Age of Information in Multihop Multicast Networks

Baturalp Buyukates, Alkan Soysal, and Sennur Ulukus

Abstract: **We consider the age of information in a multihop multicast network where there is a single source node sending time-sensitive updates to n^L end nodes, and L denotes the number of hops. In the first hop, the source node sends updates to n first-hop receiver nodes, and in the second hop each first-hop receiver node relays the update packets that it has received to n further users that are connected to it. This network architecture continues in further hops such that each receiver node in hop ℓ is connected to n further receiver nodes in hop $\ell + 1$. We study the age of information experienced by the end nodes, and in particular, its scaling as a function of n. We show that, using an earliest k transmission scheme in each hop, the age of information at the end nodes can be made a constant independent of n. In particular, the source node transmits each update packet to the earliest k_1 of the n first-hop nodes, and each first-hop node that receives the update relays it to the earliest k_2 out of n second-hop nodes that are connected to it and so on. We determine the optimum k_ℓ stopping value for each hop ℓ for arbitrary shifted exponential link delays.**

Index Terms: **Age of information, large networks, scaling laws.**
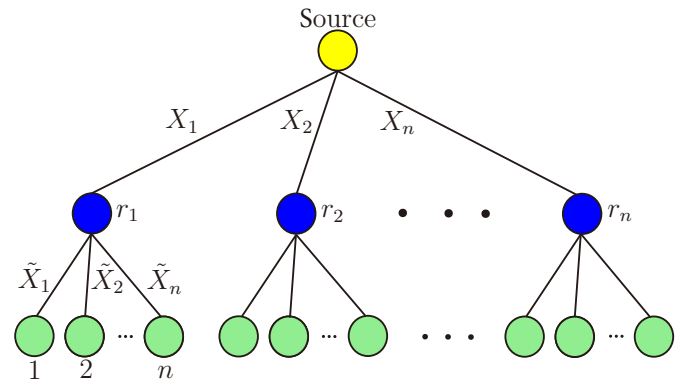
Fig. 1. Two-hop multicast network with a single source node sending updates through n middle nodes each of which is tied to further n end nodes.

I. INTRODUCTION

RECENTLY, with the increase in the number of communication network applications requiring real-time status information, timeliness of the received messages has become a critical and desirable feature for networks. Such applications include sensor networks measuring ambient temperature [1], autonomous vehicular networks where instantaneous vehicle information including velocity, position and acceleration is needed [2] and news reports from Twitter. In all these applications, information loses its value as it becomes stale.

This motivates the study of age of information, which is a metric measuring the freshness of the received information. A typical model to study age of information includes a source which acquires time-stamped status updates from a physical phenomenon. These updates are transmitted over the network to the receiver(s) and the age of information in this network, or simply the age, is the time elapsed since the most recent update at the receiver was generated at the transmitter. In other words, at time t, age $\Delta(t)$ of a packet which was generated at time $u(t)$ is $\Delta(t) = t - u(t)$.

B. Buyukates and S. Ulukus are with the Department of Electrical and Computer Engineering, University of Maryland, MD, USA, email: {baturalp, ulukus}@umd.edu.

A. Soysal is with the Department of Electrical and Electronics Engineering, Bahcesehir University, Istanbul, Turkey, email: alkan.soysal@eng.bau.edu.tr.

S. Ulukus is the corresponding author.

Most of the existing work focuses on age analysis in a queuing-theoretic setting. References [3]–[7] study the age under various arrival and service profiles. References [8] and [9] investigate packet management strategies including blocking and preemption for M/G/1/1 and G/G/1/1 queues, respectively. References [10]–[12] study multi-hop networks in which update packets are relayed from one node to another. Single-hop systems with exogenous arrivals are studied in [13] and [14]. References [15] and [16] study the stationary distribution and moments of age of information in queuing systems. In [17] and [18], the benefits of waiting in status update systems are investigated. Another line of research studies the age under the energy harvesting setting [19]–[28].

Considering dense IoT deployments and the increase in the number of users in networks supplying time-sensitive information, the scalability of age as a function of the number of nodes has become a critical issue. To this end, we need to study how the age performance of the network changes with growing network size. Reference [29] studies a mobile social network with a single service provider and n communicating users, and shows that under Poisson contact processes among users and uniform rate allocation from the service provider, the average age of the content at the users grows logarithmically in n. In contrast, reference [30] observes that in a single-hop multicast network appropriate stopping threshold k can prevent information staleness as the network grows.

Motivated by this observation, we study the scalability of the age in multihop multicast networks using similar threshold ideas. Extending the results of [30], we first analyze the single-hop problem with exogenous arrivals where the source directly communicates with the end users but cannot generate the updates itself. We then characterize the age for the two-hop case (see Fig. 1) using our single-hop with exogenous arrivals result as a building block. We show that for this two-hop multicast network under i.i.d. shifted exponential link delays and stopping thresholds k_1 and k_2 at each hop, an upper bound on the average age can be obtained. Through this upper bound, we show that
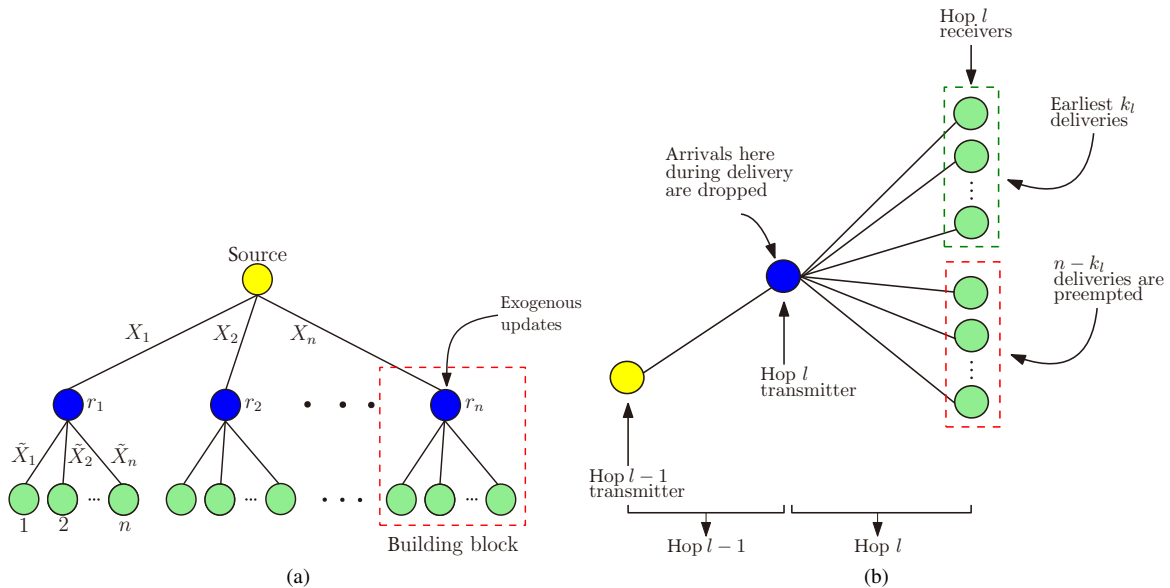
Fig. 2. Multicast network model: (a) two-hop operation and (b) details of hops $\ell - 1$ and ℓ.

the average age is limited by a constant as n increases. Utilizing this upper bound, we then extend the result of two-hop network to L-hop multicast networks for general L. We determine the optimal stopping threshold for each hop ℓ, k_ℓ, that minimizes the average age for arbitrary shifted exponential link delays.

II. SYSTEM MODEL AND AGE METRIC

We consider a multihop system where in the first hop a single source node broadcasts time-stamped updates to n first-hop receiver nodes using n links with i.i.d. random delays, and in the second hop, each first-hop receiver node relays the update packets it has received to n further nodes that are connected to it. This network architecture continues in further hops such that each receiver node in hop $\ell - 1$ acts as a transmitter in hop ℓ and it is connected to n further receiver nodes in hop ℓ. Fig. 1 shows the described model for two hops. We first consider the age of information in a two-hop ($L = 2$) multicast network and then extend our results to general multihop multicast networks with L hops, where L is arbitrary.

We start describing our system model within the simpler two-hop setting. In the two-hop multicast network, an update takes X time to reach from the source node to a particular mid-level node and $\tilde{X}$ time to reach from that mid-level node to a particular end node where X and $\tilde{X}$ are shifted exponential random variables with parameters (λ, c) and $(\tilde{\lambda}, \tilde{c})$, respectively, where c and $\tilde{c}$ are positive constants. Note that the constant shift parameters capture additional delay that may be experienced along with the exponential link delay and when they are zero random variables corresponding to service times become exponentially distributed. In each hop, the service times of individual links are i.i.d. realizations of random variables X and $\tilde{X}$, e.g., in the first hop service times of individual links are i.i.d. X_i and in the second hop service times of individual links are i.i.d. $\tilde{X}_i$.

In the two-hop scenario, age is measured for each of the n^2 end nodes and for node i at time t age is the random process $\Delta_i(t) = t - u_i(t)$ where $u_i(t)$ is the time-stamp of the most

recent update at that node. When the source node sends out update j, it waits for the acknowledgment from the earliest k_1 of n middle nodes. After it receives all k_1 acknowledgment signals, we say that update j has been completed and the source node generates update $j + 1$. At this time, transmissions of the remaining $n - k_1$ packets are terminated. Thus, if a node in the first hop is not in the earliest k_1 nodes to receive update j then service of this update is preempted. In the second hop, these earliest k_1 nodes that have received update j start transmitting this update to their end nodes and they stop whenever k_2 of their end nodes have received the current update. Middle nodes implement a blocking scheme when they are busy transmitting to the end nodes, i.e., they discard arriving packets when they are not idle. When the middle nodes finish transmitting the current update to k_2 of their children nodes, they wait for the arrival of the next update from the source node.

For general L, principles that are explained above are repeated at every hop, e.g., the transmitters in hop ℓ wait for k_ℓ of their children nodes to receive the current update before they declare that the current update has been completed; they drop all other incoming updates from transmitters in hop $\ell - 1$ as they are transmitting the current update; and when their update is received by k_ℓ receivers they preempt the remaining $n - k_\ell$ updates. When this is over, transmitters in hop ℓ wait for the next update from their parent nodes in hop $\ell - 1$; see Fig. 2(b).

The metric we use, time averaged age, is given by

$$\Delta = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^\tau \Delta(t) dt, \tag{1}$$

where $\Delta(t)$ is the instantaneous age of the last successfully received update as defined above. We will use a graphical argument similar to [30] to derive the average age at an individual end node. Since all link delays are i.i.d. for all nodes and packets, each end node i experiences statistically identical age processes and will have the same average age. Therefore, it suffices to focus on a single end user for the age analysis.
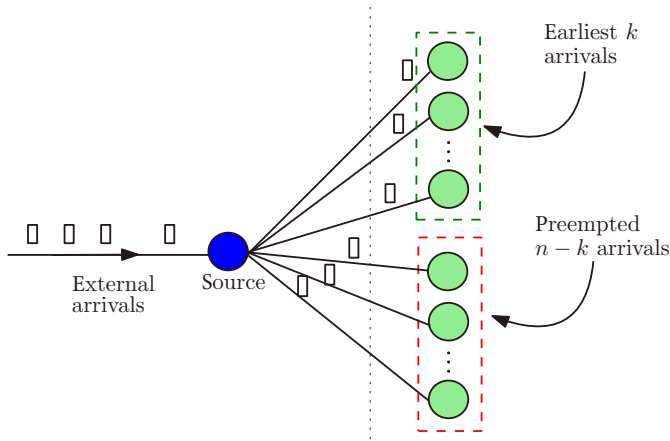
Fig. 3. The arrival and update structure of a building block.

III. BUILDING BLOCK: SINGLE-HOP NETWORK WITH EXOGENOUS ARRIVALS

We first note that at the second hop what we essentially have is n parent nodes each tied to n children nodes. Therefore, each second hop transmitter and its children nodes correspond to the single-hop network analyzed in [30] with one important difference: second hop transmitters cannot generate update packets. They can only relay packets sent from the source node. Thus, in this section, we first analyze a single-hop network in which update packets arrive exogenously with a given expected inter-arrival time $1/\mu$. Then, using this network as a building block we analyze the L-hop network described in Section II, first for $L = 2$, i.e., a two-hop network, then for general L. We have i.i.d. shifted exponential service times between the source and each of its n children nodes as in [30]. Similarly, transmission of the current update stops when k out of n nodes receive the update. Thus, in this section, we extend the results of [30] to the case of exogenous arrivals, and determine a k threshold which depends on λ, c and μ. Fig. 3 shows the arrival and update structure of a building block. Fig. 2(a) shows this building block as part of a two-hop network; and Fig. 2(b) shows it as a part of an ℓth hop in an L-hop network.

Under this model with i.i.d. link delay X, an update takes $X_{k:n}$ units of time to reach k out of n nodes where we denote the kth order statistic of random variables $X_1, \cdots, X_n$ as $X_{k:n}$. Here $X_{k:n}$ is the kth smallest of $X_1, \cdots, X_n$, e.g., $X_{1:n} = \min\{X_i\}$ and $X_{n:n} = \max\{X_i\}$. For shifted exponential random variable X, we have

$$E[X_{k:n}] = c + \frac{1}{\lambda}(H_n - H_{n-k}), \tag{2}$$

$$Var[X_{k:n}] = \frac{1}{\lambda^2}(G_n - G_{n-k}), \tag{3}$$

where $H_n = \sum_{j=1}^{n} 1/j$ and $G_n = \sum_{j=1}^{n} 1/j^2$. Using these,

$$E[X_{k:n}^2] = c^2 + \frac{2c}{\lambda}(H_n - H_{n-k})$$
$$+ \frac{1}{\lambda^2}\left((H_n - H_{n-k})^2 + G_n - G_{n-k}\right). \tag{4}$$

We say that the system is busy when a transmitted update has not been received by k out of n nodes yet. Once k end nodes

receive an update, transmitter stops the service to the remaining $n - k$ nodes. Then, the system is idle until the next update arrives. Fig. 4 shows a realization of the update process between the source node and a particular end node. Note that realizations of the update process for different end nodes might be different depending on when and for which updates these nodes have been one of the earliest k nodes. On the other hand, although realizations are different, each end node experiences the same random process. Updates arrive at the source node with an inter-arrival time, R, where $E[R] = 1/\mu$. In the most general setting, R is an arbitrary i.i.d. random variable. In Fig. 4, arrows that are above and below the source node line indicate the received and transmitted updates at the source node, respectively.

An important aspect of our model is that updates at the source node are divided into three groups, namely successful updates, dropped updates, and preempted updates. In Fig. 4, filled circles correspond to successfully received updates. They indicate that the update received at the source node started transmission to n end nodes and this particular end node received the update. We denote the time between the transmission and the reception of successful updates (filled circles) as the service time, $\bar{X}$. The order of this particular node might be smaller than k, thus $\bar{X} \leq X_{k:n}$. The source continues the service of the first filled circled update until k nodes receive the update. During this time if new updates arrive at the source, they are dropped; the dropped updates (crosses in Fig. 4) never go into service, they are lost. Once k nodes receive this update, system becomes idle and the source waits for the arrival of the next update.

We denote the waiting time until the next arrival with Z. The next arrival which is shown with an empty circle in Fig. 4 starts the service. However, for this particular realization of the update process, this particular end node is not one of the earliest k end nodes during the transmission of empty circled update. Since the source stops service once k end nodes receive the update, this transmitted update never arrives at this particular end node. We denote those updates that start a service but do not arrive at this particular end node as preempted updates. Even though preemption usually means stopping a current service in order to start a new one immediately, here we use the word preemption to mean that the current service is stopped, and a new service will start after an idle period when a new update arrives at the source. However, similar to regular meaning of preemption, in our model as well, current preempted update leaves service and a fresh update takes over.

We denote the time between two consecutive departures from the source as $Y = X_{k:n} + Z$. Note that there are a random number of dropped updates during each realization of Y. In Fig. 4, between the first successful (filled circle) update and the next update that starts the service (empty circle), there are three dropped updates (crossed). In other words, the fourth received update is able to start a new service. In addition, there are multiple realizations of Y before the next successful (filled circle) update, since this particular end node is not able to receive empty circled updates. We denote the time between two successful updates, i.e., two consecutive updates that depart from the source and successfully arrive at the end node, with S. Remember that Y is the time between two updates that depart from the source and start the service but are not guaranteed to arrive at this particu-
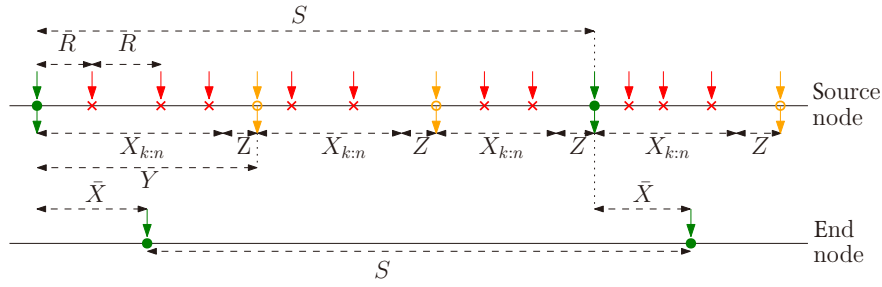
Fig. 4. Update process for an end node in the building block setting. Filled circles indicate the arrivals successfully received by this particular end node. Crosses indicate updates that arrive at the source node when it is busy transmitting another update. Empty circles show the updates that are transmitted by the source node but are not received by this particular end node, i.e., they are preempted from the perspective of this particular end node.

lar end node. We can relate the interarrival times of departing updates and arriving updates using $S = \sum_{i=1}^{M} Y_i$. In Fig. 4, this particular end node receives the first and the fourth updates that depart the source and enter the service. Therefore, in this example we have $M = 3$. In addition, remember that R is the interarrival time between the updates that arrive at the source node. For a general R, there is no closed-form expression with known variables that relates interarrival times at the end node, S, to interarrival times at the source node, R. However, when R is exponential Z is exponential as well due to the memoryless property of the exponential distribution. Thus, for exponential R we have $S = \sum_{i=1}^{M} (X_{k:n} + R)_i$.

When the current update reaches k earliest nodes, the source node terminates the remaining $n - k$ transmissions and begins to wait for the next arrival and then repeats the process upon arrival of the next update packet. Since the link delays are i.i.d., the end users receive the packet in service with probability $p = k/n$. If an end user receives update j and the next one it receives is update $j + M$, then M is geometrically distributed with p with moments

$$E[M] = \frac{1}{p} = \frac{n}{k}, \qquad E[M^2] = \frac{2-p}{p^2} = \frac{2n^2}{k^2} - \frac{n}{k}. \quad (5)$$

Similar to [30], the average age for the earliest k stopping scheme with exogenous packet arrivals with rate μ is

$$\Delta_{(k,\mu)} = \frac{E[A]}{E[S]}, \quad (6)$$

where A denotes the shaded area in Fig. 5 and S is its length. Remember from Fig. 4 that the random variable S is the interarrival time at the end nodes. Inspecting Fig. 5 to calculate A, we find $E[A] = E[S^2]/2 + E[S]E[\bar{X}]$. Here, $\bar{X}$ denotes the service time of a successful update such that $E[\bar{X}] = E[X_i|i \in \mathcal{K}]$ where $\mathcal{K}$ is the set of earliest k nodes that receive the update. Now, (6) becomes[1]

$$\Delta_{(k,\mu)} = E[\bar{X}] + \frac{E[S^2]}{2E[S]}. \quad (7)$$

[1] Since the building block studied in this section will eventually be used in the upcoming sections in the analysis of two-hop and L-hop networks, we should emphasize that the definitions of $\bar{X}$ and S do not assume any network structure. As far as (7) is considered, it is not important where an update that is received by an end node is generated. $\bar{X}$ is the time that the update spends in the system from the time it is generated (possibly by a node other than the one that relays it to the end node) until the time it is received at the end node. S is the time between two consecutive successfully received updates, regardless of whether the updates are generated at the node that transmits them or not. This reasoning is important for our derivations in multihop multicast networks.

We can write the first two moments of S in terms of Y as

$$E[S] = E[M]E[Y], \quad (8)$$
$$E[S^2] = E[M]E[Y^2] + E[Y]^2 E[M^2 - M]. \quad (9)$$

Inserting these into (7) we obtain

$$\Delta_{(k,\mu)} = E[\bar{X}] + \frac{E[M^2]}{2E[M]} E[Y] + \frac{Var[Y]}{2E[Y]}. \quad (10)$$

In the following theorem, we determine the age of an update at an end node for a single-hop building block model using (10).

Theorem 1: For a single-hop building block model with exogenous arrivals that have expected interarrival time $1/\mu$, for the earliest k stopping scheme, the average age of an update at an individual end node is

$$\Delta_{(k,\mu)} = \frac{1}{k} \sum_{i=1}^{k} E[X_{i:n}] + \frac{2n-k}{2k} (E[X_{k:n}] + E[Z])$$
$$+ \frac{Var[X_{k:n} + Z]}{2(E[X_{k:n}] + E[Z])}. \quad (11)$$

Proof: The first term comes from $E[\bar{X}]$ as

$$E[\bar{X}] = E[X_j | j \in \mathcal{K}] = \sum_{i=1}^{k} E[X_{i:n}] Pr[j = i | j \in \mathcal{K}]$$
$$= \frac{1}{k} \sum_{i=1}^{k} E[X_{i:n}]. \quad (12)$$

where we used the fact that, since we have k out of n nodes selected independently and identically in $\mathcal{K}$, we have $Pr[j = i | j \in \mathcal{K}] = 1/k$. The second and third terms are obtained upon substitution of $Y = X_{k:n} + Z$ and $E[M]$ and $E[M^2]$ expressions given in (5) into (10). ∎

When we have general interarrival times as we have in this problem, $X_{k:n}$ and Z may be dependent. However, with exponential interarrivals we can show their independence using the memoryless property, and simplify the age expression given in (11) as follows.

Corollary 1: When the arrival process is Poisson with rate μ, the age of an end node is

$$\Delta_{(k,\mu)} = \frac{1}{k} \sum_{i=1}^{k} E[X_{i:n}] + \frac{2n-k}{2k\mu} (\mu E[X_{k:n}] + 1)$$
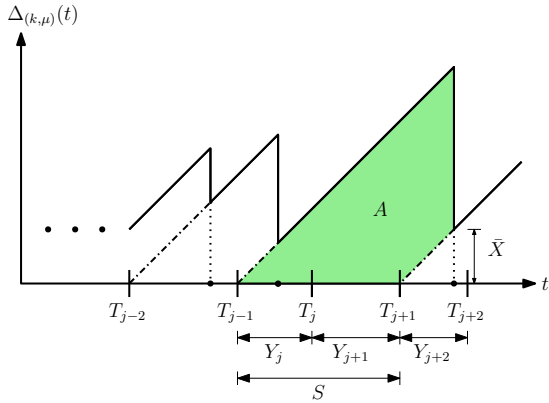
Fig. 5. Sample age evolution $\Delta_{(k,\mu)}(t)$ of an end node. Updates that find the system idle arrive at times T_i at the source. Here, update j arrives at time T_{j-1} and immediately goes into service. Successful update deliveries are marked with $\bullet$ and in this figure, updates $j-1$, j and $j+2$ are delivered successfully whereas update $j+1$ is terminated.

$$+ \frac{\mu Var[X_{k:n}]}{2(\mu E[X_{k:n}]+1)} + \frac{1}{2(\mu^2 E[X_{k:n}]+\mu)}. \quad (13)$$

Proof: When the arrival process is Poisson with μ, in other words, R is exponential with expected interarrival time $1/\mu$, the random variable Z which corresponds to the residual interarrival time is exponentially distributed with the same parameter, i.e., $Z = R$. In addition, Z is independent of X due to the memoryless property of the exponential distribution. Then,

$$Var[Y] = Var[X_{k:n} + Z] = Var[X_{k:n}] + Var[Z], \quad (14)$$

and we plug in $E[Z] = 1/\mu$ and $Var[Z] = 1/\mu^2$. $\blacksquare$

When the service times X are i.i.d. shifted exponential random variables and when n is large, we can further simplify the age expression in (13) as follows.

Corollary 2: For large n and $n > k$, set $k = \alpha n$. For shifted exponential (λ, c) service times X, the average age for the earliest k scheme with exogenous Poisson arrivals with rate μ can be approximated as

$$\Delta_{(k,\mu)} \approx \frac{c}{\alpha} + \frac{c}{2} + \frac{1}{\lambda} - \frac{1}{2\lambda}\log(1-\alpha) + \frac{1}{\alpha\mu} - \frac{1}{2\mu}$$
$$+ \frac{1}{2}\left(\mu^2 c - \frac{\mu^2\log(1-\alpha)}{\lambda} + \mu\right)^{-1}. \quad (15)$$

Proof: Using the order statistics above,

$$\delta_1 = \frac{1}{k}\sum_{i=1}^{k} E[X_{i:n}] = c + \frac{H_n}{\lambda} - \frac{1}{k\lambda}\sum_{i=1}^{k} H_{n-i}. \quad (16)$$

As in [30], we have $\sum_{i=1}^{k} H_{n-i} = \sum_{i=1}^{n-1} H_i - \sum_{i=1}^{n-k-1} H_i$ and the series identity $\sum_{i=1}^{k} H_i = (k+1)(H_{k+1}-1)$. Using these we get

$$\delta_1 = c + \frac{1}{\lambda} - \frac{n-k}{k\lambda}(H_n - H_{n-k})$$
$$\approx c + \frac{1}{\lambda} + \frac{1-\alpha}{\alpha\lambda}\log(1-\alpha), \quad (17)$$

since for large n, we have $H_i \approx \log(i) + \gamma$. Also,

$$\delta_2 = \frac{2n-k}{2\mu k}(\mu E[X_{k:n}]+1) \quad (18)$$

$$= \frac{2n-k}{2\mu k}\left(\mu\left(c + \frac{H_n - H_{n-k}}{\lambda}\right) + 1\right) \quad (19)$$

$$\approx \frac{(2-\alpha)c}{2\alpha} + \frac{\alpha-2}{2\alpha\lambda}\log(1-\alpha) + \frac{2-\alpha}{2\alpha\mu}. \quad (20)$$

Next, we note that we have

$$\lim_{n\to\infty}\frac{\mu Var[X_{k:n}]}{2(\mu E[X_{k:n}]+1)} = 0. \quad (21)$$

We see this from the expected values of order statistics,

$$\frac{\mu Var[X_{k:n}]}{2(\mu E[X_{k:n}]+1)} = \frac{\mu(G_n - G_{n-k})}{2(\mu\lambda^2 c + \mu\lambda(H_n - H_{n-k}) + \lambda^2)}. \quad (22)$$

We know that the sequence G_n converges to $\pi^2/6$. As n increases $G_{n-k} = G_{(1-\alpha)n}$ also goes to the same value making the numerator 0. Thus, as n tends to ∞ (21) is achieved. Similarly,

$$\delta_3 = \frac{1}{2(\mu^2 E[X_{k:n}]+\mu)} \approx \frac{1}{2}\left(\mu^2 c - \frac{\mu^2\log(1-\alpha)}{\lambda} + \mu\right)^{-1}. \quad (23)$$

Summing δ_1, δ_2, and δ_3 yields the expression in (15). $\blacksquare$

Note that the age expression in (15) when n is large is a function of the ratio $\alpha = k/n$ only implying that the age converges to a constant even when the packets arrive exogenously, similar to [30] where the packets are generated at will at the source.

Although there is no explicit closed form solution for the optimal α, denoted as α^*, which minimizes (15), we can calculate it numerically. For instance, when $(\lambda, c) = (1, 1)$ and Poisson arrival rate is $\mu = 1$, age minimizing α is $\alpha^* = 0.837$. This optimal value is higher than that of the original case in which the source itself generates the packets at will, which is $\alpha^* = 0.732$ found in [30]. This is because with exogenous arrivals at the source node, update interarrival time at the end nodes is larger. In this regard, we see that when the Poisson arrival rate is increased α^* decreases. This is expected because when the arrival rate is high (i.e., update packets arrive frequently), the source node prefers to wait for the *freshest* one instead of sending the current update to more and more end users. Similarly, when the arrival rate is low (i.e., update packets arrive infrequently), α^* is higher because in this case source knows that interarrival time is higher so that before it waits for the next packet it wants to update as many end nodes as it can (see Fig. 6(a)). We also note that as we take $\mu \to \infty$ in (15), we get

$$\Delta_{(k)} \approx \frac{c}{\alpha} + \frac{c}{2} + \frac{1}{\lambda} - \frac{1}{2\lambda}\log(1-\alpha), \quad (24)$$

which is the age expression in [30]. Thus, age expression under exogenous Poisson arrivals with rate μ converges to the case in which source generates the packets itself as μ tends to ∞.

Finally, we note an interesting aspect of the problem with exogenous arrivals: It is shown in [30] that when the service time random variable X is exponential (i.e., shift variable is zero), the average age is minimized when $k = 1$. However, this is not the case when updates arrive exogenously. We observe in Fig. 6(b) that the age minimizing k value can be greater than 1, and it depends on the update arrival rate, μ. As μ increases, the optimal
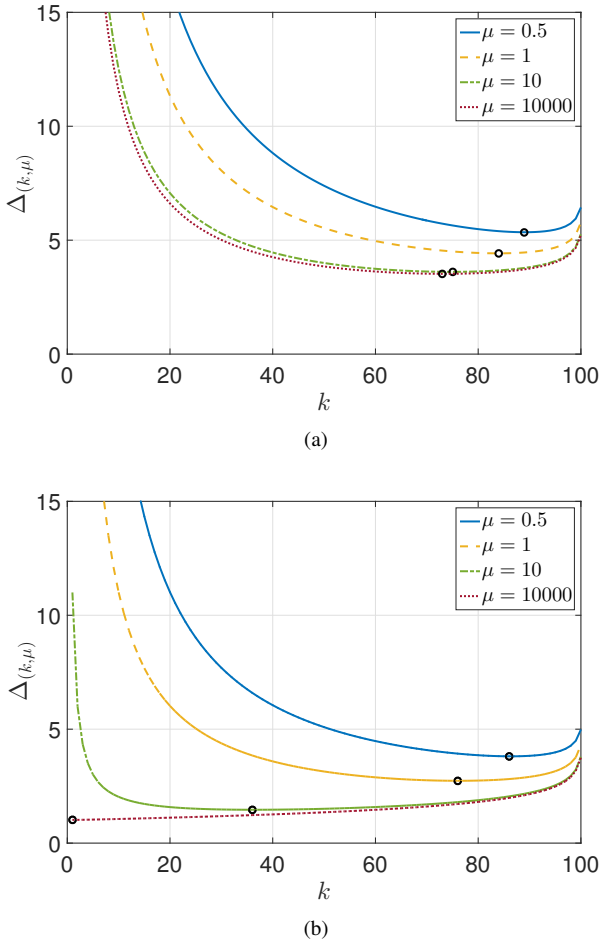
(a)

(b)

Fig. 6. $\Delta_{(k,\mu)}$ as a function of stopping threshold k for several μ values with $\lambda = 1$. $\circ$ marks the minimized $\Delta_{(k,\mu)}$: (a) When $c = 1$ and (b) when $c = 0$.

k decreases and approaches 1. The reason for this is that the random waiting time with exogenous arrivals introduces a random shift to the exponential distribution of the service time.

IV. TWO-HOP NETWORK

Using the building block problem solved in the previous section, we are now ready to solve the two-hop problem ($L = 2$) described in Section II as a preliminary step towards solving the most general case for arbitrary L. In the two-hop system, middle nodes cannot generate updates rather they receive them from the source node. Thus, each middle node and its n children nodes in the second hop can be modeled as in Section III. Since the source node sends updates to the first k_1 of its nodes, a middle node receives a certain update packet with probability $p_1 = k_1/n$. If a middle node receives update j and the next one it receives is update $j + M_1$, as in the building block problem, M_1 is a geometrically distributed random variable with parameter p_1.

Since the source generates a new update at will once the current update is delivered to k_1 middle nodes, the interarrival time between updates that start service in the first hop is $Y_1 = X_{k_1:n}$. Let random variable S_1 denote the interarrival time at the receiver nodes in the first hop. In other words, let S_1 denote the

time between updates that leave service in the first hop. Since each update cycle takes $X_{k_1:n}$ units of time, successful interarrival time at the middle nodes can be written as

$$S_1 = \sum_{i=1}^{M_1} (Y_1)_i = \sum_{i=1}^{M_1} (X_{k_1:n})_i, \qquad (25)$$

where the mean interarrival time is $E[S_1] = E[M_1]E[X_{k_1:n}]$. Note that we have $Z_1 = 0$ for the source node since it generates the updates as soon as the previous one is completed (transmitted to k_1 nodes). The receiver nodes in the first hop immediately relay the update that they have received. Therefore, the interarrival time between the successfully received updates at the receiver nodes in the first hop, S_1, is equal to update interarrival time at the transmitter nodes in the second hop, R_2.

After receiving the updates with interarrival time $R_2 = S_1$, middle nodes transmit each update until it is delivered to k_2 of their children nodes. Similar to the first hop, when a middle node transmits an update, an end node receives the update with probability $p_2 = k_2/n$. Geometrically distributed M_2 with parameter p_2 denotes the number of cycles between successive updates to an end node. The interarrival time between updates that depart from a middle node and start service in the second hop is $Y_2 = \tilde{X}_{k_2:n} + Z_2$. Let random variable S_2 denote the interarrival time between updates that successfully arrive at the receiver nodes in the second hop. Then,

$$S_2 = \sum_{i=1}^{M_2} (Y_2)_i = \sum_{i=1}^{M_2} (\tilde{X}_{k_2:n} + Z_2)_i. \qquad (26)$$

Note that in this model a successful update reaches an end node without being preempted in each of the hops. Thus, the service time of a successful update denoted by $\bar{X}$ is the sum of link delays in each hop and corresponds to the total time spent in the system by that update. Then, the total service time of a successful update delivered to some node i through middle node j is $E[\bar{X}] = E[X_j | j \in \mathcal{K}] + E[\tilde{X}_i | i \in \mathcal{K}_j]$. Here the set $\mathcal{K}$ is the set of first k_1 middle nodes that receive the update and the set $\mathcal{K}_j$ defined for each j in $\mathcal{K}$ is the set of first k_2 end nodes that receive the update. Thus, for an update to reach an end node that end node has to be among the earliest k_2 children nodes of its middle node and the corresponding middle node has to be one of the earliest k_1 middle nodes. Now, by using (10), the average age of an end node for a two-hop system is given in the following theorem.

Theorem 2: For a two-hop system with the earliest k_1, k_2 stopping scheme, the average age at an individual end node is

$$\Delta_{(k_1,k_2)} = \frac{1}{k_1} \sum_{i=1}^{k_1} E[X_{i:n}] + \frac{1}{k_2} \sum_{i=1}^{k_2} E[\tilde{X}_{i:n}]$$
$$+ \frac{2n - k_2}{2k_2} (E[\tilde{X}_{k_2:n}] + E[Z_2])$$
$$+ \frac{Var[\tilde{X}_{k_2:n} + Z_2]}{2(E[\tilde{X}_{k_2:n}] + E[Z_2])}. \qquad (27)$$

Proof: This theorem follows from Theorem 1 upon observing that the second hop is the same as the building block problem. However, successful updates in the two-hop setting spend time

in both hops to reach the end nodes. Noting also that successful updates do not wait in the system till they reach the end nodes, we find

$$E[\bar{X}] = \frac{1}{k_1} \sum_{i=1}^{k_1} E[X_{i:n}] + \frac{1}{k_2} \sum_{i=1}^{k_2} E[\tilde{X}_{i:n}]. \qquad (28)$$

Using (28) in (10) yields the theorem. ∎

This theorem is valid for any distribution for X and $\tilde{X}$. Random variable Z_2 denotes the residual interarrival time before the next update arrives to the middle node. When we no longer have exponential interarrival times, it is not easy to determine the first and second order statistics of Z_2. However, we can upper bound the average age of our model $\Delta_{(k_1,k_2)}$ with the average age under exponential interarrivals to the middle nodes $\Delta'_{(k_1,k_2)}$ using the following lemma.

Lemma 1: For a two-hop system, let $\Delta'_{(k_1,k_2)}$ denote the average age of an end node under exponential interarrivals to middle nodes with mean $E[R_2]$. Then, $\Delta_{(k_1,k_2)} \leq \Delta'_{(k_1,k_2)}$.

The proof of Lemma 1 follows from the DMRL (decreasing mean residual life) [31] property of interarrival times and NBUE (new better than used in expectation) [31] property of service times and is provided in Section VII. Note that as long as the aforementioned conditions on the interarrival and service time distributions are met, this lemma also applies to the case with $L > 2$ hops. This generalization is made in Section VII. Since $\Delta_{(k_1,k_2)} \leq \Delta'_{(k_1,k_2)}$, in order to prove that the average age in the two-hop system is bounded by a constant, all we need to show is that $\Delta'_{(k_1,k_2)}$ is upper bounded by a constant as the network grows, i.e., n increases.

Corollary 3: For a two-hop system, assuming exponential interarrivals to the middle nodes with $E[R_2]$, the average age at an end node under the earliest k_1, k_2 stopping scheme is

$$\begin{aligned} \Delta'_{(k_1,k_2)} =& \frac{1}{k_1} \sum_{i=1}^{k_1} E[X_{i:n}] + \frac{1}{k_2} \sum_{i=1}^{k_2} E[\tilde{X}_{i:n}] \\ &+ \frac{2n - k_2}{2k_2} E[\tilde{X}_{k_2:n}] + \frac{2n^2 - nk_2}{2k_1 k_2} E[X_{k_1:n}] \\ &+ \frac{k_1 Var[\tilde{X}_{k_2:n}]}{2(k_1 E[\tilde{X}_{k_2:n}] + nE[X_{k_1:n}])} \\ &+ \frac{n^2 E[X_{k_1:n}]^2}{2k_1(k_1 E[\tilde{X}_{k_2:n}] + nE[X_{k_1:n}])}. \end{aligned} \qquad (29)$$

Proof: When the interarrival times to the middle nodes, R_2 are exponential, then $Z_2 = R_2$ due to the memoryless property of the exponential distribution. In addition, we know that $R_2 = S_1$. Therefore, Z_2 is exponential with mean $E[Z_2] = E[S_1] = E[M_1]E[X_{k_1:n}]$. Then, $Var[Z_2] = E[S_1]^2 = E[M_1]^2 E[X_{k_1:n}]^2$ where M_1 is geometrically distributed with $p_1 = k_1/n$. Combining these and noting that Z_2 and $\tilde{X}_{k_2:n}$ are independent yields the result. ∎

Corollary 4: For a two-hop system, assuming n is large and $n > k_1$ and $n > k_2$ and letting $k_1 = \alpha_1 n$ and $k_2 = \alpha_2 n$, under exponential interarrival assumption to the middle nodes with mean $E[R_2]$, and shifted exponential service times X with (λ, c) and $\tilde{X}$ with $(\tilde{\lambda}, \tilde{c})$, the average age for the earliest k_1, k_2

scheme can be approximated as

$$\begin{aligned} \Delta'_{(k_1,k_2)} \approx& \frac{1}{\lambda} + \frac{1}{\tilde{\lambda}} + \frac{\tilde{c}}{\alpha_2} + \frac{\tilde{c}}{2} - \frac{1}{2\tilde{\lambda}} \log(1 - \alpha_2) \\ &+ \frac{2 - \alpha_2 + 2\alpha_1\alpha_2}{2\alpha_1\alpha_2} c + \frac{\tilde{\lambda}K_1^2}{2\alpha_1\lambda[\lambda\alpha_1 K_2 + \tilde{\lambda}K_1]} \\ &+ \frac{3\alpha_2 - 2\alpha_1\alpha_2 - 2}{2\alpha_1\alpha_2\lambda} \log(1 - \alpha_1), \qquad (30) \end{aligned}$$

where $K_1 = (\lambda c - \log(1 - \alpha_1))$ and $K_2 = (\tilde{\lambda}\tilde{c} - \log(1 - \alpha_2))$.

The proof of Corollary 4 is similar to that of Corollary 2. With Corollary 4 we have showed that $\Delta'_{(k_1,k_2)}$ derived in Corollary 3 is independent of n for large n. Since it upper bounds our age expression $\Delta_{(k_1,k_2)}$, we conclude that age under the earliest k_1, k_2 stopping scheme for a two-hop multicast network is also independent of n for large n and is bounded by a constant as the number of end nodes increases.

V. EXTENSION TO L HOPS

In this section, we extend our two-hop age results in (27), (29), and (30) to L hops. Considering an L-hop network, we have a single source node, n first hop receiver nodes, n^2 second hop receiver nodes and extending in this manner, n^L end (L hop) nodes. The network model for $L = 2$ is shown in Fig. 1 and it is generalized such that each of n nodes of the first hop is tied to n further nodes making n^2 second hop nodes and similarly, each of these n^2 second hop nodes is further connected to n nodes forming n^3 end nodes, and so on, for L hops.

Remember that in the two-hop model, we denote the first hop link delays as X, and the second hop link delays as $\tilde{X}$. In this section, in order to accommodate general L hops, we change our notation so that the first hop link delay is now denoted as $X^{(1)}$, the second hop link delay is now denoted as $X^{(2)}$, the ℓth hop is denoted as $X^{(\ell)}$, and the last hop link delay is denoted as $X^{(L)}$. For each hop ℓ, we utilize the earliest k_ℓ transmission policy such that once hop $\ell - 1$ receiver nodes receive an update they begin to act as hop ℓ transmitter nodes and relay the update they have received to k_ℓ of their n children nodes (see Fig. 2). After the packet transmission to k_ℓ children nodes are completed, hop ℓ transmitter nodes start waiting for the next update. Here, random variable Z_ℓ denotes this waiting time upon the completion of an update until the next one arrives. Thus, the interarrival time between two consecutive updates that depart from the transmitter and start service in hop ℓ is $Y_\ell = X_{k_\ell:n}^{(\ell)} + Z_\ell$.

Overall this L-hop network implements the earliest $\{k_\ell\}_{\ell=1}^L$ transmission scheme. At each hop, we have random variable M_ℓ which is geometrically distributed with parameter $p_\ell = k_\ell/n$ that represents the number of update cycles between two successive updates that arrive at a receiver in hop ℓ. The interarrival time between two consecutive updates that leave service in hop ℓ without being preempted can be written as

$$S_\ell = \sum_{i=1}^{M_\ell} (Y_\ell)_i = \sum_{i=1}^{M_\ell} (X_{k_\ell:n}^{(\ell)} + Z_\ell). \qquad (31)$$

A receiver node in hop ℓ immediately transmits an update it receives to its children nodes in hop $\ell + 1$. Therefore, the interarrival time between two consecutive successful updates that leave

service in hop ℓ, S_ℓ, is equal to the interarrival time between two consecutive updates that arrive in hop $\ell + 1$, $R_{\ell+1}$. We have $R_{\ell+1} = S_\ell$.

The last hop in this L-hop network can be seen as an application of our building block problem. In the building block problem, there is a source node and an end node. When this is applied to the last hop of the L-hop network, the source node is an arbitrary receiver node in hop $L - 1$, and the end node is an arbitrary receiver node in hop L. Interarrival time of updates that arrive exogenously at the source node in the building block problem, R, is now $R = S_{L-1}$, and the interarrival time between successful updates that arrive at the end node in the building block problem, S, is now $S = S_L$. In addition, we know that $S_L = \sum_{i=1}^{M_L}(Y_L)_i$. Finally, the service time in the building block problem, $\bar{X}$, is now $\bar{X} = \sum_{\ell=1}^{L} \bar{X}_\ell$, where $\bar{X}_\ell$ is the service time at each hop. We know that for an update packet to reach the end nodes, it has to be among the earliest $k_1, \cdots, k_L$ nodes in all hops. Thus, $\sum_{\ell=1}^{L} \bar{X}_\ell$ term captures the expected time spent in the system for a successful update without being preempted until it reaches one of the end nodes. The average age of an end node for an L-hop system is given in the following theorem.

Theorem 3: For the general L-hop network with the earliest $\{k_\ell\}_{\ell=1}^{L}$ stopping scheme, the average age at an individual end node at hop L is

$$
\Delta_{\{k_\ell\}_{\ell=1}^{L}} = \sum_{\ell=1}^{L} \frac{1}{k_\ell} \sum_{i=1}^{k_\ell} E[X_{i:n}^{(\ell)}]
$$
$$
+ \frac{2n - k_L}{2k_L}(E[X_{k_L:n}^{(L)}] + E[Z_L])
$$
$$
+ \frac{Var[X_{k_L:n}^{(L)} + Z_L]}{2(E[X_{k_L:n}^{(L)}] + E[Z_L])}. \quad (32)
$$

Proof: Since the last hop of the general L-hop network can be seen as a building block, we can apply (10) to the setting here by inserting $Y = Y_L$ and $\bar{X} = \sum_{\ell=1}^{L} \bar{X}_\ell$. Then, (32) follows after similar calculations as in Theorems 1 and 2. ∎

Similar to Corollary 3, the age of the last hop can be upper bounded by assuming that the interarrival times at each hop are exponentially distributed.

Corollary 5: For the general L-hop network, assuming exponential interarrivals to each hop with mean $E[R_\ell]$, average age at an end node under $\{k_\ell\}_{\ell=1}^{L}$ stopping scheme is given by

$$
\Delta'_{\{k_\ell\}_{\ell=1}^{L}} = \sum_{\ell=1}^{L} \frac{1}{k_\ell} \sum_{i=1}^{k_\ell} E[X_{i:n}^{(\ell)}]
$$
$$
+ \frac{2n - k_L}{2k_L} \sum_{\ell=1}^{L} E[X_{k_\ell:n}^{(\ell)}] \prod_{i=\ell}^{L-1} E[M_i]
$$
$$
+ \frac{Var[X_{k_L:n}^{(L)}] + \left(\sum_{\ell=1}^{L-1} E[X_{k_\ell:n}^{(\ell)}] \prod_{i=\ell}^{L-1} E[M_i]\right)^2}{2(\sum_{\ell=1}^{L} E[X_{k_\ell:n}^{(\ell)}] \prod_{i=\ell}^{L-1} E[M_i])}.
$$
$$
\quad (33)
$$

Proof: Using Lemma 1, (32) can be upper bounded with exponential interarrivals to the nodes at hop $L - 1$, where

mean interarrival time is $E[R_L] = E[S_{L-1}]$. When the interarrivals are exponentially distributed, we have $Z_L = S_{L-1}$ and therefore, $E[Z_L] = E[S_{L-1}]$. In addition, with exponential interarrivals, Z_L is independent of $X_{k_L:n}$. Then, we have $Var[X_{k_L:n}^{(L)} + Z_L] = Var[X_{k_L:n}^{(L)}] + Var[Z_L]$, where $Var[Z_L] = E[S_{L-1}]^2$. Now, the upper bound for (32) can be written as

$$
\Delta'_{\{k_\ell\}_{\ell=1}^{L}} = \sum_{\ell=1}^{L} \frac{1}{k_\ell} \sum_{i=1}^{k_\ell} E[X_{i:n}^{(\ell)}]
$$
$$
+ \frac{2n - k_L}{2k_L}(E[X_{k_L:n}^{(L)}] + E[S_{L-1}])
$$
$$
+ \frac{Var[X_{k_L:n}^{(L)}] + E[S_{L-1}]^2}{2(E[X_{k_L:n}^{(L)}] + E[S_{L-1}])}. \quad (34)
$$

Now, let us calculate $E[S_{L-1}]$. We can write S_{L-1} as

$$
S_{L-1} = \sum_{i=1}^{M_{L-1}}(Y_{L-1})_i = \sum_{i=1}^{M_{L-1}}(X_{k_{L-1}:n}^{(L-1)} + Z_{L-1})_i, \quad (35)
$$

where Y_{L-1} is the interarrival time between updates that start service in hop $L - 1$. Then,

$$
E[Z_L] = E[S_{L-1}]
$$
$$
= E[M_{L-1}]E[X_{k_{L-1}:n}^{(L-1)}] + E[M_{L-1}]E[Z_{L-1}]. \quad (36)
$$

Similarly, $E[Z_{L-1}]$ can be written in terms of the variables in hop $L - 2$. We continue with this recursive calculation until the second hop, where we have $E[Z_2] = E[M_1]E[X_{k_1:n}^{(1)}]$. As a result, we have

$$
E[S_{L-1}] = E[M_{L-1}]E[X_{k_{L-1}:n}^{(L-1)}] + \cdots
$$
$$
+ E[M_{L-1}]E[M_{L-2}] \cdots E[M_2]E[M_1]E[X_{k_1:n}^{(1)}]
$$
$$
= \sum_{\ell=1}^{L-1} E[X_{k_\ell:n}^{(\ell)}] \prod_{i=\ell}^{L-1} E[M_i]. \quad (37)
$$

Adding $E[X_{k_{L-1}:n}^{L-1}]$ to (37), we have

$$
E[X_{k_{L-1}:n}^{L-1}] + E[S_{L-1}] = \sum_{\ell=1}^{L} E[X_{k_\ell:n}^{(\ell)}] \prod_{i=\ell}^{L-1} E[M_i]. \quad (38)
$$

Finally, by inserting (37) and (38) into (34), we obtain (33). ∎

Corollary 5 introduces an upper bound for average age of an update that is generated at the source node and delivered to any one of the end nodes in hop L. Note that the total number of nodes in the L-hop network is $\sum_{\ell=0}^{L} n^l$. The result of Corollary 5 holds for any n. Next, we characterize the effect of increasing n on the scaling of average age.

Corollary 6: Assume n is large and $n > k_\ell$ and let $k_\ell = \alpha_\ell n$ for $\ell = 1, \cdots, L$. Under exponential interarrival assumption to each hop with mean $E[R_\ell]$, and shifted exponential service times $X^{(\ell)}$ with (λ_ℓ, c_ℓ), the average age for the earliest $\{k_\ell\}_{\ell=1}^{L}$ scheme can be approximated as

$$
\Delta'_{\{k_\ell\}_{\ell=1}^{L}} \approx \sum_{\ell=1}^{L} \left(c_\ell + \frac{1}{\lambda_\ell} + \frac{1 - \alpha_\ell}{\alpha_\ell \lambda_\ell} \log(1 - \alpha_\ell)\right)
$$

number of hops L	α_1^*	α_2^*	α_3^*	α_4^*
$L=1$	0.732	-	-	-
$L=2$	0.615	0.921	-	-
$L=3$	0.626	0.832	0.965	-
$L=4$	0.635	0.837	0.901	0.935

Table 1. Optimal α_i values for L-hop network when all link delays are shifted exponentials with $(1,1)$.

$$+ \frac{2-\alpha_L}{2\alpha_L} \sum_{\ell=1}^{L} \left[\left(c_\ell - \frac{1}{\lambda_\ell} \log(1-\alpha_\ell) \right) \prod_{i=\ell}^{L-1} \frac{1}{\alpha_i} \right]$$

$$+ \frac{\left(\sum_{\ell=1}^{L-1} \left[\left(c_\ell - \frac{1}{\lambda_\ell} \log(1-\alpha_\ell) \right) \prod_{i=\ell}^{L-1} \frac{1}{\alpha_i} \right] \right)^2}{2 \sum_{\ell=1}^{L} \left[\left(c_\ell - \frac{1}{\lambda_\ell} \log(1-\alpha_\ell) \right) \prod_{i=\ell}^{L-1} \frac{1}{\alpha_i} \right]}. \tag{39}$$

The proof of this corollary is similar to that of Corollaries 2 and 4. We use the first and second moments of order statistics of shifted exponential random variables and make necessary approximations for large n.

In conclusion, under the assumption that middle nodes receive update packets with exponential interarrivals with means equal to $E[R_\ell]$, the average age attained at the end nodes depends on n only through ratios α_ℓ. Thus, this result together with Lemma 1 imply that the average age of an end node in an L-hop multicast network implementing the earliest $\{k_\ell\}$ transmission scheme is bounded by a constant as n increases.

VI. NUMERICAL RESULTS

In this section, we provide simple numerical results. In order to optimize the age of information at the end nodes, we need to select appropriate k_i values, i.e., optimum ratios α_i^*, at each hop. From [30], we know that in the single-hop multicast network α^* is 0.732 when link delays are shifted exponential with parameters $(\lambda, c) = (1,1)$. For the two-hop network with $(\lambda, c) = (\tilde{\lambda}, \tilde{c}) = (1,1)$, we obtain $\alpha_1^* = 0.615$ and $\alpha_2^* = 0.921$. This shows that when all link delays are statistically identical, to achieve a good age performance, we need to be more aggressive in the second hop than the first hop (see [32]).

When we add a third hop, we see that the optimal threshold results follow similar trends. Let us denote the link delay parameters of the third hop as $(\tilde{\tilde{\lambda}}, \tilde{\tilde{c}})$ for tractability. When all link delays are statistically identical, i.e., shifted exponentials with parameters $(1,1)$, we have $\alpha_1^* = 0.626$, $\alpha_2^* = 0.832$ and $\alpha_3^* = 0.965$. In a similar fashion, when we add a fourth hop, we have $\alpha_1^* = 0.635$, $\alpha_2^* = 0.837$, $\alpha_3^* = 0.901$ and $\alpha_4^* = 0.935$. Thus, the observation we have made for two-hop multicast networks, i.e., that we need to be more aggressive in the further hops to achieve a lower average age holds true for four-hop multicast networks as well. These results are summarized in Table 1.

Returning to the two-hop network, we observe that α_2 is responsive to the changes in the parameters of the first hop. This is intuitive because as k_1 varies, the mean interarrival time for the second hop changes. In Figs. 7(a) and 7(b), we plot the age as a function of k_2 (equivalently α_2) for a fixed set of second hop parameters and for a fixed set of first hop parameters, respectively. As shown in Fig. 7(a), for the same $(\tilde{\lambda}, \tilde{c})$ pair, when the mean
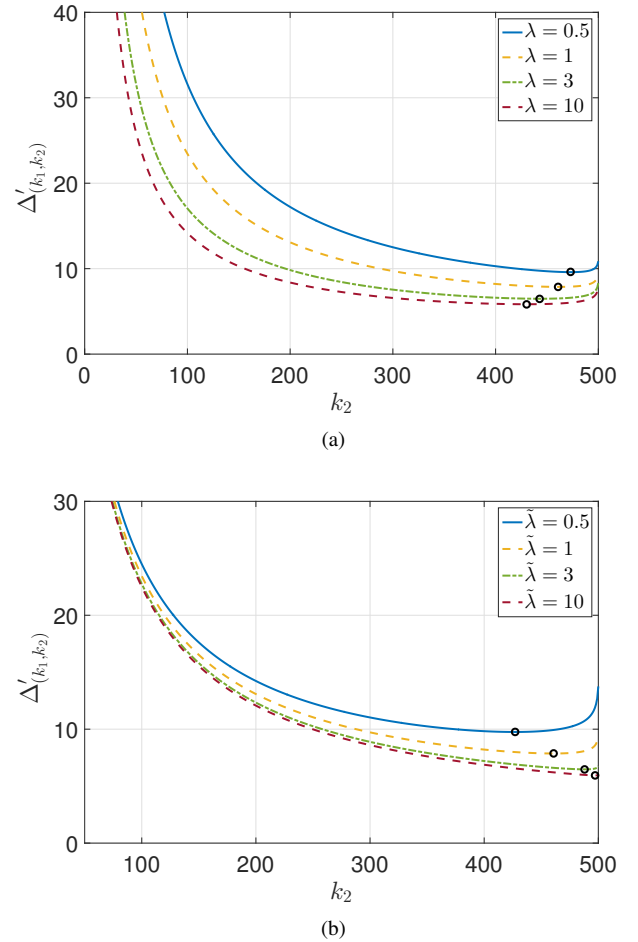

(a)

(b)

Fig. 7. $\Delta'_{(k_1,k_2)}$ as a function of k_2 for $n = 500$. $\circ$ marks the minimized average age $\Delta'_{(k_1,k_2)}$: (a) $c = 1$ and $(\tilde{\lambda}, \tilde{c}) = (1,1)$ for varying λ and (b) $\tilde{c} = 1$ and $(\lambda, c) = (1,1)$ for varying $\tilde{\lambda}$.

interarrival time gets lower by increasing λ, α_2^* gets lower as well. Knowing that the next update arrival is not far away, middle nodes tend to wait for the next one instead of sending the current packet to more and more end users when the arrivals are frequent. This is exactly what we have observed in the building block problem with exogenous update arrivals.

In Fig. 7(b) we observe that as the service rate of the second hop, $\tilde{\lambda}$, increases for a given λ, α_2^* increases as well. This is intuitive because when λ is fixed, the interarrival time to the middle nodes stays constant as opposed to their increasing service rate. Thus, they relay their current packet to more end nodes.

In Figs. 8(a) and 8(b), we plot the age as a function of k_1 (equivalently α_1) when the second hop parameters are fixed and when the first hop parameters are fixed, respectively. In Fig. 8(a), α_1^* shows a similar trend as in α_2^* in Fig. 7(b) but α_1^* experiences bigger changes as λ increases. Fig. 8(b) shows the response of α_1^* to changes in $\tilde{\lambda}$ when λ is fixed. Here, we observe rather minor reaction from α_1^* when the service performance of the second stage varies.

In Fig. 9 we repeat the numerical analysis in Fig. 7 when the link delays are exponential, i.e., shift parameters $c = 0$ and $\tilde{c} = 0$. We observe similar trends for k_2^* as in Fig. 7. The only difference is now that the link delays are all exponential, we get
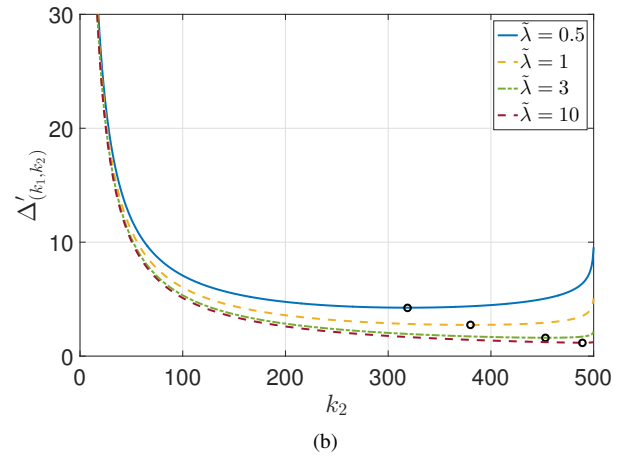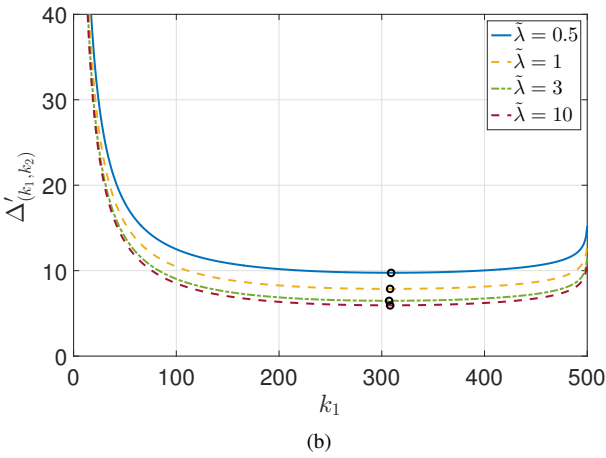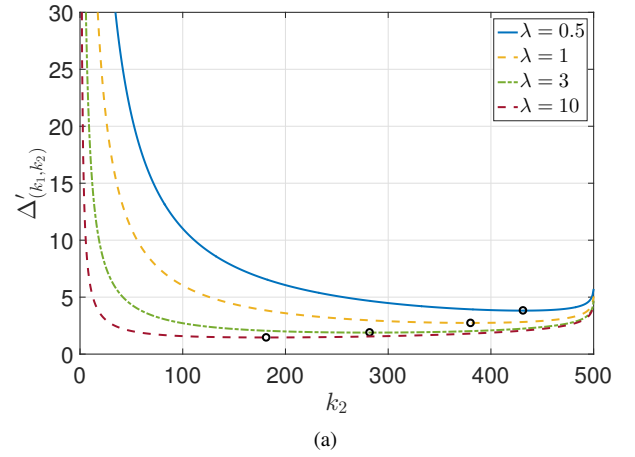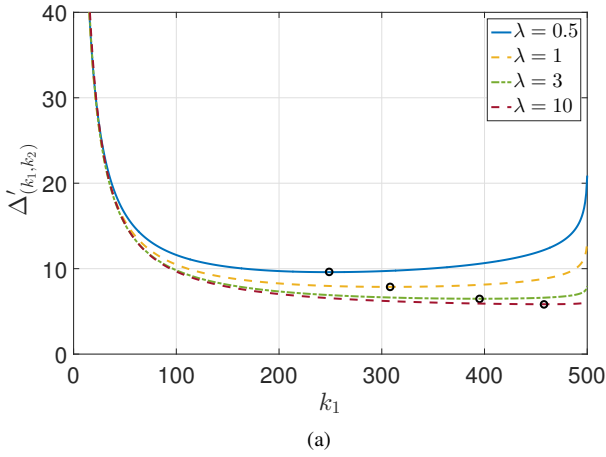
Fig. 8. $\Delta'_{(k_1,k_2)}$ as a function of k_1 for $n = 500$. $\circ$ marks the minimized average age $\Delta'_{(k_1,k_2)}$: (a) $c = 1$ and $(\tilde{\lambda}, \tilde{c}) = (1, 1)$ for varying λ and (b) $\tilde{c} = 1$ and $(\lambda, c) = (1, 1)$ for varying $\tilde{\lambda}$.

Fig. 9. $\Delta'_{(k_1,k_2)}$ as a function of k_2 for $n = 500$ when link delays are exponential. $\circ$ marks the minimized average age $\Delta'_{(k_1,k_2)}$: (a) $c = 0$ and $(\tilde{\lambda}, \tilde{c}) = (1, 0)$ for varying λ, (b) $\tilde{c} = 0$ and $(\lambda, c) = (1, 0)$ for varying $\tilde{\lambda}$.

$k_1^* = 1$ for all cases. This is also observed in [30] and in the building block problem when μ tends to ∞ and is because of the memoryless property of the exponential distribution.

In Fig. 10 we analyze the effects of the link delay parameters λ and $\tilde{\lambda}$ on $\Delta'_{(k_1,k_2)}$. In both cases we see that larger service rates (large λ and $\tilde{\lambda}$), i.e., lower link delays, lead to smaller average age at the end nodes.

In Fig. 11 we plot the three-hop version of Fig. 7. As shown in Fig. 11, when the mean interarrival time decreases at the third hop through a service rate increase in either one of the first two hops α_3^* gets smaller. One other observation from Fig. 11 is the fact that although we change λ values in Fig. 11(a) and $\tilde{\lambda}$ values in Fig. 11(b), their effects on the third hop are quite similar. In either case we observe a similar trend, i.e., k_3 value decreases as λ or $\tilde{\lambda}$ increases.

VII. APPENDIX: PROOF OF LEMMA 1

In this section, we prove Lemma 1, first, for the two-hop scenario, then we extend the proof to the L-hop case. We use [9, Thm. 2] which, for a G/G/1/1 system, requires interarrival times to have DMRL property and service times to have NBUE property. Note that the updates that depart the service at the first

hop, S_1 are considered to be the arrivals at the second hop, R_2. Then, between a transmitter and a receiver at the second hop, we have interarrivals $R_2 = \sum_{i=1}^{M_1}(X_{k_1:n})_i$ and service times $\tilde{X}_{k_2:n}$. In order to employ [9, Thm. 2], we need to show that R_2 has DMRL property and $\tilde{X}_{k_2:n}$ has NBUE property. It is sufficient to show that both R_2 and $\tilde{X}_{k_2:n}$ have log-concave density, since log-concavity implies both DMRL and NBUE properties [31].

It is given in [33, Thm. 1.C.54] that the order statistics of i.i.d. log-concave random variables is log-concave as well. Since shifted exponential has a log-concave density, we conclude that the second hop service time $\tilde{X}_{k_2:n}$ has a log-concave density.

In order to show the log-concavity of R_2, we use the fact that a function is log-concave if and only if it is a Polya Frequency function of order 2 [31, Proposition 21.B.8], which is denoted by PF_2. We know that the geometric random variable M_1 and shifted exponential random variables $X_{k_1:n}$ are log-concave, and hence they have PF_2 densities. In addition, we know that M_1 is independent of $X_{k_1:n}$. Now, we can apply [34, Thm. 6] that states that if $X_{k_1:n}$ and M_1 have PF_k densities, and M_1 is independent of $X_{k_1:n}$, then $R_2 = \sum_{i=1}^{M_1}(X_{k_1:n})_i$ has a PF_k density as well. Since [34, Thm. 6] is stated for any k, it holds for $k = 2$, and we conclude that R_2 has a PF_2 density, which, in turn, means that R_2 is log-concave.
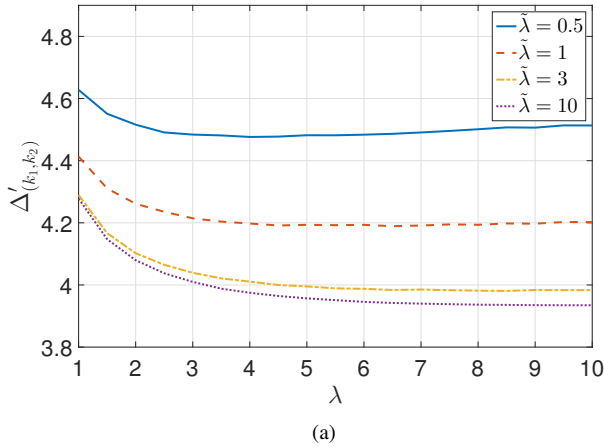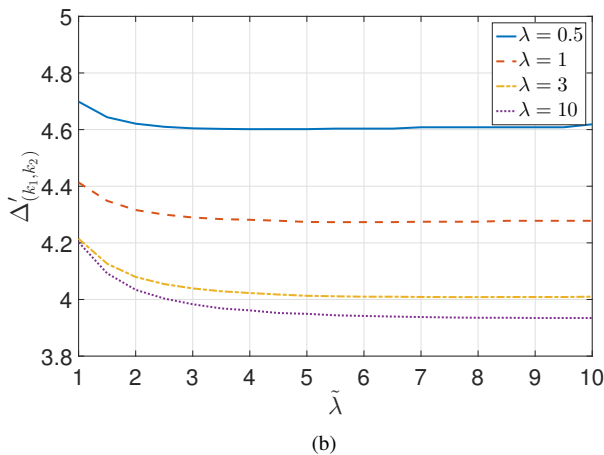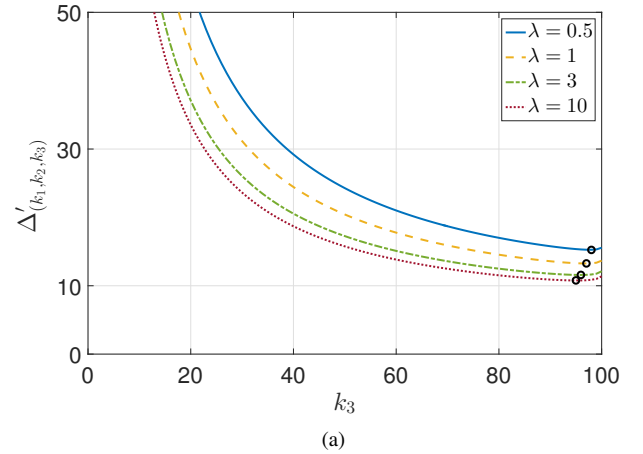
Fig. 10. $\Delta'_{(k_1,k_2)}$ as a function of link delay parameters λ and $\tilde{\lambda}$ for $n = 500$: (a) $\Delta'_{(k_1,k_2)}$ as a function of λ, (b) $\Delta'_{(k_1,k_2)}$ as a function of $\tilde{\lambda}$.

Fig. 11. $\Delta'_{(k_1,k_2,k_3)}$ as a function of k_3 for $n = 100$. $\circ$ marks the minimized average age $\Delta'_{(k_1,k_2,k_3)}$: (a) $c = 1$ and $(\tilde{\lambda}, \tilde{c}) = (\tilde{\tilde{\lambda}}, \tilde{\tilde{c}}) = (1,1)$ for varying λ and (b) $\tilde{c} = 1$ and $(\lambda, c) = (\tilde{\tilde{\lambda}}, \tilde{\tilde{c}}) = (1,1)$ for varying $\tilde{\lambda}$.
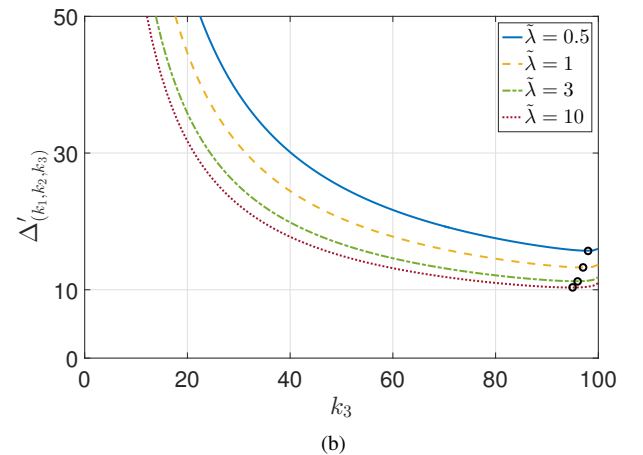
We have proved that interarrivals to the second hop are DMRL and service time of the second hop is NBUE. Hence, using [9, Thm. 2] average age of the second hop is upper bounded by a system where interarrivals are exponential with $E[R_2]$. In the third hop, service times are i.i.d. with the service times at the second hop. Therefore, service time of the third hop is NBUE as well. The interarrivals at the third hop are now $R_3 = \sum_{i=1}^{M_2} (\tilde{X}_{k_2:n} + Z_2)_i$. In order to calculate the upper bound in the second hop, we have already assumed that the interarrivals, R_2, are exponential which resulted in $Z_2 = R_2$ to be exponential as well. Then, R_3 can be shown to be log-concave using similar ideas above and for exponential Z_2. This reasoning holds for each hop in an L-hop system, which proves the applicability of Lemma 1 to an L-hop system as well.

VIII. CONCLUSIONS

We studied the age of information in a multihop multicast network with a single source updating n^L end nodes where L denotes the number of hops. We showed that when the earliest k_ℓ transmission policy is implemented at each hop ℓ, the age of information at the end nodes can be upper bounded by a constant that is independent of n. We explicitly characterized an upper bound for an L-hop multicast network, and then determined the

optimal stopping thresholds k_ℓ for arbitrary shifted exponential link delays. We note that even when the link delays are exponential we find $k_1^* = 1$ and $k_\ell^* > 1$ in hops $\ell > 1$. This is because even when there is no shift in service, the random waiting time under exogenous arrivals introduces a random shift to the exponential service distribution in hops $\ell > 1$.

REFERENCES

[1] A. Mainwaring et al, "Wireless sensor networks for habitat monitoring," in *Proc. WSNA*, Sept. 2002.
[2] P. Papadimitratos et al, "Vehicular communication systems: Enabling technologies, applications, and future outlook on intelligent transportation," *IEEE Commun. Mag.*, vol. 47, no. 11 pp. 84–95, Nov. 2009.
[3] S. K. Kaul, R. D. Yates, and M. Gruteser, "Real-time status: How often should one update?," in *Proc. IEEE INFOCOM*, Mar. 2012.
[4] M. Costa, M. Codreanu, and A. Ephremides, "On the age of information in status update systems with packet management," *IEEE Trans. Inf Theory*, vol. 62, no. 4 pp. 1897–1910, Apr. 2016.
[5] R. D. Yates and S. K. Kaul, "Real-time status updating: Multiple sources," in *Proc. IEEE ISIT*, July 2012.
[6] Y. Inoue, Analysis of the age of information with packet deadline and infinite buffer capacity, in *Proc. IEEE ISIT*, June 2018.
[7] V. Tripathi, R. Talak, and E. Modiano, "Age of information for discrete time queues," Jan. 2019. Available: arXiv:1901.10463.
[8] E. Najm, R. D. Yates, and E. Soljanin, "Status updates through M/G/1/1 queues with HARQ," in *Proc. IEEE ISIT*, June 2017.

[9] A. Soysal and S. Ulukus, "Age of information in G/G/1/1 systems," Nov. 2018. Available arXiv:1805.12586.

[10] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Age-optimal information updates in multihop networks," in *Proc. IEEE ISIT*, June 2017.

[11] A. M. Bedewy, Y. Sun, and N. B. Shroff, "The age of information in multihop networks," Dec. 2017. Available arXiv:1712.10061.

[12] R. Talak, S. Karaman, and E. Modiano, "Minimizing age-of-information in multi-hop wireless networks," in *Proc. Allerton*, Oct. 2017.

[13] Y. Hsu, E. Modiano, and L. Duan, "Scheduling algorithms for minimizing age of information in wireless broadcast networks with random arrivals," Dec. 2017. Available arXiv:1712.07419.

[14] Y. Hsu, "Age of information: Whittle index for scheduling stochastic arrivals," in *Proc. IEEE ISIT*, June 2018.

[15] R. D. Yates, "The age of information in networks: Moments, distributions, and sampling," June 2018. Available arXiv:1806.03487.

[16] Y. Inoue, H. Masuyama, T. Takine, and T. Tanaka, "A general formula for the stationary distribution of the age of information and its application to single-server queues," Apr. 2018. Available arXiv:1804.06139.

[17] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," *IEEE Trans. Inf Theory*, vol. 63, no. 11 pp. 7492–7508, Nov. 2017.

[18] P. Zou, O. Ozel, and S. Subramaniam, "Waiting before serving: A companion to packet management in status update systems," Jan. 2019. Available arXiv:1901.02873.

[19] A. Arafa and S. Ulukus, "Age minimization in energy harvesting communications: Energy-controlled delays," in *Proc. Asilomar Conference*, Oct. 2017.

[20] A. Arafa and S. Ulukus, "Age-minimal transmission in energy harvesting two-hop networks," in *Proc. IEEE GLOBECOM*, Dec. 2017.

[21] R. D. Yates, "Lazy is timely: Status updates by an energy harvesting source," in *Proc. IEEE ISIT*, June 2015.

[22] B. T. Bacinoglu, E. T. Ceran, and E. Uysal-Biyikoglu, "Age of information under energy replenishment constraints," in *Proc. UCSD ITA*, Feb. 2015.

[23] X. Wu, J. Yang, and J. Wu, "Optimal status update for age of information minimization with an energy harvesting source," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 1, pp. 193–204, Mar. 2018.

[24] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor, "Age-minimal online policies for energy harvesting sensors with incremental battery recharges," in *Proc. UCSD ITA*, Feb. 2018.

[25] A. Arafa, J. Yang, and S. Ulukus, "Age-minimal online policies for energy harvesting sensors with random battery recharges," in *Proc. IEEE ICC*, May 2018.

[26] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor, "Online timely status updates with erasures for energy harvesting sensors," in *Proc. Allerton*, Oct. 2018.

[27] A. Baknina and S. Ulukus, "Coded status updates in an energy harvesting erasure channel," in *Proc. CISS*, Mar. 2018.

[28] A. Baknina, O. Ozel, J. Yang, S. Ulukus, and A. Yener, "Sending information through status updates," in *Proc. IEEE ISIT*, June 2018.

[29] S. Ioannidis, A. Chaintreau, and L. Massoulie, "Optimal and scalable distribution of content updates over a mobile social network," in *Proc. IEEE INFOCOM*, Apr. 2009.

[30] J. Zhong, E. Soljanin, and R. D. Yates, "Status updates through multicast networks," in *Proc. Allerton*, Oct. 2017.

[31] A. W. Marshall and I. Olkin, *Life Distributions, Structure of Nonparametric, Semiparametric, and Parametric Families*. Springer, New York, 2007.

[32] B. Buyukates, A. Soysal, and S. Ulukus, "Age of information in two-hop multicast networks," in *Proc. Asilomar Conference*, Oct. 2018.

[33] M. Shaked and J. G. Shanthikumar, *Stochastic Orders and Their Applications*. Springer-Verlag New York, 2007.

[34] S. Karlin and F. Proschan, "Polya type distributions of convolutions," *The Annals of Mathematical Statistics*, vol. 31, no. 3, pp. 721–736, Sept. 1960.

Baturalp Buyukates received B.S. degree in Electrical and Electronics Engineering from Bilkent University, Turkey in 2016. He is currently pursuing the Ph.D. degree at the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, USA. His research interests include age of information, scaling laws and resource allocation for low latency wireless communication systems.

Alkan Soysal received the B.S. degree in Electrical and Electronics Engineering from the Middle East Technical University, Ankara, Turkey, in 2002 and the M.S. and Ph.D. degrees in Electrical Engineering from the University of Maryland, College Park, in 2006 and 2008, respectively. Since February 2008, he has been with Bahcesehir University, Istanbul, Turkey, first as an Assistant Professor then as an Associate Professor at the Department of Electrical and Electronics Engineering. Currently, he is a Visiting Associate Professor in the Department of Electrical and Computer Engineering at the University of Maryland at College Park. His research interests include wireless communication theory, information theory, and signal processing for wireless communications. His current research focuses on age of information, massive MIMO and energy efficiency in heterogeneous networks.

Sennur Ulukus is the Anthony Ephremides Professor in Information Sciences and Systems in the Department of Electrical and Computer Engineering at the University of Maryland at College Park, where she also holds a joint appointment with the Institute for Systems Research (ISR). Prior to joining UMD, she was a Senior Technical Staff Member at AT&T Labs-Research. She received her Ph.D. degree in Electrical and Computer Engineering from Wireless Information Network Laboratory (WINLAB), Rutgers University, and B.S. and M.S. degrees in Electrical and Electronics Engineering from Bilkent University. Her research interests are in information theory, wireless communications, machine learning, signal processing and networks, with recent focus on private information retrieval, age of information, distributed coded computation, energy harvesting communications, physical layer security, and wireless energy and information transfer.

She is a fellow of the IEEE, and a Distinguished Scholar-Teacher of the University of Maryland. She received the 2003 IEEE Marconi Prize Paper Award in Wireless Communications, the 2019 IEEE Communications Society Best Tutorial Paper Award, an 2005 NSF CAREER Award, the 2010-2011 ISR Outstanding Systems Engineering Faculty Award, and the 2012 ECE George Corcoran Outstanding Teaching Award. She is a Distinguished Lecturer of the IEEE Information Theory Society for 2018-2019. She is on the Editorial Board of the IEEE Transactions on Green Communications and Networking since 2016. She was an Editor for the IEEE Journal on Selected Areas in Communications-Series on Green Communications and Networking (2015-2016), IEEE Transactions on Information Theory (2007-2010), and IEEE Transactions on Communications (2003-2007). She was a Guest Editor for the IEEE Journal on Selected Areas in Communications (2015 and 2008), Journal of Communications and Networks (2012), and IEEE Transactions on Information Theory (2011). She is a TPC co-chair of 2019 ITW, 2017 IEEE ISIT, 2016 IEEE GLOBECOM, 2014 IEEE PIMRC, and 2011 IEEE CTW.