

Information Freshness in Cache Updating Systems

Melih Bastopcu, *Graduate Student Member, IEEE*, and Sennur Ulukus¹, *Fellow, IEEE*

Abstract—We consider a cache updating system with a source, a cache and a user. There are n files. The source keeps the freshest version of the files which are updated with known rates λ_i . The cache downloads and keeps the freshest version of the files from the source with rates c_i . The user gets updates from the cache with rates u_i . When the user gets an update, it either gets a fresh update from the cache or the file at the cache becomes outdated by a file update at the source in which case the user gets an outdated update. We find an analytical expression for the average freshness of the files at the user. Next, we generalize our setting to the case where there are multiple caches in between the source and the user, and find the average freshness at the user. We provide an alternating maximization based method to find the update rates for the cache(s), c_i , and for the user, u_i , to maximize the freshness of the files at the user. We observe that for a given set of update rates for the user (resp. for the cache), the optimal rate allocation policy for the cache (resp. for the user) is a *threshold policy*, where the optimal update rates for rapidly changing files at the source may be equal to zero. Finally, we consider a system where multiple users are connected to a single cache and find update rates for the cache and the users to maximize the total freshness over all users.

Index Terms—Age of information, information freshness, cache updating systems, multi-hop multi-user communication systems.

I. INTRODUCTION

WITH emerging technologies such as autonomous driving, augmented reality, social networking, high-frequency automated trading, and online gaming, time sensitive information has become ever more critical. Age of information has been proposed as a performance metric to quantify the *freshness* of information in communication networks. Age of information has been studied in the context of web crawling [1]–[4], social networks [5], queueing networks [6]–[16], caching systems [17]–[24], remote estimation [25]–[28], energy harvesting systems [29]–[45], fading wireless channels [46], [47], scheduling in networks [48]–[56], multi-hop multicast networks [57]–[60], lossless and lossy source coding [61]–[67], computation-intensive systems [68]–[73], vehicular, IoT and UAV systems [74]–[76], reinforcement learning [77]–[79], and so on.

Manuscript received April 18, 2020; revised August 21, 2020 and November 2, 2020; accepted November 3, 2020. Date of publication November 17, 2020; date of current version March 10, 2021. This work was supported by NSF under Grant CCF 17-13977 and Grant ECCS 18-07348. The associate editor coordinating the review of this article and approving it for publication was S. Dey. (Corresponding author: Sennur Ulukus.)

The authors are with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: bastopcu@umd.edu; ulukus@umd.edu).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TWC.2020.3037144>.

Digital Object Identifier 10.1109/TWC.2020.3037144

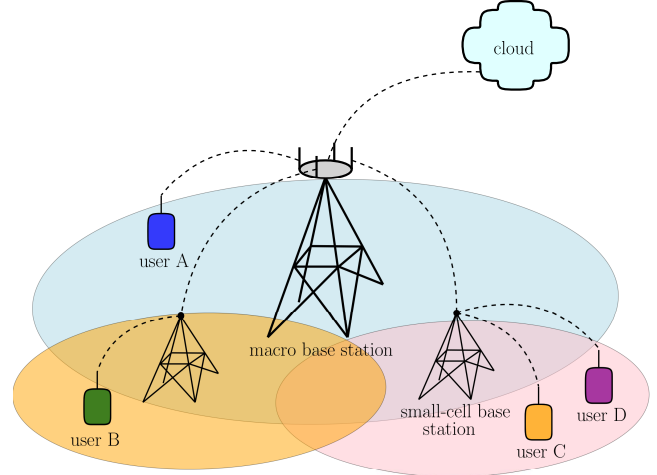


Fig. 1. A cache updating system consisting of a cloud (the source), a macro base station (the first cache), a small-cell base station (the second cache), and users. The files at the source are updated with known rates. The first cache always obtains fresh files from the source. However, depending on the file status at the first cache, the second cache may not be able to obtain a fresh file all the time; the same is true for the users as well. We consider end-to-end freshness at the users.

In this work, we consider a cache updating system that consists of a source, cache(s) and user(s). We start with the simplest system model with a source, a single cache and a single end-user in Section III (and shown in Fig. 2); generalize it to the case where there are multiple caches in between the source and the user in Section IV (and shown in Fig. 4); and further extend it to the case where there are multiple end-users in Section VII (and shown in Fig. 6). The models we study are abstractions of a real-life setting shown in Fig. 1. Specifically, the two-hop serial cache system considered in Section III is an abstraction for the communication system between the cloud, macro base station and user *A* in Fig. 1; the multi-hop serial cache system considered in Section IV is an abstraction for the communication system between the cloud, macro base station, small-cell base station and user *B* in Fig. 1 (for a three-hop system); and the multi-access caching system considered in Section VII is an abstraction for the communication system between the cloud, macro base station, small-cell base station and users *C* and *D* in Fig. 1. Example deployment scenarios for hierarchical caching systems connected wirelessly can be found in 5G-enabled vehicular networks where self-sustaining wirelessly connected caching stations are placed to enhance vehicular network capacity [80].

In all these system models, the source keeps the freshest version of all the files which are updated with known rates λ_i . The

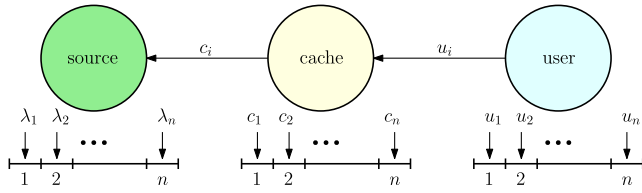


Fig. 2. A cache updating system which consists of a source, a cache and a user. The i th file at the source is updated with rate λ_i , the cache requests updates for the i th file from the source with rate c_i , and the user requests updates for the i th file from the cache with rate u_i .

cache downloads the files from the source and stores the latest downloaded versions of these files. When the cache downloads a file from the source, the file at the cache becomes fresh. After that, either the user gets the fresh file from the cache or the file at the cache becomes outdated due to a file update at the source. Thus, depending on the file status at the cache, the user may get a fresh or an outdated file. For all these system models, we derive analytical expressions for the information freshness at the end-users, and determine the updating frequencies for the intermediate caches and the end-users for maximum freshness.

References that are most closely related to our work are [4], [18]. Reference [4] studies the problem of finding optimal crawl rates to keep the information in a search engine fresh while maintaining the constraints on crawl rates imposed by the websites and also the total crawl rate constraint of the search engine. Even though the freshness metric used in [4] is similar to ours, the problem settings are different where we develop a general freshness expression for a multi-hop multi-user caching system, which differentiates our overall work from [4]. Reference [18] considers a similar model to ours where a resource constrained remote server wants to keep the items at a local cache as fresh as possible. Reference [18] shows that the update rates of the files should be chosen proportional to the square roots of their popularity indices. Different from [18] where the freshness of the local cache is considered, we consider the freshness at the user which is connected to the source via a single cache or multiple caches. Thus, our system model can be thought of as an extended version of the one-hop model in [18]. However, our freshness metric is different than the traditional age metric used in [18], and hence, the overall work in this paper is distinct compared to [18].

In this paper, we first consider a system where there is a source, a cache and a user (Fig. 2). We find an analytical expression for the average freshness of the files at the user. We then generalize our result to find the average freshness for the end-user when multiple caches are placed in between the source and the user (Fig. 4). We impose total update rate constraints for the caches and also for the user due to limited nature of resources. Our aim is to find the update rates for the cache(s) and also for the user such that the total freshness of the files at the user is maximized. We find that the average freshness of the user is a concave function of the update rates of the caches and of the user individually, but not jointly. We provide an alternating maximization based solution where the update rates of the user (resp. of the cache) are optimized

for a given set of update rates of the cache (resp. of the user). We observe that for a given set of parameters, such as update rates of the user, the optimal rate allocation policy for the other set of parameters, such as update rates at the caches, is a *threshold policy*, where the files that are updated frequently at the source may not be updated by the corresponding entity. Finally, we consider a system where multiple users are connected to a single cache (Fig. 6) and find update rates for the cache and for the users to maximize the total freshness over all users.

II. SYSTEM MODEL, FRESHNESS FUNCTION AND PROBLEM FORMULATION

We consider a cache updating system where there is an information source, a cache and a user as shown in Fig. 2. The information source keeps the freshest version of n files where the i th file is updated with exponential inter-arrival times with rate λ_i . The file updates at the source are independent of each other. A cache which is capable of storing the latest downloaded versions of all files gets the fresh files from the source. The channel between the source and the cache is assumed to be perfect and the transmission times are negligible, which is possible if the distance between the source and the cache and/or the file sizes are relatively small, as in [29]–[39]. Thus, if the cache requests an update for the i th file, it receives the file from the source right away. The inter-update request times of the cache for the i th file are exponential with rate c_i . The cache is subject to a total update rate constraint as in [18] as it is resource-constrained, i.e., $\sum_{i=1}^n c_i \leq C$.¹ The user requests the latest versions of the files stored in the cache. The inter-update request times of the user for the i th file are exponential with rate u_i . The channel between the user and the cache is also assumed to be perfect and the transmission times are negligible. Similarly, there is a total update rate constraint for the user, i.e., $\sum_{i=1}^n u_i \leq U$.

We note that each file at the source is always *fresh*. However, when a file is updated at the source, the stored versions of the same file at the cache and at the user become *outdated*. When the cache gets an update for an outdated file, the updated file in the cache becomes *fresh* again until the next update arrival at the source. When the user requests an update for an outdated file, it might still receive an outdated version if the file at the cache is not fresh. We note that since the cache and the user are unaware of the file updates at the source, they do not know whether they have the freshest versions of the files or not. Thus, they may still request an update even though they have the freshest version of a file.

In order to keep track of the *freshness*, we define the freshness function of the i th file at the cache $f_c(i, t)$ shown in Fig. 3(a) as,

$$f_c(i, t) = \begin{cases} 1, & \text{if the } i\text{th file is fresh at time } t, \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

¹We note that in practical systems, the cache's frequent update requests from the source can cause network congestion problems. Even though we impose the total update rate constraint C due to the cache's limited resources, we can also imagine it to be imposed by the source in order to avoid network congestion problems.

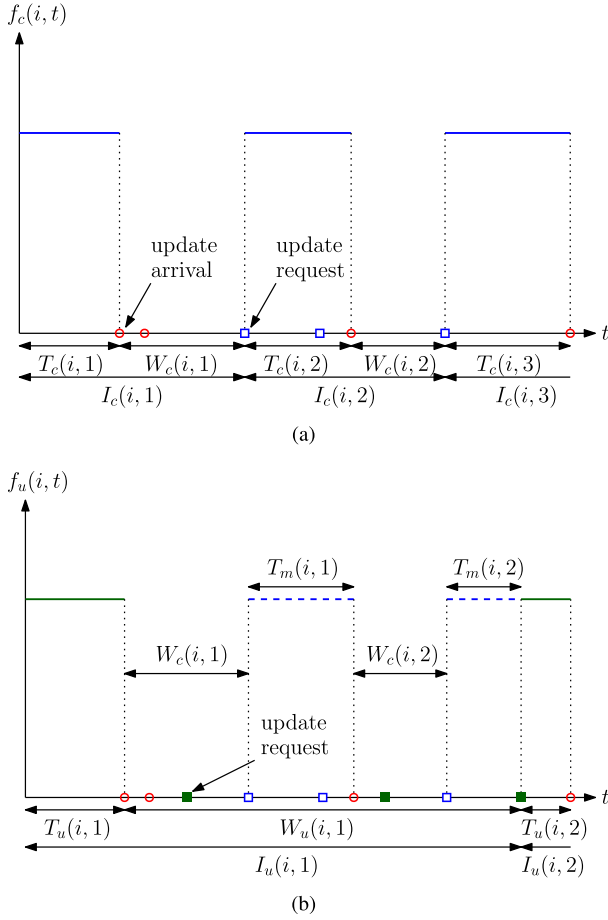


Fig. 3. Sample evolution of freshness of the i th file (a) at the cache and (b) at the user. Red circles represent the update arrivals at the source, blue squares represent the update requests from the cache, and green filled squares represent the update requests from the user.

i.e., the instantaneous freshness function is a binary function taking values of fresh, “1”, or not fresh, “0”, at any time t . The binary freshness function [1] is different than the well-known age of information metric [6]. For example, if the information at the source does not change frequently, then the receiver may still have the freshest version of the information even though it has not received an update from the source for a while. In this case, the traditional age of information metric at the receiver would be high indicating that the information at the receiver is stale. Since the traditional age of information metric does not take into account the information change rate at the source, it may not be appropriate to measure information freshness in such systems. As the binary freshness metric compares the information at the receiver with the information stored at the source, it can measure information freshness at the receiver better than the traditional age of information in such settings.

We denote file updates which replace an outdated file with the freshest version of the file as successful updates. We define the time interval between the j th and the $(j+1)$ th successful updates for the i th file at the cache as the j th update cycle and denote it by $I_c(i, j)$. We denote the time duration when the i th file at the cache is fresh during the j th update cycle as $T_c(i, j)$. Then, we define the long term average freshness of

the i th file at the cache as

$$F_c(i) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f_c(i, t) dt. \quad (2)$$

Let N denote the number of update cycles in the time duration T . Given that the system is ergodic, $\lim_{T \rightarrow \infty} \frac{N}{T}$ exists and is finite. Then, similar to [1], [6], [13], $F_c(i)$ is equal to

$$F_c(i) = \lim_{T \rightarrow \infty} \frac{N}{T} \left(\frac{1}{N} \sum_{j=1}^N T_c(i, j) \right) = \frac{\mathbb{E}[T_c(i)]}{\mathbb{E}[I_c(i)]}. \quad (3)$$

Similarly, we define $f_u(i, t)$ as the freshness function of the i th file at the user shown in Fig. 3(b). Then, the long term average freshness of the i th file at the user is equal to

$$F_u(i) = \frac{\mathbb{E}[T_u(i)]}{\mathbb{E}[I_u(i)]}. \quad (4)$$

Finally, we define F_u as the total freshness over all files at the user as

$$F_u = \sum_{i=1}^n F_u(i). \quad (5)$$

Our aim is to find the optimal update rates for the cache, c_i , and for the user, u_i , for $i = 1, \dots, n$, such that the total average freshness of the user F_u is maximized while satisfying the constraints on the total update rate for the cache, $\sum_{i=1}^n c_i \leq C$, and for the user, $\sum_{i=1}^n u_i \leq U$. Thus, our optimization problem is,

$$\begin{aligned} \max_{\{c_i, u_i\}} \quad & F_u \\ \text{s.t.} \quad & \sum_{i=1}^n c_i \leq C \\ & \sum_{i=1}^n u_i \leq U \\ & c_i \geq 0, \quad u_i \geq 0, \quad i = 1, \dots, n. \end{aligned} \quad (6)$$

In the following section, we find analytical expressions for the long term average freshness of the i th file at the cache, $F_c(i)$, and at the user, $F_u(i)$, as a function of the update rate at the source λ_i , the update rate at the cache c_i , and the update rate at the user u_i . Once we find $F_u(i)$, this will determine the objective function of (6) via (5).

III. AVERAGE FRESHNESS ANALYSIS FOR A SINGLE CACHE

In this section, we consider the system model in Fig. 2, where there is a source, a single cache and a user. First, we find an analytical expression for the long term average freshness of the i th file at the cache, i.e., $F_c(i)$ in (3). We note that due to the memoryless property of the exponential distribution, $T_c(i, j)$ which is the time duration when the i th file at the cache is fresh during the j th update cycle is exponentially distributed with parameter λ_i . Since $T_c(i, j)$ are independent and identically distributed (i.i.d.) over j , we drop index j , and denote a typical $T_c(i, j)$ as $T_c(i)$. Thus, we have $\mathbb{E}[T_c(i)] = \frac{1}{\lambda_i}$. Let $W_c(i, j)$ be the total duration when the

i th file at the cache is outdated during the j th update cycle, i.e., $W_c(i, j) = I_c(i, j) - T_c(i, j)$. Note that $W_c(i, j)$ is also equal to the time passed until the fresh version of the i th file is obtained from the source after the file is outdated at the cache. We denote typical random variables for $W_c(i, j)$ and $I_c(i, j)$ by $W_c(i)$ and $I_c(i)$, respectively. As the update request times for the cache are exponentially distributed with rate c_i , we have $\mathbb{E}[W_c(i)] = \frac{1}{c_i}$. Thus, we find

$$\mathbb{E}[I_c(i)] = \mathbb{E}[T_c(i)] + \mathbb{E}[W_c(i)] = \frac{1}{\lambda_i} + \frac{1}{c_i}. \quad (7)$$

By using (3), we find $F_c(i)$ as

$$F_c(i) = \frac{c_i}{c_i + \lambda_i}. \quad (8)$$

We note that the freshness of the i th file at the cache $F_c(i)$ in (8) is an increasing function of the cache update rate c_i , but a decreasing function of the source update rate λ_i .²

Next, we find an analytical expression for the average freshness of the i th file at the user $F_u(i)$. Similar to $\mathbb{E}[T_c(i)]$, we have $\mathbb{E}[T_u(i)] = \frac{1}{\lambda_i}$ due to the memoryless property of the exponential distribution, i.e., after the user gets the fresh file, the remaining time for the next file update at the source is still exponentially distributed with rate λ_i . Similarly, we denote the time duration when the i th file at the user is outdated during the j th update cycle as $W_u(i, j)$ which is equal to $W_u(i, j) = I_u(i, j) - T_u(i, j)$. In order for the user to get fresh updates from the cache, the cache needs to get the fresh update from the source which takes $W_c(i, j)$ time as discussed earlier. After the file at the cache becomes fresh, either the user gets the fresh update from the cache or the file at the source is updated, and thus the file at the cache becomes outdated again. We denote the earliest time that one of these two cases happens as $T_m(i)$, i.e., $T_m(i) = \min\{T_c(i), W_u(i)\}$ where $W_u(i)$ is the time for the user to obtain a new update from the cache which is exponentially distributed with rate u_i . Thus, $T_m(i)$ is also exponentially distributed with rate $u_i + \lambda_i$. We note that $\mathbb{P}[T_m(i) = W_u(i)] = \frac{u_i}{u_i + \lambda_i}$ and $\mathbb{P}[T_m(i) = T_c(i)] = \frac{\lambda_i}{u_i + \lambda_i}$.

Note that if the user gets the fresh update before the file at the cache becomes outdated which happens with probability $\mathbb{P}[T_m(i) = W_u(i)]$, an update cycle of the i th file at the user is completed and thus, $f_u(i, t)$ is equal to 1 again. However, if the file at the source is updated before the user gets the fresh update from the cache, then this process repeats itself, i.e., the cache initially needs to receive the fresh update which takes another $W_c(i, j)$ time and so on, until the user receives

the fresh update from the cache. Thus, we write $W_u(i, j)$ as

$$W_u(i, j) = \sum_{k=1}^K W_c(i, k) + T_m(i, k), \quad (9)$$

where K is a geometric random variable with rate $\frac{u_i}{u_i + \lambda_i}$. Due to [81, Prob. 9.4.1], $\sum_{k=1}^K W_c(i, k)$ and $\sum_{k=1}^K T_m(i, k)$ are exponentially distributed with rates $\frac{u_i c_i}{u_i + \lambda_i}$ and u_i , respectively. We use $W_u(i)$ and $I_u(i)$ to denote the typical random variables for $W_u(i, j)$ and $I_u(i, j)$, respectively. Thus, we have $\mathbb{E}[W_u(i)] = \frac{u_i + \lambda_i}{u_i c_i} + \frac{1}{u_i}$. Since $\mathbb{E}[I_u(i)] = \mathbb{E}[T_u(i)] + \mathbb{E}[W_u(i)]$, we get

$$\mathbb{E}[I_u(i)] = \frac{1}{\lambda_i} + \frac{1}{u_i} + \frac{u_i + \lambda_i}{u_i c_i}. \quad (10)$$

Finally, we find $F_u(i)$ as

$$F_u(i) = \frac{\mathbb{E}[T_u(i)]}{\mathbb{E}[I_u(i)]} = \frac{u_i}{u_i + \lambda_i} \frac{c_i}{c_i + \lambda_i}. \quad (11)$$

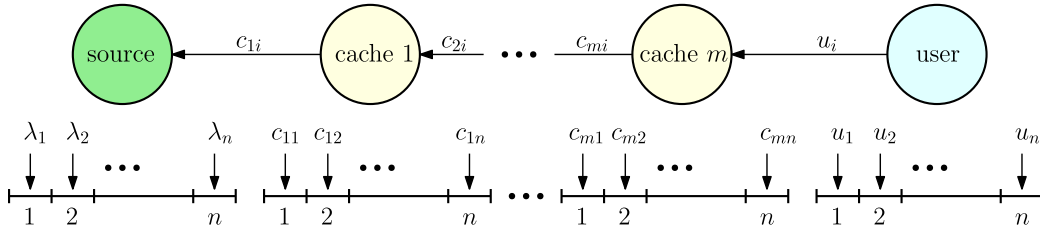
We note that the freshness of the i th file at the user $F_u(i)$ in (11) depends not only on the update rate of the user u_i and file update rate at the source λ_i but also the update rate of the cache c_i as the user obtains the fresh update from the cache.³⁴ We note that $F_u(i)$ is an increasing function of u_i and c_i , but a decreasing function of λ_i . We observe that $F_u(i)$ is an individually concave function of u_i and c_i but not jointly concave in u_i and c_i , as u_i and c_i terms appear as a multiplication in (11). If the user was directly connected to the source, its freshness would be equal to $\frac{u_i}{u_i + \lambda_i}$, i.e., the first term in (11). However, as the user is connected to the source via the cache, the freshness experienced by the user is equal to the multiplication of the freshness of the cache and the freshness of the user if the user was directly connected to the source. Note that, since $\frac{c_i}{c_i + \lambda_i} < 1$, the freshness of the user when connected to the source via a cache is smaller than the freshness it would achieve if it was directly connected to the source.

In the following section, we find the average freshness of the caches and of the user for the general case when there are m caches connected serially in between the source and the user.

³In this paper, we assume that the channel is perfect and the transmission times are negligible. We note that $F_u(i)$ in (11) can be extended to a case where the transmissions are still instantaneous but are not error free. Let p_i and q_i be the probabilities of successfully transmitting the i th file for the cache and for the user, respectively. By using [68, Prob. 9.4.1], one can show that the successful inter-update request times for the cache and for the user are exponentially distributed with rates $p_i c_i$ and $q_i u_i$, respectively. Then, the freshness of the i th file at the user becomes $F_u(i) = \frac{q_i u_i}{q_i u_i + \lambda_i} \frac{p_i c_i}{p_i c_i + \lambda_i}$. We observe that the freshness of the i th file at the user increase with p_i and q_i as expected.

⁴Even though we neglect the transmission times, as an extension to our work, in [24], we consider a cache updating system where the user is able to obtain uncached files directly from the source. However, the channel between the user and the source is imperfect, and thus there is a transmission time which is exponentially distributed with rate s_i . In [24], if the i th file is not cached and the cache is able to send all the file update requests from the user to the source, then the freshness of the i th file is equal to $F_u(i) = \frac{u_i}{u_i + \lambda_i + \frac{u_i \lambda_i}{s_i}}$. Thus, we observe that $F_u(i)$ decreases with the transmission delays as the additional term $\frac{u_i \lambda_i}{s_i}$ in the denominator decreases the freshness.

²We note that [1] investigates several updating policies, including fixed-order updating, random-order updating, and purely random updating. The freshness metric with fixed-order updating is $F_{\text{fixed-order}} = \frac{c_i}{\lambda_i} \left(1 - e^{-\frac{\lambda_i}{c_i}}\right)$, with random-order updating is $F_{\text{random-order}} = \frac{c_i}{\lambda_i} \left(1 - \frac{c_i^2}{\lambda_i^2} \left(1 - e^{-\frac{\lambda_i}{c_i}}\right)^2\right)$, and with purely random updating is $F_{\text{purely-random}} = \frac{c_i}{c_i + \lambda_i}$, as given in (8). All these three functions are monotonically decreasing in $\frac{\lambda_i}{c_i}$, and have similar forms when plotted. We adopt purely random updating in this paper due to its simplicity, and amenability to yield closed form expressions when optimized.


 Fig. 4. Generalized system model where there are m serially connected caches in between the source and the user.

IV. AVERAGE FRESHNESS ANALYSIS FOR M CACHES

In this section, we consider a system where there are m caches placed in between the source and the user, as shown in Fig. 4. We denote the r th cache's update rate for the i th file as c_{ri} . We define $I_c(r, i, j)$ as the j th update cycle for the i th file at cache r for $r = 1, \dots, m$. Similarly, we define $T_c(r, i, j)$ (and $W_c(r, i, j)$) as the time duration when the i th file at cache r is fresh (and outdated) during the j th update cycle, i.e., we have $I_c(r, i, j) = T_c(r, i, j) + W_c(r, i, j)$.

Next, we find an analytical expression for the average freshness of the i th file at the r th cache $F_c(r, i)$ and at the user $F_u(i)$. In order to obtain a fresh file at cache r , the file at cache $r - 1$ needs to be fresh for $r > 1$. Similar to the derivation of $F_u(i)$ in (11), after cache $r - 1$ obtains the fresh file, either cache r gets the fresh file from cache $r - 1$ or the file at the source is updated and the file in all the caches becomes outdated. Thus, we write $W_c(r, i, j)$ as

$$W_c(r, i, j) = \sum_{\ell=1}^{K_r} W_c(r-1, i, \ell) + T_m(r, i, \ell), \quad (12)$$

where K_r is a geometric random variable with rate $\frac{c_{ri}}{c_{ri} + \lambda_i}$ and $T_m(r, i) = \min\{T_c(r, i), \bar{W}_c(r, i)\}$ where $T_c(r, i)$ and $\bar{W}_c(r, i)$ are exponentially distributed with rates λ_i and c_{ri} , respectively. We note that $T_m(r, i)$ is also exponentially distributed with rate $c_{ri} + \lambda_i$. Then, given that $K_r = k$, we write $\mathbb{E}[W_c(r, i) | K_r = k]$ as

$$\mathbb{E}[W_c(r, i) | K_r = k] = k \left(\mathbb{E}[W_c(r-1, i)] + \frac{1}{c_{ri} + \lambda_i} \right). \quad (13)$$

Then, we find $\mathbb{E}[W_c(r, i)] = \mathbb{E}[\mathbb{E}[W_c(r, i) | K_r]]$ as

$$\mathbb{E}[W_c(r, i)] = \frac{c_{ri} + \lambda_i}{c_{ri}} \mathbb{E}[W_c(r-1, i)] + \frac{1}{c_{ri}}, \quad (14)$$

which is equal to $\mathbb{E}[W_c(1, i)] = \frac{1}{c_{1i}}$ if $r = 1$, and to

$$\mathbb{E}[W_c(r, i)] = \frac{1}{c_{ri}} + \sum_{\ell=1}^{r-1} \frac{1}{c_{\ell i}} \prod_{p=\ell+1}^r \frac{c_{pi} + \lambda_i}{c_{pi}}, \quad r = 2, \dots, m. \quad (15)$$

Then, by using $\mathbb{E}[I_c(r, i)] = \mathbb{E}[T_c(r, i)] + \mathbb{E}[W_c(r, i)]$, we have

$$\mathbb{E}[I_c(r, i)] = \begin{cases} \frac{1}{\lambda_i} + \frac{1}{c_{1i}}, & r = 1, \\ \left(\frac{1}{\lambda_i} + \frac{1}{c_{1i}} \right) \prod_{\ell=2}^r \frac{c_{\ell i} + \lambda_i}{c_{\ell i}}, & r = 2, \dots, m. \end{cases} \quad (16)$$

Finally, we find the average freshness for the i th file at cache r as

$$F_c(r, i) = \frac{\mathbb{E}[T_c(r, i)]}{\mathbb{E}[I_c(r, i)]} = \prod_{\ell=1}^r \frac{c_{\ell i}}{c_{\ell i} + \lambda_i}, \quad r = 1, \dots, m. \quad (17)$$

Similarly, we find $\mathbb{E}[I_u(i)]$ as

$$\mathbb{E}[I_u(i)] = \left(\frac{1}{\lambda_i} + \frac{1}{c_{1i}} \right) \frac{u_i + \lambda_i}{u_i} \prod_{r=2}^m \frac{c_{ri} + \lambda_i}{c_{ri}}. \quad (18)$$

Then, the average freshness of the i th file at the user is

$$F_u(i) = \frac{\mathbb{E}[T_u(i)]}{\mathbb{E}[I_u(i)]} = \frac{u_i}{u_i + \lambda_i} \prod_{r=1}^m \frac{c_{ri}}{c_{ri} + \lambda_i}. \quad (19)$$

Thus, we observe from (17) that, for the general system, the average freshness experienced by cache r for $r > 1$ is equal to the multiplication of the freshness of cache $r - 1$ with the freshness of cache r when cache r is directly connected to the source. We observe from (19) that the same structure is valid for the freshness of the user as well. We also note that the average freshness expression in (19) reduces to the expression in (11) found in Section III, when $m = 1$. Finally, as an explicit example of the expression in (19), if we have $m = 2$ caches between the source and the user, the freshness at the user is

$$F_u(i) = \frac{u_i}{u_i + \lambda_i} \frac{c_{1i}}{c_{1i} + \lambda_i} \frac{c_{2i}}{c_{2i} + \lambda_i}. \quad (20)$$

In the following section, we solve the optimization problem in (6) for the system with a single cache by using the freshness expression $F_u(i)$ found in (11) in Section III.

V. FRESHNESS MAXIMIZATION FOR A SYSTEM WITH A SINGLE CACHE

In this section, we consider the optimization problem in (6) for a system with a single cache. Using $F_u(i)$ in (11) and F_u in (5), we rewrite the freshness maximization problem as

$$\begin{aligned} \max_{\{c_i, u_i\}} \quad & \sum_{i=1}^n \frac{u_i}{u_i + \lambda_i} \frac{c_i}{c_i + \lambda_i} \\ \text{s.t.} \quad & \sum_{i=1}^n c_i \leq C \\ & \sum_{i=1}^n u_i \leq U \\ & c_i \geq 0, \quad u_i \geq 0, \quad i = 1, \dots, n. \end{aligned} \quad (21)$$

We introduce the Lagrangian function [82] for (21) as

$$\mathcal{L} = -\sum_{i=1}^n \frac{u_i}{u_i + \lambda_i} \frac{c_i}{c_i + \lambda_i} + \beta \left(\sum_{i=1}^n c_i - C \right) + \theta \left(\sum_{i=1}^n u_i - U \right) - \sum_{i=1}^n \nu_i c_i - \sum_{i=1}^n \eta_i u_i, \quad (22)$$

where $\beta \geq 0$, $\theta \geq 0$, $\nu_i \geq 0$ and $\eta_i \geq 0$. Then, we write the KKT conditions as

$$\frac{\partial \mathcal{L}}{\partial c_i} = -\frac{u_i}{u_i + \lambda_i} \frac{\lambda_i}{(c_i + \lambda_i)^2} + \beta - \nu_i = 0, \quad (23)$$

$$\frac{\partial \mathcal{L}}{\partial u_i} = -\frac{c_i}{c_i + \lambda_i} \frac{\lambda_i}{(u_i + \lambda_i)^2} + \theta - \eta_i = 0, \quad (24)$$

for all i . The complementary slackness conditions are

$$\beta \left(\sum_{i=1}^n c_i - C \right) = 0, \quad (25)$$

$$\theta \left(\sum_{i=1}^n u_i - U \right) = 0, \quad (26)$$

$$\nu_i c_i = 0, \quad (27)$$

$$\eta_i u_i = 0. \quad (28)$$

The objective function in (21) is not jointly concave in c_i and u_i since c_i s and u_i s appear as multiplicative terms in the objective function. However, for given c_i s, the objective function in (21) is concave in u_i . Similarly, for given u_i s, the objective function in (21) is concave in c_i . Thus, we apply an alternating maximization based method [67], [83]–[85] to find (c_i, u_i) pairs such that (23) and (24) are satisfied for all i .⁵

Starting with initial u_i s, we find the optimum update rates for the cache, c_i s, such that the total update rate constraint for the cache, i.e., $\sum_{i=1}^n c_i \leq C$, and the feasibility of the update rates, i.e., $c_i \geq 0$ for all i , are satisfied. Then, for given c_i s, we find the optimum update rates for the user, u_i s, such that the total update rate constraint for the user, i.e., $\sum_{i=1}^n u_i \leq U$, and the feasibility of the update rates, i.e., $u_i \geq 0$ for all i , are satisfied. We repeat these steps until the KKT conditions in (23) and (24) are satisfied.

For given u_i s with $u_i > 0$, we rewrite (23) as

$$(c_i + \lambda_i)^2 = \frac{1}{\beta - \nu_i} \frac{u_i \lambda_i}{u_i + \lambda_i}. \quad (29)$$

Then, we find c_i as

$$c_i = \frac{1}{\sqrt{\beta - \nu_i}} \sqrt{\frac{u_i \lambda_i}{u_i + \lambda_i}} - \lambda_i, \quad (30)$$

for all i with $u_i > 0$. If $c_i > 0$, we have $\nu_i = 0$ from (27). Thus, we have

$$c_i = \left(\frac{1}{\sqrt{\beta}} \sqrt{\frac{u_i \lambda_i}{u_i + \lambda_i}} - \lambda_i \right)^+, \quad (31)$$

⁵The proposed alternating maximization based method finds (c_i, u_i) pairs that satisfy the first order optimality conditions, i.e., the KKT conditions in (23)–(28). We note that as the optimization problem in (21) is not a convex optimization problem, the solutions obtained from the alternating maximization based method are not globally optimal. Similarly, solutions obtained for the optimization problems in (37) and (51) are locally optimal as well.

for all i with $u_i > 0$, where $(x)^+ = \max(x, 0)$. Note that $c_i > 0$ requires $\frac{1}{\lambda_i} \frac{u_i}{u_i + \lambda_i} > \beta$ which also implies that if $\frac{1}{\lambda_i} \frac{u_i}{u_i + \lambda_i} \leq \beta$, then we must have $c_i = 0$. Thus, for given u_i s, we see that the optimal rate allocation policy for the cache is a *threshold policy* in which the optimal update rates are equal to zero when the files are updated too frequently at the source, i.e., when the corresponding λ_i s are too large.^{6,7}

Next, we solve for u_i s for given c_i s with $c_i > 0$. We rewrite (24) as

$$(u_i + \lambda_i)^2 = \frac{1}{\theta - \eta_i} \frac{c_i \lambda_i}{c_i + \lambda_i}. \quad (32)$$

Then, we find u_i as

$$u_i = \frac{1}{\sqrt{\theta - \eta_i}} \sqrt{\frac{c_i \lambda_i}{c_i + \lambda_i}} - \lambda_i, \quad (33)$$

for all i with $c_i > 0$. If $u_i > 0$, we have $\eta_i = 0$ from (28). Thus, we have

$$u_i = \left(\frac{1}{\sqrt{\theta}} \sqrt{\frac{c_i \lambda_i}{c_i + \lambda_i}} - \lambda_i \right)^+. \quad (34)$$

Similarly, $u_i > 0$ requires $\frac{1}{\lambda_i} \frac{c_i}{c_i + \lambda_i} > \theta$ which implies that if $\frac{1}{\lambda_i} \frac{c_i}{c_i + \lambda_i} \leq \theta$, then we must have $u_i = 0$. Thus, for given c_i s, we see that the optimal rate allocation policy for the user is also a threshold policy in which the optimal update rates are equal to zero when the files are updated too frequently at the source, i.e., when the corresponding λ_i s are too large.

In the following lemma, we show that if the update rate of the cache c_i (resp. of the user u_i) is equal to zero for the i th file, then the update rate of the user u_i (resp. of the cache c_i) for the same file must also be equal to zero.

Lemma 1: *In the optimal policy, if $c_i = 0$, then we must have $u_i = 0$; and vice versa.*

Proof: Assume for contradiction that in the optimal policy, there exist update rates with $c_i = 0$ and $u_i > 0$. We note that average freshness of this file at the user is equal to zero, i.e., $F_u(i) = 0$, as $c_i = 0$. We can increase the total freshness of the user F_u by decreasing u_i to zero and increasing one of u_j s with $c_j > 0$. Thus, we reach a contradiction and in the optimal policy, if $c_i = 0$, we must have $u_i = 0$. For the update rates with $c_i > 0$ and $u_i = 0$, one can similarly show that if $u_i = 0$, then we must have $c_i = 0$. ■

In the following lemma, we show that the total update rate constraints for the cache, i.e., $\sum_{i=1}^n c_i \leq C$, and for the user, i.e., $\sum_{i=1}^n u_i \leq U$, must be satisfied with equality.

⁶As the stored versions of the files that change too frequently at the source become obsolete too quickly at the user, the freshness of these files at the user will be small. That is why, instead of updating these files, with the threshold policy, the files that change less frequently at the source are updated more which brings higher contribution to the overall freshness at the user.

⁷As a result of the threshold policy, the user may not receive the fresh versions of the files that change too frequently at the source which can be undesirable especially if obtaining the fresh versions of some files is more important than the others. In order to address this problem, we can introduce an importance factor for each file μ_i . Then, we can rewrite the overall freshness at the user F_u in (5) as $F_u = \sum_{i=1}^n \mu_i F_u(i)$. We note that by solving the optimization problem in (6) with a freshness expression with importance factors, the important files that change too frequently at the source may be updated by the cache and also by the user.

Lemma 2: In the optimal policy, we must have $\sum_{i=1}^n c_i = C$ and $\sum_{i=1}^n u_i = U$.

Proof: Assume for contradiction that in the optimal policy, we have $\sum_{i=1}^n c_i < C$. As the objective function in (21) is an increasing function of c_i , we can increase the total freshness of the user F_u by increasing one of c_i with $u_i > 0$ until the total update rate constraint for the cache is satisfied with equality, i.e., $\sum_{i=1}^n c_i = C$. Thus, we reach a contradiction and in the optimal policy, we must have $\sum_{i=1}^n c_i = C$. By using a similar argument, we can also show that in the optimal policy, we must have $\sum_{i=1}^n u_i = U$. ■

In the following lemma, we identify a property of the optimal cache update rates c_i for given user update rates u_i . To that end, for given u_i s, let us define ϕ_i s as

$$\phi_i = \frac{1}{\lambda_i} \frac{u_i}{u_i + \lambda_i}. \quad (35)$$

This lemma will be useful for solving for c_i given u_i using (31).

Lemma 3: For given u_i s, if $c_i > 0$ for some i , then we have $c_j > 0$ for all j with $\phi_j \geq \phi_i$.

Proof: As we noted earlier, from (31), $c_i > 0$ implies $\phi_i > \beta$. Thus, if $\phi_j \geq \phi_i$, then we have $\phi_j > \beta$, which further implies $c_j > 0$. ■

Next, we describe the overall solution for the single cache setting. We start with a set of initial u_i s. We obtain ϕ_i from u_i using (35). We will utilize Fig. 5 to describe the steps of the solution visually. We plot ϕ_i in Fig. 5. Note that if $u_i = 0$ then $\phi_i = 0$, and vice versa. First, we choose $c_i = 0$ for the files with $u_i = 0$ due to Lemma 1, i.e., in Fig. 5, we choose c_3 and c_6 as zero. Next, we find the remaining c_i s with $u_i > 0$. For that, we rewrite (31) as

$$c_i = \frac{\lambda_i}{\sqrt{\beta}} \left(\sqrt{\phi_i} - \sqrt{\beta} \right)^+. \quad (36)$$

Due to Lemma 2, in the optimal policy, we must have $\sum_{i=1}^n c_i = C$. Assuming that $\phi_i \geq \beta$ for all i , i.e., by ignoring $(\cdot)^+$ in (31) and (36), we solve $\sum_{i=1}^n c_i = C$ for β . Then, we compare the smallest ϕ_i with β . If the smallest ϕ_i is larger than or equal to β , it implies that $c_i > 0$ for all i due to Lemma 3, and we have obtained the optimal c_i values for given u_i s. If the smallest ϕ_i is smaller than β , it implies that the corresponding c_i was negative and it must be chosen as zero. In this case, we choose $c_i = 0$ for the smallest ϕ_i . In the example in Fig. 5, if the β we found is β_1 , then since $\phi_7 < \beta_1$ we choose c_7 as zero. Then, we repeat this process again until the smallest ϕ_i among the remaining c_i s satisfies $\phi_i \geq \beta$. For example, in Fig. 5, the next β found by using only indices 1, 2, 4, 5, 8 may be β_2 . Since $\phi_5 < \beta_2$, we choose $c_5 = 0$. In the next iteration, the β found by using indices 1, 2, 4, 8 may be β_3 . Since $\phi_8 > \beta_3$, we stop the process and find c_i for $i = 1, 2, 4, 8$ from (31) or (36) by using β_3 in Fig. 5. This concludes finding c_i s for given u_i s. Next, for given c_i s, we find u_i s by following a similar procedure. We keep solving for c_i s for given u_i s, and u_i s for given c_i s, until (c_i, u_i) pairs converge.

In the following section, we provide a solution for the general system with multiple caches.

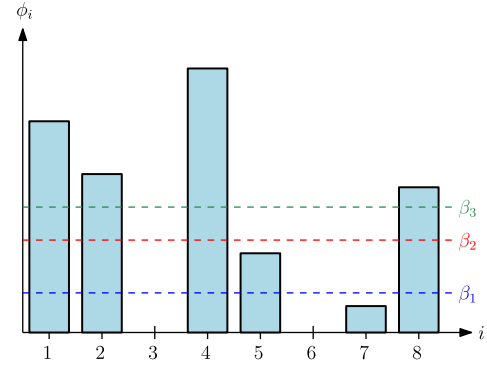


Fig. 5. For given u_i s, we show ϕ_i s calculated in (35) for $n = 8$.

VI. FRESHNESS MAXIMIZATION FOR THE GENERAL SYSTEM

In this section, we provide a solution for the general system shown in Fig. 4, where there are m caches in between the source and the user. We define C_r as the total update rate of cache r , i.e., $\sum_{i=1}^n c_{ri} \leq C_r$. Using $F_u(i)$ in (19), we rewrite the optimization problem in (6) as

$$\begin{aligned} \max_{\{c_{ri}, u_i\}} \quad & \sum_{i=1}^n \frac{u_i}{u_i + \lambda_i} \prod_{r=1}^m \frac{c_{ri}}{c_{ri} + \lambda_i} \\ \text{s.t.} \quad & \sum_{i=1}^n c_{ri} \leq C_r, \quad r = 1, \dots, m \\ & \sum_{i=1}^n u_i \leq U \\ & c_{ri} \geq 0, \quad u_i \geq 0, \quad r = 1, \dots, m, \quad i = 1, \dots, n. \end{aligned} \quad (37)$$

We introduce the Lagrangian function for (37) as

$$\begin{aligned} \mathcal{L} = & - \sum_{i=1}^n \frac{u_i}{u_i + \lambda_i} \prod_{r=1}^m \frac{c_{ri}}{c_{ri} + \lambda_i} + \sum_{r=1}^m \beta_r \left(\sum_{i=1}^n c_{ri} - C_r \right) \\ & + \theta \left(\sum_{i=1}^n u_i - U \right) - \sum_{r=1}^m \sum_{i=1}^n \nu_{ri} c_{ri} - \sum_{i=1}^n \eta_i u_i, \end{aligned} \quad (38)$$

where $\beta_r \geq 0$, $\theta \geq 0$, $\nu_{ri} \geq 0$ and $\eta_i \geq 0$. Then, we write the KKT conditions as

$$\frac{\partial \mathcal{L}}{\partial c_{ri}} = - \frac{u_i}{u_i + \lambda_i} \prod_{\ell \neq r} \frac{c_{\ell i}}{c_{\ell i} + \lambda_i} \frac{\lambda_i}{(c_{ri} + \lambda_i)^2} + \beta_r - \nu_{ri} = 0, \quad (39)$$

$$\frac{\partial \mathcal{L}}{\partial u_i} = - \frac{\lambda_i}{(u_i + \lambda_i)^2} \prod_{r=1}^m \frac{c_{ri}}{c_{ri} + \lambda_i} + \theta - \eta_i = 0, \quad (40)$$

for all r and i . The complementary slackness conditions are

$$\beta_r \left(\sum_{i=1}^n c_{ri} - C_r \right) = 0, \quad (41)$$

$$\theta \left(\sum_{i=1}^n u_i - U \right) = 0, \quad (42)$$

$$\nu_{ri} c_{ri} = 0, \quad (43)$$

$$\eta_i u_i = 0. \quad (44)$$

The objective function in (37) is not jointly concave in c_{ri} and u_i . However, for given c_{ri} s, the objective function in (37) is concave in u_i , and for a given u_i and $c_{\ell i}$ s for all $\ell \neq r$, the objective function in (37) is concave in c_{ri} . Thus, similar to the solution approach in Section V, we apply an alternating maximization based method to find $(c_{1i}, \dots, c_{ri}, u_i)$ tuples such that (39) and (40) are satisfied for all r and i .

Starting with initial u_i and $c_{\ell i}$ s for $\ell \neq r$, we find the optimum update rates for cache r , c_{ri} s, such that the total update rate constraint for the cache, i.e., $\sum_{i=1}^n c_{ri} \leq C_r$, and the feasibility of the update rates, i.e., $c_{ri} \geq 0$ for all i , are satisfied. We repeat this step for all r . Then, for given c_{ri} s, we find the optimum update rates for the user, u_i s, such that the total update rate constraint for the user, i.e., $\sum_{i=1}^n u_i \leq U$, and the feasibility of the update rates, i.e., $u_i \geq 0$ for all i , are satisfied. We repeat these steps until the KKT conditions in (39) and (40) are satisfied.

For given u_i s with $u_i > 0$, and $c_{\ell i}$ with $c_{\ell i} > 0$, for $\ell \neq r$, we rewrite (39) as

$$(c_{ri} + \lambda_i)^2 = \frac{\sigma_i \lambda_i}{\beta_r - \nu_{ri}}, \quad (45)$$

where $\sigma_i = \frac{u_i}{u_i + \lambda_i} \prod_{\ell \neq r} \frac{c_{\ell i}}{c_{\ell i} + \lambda_i}$. Then, we find c_{ri} as

$$c_{ri} = \sqrt{\frac{\sigma_i \lambda_i}{\beta_r - \nu_{ri}}} - \lambda_i, \quad (46)$$

for all i with $u_i > 0$ and $c_{\ell i} > 0$ for $\ell \neq r$. If $c_{ri} > 0$, we have $\nu_{ri} = 0$ from (43). Thus, we have

$$c_{ri} = \left(\sqrt{\frac{\sigma_i \lambda_i}{\beta_r}} - \lambda_i \right)^+. \quad (47)$$

Note that $c_{ri} > 0$ requires $\frac{\sigma_i}{\lambda_i} > \beta_r$, i.e., $\frac{1}{\lambda_i} \frac{u_i}{u_i + \lambda_i} \prod_{\ell \neq r} \frac{c_{\ell i}}{c_{\ell i} + \lambda_i} > \beta_r$ which also implies that if $\frac{1}{\lambda_i} \frac{u_i}{u_i + \lambda_i} \prod_{\ell \neq r} \frac{c_{\ell i}}{c_{\ell i} + \lambda_i} \leq \beta_r$, then we must have $c_{ri} = 0$. We repeat this step for $r = 1, \dots, m$.

Next, we solve for u_i s for given c_{ri} s for all r with $c_{ri} > 0$. We rewrite (40) as

$$(u_i + \lambda_i)^2 = \frac{\rho_i \lambda_i}{\theta - \eta_i}, \quad (48)$$

where $\rho_i = \prod_{r=1}^m \frac{c_{ri}}{c_{ri} + \lambda_i}$. Then, we find u_i as

$$u_i = \sqrt{\frac{\rho_i \lambda_i}{\theta - \eta_i}} - \lambda_i, \quad (49)$$

for all i with $c_{ri} > 0$ for all r . If $u_i > 0$, we have $\eta_i = 0$ from (44). Thus, we have

$$u_i = \left(\sqrt{\frac{\rho_i \lambda_i}{\theta}} - \lambda_i \right)^+. \quad (50)$$

Similarly, $u_i > 0$ requires $\frac{\rho_i}{\lambda_i} > \theta$, i.e., $\frac{1}{\lambda_i} \prod_{r=1}^m \frac{c_{ri}}{c_{ri} + \lambda_i} > \theta$ which implies that if $\frac{1}{\lambda_i} \prod_{r=1}^m \frac{c_{ri}}{c_{ri} + \lambda_i} \leq \theta$, then we must have $u_i = 0$. Thus, similar to the results in Section V, for given c_{ri} s (resp. for given u_i s and $c_{\ell i}$ s for all $\ell \neq r$), we observe that the optimal rate allocation policy for the user (resp. for cache r) is a threshold policy and the update rates for the user

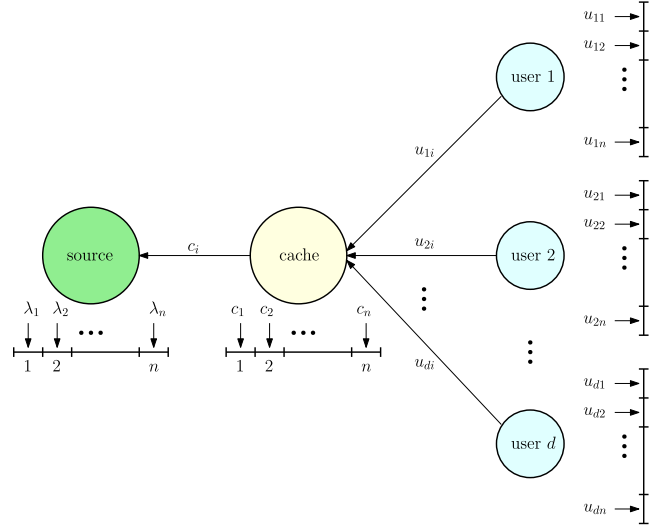


Fig. 6. A cache updating system with a source, a single cache and d users.

(resp. for cache r) are equal to zero for the files with very large λ_i s.

Similar to Lemma 1, one can show that in the optimal policy, if $c_{ri} = 0$ for some r , then we must have $c_{\ell i} = 0$ for all $\ell \neq r$ and $u_i = 0$. Furthermore, if $u_i = 0$ for some i , then we must have $c_{ri} = 0$ for all r . One can also show that the total update rate constraints for cache r , i.e., $\sum_{i=1}^n c_{ri} \leq C_r$, for all r and for the user, i.e., $\sum_{i=1}^n u_i \leq U$, should be satisfied with equality as the objective function in (37) is an increasing function of c_{ri} and u_i . Thus, in the optimal policy, we must have $\sum_{i=1}^n c_{ri} = C_r$ for all r and $\sum_{i=1}^n u_i = U$.

We note from (47) that the update rates of cache r directly depend on the update rates of the other caches as well as the update rates of the user. Similarly, we note from (50) that the update rates of the user directly depend on the update rates of all caches. In order to find the overall solution, for given initial u_i s and $c_{\ell i}$ for all $\ell \neq r$, we choose $c_{ri} = 0$ for the files with $u_i = 0$ or $c_{\ell i} = 0$ for some ℓ . We solve $\sum_{i=1}^n c_{ri} = C_r$ for β_r and then, find c_{ri} s by using (47) similar to the solution method in Section V. We repeat this step for $r = 1, \dots, m$. Next, for given c_{ri} s, we find u_i s by following a similar procedure. We keep updating these parameters until $(c_{1i}, \dots, c_{ri}, u_i)$ tuples converge.

In the following section, we find rate allocations for a system with a source, a single cache, and multiple users.

VII. FRESHNESS MAXIMIZATION FOR A SYSTEM WITH MULTIPLE USERS

In this section, we consider a system where there is a source, a single cache and d users connected to the cache, as shown in Fig. 6. Our aim is to find the update rates for the cache and for the users such that the overall freshness experienced by the users is maximized.

Let the k th user's update rate for the i th file be u_{ki} . Each user is subject to a total update rate constraint as $\sum_{i=1}^n u_{ki} \leq U_k$, for $k = 1, \dots, d$, and the cache is subject to a total update rate constraint as $\sum_{i=1}^n c_i \leq C$. From

Section III, the average freshness of the i th file at the k th user is $F_u(k, i) = \sum_{i=1}^n \frac{u_{ki}}{u_{ki} + \lambda_i} \frac{c_i}{c_i + \lambda_i}$. Then, we write the freshness maximization problem as

$$\begin{aligned} \max_{\{c_i, u_{ki}\}} \quad & \sum_{k=1}^d \sum_{i=1}^n \frac{u_{ki}}{u_{ki} + \lambda_i} \frac{c_i}{c_i + \lambda_i} \\ \text{s.t.} \quad & \sum_{i=1}^n c_i \leq C \\ & \sum_{i=1}^n u_{ki} \leq U_k, \quad k = 1, \dots, d \\ & c_i \geq 0, \quad u_{ki} \geq 0, \quad k = 1, \dots, d, \quad i = 1, \dots, n. \end{aligned} \quad (51)$$

We introduce the Lagrangian function for (51) as

$$\begin{aligned} \mathcal{L} = & - \sum_{k=1}^d \sum_{i=1}^n \frac{u_{ki}}{u_{ki} + \lambda_i} \frac{c_i}{c_i + \lambda_i} + \beta \left(\sum_{i=1}^n c_i - C \right) \\ & + \sum_{k=1}^d \theta_k \left(\sum_{i=1}^n u_{ki} - U_k \right) - \sum_{i=1}^n \nu_i c_i - \sum_{k=1}^d \sum_{i=1}^n \eta_{ki} u_{ki}, \end{aligned} \quad (52)$$

where $\beta \geq 0$, $\theta_k \geq 0$, $\nu_i \geq 0$ and $\eta_{ki} \geq 0$. We write the KKT conditions as

$$\frac{\partial \mathcal{L}}{\partial c_i} = - \frac{\lambda_i}{(c_i + \lambda_i)^2} \sum_{k=1}^d \frac{u_{ki}}{u_{ki} + \lambda_i} + \beta - \nu_i = 0, \quad (53)$$

$$\frac{\partial \mathcal{L}}{\partial u_{ki}} = - \frac{\lambda_i}{(u_{ki} + \lambda_i)^2} \frac{c_i}{c_i + \lambda_i} + \theta_k - \eta_{ki} = 0, \quad (54)$$

for all k and i . The complementary slackness conditions are

$$\beta \left(\sum_{i=1}^n c_i - C \right) = 0, \quad (55)$$

$$\theta_k \left(\sum_{i=1}^n u_{ki} - U_k \right) = 0, \quad (56)$$

$$\nu_i c_i = 0, \quad (57)$$

$$\eta_{ki} u_{ki} = 0. \quad (58)$$

The objective function in (51) is not jointly concave in c_i and u_{ki} . However, for given u_{ki} s, the objective function in (51) is concave in c_i , and for given c_i s, the objective function in (51) is jointly concave in all u_{ki} . Thus, similar to the solution approach used in Section V, we apply an alternating maximization based method to find (c_i, u_{ki}) such that (53) and (54) are satisfied for all k and i .

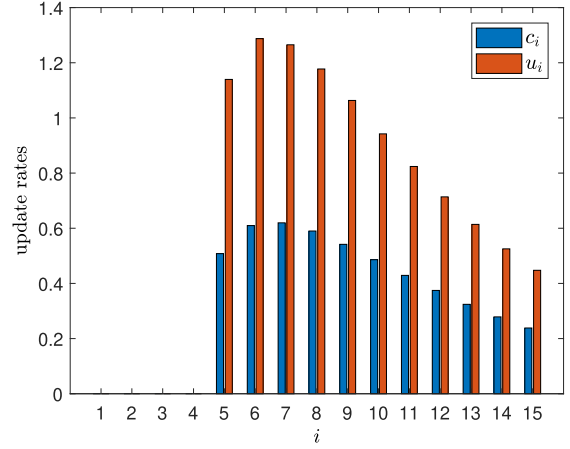
For given u_{ki} s, we find c_i as

$$c_i = \frac{1}{\sqrt{\beta - \nu_i}} \sqrt{\sum_{k=1}^d \frac{u_{ki} \lambda_i}{u_{ki} + \lambda_i}} - \lambda_i, \quad (59)$$

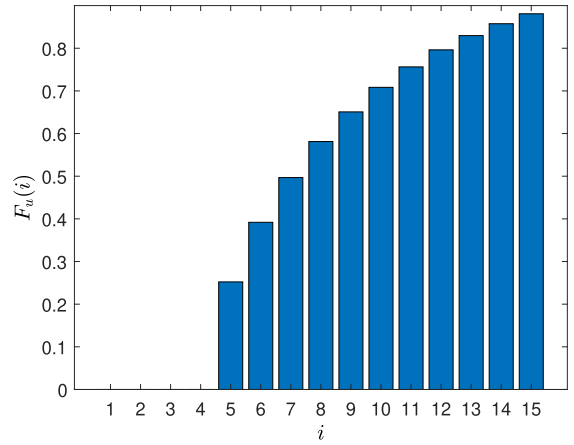
for all i . If $c_i > 0$, we have $\nu_i = 0$ due to (57). Thus, we have

$$c_i = \left(\frac{1}{\sqrt{\beta}} \sqrt{\sum_{k=1}^d \frac{u_{ki} \lambda_i}{u_{ki} + \lambda_i}} - \lambda_i \right)^+. \quad (60)$$

Similarly, the optimal rate allocation policy is a threshold policy where if $\frac{1}{\lambda_i} \left(\sum_{k=1}^d \frac{u_{ki}}{u_{ki} + \lambda_i} \right) < \beta$, then we have $c_i = 0$.



(a)



(b)

Fig. 7. (a) Update rate allocation for the cache and the user for each file, and (b) the corresponding freshness $F_u(i)$, when $U = 10$ and $C = 5$ with the file update rates at the source λ_i given in (63), with $a = 10$ and $q = 0.7$ for $n = 15$.

Next, for a given c_i with $c_i > 0$, we find u_{ki} as

$$u_{ki} = \frac{1}{\sqrt{\theta_k - \eta_{ki}}} \sqrt{\frac{c_i \lambda_i}{c_i + \lambda_i}} - \lambda_i, \quad (61)$$

for all i . If $u_{ki} > 0$, we have $\eta_{ki} = 0$ due to (58). Thus, we have

$$u_{ki} = \left(\frac{1}{\sqrt{\theta_k}} \sqrt{\frac{c_i \lambda_i}{c_i + \lambda_i}} - \lambda_i \right)^+. \quad (62)$$

Thus, we note that the optimal rate allocation policy is a threshold policy where if $\frac{1}{\lambda_i} \frac{c_i}{c_i + \lambda_i} < \theta_k$, then we have $u_{ki} = 0$. We observe that the optimal rates for the users depend directly on λ_i and c_i . As the update rates for the cache depend on the update rates of all the users, update rates of the users affect each other indirectly. We also note that in Section VI, where the caches are connected serially, we see this effect as a product of the terms, i.e., $\prod_{r=1}^m \frac{c_{ri}}{c_{ri} + \lambda_i}$. In this section, as the users are connected in parallel to a single cache, we see this effect as a summation of the terms, i.e., $\sum_{k=1}^d \frac{u_{ki}}{u_{ki} + \lambda_i}$.

In the next section, we provide numerical results for the system with a single cache in Section V, with multiple caches in Section VI, and with multiple users in Section VII.

VIII. NUMERICAL RESULTS

In this section, we provide five numerical results. For these results, we use the following update arrival rates at the source

$$\lambda_i = bq^i, \quad i = 1, \dots, n, \quad (63)$$

where $b > 0$ and $0 < q \leq 1$ such that $\sum_{i=1}^n \lambda_i = a$. Note that with the update arrival rates at the source in (63), we have $\lambda_i \geq \lambda_j$ for $i \leq j$.

In example 1, we take $a = 10$, $q = 0.7$, and $n = 15$ in (63). For this example, we consider the system with a source, a single cache and a user. We choose the total update rate constraint for the cache as $C = 5$, i.e., $\sum_{i=1}^n c_i \leq 5$, and for the user as $U = 10$, i.e., $\sum_{i=1}^n u_i \leq 10$. We initialize the file update rates at the user as $u_i = \frac{U}{n}$ for all i . We apply the alternating maximization method in Section V to find the update rates for the cache and for the user until the KKT conditions in (23) and (24) are satisfied. We see in Fig. 7(a) that the first four files which are updated most frequently at the source are not updated by the cache and the user. As these files change too frequently at the source, their stored versions at the user become obsolete very quickly. In other words, updating files that change less frequently at the source brings more contribution to the overall information freshness at the user. The distributions of the update rates for the cache and for the user have similar trends as the update rates increase up to the seventh file for the cache and the sixth file for the user, and then update rates for the cache and the user decrease. Even though the update rates of the cache and the user for the slowly changing files are low, we see in Fig. 7(b) that the freshness of the slowly varying files is higher compared to the rapidly changing files.

In example 2, we consider the same system as in example 1, and compare the performance of the proposed updating policy which maximizes the freshness at the user with two other baseline updating policies. As the first baseline policy, we consider λ -proportional update policy where the update rates for the cache and for the user are chosen as $c_i = \frac{\lambda_i C}{\sum_{i=1}^n \lambda_i}$ and $u_i = \frac{\lambda_i U}{\sum_{i=1}^n \lambda_i}$ for all i , respectively. With this policy, the freshness of the i th file at the user is $F_u(i) = \frac{U}{U+a} \frac{C}{C+a}$. Thus, this update policy may be desirable if the user wants to have the same level of freshness for all files. As the second baseline policy, we consider λ -inverse update policy where the update rates for the cache and for the user are given by $c_i = C \frac{\frac{1}{\lambda_i}}{\sum_{i=1}^n \frac{1}{\lambda_i}}$ and $u_i = U \frac{\frac{1}{\lambda_i}}{\sum_{i=1}^n \frac{1}{\lambda_i}}$ for all i , respectively. The motivation for λ -inverse update policy stems from the fact that updating slowly varying files at the source with higher update rates brings more contribution to the overall freshness. In this example, we take $C = 15$, $U = 10$, and λ_i in (63) with $n = 20$. In Fig. 8(a), we fix the total file update rate at the source, i.e., $a = 10$, and vary the distribution of the file update rate at the source by choosing $0 < q \leq 1$. We observe in Fig. 8(a) that proposed update policy achieves

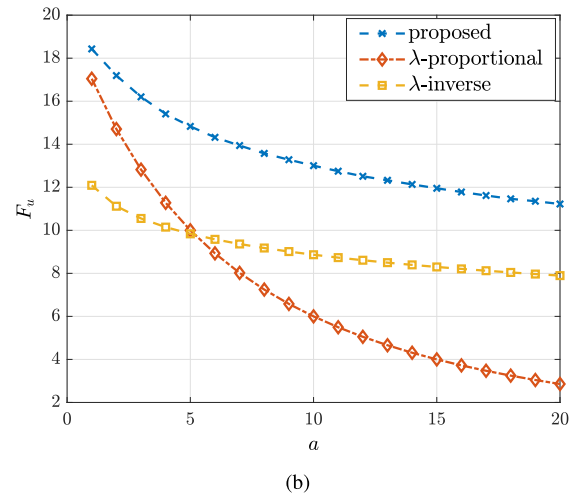
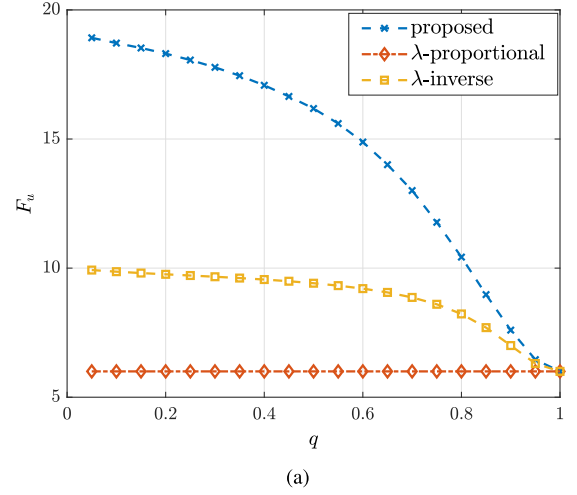


Fig. 8. We compare the proposed update policy with the λ -proportional and the λ -inverse updating policies when $C = 15$, $U = 10$. We use λ_i in (63) for $n = 20$, (a) $a = 10$, and $0 < q \leq 1$ and (b) $q = 0.7$ and $a = 1, \dots, 20$.

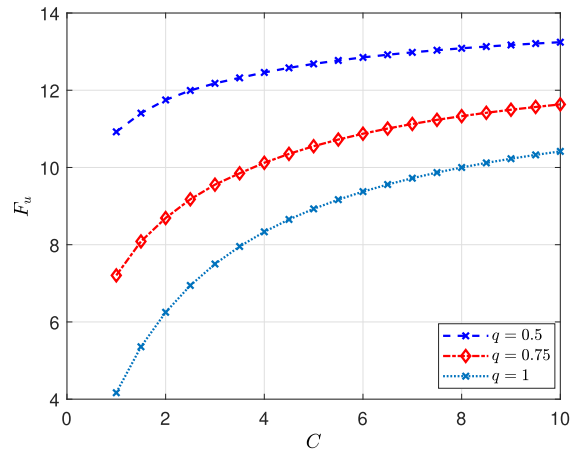


Fig. 9. Total freshness of the user F_u with respect to C , when λ_i are given in (63), with $a = 2$ and $q = 0.5, 0.75, 1$ for $n = 15$.

the highest freshness compared to the considered baseline policies. We see that the λ -proportional policy achieves a constant level of freshness independent of the distribution of the file change rate at the source, but this policy gives the

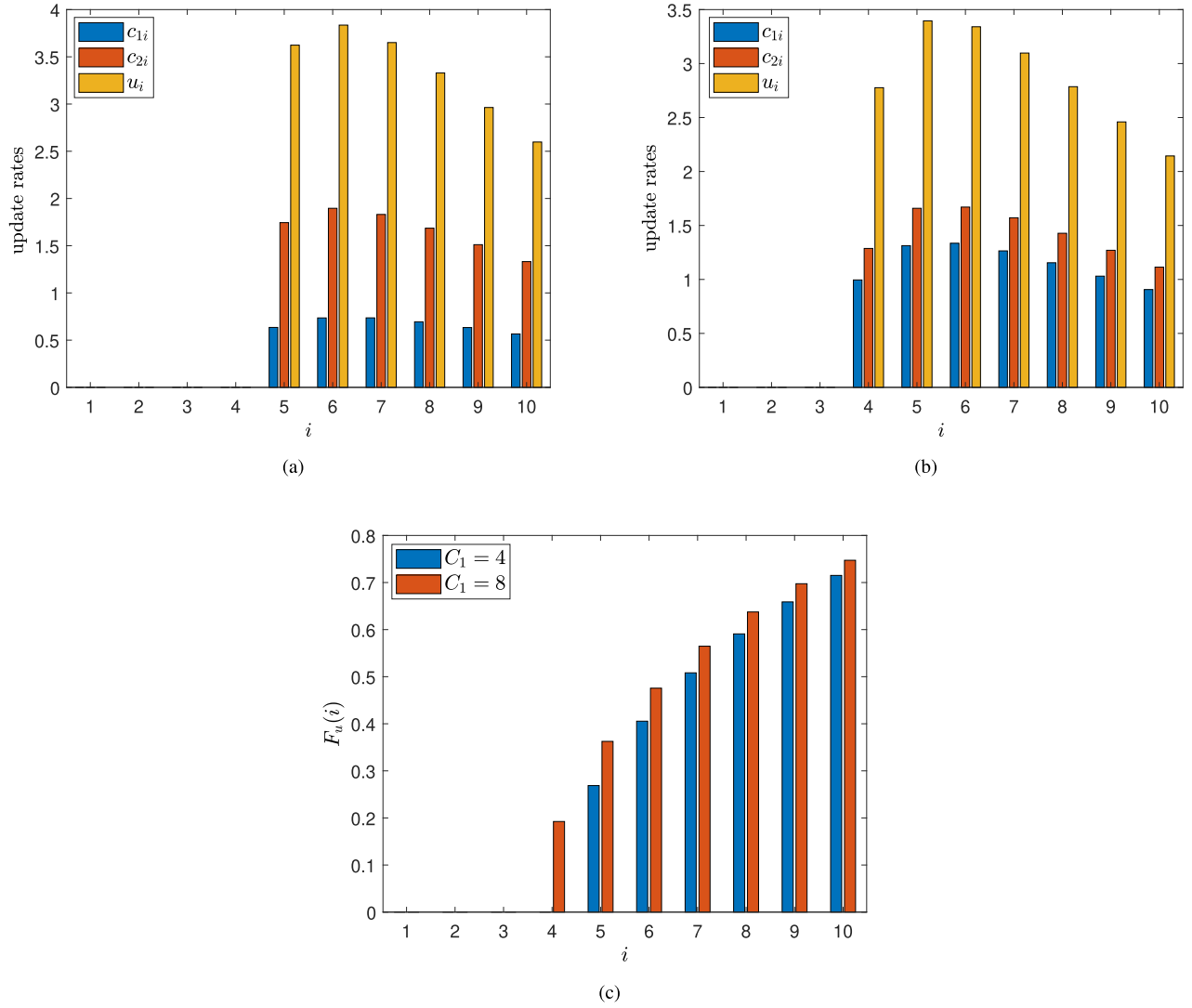


Fig. 10. The update rates of the caches and the user when the total update rate constraint for the user is $U = 20$ and, for the second cache is $C_2 = 10$. Total update rate constraint for the first cache is (a) $C_1 = 4$, and (b) $C_1 = 8$. The freshness of the files at the user is shown in (c).

lowest freshness compared to others. We note that when λ_i s are evenly distributed, i.e., when q is close to 1, we see that all three updating policies achieve similar levels of freshness. In Fig. 8(b), we fix $q = 0.7$ and increase the file change rate at the source, i.e., $a = 1, \dots, 20$. We observe in Fig. 8(b) that the proposed update policy achieves the highest freshness. When the total file change rate at the source is low, λ -proportional update policy achieves freshness close to the proposed update policy, whereas when the total file change rate at the source is high, λ -inverse achieves freshness close to the proposed update policy.

In example 3, we consider the same system as in examples 1 and 2 but this time we examine the effect of file update rates at the source over the information freshness at the user. We take λ_i in (63) with $a = 2$, $n = 15$ for $q = 0.5, 0.75, 1$. We note that a smaller q corresponds to a less even (more polarized) distribution of file change rates at the source. We choose the total update rate for the user as $U = 10$ and vary the total update rate for the cache as $C = 1, 1.5, \dots, 10$. For each C and q values, we initialize $u_i = \frac{U}{n}$ for all i and apply the alternating maximization method proposed in

Section V until convergence. We see in Fig. 9 that when the distribution of the change rates of the files are more polarized, i.e., when q is small, the overall information freshness at the user is larger as the freshness contribution from the slowly varying files can be utilized more with the more polarized distributions of file change rates at the source. We also note that for a fixed λ_i distribution, the freshness of the user increases with the total update rate at the cache C due to the fact that the user gets fresh files more frequently from the cache as the freshness of the files at the cache increases with C .

In example 4, we consider a system where there are two caches placed in between the source and the user with λ_i as given in (63) with $a = 10$, $q = 0.7$, and $n = 10$. We take the total update rate constraint for the user as $U = 20$ and, for the second cache as $C_2 = 10$. We initialize the update rates at the user and at the second cache as $u_i = \frac{U}{n}$ and $c_{2i} = \frac{C_2}{n}$ for all i , respectively, and apply the alternating maximization method proposed in Section VI, for the total update rate constraint for the first cache as $C_1 = 4, 8$. We observe that the update rates for the caches and for the user have similar

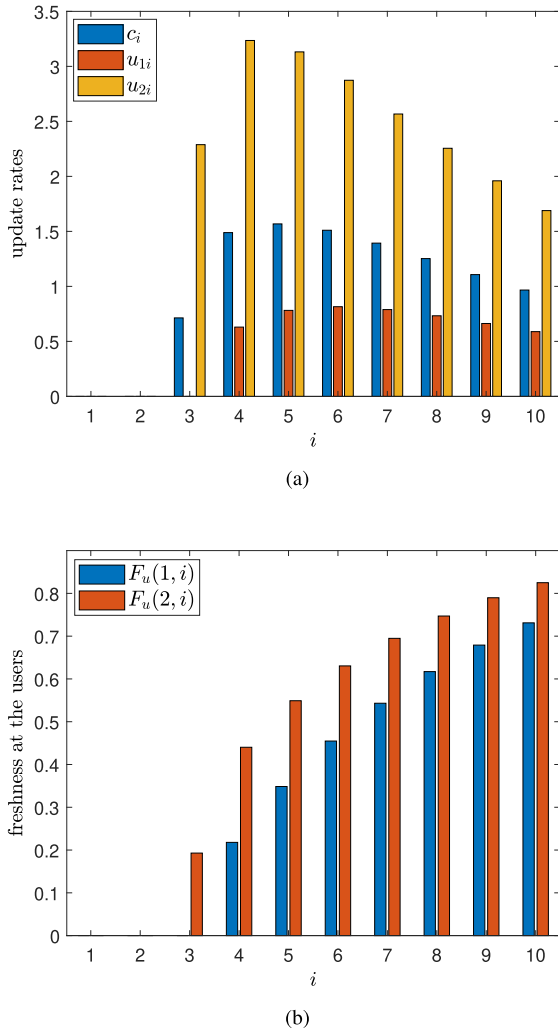


Fig. 11. A system with a source, a single cache and two users: (a) The update rates of the cache and of the users, (b) the freshness of each file at the users.

trends as shown in Fig. 10(a) for $C_1 = 4$ and in Fig. 10(b) for $C_1 = 8$. We note that the update rates of the caches in (47) depend directly on the update rates of the other caches and also of the user. Similarly, the update rates of the user in (50) depend directly the update rates of all caches. That is why even though the total update rate constraints for the second cache and for the user remain the same in Fig. 10(a)-(b), we see that the update rates at the second cache and at the user also change depending on the update rates at the first cache. In Fig. 10(c), we observe that increasing the total update rate constraint for the first cache improves the freshness of every file except the first three files as the total update rate constraints for the caches and for the user are not high enough to update the most rapidly changing files. Furthermore, the improvement on the freshness of the rapidly changing files is more significant than the others as the freshness of the files at the user is a concave increasing function of c_{1i} .

In example 5, we consider the system where there is a source, a cache and two users connected to the cache. For this example, we use λ_i in (63) with $a = 10$, $q = 0.7$ and $n = 10$. We take the total update rate constraint for the cache

as $C = 10$, for the first user as $U_1 = 5$ and for the second user as $U_2 = 20$. We initialize the cache update rates as $c_i = \frac{C}{n}$ for all i , and apply the proposed alternating maximization based method to find (c_i, u_{1i}, u_{2i}) that satisfy the KKT conditions in (53) and (54). The update rates of the cache and the users are shown in Fig. 11(a) where we see that even though the update rate of the third file at the first user is equal to zero, the cache still updates the third file for the second user. We see in Fig. 11(b) that the freshness of the files at the second user is higher as the total update rate of the second user is higher compared to the first user. The freshness difference between the users decreases for the slowly changing files at the source.

IX. CONCLUSION AND FUTURE DIRECTIONS

In this paper, we first considered a cache updating system with a source, a single cache and a user. We found an analytical expression for the average freshness of the files at the cache and also at the user. We generalized this setting to the case where there are multiple caches placed in between the source and the user. Then, we provided an alternating maximization based method to find the update rates for the cache(s) and for the user to maximize the total freshness of the files at the user. We observed that for a given set of update rates of the user (resp. of the cache), the optimal rate allocation policy for the cache (resp. for the user) is a threshold policy where the frequently changing files at the source may not be updated by the user and also by the cache. Finally, we considered a system with multiple users refreshed via a single cache and found update rates for the cache and the users to maximize the overall freshness of the users.

As a future direction, one can consider a system where a user has access to multiple caches with different total update rate constraints. As a fresh file can be obtained through multiple caches, one can formulate a problem of finding the optimal rate allocation policies for the caches and for the user to maximize the freshness at the user. Another interesting direction to consider is the case where the caches can store only a portion of the files. In this case, one can consider the problem of joint content placement/assignment to the caches as well as the optimization of the update rates of each cache to maximize the information freshness for multiple users. While we considered only serially and parallel connected caches and users, one can investigate information freshness in arbitrarily connected networks.

REFERENCES

- [1] J. Cho and H. Garcia-Molina, "Effective page refresh policies for Web crawlers," *ACM Trans. Database Syst.*, vol. 28, no. 4, pp. 390–426, Dec. 2003.
- [2] B. E. Brewington and G. Cybenko, "Keeping up with the changing Web," *Computer*, vol. 33, no. 5, pp. 52–58, May 2000.
- [3] Y. Azar, E. Horvitz, E. Lubetzky, Y. Peres, and D. Shahaf, "Tractable near-optimal policies for crawling," *Proc. Nat. Acad. Sci. USA*, vol. 115, no. 32, pp. 8099–8103, Aug. 2018.
- [4] A. Kolobov, Y. Peres, E. Lubetzky, and E. Horvitz, "Optimal freshness crawl under politeness constraints," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2019, pp. 495–504.
- [5] 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, pp. 1422–1430.

- [6] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 2731–2735.
- [7] M. Costa, M. Codreanu, and A. Ephremides, "Age of information with packet management," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2014, pp. 1583–1587.
- [8] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Optimizing data freshness, throughput, and delay in multi-server information-update systems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 2569–2573.
- [9] Q. He, D. Yuan, and A. Ephremides, "Optimizing freshness of information: On minimum age link scheduling in wireless systems," in *Proc. 14th Int. Symp. Model. Optim. Mobile, Ad Hoc, Wireless Netw. (WiOpt)*, May 2016, pp. 1–8.
- [10] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, "Age of information with a packet deadline," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 2564–2568.
- [11] 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.
- [12] E. Najm and E. Telatar, "Status updates in a multi-stream M/G/1/1 preemptive queue," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr. 2018.
- [13] E. Najm, R. Yates, and E. Soljanin, "Status updates through M/G/1/1 queues with HARQ," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 131–135.
- [14] A. Soysal and S. Ulukus, "Age of information in G/G/1/1 systems," in *Proc. Asilomar Conf.*, Nov. 2019, pp. 2022–2027.
- [15] A. Soysal and S. Ulukus, "Age of information in G/G/1/1 systems: Age expressions, bounds, special cases, and optimization," May 2019, *arXiv:1905.13743*. [Online]. Available: <http://arxiv.org/abs/1905.13743>
- [16] R. D. Yates, "The age of information in networks: Moments, distributions, and sampling," *IEEE Trans. Inf. Theory*, vol. 66, no. 9, pp. 5712–5728, Sep. 2020.
- [17] W. Gao, G. Cao, M. Srivatsa, and A. Iyengar, "Distributed maintenance of cache freshness in opportunistic mobile networks," in *Proc. IEEE 32nd Int. Conf. Distrib. Comput. Syst.*, Jun. 2012, pp. 132–141.
- [18] R. D. Yates, P. Ciblat, A. Yener, and M. Wigger, "Age-optimal constrained cache updating," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 141–145.
- [19] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, "Information freshness and popularity in mobile caching," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 136–140.
- [20] J. Zhong, R. D. Yates, and E. Soljanin, "Two freshness metrics for local cache refresh," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 1924–1928.
- [21] S. Zhang, J. Li, H. Luo, J. Gao, L. Zhao, and X. S. Shen, "Towards fresh and low-latency content delivery in vehicular networks: An edge caching aspect," in *Proc. 10th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2018, pp. 1–6.
- [22] H. Tang, P. Ciblat, J. Wang, M. Wigger, and R. Yates, "Age of information aware cache updating with file-and age-dependent update durations," in *Proc. IEEE WiOPT*, Jun. 2020, pp. 1–6.
- [23] L. Yang, Y. Zhong, F.-C. Zheng, and S. Jin, "Edge caching with real-time guarantees," 2019, *arXiv:1912.11847*. [Online]. Available: <http://arxiv.org/abs/1912.11847>
- [24] M. Bastopcu and S. Ulukus, "Maximizing information freshness in caching systems with limited cache storage capacity," in *Proc. Asilomar Conf.*, Nov. 2020.
- [25] M. Wang, W. Chen, and A. Ephremides, "Reconstruction of counting process in real-time: The freshness of information through queues," in *Proc. IEEE ICC*, Jul. 2019, pp. 1–6.
- [26] Y. Sun, Y. Polyanskiy, and E. Uysal-Biyikoglu, "Remote estimation of the Wiener process over a channel with random delay," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 321–325.
- [27] Y. Sun and B. Cyr, "Information aging through queues: A mutual information perspective," in *Proc. IEEE 19th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2018, pp. 1–5.
- [28] J. Chakravorty and A. Mahajan, "Remote estimation over a packet-drop channel with Markovian state," 2018, *arXiv:1807.09706*. [Online]. Available: <http://arxiv.org/abs/1807.09706>
- [29] B. T. Bacinoglu, E. T. Ceran, and E. Uysal-Biyikoglu, "Age of information under energy replenishment constraints," in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2015, pp. 25–31.
- [30] B. T. Bacinoglu and E. Uysal-Biyikoglu, "Scheduling status updates to minimize age of information with an energy harvesting sensor," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 1122–1126.
- [31] S. Feng and J. Yang, "Optimal status updating for an energy harvesting sensor with a noisy channel," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr. 2018, pp. 348–353.
- [32] S. Feng and J. Yang, "Minimizing age of information for an energy harvesting source with updating failures," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 2431–2435.
- [33] 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.
- [34] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor, "Age-minimal online policies for energy harvesting sensors with incremental battery recharges," in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2018, pp. 1–10.
- [35] A. Arafa, J. Yang, and S. Ulukus, "Age-minimal online policies for energy harvesting sensors with random battery recharges," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.
- [36] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor, "Age-minimal transmission for energy harvesting sensors with finite batteries: Online policies," *IEEE Trans. Inf. Theory*, vol. 66, no. 1, pp. 534–556, Jan. 2020.
- [37] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor, "Online timely status updates with erasures for energy harvesting sensors," in *Proc. 56th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2018, pp. 966–972.
- [38] A. Baknina and S. Ulukus, "Coded status updates in an energy harvesting erasure channel," in *Proc. 52nd Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2018, pp. 1–6.
- [39] A. Baknina, S. Ulukus, O. Ozel, J. Yang, and A. Yener, "Sending information through status updates," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 2271–2275.
- [40] A. Arafa and S. Ulukus, "Age minimization in energy harvesting communications: Energy-controlled delays," in *Proc. 51st Asilomar Conf. Signals, Syst., Comput.*, Oct. 2017, pp. 1801–1805.
- [41] A. Arafa and S. Ulukus, "Age-minimal transmission in energy harvesting two-hop networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–6.
- [42] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor, "Using erasure feedback for online timely updating with an energy harvesting sensor," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2019, pp. 607–611.
- [43] S. Farazi, A. G. Klein, and D. R. Brown, "Average age of information for status update systems with an energy harvesting server," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr. 2018, pp. 112–117.
- [44] S. Leng and A. Yener, "Age of information minimization for an energy harvesting cognitive radio," *IEEE Trans. Cognit. Commun. Netw.*, vol. 5, no. 2, pp. 427–439, Jun. 2019.
- [45] Z. Chen, N. Pappas, E. Björnson, and E. G. Larsson, "Age of information in a multiple access channel with heterogeneous traffic and an energy harvesting node," 2019, *arXiv:1903.05066*. [Online]. Available: <http://arxiv.org/abs/1903.05066>
- [46] R. Vishweshwar Bhat, R. Vaze, and M. Motani, "Throughput maximization with an average age of information constraint in fading channels," Mar. 2019, *arXiv:1911.07499*. [Online]. Available: <http://arxiv.org/abs/1911.07499>
- [47] J. Ostman, R. Devassy, G. Durisi, and E. Uysal, "Peak-age violation guarantees for the transmission of short packets over fading channels," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr. 2019, pp. 109–114.
- [48] S. Nath, J. Wu, and J. Yang, "Optimizing age-of-information and energy efficiency tradeoff for mobile pushing notifications," in *Proc. IEEE 18th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2017, pp. 1–5.
- [49] Y.-P. Hsu, "Age of information: Whittle index for scheduling stochastic arrivals," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 2634–2638.
- [50] I. Kadota, A. Sinha, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Scheduling policies for minimizing age of information in broadcast wireless networks," *IEEE/ACM Trans. Netw.*, vol. 26, no. 6, pp. 2637–2650, Dec. 2018.
- [51] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "Age and value of information: Non-linear age case," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 326–330.
- [52] M. Bastopcu and S. Ulukus, "Age of information with soft updates," in *Proc. 56th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2018, pp. 378–385.
- [53] M. Bastopcu and S. Ulukus, "Minimizing age of information with soft updates," *J. Commun. Netw.*, vol. 21, no. 3, pp. 233–243, Jun. 2019.

- [54] B. Buyukates, A. Soysal, and S. Ulukus, "Age of information scaling in large networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.
- [55] B. Buyukates, A. Soysal, and S. Ulukus, "Age of information scaling in large networks with hierarchical cooperation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.
- [56] M. Bastopcu and S. Ulukus, "Who should Google scholar update more often?" in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Jul. 2020, pp. 1–6.
- [57] J. Zhong, E. Soljanin, and R. D. Yates, "Status updates through multicast networks," in *Proc. 55th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2017, pp. 463–469.
- [58] B. Buyukates, A. Soysal, and S. Ulukus, "Age of information in two-hop multicast networks," in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 513–517.
- [59] B. Buyukates, A. Soysal, and S. Ulukus, "Age of information in multicast networks with multiple update streams," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 1977–1981.
- [60] B. Buyukates, A. Soysal, and S. Ulukus, "Age of information in multihop multicast networks," *J. Commun. Netw.*, vol. 21, no. 3, pp. 256–267, Jun. 2019.
- [61] J. Zhong and R. D. Yates, "Timeliness in lossless block coding," in *Proc. Data Compress. Conf. (DCC)*, Mar. 2016, pp. 339–348.
- [62] J. Zhong, R. D. Yates, and E. Soljanin, "Timely lossless source coding for randomly arriving symbols," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Nov. 2018, pp. 1–5.
- [63] P. Mayekar, P. Parag, and H. Tyagi, "Optimal lossless source codes for timely updates," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 1246–1250.
- [64] D. Ramirez, E. Erkip, and H. V. Poor, "Age of information with finite horizon and partial updates," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 4965–4969.
- [65] M. Bastopcu, B. Buyukates, and S. Ulukus, "Optimal selective encoding for timely updates," in *Proc. 54th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2020, pp. 1–6.
- [66] B. Buyukates, M. Bastopcu, and S. Ulukus, "Optimal selective encoding for timely updates with empty symbol," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2020, pp. 1–6.
- [67] M. Bastopcu and S. Ulukus, "Partial updates: Losing information for freshness," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2020, pp. 1–6.
- [68] J. Gong, Q. Kuang, X. Chen, and X. Ma, "Reducing age-of-information for computation-intensive messages via packet replacement," Jan. 2019, *arXiv:1901.04654*. [Online]. Available: <http://arxiv.org/abs/1901.04654>
- [69] B. Buyukates and S. Ulukus, "Timely distributed computation with stragglers," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5273–5282, Sep. 2020.
- [70] P. Zou, O. Ozel, and S. Subramaniam, "Trading off computation with transmission in status update systems," in *Proc. IEEE 30th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2019, pp. 1–6.
- [71] A. Arafat, R. D. Yates, and H. V. Poor, "Timely cloud computing: Preemption and waiting," in *Proc. 57th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2019, pp. 528–535.
- [72] M. Bastopcu and S. Ulukus, "Age of information for updates with distortion," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Aug. 2019, pp. 1–5.
- [73] M. Bastopcu and S. Ulukus, "Age of information for updates with distortion: Constant and age-dependent distortion constraints," 2020, *arXiv:1912.13493*. [Online]. Available: <http://arxiv.org/abs/1912.13493>
- [74] M. A. Abd-Elmagid and H. S. Dhillon, "Average peak age-of-information minimization in UAV-assisted IoT networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 2003–2008, Feb. 2019.
- [75] J. Liu, X. Wang, B. Bai, and H. Dai, "Age-optimal trajectory planning for UAV-assisted data collection," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr. 2018, pp. 553–558.
- [76] M. A. Abd-Elmagid, N. Pappas, and H. S. Dhillon, "On the role of age of information in the Internet of Things," *IEEE Commun. Mag.*, vol. 57, no. 12, pp. 72–77, Dec. 2019.
- [77] E. T. Ceran, D. Gunduz, and A. Gyorgy, "A reinforcement learning approach to age of information in multi-user networks," in *Proc. IEEE 29th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2018, pp. 1967–1971.
- [78] H. B. Beytur and E. Uysal-Biyikoglu, "Age minimization of multiple flows using reinforcement learning," in *IEEE ICNC*, Feb. 2019, pp. 339–343.
- [79] M. A. Abd-Elmagid, H. S. Dhillon, and N. Pappas, "A reinforcement learning framework for optimizing Age-of-Information in RF-powered communication systems," Aug. 2019, *arXiv:1908.06367*. [Online]. Available: <http://arxiv.org/abs/1908.06367>
- [80] S. Zhang, N. Zhang, X. Fang, P. Yang, and X. S. Shen, "Self-sustaining caching stations: Toward cost-effective 5G-enabled vehicular networks," *IEEE Commun. Mag.*, vol. 55, no. 11, pp. 202–208, Nov. 2017.
- [81] R. D. Yates and D. J. Goodman, *Probability and Stochastic Processes*. Hoboken, NJ, USA: Wiley, 2014.
- [82] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [83] D. P. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1989.
- [84] J. A. O'Sullivan, "Alternating minimization algorithms: From Blahut-Arimoto to expectation-maximization," in *Codes, Curves, and Signals*, vol. 485. Boston, MA, USA: Springer, 1998, pp. 173–192.
- [85] A. Yener, R. D. Yates, and S. Ulukus, "Interference management for CDMA systems through power control, multiuser detection, and beamforming," *IEEE Trans. Commun.*, vol. 49, no. 7, pp. 1227–1239, Jul. 2001.



Melih Bastopcu (Graduate Student Member, IEEE) received the B.S. degree in electrical and electronics engineering from Bilkent University, Turkey, in 2016. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Maryland, College Park. His research interests include age of information, resource allocation, low latency systems, and communication network design.



Sennur Ulukus (Fellow, IEEE) received the B.S. and M.S. degrees in electrical and electronics engineering from Bilkent University, and the Ph.D. degree in electrical and computer engineering from the Wireless Information Network Laboratory (WINLAB), Rutgers University.

She is currently the Anthony Ephremides Professor of information sciences and systems with the Department of Electrical and Computer Engineering, University of Maryland, 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 with AT&T Labs-Research. Her research interests include 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 also 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, a 2005 NSF CAREER Award, the 2010–2011 ISR Outstanding Systems Engineering Faculty Award, and the 2012 ECE George Corcoran Outstanding Teaching Award. She was a Distinguished Lecturer of the IEEE Information Theory Society for 2018–2019. She is also a TPC Chair/Co-Chair of 2021 IEEE Globecom, 2019 IEEE ITW, 2017 IEEE ISIT, 2016 IEEE Globecom, 2014 IEEE PIMRC, and 2011 IEEE CTW. She has been an Area Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS since 2019 and a Senior Editor of the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING since 2020. She has been an Area Editor of the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING since 2016, an Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS—Series on Green Communications and Networking from 2015 to 2016, an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY from 2007 to 2010, and an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS from 2003 to 2007. She was a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS in 2008 and 2015, the IEEE TRANSACTIONS ON INFORMATION THEORY in 2011, and *Journal of Communications and Networks* in 2012.