Pricing-based Load Control of M2M Traffic for the LTE-A Random Access Channel

Mehmet Koseoglu, *Member, IEEE*

***Abstract*—Existing cellular infrastructures have to be revisited for emerging machine-to-machine (M2M) traffic as semi-synchronized M2M arrivals may create a significant congestion resulting in a high access delay. In such a case, there is a strong need for service differentiation since the delay requirements of IoT applications may vary greatly from delay-tolerant metering applications to security applications with stringent requirements. This problem has been receiving significant interest from the research community in the context of the LTE-A random access channel. Most studies, however, consider load control schemes with few fixed service classes which can provide limited service differentiation. We propose an alternative scheme where the load is controlled by the price announced by the base station. The proposed method controls the load effectively and provides negligible delay for messages with the highest priority. It also enables low-cost wireless access to delay-tolerant messages by generating most of the revenue from high-priority messages. We derive pricing-based load control schemes for throughput and revenue maximization; and, present the relationship between delay, revenue and cost in both schemes. Our results suggest that dynamic pricing is a promising solution for major problems associated with cellular M2M traffic.**

I. Introduction

According to the OECD estimates, there will be 50 connected devices per household in 2020 [1]. This number, which does not include the devices outside of a household, indicates the expected predominance of the Internet of Things (IoT) devices. Most of these devices will not directly be controlled by a human being and will communicate autonomously, which is enabled through machine-to-machine (M2M) communications. This increase in M2M communications necessitates adapting the existing cellular infrastructures which have been optimized for human generated traffic.

Several requirements of M2M communications are not successfully addressed in current cellular standards: First of all, M2M communications may create a significant congestion when a massive number of nodes transmit semi-synchronously [2]–[7]. There is also a strong need for service differentiation since the requirements of IoT applications may vary greatly from delay-tolerant metering applications to security applications with stringent requirements [7]. Moreover, to enable ubiquitous wireless access for all these devices, providing affordable pricing schemes are essential [8].

There has been a growing interest on load control and service differentiation for LTE random access channel (RACH) which is the bottleneck for M2M cellular communications. Legacy techniques such as Access Class Barring and Extended Access Barring define several service classes according to the delay tolerance of users [5], [9]. At times of congestion, some access classes are barred based on an access probability announced by the base station. There are also numerous recent studies of LTE RACH focusing on load estimation [10]–[17], performance analysis [18]–[24] and service differentiation [25]–[27]. In line with the current standards, most of these studies consider load control schemes where the base station announces an access probability and they assume fixed service classes.

We propose an alternative approach for load control and service differentiation for random access based cellular M2M communications. In the proposed scheme, the base station announces a price which has to be paid for successful channel access. The load is controlled by adapting the price according to the channel congestion. The proposed scheme addresses major problems associated with LTE random access channel: First, it is possible to control the load in the network by incentivizing delay-tolerant nodes to transmit when the congestion relieved. Second, it is possible to provide a lower delay to high-priority messages by eliminating competition from low-priority price-sensitive devices. Third, it is possible to provide discounted network access for low-priority messages during off-peak times by obtaining most of the revenue from high-priority messages at times of congestion. Lastly, such a scheme eliminates the need for defining fixed service classes and it allows users to associate varying priority to different messages.

We develop the proposed pricing scheme for a multi-channel slotted ALOHA system which is extensively used in the analysis of LTE RACH [14], [15], [22], [23]. Our proposal is also applicable for next-generation connectionless M2M architectures for the cellular network [28]–[32] which effectively convert the spectrum into a multichannel ALOHA system [32].

We propose a pricing-based scheme for random access based M2M communications and it is one of the first such attempts along with our prior work [33]. Specific contributions of the present paper are as follows:

- We present an algebraic derivation of throughput-optimum and revenue-optimum pricing schemes for massive M2M arrivals.
- We analytically show the relationship between the valuation of a message and the delay it experiences for the throughput-optimum pricing scheme. The proposed scheme performs as a service differentiation mechanism by providing lower access delay for high-priority mes-

M. Koseoglu is with the Department of Computer Engineering, Hacettepe University, Ankara, 06800 Turkey. e-mail: mkoseoglu@cs.hacettepe.edu.tr.
This work is supported by the Scientific and Technological Research Council of Turkey (Tubitak) under project EEEAG-115E459 and by Hacettepe University Scientific Research Coordination Unit under project FDS-2015-7507.

sages. For messages willing to pay above a threshold, the proposed scheme provides negligible delay. For other messages, the delay linearly decreases with the price paid for the message.

- We show that revenue-optimum pricing results in increased revenues for the operator at the expense of higher delay. The difference between these two policies is more apparent at lower loads where both the delay and the revenue may be several times higher for the revenue-optimum policy.
- We show that most of the revenue is obtained from high-priority messages at times of congestion and low-priority messages can be provided low-cost access after the congestion relieves.
- We show that if a binary exponential backoff mechanism is employed by the messages, they pay lower prices at the expense of higher delay.

By presenting a mathematical analysis of optimum pricing, we complement our prior work which proposes a heuristic algorithm for price determination without any mathematical analysis [33]. Our results suggest that dynamic pricing is a promising solution for some major problems associated with M2M communications. We also argue that dynamic pricing is a more realizable solution for M2M communications rather than human-to-human communications because autonomous IoT devices can respond to dynamic pricing better than humans.

The rest of the paper is organized as follows: Next section presents the related work. Sec. III presents the studied system model. Throughput-optimum and revenue-optimum pricing schemes are presented in Secs. IV and V, respectively. Sec. VI investigates the case for non-uniform arrivals. Numerical results are presented in Sec. VII. A discussion on the feasibility of dynamic pricing is presented in Sec. VIII and the conclusions are presented in Sec. IX.

II. Related Work

The impact of M2M communications on the existing infrastructures has recently been an important focus of research. The most important problem associated with M2M traffic is the sporadic congestions created by a massive number of M2M devices activating semi-synchronously. For the LTE standard, the main bottleneck is the random access procedure through which a device request channel access [2]–[7]. 3GPP has been actively working on the analysis of M2M load on the cellular architectures and proposing improvements [34]–[37].

There are numerous recent studies on the design and analysis of load control systems for M2M communications [9]–[25]. Duan et al. proposed a heuristic method for controlling the load [10]. Estimation techniques based on the history of transmissions are also proposed [12]–[17]. Performance analysis of LTE RACH for real-time traffic [19], for energy consumption [18] has been proposed along with several other studies on modeling and analysis of the LTE RACH [20]–[23]. Performance limits of the load control are investigated in [24]. Lin et al. proposed a prioritized access scheme for RACH [25].

Since the channel access procedure of the LTE standard creates a significant overhead for short M2M messages, there are also recent proposals for connectionless random-access based transmissions [28]–[32]. In these schemes, collisions are unavoidable as there is no channel reservation. As a result, a connectionless M2M system turns into a multichannel slotted ALOHA system [32].

For both frameworks, the study of the multichannel ALOHA is crucial. Different from most of the previous studies on the load control of this channel, our proposal is to use a pricing-based load control and service differentiation method for multi-channel slotted ALOHA. To the best of our knowledge, this is one of the first such proposals along with our prior work [33]. Another recent work by Lin and Wei proposes a two-stage auction based pricing mechanism for the random access channel [27]. As opposed our proposal where the base station simply announces a price which would be paid for each successful message, this proposal requires M2M applications to bid for preamble transmission rates for specific periods which significantly increases the implementation complexity.

Our proposal has implications from an operator's perspective: Business models for M2M communications have not been settled [38]. Operators currently generate their revenue from humans which usually subscribe to monthly plans with data caps [39]. The operators generally employ tiered pricing where the users pay an extra fee to increase their data caps. Hence, the revenue collected from a user is roughly proportional to her data consumption.

M2M communications, however, require a new approach to pricing. M2M is characterized by small message transmissions in the uplink such as the measurement updates transmitted by a smart meter or the location-tracking messages of a vehicle tracking application. Traditional pricing methods may not be suitable for such applications as the amount of data consumed by these nodes is insignificant. While M2M communications have to be inexpensive to enable ubiquitous access, it should also be profitable for operators. We believe that dynamic pricing can suggest a new business model for operators to generate revenue from high-priority M2M devices while providing low-cost access for the rest of the devices.

Dynamic pricing has been applied to various domains with different success levels. Hospitality and airline industries have been applying dynamic pricing techniques to increase revenue at times of high demand [40]. Similar approaches have been proposed for sharing network resources since the early days of the Internet [41]. Despite many theoretical efforts towards the development of pricing schemes for networks [39], flat-rate pricing schemes continue to dominate due to psychological factors associated with price variations [42]. Here we also argue that M2M communications have the potential to benefit from dynamic pricing as the purchasing decisions (i.e. when to transmit data) can be made by autonomous IoT devices instead of their owners.

From a technical point-of-view, pricing of M2M communications necessitates a different framework from pricing studies focusing on bandwidth sharing among users. Early studies on pricing consider wireline network bandwidth sharing [43]–[46]. More recent studies consider pricing in cellular wireless networks especially for admission and power control [47]. The present study differs from these works significantly as we deal with a setting which is dominated by small message transmissions instead of users with bandwidth requirements.

III. System Model

We consider a scenario where the messages are transmitted through a multichannel slotted Aloha channel where each message has a duration of one slot. This is similar to a smart metering environment where there are a massive number of nodes with small message requirements [48]. The channels may correspond to different frequency bands in a connectionless architecture; or, they can be considered as different preambles in an LTE framework. A node transmits its message through a randomly selected channel from a set of M channels. If two or more messages are transmitted using the same channel in the same slot, all of these messages are failed. We later relax this assumption and present simulation results for the case of capture in Sec. VII-D.

In the LTE context, the channels in our system model can be considered as different preambles. In this case, we assume that the successful preamble transmission is followed by the successful transmission of a small payload and the successful reception of an acknowledgment message. Since the main bottleneck of the LTE architecture for M2M transmissions is the channel access procedure, we believe that pricing based only on the successful preamble transmissions is suitable. We only consider uplink transmissions, as M2M devices create traffic mostly on the uplink.

A pricing method for M2M communications must incur a low overhead. Complex pricing schemes such as auctions require message exchanges which would create a significant overhead. Hence, we consider a scenario where the price is determined by the base station and broadcasted to all nodes.

In the proposed system, the base station announces a price $p \in [0, 1]$ before each slot. The price will be paid only if a message is successfully transmitted; that is, if it does not experience a collision. Given a price p, a message is decided to be transmitted depending on its priority indicated by its valuation parameter $\theta \in [0, 1]$ which also corresponds to the willingness to pay [49]. The message is transmitted if its valuation is greater than or equal to the announced price, $\theta \geq p$, and its transmission is deferred otherwise. Messages with a high valuation correspond to emergency messages which require low delay and are willing to pay a higher price for immediate access. Messages with a low valuation can be considered as delay-tolerant messages which can be transmitted any time when the price drops sufficiently.

In the following analysis, we assume that the valuation parameter of messages, θ, is uniformly distributed between 0 and 1 as frequently used in the network economics literature [50], [51]. In our analysis, we first assume that a message is ready to be transmitted in the next slot if $\theta < p$ or the transmission has collided. We later investigate the case where the messages employ binary exponential back-off if $\theta < p$.

The proposed scheme can also be considered a class-based service differentiation scheme instead of a pricing mechanism: The valuation parameters can be considered as the predefined classes of nodes and the price announced by the base station can be thought as the minimum access class that is allowed to transmit. Although we will follow the jargon of pricing throughout the paper, it should be noted that the proposed scheme can also be implemented as an access class barring system.

IV. Throughput-Optimum Pricing

A. Single Time Slot

We first analyze throughput-optimum pricing for a single slot. Assume that there are B backlogged messages and their valuation parameters, θ, are uniformly distributed between 0 and 1. For a given price p, the expected number of transmitted messages is $B_t = B(1-p)$. If there are B_t messages randomly transmitted over M channels, the probability of success for a message can be written as:

$$P_s(B_t) = \left(1 - \frac{1}{M}\right)^{B_t - 1} \quad (1)$$

which is the probability of no other message selecting the same channel. Then, the channel throughput can be written as:

$$T(p) = B_t P_s(B_t) = B_t \left(1 - \frac{1}{M}\right)^{B_t - 1} \quad (2)$$

The maximum throughput is obtained when the number of transmitted messages equals to the number of channels; i.e. $B_t = M$ [10], [14], [24], [52]. So, the throughput-optimum pricing aims to have M transmitted messages at each slot:

$$p = \max\left(0, 1 - \frac{M}{B}\right). \quad (3)$$

For $B > M$, the expected number of successful transmissions under this throughput-optimum scheme can be approximated as M/e where e is the Euler's number [24].

This optimum price expression is derived for messages with uniformly distributed θ over a single slot. If this scheme is applied over successive slots, the distribution of θ among the backlogged messages will differ from the initial distribution due to the following: At each slot, messages with $\theta > p$ will be transmitted and some of them will become successful and exit the system. As a result, the proportion of messages with $\theta < p$ will increase among the backlogged messages. Hence, a selective accumulation of low-priority messages will be observed as time progresses; and, the distribution of θ will deviate from the initial distribution. We next study throughput-optimum pricing over successive slots.

B. Simultaneous Arrivals

Let us now consider the special case where all messages arrive at time $t = 0$. This analysis will be useful for analyzing the delay experienced by the backlogged messages in the later parts.

1) Pricing: Assume that there are N backlogged messages at time $t = 0$ with uniformly distributed θ between 0 and $\theta_{\max}$. How should the price change over time to eliminate this backlog as fast as possible? At the first slot, the price should be

$$p_s(0) = \theta_{\max} - \frac{M}{N/\theta_{\max}} \tag{4}$$

so that M messages are transmitted. As a result, the backlog is reduced by M/e messages on the average. In the second slot, the price should be decreased so that M/e new messages are transmitted along with the $M - M/e$ unsuccessful messages from the first slot. Then, for the i^{th} slot, the optimum price can be written as

$$p_s(i) = \theta_{\max} - \frac{M}{N/\theta_{\max}} - i\frac{M/e}{N/\theta_{\max}} \tag{5}$$

which indicates that the optimum price reduces linearly till the backlog is eliminated.

It should be noted that it is not certain that the unsuccessful messages from a slot will be cleared in the successive slot. It is possible that a message may be unsuccessful repeatedly over successive slots.

2) Delay: We now investigate the delay experienced by a message as a function of its θ. A message is first transmitted when the price reduces below its θ. It would be the j^{th} slot which is the solution of $\theta = p_s(j)$ as given by

$$j = \left(\theta_{\max} - \frac{M}{N/\theta_{\max}} - \theta\right)\left(\frac{N/\theta_{\max}}{M/e}\right). \tag{6}$$

After the price reduces below θ, it is not certain that the message will be successful on its first transmission. A message is needed to be transmitted $e-1$ times on the average since the success probability is $1/e$. Then, the expected delay is given by

$$d_s(\theta) = \left(\theta_{\max} - \frac{M}{N/\theta_{\max}} - \theta\right)\left(\frac{N/\theta_{\max}}{M/e}\right) + (e-1). \tag{7}$$

For the case of simultaneous arrivals, the delay experienced by a message linearly decreases with its valuation parameter.

C. Uniformly Distributed Arrivals

We now analyze the throughput-optimum pricing scheme for the uniformly distributed arrival of users over a fixed time interval.

1) Pricing: Assume that there are N messages arriving according to the uniform distribution over an interval of T slots and $n = N/T$ be the average number of arrivals at a slot. Assume that the valuations of the messages are uniformly distributed between 0 and 1.

If n is less than the maximum possible number of successful transmissions at a slot, i.e. $n < M/\mathrm{e}$, the arrivals do not overload the channel. In this case, the price should remain zero to allow all messages to be transmitted:

$$p_u(i) = 0, \quad \text{if } n \le M/e. \tag{8}$$

If the arrival rate is larger than the channel capacity, $n > M/\mathrm{e}$, the price should be greater than zero to prevent a congestion in the channel.

Since the arrivals are uniformly distributed, there is an average steady-state throughput-optimum price, p_u^*. Assume that this optimum price is announced before a slot and M messages are transmitted. In that slot, M/e of them will be successful on the average. In the next slot, n new messages will arrive. Then, the throughput-optimum price should ensure that only M/e out of these n messages will be transmitted so that the total number of transmissions is kept at M.

Since the θ of the new arrivals is uniformly distributed between 0 and 1, the solution of $n(1 - p_u^*) = M/e$ yields the throughput-optimum price:

$$p_u(i) = 1 - \frac{M/e}{n} \triangleq p_u^*, \quad \text{if } n > M/e \text{ and } i \le T. \tag{9}$$

In this analysis, we follow a mean-value analysis where the number of newly arriving users with valuation greater than p_u^* is assumed to be constant at $n(1 - p_u^*)$ for each slot. However, the number of users arriving at a specific slot is stochastic. Due to this randomness, price fluctuations are observed and our results indicate that the proposed analysis slightly underestimates the actual prices. In-depth analysis of the fluctuations in the price and the implications of our assumptions are presented in Sec. VII-A.

At the end of arrivals, at time T, there will be an accumulated backlog of messages which could not be transmitted due to non-zero prices. The number of backlogged messages will be $N_b = N - T(M/e)$ as M/e messages transmitted on the average at each of the T slots. The distribution of θ among the backlog will be uniformly distributed between 0 and p_u^*. Then, the analysis is similar to the simultaneous arrivals case given by (5) and the optimum price for the rest of the time can be written as:

$$p_u(i) = p_u^* - \frac{M}{N_b/p_u^*} - (i-T)\frac{(M/e)}{N_b/p_u^*}, \quad \text{if } n > M/e \text{ and } i > T. \tag{10}$$

In summary, the optimum price stays around an average steady-state level during the course of uniform-distributed arrivals and decreases linearly after the end of arrivals.

In this pricing scheme, the messages with $\theta < p_u^*$ pay a price of θ whereas messages with $\theta > p_u^*$ pay a price of p_u^*:

$$c_u(\theta) = \begin{cases} \theta, & \text{if } \theta < p_u^* \\ p_u^*, & \text{if } \theta \ge p_u^*. \end{cases} \tag{11}$$

This result states that the messages with the highest valuation pay a price lower than they are willing to depending on the congestion level.

2) Delay: We now express the relationship between the delay experienced and the valuation of the message. For messages with a valuation parameter greater than the steady-state price, the expected delay can be written as:

$$d_u(\theta) = e - 1, \quad \text{if } \theta \ge p_u^*. \tag{12}$$

since they are immediately transmitted with a success probability of $1/e$. Other messages, however, need to wait at least for the end of arrivals, T. Since the expected arrival time of messages is $T/2$, the expected waiting time till the end of arrivals is $T/2$.

After the end of arrivals, a message has to further wait for the price to reduce below its θ. The analysis is similar to the simultaneous arrivals case given by (7) and the average delay for the backlogged messages can be written as:

$$d_u(\theta) = T/2 + \left(p_u^* - \frac{M}{N_b/p_u^*} - \theta\right)\left(\frac{N_b/p_u^*}{M/e}\right) + (e-1), \quad \text{if } \theta < p_u^*. \tag{13}$$

For uniform arrivals, the delay experienced by a message with a valuation parameter greater than the steady-state price, the delay turns out to be a few slots. Other messages, however, need to wait till the end of congestion and also experience additional delay inversely proportional to their valuation parameter. So, the proposed policy acts as an effective service differentiation mechanism.

3) Revenue: In the proposed system, most of the revenue is collected from messages which experience negligible delay. $N(1-p_u^*)$ messages with $\theta > p_u^*$ pays a price of p_u^*. The other Np_u^* messages, which can depart only after the congestion ends, pay an average price of $p_u^*/2$. Then, the ratio of revenue collected from the messages with negligible delay can be written as:

$$\frac{N(1-p_u^*)p_u^*}{N(1-p_u^*)p_u^* + Np_u^*(p_u^*/2)} = \frac{1-p_u^*}{1-p_u^*/2} \tag{14}$$

As an example, for $M = 54$ and $n = 30$, $p_u = 0.34$ and the 66% of the messages with $\theta > p_u$ experience negligible delay and pay 80% of the total revenue. For a worse congestion, $n = 60$, $p_u = 0.67$, 33% of the messages pay 50% of the total revenue. Hence, the proposed mechanism allows charging messages according to their priority. Such a differentiation also enables low-cost access for messages willing to tolerate delay as the most of the revenue will be collected from messages with higher priority.

V. Revenue-Optimum Pricing

So far, we have studied a pricing scheme which maximizes the channel throughput and, as a result, minimizes the access delay. From a provider point of view, however, the main goal can be revenue maximization at the expense of increased delay. In that case, the provider should announce a price higher than the throughput-optimum price to increase the cost of channel access and to increase the probability of successful transmission of messages with a higher valuation.

Let P_s be the probability of success defined by (1) and θ^s is the sorted list of valuation parameters of backlogged messages in decreasing order. If the provider aims to maximize expected revenue per slot, it should announce a price $p^r = \theta^s(x)$ where x can be found as follows:

$$\arg\max_x \sum_{i=1}^{x} \theta^s(i) P_s(i) \tag{15}$$

maximizing the sum of the expected revenue collected from each message. If θ is distributed uniformly, the optimum p which maximizes the total revenue for a single slot can be derived as follows: For a price p, $B(1-p)$ messages will be transmitted with a success probability of $P_s\left(B(1-p)\right)$. Then, the total expected revenue can be written as:

$$\pi(p) = pB_tP_s = pB(1-p)\left(1-\frac{1}{M}\right)^{B(1-p)-1} \tag{16}$$

which is maximized at the following revenue-optimum price:

$$p^r = \frac{-\sqrt{B^2\log^2\left(1-\frac{1}{M}\right)+4} + B\log\left(1-\frac{1}{M}\right)+2}{2B\log\left(1-\frac{1}{M}\right)} \tag{17}$$

Approximating $\log\left(1-\frac{1}{M}\right) \approx -\frac{1}{M}$, it can be rewritten as:

$$p^r \approx \frac{\frac{B}{M} + \sqrt{\frac{B^2}{M^2}+4} - 2}{2\frac{B}{M}}. \tag{18}$$

In contrast to the throughput-optimum price which is zero for $B < M$, the revenue-optimum price approaches to $1/2$ as $B/M \to 0$. From a revenue perspective, it is not rational to announce a price of zero even if the channel is underutilized. For the evaluation of revenue-optimum pricing against throughput-optimum pricing under uniformly-distributed arrivals, we present simulation results in Sec. VII-B.

VI. Non-Uniform Arrivals

The uniform arrival distribution results in a steady-state average price during the course of arrivals but the price does not have a steady-state average value for non-uniform arrivals. In this section, we first analyze the case where the arrival rate monotonically increases, and, then, the case where the messages arrive according to a triangular distribution.

A. Monotonically increasing arrival rate

Assume that the arrival rate $n(i)$ is a monotonically increasing function of i. In such a case, the throughput-optimum price should be selected such that M/e messages departed in the previous slot should be replaced from the $n(i)$ newly arriving messages. Hence, the throughput-optimum price can be written as

$$p_m^*(i) = 1 - \frac{M/e}{n(i)} \tag{19}$$

as long as the arrival rate increases.

At the end of arrivals, there will be a backlog of users; and, different from the uniform arrivals case, the distribution of θ among this backlog will not be uniform due to increasing prices. The distribution will be skewed towards lower values; hence, the optimum price will not decrease linearly during the elimination of the backlog unlike the uniform arrivals case.

B. Triangularly distributed arrivals

In the LTE M2M communications literature, there are two commonly investigated arrival distributions: uniformly distributed arrivals and beta-distributed arrivals [35]. Triangular distribution provides an analytically tractable approximation to the Beta distribution [53]. We have also previously shown that the triangular distribution approximates the Beta distribution very well from the point of M2M communications [24]. Similarly, we here analyze triangularly distributed arrivals.

Let us consider the following triangular arrival distribution for N messages arriving over an interval of T slots:

$$n_t(i) = \begin{cases} \frac{2N}{T}\frac{i}{T/2} & 0 \le i/T \le 1/2, \\ \frac{2N}{T}\left(1 - \frac{i-T/2}{T/2}\right) & 1/2 \le i/T \le 1. \end{cases} \quad (20)$$

As the arrival rate is low at the beginning, the optimum price stays at zero as long as $n_t(i) < M/e$:

$$p_t(i) = 0, \quad \text{if } i < \tfrac{(M/e)(T/2)T}{2N}. \quad (21)$$

Using (19), the optimum price after this point till the peak of arrivals can be written as

$$p_t(i) = 1 - \frac{M/e}{n_t(i)}, \quad \text{if } \tfrac{(M/e)(T/2)T}{2N} < i/T < 1/2. \quad (22)$$

reaching a peak at

$$p_t(i = T/2) = 1 - \frac{M/e}{2N/T} \triangleq p_t^*. \quad (23)$$

After the arrival rate reaches its peak, the price starts to reduce. Each time the price is reduced, some of the backlogged messages will be transmitted along with some of the newly arriving messages. So, the reduction in the price will not be as steep as its increase.

Different from the uniformly distributed arrivals, the distribution of θ among the backlogged messages at the peak of arrivals will be skewed towards lower values. To simplify the analysis, at the peak of arrivals, we assume that there is a hypothetical backlog composed of both backlogged messages by that time and the future arrivals. Colored area in Fig. 1 illustrates this backlog at time $t = T/2$. Similar to the analysis in [24], the area marked as D in the figure represents the number of messages that have transmitted prior to time $t = T/2$ as M/e messages can be successfully transmitted at each slot. The area marked as A denotes the backlogged messages till time T/2 and areas marked by B and C denotes the future arrivals.

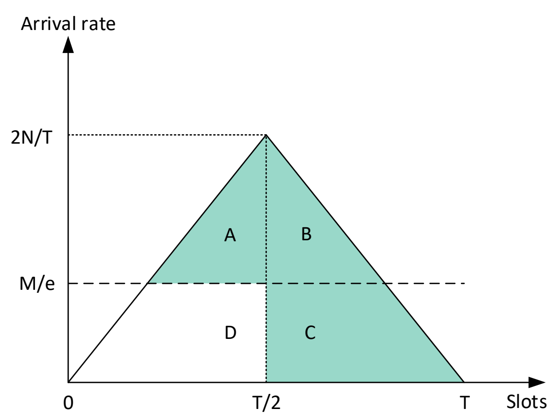

Fig. 1. The hypothetical backlog at time $t = T/2$ which both consists of the backlog before time $t = T/2$ (marked by area A) and the future arrivals (marked by B and C).

Then, the total number of messages in this hypothetical backlog can be computed as the area of colored regions in the Fig. 1 as follows:

$$N_b = \frac{N}{2} + \left(\frac{T}{2} - \frac{\frac{T}{2}(M/e)}{2\frac{N}{T}}\right)\left(2\frac{N}{T} - M/e\right)/2. \quad (24)$$

We assume θ among this backlog is uniformly-distributed between 0 and p_t^*. Then, similar to (5), the price during the second half of arrival duration can be written as:

$$p_t(i) = p_t^* - \frac{M}{N_b/p_t^*} - \left(i - \frac{T}{2}\right)\frac{M/e}{N_b/p_t^*}, \quad \text{if } i/T > 1/2. \quad (25)$$

This assumption results in a minor inaccuracy as shown in Sec. VII-C.

C. Delay

For messages with $\theta > p_t^*$, the expected number of retransmissions is negligible as given by

$$d_b(\theta) = e - 1, \quad \text{if } \theta > p_t^* \quad (26)$$

For messages with $\theta < p_t^*$, the delay strongly depends on the time of arrival. Before the arrival rate exceeds the capacity, $i < \frac{(M/e)(T/2)T}{2N}$, the delay experienced by the messages are $e-1$ regardless of their θ. For messages arriving after this point, the expected delay is still low if θ is higher than the price at the time of arrival. Otherwise, the message waits till the price drops below its θ.

VII. Simulation Results

We have evaluated the proposed pricing scheme and analysis using the simulation software that we developed. The following results are for throughput-optimum or revenue-optimum prices announced by a base station which has the perfect information about the valuation parameters of messages in the network. In a situation where the base station has limited information about the number of messages and their valuation parameters, heuristic price determination methods similar to the one that we have previously proposed [33] are needed. Here our aim is to define the optimum pricing schemes and to evaluate their delay, service differentiation and revenue collection performance.

In the simulations, a message is transmitted if its valuation is greater than or equal to the price announced. When a collision occurs at a slot, none of the collided messages is successful; i.e. we ignore the capture effect. We later relax this assumpton in Sec. VII-D. Before each slot, the base station announces a new price using the information about the backlogged messages. If the throughput-optimum scheme is employed, the base station announces a price such that only M messages are transmitted at a slot. If the revenue-optimum scheme is employed, the base station calculates the expected revenue for different possible prices as explained in Sec. V and selects the price which maximizes it. We first performed simulations for persistent messages which stay backlogged if they discover that the price is higher than their valuation. Then, we consider the effect of the binary-exponential backoff in Sec. VII-E.

To compare the delay performance of the proposed policies, we have used the optimum probabilistic access barring scheme as the benchmark. In this scheme, the base station broadcasts an access probability of $q = \min(1, M/N)$ before each slot to maximize the channel throughput. This scheme has also been used in several studies as a benchmark [10], [15], [24].

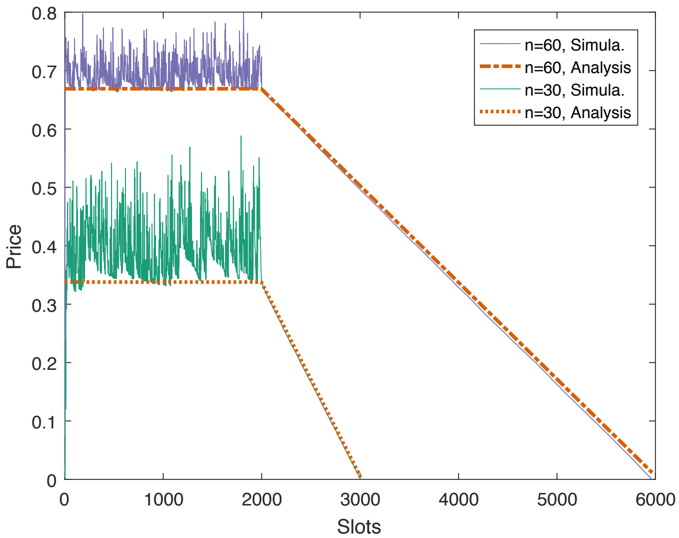

Fig. 2. Throughput-optimum price over the course of uniformly-distributed arrivals over 2000 slots.

A. Throughput-Optimum Pricing for Uniformly Distributed Arrivals

We first evaluate the analysis for uniformly distributed arrivals presented in Sec. IV-C. Fig. 2 denotes the change in the price during two sample simulations for the arrival of $N = 60\,000$ ($n = 30$) and $N = 120\,000$ ($n = 60$) messages over a duration of $T = 2000$ slots where there are $M = 54$ channels. These parameters have been inspired by the typical LTE RACH case studies where a massive number of arrivals occur over a duration of 10 s where the random access opportunities occur every 5 ms [35].

The analysis given by (9) and (10) matches the simulation results well. Although there are price fluctuations in the simulations, the average price shortly reaches a steady state value closely predicted by our analysis and stays around the same level till the end of arrivals after which it decreases linearly. The steady-state price predicted by the proposed analysis is 13% lower than the average price during the course of arrivals for $n = 30$ and 4% lower for $n = 60$. After the end of arrivals, the price difference between the analysis and simulations is lower and bounded by 0.003 for both n.

Fig. 3 shows the histogram of the throughput-optimum price at each slot during the first 2000 slots of the simulation for $n = 30$ along the with the predicted steady-state price. The difference between the analysis and the simulation is caused by the stochastic behavior of arrivals which has been omitted in our analysis. The price during the simulations tends to be greater than the price predicted by the analysis due to the following: At a given slot, there is a massive backlog of

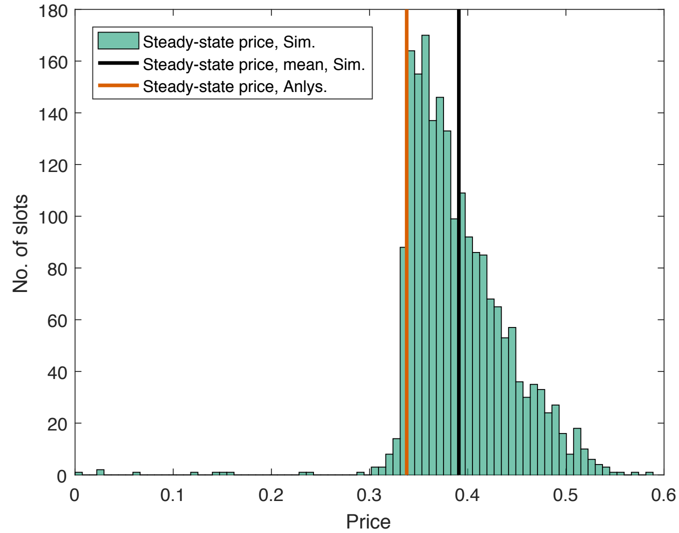

Fig. 3. Histogram of the throughput-optimum price during the arrivals (the first 2000 slots) along with the predicted steady-state price.

messages with valuations below the steady-state price which have never found a chance to be transmitted due to high prices. For example, Fig. 4 shows the histogram of backlogged messages at the 1000th slot for $n = 30$.

This backlog prevents the price to drop significantly. Suppose that there are few new arrivals in a slot: The price will drop slightly since the valuations of messages are densely distributed just below the steady state price. On the other hand, in the opposite case where there are too many new arrivals at a slot, a spike in the price will be observed. Hence, a significant increase in the price is more likely than a significant decrease in the price. For that reason, proposed analysis tends to underestimate the price slightly.

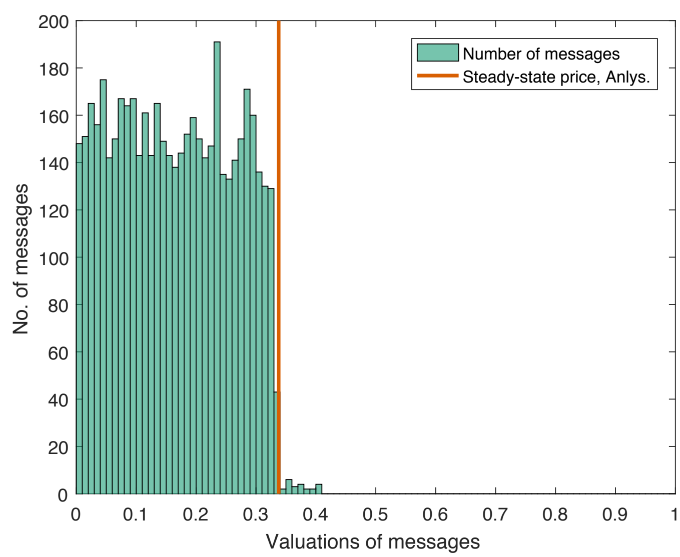

Fig. 4. Histogram of the valuations of backlogged messages at the 1000th slot.

Fig. 5 plots the change in the delay experienced by the messages as a function of their valuation parameter under

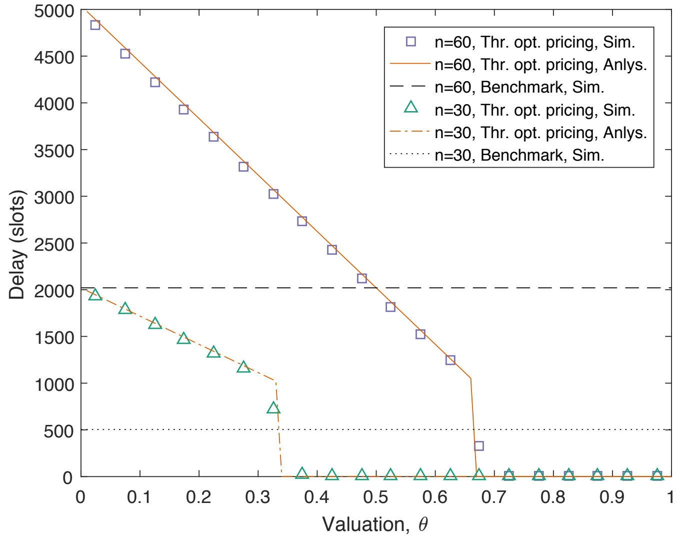

Fig. 5. Delay as a function of valuation parameter of messages for uniformly-distributed arrivals under the throughput-optimum pricing and the benchmark algorithm.

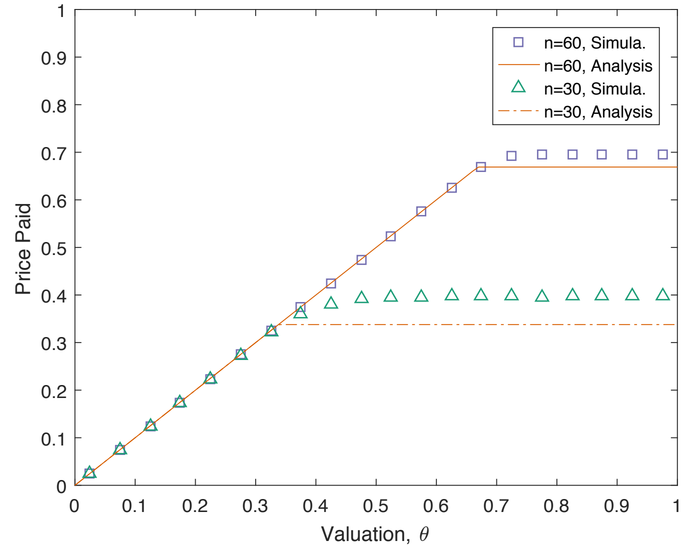

Fig. 6. Price paid as a function of valuation parameter of messages for uniformly-distributed arrivals under throughput-optimum pricing.

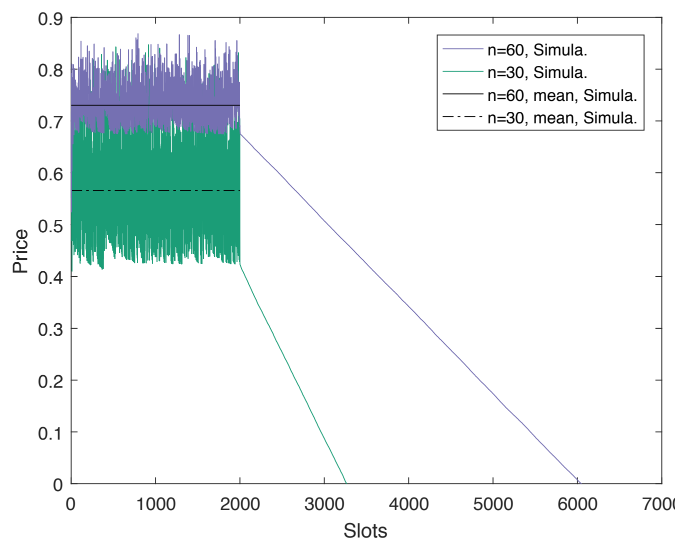

Fig. 7. Revenue-optimum price for uniformly-distributed arrivals over 2000 slots.

the throughput-optimum policy along with the analysis given by (12) and (13). Points in the plot denote the average delay experienced by messages with $k \leq \theta < k + 0.05$ where $k = 0, 0.05, ..., 0.95$. Results indicate that the delay is negligible for messages with θ greater than the steady state price. For other messages, the average delay starts at $T/2 = 1000$ slots and increases linearly as the valuation parameter reduces. These results suggest that the proposed mechanism successfully differentiates among the messages according to their priority.

Fig. 5 also plots the delay achieved by the benchmark throughput-optimum probabilistic barring algorithm. Since the benchmark algorithm does not differentiate according to the valuations of messages, the delay experienced by a message is independent of its valuation parameter. The average delay achieved by the throughput-optimum pricing algorithm is close to the delay achieved by the benchmark algorithm. Average delay comparison among these two algorithms is presented in more detail in Fig. 9.

Fig. 6 shows the price paid by messages according to their valuation parameter along with the analysis given by (11). For messages with $\theta < p_u^*$, the cost equals to their valuation parameter. Other messages pay a maximum price of p_u^* which is lower below their valuation. From the provider's perspective, a more severe congestion is better for generating higher revenues by charging messages with prices closer to their valuation parameters.

B. Revenue-Optimum vs. Throughput-Optimum Pricing

We now compare the throughput-optimum pricing scheme against the revenue-optimum pricing scheme defined in Sec. V. Fig. 7 shows the change in the revenue-optimum price for the same parameters with Fig. 2. The revenue-optimum prices are higher than the throughput-optimum prices.

Fig. 8 shows the change in the average throughput-optimum and revenue-optimum prices during the course of arrivals (first 2000 slots) for different load levels along with the analysis given by (9) and (10). The proposed expressions for throughput-optimum prices slightly underestimates the simulation results by an average price of 0.04. The main difference between throughput-optimum and revenue-optimum pricing is observed at lower load levels; and, they approach each other as the network load increases. As derived in Sec. V, the revenue-optimum price does not fall below 0.5 for a single slot when the θ is uniformly distributed. In accordance with this result, the simulations also indicate that the average revenue-optimum price does not fall below 0.47 even at very low loads. From a provider point-of-view, prices should not be too low in order to optimize profits despite it may reduce channel utilization.

Fig. 9 shows the change in delay for throughput-optimum and revenue-optimum pricing along with the benchmark al-

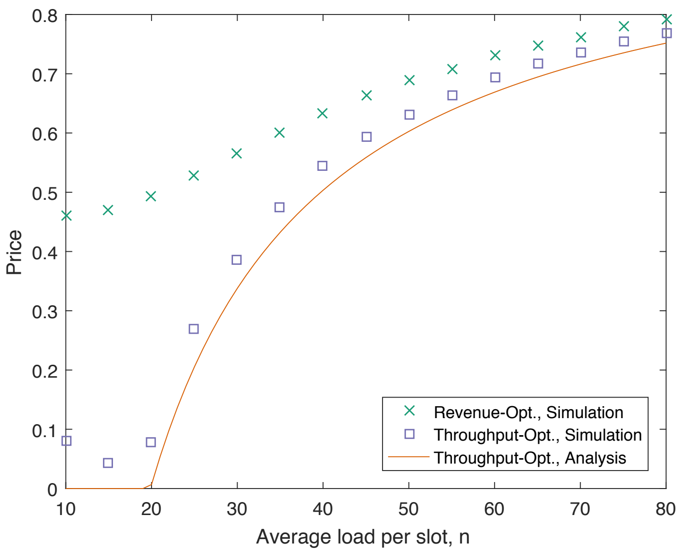

Fig. 8. Average price over the course of uniformly-distributed arrivals as a function of the channel load.

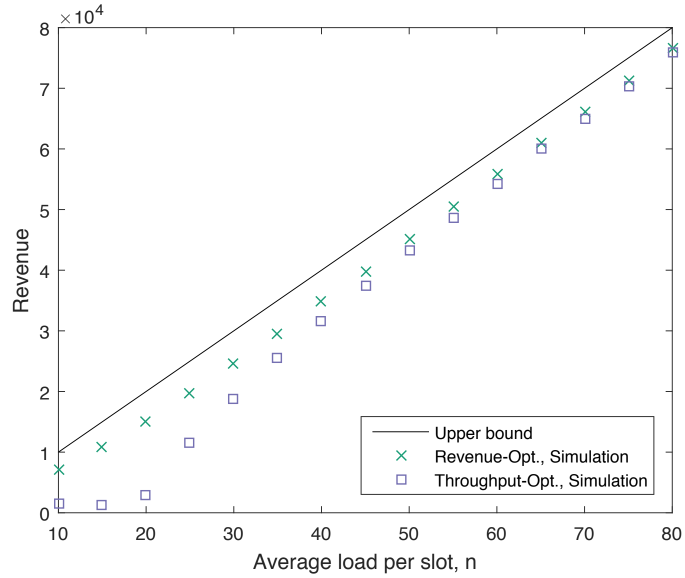

Fig. 10. The total revenue for throughput-optimum and revenue-optimum pricing as a function of the channel load.

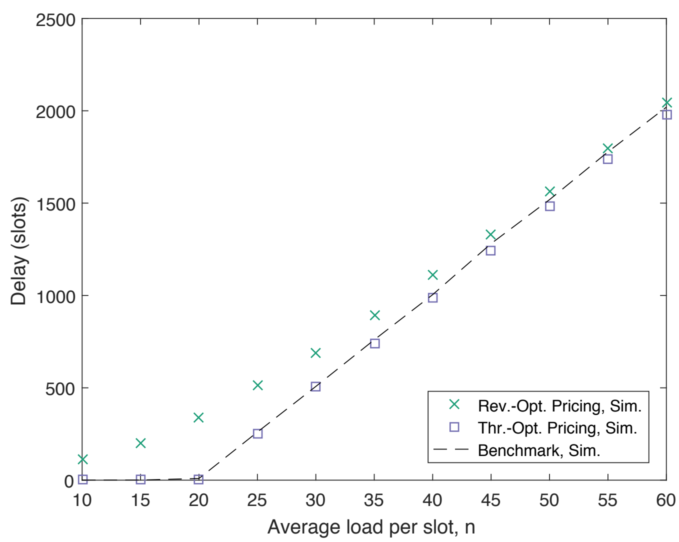

Fig. 9. Average delay for throughput-optimum and revenue-optimum pricing along with the benchmark algorithm as a function of the channel load.

gorithm as a function of the network load. In comparison to the throughput-optimum policy, revenue-optimum pricing results in increased delay. The increase in delay is more apparent at lower loads due to the significant price difference between the policies. For $n = 20$, for example, the average delay is approximately 3 slots for the throughput-optimum policy whereas it is 346 slots for the revenue-optimum policy. The difference in delay reduces as the load increases since both policies result in similar prices at higher loads. Delay obtained by the throughput-optimum policy is very close to the benchmark algorithm.

Fig. 10 shows the change in the total revenue collected for both pricing policies along with the upper bound on the revenue. The upper bound is the sum of valuation parameters of messages which indicates the maximum amount that a message is willing to pay. At the expense of increased delay, the revenue-optimum policy collects more revenue. For lower loads, the difference between the upper bound and the revenue collected is more significant: For $n = 10$, the revenue generated by the revenue-optimum policy is 29% lower than the upper bound and the revenue generated by the throughput-optimum is 85% lower than the upper bound. At higher loads, the revenue approaches to the upper bound: For $n = 80$, revenue collected by the revenue-optimum and throughput-optimum policies is 4% and 5% lower than the upper bound, respectively. Since there is less competition from the messages at lower loads, most messages pay lower than their valuation parameter. At higher loads, however, the price paid for a message is closer to its valuation.

C. Non-uniform Arrivals

We next evaluate the analysis for triangularly distributed arrivals presented in Sec. VI-B. Fig. 11 denotes the change in the price for $N = 60\,000$ ($n = 30$) and $120\,000$ ($n = 60$) for $M = 54$. The analysis given by (19) predicts the prices as the arrival rate increases till time $t = 5\,\mathrm{s}$ very well. Price change during the decreasing arrival rate is predicted by (25) with less accuracy due to the assumptions made in the analysis. Since we assume a hypothetical backlog with uniformly distributed valuation to analyze the price behavior after time $t = T/2$, there is a difference between the simulation results and the analysis.

D. Capture Effect

So far, we have assumed that if two or messages are transmitted over the same channel at the same slot, none of them can be correctly received. In practice, however, it is possible to recover a message in the case of a collision if its received SNR is sufficiently high. The capture effect increases the throughput of the channel, hence affects the

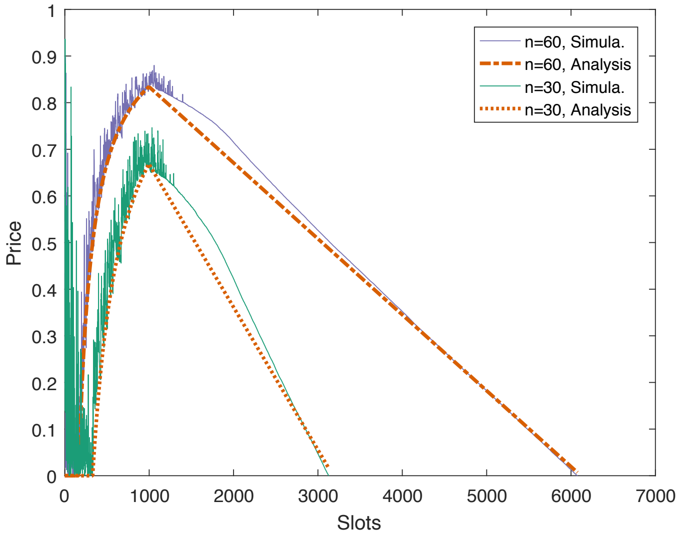

Fig. 11. Throughput-optimum price over the course of triangularly-distributed arrivals over 2000 slots.

performance of the proposed pricing policies. To take this effect into consideration, here we define a simple model of capture where one of the colliding messages in a slot can be recovered with a probability of p_c. The case with no capture corresponds to $p_c = 0$ and the perfect capture case is $p_c = 1$.

Fig. 12 shows how the price and delay change as the probability of capture increases for $n = 30$ under the throughput-optimum policy. Since the capture effect increases the channel capacity, the delay reduces as the capture probability increases. As a result of the increased throughput, there is a smaller backlog of users which results in lower prices. This result implies that the increased throughput may not result in greater revenue for the provider in the single-provider scenario.

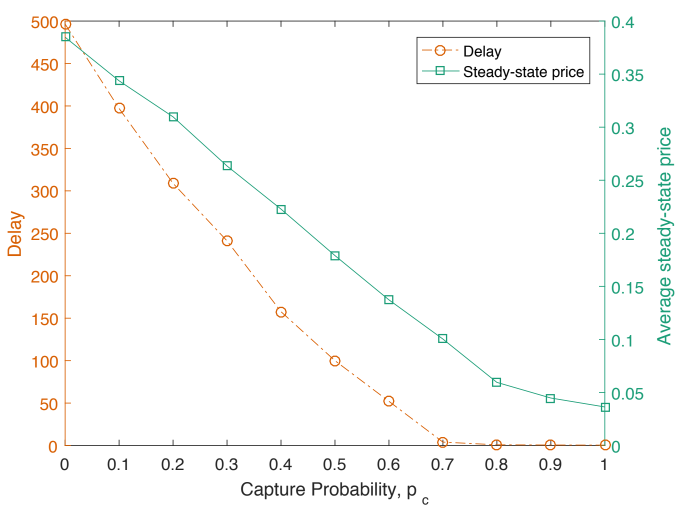

Fig. 12. Delay and the average steady-state price as a function of the capture probability for $n = 30$.

E. Binary exponential back-off

So far, we assumed that the nodes continue to listen to the price announced by the base station and transmit their messages as soon as the price falls below their valuation. In practice, however, the nodes may decide to backoff when they notice that the price is higher than the valuation of the message. Backing off may help nodes to preserve energy by eliminating idle listening and transmitting at a lower cost when the price is lower. We here investigate the effect of binary exponential backoff (BEB) on the delay experienced by the messages and the revenue collected.

In the BEB algorithm, a message waits for a random number of slots selected between 0 and $2^f - 1$ where f is the number of times that the node discovers that the price is higher than the valuation of the message. We assume a message is retransmitted in the next slot if it failed due to a collision. Since the success probability is $1/e$ for the throughput-optimum scheme, a message will be transmitted successfully in a short time as long as the price is lower than its valuation.

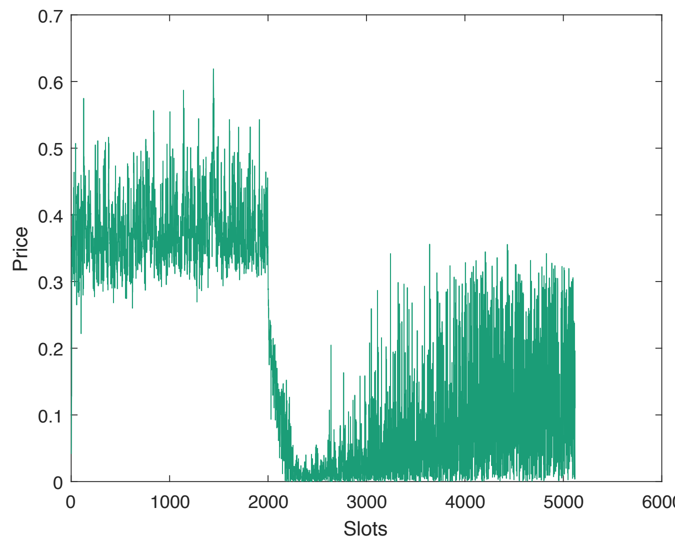

Fig. 13. Throughput-optimum price over the course of uniformly-distributed arrival of $N = 60000$ ($n = 30$) messages when there is a binary exponential backoff algorithm.

Fig. 13 shows the change in throughput-optimum price when the messages employ BEB. When compared with Fig. 2, it can be seen that the price does not decrease linearly after the end of arrivals. In this case, most of the messages backed off during the congestion and they became active after their backoff duration ends. BEB results in increased delay: Backlog is eliminated at the 5120th slot when the BEB algorithm is employed; but, it was eliminated at the 3003rd slot when the messages persist.

Figs. 14 and 15 shows the change in the delay and the price paid as function of the valuation parameter of the messages, respectively. For $n = 30$, the average delay is 25% higher when BEB is employed whereas the average price paid is 13% lower. For $n = 60$, the delay is 15% higher and the price paid is 23% lower when BEB is used. At the expense of increased delay, BEB allowed price-sensitive low-priority messages a lower-cost channel access. For high-priority messages having a valuation greater than the maximum price, both the delay and the cost of channel access stays the same.

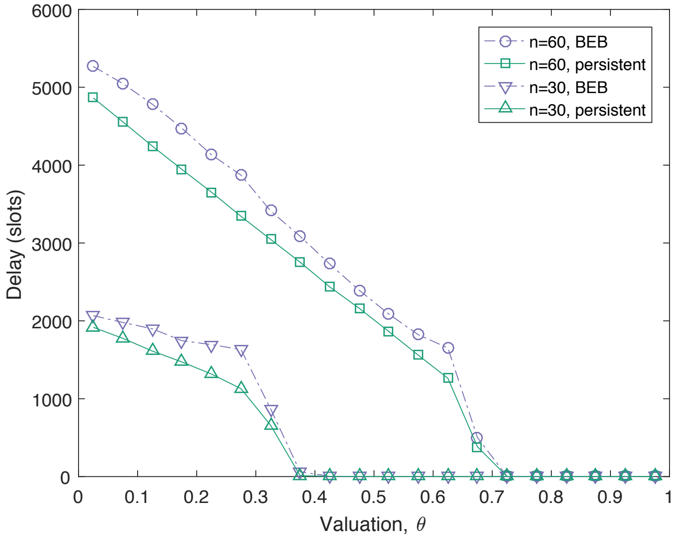

Fig. 14. Delay as a function of the valuation parameter of messages both when the BEB algorithm is employed and the messages persist.

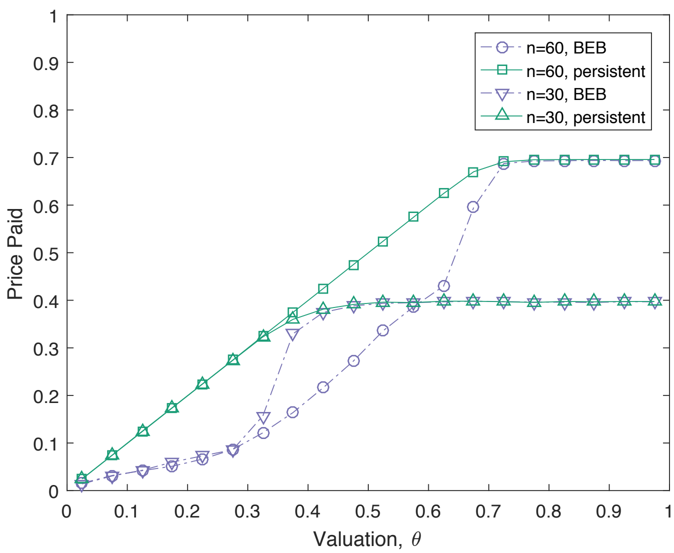

Fig. 15. Delay as a function of the valuation parameter of messages both when the BEB algorithm is employed and the nodes persist.

VIII. Feasibility of dynamic pricing for M2M communications

Here we discuss several factors both in favor and against dynamic pricing for M2M communications.

Elimination of behavioral factors: Complex pricing schemes are known to be disliked by human users [42] and dynamic pricing has been met by aversion in several different contexts [54]. In M2M communications, however, purchasing decisions will be made by smart devices which would eliminate the emotional aspects of pricing. Google Adwords is a good example of how complex pricing schemes be successful when the decisions are made by autonomous agents [55]. Human users configure several parameters in AdWords and individual decisions are made on behalf of the user. Similarly, in M2M communications, fundamental parameters can be configured by the users and the rest of the decisions can be left to the autonomous agents. Still, the idea of purchasing decisions made by their devices may be disengaging for users.

High-frequency dynamic pricing: It is possible to implement high-frequency dynamic pricing algorithms when the decisions are made by autonomous agents. M2M communications may suffer from short-term severe congestion due to some external event such as power recovery. Such congestions require a quick reaction by the base station as it may result in severe access delays both for M2M and human-to-human communications. Since M2M nodes can react quickly, it is possible to manage congestions by high-frequency pricing updates and shifting the traffic of delay tolerant nodes at times of congestion. The downside here is that it requires frequent processing by IoT devices which may be constrained in terms of processing power and energy consumption.

The wide range of QoS requirements: The applications in the IoT ecosystem will have a wide range of QoS requirements ranging from alarm sensors to statistical monitoring applications. Moreover, it is possible that the service requirements of various messages belonging to the same device can differ. A temperature monitor, for example, would require immediate service if there is a significant temperature increase, otherwise it can delay the transmission of other messages to off-peak times. Hence, preconfigured QoS plans for IoT devices may not be adequate for next generation networking. Dynamic pricing may be a viable alternative for static access class schemes.

Regulations: Pricing is heavily regulated by governments and such regulations may prevent the adoption of dynamic pricing. Although deregulation has been occurring in several markets such as in electricity [56], it may delay the adoption of dynamic pricing.

Theoretical developments: Since the M2M requirements differ significantly from human-based communications, pricing techniques has to be revisited. The present study is one of the first steps towards such an aim; but, there are many open issues such as pricing schemes for strategic users/providers or multi-provider environments.

IX. Conclusions

We have proposed a pricing scheme for multichannel slotted Aloha systems which captures the behavior of LTE random-access channel. We have proposed a mathematical derivation of throughput-optimum and revenue-optimum pricing schemes. In comparison to previous approaches which are based on probabilistic barring of nodes based on predefined access classes, the proposed scheme uses price incentives to control the load. As well as controlling the load effectively, the proposed scheme provides service differentiation among the messages by providing a negligible delay for highest-priority messages. Since these high-priority messages pay most of the revenue by paying higher prices, it is possible to provide low-cost communications for the rest of the devices. As low-cost wireless access is crucial to enable an IoT ecosystem, we believe dynamic pricing can be a feasible tool for enabling ubiquitous wireless access for all devices. We also believe that M2M communications will not suffer from psychological downsides of dynamic pricing as the decisions will be made by smart devices on behalf of their users.

References

[1] OECD, "Machine-to-machine communications: Connecting billions of devices," *OECD Digital Economy Papers*, no. 192, 2012.

[2] S.-Y. Lien, K.-C. Chen, and Y. Lin, "Toward ubiquitous massive accesses in 3GPP machine-to-machine communications," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 66–74, 2011.

[3] A. Ksentini, Y. Hadjadj-Aoul, and T. Taleb, "Cellular-based machine-to-machine: overload control," *IEEE Network*, vol. 26, no. 6, pp. 54–60, 2012.

[4] M.-Y. Cheng, G.-Y. Lin, H.-Y. Wei, and A.-C. Hsu, "Overload control for machine-type-communications in LTE-Advanced system," *IEEE Communications Magazine*, vol. 50, no. 6, pp. 38–45, 2012.

[5] M. Hasan, E. Hossain, and D. Niyato, "Random access for machine-to-machine communication in LTE-advanced networks: issues and approaches," *IEEE Communications Magazine*, vol. 51, no. 6, pp. 86–93, 2013.

[6] A. Laya, L. Alonso, and J. Alonso-Zarate, "Is the random access channel of LTE and LTE-A suitable for M2M communications? A survey of alternatives," *IEEE Comm. Surveys Tutorials*, vol. 16, no. 1, pp. 4–16, First 2014.

[7] K. Zheng, S. Ou, J. Alonso-Zarate, M. Dohler, F. Liu, and H. Zhu, "Challenges of massive access in highly dense LTE-advanced networks with machine-to-machine communications," *IEEE Wireless Communications*, vol. 21, no. 3, pp. 12–18, June 2014.

[8] J.-M. Bohli, C. Sorge, and D. Westhoff, "Initial observations on economics, pricing, and penetration of the internet of things market," *SIGCOMM Comput. Commun. Rev.*, vol. 39, no. 2, pp. 50–55, Mar. 2009.

[9] U. Phuyal, A. Koc, M.-H. Fong, and R. Vannithamby, "Controlling access overload and signaling congestion in M2M networks," in *Asilomar Conference on Signals, Systems and Computers*, 2012, pp. 591–595.

[10] S. Duan, V. Shah-Mansouri, and V. Wong, "Dynamic access class barring for M2M communications in LTE networks," in *IEEE GLOBECOM*, 2013, pp. 4747–4752.

[11] G. C. Madueno, N. K. Pratas, C. Stefanovic, and P. Popovski, "Massive M2M access with reliability guarantees in LTE systems," in *IEEE ICC*, 2015.

[12] H. Wu, C. Zhu, R. La, X. Liu, and Y. Zhang, "FASA: Accelerated S-ALOHA using access history for event-driven M2M communications," *IEEE/ACM Transactions on Networking*, vol. 21, no. 6, pp. 1904–1917, 2013.

[13] M. Tavana, V. Shah-Mansouri, and V. W. Wong, "Congestion control for bursty M2M traffic in LTE networks," in *IEEE ICC*, 2015.

[14] O. Galinina, A. Turlikov, S. Andreev, and Y. Koucheryavy, "Stabilizing multi-channel slotted aloha for machine-type communications," in *IEEE ISIT*, 2013, pp. 2119–2123.

[15] O. Arouk and A. Ksentini, "Multi-channel slotted aloha optimization for machine-type-communication," in *ACM MSWIM*, 2014, pp. 119–125.

[16] C.-Y. Oh, D. Hwang, and T.-J. Lee, "Joint access control and resource allocation for concurrent and massive access of M2M devices," *IEEE Transactions on Wireless Communications*, vol. 14, no. 8, pp. 4182–4192, Aug 2015.

[17] G. Y. Lin, S. R. Chang, and H. Y. Wei, "Estimation and adaptation for bursty LTE random access," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 4, pp. 2560–2577, April 2016.

[18] M. Gerasimenko, V. Petrov, O. Galinina, S. Andreev, and Y. Koucheryavy, "Impact of machine-type communications on energy and delay performance of random access channel in LTE-advanced," *Trans. on Emerging Telecomm. Tech.*, vol. 24, no. 4, pp. 366–377, 2013.

[19] H. Thomsen, N. Pratas, C. Stefanovic, and P. Popovski, "Analysis of the LTE access reservation protocol for real-time traffic," *IEEE Communications Letters*, vol. 17, no. 8, pp. 1616–1619, 2013.

[20] R.-G. Cheng, J. Chen, D.-W. Chen, and C.-H. Wei, "Modeling and analysis of an extended access barring algorithm for machine-type communications in LTE-A networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 2956–2968, 2015.

[21] J. J. Nielsen, D. M. Kim, G. C. Madueno, N. K. Pratas, and P. Popovski, "A tractable model of the lte access reservation procedure for machine-type communications," in *2015 IEEE Global Communications Conference (GLOBECOM)*, Dec 2015, pp. 1–6.

[22] C.-H. Wei, G. Bianchi, and R.-G. Cheng, "Modeling and analysis of random access channels with bursty arrivals in OFDMA wireless networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 4, pp. 1940–1953, April 2015.

[23] C.-H. Wei, R.-G. Cheng, and Y.-S. Lin, "Analysis of slotted-access-based channel access control protocol for LTE-advanced networks," *Wireless Personal Communications*, vol. 85, no. 1, pp. 9–27, 2015.

[24] M. Koseoglu, "Lower bounds on the LTE-A average random access delay under massive M2M arrivals," *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 2104–2115, May 2016.

[25] T. M. Lin, C. H. Lee, J. P. Cheng, and W. T. Chen, "PRADA: Prioritized random access with dynamic access barring for MTC in 3GPP LTE-A networks," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 5, pp. 2467–2472, Jun 2014.

[26] Y. C. Pang, S. L. Chao, G. Y. Lin, and H. Y. Wei, "Network access for M2M/H2H hybrid systems: a game theoretic approach," *IEEE Communications Letters*, vol. 18, no. 5, pp. 845–848, 2014.

[27] G. Y. Lin and H. Y. Wei, "Auction-based random access load control for time-dependent machine-to-machine communications," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 658–672, Oct 2016.

[28] S. Andreev, A. Larmo, M. Gerasimenko, V. Petrov, O. Galinina, T. Tirronen, J. Torsner, and Y. Koucheryavy, "Efficient small data access for machine-type communications in LTE," in *IEEE International Conference on Communications (ICC)*, June 2013, pp. 3569–3574.

[29] A. Ahmadian, O. Galinina, S. Andreev, and Y. Koucheryavy, "Modeling contention-based M2M transmissions over 3GPP LTE cellular networks," in *IEEE ICC Workshops*, June 2014, pp. 441–447.

[30] H. Dhillon, H. Huang, H. Viswanathan, and R. Valenzuela, "Fundamentals of throughput maximization with random arrivals for M2M communications," *IEEE Transactions on Communications*, vol. 62, no. 11, pp. 4094–4109, 2014.

[31] C. Kahn and H. Viswanathan, "Connectionless access for mobile cellular networks," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 26–31, September 2015.

[32] R. P. Jover and I. Murynets, "Connection-less communication of IoT devices over LTE mobile networks," in *12th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, June 2015, pp. 247–255.

[33] M. Koseoglu, "Smart pricing for service differentiation and load control of the LTE-A IoT system," in *IEEE 2nd World Forum on Internet of Things (WF-IoT)*, Dec 2015, pp. 187–192.

[34] 3GPP, "Rach intensity of time controlled devices," Tech. Rep., 2010.

[35] ——, "Study on RAN improvements for machine-type communications (release 11)," Tech. Rep., 2011.

[36] ——, "Study on machine-type communications (MTC) and other mobile data applications communications enhancements (TR 23.887)," Tech. Rep., 2013.

[37] ——, "Service requirements for machine-type communications (MTC)," Tech. Rep., 2014.

[38] H. Viswanathan, *M2M Communications: A Systems Approach*. John Wiley & Sons, Ltd, 2012, ch. The Business of M2M, pp. 23–36.

[39] S. Sen, C. Joe-Wong, S. Ha, and M. Chiang, "A survey of smart data pricing: Past proposals, current plans, and future trends," *ACM Comput. Surv.*, vol. 46, no. 2, pp. 15:1–15:37, Nov. 2013.

[40] McAfeea, R. Preston, and V. te Velde, *Economics and information systems*. Elsevier, 2007, ch. Dynamic pricing in the airline industry, pp. 527–570.

[41] R. J. Edell, N. McKeown, and P. P. Varaiya, "Billing users and pricing for TCP," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1162–1175, Sep 1995.

[42] A. Odlyzko, "Should flat-rate internet pricing continue," *IT Professional*, vol. 2, no. 5, pp. 48–51, Sep 2000.

[43] J. K. MacKie-Mason and H. R. Varian, "Pricing congestible network resources," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1141–1149, Sep 1995.

[44] F. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, no. 1, pp. 33–37, 1997.

[45] R. Gibbens and F. Kelly, "Resource pricing and the evolution of congestion control," *Automatica*, vol. 35, no. 12, pp. 1969 – 1985, 1999.

[46] L. A. DaSilva, "Pricing for QoS-enabled networks: A survey," *IEEE Communications Surveys Tutorials*, vol. 3, no. 2, pp. 2–8, Second 2000.

[47] B. Al-Manthari, N. Nasser, and H. Hassanein, "Congestion pricing in wireless cellular networks," *IEEE Communications Surveys Tutorials*, vol. 13, no. 3, pp. 358–371, Third 2011.

[48] G. C. Madueo, 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.

[49] P. Nuggehalli, J. Price, and T. Javidi, "Pricing and QoS in wireless random access networks," in *IEEE GLOBECOM*, Nov 2008, pp. 1–5.

[50] J. Walrand, *Performance Modeling and Engineering*. Boston, MA: Springer US, 2008, ch. Economic Models of Communication Networks, pp. 57–89.

[51] C. K. Chau, Q. Wang, and D. M. Chiu, "On the viability of paris metro pricing for communication and service networks," in *IEEE INFOCOM Proceedings*, March 2010, pp. 1–9.

[52] D. Shen and V. Li, "Performance analysis for a stabilized multi-channel slotted ALOHA algorithm," in *IEEE PIMRC*, 2003.

[53] D. Johnson, "The triangular distribution as a proxy for the beta distribution in risk analysis," *Journal of the Royal Statistical Society. Series D (The Statistician)*, vol. 46, no. 3, pp. 387–398, 1997.

[54] K. L. Haws and W. O. Bearden, "Dynamic pricing and consumer fairness perceptions," *Journal of Consumer Research*, vol. 33, no. 3, pp. 304–311, 2006.

[55] B. Edelman, M. Ostrovsky, and M. Schwarz, "Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords," National Bureau of Economic Research, Working Paper 11765, November 2005.

[56] J. M. Griffin and S. L. Puller, Eds., *Electricity Deregulation*, 1st ed. University of Chicago Press, 2005.