Timely Updates with Priorities: Lexicographic Age Optimality

Ali Maatouk*, Yin Sun§, Anthony Ephremides†, and Mohamad Assaad*

*Laboratoire des Signaux et Systèmes, CentraleSupélec, Gif-sur-Yvette, France

§Department of ECE, Auburn University, Auburn, AL 36849

†ECE Dept., University of Maryland, College Park, MD 20742

Abstract

In this paper, we consider a scheduling problem, in which several streams of status update packets with different priority levels are sent through a shared channel to their destinations. We introduce a notion of *lexicographic age optimality*, or simply *lex-age-optimality*, to evaluate the performance of multi-class status update policies. In particular, a lex-age-optimal scheduling policy first minimizes the Age of Information (AoI) metrics for high-priority streams, and then, within the set of optimal policies for high-priority streams, achieves the minimum AoI metrics for low-priority streams. We propose a new scheduling policy named Preemptive Priority, Maximum Age First, Last-Generated, First-Served (PP-MAF-LGFS), and prove that the PP-MAF-LGFS scheduling policy is lex-age-optimal. This result holds (i) for minimizing any time-dependent, symmetric, and non-decreasing age penalty function; (ii) for minimizing any non-decreasing functional of the stochastic process formed by the age penalty function; and (iii) for the cases where different priority classes have distinct arrival traffic patterns, age penalty functions, and age penalty functionals. For example, the PP-MAF-LGFS scheduling policy is lex-age-optimal for minimizing the probability of age violation of a high-priority stream and the time-average age of a low-priority stream. Numerical results are provided to illustrate our theoretical findings.

This paper was presented in part at the 18th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt 2020) [1].

This work has been supported by ONR N000141812046, NSF CCF1813078, NSF CNS1551040, NSF CCF1420651, and ONR N00014-17-1-2417. Yin Sun's work has been supported in part by NSF grant CCF-1813050, ONR grant N00014-17-1-2417, and ARO grant W911NF-21-1-0244.

I. Introduction

Due to the proliferation of cheap hardware, remote monitoring has been adopted in many cyber-physical systems. In these applications, a monitor is interested in timely updates about the status of a remote system. These status updates range from the position and velocity of vehicles in autonomous driving to the temperature and humidity levels of a certain area in environmental monitoring. In [2], the Age of Information (AoI) was proposed to measure the timeliness of status updates. Due to its widespread application range and its ability to quantify the freshness of information, the AoI has attracted a significant surge of interest in recent years [2]–[16]. In particular, transmission scheduling of multiple update streams in both centralized and distributed settings has been explored in [5], [17]–[26]. For example, the authors in [17] proposed both age-optimal and near age-optimal scheduling policies for the single and multi-server cases, respectively (we refer the readers to [27, Chapter 2]).

In a variety of real-life applications, information streams are assigned different priorities based on how crucial and time-sensitive their data are. A simple example is a vehicular network where data can be divided into two categories: crucial safety data and non-safety-related information. As the former is more time-sensitive than the latter, it should always be given a higher priority by the service facility [28]. Accordingly, priority-based scheduling problems have been extensively studied in the queuing theory literature for different performance measures (e.g., delay, and throughput). In [29], a notion of lexicographic optimality, or simply, lex-optimality, was introduced for throughput maximization in multi-class scheduling scenarios. The idea of lex-throughput-optimality is to first find a class of optimal scheduling policies Π_{opt} that maximize the throughput of a high priority class, and then find the optimal scheduling policies within Π_{opt} that maximize the throughput of the low priority class. Therefore, it provides high priority streams the best possible service, and meanwhile, optimizes the performance of the low priority streams without affecting that of high priority ones.

There are few recent studies on status updates with multiple priority classes. In [30], multiple information streams with distinct priorities are processed at a common service facility. The facility can have one waiting room for storing a status update packet and the waiting room is shared by all the streams. The authors studied a case where low priority packets are preempted by high priority ones and the preempted packets are discarded. Using a tool named Stochastic Hybrid Systems (SHS), the authors found an expression for the average age of each stream. The arrival rate of each stream was then optimized accordingly. In another work [31], the authors investigated the same settings of [30] but by letting each stream have its own buffer space. Most recently, closed forms of the average Peak Age of Information (PAoI) were found in M/M/1/1 settings where streams are assigned different priorities [32]. Existing research efforts have been focused on finding closed-form expressions of the average AoI/PAoI in a particular scenario and for a specific arrival traffic model, providing insights on the performance of the system. Nonetheless, how to optimize the age performance of status updates with multiple priorities remains to be a fundamental open question. In this paper, we provide an answer to this question. The key contributions of this paper are summarized as follows: :

- We introduce the notion of lexicographic optimality for age minimization, which we will refer to as the lex-age-optimality. The lex-age-optimality captures both the age-optimality and priorities in a multi-class scheduling for minimizing the AoI. It guarantees that the age performance of low priority streams is optimized while ensuring that high priority streams are not affected.
- In the case of a single server with i.i.d. exponential service times, we propose a scheduling policy named *Preemptive Priority, Maximum Age First, Last-Generated, First-Served* (PP-MAF-LGFS). Using a sample-path argument, we show that this policy is lex-age-optimal. Our lex-age-optimality results are not constrained to the minimization of the average AoI previously adopted in [30]–[32]. In fact, they hold for (i) minimizing any time-dependent, symmetric, and non-decreasing penalty function of the ages, and (ii) minimizing any non-decreasing functional of the age penalty process. Unlike the previous works on multi-class status updates [30], [32], our lex-age-optimality results are not limited to a given traffic arrival process. Moreover, they hold when the priority classes have distinct traffic patterns and age penalty functionals. For example, we could be interested in minimizing the probability of age violation for a class and the average AoI for another.

The rest of the paper is organized as follows: Section II is dedicated to the system model where the required definitions and the queuing model are presented. In Section III, we introduce the notion of lex-age-optimality and propose a lex-age-optimal policy in the single exponential server settings. Numerical results that corroborate these findings are laid out in Section IV while the paper is concluded in Section V.

II. System Model

A. Notations and Definitions

We let x and $\boldsymbol{x}$ denote deterministic scalars and vectors, respectively. Similarly, we will use X and $\boldsymbol{X}$ to denote random scalars and vectors, respectively. Let x^i denote the i-th element of vector $\boldsymbol{x}$, and let $x^{[i]}$ denote the i-th largest element of vector $\boldsymbol{x}$. Hence, $x^{[1]}$ and $x^{[N]}$ denote the largest and smallest elements of vector $\boldsymbol{x}$, respectively. We denote by $[\boldsymbol{x}]$ the sorted version of vector $\boldsymbol{x}$ (i.e. $[\boldsymbol{x}]^i = x^{[i]}$). Vector $\boldsymbol{x} \in \mathbb{R}^N$ is said to be smaller than $\boldsymbol{y} \in \mathbb{R}^N$, denoted by $\boldsymbol{x} \leq \boldsymbol{y}$, if $x^i \leq y^i$ for $i = 1, \ldots, N$. The composition of two functions f and g is denoted by $f \circ g(\boldsymbol{x}) = f(g(\boldsymbol{x}))$. A function $p : \mathbb{R}^N \mapsto \mathbb{R}$ is said to be symmetric if $p(\boldsymbol{x}) = p([\boldsymbol{x}])$ for all $\boldsymbol{x} \in \mathbb{R}^N$. Next, we define stochastic ordering, which we will use in subsequent age-optimality analysis.

Definition 1. *Stochastic Ordering of Random Variables* [33]: A random variable X is said to be stochastically smaller than another random variable Y, denoted by $X \leq_{st} Y$, if $\Pr(X > t) \leq \Pr(Y > t)\ \forall t \in \mathbb{R}$.

Definition 2. *Stochastic Ordering of Random Vectors* [33]: A set $\mathscr{U} \subseteq \mathbb{R}^N$ is called upper if $\boldsymbol{y} \in \mathscr{U}$ whenever $\boldsymbol{x} \leq \boldsymbol{y}$ and $\boldsymbol{x} \in \mathscr{U}$. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two n-dimensional random vectors, $\boldsymbol{X}$ is said to be stochastically smaller than $\boldsymbol{Y}$, denoted by $\boldsymbol{X} \leq_{st} \boldsymbol{Y}$, if

$$\Pr(\boldsymbol{X} \in \mathscr{U}) \leq \Pr(\boldsymbol{Y} \in \mathscr{U}) \quad \text{for any upper set } \mathscr{U} \subseteq \mathbb{R}^N. \tag{1}$$

Definition 3. *Stochastic Ordering of Stochastic Processes* [33]: A stochastic process $\{X(t), t \geq 0\}$ is said to be stochastically smaller than another stochastic process $\{Y(t), t \geq 0\}$, denoted by $\{X(t), t \geq 0\} \leq_{st} \{Y(t), t \geq 0\}$, if for any sequence of time instants $t_1 < t_2 < \ldots < t_m \in \mathbb{R}^+$

$$(X(t_1), X(t_2), \ldots, X(t_m)) \leq_{st} (Y(t_1), Y(t_2), \ldots, Y(t_m)). \tag{2}$$

Let $\mathbb{V}$ be the set of Lebesgue measurable functions on $[0, \infty)$, i.e.,

$$\mathbb{V} = \{g : [0, \infty) \mapsto \mathbb{R} \text{ is Lebesgue measurable}\}. \tag{3}$$

A functional $\phi : \mathbb{V} \mapsto \mathbb{R}$ is said to be non-decreasing if $\phi(g_1) \leq \phi(g_2)$ holds for all $g_1, g_2 \in \mathbb{V}$ that satisfy $g_1(t) \leq g_2(t)$ for $t \in [0, \infty)$. We note that $\{X(t), t \geq 0\} \leq_{st} \{Y(t), t \geq 0\}$ if, and only if, [33]

$$\mathbb{E}[\phi(\{X(t), t \geq 0\})] \leq \mathbb{E}[\phi(\{Y(t), t \geq 0\})] \tag{4}$$

holds for every non-decreasing functional ϕ for which the expectations in (4) exist.

B. Queuing Model

Consider the status-update system illustrated in Fig. 1, where N streams of status update packets are sent through a shared service facility. Each update stream has a queue, which could have an infinite or finite buffer space. The server can process at most one packet at a time. The packet service times are i.i.d. across streams and packets. The packet streams are divided into I priority classes, with streams from the same class having the same priority. Each stream is indexed by two numbers (i, j), where i is the class index and j is the stream index within class i. The classes are indexed in a decreasing order of priority. In other words, classes 1 and I are the highest and lowest priority classes, respectively. Let J_i be the number of steams in class i. Let $s_{i,j}$ and $d_{i,j}$ denote the source and destination nodes of stream (i, j), respectively. Different streams may have different source and/or destination nodes.

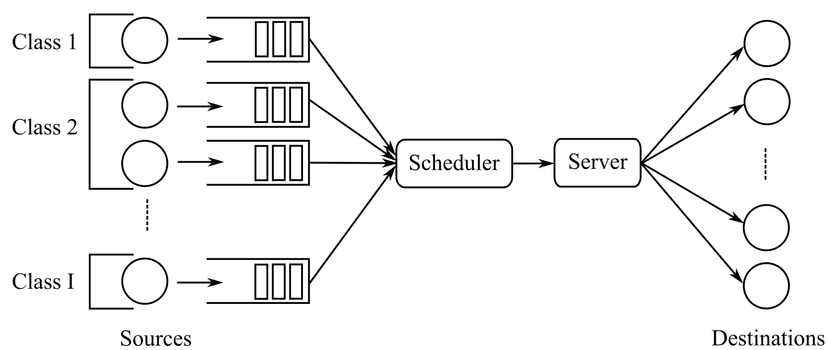


Fig. 1: System model.

The system starts operating at time $t = 0$. The n-th update packet of stream (i, j) is generated at time $S_n^{i,j}$, arrives to the stream's buffer at time $A_n^{i,j}$, and is delivered to the destination $d_{i,j}$ at time $D_n^{i,j}$. Accordingly, we always have $0 \leq S_1^{i,j} \leq S_2^{i,j} \leq \ldots$ and $S_n^{i,j} \leq A_n^{i,j} \leq D_n^{i,j}$. We consider the following synchronized packet generation and arrival processes.

Definition 4. *Intra-class Synchronized Generations and Arrivals*: The packet generation and arrival times are said to be synchronized across streams within each class, if for each class $i = 1, \ldots, I$, there exist two sequences $\{S_1^i, S_2^i, \ldots\}$ and $\{A_1^i, A_2^i, \ldots\}$ such that for all $n = 1, 2, \ldots$ and $j = 1, \ldots, J_i$

$$S_n^{i,j} = S_n^i, \quad A_n^{i,j} = A_n^i. \tag{5}$$

Note that we let each class have its unique traffic pattern as we do not impose *inter-class* synchronization. In practice, the synchronization between streams within each class occurs when these streams are synchronized by the same clock, as in monitoring and control applications [34], [35]. Another example is a vehicular network where safety-related data (e.g., position and velocity) are generated every T time units, while other data of lower priority can have a different traffic pattern (e.g., updates on the traffic are generated every T' time units) [28]. It is important to note that, if $J_i = 1$, there is only one stream in class i and the stream could have *arbitrary* packet generation and arrival processes. For example, our study holds even for out-of-order packet arrivals, i.e., $S_n^i < S_{n+1}^i$ but $A_n^i > A_{n+1}^i$, which were not allowed in the models of [30], [32]. In the sequel, we let

$$\mathscr{I} = \{(S_n^i, A_n^i),\ i = 1, \ldots, I, \quad n = 1, 2, \ldots\} \tag{6}$$

denote the sequence of generation/arrival times for all the classes of streams. We suppose that $\mathscr{I}$ is independent of the service times of the packets and is not altered by the choice of the scheduling policy.

Let π represents a scheduling policy that determines which packets from which streams to send over time. Let Π denotes the set of all *causal* scheduling policies, in which the decisions are made without using any future knowledge. A policy is said to be *work-conserving* if the service facility is kept busy whenever there exists one or more unserved packets in the queues. We let Π_{wc} denote the set of work-conserving causal policies. A policy is said to be *preemptive* if it allows the service facility to switch to transmitting another packet at any time.

C. Age Penalty Functions and Functionals

The age of information of stream (i, j) at time t is:

$$\Delta^{i,j}(t) = t - \max\{S_n^{i,j} : D_n^{i,j} \leq t,\ n = 1, 2, \ldots\}, \tag{7}$$

which is the difference between the current time t and the generation time of the freshest packet that has been delivered to the destination $d_{i,j}$. Let $\boldsymbol{\Delta}^i(t) = (\Delta^{i,1}(t), \ldots, \Delta^{i,J_i}(t))$ denote the age vector at time t of all streams belonging to class i, and let $\boldsymbol{\Delta}(t) = (\boldsymbol{\Delta}^1(t), \ldots, \boldsymbol{\Delta}^I(t))$ denote the age vector of all streams at time t.

We introduce an age penalty function $p_t \circ \boldsymbol{\Delta}^i(t)$ that represents the level of dissatisfaction with the aged information of class i at time t, where $p_t : \mathbb{R}^{J_i} \mapsto \mathbb{R}$ is a non-decreasing function of $\boldsymbol{\Delta}^i(t)$. Some commonly used age penalty functions are listed below.

- The sum age of the J_i streams:

$$p_{\text{sum}} \circ \mathbf{\Delta}^i(t) = \sum_{j=1}^{J_i} \Delta^{i,j}(t). \tag{8}$$

- The maximum age of the J_i streams:

$$p_{\text{max}} \circ \mathbf{\Delta}^i(t) = \max_{j=1,\ldots,J_i} \Delta^{i,j}(t). \tag{9}$$

- The average age threshold violation of the J_i streams:

$$p_{\text{exceed}-\alpha} \circ \mathbf{\Delta}^i(t) = \frac{1}{J_i} \sum_{j=1}^{J_i} \mathbb{1}_{\{\Delta^{i,j}(t)>\alpha\}}. \tag{10}$$

  where $\mathbb{1}_{\{.\}}$ is the indicator function, and α is a fixed age threshold that should not be violated.
- The sum age penalty function of the J_i streams:

$$p_{\text{pen}} \circ \mathbf{\Delta}^i(t) = \sum_{j=1}^{J_i} g(\Delta^{i,j}(t)), \tag{11}$$

  where $g : \mathbb{R}^+ \mapsto \mathbb{R}$ is a non-decreasing function. For instance, an exponential function $g(\Delta^{i,j}) = \exp(a\Delta^{i,j})$ with $a > 0$ can be used for control applications where the system is vulnerable to outdated information and the need for fresh information grows quickly with respect to the age [11], [36].

We focus on the family of symmetric and non-decreasing penalty functions:

$$\mathscr{P}_{\text{sym}} = \{p : [0,\infty)^N \mapsto \mathbb{R} \text{ is symmetric and non-decreasing}\}.$$

This class of penalty functions $\mathscr{P}_{\text{sym}}$ is fairly large, and include the age penalty functions in (8)-(11). Furthermore, we point out that p_t can change over time, which represents the time-variant importance of the information streams. This highlights the generality of the class of penalty functions that we consider.

In addition to age penalty functions, we use non-decreasing functionals $\phi(\{p_t \circ \mathbf{\Delta}^i(t), t \geq 0\})$ of the age penalty process $\{p_t \circ \mathbf{\Delta}^i(t), t \geq 0\}$ to represent the level of dissatisfaction with the aged information of class i, which is called age penalty functionals. One example of age penalty functionals is the time-average age penalty:

$$\phi_{\text{avg}}(\{p_t \circ \mathbf{\Delta}^i(t), t \geq 0\}) = \frac{1}{T} \int_0^T p_t \circ \mathbf{\Delta}^i(t) dt. \tag{12}$$

Different priority classes can have distinct age penalty functions and functionals. This is important because each priority class typically represents a different application and has its own data timeliness requirements. For example, in vehicular networks, time-crucial safety data related to vehicle position should be delivered promptly. Typically, the system performance is affected by the probability of violation of an age threshold. Accordingly, we can choose the excess age penalty function $p_{\text{exceed}-\alpha}$ and the time-average age penalty functional ϕ_{avg} for this class of traffic. On the other hand, for example, we can choose the penalty function p_{sum} and the time-average age penalty functional ϕ_{avg} for the updates on gas tank levels. This further highlights the generality of our considered framework.

In the sequel, we use $\{\mathbf{\Delta}_\pi^i(t), t \geq 0\}$ and $\{p_t \circ \mathbf{\Delta}_\pi^i(t), t \geq 0\}$ to represent the stochastic age process and age penalty process of class i, respectively, when policy π is adopted. We assume that the initial age $\mathbf{\Delta}_\pi(0^-)$ at time $t = 0^-$ is the same for all $\pi \in \Pi$.

III. Multi-class Multi-stream Scheduling

A. Lexicographic Age Optimality

In the sequel, we introduce the notion of lexicographic optimality for age minimization, which is referred to as the lex-age-optimality. Recent efforts on age analysis in multi-class systems focused mainly on finding closed-form expressions of the average AoI/PAoI a particular scheduling policy [30]–[32]. Note that, in the multi-class system case, the minimization of the total average AoI/PAoI falls short in capturing the differences in time-cruciality between the classes. In fact, this approach lets each stream contribute equally to the penalty of the system regardless of its class. As it will be shown below, the lex-age-optimality not only solves this issue but also provides a new research direction for age analysis in multi-class timely status-update systems.

Definition 5. *Lex-age-optimality*: A scheduling policy $P \in \Pi$ is said to be level 1 lex-age-optimal within Π if for all $\mathscr{I}$, $p_t \in \mathscr{P}_{\text{sym}}$ and $\pi \in \Pi$

$$[\{p_t \circ \mathbf{\Delta}_P^1(t), t \geq 0\}|\mathscr{I}] \leq_{st} [\{p_t \circ \mathbf{\Delta}_\pi^1(t), t \geq 0\}|\mathscr{I}]. \tag{13}$$

We let $\Pi^1_{\text{lex-opt}} \subseteq \Pi$ denote the set of scheduling policies that are level 1 lex-age-optimal. In addition, policy P is said to be level k lex-age-optimal for $k = 2, \ldots, I$ if it is level $k-1$ lex-age-optimal, and for all $\mathscr{I}$, $p_t \in \mathscr{P}_{\text{sym}}$ and $\pi \in \Pi^{k-1}_{\text{lex-opt}}$

$$[\{p_t \circ \mathbf{\Delta}_P^k(t), t \geq 0\}|\mathscr{I}] \leq_{st} [\{p_t \circ \mathbf{\Delta}_\pi^k(t), t \geq 0\}|\mathscr{I}], \tag{14}$$

where $\Pi_{\text{lex-opt}}^{k-1}$ is the set of scheduling policies that are level $k-1$ lex-age-optimal. If policy P is level k lex-age-optimal for all $k = 1, \ldots, I$, it is said to be lex-age-optimal.

According to (4), (13) can be equivalently expressed as

$$\begin{aligned}&\mathbb{E}[\phi(\{p_t \circ \mathbf{\Delta}_P^1(t), t \geq 0\})|\mathscr{I}]\\ =&\min_{\pi \in \Pi} \mathbb{E}[\phi(\{p_t \circ \mathbf{\Delta}_\pi^1(t), t \geq 0\})|\mathscr{I}],\end{aligned} \tag{15}$$

for all $\mathscr{I}$, $p_t \in \mathscr{P}_{\text{sym}}$, and non-decreasing functional $\phi : \mathbb{V} \mapsto \mathbb{R}$, provided that the expectations in (15) exist. Similarly, an equivalent formulation of the level k lex-age-optimality (14) of a policy $P \in \Pi_{\text{lex-opt}}^{k-1}$ is

$$\begin{aligned}&\mathbb{E}[\phi(\{p_t \circ \mathbf{\Delta}_P^k(t), t \geq 0\})|\mathscr{I}]\\ =&\min_{\pi \in \Pi_{\text{lex-opt}}^{k-1}} \mathbb{E}[\phi(\{p_t \circ \mathbf{\Delta}_\pi^k(t), t \geq 0\})|\mathscr{I}],\end{aligned} \tag{16}$$

for all $\mathscr{I}$, $p_t \in \mathscr{P}_{\text{sym}}$, and non-decreasing functional $\phi : \mathbb{V} \mapsto \mathbb{R}$, provided that the expectations in (16) exist.

The goal of the lex-age-optimality is to first guarantee the age-optimality of high priority classes, and then optimize the age performance of the low priority classes accordingly. To see how this is achieved, we recall from (15) that a level 1 lex-age-optimal policy P achieves the smallest possible expected value of any non-decreasing functional ϕ of the stochastic age penalty process $[\{p_t \circ \mathbf{\Delta}^1(t), t \geq 0\})|\mathscr{I}]$ among all causal policies. Next, to maintain the age-optimality of the highest priority class, our attention is restricted to the set of scheduling policies that are level 1 lex-age-optimal. We have denoted this set by $\Pi_{\text{lex-opt}}^1$. To that end, and as seen in (16), a policy P is level 2 lex-age-optimal if it achieves the smallest possible expected value of any non-decreasing functional ϕ of the stochastic age penalty process $[\{p_t \circ \mathbf{\Delta}^2(t), t \geq 0\})|\mathscr{I}]$ among all level 1 lex-age-optimal policies. This showcases how the lex-age-optimality captures the time-cruciality of streams since, by definition, lex-age-optimal policies grant high priority streams the best possible age performance without being influenced by low priority streams. Then, while ensuring the age-optimality of the high priority streams, the age performance of the low priority streams is optimized.

B. Lex-Age-Optimal Scheduling

We consider the case where the service time of each packet is exponentially distributed with service rate μ. To address this multi-stream scheduling problem, we first lay out the notion of informative packets.

Definition 6. *Informative and Non-informative Packets*: Consider a packet of stream (i, j) that is generated at time $S_n^{i,j} \leq t$. The packet is said to be informative (non-informative) at time t if $t - S_n^{i,j} < \Delta^{i,j}(t)$ $(t - S_n^{i,j} \geq \Delta^{i,j}(t))$, i.e., the age of the packet is (not) smaller than $\Delta^{i,j}(t)$.

Equipped with the above definition, we consider in the following several scheduling disciplines that are based on informative packets.

Definition 7. *Preemptive Priority (PP) policy based on Informative Packets*: Among the streams with informative packets, the streams with the highest priority are served first. When an informative packet arrives, if it has a higher priority than the packet that is being served, it will preempt the packet under service. If the preempted packet is informative, it is stored back to the queue; if the preempted packet is non-informative, it can be either stored back to the queue or dropped.

Definition 8. *Maximum Age First (MAF) policy*: Among the streams from the same priority class, the stream with the maximum age is served first, with ties broken arbitrarily.

Definition 9. *Last-Generated, First-Served (LGFS) policy*: Among the informative packets from a stream, the last generated informative packet is served first, with ties broken arbitrarily.

By combining the above three service disciplines, we propose a new scheduling policy called Preemptive Priority, Maximum Age First, Last-Generated, First-Served (PP-MAF-LGFS), which is defined as follows.

Definition 10. *Preemptive Priority, Maximum Age First, Last-Generated, First-Served*: This policy is preemptive, work-conserving and obeys the following set of scheduling rules:

- If there exist informative packets, the system will serve an informative packet that is selected as follows
  - among all streams with informative packets, pick the class of streams with the highest priority;
  - among the streams from the selected priority class, pick the stream with the maximum age, with ties broken arbitrarily;
  - among the informative packets from the selected stream, pick the last generated informative packet, with ties broken arbitrarily;

- if there exists no informative packet, the system can serve any non-informative packet.

As can be seen above, a crucial aspect of the policy is that informative packets are not dropped, which will be shown later to be essential for lex-age-optimality. Because of this, the packet management strategies in [10] are not suitable for multi-class age-based scheduling. As for non-informative packets, our proposed policy does not drop them as well. Although these packets are not necessary for reducing the age, but in many applications, they may still be needed at the monitor (e.g., social updates). Note that in the case of a single priority class (i.e., $I = 1$), the proposed policy coincides with the preemptive Maximum Age First, Last-Generated, First-Served (MAF-LGFS) scheduling policy proposed in [17].

By definition, our policy ensures that the service of high priority informative packets is not interrupted or influenced by any lower priority packets. This grants crucial timely packets the best possible service. Note that informative packets play a key role in our policy. In particular, the preemptive priority discipline that we consider is a dynamic priority rule in which the priority level depends on the existence of informative packets: If a stream from class 1 has informative packets, the stream has the highest priority; otherwise, if the stream does not have any informative packets, the stream has the lowest priority, even lower than the streams in the lowest class I that have informative packets. This non-trivial aspect of our policy ensures that low priority classes are provided with the best possible opportunity for transmission while not affecting the age of the high priority streams. On another note, our policy ensures that the freshest packets from the selected priority class are delivered first. These key observations are crucial and will be used to establish the lex-age-optimality of the PP-MAF-LGFS policy.

Theorem 1 (Lex-age-optimality of PP-MAF-LGFS)**.** *If (i) the packet generation and arrival times are synchronized across streams within each class, and (ii) the packet service times are exponentially distributed and i.i.d. across streams and time, then the policy* PP-MAF-LGFS *is lex-age-optimal.*

Proof: This theorem is proven by using an inductive sample-path comparison. Specifically, we show by induction that the set of scheduling rules that the PP-MAF-LGFS policy satisfies are sufficient and necessary for level k lex-age-optimality for $k = 1, \ldots, I$. Contrary to previous sample-path proofs in the literature (e.g., in [17]), showing these scheduling rules are sufficient for optimality is not enough for proving Theorem 1. In fact, in order to prove level $k+1$ lex-age-optimality, one need to exactly characterize all the level k lex-age-optimal scheduling policy in $\Pi^k_{\text{lex-opt}}$. This requirement poses several technical difficulties, which are solved in our proof by showing the necessity of the scheduling rules in Definition 10 for level k lex-age-optimality. The details can be found in Appendix A. ■

Given the above theorem, we can also deduce the following corollary.

Corollary 1. *If (i) the packet service times are exponentially distributed and i.i.d. across streams and time, and (ii) each priority class has only one stream, then the* PP-MAF-LGFS *is lex-age-optimal for* arbitrary *packet generation and arrival processes.*

To the best of our knowledge, this is the first lex-age-optimality result for multi-class status updates. Our result is strong as the optimality is established in terms of stochastic ordering of stochastic processes for all symmetric non-decreasing penalty functions, and for all non-decreasing age penalty functionals. What makes these results further interesting is that the priority classes can have different traffic patterns, age penalty functions, and age penalty functionals. As it was previously explained in Section II-C, this is of paramount importance as priority classes typically represent different applications, each with their own traffic arrivals and data timeliness requirements. For example, in a certain scenario, we can be interested in minimizing the the time-average sum-age for class 1 and the probability of age threshold violation for class 2. Theorem 1 guarantees that our proposed policy achieves the required data timeliness goal for any of these cases, despite the differences in age penalty functions and functionals between the classes.

IV. Numerical Results

We consider a vehicle in a V2X (Vehicle-To-Everything) network that sends packets to nearby vehicles and roadside units (see [28], [37] for two surveys). In the aforementioned surveys, a list of possible packets use cases are presented, each of which having multiple priorities in the status-updating network. We pick 3 data categories in our simulations:

1) **Road Safety Data**: These are the data primarily employed to reduce the number of traffic accidents. These packets are generated periodically with a minimum frequency of $10\,\text{Hz}$. We assume in our settings that the packets' generation frequency is set to $10\,\text{Hz}$. This class of streams has the highest priority among all data types. We consider in our simulations that two streams belong to this class (e.g., the vehicle's position and speed).

2) **Traffic Management Data**: The goal of these data is to optimize the traffic stream and reduce the travel time in the network. We consider in our simulations that one stream belongs to this class (e.g., updates concerning the destination of the vehicle). The generation frequency of these packets is set to $1\,\mathrm{Hz}$. The priority of this class is secondary to road safety data.
3) **Convenience and Entertainment Data**: The data in this class are considered to be the least crucial as their aim is to provide entertainment and convenience solely for improving the quality of travel. We consider in our simulations that two streams belong to this class and we suppose that the generation frequency of their packets is $5\,\mathrm{Hz}$.

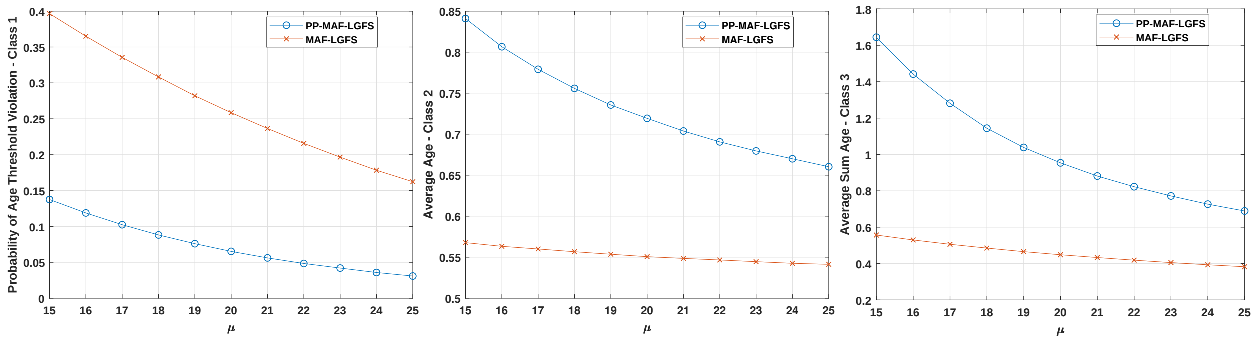
Based on the above, we can conclude that the arrival rate to our considered system is $\lambda_{tot} = 31$ packets per second. The service facility of the vehicle is supposed to be constituted of 1 server with the transmission times being i.i.d. across streams and time. Moreover, the transmission times are considered to be exponentially distributed with service rate μ.

We compare our proposed policy to the preemptive MAF-LGFS[1] policy proposed in [17]. The preemptive MAF-LGFS policy schedules the packet of the stream with the highest age, regardless of the class it belongs to. As for the age penalty function and functional for each class, we choose $p_{\text{exceed}-\alpha}$ and ϕ_{avg} as the age penalty function and functional for class 1 respectively, where α is set to $250\,\mathrm{ms}$. By doing so, we get

$$\begin{aligned}&\mathbb{E}[\phi_{\text{avg}}(\{p_{\text{exceed}-\alpha} \circ \boldsymbol{\Delta}^1(t), t \geq 0\})]\\ =&\frac{1}{2}\sum_{j=1}^{2}\frac{1}{T}\int_0^T \Pr(\Delta^{1,j}(t) > \alpha)dt,\end{aligned} \tag{17}$$

where $\Pr(\Delta^{1,j}(t) > \alpha)$ is the probability of violation of the maximum tolerated age $250\,\mathrm{ms}$ by stream $(1, j)$ at time t. The interest in this time-average age penalty function is that in vehicular networks, small age for the velocity and position data can be tolerated but, after a certain value, the performance of the system starts deteriorating due to this aging. For class 2, we choose ϕ_{avg} as the age penalty functional. Lastly, we choose p_{sum} and ϕ_{avg} for class 3. We iterate over a range of the service rate μ and we run the simulations for $10^5\,\mathrm{s}$. We report in Fig. 2 the simulations results that showcase the performance of each policy. We can conclude from these results the following:

[1]First-Come-First-Served (FCFS) policies are omitted from our simulations as they will always be outperformed by LGFS policies since queuing will lead to unnecessary staleness of the packets.



(a) Probability of age violation for class 1.

(b) Time-averaged age of class 2.

(c) Time-averaged sum age of class 3.

Fig. 2: Comparison between the two policies in function of the service rate μ.

- As seen in Fig. 2a, our proposed policy always outperforms the preemptive MAF-LGFS policy for class 1 at any service rate. Specifically, the probability of the age threshold violation by the preemptive MAF-LGFS policy is 3 times higher than the one achieved by our policy. This is a consequence of our proposed policy's goal as it gives priority to minimizing the time-average age penalty of class 1 regardless of the other remaining classes.
- On the other hand, we can see in Fig. 2b-2c that the preemptive MAF-LGFS policy outperforms our proposed policy for classes 2 and 3. In fact, in our policy, giving the priority to class 1 leads to a penalty for the remainder of the classes. However, we recall that the probability of violation of the age threshold in class 1 for our policy is 3 times less than the preemptive MAF-LGFS. Accordingly, the penalty incurred by the remaining classes is justified. Moreover, we can see that as μ increases, the gap between the two curves in both figures shrinks. The reason behind this is that class 1's packets finish transmission much faster the higher μ is. Consequently, in our proposed policy, the server will be able to finish serving class 1 fast enough that it can start serving the other classes before new packets for class 1 arrive to the system. This reduces the incurred penalty by the low priority classes due to the presence of the high priority streams.

The above results highlight the performance of our proposed lex-age-optimal policy, and showcase how it grants high priority streams the best possible age performance while optimizing the age performance of the low priority classes accordingly.

V. Conclusion

In this paper, we have introduced the notion of lex-age-optimality that captures both the age-optimality and priorities of the streams in a general multi-class priority-based scheduling scenario. To that end, we have proposed a scheduling policy in a general multi-class, multi-stream scheduling scenario named Preemptive Priority, Maximum Age First, Last-Generated, First-Served (PP-MAF-LGFS). Using a sample-path argument, we were able to prove the lex-age-optimality of the PP-MAF-LGFS policy in the single exponential server case for any symmetric non-decreasing penalty function, and for all non-decreasing age penalty functionals. Numerical results were then presented to highlight the performance of our proposed policy.

References

[1] A. Maatouk, Y. Sun, A. Ephremides, and M. Assaad, “Status updates with priorities: Lexicographic optimality,” in *2020 18th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*, 2020, pp. 1–8.

[2] S. Kaul, R. Yates, and M. Gruteser, “Real-time status: How often should one update?” in *Proceedings IEEE INFOCOM*, March 2012, pp. 2731–2735.

[3] R. Talak, S. Karaman, and E. Modiano, “Minimizing age-of-information in multi-hop wireless networks,” in *55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct 2017, pp. 486–493.

[4] S. Farazi, A. G. Klein, J. A. McNeill, and D. R. Brown, “On the age of information in multi-source multi-hop wireless status update networks,” in *IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2018, pp. 1–5.

[5] A. M. Bedewy, Y. Sun, and N. B. Shroff, “The age of information in multihop networks,” *IEEE/ACM Transactions on Networking*, vol. 27, no. 3, pp. 1248–1257, June 2019.

[6] P. Zou, O. Ozel, and S. Subramaniam, “Waiting before Serving: A Companion to Packet Management in Status Update Systems,” *arXiv e-prints*, p. arXiv:1901.02873, Jan 2019.

[7] B. Buyukates and S. Ulukus, “Timely Distributed Computation with Stragglers,” *arXiv e-prints*, p. arXiv:1910.03564, Oct 2019.

[8] A. Maatouk, M. Assaad, and A. Ephremides, “The age of updates in a simple relay network,” in *IEEE Information Theory Workshop (ITW)*, Nov 2018, pp. 1–5.

[9] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, “Age of information performance of multiaccess strategies with packet management,” *Journal of Communications and Networks*, vol. 21, no. 3, pp. 244–255, June 2019.

[10] M. Costa, M. Codreanu, and A. Ephremides, “Age of information with packet management,” in *IEEE International Symposium on Information Theory*, June 2014, pp. 1583–1587.

[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 Transactions on Information Theory*, vol. 63, no. 11, pp. 7492–7508, Nov 2017.

[12] J. Zhong, R. D. Yates, and E. Soljanin, “Two freshness metrics for local cache refresh,” in *IEEE International Symposium on Information Theory (ISIT)*, June 2018, pp. 1924–1928.

[13] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor, “Age-minimal transmission for energy harvesting sensors with finite batteries: Online policies,” *IEEE Transactions on Information Theory*, vol. 66, no. 1, pp. 534–556, Jan 2020.

[14] Y. Sun, Y. Polyanskiy, and E. Uysal, “Sampling of the wiener process for remote estimation over a channel with random delay,” *IEEE Transactions on Information Theory*, vol. 66, no. 2, pp. 1118–1135, 2020.

[15] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides, “The age of incorrect information: A new performance metric for status updates,” *IEEE/ACM Transactions on Networking*, vol. 28, no. 5, pp. 2215–2228, 2020.

[16] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, “Towards an effective age of information: Remote estimation of a markov source,” in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, April 2018, pp. 367–372.

[17] Y. Sun, E. Uysal-Biyikoglu, and S. Kompella, “Age-optimal updates of multiple information flows,” in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, April 2018, pp. 136–141.

[18] 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.

[19] A. M. Bedewy, Y. Sun, and N. B. Shroff, “Minimizing the age of information through queues,” *IEEE Transactions on Information Theory*, vol. 65, no. 8, pp. 5215–5232, 2019.

[20] A. Maatouk, M. Assaad, and A. Ephremides, “Minimizing the age of information: Noma or oma?” in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, April 2019, pp. 102–108.

[21] A. Maatouk, S. Kriouile, M. Assad, and A. Ephremides, “On the optimality of the whittle’s index policy for minimizing the age of information,” *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 1263–1277, 2021.

[22] B. Zhou and W. Saad, “Minimum age of information in the internet of things with non-uniform status packet sizes,” *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 1933–1947, 2020.

[23] Y.-P. Hsu, E. Modiano, and L. Duan, “Scheduling algorithms for minimizing age of information in wireless broadcast networks with random arrivals,” *IEEE Transactions on Mobile Computing*, vol. 19, no. 12, pp. 2903–2915, 2020.

[24] Z. Jiang, B. Krishnamachari, S. Zhou, and Z. Niu, “Decentralized status update for age-of-information optimization in wireless multiaccess channels,” in *2018 Information Theory and Applications Workshop (ITA)*, 2018, pp. 1–1.

[25] A. Maatouk, M. Assaad, and A. Ephremides, “Minimizing the age of information in a csma environment,” in *2019 International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*, 2019, pp. 1–8.

[26] ——, “On the age of information in a csma environment,” *IEEE/ACM Transactions on Networking*, vol. 28, no. 2, pp. 818–831, 2020.

[27] Y. Sun, I. Kadota, R. Talak, and E. Modiano, “Age of information: A new metric for information freshness,” *Synthesis Lectures on Communication Networks*, vol. 12, no. 2, pp. 1–224, 2019.

[28] G. Karagiannis, O. Altintas, E. Ekici, G. Heijenk, B. Jarupan, K. Lin, and T. Weil, “Vehicular networking: A survey and tutorial on requirements, architectures, challenges, standards and solutions,” *IEEE Communications Surveys Tutorials*, vol. 13, no. 4, pp. 584–616, Fourth 2011.

[29] B. E. Hajek and P. Seri, “Lex-optimal multiclass scheduling with deadlines,” 2000.

[30] S. K. Kaul and R. D. Yates, “Age of information: Updates with priority,” in *2018 IEEE International Symposium on Information Theory (ISIT)*, June 2018, pp. 2644–2648.

[31] A. Maatouk, M. Assaad, and A. Ephremides, “Age of information with prioritized streams: When to buffer preempted packets?” in *2019 IEEE International Symposium on Information Theory (ISIT)*, July 2019, pp. 325–329.

[32] J. Xu and N. Gautam, “Peak age of information in priority queuing systems,” *IEEE Transactions on Information Theory*, vol. 67, no. 1, pp. 373–390, 2021.

[33] M. Shaked and J. G. Shanthikumar, Eds., *Stochastic Orders*. Springer New York, 2007.

[34] A. G. Phadke, B. Pickett, M. Adamiak, M. Begovic, G. Benmouyal, R. O. Burnett, T. W. Cease, J. Goossens, D. J. Hansen, M. Kezunovic, L. L. Mankoff, P. G. McLaren, G. Michel, R. J. Murphy, J. Nordstrom, M. S. Sachdev, H. S. Smith, J. S. Thorp, M. Trotignon, T. C. Wang, and M. A. Xavier, "Synchronized sampling and phasor measurements for relaying and control," *IEEE Transactions on Power Delivery*, vol. 9, no. 1, pp. 442–452, Jan 1994.

[35] F. Sivrikaya and B. Yener, "Time synchronization in sensor networks: a survey," *IEEE Network*, vol. 18, no. 4, pp. 45–50, July 2004.

[36] Y. Sun and B. Cyr, "Sampling for data freshness optimization: Non-linear age functions," *Journal of Communications and Networks*, vol. 21, no. 3, pp. 204–219, 2019.

[37] H. Vahdat-Nejad, A. Ramazani, T. Mohammadi, and W. Mansoor, "A survey on context-aware vehicular network applications," *Vehicular Communications*, vol. 3, pp. 43 – 57, 2016.

Appendix A
Proof of Theorem 1

To establish this theorem, we first provide a set of scheduling rules and prove by induction, and using a sample-path comparison, that they are necessary and sufficient for level k lex-age-optimality for $k = 1, \dots, I$. Afterward, we show that the PP-MAF-LGFS policy satisfies these rules for all $k = 1, \dots, I$, and we can therefore conclude that it is lex-age-optimal. Before proceeding in this direction, we lay out some preliminaries on stochastic ordering that will be useful to our proof.

• **Preliminaries**: Let us consider two scheduling policies $P, \pi \in \Pi$. In general, for any class i, a direct comparison between two processes $\{p_t \circ \mathbf{\Delta}^i_P(t), t \geq 0\}$ and $\{p_t \circ \mathbf{\Delta}^i_\pi(t), t \geq 0\}$ to establish a stochastic ordering between the two is complex, as it involves comparing their probability distributions. To circumvent this difficulty, the following approach can be adopted:

- Define two policies $P_1, \pi_1 \in \Pi$ on the same probability space such that $\{p_t \circ \mathbf{\Delta}^i_{P_1}(t), t \geq 0\}$ and $\{p_t \circ \mathbf{\Delta}^i_P(t), t \geq 0\}$ (respectively $\{p_t \circ \mathbf{\Delta}^i_{\pi_1}(t), t \geq 0\}$ and $\{p_t \circ \mathbf{\Delta}^i_\pi(t), t \geq 0\}$) have the same distribution.
- Proceed with a direct comparison between $\{p_t \circ \mathbf{\Delta}^i_{P_1}(t), t \geq 0\}$ and $\{p_t \circ \mathbf{\Delta}^i_{\pi_1}(t), t \geq 0\}$.

This approach is called *coupling* in the scheduling literature, and we will adopt it in our proof. To that end, and using the memoryless property of the exponential distribution, we can obtain the following coupling lemma.

Lemma 1 ([17, Lemma 1] Stochastic Coupling)**.** *For any given $\mathscr{I}$, consider two* work-conserving *policies $P, \pi \in \Pi_{wc}$. If the service times are exponentially distributed and i.i.d. across streams and time, then the following holds:*

1) *There exists a* work-conserving *policy* P_1 *such that* $\{\mathbf{\Delta}_{P_1}(t), t \geq 0\}$ *and* $\{\mathbf{\Delta}_{P}(t), t \geq 0\}$ *have the same distribution.*
2) *There exists a* work-conserving *policy* π_1 *such that* $\{\mathbf{\Delta}_{\pi_1}(t), t \geq 0\}$ *and* $\{\mathbf{\Delta}_{\pi}(t), t \geq 0\}$ *have the same distribution.*
3) P_1 *and* π_1 *are defined on the same probability space and, if a packet is delivered in policy* π_1 *at time* t*, then with probability* 1*, a packet is delivered in policy* P_1 *at time* t*.*

Next, we present in the following proposition a set of scheduling rules for the first k classes with $k \in \{1, \ldots, I\}$. We show that a policy P is level k lex-age-optimal if, and only if, these rules hold for the first k classes. Note that, throughout this proof, we refer to classes 1 till k as the first k classes. Before laying out the proposition, we define the notion of work-conserving policies for the informative packets of a class k.

Definition 11. *Work-conserving policies for the informative packets of a class* k: A scheduling policy P is said to be work-conserving for the informative packets of a class k if the service facility is kept busy whenever there exist one or more informative packet in the queues of class k.

Proposition 1 (Lex-age-optimal Scheduling Rules)**.** *If (i) the packet generation and arrival times are synchronized across streams within each class, and (ii) the packet service times are exponentially distributed and i.i.d. across streams and time, a scheduling policy* P *is level* k *lex-age-optimal for* $k \in \{1, \ldots, I\}$ *if, and only if, the following four* ***rules*** *are satisfied*

1) *Policy* P *is work-conserving for the informative packets of the first* k *classes;*
2) *Among the streams with informative packets,* P *serves the streams belonging to the first* k *classes first. Among these classes with informative packets, the class of streams with the highest priority are preemptively served first;*
3) *Among the streams of each of the first* k *classes with informative packets, the stream with the maximum age is served first, with ties broken arbitrarily;.*
4) *Among the informative packets from a stream of the first* k *classes, the last generated informative packet is preemptively served first, with ties broken arbitrarily.*

Proof: We prove this proposition by induction. Specifically, we show in step 1 that a policy is level 1 lex-age-optimal if, and only if, Rules 1)-4) hold for class 1. Then, by assuming that they are necessary and sufficient for level k lex-age-optimality, we use this assumption to prove in step 2 that these rules are sufficient and necessary for level $k+1$ lex-age-optimality.

• **Step 1**: We prove in this step that these rules for $k=1$ are sufficient and necessary for level 1 lex-age-optimality.

1) ***Sufficiency***: Let us consider a work-conserving policy $P \in \Pi_{wc}$ that satisfies these rules for class 1. We compare its performance to any *work-conserving* policy $\pi \in \Pi_{wc}$. As both policies are work-conserving, we consider the two policies P_1 and π_1 that are defined on the same probability space and originate from Lemma 1. Next, we provide the following lemma that describes the evolution of the age vector of class 1 upon a packet delivery by both P_1 and π_1.

Lemma 2 (Packet Delivery)**.** *Suppose that a packet is delivered at time t by both policies π_1 and P_1. The age vector changes at time t from $\mathbf{\Delta}_{P_1}$ and $\mathbf{\Delta}_{\pi_1}$ to $\mathbf{\Delta}'_{P_1}$ and $\mathbf{\Delta}'_{\pi_1}$, respectively. If*

$$\Delta_{P_1}^{1,[j]} \leq \Delta_{\pi_1}^{1,[j]}, \quad j = 1, \ldots, J_1, \tag{18}$$

then

$$(\Delta_{P_1}^{1,[j]})' \leq (\Delta_{\pi_1}^{1,[j]})', \quad j = 1, \ldots, J_1, \tag{19}$$

where $\Delta_{P_1}^{1,[j]}$ and $\Delta_{\pi_1}^{1,[j]}$ refers to the j-th largest element of the age vector of class 1 in policy P_1 and π_1, respectively.

Proof: The proof can be found in Appendix B. ■

We can now proceed to prove that P is level 1 lex-age-optimal. To do so, we compare the age vector $\mathbf{\Delta}^1$ on a sample-path of the policies P_1 and π_1. We note that for any sample-path, $\mathbf{\Delta}_{P_1}(0^-) = \mathbf{\Delta}_{\pi_1}(0^-)$. To that end, We consider two cases:

Case 1: When there are no packets deliveries by any of the policies, the age of each stream belonging to class 1 increases at a unit rate.

Case 2: When a packet is delivered by π_1, the evolution of the age vector of class 1 is dictated by Lemma 2. By induction over time, we obtain

$$\Delta_{P_1}^{1,[j]}(t) \leq \Delta_{\pi_1}^{1,[j]}(t), \quad j = 1, \ldots, J_1, \quad t \geq 0. \tag{20}$$

For any symmetric non-decreasing function p_t, and for $t \geq 0$, it holds from (20) that

$$\begin{aligned} & p_t \circ \mathbf{\Delta}^1_{P_1}(t) \\ =& p_t(\Delta^{1,1}_{P_1}(t), \ldots, \Delta^{1,J_1}_{P_1}(t)) \\ =& p_t(\Delta^{1,[1]}_{P_1}(t), \ldots, \Delta^{1,[J_1]}_{P_1}(t)) \\ \leq& p_t(\Delta^{1,[1]}_{\pi_1}(t), \ldots, \Delta^{1,[J_1]}_{\pi_1}(t)) \\ =& p_t(\Delta^{1,1}_{\pi_1}(t), \ldots, \Delta^{1,J_1}_{\pi_1}(t)) \\ =& p_t \circ \mathbf{\Delta}^1_{\pi_1}(t). \end{aligned} \tag{21}$$

By Lemma 1, the processes $\{\mathbf{\Delta}_{P_1}(t), t \geq 0\}$ and $\{\mathbf{\Delta}_{P}(t), t \geq 0\}$ (respectively the processes $\{\mathbf{\Delta}_{\pi_1}(t), t \geq 0\}$ and $\{\mathbf{\Delta}_{\pi}(t), t \geq 0\}$) have the same distribution. Accordingly, using (21) and Theorem 6.B.30 in [33], we can deduce that

$$[\{p_t \circ \mathbf{\Delta}^1_{P}(t), t \geq 0\} | \mathscr{I}] \leq_{st} [\{p_t \circ \mathbf{\Delta}^1_{\pi}(t), t \geq 0\} | \mathscr{I}], \tag{22}$$

for all $\mathscr{I}$, $p_t \in \mathscr{P}_{\text{sym}}$ and $\pi \in \Pi_{wc}$. The extension of (22) to the case where π is non-work-conserving is straightforward due to the exponential distribution of the service time and its independence across streams and time. In fact, due to the memoryless property offered by the exponential distribution, letting the server idle before a transmission will lead to unnecessary staleness of the available packets. This can be shown by a stochastic ordering argument but the details are omitted for the sake of space. Consequently, (22) holds for any $\pi \in \Pi$ and, accordingly, P is level 1 lex-age-optimal.

2) ***Necessity***: In this part, we prove that every level 1 lex-age-optimal policy satisfies these 4 scheduling rules for class 1. We do so by contradiction. Specifically, we consider a level 1 lex-age-optimal policy $\pi \in \Pi^1_{\text{lex-opt}}$. We show that if π violates any of these 4 rules for class 1, then it cannot be level 1 lex-age-optimal.

- **Violation of Rule 1**: Let us consider that π is not work-conserving for the informative packets of class 1. Due to the memoryless property of the exponential distribution of the service time and its independence across streams and time, letting the server idle before a transmission will lead to unnecessary staleness of the available informative packets. This can be shown by a stochastic ordering argument but the details are omitted for the sake of space. Accordingly, π cannot be level 1 lex-age-optimal.

- **Violation of Rule $2-4$**: As shown in the proof of necessity of Rule 1, we can affirm that π

has to be work-conserving for the informative packets of class 1. Note that when there are no informative packets for class 1 in the system, the performance of class 1's streams is not affected by the scheduling rules adopted. Accordingly, and without loss of generality, let us consider that π is *work-conserving*. In other words, we have $\pi \in \Pi_{wc} \cap \Pi^1_{\text{lex-opt}}$. By Definition 5 and (15), we have

$$\begin{aligned} &\mathbb{E}[\phi(\{p_t \circ \mathbf{\Delta}^1_{\pi}(t), t \geq 0\})|\mathscr{I}] \\ =&\min_{\pi' \in \Pi} \mathbb{E}[\phi(\{p_t \circ \mathbf{\Delta}^1_{\pi'}(t), t \geq 0\})|\mathscr{I}], \end{aligned} \tag{23}$$

for all $\mathscr{I}$, $p_t \in \mathscr{P}_{\text{sym}}$ and non-decreasing functional $\phi : \mathbb{V} \mapsto \mathbb{R}$, provided that the expectations in (23) exist. We show by contradiction that if π violates any of the rules $2-4$ for class 1, then there exists a policy P, a symmetrical non-decreasing penalty function p', and a non-decreasing functional ϕ_1 such that

$$\begin{aligned} &\mathbb{E}[\phi_1(\{p' \circ \mathbf{\Delta}^1_{P}(t), t \geq 0\})|\mathscr{I}] \\ <&\mathbb{E}[\phi_1(\{p' \circ \mathbf{\Delta}^1_{\pi}(t), t \geq 0\})|\mathscr{I}]. \end{aligned} \tag{24}$$

To that end, let us consider a work-conserving policy P that satisfies these 4 rules for class 1. Note that P and π are both work-conserving. Accordingly, we consider the two coupled policies P_1 and π_1 that are defined on the same probability space and originate from Lemma 1. From the sufficiency proof, (20) holds for our case. In other words,

$$\Delta^{1,[j]}_{P_1}(t) \leq \Delta^{1,[j]}_{\pi_1}(t), \quad j = 1, \ldots, J_1, \quad t \geq 0. \tag{25}$$

Accordingly, for any symmetrical non-decreasing function $p_t \in \mathscr{P}_{\text{sym}}$, and for $t \geq 0$

$$p_t \circ \mathbf{\Delta}^1_{P_1}(t) \leq p_t \circ \mathbf{\Delta}^1_{\pi_1}(t). \tag{26}$$

Next, let us consider a delivery time t_s such that (i) the age of streams of class 1 are not all equal to one another[2], and (ii) there exist informative packets for $l_1 > 0$ and $l_2 > 0$ streams of class 1 in the system just before t_s for policy π_1 and P_1, respectively. As P_1 follows the 4 rules of the proposition for class 1, we have $l_2 \leq l_1$. We recall that, according to Lemma 1, if a packet is delivered in policy π_1 at time t, then with probability 1, a packet is delivered in policy P_1 at

[2]We avoid this scenario since, in the case where all streams have the same age, all streams of class 1 are considered to have the highest age.

time t. Hence, we describe the evolution of the age vector of class 1 upon a packet delivery by both policies π_1 and P_1 at time t_s.

Lemma 3 (Packet Delivery)**.** *Suppose that a packet is delivered at time t_s by both policies π_1 and P_1. The age vector changes at time t_s from $\mathbf{\Delta}_{P_1}$ and $\mathbf{\Delta}_{\pi_1}$ to $\mathbf{\Delta}'_{P_1}$ and $\mathbf{\Delta}'_{\pi_1}$, respectively. If π_1 breaks any of the scheduling rules $2-4$ for class 1 at time t_s, then there exists a stream j of class 1 such that*

$$(\Delta_{P_1}^{1,[j]})' < (\Delta_{\pi_1}^{1,[j]})'. \tag{27}$$

Proof: The proof can be found in Appendix C. ■

Next, to prove (24), let us consider the symmetrical non-decreasing penalty function $p' = p_{\text{sum}} \in \mathscr{P}_{\text{sym}}$ and the non-decreasing age penalty functional $\phi_1 = \phi_{avg}$. By taking Lemma 3 into account, along with (25), and the fact that the service rate μ is finite, we can affirm that there exists a time interval $\mathcal{T} \subseteq [0, \infty)$ such that

$$p' \circ \mathbf{\Delta}_{P_1}^1(t) < p' \circ \mathbf{\Delta}_{\pi_1}^1(t) \quad \forall t \in \mathcal{T}. \tag{28}$$

By Lemma 1, we have that the processes $\{\mathbf{\Delta}_{P_1}(t), t \geq 0\}$ and $\{\mathbf{\Delta}_{P}(t), t \geq 0\}$ (respectively the processes $\{\mathbf{\Delta}_{\pi_1}(t), t \geq 0\}$ and $\{\mathbf{\Delta}_{\pi}(t), t \geq 0\}$) have the same distribution. By taking this into account, and by using (26) and (28), we obtain:

$$\mathbb{E}[\phi_1(\{p' \circ \mathbf{\Delta}_P^1(t), t \geq 0\} | \mathscr{I})] < \mathbb{E}[\phi_1(\{p' \circ \mathbf{\Delta}_\pi^1(t), t \geq 0\} | \mathscr{I})] \tag{29}$$

Therefore, π is not level 1 lex-age-optimal if it breaks any of the 4 scheduling rules of the proposition for class 1.

This concludes our proof that this set of rules for class 1 are sufficient and necessary to have level 1 lex-age-optimality.

•**Step 2**: Next, we will prove the induction step: Assume that this set of rules for the first k classes are necessary and sufficient for level k lex-age-optimality. In other words, every policy $\pi \in \Pi_{\text{lex-opt}}^k$ follows these scheduling rules for the first k classes. Our goal is to use this assumption to prove that a policy P is level $k+1$ lex-age-optimal if, and only if, it follows these rules for the first $k+1$ classes.

1) ***Sufficiency***: Let us consider a work-conserving policy P that satisfies the depicted set of rules for the first $k+1$ classes. We compare its performance to any *work-conserving* policy

$\pi \in \Pi_{wc} \cap \Pi^k_{\text{lex-opt}}$. As both policies are work-conserving, we consider the two policies P_1 and π_1 that are defined on the same probability space and originate from Lemma 1. Next, we provide the following Lemma that describes the evolution of the age vector of classes $i = 1, \dots, k+1$ upon a packet delivery by both π_1 and P_1.

Lemma 4 (Packet Delivery)**.** *Suppose that a packet is delivered at time t by both policies π_1 and P_1. The age vector changes at time t from $\mathbf{\Delta}_{P_1}$ and $\mathbf{\Delta}_{\pi_1}$ to $\mathbf{\Delta}'_{P_1}$ and $\mathbf{\Delta}'_{\pi_1}$, respectively. If*

$$\Delta^{i,[j]}_{P_1} = \Delta^{i,[j]}_{\pi_1}, \quad i = 1, \dots, k, \;\; j = 1, \dots, J_i, \tag{30}$$

$$\Delta^{k+1,[j]}_{P_1} \le \Delta^{k+1,[j]}_{\pi_1}, \quad j = 1, \dots, J_{k+1}, \tag{31}$$

then

$$(\Delta^{i,[j]}_{P_1})' = (\Delta^{i,[j]}_{\pi_1})', \quad i = 1, \dots, k, \;\; j = 1, \dots, J_i, \tag{32}$$

$$(\Delta^{k+1,[j]}_{P_1})' \le (\Delta^{k+1,[j]}_{\pi_1})', \quad j = 1, \dots, J_{k+1}. \tag{33}$$

Proof: The proof can be found in Appendix D. ■

We can now show that P is level $k+1$ lex-age-optimal. To do so, we compare the age vector $\mathbf{\Delta}^i$ for $i = 1, \dots, k+1$ on a sample-path of the policies P_1 and π_1. We note that for any sample-path, $\mathbf{\Delta}_{P_1}(0^-) = \mathbf{\Delta}_{\pi_1}(0^-)$. To that end, we consider two cases:

Case 1: When there are no packets deliveries by any of the policies, the age of each stream of the first $k+1$ classes increases at a unit rate.

Case 2: When a packet is delivered by π_1, the evolution of the age vector of the first $k+1$ classes is dictated by Lemma 4. By induction over time, we obtain for all $t \ge 0$:

$$\Delta^{i,[j]}_{P_1}(t) = \Delta^{i,[j]}_{\pi_1}(t), \quad i = 1, \dots, k, \;\; j = 1, \dots, J_i, \tag{34}$$

$$\Delta^{k+1,[j]}_{P_1}(t) \le \Delta^{k+1,[j]}_{\pi_1}(t), \quad j = 1, \dots, J_{k+1}. \tag{35}$$

For any symmetric non-decreasing function p_t, and for $t \ge 0$, it holds from (34) and (35)

$$p_t \circ \mathbf{\Delta}^i_{P_1}(t) = p_t \circ \mathbf{\Delta}^i_{\pi_1}(t), \quad i = 1, \dots, k, \quad t \ge 0, \tag{36}$$

$$p_t \circ \mathbf{\Delta}^{k+1}_{P_1}(t) \le p_t \circ \mathbf{\Delta}^{k+1}_{\pi_1}(t), \quad t \ge 0. \tag{37}$$

By Lemma 1, the processes $\{\boldsymbol{\Delta}_{P_1}(t), t \geq 0\}$ and $\{\boldsymbol{\Delta}_{P}(t), t \geq 0\}$ (respectively the processes $\{\boldsymbol{\Delta}_{\pi_1}(t), t \geq 0\}$ and $\{\boldsymbol{\Delta}_{\pi}(t), t \geq 0\}$) have the same distribution. Accordingly, using (36)-(37) and Theorem 6.B.30 in [33], we can deduce that

$$\begin{aligned}&[\{p_t \circ \boldsymbol{\Delta}^i_P(t), t \geq 0\}|\mathscr{I}]\\ =_{st}&[\{p_t \circ \boldsymbol{\Delta}^i_\pi(t), t \geq 0\}|\mathscr{I}], \quad i = 1, \ldots, k,\end{aligned} \tag{38}$$

and

$$\begin{aligned}&[\{p_t \circ \boldsymbol{\Delta}^{k+1}_P(t), t \geq 0\}|\mathscr{I}]\\ \leq_{st}&[\{p_t \circ \boldsymbol{\Delta}^{k+1}_\pi(t), t \geq 0\}|\mathscr{I}],\end{aligned} \tag{39}$$

for all $\mathscr{I}$, $p_t \in \mathscr{P}_{\text{sym}}$ and $\pi \in \Pi_{wc} \cap \Pi^k_{\text{lex-opt}}$. The extension of (38)-(39) to the case where $\pi \in \Pi^k_{\text{lex-opt}}$ but is not necessarily work-conserving is straightforward due to the exponential distribution of the service time and its independence across streams and time. As it was previously explained, due to the memoryless property offered by the exponential distribution, letting the server to idle before a transmission will lead to unnecessary staleness of the packets. This can be shown by a stochastic ordering argument but the details are omitted for the sake of space. Consequently, (38)-(39) hold for any $\pi \in \Pi^k_{\text{lex-opt}}$ and, therefore, P is level $k+1$ lex-age-optimal.

2) ***Necessity***: In this part, we leverage our inductive assumption for level k lex-age-optimality and prove that every level $k+1$ lex-age-optimal policy follow these 4 scheduling rules for the first $k+1$ classes. We prove this by contradiction. Specifically, let us consider a level k lex-age-optimal policy $\pi \in \Pi^k_{\text{lex-opt}}$. We know by our inductive assumption that π has to follow this set of rules for the first k classes. We show that if π violates any of the 4 rules for class $k+1$, then it cannot be level $k+1$ lex-age-optimal.

- **Violation of Rule 1**: Let us consider that π is not work-conserving for the informative packets of class $k+1$. Due to the memoryless property of the exponential distribution of the service time and its independence across streams and time, letting the server idle before a transmission will lead to unnecessary staleness of the available packets. This can be shown by a stochastic ordering argument but the details are omitted for the sake of space. Accordingly, π cannot be level $k+1$ lex-age-optimal.

- **Violation of Rule $2-4$**: The proof follows the same line of work done in the necessity proof of Step 1. Specifically, and as it was previously explained, we can consider that $\pi \in \Pi_{wc} \cap \Pi^k_{\text{lex-opt}}$. Next, we consider a work-conserving policy P that satisfies the 4 scheduling rules for the first

$k+1$ classes. Note that P and π are both work-conserving. Accordingly, we consider the two coupled policies P_1 and π_1 that are defined on the same probability space and originate from Lemma 1. From the sufficiency proof for level $k+1$ lex-age-optimality, we have that for all $t \geq 0$:

$$\Delta_{P_1}^{i,[j]}(t) = \Delta_{\pi_1}^{i,[j]}(t), \quad i = 1, \ldots, k, \;\; j = 1, \ldots, J_i, \tag{40}$$

$$\Delta_{P_1}^{k+1,[j]}(t) \leq \Delta_{\pi_1}^{k+1,[j]}(t), \quad j = 1, \ldots, J_{k+1}. \tag{41}$$

Accordingly, for any symmetric non-decreasing function p_t:

$$p_t \circ \boldsymbol{\Delta}_{P_1}^{i}(t) = p_t \circ \boldsymbol{\Delta}_{\pi_1}^{i}(t), \quad i = 1, \ldots, k \quad t \geq 0, \tag{42}$$

$$p_t \circ \boldsymbol{\Delta}_{P_1}^{k+1}(t) \leq p_t \circ \boldsymbol{\Delta}_{\pi_1}^{k+1}(t), \quad t \geq 0. \tag{43}$$

Next, as per our inductive assumption, we have that π_1 and P_1 follow the same scheduling discipline for the first k classes. Accordingly, the streams of the first k classes will have no informative updates at the same time in both policies π_1 and P_1. This allows us to consider a delivery time t_s such that (i) there are no informative packets for the first k classes, (ii) the age of streams of class $k+1$ are not all equal to one another, and (iii) there exist informative packets for $l_1 > 0$ and $l_2 > 0$ streams of class $k+1$ in the system just before t_s for policy π_1 and P_1, respectively. As P_1 follows the 4 scheduling rules of the proposition for the first $k+1$ classes, we have $l_2 \leq l_1$. By proceeding similarly to Lemma 3, we can show that if π_1 breaks any of the scheduling rules $2-4$ for class $k+1$ at time t_s, then there exists a stream j of class $k+1$ such that

$$\Delta_{P_1}^{k+1,[j]}(t_s^+) < \Delta_{\pi_1}^{k+1,[j]}(t_s^+). \tag{44}$$

Afterward, we consider the symmetric non-decreasing penalty function $p' = p_{\text{sum}} \in \mathscr{P}_{\text{sym}}$ and the non-decreasing age penalty functional $\phi_1 = \phi_{avg}$. By taking (44) into account, along with (40)-(41), and the fact that the service rate μ is finite, we can affirm that there exists a time interval $\mathcal{T} \subseteq [0, \infty)$ such that

$$p' \circ \boldsymbol{\Delta}_{P_1}^{k+1}(t) < p' \circ \boldsymbol{\Delta}_{\pi_1}^{k+1}(t), \quad \forall t \in \mathcal{T}. \tag{45}$$

By Lemma 1, we have that the processes $\{\mathbf{\Delta}_{P_1}(t), t \geq 0\}$ and $\{\mathbf{\Delta}_{P}(t), t \geq 0\}$ (respectively the processes $\{\mathbf{\Delta}_{\pi_1}(t), t \geq 0\}$ and $\{\mathbf{\Delta}_{\pi}(t), t \geq 0\}$) have the same distribution. By taking this into consideration, and by using (42), (43), and (45), we obtain:

$$
\begin{aligned}
&[\{p_t \circ \mathbf{\Delta}^i_P(t), t \geq 0\}|\mathscr{I}] \\
=_{st}&[\{p_t \circ \mathbf{\Delta}^i_\pi(t), t \geq 0\}|\mathscr{I}], \quad i = 1, \ldots, k,
\end{aligned} \tag{46}
$$

and

$$
\begin{aligned}
&\mathbb{E}[\phi_1(\{p' \circ \mathbf{\Delta}^{k+1}_P(t), t \geq 0\}|\mathscr{I})] \\
<&\mathbb{E}[\phi_1(\{p' \circ \mathbf{\Delta}^{k+1}_\pi(t), t \geq 0\}|\mathscr{I})].
\end{aligned} \tag{47}
$$

Therefore, π is not level $k+1$ lex-age-optimal if it breaks any of the 4 scheduling rules of the proposition for class $k+1$. ■

By Definition 10, the PP-MAF-LGFS policy is the only policy that satisfies the scheduling rules depicted in this proposition for the first k classes simultaneously for any $k = 1, \ldots, I$. Accordingly, the PP-MAF-LGFS policy is lex-age-optimal, which concludes the proof of the theorem.

APPENDIX B
PROOF OF LEMMA 2

Let us denote by $W^1_j(t) = \max\{S^{1,j}_n : A^{1,j}_n \leq t\}$ the time-stamp of the freshest packet that has arrived to the queue of stream j of class 1 at time t. Since the generation/arrival sequences are synchronized across streams within each class, there exists a $W^1(t)$ such that $W^1_j(t) = W^1(t)$ for $j = 1, \ldots, J_1$. We distinguish between three cases that can happen at time t. The proof of Case 3 is adopted from the proof of Lemma 2 of [17]. For the sake of completeness, we provide a proof of all 3 cases.

1) Case 1: There was no transmission of packets for class 1 by policy P_1, or a non-informative packet of class 1 has just finished transmission. In other words, prior to time t, policy P_1 has already finished the transmission of all class 1's informative packets. To that end:

$$(\Delta^{1,[j]}_{P_1})' = \Delta^{1,[j]}_{P_1} = t - W^1(t), \quad j = 1, \ldots, J_1. \tag{48}$$

On the other hand, in policy π_1, the delivered packet can be any packet from any information stream. Consequently, we can conclude:

$$\Delta^{1,[j]}_{\pi_1} \geq (\Delta^{1,[j]}_{\pi_1})' \geq t - W^1(t), \quad j = 1, \ldots, J_1. \tag{49}$$

Therefore, (19) holds for this case.

2) Case 2: An informative packet belonging to a stream of class 1 finishes transmission by policy P_1 at time t. On the other hand, policy π_1 delivers a non-informative packet of class 1 or a packet belonging to one of the $I-1$ remaining classes at time t. Consequently, $(\boldsymbol{\Delta}_{\pi_1}^1)' = \boldsymbol{\Delta}_{\pi_1}^1$ and (19) holds trivially in this scenario.

3) Case 3: An informative packet belonging to a stream of class 1 finishes transmission by both policies P_1 and π_1 at time t. By definition, the following always holds:

$$\Delta_{P_1}^{1,j} \geq (\Delta_{P_1}^{1,j})' \geq t - W^1(t), \quad j = 1, \ldots, J_1, \tag{50}$$

$$\Delta_{\pi_1}^{1,j} \geq (\Delta_{\pi_1}^{1,j})' \geq t - W^1(t), \quad j = 1, \ldots, J_1. \tag{51}$$

We recall that P_1 schedules the stream of class 1 with the highest age. Consequently, the stream of class 1 having the age $\Delta_{P_1}^{1,[1]}$ is the one that finishes transmission at time t by P_1. Since the transmitted packet has $W^1(t)$ as time-stamp, the age of this stream becomes the smallest among the streams of class 1. To that end,

$$(\Delta_{P_1}^{1,[J_1]})' = t - W^1(t). \tag{52}$$

As there is only one server, the age of the remaining $J_1 - 1$ streams of class 1 stay the same. By taking this into account, along with (52), we get:

$$(\Delta_{P_1}^{1,[j]})' = \Delta_{P_1}^{1,[j+1]}, \quad j = 1, \ldots, J_1 - 1. \tag{53}$$

On the other hand, since the packet delivered by π_1 can belong to any stream of class 1, the following always holds:

$$(\Delta_{\pi_1}^{1,[j]})' \geq \Delta_{\pi_1}^{1,[j+1]}, \quad j = 1, \ldots, J_1 - 1. \tag{54}$$

Combining (18), (53) and (54), we obtain:

$$(\Delta_{\pi_1}^{1,[j]})' \geq \Delta_{\pi_1}^{1,[j+1]} \geq \Delta_{P_1}^{1,[j+1]} = (\Delta_{P_1}^{1,[j]})', \;\; j = 1, \ldots, J_1 - 1. \tag{55}$$

Also, using (51) and (52), we can deduce that $(\Delta_{\pi_1}^{1,[J_i]})' \geq t - W^1(t) = (\Delta_{P_1}^{1,[J_i]})'$ which concludes the proof.

APPENDIX C
PROOF OF LEMMA 3

To prove this lemma, we recall that (25) always holds from our sufficiency results on P. Next, we distinguish between 3 cases.

Case 1: Suppose that π_1 breaks Rule 2 and delivers at time t_s a packet that does not belong to class 1. We know that P_1 will deliver at time t_s an informative packet for one of the l_2 streams belonging to class 1. Consequently, (27) holds trivially in this case.

Case 2: Suppose that π_1 delivers a packet from class 1. However, at time t_s, π_1 breaks Rule 3 for class 1 and delivers a packet that does not belong to the stream of class 1 with the highest age. To tackle this case, we define the rank of a stream within a class.

Definition 12. *Rank of a stream*: The rank of a stream (i, j) within the class i is defined as its position in the ordered age vector $[\mathbf{\Delta}^i]$. In other words, if stream (i, j) has a rank $1 \leq r \leq J_i$, then:

- There exist $J_i - r$ streams in the same class having an age that is smaller or equal to $\Delta^{i,j}$.
- There exist $r - 1$ streams in the same class having an age that is larger or equal to $\Delta^{i,j}$.

We know that P_1 delivers the freshest packet from the stream of class 1 with the highest age at time t_s (i.e., the stream with rank 1). Therefore, after delivery, the served stream will have the smallest age among all streams of class 1. Moreover, the age of the remaining $J_1 - 1$ streams of class 1 is not altered at the delivery time. Accordingly, these $J_1 - 1$ streams gain a single rank in the sorted age vector $[\mathbf{\Delta}^1_{P_1}]$. On the other hand, let us suppose that the served stream by π_1 has a rank $r > 1$ in the sorted age vector $[\mathbf{\Delta}^1_{\pi_1}]$. After being served, this stream will have a rank $r' \leq r$. Consequently, $r' - r$ streams will gain a rank at time t_s and the rank of all the remaining streams stays the same. Therefore, we can assert that (27) holds. We provide in the following an example to showcase this. Suppose that the ordered age vector of class 1 just before t_s is:

$$[\mathbf{\Delta}^1_{\pi_1}](t_s^-) = (10, 9, 8, 1),$$

$$[\mathbf{\Delta}^1_{P_1}](t_s^-) = (10, 9, 8, 1). \tag{56}$$

Suppose that the age of the available informative packets of class 1 is equal to 1 at time t_s. If we consider that π_1 delivers a packet from stream $(1, [3])$, and knowing that P_1 will deliver a packet from stream $(1, [1])$, we get:

$$[\mathbf{\Delta}^1_{\pi_1'}](t_s^+) = (10, 9, 1, 1)$$

$$[\boldsymbol{\Delta}^1_{\pi_1}](t_s^+) = (9, 8, 1, 1) \tag{57}$$

Accordingly, we can easily see that $j = 1$ or $j = 2$.

Case 3: Suppose that π_1 delivers a packet from the stream of class 1 with the highest age at time t_s. However, suppose that π_1 breaks Rule 4 for class 1 and does not deliver the freshest available informative packet. Accordingly, at time t_s, the served stream by P_1 will have a strictly smaller age when compared to the stream served by π_1. Consequently, (27) holds.

APPENDIX D
PROOF OF LEMMA 4

We proceed with our proof by distinguishing between two possible scenarios at time t:

- *The served packet by π_1 is an informative packet belonging to any of the first k classes*: We recall that, as per our inductive assumption till level k, policy π_1 and P_1 follow the same set of scheduling rules for the first k classes. Accordingly, when an informative packet from one of these classes is delivered by π_1, the same packet (or an informative packet of another stream of the same class that has the same age) is delivered by P_1. Consequently, we can affirm the validity of (32). Moreover, as the age vector of class $k+1$ remains unchanged for both policies in this case, (33) holds naturally.
- *The served packet by π_1 is not an informative packet belonging to the k first classes*: As π_1 and P_1 follow the same set of scheduling rules for the first k classes, this case can only occur when the buffers of streams belonging to the first k classes are either empty or contain non-informative packets for *both* policies. Therefore, (32) holds naturally. Next, to obtain (33), we can proceed similarly to Lemma 2 for class $k+1$. The details are therefore omitted for the sake of space.