterized by a massive number of machine-type devices that connect to the network to transmit small data payloads. Those features present a significant challenge to cellular networks, whose radio access part is traditionally designed to deal with a rather low number of connections with high data requirements. Specifically, current cellular networks, such as LTE-A, are connection-oriented [1], requiring a connection establishment between the device and the Base Station (BS) before the device can transmit its data packet. As an example, the connection establishment in LTE-A involves a high amount of signaling overhead, which is particularly emphasized when the data payload is small, e.g., less than 1000 bytes [2]. Therefore, in 3GPP it was proposed an approach to optimize the connection establishment by reducing the signaling overhead [3]. The resulting simplified connection establishment protocol starts with the contention-based Access Reservation Protocol (ARP) [4], depicted in the first four steps in Fig. 1(a), followed by a fifth message where the signaling and a small data payload are concatenated. The signaling exchanges related to the security mechanisms are omitted in the optimized version of the LTE-A connection establishment, by reusing an a-priori established security context [2].

The throughput and blocking probability of the ARP are rather sensitive to the number of contending devices. Specifically, the devices contend for access by sending their preambles in a designated and periodically occurring uplink subframe, here termed as random access opportunity (RAO). When the number of contending devices is high [5], multiple devices activate the same preamble in a RAO, which leads to collisions of their RRC Connection Requests, see Fig. 1(a). Consequently, most devices are unable to establish a connection in the first attempt and perform subsequent attempts that, due to the high load, are also likely to result in collisions. and may lead to the collapse of the ARP protocol. A potential solution has been seen in using extended access class barring (EAB) [6], where certain classes of devices are blocked from participating in the ARP at the cost of an increased access latency. Another drawback of the ARP is that the network learns the devices' identities and connection establishment causes only after the RRC Connection Request is successfully received, as the contention is performed via randomly chosen preambles that do not carry information. A solution that allows the network to learn the identities and connection establishment causes of the contending devices already at the beginning of the ARP, could enable their differentiated treatment in later phases of the connection establishment and even skip some of the steps in the random access process of LTE-A, as indicated in Fig. 1.

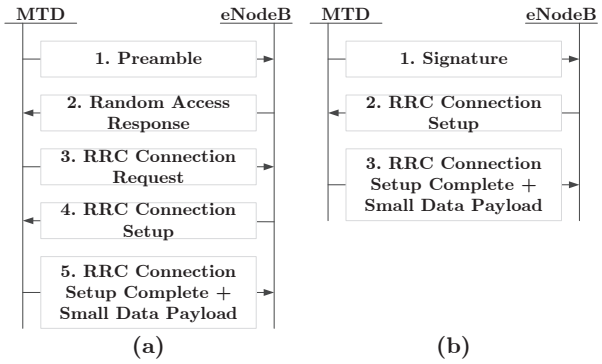

Fig. 1. (a) LTE-A connection establishment protocol optimized for MTC [3] and (b) signature-based modification of LTE-A connection establishment.

In this paper we propose a new access method based on signatures and Bloom filtering[7]. The method is demonstrated in the context of the LTE-A ARP, however, we note that it can be employed in the next generation ARPs [8] following similar principles. In the proposed method, instead of contending with a single preamble in a RAO, the devices contend by transmitting a predefined sequence of preambles in a frame

composed of several RAOs, The transmitted sequence of preambles is denoted as the *device signature*. The presented ideas are a conceptual extension of the work [9], where the devices contend for access by selecting a random signature, generated by combining random preambles over consecutive RAOs. In contrast, in the method described here, each device contends with a unique signature generated using the International Mobile Subscriber Identity (IMSI) of the device and its connection establishment cause, in further text referred to as the device's identification.[1] Specifically, we apply the Bloom-filter [7] principles for signature generation, where the device's identification is hashed over multiple independent hash functions and the resulting output used to select which preamble in which RAO to activate. We introduce an analytical framework through which we tune the signature properties, i.e., its length and the number of activated preambles, based on the number of expected arrivals and the target efficiency of the use of system resources, denoted as the goodput. We also investigate the expected latency and signature detection probability of the proposed method. Finally, we show that, when the arrivals are synchronous, the proposed method outperforms the LTE-A connection establishment procedure in terms of goodput, while achieving similar or lower average latency.

The rest of the paper is organized as follows. Section II summarizes the standard ARP in LTE-A. Section III describes the proposed access method and Section IV presents the corresponding analysis. Section V evaluates the performance of the proposed method, comparing it with the reference LTE-A procedure for MTC traffic. Section VI concludes the paper.

II. LTE-A Access Reservation Procedure

A successful LTE-A access reservation entails the exchange of four messages[2], as depicted in Fig. 1(a). Initially, a device randomly choses a preamble to be transmitted in a RAO from a set of available preambles generated using Zadoff-Chu sequences [10]. The preambles are orthogonal and can be simultaneously detected by the BS. We also note that the BS is able to detect a preamble even when it is transmitted by multiple devices [1], [9], i.e., a collision in the "preamble space" is still interpreted as an activated preamble. This represents a logical OR operation, since the preamble is detected as activated if there is *at least* one device that transmits the preamble. This observation motivates the use of Bloom filter, a data structure based on OR operation for testing set membership.

The devices whose preambles are detected are notified via a Random Access Response (RAR) in the downlink and assigned a temporary network identifier. The reception of the RAR triggers the transmission of the RRC Connection Request in the allocated uplink sub-frame. At this point, the BS is able to detect the collision of the multiple connection requests, sent by the devices that originally sent the same preamble. The

[1]We note that the proposed method can be straightforwardly applied to cases where some other information is used for signature generation.

[2]For the sake of brevity, we omit the details that are nonessential for the proposed method, such as the power ramping procedure etc.

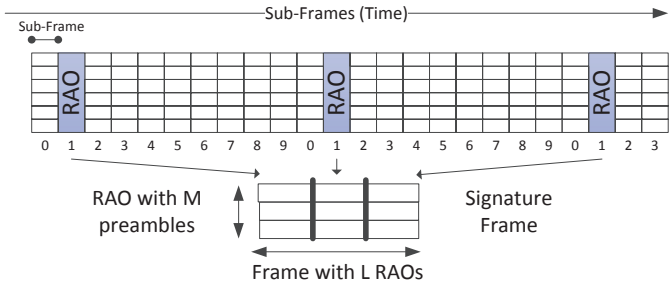

Fig. 2. Illustration of the mapping of the LTE-A preambles into a signature frame composed by multiple RAOs.

successfully received connection requests are acknowledged, marking the start of the data transmission phase. On the other hand, the devices whose connection requests collided, do not receive the feedback and either contend again by sending a new preamble or end up in outage when the number of connection attempts reaches the predefined limit. In the RRC Connection Request, the device informs the network of its temporary identifier, IMSI, and the connection establishment cause. From these, the network can confirm if the device is authorized for access, track the device's subscribed services and reestablish the preexisting security context [2].

As already mentioned, the channel over which the devices contend can be modeled as an OR multiple access channel (OR-MAC). By $A = \{a_i, i = 0, 1, ..., M\}$, denote the set of available preambles, where the absence of preamble activation is denoted by the idle preamble a_0. Assume that there are T devices in total. We model the contention by assuming that the device h, $h = 1, \ldots, T$, transmits a binary word

$$\mathbf{x}^{(h)} = [x_0^{(h)}, x_1^{(h)}, \cdots, x_M^{(h)}], \tag{1}$$

where bit $x^{(h)} = 1$ indicates if the device h transmitted preamble a_i. Note that only a single entry $x_i^{(h)}$, $0 \leq i \leq M$, can be set to 1 since a device can only transmit a single preamble in a single RAO. The BS observes

$$\mathbf{y} = \bigoplus_{h=1}^{T} \hat{\mathbf{x}}^{(h)}, \tag{2}$$

where $\bigoplus$ denotes a bit-wise OR operator and $\hat{\mathbf{x}}^{(h)}$ is the detected binary word of device h. In particular, the BS detects a transmitted preamble with probability $p_d \leq 1$ and with probability $p_f \geq 0$ falsely detects a non-transmitted preamble, which may cause that $\mathbf{x}^{(h)} \neq \hat{\mathbf{x}}^{(h)}$. In practice, the preamble detection at the BS should ensure that $p_d > 0.99$ and $p_f < 10^{-3}$ [11][3]. Finally, every non-zero entry in $\mathbf{y}$ implies a detection of the corresponding preamble. Obviously, in the best-case scenario, the BS can detect up to M different devices in a RAO.

III. The Proposed Method

The essence of the proposed method lies in the idea of devices contending with combinations of K preambles trans-

[3]The p_d requirement in [11] corresponds to the single activation of a preamble. When a preamble is activated by multiple devices it is expected that the effective p_d will be higher [1].

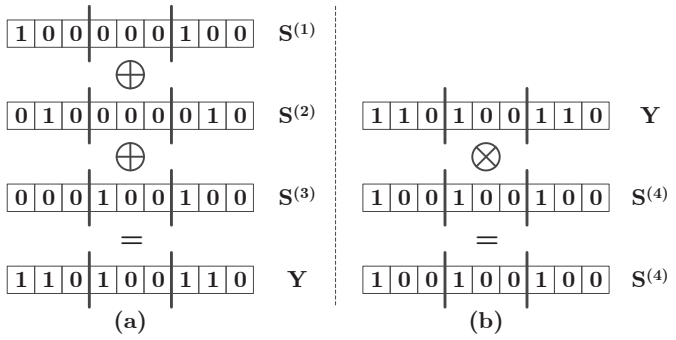

Fig. 3. Example of: (a) synchronous transmission of 3 signatures when $L = 3$ and $M = 3$ and (b) erroneous decoding of a signature which was not present in the original transmission ($p_d = 1$ and $p_f = 0$).

mitted over L RAOs, denoted as signatures. Each preamble of a signature is sent in a separate RAO, while L RAOs define a signature frame, see Fig. 2. Extending the model introduced in Section II, the device h contends by transmitting its signature

$$\mathbf{s}^{(h)} = [\mathbf{x}_1^{(h)}, \mathbf{x}_2^{(h)}, \cdots, \mathbf{x}_L^{(h)}], \tag{3}$$

where the binary words $\mathbf{x}_i^{(h)}$, $i = 1, \ldots, L$, follow the structure introduced in (1). Obviously, the number of available signatures is $\binom{L}{K} M^K$, potentially allowing for the detection of exponentially more contenders compared to the case in which the preambles sent in each of the L RAOs are treated independently and where the maximal number of detected contenders is $L\,M$.

Similarly to (2), the BS observes

$$\mathbf{y} = \bigoplus_{h=1}^{N} \hat{\mathbf{s}}^{(h)}, \tag{4}$$

where $\hat{\mathbf{s}}^{(h)}$ is the detected version of $\mathbf{s}^{(h)}$. The BS decodes all signatures $\mathbf{s}$ for which the following holds

$$\mathbf{s} = \mathbf{s} \bigotimes \mathbf{y}, \tag{5}$$

where $\bigotimes$ is the bit-wise AND.

At this point, we turn to a phenomenon intrinsically related to the proposed contention method [9]. Namely, even in the case of perfect preamble detection ($p_d = 1$) and no false detections ($p_f = 0$), the BS may in general case also decode signatures that have *not* been transmitted but for which (5) also holds. In other words, the BS may decode *false positives*. An example of this is shown in Fig. 3. The performance of the random signature construction in terms of probability of decoding false positives was first analyzed in [9], where they are referred to as phantom sequences. On the other hand, there is an extensive work on the construction of OR-MAC signatures [12] based on the following criterion: if up to N-out-of-T signatures are active, then there are no false positives. However, these constructions are not directly applicable to the LTE-A access, as they would (1) require that a device sends multiple preambles in the same RAO, and (2) imply rather long signature lengths, i.e., $\frac{N^2 \log_2 T}{2M \log_2 N} \leq L \leq \frac{N^2 \log_2 T}{M \ln 2}$, which implies an increased access latency. Inspired by Bloom filters [7], we propose a novel signature construction that uses much lower signature lengths, at the expense of introducing false positives in a controlled manner.

Signature Construction based on Bloom Filtering

In the proposed method, the device signature is constructed in such a way that it provides a representation of the device's identification, which is assumed to be a-priori known to the network. To illustrate how a signature is constructed, we first consider the case where a single preamble is available at each of the L RAOs dedicated to the signature transmission, i.e., $M = 1$. Taking the view of the device h, we start with the binary array $\mathbf{s}^{(h)}$ of length L, indexed from 1 to L, where all the bits are initially set to 0. We then activate K index positions in this array, i.e., we set them to 1; note that K is a predefined constant valid for all devices. This is done by using K independent hash functions, $f_j(\mathbf{u}^h)$, $j = 1, \ldots, K$, whose output is an integer value between 1 and L, corresponding to an index position in the array, and where $\mathbf{u}^{(h)}$ is representation of the device identity. The resulting binary array becomes the device signature. This construction follows the same steps as the object insertion operation in a Bloom filter [7].

When $M > 1$, the signature construction occurs in two stages. The first stage corresponds to the selection of the K active RAOs using hash functions $f_j(\mathbf{u}^h)$, $j = 1, \ldots, K$, as described previously. In the second stage, for each of the activated RAOs, a contending device selects and transmits randomly one of M preambles. This is performed by hashing the device identity using another set of independent hash functions $g_j(\mathbf{u}^h)$, $j = 1, \ldots, K$, i.e., a separate hash function for each RAO, whose output is an integer between 1 and M that corresponds to one of the available preambles.

Signature-Based ARP

The signature-based access reservation protocol is depicted in Fig. 1(b), which starts by the devices transmitting their signatures. Upon the successful decoding of a signature, the BS transmits the *RRC Connection Setup* message. In contrast with the LTE-A ARP depicted in Fig. 1(a), the messages 2 and 3 are not required in the signature based access, since the BS is able to determine from the signature the IMSI of the device and the connection establishment cause. The protocol concludes with the transmission of the small data payload together with the completion of the RRC connection message.

Practical Considerations

The described signature generation raises two important issues: (i) out of K hash functions $f_j(\mathbf{u}^h)$, $j = 1, \ldots, K$, there is a probability of $1 - \frac{K!\binom{L}{K}}{L^K}$ that at least two of these functions generating the same output, leading to less than K distinct RAOs active in a signature; (ii) there is a non-zero probability that two or more devices share the same signature, given by

$$\sum_{i=2}^{T} \binom{T}{i} p^i (1-p)^{T-i} \text{ with } p = \left[\binom{L}{K} (M)^K \right]^{-1} \tag{6}$$

Algorithm 1: Signature generation for h^{th} device, where $\mathbf{u}^{(h)}$ is the device's identification and $x_{i,m}^{(h)}$ indicates activation of m^{th} preamble in i^{th} RAO of the signature $\mathbf{s}^{(h)}$.

Input: $\mathbf{u}^{(h)}$, L, M, K;
Initialize: $\mathbf{s}^{(h)} \leftarrow \mathbf{0}$, $\mathbf{L} \leftarrow 1...L$, $\mathbf{M} \leftarrow 1...M$;
for $j : 1 \cdots K$ **do**
  $i \leftarrow \mathbf{L}(\text{mod}(\mathbf{u}^{(h)}, L+1-j))$;
  $\mathbf{L} = \mathbf{L} \setminus \{i\}$;
  $m \leftarrow \mathbf{M}(\text{mod}(\mathbf{u}^{(h)}, M+1-j))$;
  $\mathbf{M} = \mathbf{M} \setminus \{m\}$;
  $x_{i,m}^{(h)} = 1$;
Output $\mathbf{s}^{(h)}$;

and T as the total number of devices. The above probabilities can be minimized by increasing the signature length L, which is the reason why these issues are commonly ignored within the Bloom filter related literature, where L is of the order of 10^4. Although we do not use such large ranges for L, we note that for values of $L > 10$ and $5 < K < L$ that are used in the performance evaluation in Section V, the second probability can be neglected, as in this case $T \ll \binom{L}{K}(M)^K$.

The first issue can be addressed by a signature construction that enforces K distinct active RAOs per signature. We provide in Alg. 1 a description of a practical signature construction that uses the modulus operation as basis for hashing. This construction ensures that K distinct RAOs are active per signature, by removing the RAOs selected in previous iterations from the set of available RAOs. Further, the preambles activated in previously selected RAOs are removed from the set preambles available for the next iteration. This operation limits the generation of signatures to $K \leq \min(M, L)$ active RAOs; however, this is within the operating range of interest where $K < M$ and allows us to apply probabilistic tools, as presented in the analysis in Section IV, to design the signatures length L and number of active RAOs K. As it will be shown in Section V, the proposed signature generation algorithm matches well the derived analytical model.

Finally, we note that an essential prerequisite for the proposed signature access scheme is that the signature generation algorithm and all hash functions are known to all devices, including the BS. This can be accomplished via the existing periodic broadcasts that include the network configuration; an alternative would be to include this information already in the device's subscriber identity module.

IV. Analysis

We analyze a single instance of the contention process, assuming a synchronous batch arrival of N_a devices. We assume that the probability of an arrival of a device is $p_a = \frac{\text{E}[N_\text{a}]}{T}$, and denote the expected number of arrivals as $N = \text{E}[N_\text{a}]$. The parameters of the proposed scheme are the signature frame size, denoted by L, the number of active RAOs in the signature, denoted by K, and the number of preambles per RAO that are available for signature construction, denoted by M. The first two parameters are subject to design, and we analyze their dimensioning when on average N-out-of-T signatures are active, such that the false positive rate is below a threshold. In contrast, M is assumed to be fixed, which corresponds to the typical scenario in LTE-A systems.

We start by establishing the relationship between the correctly detected signatures and all detected signatures, which also includes the false positives, after all the contenders have completed step 2 of the proposed method, see Fig. 1(b). We denote this metric as the goodput G. In essence, the goodput reflects the efficiency of the subsequent small data transmission, as the BS will also attempt to serve the falsely detected signatures. The expected goodput is

$$\text{E}[G] = \text{E}\left[\frac{N_\text{a}}{N_\text{a}+P}\right] \approx \frac{\text{E}[N_\text{a}]}{\text{E}[N_\text{a}]+\text{E}[P]} = \frac{N}{N+\text{E}[P]}. \quad (7)$$

where P is the number of false positives. From (7) it follows

$$\frac{N}{T} \leq \text{E}[G] \leq 1, \quad (8)$$

as there can be no more than T detected signatures. The mean number of false positives $E[P]$ can be approximated as

$$E[P] \approx p_\text{fa}(T-N),$$

where $T-N$ corresponds to the mean number of inactive signatures, while p_fa denotes the false positive probability, i.e., the probability of an inactive signature being perceived as active. Eq. (7) now becomes

$$\text{E}[G] \approx \frac{N}{N+p_\text{fa}(T-N)}. \quad (9)$$

Using (9), we proceed by setting the target goodput $\hat{G}$ and establishing the relation between $\hat{G}$ and the corresponding target $\hat{p}_\text{fa}$

$$\hat{p}_\text{fa} = \frac{N(1-\hat{G})}{(T-N)\hat{G}}. \quad (10)$$

To compute p_fa, we rely on approximations that hold when the number of simultaneously active signatures N is high enough. Specifically, p_fa is the probability that all K preambles associated with an inactive signature, are detected as activated by the BS. Each of these K preambles can be (i) actually activated by an active signature and detected as such by the BS, or (ii) not activated by any of the active signatures, but falsely detected as activated by the BS. Now, the probability that a particular preamble in a particular RAO is not activated by any of the signatures, denoted by p_idle, is

$$p_\text{idle} = \left(1 - \frac{K}{L\,M}\right)^N, \quad (11)$$

where $L \cdot M$ is the total number of preambles in L RAOs, K is the number of preamble activations per user, N is the number of active signatures, and it is assumed that the selection of any preamble in any RAO is equally likely. The detection of a preamble is non-ideal and therefore we have to distinguish between two events: (i) detection of a preamble transmitted by

Algorithm 2: Iterative signature decoding where $\mathbf{S}$ is the set of signatures and $\mathbf{D}$ is the set of decoded signatures.

1 **Input**: $\mathbf{S}$, $\mathbf{y}$, L, M, K;
2 **Initialize**: $\mathbf{V} = \mathbf{S}$, $\mathbf{D} = \emptyset$;
3 **for** $i : 1 \cdots L\,M$ **do**
4 **for** $\mathbf{s^{(h)}} \in \mathbf{V} \setminus \mathbf{D}$ **do**
5 **if** $\mathbf{s^{(h)}}(1:i) \neq \mathbf{s^{(h)}}(1:i) \bigotimes \mathbf{y}(1:i)$ **then**
6 $\mathbf{V} = \mathbf{V} \setminus \{\mathbf{s^{(h)}}\}$;
7 **if** $\left(\mathbf{V} \setminus \mathbf{s^{(h)}}(1:i)\right) \bigotimes \mathbf{y}(1:i) \neq \mathbf{y}(1:i)$ **then**
8 $\mathbf{D} = \mathbf{D} \cup \{\mathbf{s^{(h)}}\}$;
9 Report to $\mathbf{u^{(h)}}$ that $\mathbf{s^{(h)}}$ is decoded;
10 **for** $\mathbf{s^{(h)}} \in \mathbf{V} \setminus \mathbf{D}$ **do**
11 $\mathbf{D} = \mathbf{D} \cup \{\mathbf{s^{(h)}}\}$; Report to $\mathbf{u^{(h)}}$ that $\mathbf{s^{(h)}}$ is decoded;

at least one device with probability p_d; (ii) false detection of a non-transmitted preamble with probability p_f. We approximate p_{fa} as

$$p_{\text{fa}} \overset{(a)}{\approx} \left[(1 - p_{\text{idle}}) \cdot p_d + p_{\text{idle}} \cdot p_f\right]^K = \left[p_d + (p_f - p_d) \cdot p_{\text{idle}}\right]^K, \tag{12}$$

and where (a) becomes a lower bound when $M = 1$ and $p_d = 1$ and $p_f = 0$ [13]. From (12), the required signature frame size $\hat{L}$ to meet the target $\hat{p}_{\text{fa}}$ is

$$\hat{L} = \frac{K}{M}\left[1 - \left(\frac{\hat{p}_{\text{fa}}^{1/K} - p_d}{p_f - p_d}\right)^{1/N}\right]^{-1} \tag{13}$$

To compute the K that minimizes $\hat{L}$ in (13), we assume $p_d = 1$ and $p_f = 0$. Then, for a given N and L, the value of K that minimizes p_{fa} is given by [14]

$$K_{\min} = \frac{L\,M}{N} \ln 2 \tag{14}$$

We use (14) to find the minimal required $\hat{L}$ via (13). Furthermore, recall that each device can only activate up to a single preamble per RAO, resulting in the constraint

$$K_{\min} = L \min\left(1, \frac{M}{N} \ln 2\right), \tag{15}$$

where we assume to work in the regime in which $\frac{M}{N} \ln 2 < 1$, i.e., where $N > M \ln 2$. Now, the minimum $\hat{L}$ can be obtained by solving iteratively the following fixed-point equation obtained from combining (13) and (14)

$$\hat{L} = \left\lceil \frac{\lceil K_{\min} \rceil}{M}\left[1 - \left(\frac{p_{fa}^{1/\lceil K_{\min} \rceil} - p_d}{p_f - p_d}\right)^{1/N}\right]^{-1} \right\rceil, \tag{16}$$

which converges for $p_d \geq 0.99$ and $p_f \leq 10^{-3}$, i.e., the prescribed preamble detection performance [11].

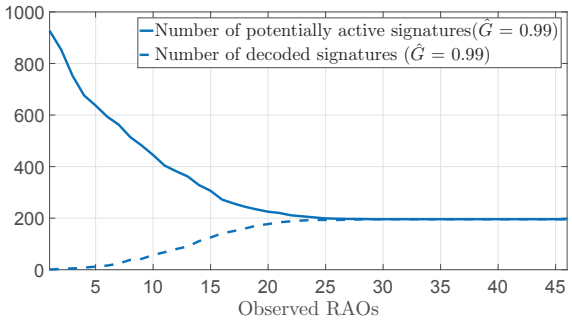

Fig. 4. Evolution of the number of potentially active and already decoded signatures by the BS as the RAOs of the signature frame elapse, for $T = 1000$, $N = 200$, $\hat{G} = 0.99$, $p_d = 0.99$, $p_f = 10^{-3}$, and $\hat{L} = 47$ from (16).

A. Signature Decoding

A straightforward approach for signature decoding is to perform it after all RAOs of the signature frame have been received, i.e., after BS has observed the whole signature frame. An alternative is to perform the decoding iteratively after every received signature RAO, i.e., the BS attempts to decode a signature while only having access to a partial observation of the signature frame. The latter strategy is inspired with the fact that K active RAOs constituting a signature are randomly spread over the signature frame and, in principle, the BS not have to wait until the end of the frame to detect a signature. The decoding performance is the same for both strategies when all L RAOs in the signature frame have been received, but the average latency in the latter approach is lower. We provide in Alg. 2 an algorithmic description of the iterative signature decoding, where the notation $\mathbf{z}(1:i)$ corresponds to the first i entries of vector $\mathbf{z}$. The key steps of the Alg. 2 are steps 5 and 7. In particular, in step 5 the BS discards the signatures that could not have generated the partial observation $\mathbf{y}(1:i)$ from the set of potentially active signatures V. Obviously, it is expected that V will decrease with the additional received RAOs. In step 7, the BS detects the signatures whose combinations of active RAOs and preambles are uniquely contributing to the partial observation $\mathbf{y}(1:i)$. Then the BS reports to the respective device that its signature has been decoded, which in the LTE-A protocol realization would correspond to the RRC Connection Setup message, as shown in Fig. 1(b). Finally, in steps 10–12, when all RAOs have been received, the BS reports all the signatures within the set $\mathbf{V} \setminus \mathbf{D}$ as decoded.

In Fig. 4, we provide a simulation snapshot showing how many signatures are considered potentially active and how many have actually been decoded as the RAOs of the signature frame elapse. Obviously, the iterative signature decoding occurs in a spread manner, which leads to the spreading of the feedback messages acknowledging the decoding of each signature, i.e., the RRC Connection Setup message in Fig. 1(b). In this way, the scenario in which a high number of devices attempt to complete the access reservation protocol simultaneously is avoided, i.e., the occurrence of congestion at the later stages of the ARP is reduced. Another important

observation is that most of the signatures become decoded well before the end of the signature frame.

V. Performance Evaluation

A. Scenario description

In order to evaluate the performance of the proposed signature based access and compare it with the proposed 3GPP LTE-A solution for MTC traffic [2], we have implemented an event driven simulator where the main downlink and uplink LTE channels are modeled. Specifically, the simulator implements the both procedures depicted in Fig. 1(a) and Fig. 1(b), while the downlink control and data channels (PDCCH and PDSCH respectively) and the uplink data and random access channels (PUSCH and PRACH) are modeled as in [2].

We consider a typical cell, configured with one RAO every 5 ms, $M = 54$ available preambles for contention [2]. We assume a total population of size $T = 1000$, and a batch arrival of N_a devices with a payload of 100 bytes to transmit, The arrival probability of an individual device is given by $p_a = N/T$, i.e., N_a is a binomially distributed random variable with mean $\mathrm{E}[N_a] = N$. The mean number of arrivals N is assumed to be known, and the signature based scheme is dimensioned for it.[4] The probability of preamble detection by the BS is set to $p_d = 0.99$ and the probability of false detection of a preamble is set to $p_f = 10^{-3}$ [11].

In the baseline, i.e., 3GPP scheme, we assume the typical values for the backoff window of 20 ms and the maximum number of 10 connection attempts [2]. The devices upon becoming active contend for access by activating randomly one preamble in one of the available RAOs within the backoff interval, i.e., the batch arrival is spread with the backoff interval.[5] In case that a device is the only one that selected a given preamble in a given RAO and that this preamble has been detected, then the access procedure, as depicted in Fig 1(a), proceeds until completion. Otherwise, the device will reattempt the access within the back-off window after the timer to receive the RAR as elapsed. When multiple devices select the same preamble within a RAO, the resources assigned by the BS corresponding to the step 3 in the protocol are wasted due to the collided devices; and the collided devices re-attempt access later by selecting a random RAO within the backoff interval. The devices re-attempt access until either successful or until exceeding the allowed number of retransmissions.

In the proposed method, the devices contend by transmitting their signatures, where the signature frame length L is obtained from (16). For the sake of comparison, we also evaluate the performance of the random signature construction [9], where $K = L$. Each device upon its signature being decoded, even in the case of false positive, receives the feedback RRC connection setup message and is assigned uplink data resources for the transmission of the third and final message, see Fig 1(b).

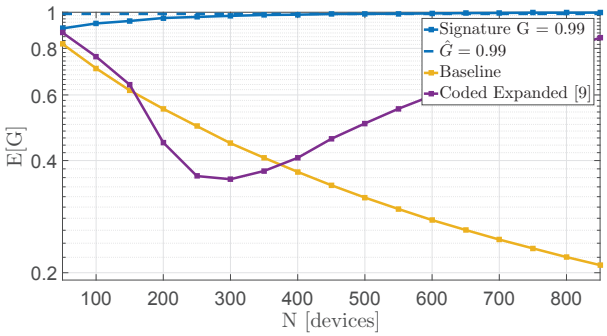

Fig. 5. $E[G]$ observed with increasing N, for the 3GPP scheme, random signature construction [9] and the proposed signature construction. ($T = 1000$)

The performance is evaluated in terms of: (i) the average goodput $E[G]$; (ii) the average latency until the first step in both access schemes is successful, corresponding to a singleton preamble in the baseline and a successfully decoded signature in the proposed scheme; (iii) the average latency until the small data transmission takes place, corresponding to step 5 in the baseline and to step 3 in the proposed scheme, see Fig 1; and (iv) probability of device being successfully detected upon the completion of the access protocol.

The average goodput $E[G]$ is evaluated as the ratio between the successfully used resources and the total resources spent in the third step of both access protocols. It directly relates to the efficient use of resources, since the BS is only able to discern if there is a correctly detected device upon successful completion of the third step. In the baseline scheme, the system resources are wasted whenever two or more devices select the same preamble within a RAO; the goodput in this case is given as the ratio between the total number of messages that are exchanged successfully and the total number of exchanged messages at the third step, including the failed ones due to collisions. In the case of the signature based access, the wasted resources in the third step occur whenever a false positive signature occurs, and the goodput is given by (7).

B. Results

The expected goodput is depicted in Fig. 5, where for the goodput target for the proposed method (10) is set to $\hat{G} = 0.99$. We observe that the proposed method meets the actual goodput meets the design target at higher access loads. On the other hand, at lower N, the performance deviates from the target value $\hat{G} = 0.99$. This is due to the assumption that the false positive signatures are independently and uniformly generated from the idle signatures, which is the basis of the approximation in (12). We can also observe that the goodput performance of the proposed method is always superior to the 3GPP scheme. Specifically, In the 3GPP scheme the devices re-attempt retransmission upon colliding and until they are either successful or the number of retransmissions is exceeded. Each subsequent failed retransmission results in additional wasted system resources, which results in the

[4] N can be estimated, e.g., using techniques that take advantage of the LTE-A ARP, such as the one proposed in [15].

[5] Note that this initial backoff is a modification of the original LTE-A access procedure, in which the devices contend by activating a preamble in the nearest RAO [16]. The purpose of this modification is to force a spread in the batch arrival and prevent the consequent imminent collision; the resulting performance of the baseline scheme is actually better than it could be expected.

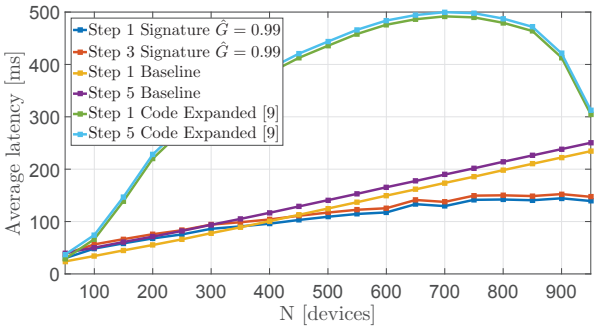

Fig. 6. Mean latency of the 3GPP scheme, random signature construction and the proposed signature construction with optimal K and minimum $\hat{L}$ computed from (16), at different stages of the access procedures. ($T = 1000$)

observed degradation of the baseline goodput with increasing number of active devices. Finally, the goodput achieved with the random signature construction [9] is quite low, due to the high number of false positives.

In Fig. 6 we depict the mean latency at step 1 in all schemes, as well as in steps 3 and 5 in the signature and 3GPP schemes, respectively. An important observation is that the latency of the proposed method at first follows the same trend as the latency of the 3GPP scheme, being only modestly higher. After certain threshold in N is surpassed, the latency of the proposed method actually becomes lower and even starts to drop. This is a consequence of the more efficient detection of active users, as can be seen when comparing the latency of these two schemes at step 1. Furthermore, the random signature construction has the worst performance, the reason being that a signature cannot be decoded before all L RAOs of the signature frame have been received [9].

Finally, in Tab. I we show the probability of a device being successfully detected at end of the access protocol. Here the proposed method has a slight performance degradation compared to the 3GPP scheme, but this degradation diminishes higher access loads. The 3GPP scheme achieves higher detection performance due to only requiring one transmission out of all preamble retransmissions to be successful, making it more robust but at the cost of lower goodput and higher latency. On the other hand, the random signature construction leads to a very low detection performance, as it requires the successful detection of all the active preambles [9].

VI. Discussion and Conclusions

Following the insights provided by Bloom filters, we have introduced the concept of signatures with probabilistic guarantees and applied it to a system model derived from the LTE-A access reservation protocol. The most important feature of the proposed method is in allowing the device to be identified already at the access stage. Moreover, the method is very efficient in terms of use of the system resources and has a favorable performance in terms of decoding latency.

In the paper we assumed that the base station serves the successfully connected devices without preferences. Nevertheless, it is straightforward to modify the proposed solution to scenarios in which the BS serves devices based on the identifications inferred from the decoded signatures, i.e., IMSIs and/or connection establishment causes. In such cases, the proposed access method enables differentiated treatment by the BS from the very beginning.

N	100	300	500	700	900
Proposed method	96	98	98	98	98
3GPP scheme	100	100	100	100	100
Random construction [9]	86	53	42	37	44

TABLE I
Probability of successfully detecting a device [%]. (T = 1000)

Finally, we note that in the paper we assessed a simplified scenario of a synchronous bath arrival in order to present the key concepts and the related analysis. Tuning the proposed scheme for the other typical models, like the Beta arrival model for synchronous arrivals or the Poisson arrival model for asynchronous arrivals, is left for further work.

Acknowledgment

This work was performed partly in the framework of H2020 project FANTASTIC-5G (ICT-671660), partly supported by the Danish Council for Independent Research grant no. DFF-4005-00281 "Evolving wireless cellular systems for smart grid communications" and by the European Research Council (ERC Consolidator Grant Nr. 648382 WILLOW) within the Horizon 2020 Program. The authors acknowledge the contributions of the colleagues in FANTASTIC-5G.

References

[1] D. T. Wiriaatmadja and K. W. Choi, "Hybrid Random Access and Data Transmission Protocol for Machine-to-Machine Communications in Cellular Networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, p. 3346, Jan. 2015.
[2] 3GPP, "TR37.869 - Study on enhancements to machine-type communications (MTC),", 2013.
[3] ——, "TR36.888 - Study on provision of low-cost Machine-Type Communications (MTC) User Equipments (UEs) based on LTE,", 2012.
[4] ——, "TS36.321 - Medium Access Control (MAC) protocol specification.", 2014.
[5] G. C. Madueno, J. J. Nielsen, D. M. Kim, N. K. Pratas, C. Stefanovic, and P. Popovski, "Assessment of LTE wireless access for monitoring of energy distribution in the smart grid," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 675–688, March 2016.
[6] 3GPP, "Radio Resource Control (RRC)," TS TS-36331, 2010.
[7] B. H. Bloom, "Space/Time Trade-offs in Hash Coding with Allowable Errors," *Commun. ACM*, vol. 13, no. 7, pp. 422–426, Jul. 1970.
[8] FANTASTIC-5G, "IR4.1 - technical results for service specific multi-node/multi-antenna solutions,", 2016.
[9] H. Thomsen, N. K. Pratas, C. Stefanovic, and P. Popovski, "Code-expanded radio access protocol for machine-to-machine communications," *Transactions on Emerging Telecommunications Technologies*, vol. 24, no. 4, pp. 355–365, 2013.
[10] D. Chu, "Polyphase codes with good periodic correlation properties (Corresp.)," *IEEE Trans. Info. Theory*, vol. 18, no. 4, pp. 531–532, Jul 1972.
[11] 3GPP, "TS36.141 - Base Station (BS) conformance testing,", 2016.
[12] S. Gyori, "Coding for a multiple access OR channel: A survey," *Discrete Applied Mathematics*, vol. 156, no. 9, pp. 1407 – 1430, 2008.
[13] K. Christensen, A. Roginsky, and M. Jimeno, "A New Analysis of the False Positive Rate of a Bloom Filter," *Inf. Process. Lett.*, vol. 110, no. 21, pp. 944–949, Oct. 2010.
[14] M. Mitzenmacher, "Compressed Bloom Filters," *IEEE/ACM Trans. Networking*, no. 5, pp. 604–612, Dec. 2002.
[15] G. C. Madueño, N. K. Pratas, C. Stefanovic, and P. Popovski, "Massive M2M Access with Reliability Guarantees in LTE Systems," *IEEE International Conference on Communications (ICC), 2015*, 2015.
[16] 3GPP, "TR37.868 - Study on ran improvements for Machine-Type Communications (MTC) User Equipments (UEs) based on LTE,", 2011.