On the Role of Early-Termination for Age of Information in Tree-Based Random Access Protocols

Andrea Munari[†], Čedomir Stefanović[‡]

[†] Institute of Communications and Navigation, German Aerospace Center (DLR), Weßling, Germany
[‡] Department of Electronic Systems, Aalborg University, Denmark.
email: andrea.munari@dlr.de, cs@es.aau.dk

Abstract—Age of Information (AoI) has emerged as a key metric for assessing data freshness in IoT applications, where a large number of devices report time-stamped updates to a monitor. Such systems often rely on random access protocols based on variations of ALOHA at the link layer, where collision resolution algorithms play a fundamental role to enable reliable delivery of packets. In this context, we provide the first analytical characterization of average AoI for the classical Capetanakis tree-based algorithm with gated access [1] under exogenous traffic, capturing the protocol's dynamics, driven by sporadic packet generation and variable collision resolution times. We also explore a variant with early termination, where contention is truncated after a maximum number of slots even if not all users are resolved. The approach explores a fundamental trade-off between reliability and timeliness, allowing stale packets to be dropped to improve freshness.

I. INTRODUCTION

Age of information (AoI) has recently emerged as an insightful performance indicator for many communications networks [2]. Defined as the time elapsed since the generation of the last received message, the metric captures in a simple way how fresh the knowledge available at the receiver is on the status of a source. AoI is especially relevant in Internet of things (IoT) systems, where devices sample a physical process of interest and send time-stamped updates to a central node for monitoring/actuation, and where timeliness is paramount.

IoT networks are often characterized by a large number of reporting terminals, which generate messages in a sporadic and unpredictable fashion. To accommodate such traffic, many commercial solutions, e.g., LoRa, Sigfox or, partly, NB-IoT, implement random access protocols based on variations of ALOHA [3], [4] at the link layer. In parallel, research has extensively focused on devising solutions that ease the well-known issue of collisions in grant-free channels. On the one hand, a vast wave of new schemes, often referred to as modern- or unsourced-random access [5], constructively embrace inter-ference and decode users resorting to advanced physical layer techniques. On the other hand, a number of solutions have been proposed to efficiently resolve users via retransmissions of collided packets. Among these, a smart approach is offered by tree-based algorithms. Remaining true to the random-access paradigm, these schemes rely on feedback and provide simple, recursive, procedures that users can implement after a collision, until each of the involved senders manages to deliver its message. Originally introduced by Capetanakis [1], the

idea has been studied thoroughly, e.g., [6]–[9] and references therein, with particular attention to improving throughput.

Notably, while the performance of random access solutions is well-understood for traditional metrics, more limited results are available in terms of AoI. Initial works explored the behavior of ALOHA and some age-specific optimizations, e.g., [10]–[13], whereas some recent contributions tackled modern random access schemes, e.g., [14], [15]. On the other hand, a characterization of tree-based algorithms remains largely unexplored, with early results [16] considering the generate-at-will case where nodes always have new messages to transmit.

In the present work, we instead tackle the more practical IoT condition of *exogenous* traffic generation, in which a device may not have fresh data to send at all times. For this setting we provide the first analytical characterization of the average AoI of the Capetanakis scheme with gated access [1], resorting to a Markovian approach. Our study captures the non-trivial and fundamentally dynamic behavior of the protocol, where the number of users trying to access the channel is driven by the duration of the previous collision resolution time, and explores its implication on AoI. To this aim, we also consider a variation of the strategy, and study the idea of *early termination*. Accordingly, a contention resolution phase can be truncated once a maximum number of slots is reached, even if not all the users originally involved in the collision have been decoded. The solution triggers a fundamental trade-off between reliability and latency, and allows us to shed light on the value of keeping retransmitting packets that progressively become stale, under the more age-challenging exogenous traffic. By providing the optimal maximum duration of contention resolution under any packet generation rate, the analysis offers some useful protocol design guidelines.

II. SYSTEM MODEL

We focus on a population of U terminals (users), which share a common wireless channel to send time-stamped messages to a common receiver. Time is slotted, and the slot duration is set to accommodate the transmission of a packet. Channel access follows a random-access approach, and a collision channel model is considered. Accordingly, if a single packet is sent over a slot, it is successfully decoded, whereas if two or more nodes access the channel simultaneously, no data is retrieved. Terminals implement the Capetanakis-Tsybakov-Mikhailov algorithm with gated access (CTM) [1].

In the interest of space, we provide in the following a simple description of the scheme, referring the reader to the vast available literature for further details, e.g., [6]–[8].

The protocol works as sequence of contention resolution intervals (CRIs), composed by a variable number of slots. At the end of a CRI, all nodes that have a new packet to send, will transmit with probability 1 in the next slot. Three cases are possible: i) no terminal has data and the slot remains *idle*; ii) only one user has data (*singleton*); iii) two or more terminals transmit, leading to a *collision*. The outcome is broadcasted by the receiver to all nodes via an instantaneous and ideal ternary feedback. If an idle or singleton slot was experienced, the newly started CRI immediately comes to an end. If, instead, a collision took place, all nodes involved in it initiate a resolution procedure, wheras any other terminal remains silent until the CRI will be concluded (gated access [1], [8]). Each terminal that collided, flips a binary fair coin, choosing whether to (a) attempt transmission again, or (b) keep idle. In the latter case, the terminal will refrain from accessing the channel until *all* the users who decided for option (a) successfully delivered their message. When this condition is met, the user will transmit in the next slot with probability 1. Instead, any user in the original collision who decided to attempt again (option (a)), will re-transmit its message in the very next slot. This set of rules is iterated, with transmitting nodes splitting in two groups after a collision, and with any user who experiences a singleton slot exiting the contention and refraining from access until the next CRI. The algorithm allows for a simple implementation, proposed by Gallager [6], using two counters that track the evolution of the contention and are updated based on the feedback after each slot.

In the remainder, we will refer to this classical scheme as the *plain CTM* algorithm. We remark that the protocol has two inherent properties. On the one hand, all nodes that join a CRI are resolved, i.e., any packet that is sent is eventually delivered. On the other hand, this leads to potentially long resolution intervals, as the duration of a CRI is unbounded.[1] To explore this trade-off, we also study a variation of the scheme, denoted as *CTM with early termination* (CTM-ET). The protocol works as the plain CTM, with the only difference that, if the CRI duration reaches a maximum length of L_{m} slots, the contention is terminated even if not all nodes have yet been resolved. When this happens, users that did not succeed drop their packet, and do not retransmit it. The CRI is concluded, and the next one starts as before (with all counters being reset).

The system description is completed by considering how terminals generate traffic. In this respect, we assume that, at each slot, a user independently generates a new time-stamped message with probability ρ, and places it in a one-packet sized buffer. The message is removed from the buffer at the start of the CRI over which it is transmitted, or if it is replaced by the generation of a new packet. Accordingly, at the end of a CRI of duration ℓ slots, a node will have a message to send (and

---

[1]We note that the system does not undergo the well-studied instability issues of CTM [6], as we deal with a finite population.

will thus contend in the next CRI) if it has generated at least once in the past ℓ slots, i.e., with probability

$$\Gamma_\ell := 1 - (1 - \rho)^\ell.$$

The model is inspired by practical IoT networks, in which a transmitter is often fed with sensor readings and cannot control when these are produced. We are thus concerned with an *exogenous* rather than the classical *generate-at-will* traffic.

In this setting, we are interested in the ability of the access schemes to maintain an up-to-date perception at the receiver. We thus focus without loss of generality on a reference user, and consider its instantaneous AoI [2]. This is defined at time t as $\delta(t) = t - \tau(t)$, where $\tau(t)$ is the generation time of the last packet successfully delivered by the user. We then measure performance in terms of the average AoI: $\Delta = \lim_{t \to \infty} \frac{1}{t} \int_0^t \delta(v) dv$.

Notation: We denote a discrete random variable (r.v.) and its realization by upper- and lower-case letters, respectively, e.g. X, x. The probability mass function of X is $p_X(x)$, with straightforward extension to conditional distributions. Whenever clear, we will omit the subscript for brevity. Moreover, we denote by calligraphic font the probability generating function (PGF) of X, i.e., $\mathcal{X}(z) = \mathbb{E}\left[z^X\right] = \sum_x p(x)z^x$, where the summation spans the whole alphabet of X.

III. PRELIMINARIES

We start by stating some results that will be useful in the computation of the average AoI.

Lemma 1. *For plain CTM, let $\tilde{L}$ be the r.v. denoting the duration in slots of a CRI, and let $\mathcal{L}_u(z)$ be its PGF conditioned on having u devices contending at the start of the CRI. Then, the following recursion holds:*

$$\mathcal{L}_u(z) = \frac{z}{2(2^{u-1} - z^2)} \sum_{i=1}^{u-1} \binom{u}{i} \mathcal{L}_i(z) \mathcal{L}_{u-i}(z) \quad (1)$$

with $\mathcal{L}_0(z) = \mathcal{L}_1(z) = z$.

Proof: Following standard arguments, e.g., [7], the PGF for a binary fair split can be expressed as

$$\mathcal{L}_u(z) = \sum_{i=0}^{u} \binom{u}{i} \frac{z}{2^u} \mathcal{L}_i(z) \mathcal{L}_{u-i}(z). \quad (2)$$

The statement follows by rearranging (2), and by observing that when a single or no user contend the CRI is terminated after one slot, so that $\sum_{\tilde{\ell}=1}^{\infty} p(\tilde{\ell})z^{\tilde{\ell}} = z$. ∎

Remark 1. *The recursion in (1) allows computing the conditional PGF of $\tilde{L}$ for any value of u. Moreover, for any K, evaluating the PGF in $z = e^{-j2\pi k/K}$, $k \in \{1, \dots, K\}$, provides the discrete Fourier transform (DFT) of the sequence $p_{\tilde{L}|U}(\tilde{\ell}|u)$ for $\tilde{\ell} \in \{1, \dots, K\}$. These PMF values can thus be obtained by taking the inverse DFT (IDFT) of the sequence.*

Lemma 2. *For plain CTM, focus on a node contending over a CRI, and let $\tilde{D}$ denote the number of slots between the start of the CRI and the moment the node is decoded, conditioned*

on having $m\in\{0,\ldots,\mathsf{U}-1\}$ other users contending. The PGF of $\tilde{D}$ conditioned on M satisfies the recursion $\mathcal{D}_1(z)=z$, and

$$\mathcal{D}_{m+1}(z) = \frac{z}{2^{m+1}-z(z+1)}\Big[z(1+\mathcal{L}_m(z))$$
$$+ \sum_{i=1}^{m-1}\binom{m}{i}\left(\mathcal{D}_{i+1}(z)+\mathcal{L}_i(z)\,\mathcal{D}_{m-i+1}(z)\right)\Big]$$

Proof: The conditional PGF of $\tilde{D}$ was originally derived in [8]. Computing it for the binary fair splitting case and rearranging the terms leads to the recursion. ∎

Also in this case, the PMF $p_{\tilde{D}|M}(\tilde{d}|m)$, can be computed efficiently by taking an IDFT of the sequence $\mathcal{D}_m(e^{-j2\pi k/\mathsf{U}})$ for $k\in\{0,\ldots,\mathsf{U}-1\}$, as streamlined in Remark 1.

A. CTM with Early Termination (CTM-ET)

Let us now focus on the case of early termination, and let L be the r.v. denoting the duration of a CRI. The PMF of L conditioned on the number of contending users readily follows from the plain CTM case, as

$$p_{L|U}(\ell\,|\,u) = \begin{cases} p_{\tilde{L}|U}(\ell\,|\,u) & \ell < L_{\mathsf{m}} \\ \sum_{\ell=L_{\mathsf{m}}}^{\infty} p_{\tilde{L}|U}(\ell\,|\,u) & \ell = L_{\mathsf{m}} \end{cases}$$

The dynamic behavior of CTM over subsequent CRIs can further be captured by means of a few conditional PMFs. The first characterizes the random number of users U_i accessing the i-th CRI, conditioned on the duration L_{i-1} of the preceding contention period. Recalling that each node transmits at the start of a CRI if it generated at least one packet over the previous one, we obtain the binomial distribution

$$p_{U|L}(u_i\,|\,\ell_{i-1}) = \binom{\mathsf{U}}{u_i}(\Gamma_{\ell_{i-1}})^{u_i}\left(1-\Gamma_{\ell_{i-1}}\right)^{\mathsf{U}-u_i}. \quad (3)$$

Leaning on (3), the duration of the i-th CRI conditioned on the previous contention follows as

$$p(\ell_i\,|\,\ell_{i-1}) = \sum_{u_i=0}^{\mathsf{U}} p_{L|U}(\ell_i\,|\,u_i)\,p_{U|L}(u_i\,|\,\ell_{i-1}). \quad (4)$$

In other words, the stochastic process L_n describing the duration of subsequent CRIs is Markovian, with one-step transition probabilities given by (4). The stationary distribution, derived via standard methods, will be denoted by π_ℓ, $\ell\in\{1,\ldots,L_{\mathsf{m}}\}$.

In the remainder, we often focus on a user of interest, and track the number of other nodes M that access a CRI. In this case, the distribution conditioned on the previous CRI duration, $p_{M|L}(m_i\,|\,\ell_{i-1})$, is also binomial, i.e., $\mathrm{Bin}(\mathsf{U}-1,\Gamma_{\ell_{i-1}})$.

Finally, we introduce the following result:

Lemma 3. *Focus on a reference user contending over a CRI, and let $\varphi(m)$ be the joint probability that the CRI terminates after L_{m} slots without resolving all terminals, and the reference user is* not *decoded. Then,*

$$\varphi(m) = 1 - \sum_{\tilde{d}=1}^{L_{\mathsf{m}}} p_{\tilde{D}|M}(\tilde{d}\,|\,m). \quad (5)$$

Proof: The result follows observing that the event corresponds to having the user decoded after slot L_{m} if plain
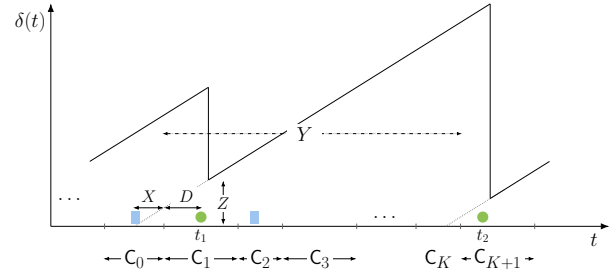

Fig. 1. Time evolution example of AoI. Blue rectangles denote generation of a new packet at the user, whereas green dots a successful delivery. Due to early termination, the packet generated in CRI C_2 is not delivered over C_3.

CTM without early termination were employed. Leaning on the conditional PMF of $\tilde{D}$ obtained from Lemma 2, this can be expressed as $\sum_{d=L_{\mathsf{m}}+1}^{\infty} p_{\tilde{D}|M}(\tilde{d}\,|\,m)$, or, equivalently (5). ∎

IV. AVERAGE AoI CHARACTERIZATION

We treat the case of early termination, clarifying in Remark 2 how the analysis can be adapted to plain CTM. Let us focus without loss of generality on a node of interest, and denote by Y_n the stochastic process describing its AoI inter-refresh time, i.e., the number of slots between two successive update deliveries by the device. Moreover, let Z_n be the value at which the AoI is reset at the beginning of the n-th inter-refresh interval (see Fig. 1). Under the assumption that the processes are ergodic, we focus on their stationary behavior, and drop the subscript n for readability. With this notation, following standard geometric arguments [2], [15], the average AoI of the node of interest can be expressed as

$$\Delta = \frac{\mathbb{E}[ZY]+\mathbb{E}[Y^2]/2}{\mathbb{E}[Y]}. \quad (6)$$

The calculation of (6) is non-trivial. Indeed, the different CRIs within an inter-refresh period have durations that are not independent of each other, due to the dynamic behavior of CTM. Moreover, Z and Y are also not independent, as the value of Z is influenced by the duration of the CRI which leads to the AoI reset, and this in turn impacts the first CRI in Y.

To tackle the problem, the timeline reported in Fig. 1 shows the structure of an inter-refresh period. At the start, the reference node is idle, and generates a new update during CRI C_0. The blue rectangle shows the slot of the last message generated by the node during C_0, i.e., the time-stamp of the packet it will send over the subsequent CRI C_1. In turn, the green dot shows the slot in C_1 at which the device is decoded, resetting its AoI value to Z. This is followed by a number of CRIs during which the device does not deliver, until C_K is reached, which sees the generation of a new update, sent and decoded over C_{K+1}. In this example, the inter-refresh time is $Y=t_2-t_1$. To ease calculations, we introduce here a simplification, and express $Y \simeq \sum_{i=1}^{K} L'_i$ where the r.v. L'_i denotes the number of slots composing CRI C_i. The approach is motivated by noting that the formulation of Y disregards the

time between the end of C_K and t_2, yet compensates this by adding the same statistical quantity for the previous refresh, i.e., the time between the start of C_1 and t_1. The tightness of the assumption will be verified in Sec. V.

Before tackling (6), it is useful to derive the joint distribution of the duration of CRIs C_0 and C_1, i.e., $p(\ell_0', \ell_1')$. To this aim, consider the conditions leading to the start of an inter-refresh period: i) a CRI of duration ℓ_i slots, with at least one packet generated by the node, followed by ii) a CRI of duration ℓ_{i+1} in which the node successfully delivers. We denote the event probability as $\theta(\ell_i, \ell_{i+1})$, and have, for $\ell_{i+1} < L_m$

$$\theta(\ell_i, \ell_{i+1}) = \pi_{\ell_i} \Gamma_{\ell_i} \sum_{m=0}^{U-1} p_{M|L}(m|\ell_i) p_{L|U}(\ell_{i+1}|m+1). \quad (7)$$

In (7), the factor $\pi_{\ell_i} \Gamma_{\ell_i}$ captures the probability of event i), whereas the rest accounts for the probability of ii), i.e., having a CRI of ℓ_{i+1} slots, given ℓ_i, and decoding the user of interest. The latter is obtained in turn by conditioning on the number m of nodes other than the user that become active over the first CRI, and noting that for $\ell_{i+1} < L_m$, all users are decoded. When the maximum CRI duration is reached, instead, we have to account for two possibilities: all users are resolved in exactly L_m slots, or the CRI is terminated, but the user of interest is decoded. Leaning on Lemma 3, we can then write

$$\theta(\ell_i, L_m) = \pi_{\ell_i} \Gamma_{\ell_i} \sum_{m=0}^{U-1} p_{M|L}(m|\ell_i) \big[p_{L|U}(L_m|m+1) - \varphi(m) \big]. \quad (8)$$

Based on (7)-(8), we finally have the overall distribution

$$p(\ell_0', \ell_1') = \frac{\theta(\ell_0', \ell_1')}{\sum_{\ell_i, \ell_{i+1}} \theta(\ell_i, \ell_{i+1})} \quad (9)$$

where the summation is taken for $\ell_i, \ell_{i+1} \in \{1, \ldots, L_m\}$.

A. Statistical moments of the inter-refresh time

Let us tackle the expected value of Y, conditioned on the duration of the first CRI C_1, denoted as $Y_{\ell_1'}^1 := \mathbb{E}[Y \mid L_1' = \ell_1']$. Leaning on the Markovian evolution of the contention durations, a first step analysis [17] can be applied, to obtain

$$Y_{\ell_1'}^1 = \ell_1' + \sum_{\ell'=1}^{L_m} Y_{\ell'}^1 \cdot (1 - \Gamma_{\ell_1'}) \sum_{m=0}^{U-1} p_{M|L}(m|\ell_1') \, p_{L|U}(\ell'|m)$$
$$+ Y_{L_m}^1 \Gamma_{\ell_1'} \sum_{m=0}^{U-1} p_{M|L}(m|\ell_1') \, \varphi(m). \quad (10)$$

In (10), the overall duration is expressed as the sum of the first CRI (ℓ_1' slots), and of the expected duration of an inter-refresh period starting with a CRI of length ℓ', weighted by the probability for this to be undergone after C_1. Consider the first row of (10). In this case, the reference user does not generate a packet over C_1 (probability $1 - \Gamma_{\ell_1'}$), and the inter-refresh period will certainly continue. Following the usual approach, the probability for the next CRI to last ℓ' can directly be obtained by conditioning on the number of nodes that will be active among the remaining $U-1$. If, instead, the reference node does generate at least one packet during C_1 (second row

of the expression), the inter-refresh period will continue only if the user will not successfully deliver the message over the next CRI. This requires the CRI to be of duration L_m slots, to be terminated, and for the user not to be decoded. Conditioning on the number of other contenders, this probability is exactly captured by $\varphi(m)$ in Lemma 3. Overall, (10) can be computed for any starting value of ℓ_1', providing a full-rank system of L_m linear equations in L_m unknowns, whose solution can be derived with standard methods and provides the sought first order moment of Y conditioned on the length of C_1.

The same reasoning can be applied to derive the second order moment of Y conditioned on L_1', denoted as $Y_{\ell_1'}^2$. In this case, the first-step approach leads to the system of equations

$$Y_{\ell_1'}^2 = (\ell_1')^2 + \sum_{\ell'=1}^{L_m} \alpha(\ell_1', \ell')(1 - \Gamma_{\ell_1'}) \sum_{m=0}^{U-1} p_{M|L}(m|\ell_1') p_{L|U}(\ell'|m)$$
$$+ \alpha(\ell_1', L_m) \Gamma_{\ell_1'} \sum_{m=0}^{U-1} p_{M|L}(m|\ell_1') \, \varphi(m) \quad (11)$$

where we have introduced for compactness the ancillary quantity $\alpha(\ell_1', \ell') := (2\ell_1' Y_{\ell'}^1 + Y_{\ell'}^2)$.

B. AoI reset value

Consider now $\mathbb{E}[ZY]$, conveniently expressed as

$$\mathbb{E}[ZY] = \sum_{\ell_0', \ell_1'} \mathbb{E}[ZY \mid L_0' = \ell_0', L_1' = \ell_1'] \, p(\ell_0', \ell_1'). \quad (12)$$

Conditioned on the duration of the two CRIs, the age-reset value Z and the inter-refresh time Y become in fact independent. This comes from the fact that the former refers to a packet generated in C_0 and decoded in C_1, whereas the latter is driven by the generation and delivery of subsequent packets, which is an independent process once the duration L_1' is fixed. The overall computation thus simplifies into the product of the two conditional expectations of Y and Z.

We focus first on the age-reset value, observing from Fig. 1 that it can be expressed as $Z = X + D$. Here, X is the number of slots elapsed between the packet generation and the end of C_0, and D is the time between the start of C_1 and the delivery of the message. To derive the statistics of X, we note that the variable by definition does not depend on L_1'. In turn, the PMF conditioned on L_0' can be written for any $x \in \{1, \ldots, \ell_0'\}$ as

$$p_{X|L_0'}(x \mid \ell_0') = \big[\rho(1-\rho)^{x-1} \big] / \Gamma_{\ell_0'} \quad (13)$$

where the numerator captures the probability of generating the message and then not producing updates for the remaining $x-1$ slots of C_0, whereas the denominator normalizes the distribution to having generated at least one packet.

On the other hand, the conditional distribution of D given L_0' and L_1' is non-trivial. In particular, determining the decoding slot of a message conditioned on the duration of the CRI it is sent over introduces inter-dependencies that prevent a direct application of the standard recursions employed for tree-based algorithms. In view of this, we resort to an approximation, and consider the term $\mathbb{E}[D|L_0' = \ell_0', L_1' = \ell_1'] \simeq \mathbb{E}[D|L_0' = \ell_0']$. The approach stems by observing that conditioning on L_0'
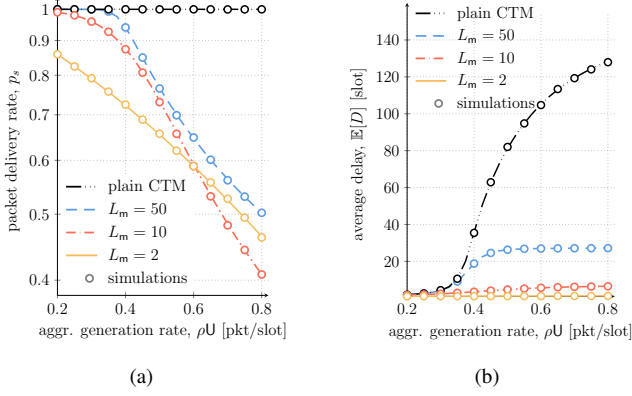
Fig. 2. (a) packet delivery rate, p_s and (b) average delay, $\mathbb{E}[D]$, plotted against the aggregate generation rate ρU.

already determines the number of contenders over C_1, and thus drives its duration. The approximation proves to be extremely tight for all configurations of arrival rate and L_m, as highlighted in Fig. 2(b). Following this assumption, we work with the conditional distribution $p_{D|L_0'}(d|\ell_0')$, which, in turn, can be obtained from $p_{\tilde{D}|M}(\tilde{d}|m)$ (Lemma 2) as

$$p_{D|L_0'}(d \mid \ell_0') = \frac{\sum_{m=0}^{U-1} p_{\tilde{D}|M}(\tilde{d}|m)\, p_{M|L}(m|\ell_0')}{\sum_{i=1}^{L_m} \sum_{m=0}^{U-1} p_{\tilde{D}|M}(\tilde{i}|m)\, p_{M|L}(m|\ell_0')}. \quad (14)$$

For $d\in\{1,\dots,L_m\}$, the numerator in (14) reflects the condition on L_0' in terms of the number of contenders M over C_1, while the denominator provides the normalization for considering CRIs (and thus decoding) of maximum L_m slots.

The distributions in (13) and (14) allow computing the (approximated) conditional expectation $\mathbb{E}[Z|L_0'=\ell_0', L_1'=\ell_1']$. On the other hand, Y only depends on L_1', whose conditional expectation has been captured in Sec. IV-A. Plugging the product of the two into (12) and leveraging the joint PMF $p(\ell_0', \ell_1')$ in (9) finally provides the sought first order moment, concluding the calculation of the average AoI in (6).

Remark 2. *The analysis presented so far for early-termination can be extended straightforwardly to plain CTM. In this case, besides considering summations that span $\ell\in\mathbb{N}$ and referring to the r.v.s $\tilde{L}$ and $\tilde{D}$, an active user is always decoded within a CRI. This simplifies the analysis, as, for instance, the final summation terms in (10) and (11) can be omitted, describing early-terminated CRIs in which an active node does not deliver. For all practical purposes, however, the plain CTM performance can be obtained by implementing the reported calculations for a sufficiently large value of L_m.*

V. RESULTS AND DISCUSSION

To evaluate performance, we consider $U=100$ users. Different U values were also studied, leading to similar trends, and are not reported for brevity. In the plots, lines report the analysis of Sec. IV, and circle markers the outcome of Montecarlo simulations, fully implementing the CTM protocol.

To gather insights on the trade-offs induced by early termination, we report in Fig. 2(a) the packet delivery rate p_s, i.e., the probability that a contending node succeeds within L_m, and in Fig. 2(b) the average delay undergone by a delivered message, $\mathbb{E}[D]$. Both quantities are shown against the aggregate packet generation rate, ρU. The black, dash-dotted line denotes plain CTM, whereas the colored lines refer to CTM-ET, for different L_m. Consider first the packet delivery rate. The metric can be derived analytically as

$$p_s = \big[\sum_{\ell_i, \ell_{i+1}} \theta(\ell_i, \ell_{i+1})\big] / \sum_{\ell_i} \pi_{\ell_i} \Gamma_{\ell_i}. \quad (15)$$

In (15), the numerator captures the probability for a node to generate a packet over the i-th CRI, and to deliver it over the next one, as computed in (7)-(8), whereas the denominator normalizes to having the node transmit over the $(i+1)$-th CRI. By construction, all packets are delivered with the plain CTM ($p_s=1$). Instead, early termination may lead to losses, and the effect is more pronounced as L_m reduces. Eventually, for $L_m=2$, the protocol approaches slotted ALOHA, where each slot would see independent contention, without resolution attempts, and attain the well-known $p_s=e^{-\rho U}$ [3]. More interestingly, the configuration $L_m=10$ experiences an even higher loss rate for large ρ, in spite of allowing more time for collision resolution. This effect is rooted in the dynamic behavior of CTM across CRIs. An intuition is obtained recalling that, at the end of a CRI of ℓ slots, on average $U(1-(1-\rho)^\ell)$ users will have updates to send in the next slot. Assume that CRIs run until they are terminated. When $L_m=2$, on average ~ 1.6 nodes will contend, whereas for $L_m=10$ this increases to ~ 7.8. As a result, in the latter case, a much larger collision set has to be resolved, and it is likely that not all the user will be retrieved within the available 10 slots, leading to a lower delivery rate. The increased losses triggered by an early termination, however, also reduce the expected delay of delivered messages, as shown by Fig. 2(b).[2] In this respect, the excellent match between simulations and analysis confirms the validity of the approximation introduced in Sec. IV.

The fundamental trade-off between latency and update delivery eventually drives the AoI performance. The behavior of the metric, normalized to the number of nodes, is reported in Fig. 3 as a function of ρU. For plain CTM and for low arrival rates, large values of Δ are experienced, as nodes sporadically have new updates to send, leaving the receiver with stale information. In turn, high packet generation rates lead the system to operate over longer CRIs, where more nodes contend. Congestion leads to larger latencies, and thus to an increase of AoI. Notably, the minimum Δ is attained by plain CTM for an aggregate generation rate ~0.347, corresponding to peak throughput conditions [6]. When early termination is implemented, two effects can be noticed. On the one hand, a worse performance is attained for low ρ, as AoI deteriorates of more than 15% for $L_m=2$ compared to plain CTM. A complete collision resolution is thus especially beneficial in

---

[2]Analytical results were in this case obtained by removing the conditioning on L_0' in (14) via the marginalization of (9), and by taking the expectation.
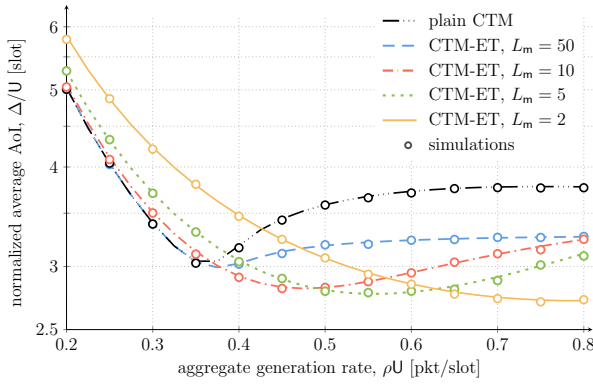
Fig. 3. Normalized average AoI, Δ/U vs aggregate generation rate, ρU.

(a)   (b)

Fig. 4. (a) minimum normalized average AoI for the optimized CTM-ET, and (b) corresponding optimal L_m value.

such conditions, allowing all nodes to deliver their message and avoiding potentially long periods of time without further updates generation. On the other hand, a truncation becomes convenient for larger values of ρ. As highlighted by the plot, an AoI reduction of roughly 30% is gained for $L_m=2$ compared to resolving all nodes. The result can be understood observing how plain CTM spends more time to resolve the larger initial collision sets experienced for high generation rates, leading to delivery of messages that have in the meantime become stale. In this respect, CTM-ET would drop packets whose time stamp is growing old, allowing nodes to attempt transmission of fresher updates that may have been generated in the meantime.

The presented result provide thus a key take-away, pinpointing the value of a thorough collision resolution under sporadic generation, and of a more aggressive early termination in harsher traffic conditions. This, in turn, triggers the natural question of how L_m shall be tuned in order to minimize AoI. We tackle this aspect in Fig. 4(a), which shows, for any ρU, the minimum AoI obtained by optimizing over L_m (solid line). For reference, the performance of the plain CTM (dashed line) as well as of a standard slotted ALOHA with no contention resolution (dash-dotted line) are also reported. The latter is obtained via the well-known formulation $\Delta = 1/2 + [\rho(1-\rho)^{U-1}]^{-1}$ [10], [14]. The improvement triggered by a proper application of the CTM-ET approach is evident, with significant reductions in AoI with respect to both benchmarks especially for intermediate values of ρ. The corresponding optimal choices of L_m are shown in Fig. 4(b), where the gray-shaded region to the left corresponds to having $L_m = \infty$ (plain CTM). From this standpoint, the presented framework provides a useful tool for protocol parameter tuning.

VI. CONCLUSIONS

We presented an analytical characterization of the average AoI under the Capetanakis tree-based random access algorithm with gated access, in the presence of exogenous traffic. Via a Markovian approach, we capture the coupling between contention dynamics and information freshness. Moreover, we proposed a truncated resolution mechanism, revealing a trade-off between latency and r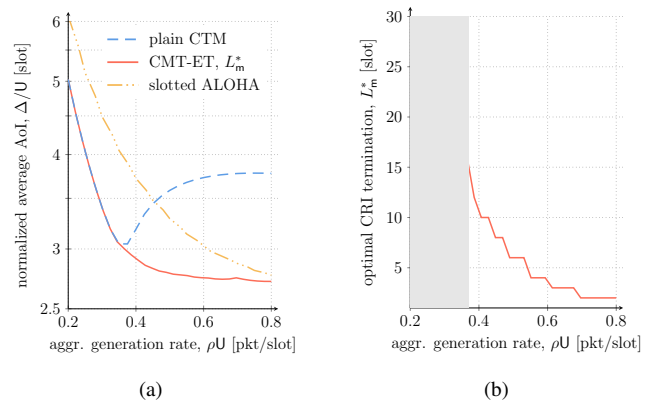eliability that becomes critical under age-sensitive traffic. The resulting insights provide design guidelines for IoT systems employing feedback-based collision resolution in scenarios where timely updates are crucial.

REFERENCES

[1] J. Capetanakis, "Tree algorithms for packet broadcast channels," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 505–515, 1979.

[2] R. Yates, Y. Sun, D. Brown, S. Kaul, E. Modiano, and S. Ulukus, "Age of information: An introduction and survey," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1183–1210, May 2021.

[3] N. Abramson, "The throughput of packet broadcasting channels," *IEEE Trans. Commun.*, vol. COM-25, no. 1, pp. 117–128, 1977.

[4] L. Iiyambo, G. Hancke, and A. Abu-Mahfouz, "A survey on NB-IoT random access: Approaches for uplink radio access network congestion management," *IEEE Access*, vol. 12, pp. 95 487–95 506, 2024.

[5] G. Liva and Y. Polyanskiy, "Unsourced multiple access: A coding paradigm for massive random access," *Proceedings of the IEEE*, vol. 112, no. 9, pp. 1214–1229, 2024.

[6] J. Massey, "Collision-resolution algorithms and random-access communications," *Multi-user communications systems*, pp. 73–137, 1981.

[7] P. Mathys and P. Flajolet, "Q-ary collision resolution algorithms in random-access systems with free or blocked channel access," vol. 31, no. 2, pp. 217–243, 1985.

[8] Y. Yu and G. B. Giannakis, "High-throughput random access using successive interference cancellation in a tree algorithm," vol. 53, no. 12, pp. 4628–4639, 2007.

[9] Q. Vogel, Y. Deshpande, C. Stefanović, and W. Kellerer, "Analysis of d-ary tree algorithms with successive interference cancellation," *Journal of Applied Probability*, vol. 61, no. 3, pp. 1075–1105, Sep. 2024.

[10] R. Yates and S. Kaul, "Status updates over unreliable multiaccess channels," in *Proc. IEEE ISIT*, 2017.

[11] ——, "Age of information in uncoordinated unslotted updating," in *Proc. IEEE ISIT*, 2020.

[12] O. Yavaskan and E. Uysal, "Analysis of slotted ALOHA with an age threshold," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1456–1470, 2021.

[13] X. Chen, K. Gatsis, H. Hassani, and S. Bidokhti, "Age of information in random access channels," *IEEE Trans. Inf. Theory*, vol. 68, no. 10, pp. 6548–6568, 2022.

[14] A. Munari, "Modern random access: an age of information perspective on irregular repetition slotted ALOHA," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 3572–3585, 2021.

[15] A. Munari, F. Lázaro, G. Durisi, and G. Liva, "The dynamic behavior of frameless ALOHA: Drift analysis, throughput, and age of information," *IEEE Trans. Commun.*, vol. 71, no. 12, pp. 6914–6927, 2023.

[16] H. Pan, T.-T. Chan, J. Li, and V. C. M. Leung, "Age of information with collision-resolution random access," *IEEE Trans. Veh. Technol.*, vol. 71, no. 10, pp. 11 295–11 300, 2022.

[17] R. Gallager, *Stochastic Processes: Theory for Applications*. Cambridge University Press, 2013.